58. Jahrgang, Nr. 8 – November/Dezember 2007 – NADOAW 58 (8) 385-464 – ISSN 1434-4653 – D 11721

Information Wissenschaft & Praxis

8/2007

Intellektuelles Indexieren

Intuitive Registererstellung

Übersetzung von Registern

Revision von Registern

Software-Unterstützung beim Indexieren

Aus- und Weiterbildungskurse Indexieren

Software zur Registererstellung

Indexieren von Websites

Aktuelle Normung für kontrollierte Vokabulare

Call for Papers Oberhof 2008

30 Jahre FIZ Karlsruhe

Jahresregister 2007

iwp Herausgeber: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V. DGI

DABIS.eu

Gesellschaft für Datenbank-InformationsSysteme mbH

Ihr Partner für Archiv-, Bibliotheks- und DokumentationsSysteme

BIS-C 2000

Archiv- und Bibliotheks-InformationsSystem

DABIS.eu · alle Aufgaben · ein Team

Synergien: Qualität und Kompetenz
Software: Innovation und Optimierung
Web · SSL · Warenkorb und Benutzeraccount
Lokalsystem zu Aleph-Verbünden

Software · State of the art · Open Source

Leistung	Sicherheit
Standards	Offenheit
Stabilität	Verläßlichkeit
Generierung	Adaptierung
Service	Erfahrenheit
Outsourcing	Support
Dienstleistungen	Zufriedenheit

GUI · Web · Wap · XML · Z 39.50

Archiv	System	Bibliothek
singleUser		multiUser
Lokalsystem		Verbund
multiDatenbank		multiServer
multiProcessing		multiThreading
skalierbar		stufenlos
Unicode		multiLingual
Normdaten		redundanzfrei
multiMedia		Integration

DABIS.com
Heiligenstädter Straße 213
1190 · Wien, Austria
Tel.: +43-1-318 9 777-10
Fax: +43-1-318 9 777-15
eMail: office@dabis.com
http://www.dabis.com

DABIS.de
Herrgasse 24
79294 · Sölden/Freiburg, Germany
Tel.: +49-761-40983-21
Fax: +49-761-40983-29
eMail: office@dabis.de
http://www.dabis.de

Zweigstellen: 61350 · Bad Homburg vdH, Germany / 1147 · Budapest, Hungary / 39042 · Brixen, Italy

Editorial

In Zeiten, in denen das semantische Web, Ontologien und automatische Indexierung oft im Vordergrund des Interesses stehen, ist es ratsam, seine ursprünglichen Wurzeln nicht aus den Augen zu verlieren. Dieses IWP-Heft beleuchtet mit dem intellektuellen Indexieren einen klassischen Kernbereich in der Informationswissenschaft. Dabei bezieht sich dieses Schwerpunktheft im Wesentlichen, aber nicht darauf beschränkt, auf die Indexierung der Inhalte geschlossener Dokumente (wie z. B. Fachbücher) – in Anlehnung an die Unterscheidung in Susan Klements Artikel „Open-system versus closed-system indexing: A vital distinction" in *The Indexer* (Vol. 23, No. 1, April 2002, S. 23–31).

Das Schwerpunktheft zielt darauf, die vielfältigen Aspekte der Registererstellung als eigenständige Disziplin und praktisches Handwerk aufzuzeigen – einschließlich der Erläuterung häufig auftretender Fehleinschätzungen. Dabei bietet es sich auch als Brückenschlag zwischen der Informationswissenschaft und dem deutschsprachigen Publikationswesen an. Letzteres leidet nach Beobachtungen des Deutschen Netzwerks der Indexer (DNI) an eklatanten Defiziten bei der Erstellung adäquater Register für Fach- und Sachbücher – sowohl was die methodische Vorgehensweise beim Indexieren als auch die äußere Form von Registern angeht.

Da die anglo-amerikanische Indexing-Szene die derzeit weltweit führende ist – seien es Fachverbände, Fachliteratur oder Software, war es dem Gastherausgeber wichtig, Autoren – viele davon mit langjährigen Erfahrungen – aus den dortigen Fachverbänden zu gewinnen, weshalb auch die meisten Artikel in englischer Sprache vorliegen. Die Beiträge kommen von Mitgliedern folgender Indexing-Fachverbände: Society of Indexers (SI), American Society of Indexers (ASI), Australian and New Zealand Society of Indexers (ANZSI), Nederlands Indexers Netwerk (NIN) sowie DNI.

Der erste Block gibt Einblicke in grundsätzliche Bereiche des Indexing. Jutta Bertram beschreibt die Herausforderungen der Registererstellung zu ihrem eigenen Buch *Einführung in die inhaltliche Erschließung*. Danach erläutert Pierke Bosschieter anhand eines vor kurzem durchgeführten Projektes die Probleme der Registererstellung bei übersetzten Büchern, inklusive der häufig naiven Annahme, Register könnten auf jeden Fall 1:1 übersetzt werden. Ähnlich gelagert ist das Thema der Index-Erstellung bzw. -Wiederverwendung für neu aufgelegte Werke, das Peter Rooney ebenfalls anhand eines eigenen Projektes beschreibt. Die Vor- und Nachteile des automatischen und intellektuellen Indexierens und deren mögliche Integrierung diskutiert Seth Maislin.

Es folgt ein Artikel von Maureen MacGlashan über die Geschichte, Gegenwart und Zukunft der seit einem halben Jahrhundert erscheinenden Fachzeitschrift *The Indexer* – dem Flaggschiff der Indexing-Fachliteratur – die ab 2008 auch online erhältlich ist.

Der nächste Block besteht aus zwei Beiträgen über drei populäre Indexing-Kurse, die seit den 1980er Jahren bestehen. Ann Hudson erläutert die Konzeption des Lehrgangs der Society of Indexers, und Kari Kells beschreibt die beiden über das U.S. Department of Agriculture (USDA) angebotenen Kurse.

Caroline Diepeveen, Michael Robertson und Jochen Fassbender stellen die drei von Indexierern weltweit am meisten benutzten Indexierungs-Programme vor (so genannte Dedicated Indexing-Programme) und beschreiben die immensen Vorteile dieser Art von Software. Vorgestellt werden zudem zwei Programme zur Indexierung von Websites.

Ein aus zwei Beiträgen bestehender Block behandelt exemplarisch die Registererstellung in Fachgebieten. Janyne Ste Marie gibt Einblicke ins Indexieren von medizinischen Fachbüchern, und Richard Evans erläutert umfassend die Aspekte der Registererstellung von Computer- und Software-Büchern.

In zwei Artikeln stellen die auf die Index-Erstellung von Websites spezialisierten Heather Hedden und Glenda Browne sowohl grundsätzliche Aspekte als auch den neuesten Stand dieser speziellen Form des Indexierens dar, die im deutschsprachigen Raum bisher kaum wahrgenommen wurde.

Ein auch für das Indexieren interessanter Extra-Block schließt sich mit zwei Beiträgen über die neuesten Entwicklungen von Normen im Bereich kontrollierter Vokabulare an, den Nachfolgern der alten Thesaurus-Normen. Einen Ausblick auf die neue ISO 25964 gibt Stella Dextre Clarke. Emily Fayen erläutert die Überarbeitung der ANSI/NISO-Norm Z39.19.

Mit den für Register einschlägigen Aspekten der in „Informationstheorie" umbenannten „Mathematical Theory of Communication" von Shannon und Weaver setzt sich Robert Fugmann in Teil 1 seines Artikels kritisch auseinander und beleuchtet darin grundsätzliche Wissensdefizite, die häufig außerhalb der Informationswissenschaft anzutreffen sind, wie z. B. Unkenntnis über die Unterschiede von Begriff und Wort oder Erinnerungs- und Entdeckungsrecherche.

Der Gastherausgeber dankt besonders allen Autoren und Korrekturlesern, die weltweit beim Zustandekommen dieses IWP-Schwerpunkthefts halfen. Ein spezieller Dank geht auch an die DGI und die IWP, die dieses Schwerpunktheft ermöglichten.

Jochen Fassbender (DNI-Koordinator)

In the age of the semantic web, ontologies, and automatic indexing, it is a good idea not to lose track of one's original roots. This issue of IWP therefore focuses on a different topic and highlights one of the classic areas of information science: human indexing. The contributions in this themed issue mainly refer to, but are not restricted to, closed-system documents (e.g. specialised books), using the distinction established by Susan Klement in her article 'Open-system versus closed-system indexing: A vital distinction' in *The Indexer* 23 (1), April 2002, pp. 23–31.

This special issue aims to present the various aspects of indexing as an independent discipline and craft, including the explanation of frequent misconceptions. At the same time, the issue can be regarded as bridging the gap between information science and the German publishing industry. In the view of the German Network of Indexers (DNI), publishers in Germany suffer from serious deficiencies in relation to the compilation of adequate back-of-the-book indexes, both with regard to indexing procedure and the presentation of indexes.

As the Anglo-American indexing world, with its professional societies, literature and software, has a leading role in this field, the guest editor has looked for contributors from the societies concerned, many of them long-standing practitioners of indexing. This is also the reason why most of the articles are in English. These contributions are by members of the following organisations: Society of Indexers (SI), American Society of Indexers (ASI), Australian and New Zealand Society of Indexers (ANZSI), Nederlands Indexers Netwerk (NIN) and DNI.

The first four articles deal with some basic aspects of indexing. Jutta Bertram describes the challenges of indexing her own book *Einführung in die inhaltliche Erschliessung*. Pierke Bosschieter explains the problems she faced in one of her recent projects when indexing a translated book, including the commonly encountered assumption that indexes can simply be translated verbatim. Related to this is the creation of new indexes or reuse of old indexes for new editions, as described by Peter Rooney on the basis of one of his own projects. The pros and cons of automatic vs. human indexing, and ways of integrating the two approaches, are discussed by Seth Maislin.

This is followed by an article by Maureen MacGlashan about the past, present and future of the journal *The Indexer*, the flagship of literature on indexing, which has been published continuously for almost half a century now. It will be available online from 2008.

The next two articles are about three popular indexing courses which have existed since the 1980s. Ann Hudson explains the structure of the Society of Indexers' training course, and Kari Kells introduces two courses available through the U.S. Department of Agriculture (USDA).

Caroline Diepeveen, Michael Robertson and Jochen Fassbender introduce the three dedicated indexing programs that are most commonly used by indexers worldwide and describe the immense advantages of this kind of software. Two website indexing programs are also presented.

Two articles present examples illustrating the indexing of subject specialities. Janyne Ste Marie provides an insight into the indexing of medical books, and Richard Evans describes in detail the challenges of indexing computer books.

Website indexing, which is hardly known in German-speaking countries, is discussed by Heather Hedden and Glenda Browne; the former introduces basic aspects of website indexing, while the latter describes recent developments in this indexing speciality.

There are two interesting contributions about the latest developments in standards covering controlled vocabularies, the successors to the old thesaurus standards. Stella Dextre Clarke offers an overview of the forthcoming ISO 25964, while Emily Fayen explains the revision of ANSI/NISO Z39.19.

In part 1 of his article, Robert Fugmann analyses the indexing-relevant aspects of Shannon and Weaver's 'Mathematical Theory of Communication', later renamed 'information theory'. He highlights the lack of knowledge common among people from outside the information science field – for example, with regard to the differences between concept and term, or known-item vs. unknown-item search.

The guest editor would like to offer sincere thanks to all of the contributors and proofreaders worldwide who have helped to put this themed issue of IWP together. Special thanks go to the German Society for Information Science and Practice (DGI) and the IWP journal for making this themed issue possible.

Jochen Fassbender (DNI coordinator)

Neue Plattform Web-Information-Retrieval.de

Web-Information-Retrieval.de ist eine offene, nicht-kommerzielle Plattform zu den Themenfeldern Internetsuchdienste, Suchmaschinenmarketing und Internetrecherche. Sie wurde am 11.Oktober 2007 im Rahmen der 59. DGI-Jahrestagung „Information in Wissenschaft, Bildung und Wirtschaft – IWBW" erstmals einem breiteren Publikum vorgestellt.

Alle Interessierten aus dem Bereich Informationswissenschaft und verwandten Themengebieten sind herzlich eingeladen, sich inhaltlich und konzeptuell einzubringen. Beispielsweise kann die Website genutzt werden, um unter „Mitteilungen" auf eigene Arbeiten, Veranstaltungen u.Ä. hinzuweisen.

Ziel von Web-Information-Retrieval.de ist es, einen Beitrag zur Transparenz im Suchdienste- und Retrievalbereich und im Suchmaschinenmarketing für den deutschsprachigen Raum zu schaffen, indem Mitteilungen, Artikel und Weblinks zu den genannten Themenfeldern angeboten werden. Das Projekt steht am Anfang der Entwicklung, kann aber schon derzeit genutzt werden. U.a. werden im Webkatalog wöchentlich mehrere suchrelevante Weblinks eingetragen. Mitteilungen und Artikel sind über RSS abonnierbar. Web-Information-Retrieval.de ist sowohl konzeptionell als auch bzgl. der zugrunde liegenden Software und Funktionalität noch weitgehend offen.

Kontakt: Dr. Joachim Griesbaum, Informationswissenschaft, Universität Konstanz, joachim.griesbaum@web.de, www.inf-wiss.uni-konstanz.de/People/jg.html

Fehlende Information kommt teuer zu stehen...

... deshalb jetzt mit GENIOS sparen! Profitieren auch Sie von dem größten Anbieter seriöser, deutschsprachiger Wirtschaftsinformationen. Wir bieten Ihnen zum Beispiel das Wissen von ● 800 Datenbanken ● 180 Pressearchiven und ● über 420 Fachzeitschriften, außerdem ● 60 Millionen Firmeninformationen ● 36 Millionen Personeninformationen sowie die ● Originaldaten des Bundesanzeigers.

Testen Sie jetzt GENIOS firmenweit für nur € 500,– pro Monat 3 Monate lang!*

Mehr dazu unter
www.genios.de → Einführungsaktion!

GENIOS
German Business Information

GBI-Genios Deutsche Wirtschaftsdatenbank GmbH
Ein Unternehmen der Frankfurter Allgemeine Zeitung GmbH
und der Verlagsgruppe Handelsblatt GmbH

* Preis zuzüglich MwSt. Die Aktion gilt nur für neue Firmenkunden und ist zeitlich begrenzt.

Inhalt

8/2007

SCHWERPUNKTTHEMA INTELLEKTUELLES INDEXIEREN

Gastherausgeber: Jochen Fassbender

385 **EDITORIAL**
Jochen *Fassbender*

NACHRICHTEN

386 Neue Plattform Web-Information-Retrieval.de

458 140 Jahre Chemie-Forschung vollständig digitalisiert

INTELLEKTUELLES INDEXIEREN

389 Jutta *Bertram*
I would do it again. Erfahrungen mit der intuitiven Registererstellung

391 Pierke *Bosschieter*
Translate the index or index the translation?

394 Peter *Rooney*
How I reused my own index

399 Seth *Maislin*
Cyborg indexing: Half-human, half-machine

402 Maureen *MacGlashan*
The Indexer: past, present and future

407 Ann *Hudson*
Training in indexing: The Society of Indexers' course

410 Kari *Kells*
Indexing classes offered by the Graduate School (USDA)

413 Caroline *Diepeveen,* Jochen *Fassbender*, Michael *Robertson*
Indexing software

421 Janyne *Ste Marie*
Medical indexing in the United States

425 Richard *Evans*
Indexing computer books: Getting started

433 Heather *Hedden*
Book-style indexes for websites

437 Glenda *Browne*
Changes in website indexing

441 Stella G. *Dextre Clarke*
Evolution towards ISO 25964: an international standard with guidelines for thesauri and other types of controlled vocabulary

445 Emily Gallup *Fayen*
Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39.19 for the 21st century

449 Robert *Fugmann*
Informationstheorie: Der Jahrhundertbluff. Eine zeitkritische Betrachtung (Teil 1)

459 **INFORMATIONSWIRTSCHAFT**
Vera *Münch*
Förderpolitik sorgt für 30 Jahre bewegtes Leben

INFORMATIONEN

412 24. Oberhofer Kolloquium – Call for Papers

424 Vom „point of information" zum „point of communication". Integration eines social networks in wiso

438 **REZENSION**
Huber, S.: Neue Erlösmodelle für Zeitungsverlage
(W. *Ratzek*)

462 **JAHRESREGISTER 2007**

464 **IMPRESSUM**

U3 **TERMINKALENDER**

I would do it again

Erfahrungen mit der intuitiven Registererstellung

Jutta Bertram, Eisenstadt (Österreich)

Der Artikel behandelt, was man beim Registermachen typischerweise tun und lassen sollte. Er widmet sich zunächst der Frage, welches Vokabular man wie in das Register aufnehmen sollte. Dann erörtert er, wann und wie Beziehungen zwischen den Vokabularelementen des Registers herzustellen sind.

I would do it again

The article deals with some typical dos and don'ts of book indexing. At first it addresses the question which vocabulary an index should contain and how it should be displayed. Then it discusses when and how one should define relations among the index' vocabulary.

Mit den folgenden Ausführungen möchte ich rückblickend einige der schwierigsten Fragen Revue passieren lassen, die die erstmalige Erstellung eines Buchregisters für mich aufwarf. Und ich möchte die Antworten vorstellen, die ich darauf gefunden habe. Im Nachhinein ist man ja bekanntlich immer schlauer. Dass ich das jetzt bin, habe ich wesentlich Jochen Fassbender, dem Gründer des Deutschen Netzwerks der Indexer (DNI), zu verdanken. Er machte sich nach der Publikation die Mühe, mein Register mit kritischen Anmerkungen zu versehen. Zu meiner nicht geringen Erleichterung bescheinigte er mir damals, „keine großen Katastrophen" produziert zu haben[1] – von den kleinen soll im folgenden die Rede sein.

Meine Ausgangsvoraussetzungen bei der Registererstellung waren die folgenden: Erstens: Das Buch, das es mit einem Register zu bestücken galt, war mein eigenes. Zweitens: Es war ein Buch über Inhaltserschließung, das sogar ein (allerdings recht summarisches) Kapitel über Register enthielt.[2] Drittens: Problembewusstsein war dadurch ausreichend vorhanden, aber viertens nicht die geringste Erfahrung mit der Registererstellung und fünftens auch nicht mehr die Zeit, mir das Know-how durch einschlägige englischsprachige Literatur anzulesen (in Ermangelung diesbezüglicher deutscher).[3]

Der Gedanke an ein reines Stichwortregister verbot sich vor diesem Hintergrund von allein – es musste schon etwas Gescheites sein. Register der elaborierteren Art bestehen nun letztlich wie kontrolliertes Vokabular aus Vokabularelementen und den Beziehungen, die man zwischen ihnen herstellt. Auf beide Ebenen möchte ich mit einigen Beispielen eingehen.

Das Vokabular: Was kommt rein, was bleibt draußen?

Diese Frage aller Fragen stellte sich grundlegend zunächst einmal in folgendem Sinne: Welche Buchbestandteile berücksichtige ich überhaupt? Was mache ich mit dem Inhalt von Fußnoten, Abbildungen und Tabellen? Und: wie mache ich diesen gegebenenfalls in den Fundstellen kenntlich? – Aus pragmatischen Gründen entschloss ich mich dazu, nur Abbildungs- und Tabelleninhalte zu berücksichtigen, die Fußnoten als notfalls entbehrliche Bestandteile des Haupttexts hingegen nicht. Dem Inhalt von Abbildungen und Tabellen stellte ich ein A bzw. ein T bei den Fundstellen voran, also z.B.:

```
Alphabetische Vokabularanordnung
   131, 136-137, A137, T139
```

Leserlicher wäre es aber mit nachgestellten Kleinbuchstaben gewesen, also so:

```
Alphabetische Vokabularanordnung
   131, 136-137, 137a, 139t
```

Um den Haupttext terminologisch nicht zu überladen, hatte ich in die Fußnoten oft alternative Fachbegriffe ausgelagert. Durch die Nichtberücksichtigung der Fußnoten im Register ging folglich Vokabular verloren, das als zusätzlicher Einstiegspunkt in das Buch hätte dienen können. So fehlt z.B. ein Ausdruck wie *Taxonomie*, der auf das Thema Klassifikation hätte hinführen können. Nun kann der Inhalt von Abbildungen bzw. Tabellen notfalls durch Verzeichnisse aufgefangen werden, derjenige von Fußnoten hingegen nicht. Daher wäre es wahrscheinlich klüger gewesen, andersherum zu entscheiden, wenn man der Kürze halber selektiv vorgehen muss: Den Inhalt von Abbildungen und Tabellen durch Verzeichnisse erschließen und herausnehmen aus dem Register, den Inhalt der Fußnoten hineinnehmen.

Die zweite grundlegende Frage war für mich, inwieweit kategorielle Brüche im Registervokabular erlaubt und zweckmäßig, vielleicht sogar geboten sind. Das betrifft vor allem die vielen im Buch angeführten Beispiele: So wird das Prinzip der Begriffszerlegung am Beispiel von Briefmarken erläutert, die Facettenklassifikation am Eierkauf illustriert, die systematische Ordnung am Beispiel von Schuhen problematisiert. Und es ist doch gut vorstellbar, dass jemand später nicht mehr weiß, wie denn der korrekte Fachbegriff hieß, von dem da die Rede war, sich aber noch darin erinnert, dass er irgendwas mit Briefmarken zu tun hatte. Beispiele dürften mutmaßlich das sein, was am ehesten im Lesergedächtnis hängenbleibt. Dafür sind sie ja schließlich da. Aber kann ich denn *Briefmarken* zwischen *Bottom-up-Prinzip* und *Browsing* platzieren? – Ich habe mich schließlich gegen die Aufnahme von Beispielen entschieden. Nutzerfreundlicher wäre indes gewiss eine Entscheidung dafür gewesen. So wäre es vielleicht gegangen:

```
Bottom-up-Prinzip
Briefmarken (Beispiel)
Browsing
```

Die Begriffsbeziehungen: Wie mit Hierarchien und Ausdrucksvielfalt umgehen?

Der erste Fragekomplex, der hier angerissen werden soll, ist der Umgang mit Schreibweisen-, Ansetzungs- und Ausdrucksvarianten. Ich habe das Register an der einen oder anderen Stelle mit Varianten schlichtweg überfrachtet, zumal dann, wenn sie, wie die folgenden Beispiele, am Wortanfang übereinstimmen:

1 Zitat aus einer Mail an mich vom 15.3.2007.
2 Es handelt sich um das Buch: Einführung in die inhaltliche Erschließung: Grundlagen – Methoden – Instrumente, Würzburg 2005.
3 Fugmanns Buch war zu diesem Zeitpunkt noch nicht erschienen (Fugmann, Robert: Das Buchregister. Methodische Grundlagen und praktische Anwendungen, Frankfurt am Main 2006) und Kunzes Klassiker (Kunze, Horst: Über das Registermachen. 4. Aufl., München 1992), gibt letztlich wenig konkrete Anhaltspunkte für das methodische Vorgehen bei der Registererstellung.

```
Äquivalenzrelation (Äquivalenz)
Bilddokument (Bild)
Inhaltliche Erschließung (Inhalts-
  erschließung)
```

Bei der Darstellung habe ich mich am Umgang mit Äquivalenzrelationen orientiert, wie er in einem Thesaurus praktiziert wird. Dort entscheidet man sich stets für eine Vorzugsbenennung. Bei Varianten, die sich als Folge der Koexistenz von Kurzform und Langform ergeben, sieht das dann beispielsweise so aus:

```
Regensburger Verbundklassifikation
  (RVK)  199-200, 254
(…)
RVK siehe Regensburger Verbund-
  klassifikation
```

Stattdessen bietet sich hier der Doppeleintrag als nutzerfreundlichere Alternative an – zumal dann, wenn er nicht mehr Platz beansprucht, als es der Verweis täte:

```
Regensburger Verbundklassifikation
  (RVK)  199-200, 254
(…)
RVK (Regensburger Verbundklassifi-
  kation)  199-200, 254
```

Der zweite und ungleich schwierigere Fragekomplex betraf den Umgang mit hierarchischen Beziehungen. Nämlich erstens: Soll ich sie abbilden? Und wenn ja: immer oder nur fallweise? Und wenn nur fallweise: wann und wann nicht? Und zweitens: In welche Richtung? Kann ich mich darauf beschränken, vom übergeordneten auf untergeordnete Begriffe zu verweisen? Oder auch andersrum? – Um es gleich vorwegzunehmen: Vor den hierarchischen Beziehungen habe ich weitgehend kapituliert, sie also nur ausnahmsweise im Register angeführt. Diese Entscheidung hat mir einiges an Zeit, Platz und vor allem Nerven erspart, nutzerfreundlich ist sie natürlich nicht. Denn durch die Ausweisung von Hierarchien werden Zusammenhänge sichtbar, die sich dem Nutzer möglicherweise nicht von allein erschließen und die ihm einen präziseren Zugriff auf die Buchinhalte ermöglichen.

Ich verweise also z.B. bei *Schlagwort* nicht auf *Gebundenes Schlagwort* und *Freies Schlagwort* und bei *Thesauri* nicht auf die einzelnen konkreten Thesauri. Auf die wiederum kommt man nicht ohne Weiteres, wenn sie *Thesaurus* nicht am Anfang ihres Namens führen. So mag man den *Thesaurus Ethik* in den *Biowissenschaften* oder den *Thesaurus Sozialwissenschaften* noch umstandslos finden, nicht aber den *Art & Architecture Thesaurus*, den *MeSH* und den *Standard-Thesaurus Wirtschaft*.

4 Bernard Levin zitiert nach einem Flyer der britischen Society of Indexers.

Was also tun mit den hierarchischen Beziehungen? Sie abbilden. Zumindest so, dass von einem übergeordneten Begriff auf einen untergeordneten verwiesen wird. Und zumindest dort, wo es keine morphologischen Übereinstimmungen am Wortanfang gibt. Bei der Darstellung sind zwei Varianten naheliegend: Entweder den Haupteintrag mit einem Siehe-auch-Verweis auf den untergeordneten Begriff versehen, also so:

```
Schlagwort  57, 68-69, 78, 132a
  siehe auch Gebundenes Schlag-
  wort, Freies Schlagwort
```

oder den untergeordneten Begriff als Untereintrag zu einem Haupteintrag ansetzen und mit Fundstellen ausstatten:

```
Schlagwort  57, 68-69, 78, 132a
  freies  71, 80, 82, 90, 94
  gebundenes  80, 82
```

Wann ist nun welcher dieser beiden Möglichkeiten der Vorzug zu geben? Der Siehe-auch-Verweis bietet sich dann an, wenn der untergeordnete Begriff selbst wieder sehr viele Fundstellen aufweist. Der Untereintrag wiederum ist das Mittel der Wahl, wenn der untergeordnete Begriff nur wenige Fundstellen aufweist und es zugleich mehrere Untereinträge zum Haupteintrag gibt. Nur einer allein ist unschön. Das obige Beispiel legt Untereinträge nahe. Dabei sind zwei Dinge wichtig: Erstens sollen sich die Fundstellen für den Haupteintrag nach Möglichkeit nicht mit denjenigen für die jeweiligen Untereinträge überlappen. Zweitens müssen die Untereinträge an der entsprechenden alphabetischen Stelle des Registers noch einmal vorkommen, nunmehr als Haupteinträge:

```
Freies Schlagwort  71, 80, 82, 90,
  94
(…)
Gebundenes Schlagwort  80, 82
(…)
Schlagwort  57, 68-69, 78, 132a
  freies  71, 80, 82, 90, 94
  gebundenes  80, 82
```

Im Falle, dass es sehr viele untergeordnete Begriffe gibt, kann man sich auch mit einem Pauschal-siehe-auch-Verweis behelfen, also z.B.:

```
Abstract
  siehe auch einzelne Abstract-
  Arten

Thesauri
  siehe auch einzelne Thesauri
```

Solche Verweise sind für den Nutzer freilich nur im Falle einer soliden Kenntnis der Materie hilfreich.

Fazit

Registererstellung ist schrecklich nervenaufreibend und zugleich wunderbar herausfordernd – wenn man ein Faible für Inhaltserschließung hat. Man erfährt hierbei sehr nachdrücklich, wie berechtigt es ist, dass es im Englischen nur ein einziges Wort für die Prozesse der Verschlagwortung und der Registererstellung gibt (nämlich ‚indexing') und wie berechtigt es zugleich ist, dass wir im Deutschen zwei Wörter dafür haben. Denn beim Indexieren im Sinne von ‚Verschlagworten' (im Englischen auch als ‚database indexing' konkretisiert) reicht es i.d.R. aus, wenn man solides Problembewusstsein für die Tücken der natürlichen Sprache und zusätzlich Kompetenz in der Anwendung kontrollierter Vokabularien mitbringt. Bei der Registererstellung (‚book indexing') braucht man hingegen Anwendungs- und konzeptionelle Kompetenz. Hier benötigt man also zudem das Wissen, wie man kontrollierte Vokabulare aufbaut und strukturiert. Die Registererstellung ist für mich daher gewissermaßen die Königsdisziplin der Inhaltserschließung. Alles in allem habe ich im Nachhinein großes Verständnis für den geplagten und vielzitierten Kollegen, der nach der Erstellung seines ersten Buchregisters schrieb: „I would rather be dead than do it again."[4] Allein: Ich würde es wieder tun. Aber vieles anders.

Book, Index, Indexing, Requirements

Buch, Register, Indexieren, Inhaltliche Erschließung. Anforderung

Die Autorin

Jutta Bertram, M.A.

studierte Soziologie an der Universität Bielefeld und war mehrere Jahre Dozentin für Inhaltserschließung am Potsdamer Institut für Information und Dokumentation (IID). Seit 2005 ist sie Hochschullehrerin an den Fachhochschulstudiengängen Informationsberufe und Angewandtes Wissensmanagement in Eisenstadt (Österreich).

Fachhochschulstudiengänge Burgenland
Campus 1
A-7000 Eisenstadt
Telefon: +43 2682 9010-60261
jutta.bertram@fh-burgenland.at

Translate the index or index the translation?

Pierke Bosschieter, Stitswerd (The Netherlands)

In this article the author describes the procedure of providing page numbers for a verbatim translated index. She explains the difficulties she encountered while compiling such an index. With this experience in mind she strongly argues against this method, in favour of indexing from scratch.

Besser ein Register übersetzen oder die Übersetzung indexieren?

Die Autorin beschreibt, wie sie einem wörtlich übersetzten Register Seitenzahlen zugewiesen hat, und erläutert die Schwierigkeiten, denen sie dabei begegnete. Aus dieser Erfahrung heraus spricht sie sich klar gegen diese Methode aus und plädiert stattdessen für eine Neuindexierung.

The Context

Publishing of translated non-fiction books

Many newly-published non-fiction books in Holland and Germany are translated from English. Statistics are hard to come by. In 2006, of the 81,177 newly published books in Germany, 3,781 were translations from the English, but according to the *brancheninfo*[1] these were mostly novels. Taking the Royal Dutch Library, NetUit[2], Website as my basis, I tried to establish some figures for the Dutch publishing industry. Looking at the A-list and only at non-fiction, I came up with the rough estimate that about 7% of the non-fiction books published in the previous couple of months were books translated from English. Even more telling was that some of these books had indexes in the original publication but not in the Dutch version. As I only checked a few on Amazon.com, with the 'Look Inside this Book' feature this is not a well-researched statement and should be read with caution. I could not find any total figure for books published in Holland later than the 2002 figure of 19,061.

In the absence of firm figures, for the purposes of this article, I will assume that the translation of English non-fiction books is a normal occurrence in both countries and is done on a regular basis.

The Indexing Profession in Germany and in The Netherlands

In her article Continental European Indexing, Caroline Diepeveen makes some general remarks about indexing practice in Germany and in The Netherlands. Her researches suggest that about 52.5% of non-fiction books in Holland have an index, approximately the same percentage as Martin Tulic quotes for the US in the first quarter of 2006[3]. Quantitatively equal but qualitatively not. Over the last 50 years or so, Anglo-Saxon countries have developed indexing into a science (if not an art form), while on the European mainland indexing has had little attention. In Great Britain, as in the USA, there are many ways open to students to follow specialized training courses at private institutions or universities. In addition, indexers in the English-speaking world can turn to their national indexing societies for help with on-going professional development and for general professional support.

Germany and The Netherlands now both have their own indexing networks, the Deutsches Netzwerk der Indexer (DNI)[4] and the Nederlands Indexers Netwerk (NIN)[5], which are signatories, as fully-fledged affiliates, of the International agreement of indexing societies.[6] Both networks hope to draw attention to the indexing profession and to raise the standard of indexing in the two countries.

Why opt for a translation

In English-language books, an index usually is a subject or analytical index, while indexes in German and Dutch publications are often poor quality indexes or just names indexes. In addition, these indexes usually have long strings of page numbers attached to each entry, which is not exactly user-friendly. For the reader of a non-fiction book a general subject index is of course much more helpful than a subject index with long strings of locators or an index containing only names of persons.

One of the first things publishers deciding to publish a translation of an English or American non-fiction book come up against is the realization that the index in the original book is much more helpful than the average Dutch or German index. Not knowing much about the indexing process and unaware of the professionalism of the indexers in the German and Dutch Networks, some publishers choose the option of verbatim translation of the index. This choice is often prompted out off reverence for the book (and its index) and/or because of financial constraints. They assume that this method will produce an index of the same quality as the English index, as cost-effectively as possible.

What publishers should be aware of is the intricate process of writing an index. A professional indexer starts reading on page one, works his or her way through the complete text and while doing so analyses the text and picks out the concepts, subjects and names that should be included in the index. This sounds simple and straightforward, but it is actually a rather complex intellectual and creative process. After identifying the indexable matter, the indexer needs to find terminology for the index that is appropriate to the prospective readers of the book. It's not a matter of merely underlining important names and keywords in a text and putting these in alphabetical order, for that could be done by a computer. Because of syntactic and semantic ambiguity, the understanding of a text is far from straightforward. Think of

1 www.buchhandel-bayern.de/brancheninfo/wirtschafts-zahlen.shtml.
2 http://netuit.kb.nl/kozijn.htm?A-lijst.ned.
3 www.anindexer.com/about/bestsellers/bestindex.html
4 www.d-indexer.org
5 www.indexers.nl
6 *The Indexer* 25 (3) Centrepiece 2, C16

the use of synonyms and homographs, of the fact that certain concepts are not implicitly mentioned and of the fact that sometimes the less important words hold the key to understanding a text. That is the indexing challenge.

In addition indexers have to adhere to rules, formulated in indexing standards and intended to help the user, about the construction, lay-out, typography and other technical details of an index. The Dutch standard NEN 3547 was published in 1981 and withdrawn in 2001. Since then the official standard in Holland is the international ISO 999 which was published in 1996 and is widely used as an authoritative guide worldwide. Germany, on the other hand, has the DIN 31630, which has been published in 1988. According to Jochen Fassbender the standard covers the important aspects of indexing, although some of it is now out of date (for instance the recommended use of the abbreviations f. for following page and ff. for following pages). Moreover, it does not cover a number of important aspects, such as the limitation of the number of locators or page numbers to be attached to a heading and it does not say anything about passing mentions[7], a subject I will return to later.

A case study

Preliminaries

In August 2007 I was asked by the Dutch publisher Nieuw Amsterdam to provide page numbers to the verbatim translation of the index to 'The Years of Extermination 1939–1945, Nazi-Germany and the Jews by Saul Friedländer, which was published to much acclaim in the US in 2006 by HarperCollins. It is the second and last part of Friedländer's comprehensive study of the Holocaust and the English version has about 665 indexable pages. This was the first time I was ever offered such a job and I thought it would be a nice challenge, especially since I had to deliver the index in a week.

A German version of the book had already been published by C. H. Beck Verlag, at the beginning of 2007. Both the American and German versions have an index.

Nieuw Amsterdam had the index translated in parallel with the translation of the text, by a different translator. This second translator had a solid knowledge of the subject and kept in touch with the translator of the text. After completing the translation of the English index, he compared it with the German index. Entries that appeared only in the German index were duly translated and added to the 'master list'. The idea behind this was that they would now have a comprehensive 'master list', equal the level of detail of the book. It would also, thought Nieuw Amsterdam, be a time- and money-saver, for on completion of the Dutch translation, this 'master list' needed only page numbers to turn it into a first-rate index.

I was presented with a 'master list' of 2763 lines, each line representing one main entry or subentry. The list was not in alphabetical order and only the entries copied from the German version had page numbers attached (of course referring to the German text). I was also provided with PDF files and hard copies of the book in all three languages. As I said earlier, an index cannot be generated automatically by computer, but the computer can greatly assist the human indexer. There are three major software packages available at the moment: CINDEX, MACREX and SKY Index. As I use SKY, my remarks about software are mainly references to SKY. SKY's main window is divided horizontally in two; the upper part showing the index in progress, the lower part showing a database grid where you can enter and modify your entries. The software has powerful tools to help with building an index, including fast data entering, tools to help with sorting order and repagination, lay-out and typography, and tools for checking inconsistencies.

Procedure

I started by loading the 'master list' into SKY, so that I could use all the alphabetization and lay-out tools. I then made a transposing chart, listing the page numbers of the English edition with the corresponding page numbers of the Dutch edition. Of course this was a rough chart, because I couldn't include all 661 page numbers, for lack of time and because the list would have become too unwieldy. The list fitted an A4 page and hung below my main screen. Another page, also fastened to the main screen, had a list of corresponding page numbers of the chapters and paragraphs of the Dutch edition. My big main screen had an open SKY window and Adobe reader with the Dutch text. On my second screen I had the second Adobe reader with the English PDF file opened. In addition I had printed versions of the 'master list', the English index and the German index on my desk in front of me. Then I took the 'master list' and started with the first heading.

What was of course obvious from the start was that it was easy enough, using the search option in Adobe Reader, to find names of persons, geographical entities and corporations. But what became clear when working my way down the 'master list' was that the names inserted from the German index were very often passing mentions and therefore omitted by the American indexer. It is one of the discrepancies between the DIN 31630 rules and the ISO 999 rules. According to the latter, an indexer should discriminate between information on a subject and the passing mention of a subject and exclude passing mentions that offer nothing significant to the potential user.[8]

Another significant difference in the indexing practice of English and German indexers, relates to the use of long strings of undifferentiated page numbers. Although ISO 999 states that users should be able to retrieve information quickly[9], the DIN 31630 doesn't touch on this subject at all. Leading American and English literature on indexing clearly explains the usual upper limit of 5 or 6 locators per entry. In fact both indexes had long strings of locators. In the German index it was a very common occurrence, in the American index an exception.

As the translated index can only be as good as the original version, the publisher should have ascertained the quality of the indexes, before he decided to have them translated. The German one had a very simple structure, made no use of see and see also references[10], used very long strings of locators, made use of 'f.' and 'ff.'(which is not recommended by ISO 999 because it gives the user incomplete information[11]), contained a lot of passing mentions and had too few subheadings. On the other hand, the American one had a lot of merits, like an intricate structure with see and see also references, an appropriate number of subheadings and no passing mentions. In this case, the American index seemed on the face of it, to be a good one and should not have been augmented by mostly unnecessary headings from the German index.

The structure of an index is very intricate and develops slowly as the indexer makes his or her way through the text. Usually the structure is created simultaneously with the selection of main and subheadings, so that interrelated terms can be created on the fly. This can mean that one page reference may be attached to many headings and subheadings. For example:

7 Fassbender, Jochen: German indexing: some observations on typographical practice, and a personal email to the author.
8 ISO 999 rules 4b and 4c
9 ISO 999 rule 4e
10 ISO 999 rules 7.5 to 8
11 ISO 999 rule 7.4.3

```
Amsterdam, 121
anti-Jewish measures
   in Belgium, 121-122
   in Holland, 121-123
   Jewish council and, 121
anti-Semitism
   in Nazi-soldiers' letters, 121
Belgium
   anti-Jewish measures in, 121-122
Bieberstein, Wilhem, 121
Comité of Coordination, 121
Comité de Défense des Juives, 121
Danneker, Theodore, 121
Holland
   anti-Jewish measures in, 121-123
Jewish councils
   in France, 121
   in Holland, 121
letters, German soldiers'
   anti-Semitism of, 121
suicides, 121
```

The above example shows the interrelationships between headings created when the indexer read page 121. It shows that when Belgium is mentioned under anti-Jewish measures, the main heading Belgium should have a subheading anti-Jewish measures and the page range for both should be identical. The same applies to anti-Semitism in Nazi soldiers' letters. While working through the 'master list' I discovered that the American index wasn't as well structured as I first thought, for a lot of the above mentioned interrelationships were missing or the page ranges for some of the interrelated terms were not identical in instances were they should have been.

My example also makes it abundantly clear, why compiling an index by simple translation, is the most ineffective method imaginable. For an indexer working from scratch reads page 121 once and makes the 15 entries and moves on to page 122. Working with the 'master-list' I did encounter the page references to page 121 fifteen times and had to scan the page anew each time to see where a concept ended and began. Coming from any number of other pages I had to familiarize myself yet again with the content of page 121.

My finished index has 5,300 locators. I checked at least 1,000 locators that turned out to be passing mentions and didn't end up in the finished index. The book had about 625 indexable pages. So in the end I looked at each page an average of ten times. In addition, although I had the transposing list to work from, it was time-consuming to find the exact passage the page reference was referring to, especially so with underlying concepts and ideas like 'aangeven van Joden' (denouncing/betraying Jews). I had also to repeatedly read/scan the next pages to see where the passage ended.

Another aspect that slowed me down was the fact that the 'master list' contained words that simply did not appear in the text, due to the fact that the translator of the index was not the translator of the text.

In the end it took me 100 hours to complete the index. Because of time constraints – six working days – the final index still had a lot of main headings with long strings of page numbers attached. The structure was better, but not flawless and there wasn't enough time to sift out all passing mentions. Indexing from scratch, taking no more and probably much less time, would have resulted in a higher quality index. As there would have been no need for translation, this method would also have been more cost-effective.

Copyright issues

An afterthought about copyright

Who is to take the credit or the blame for this translated index: The original indexer, the translator of the index, or me, the 'minion' who dealt with all the drudgery of converting the page references? That is to say, I don't believe this latter job could have been done without a thorough knowledge of indexing. Even re-indexing from scratch with the original index as a reference, can in practice result in an index close to a verbatim translation, in which case, the original indexer might also have copyright claims, but these will be extremely difficult to enforce.

What if the end result is a very poor index (which I hope is not the case with the Friedländer book)? Aside from any questions about copyright, this might be a breach of moral rights, since the resulting index could lay the original indexer open to ridicule[12].

Conclusion

The translation of an existing index may well seem the easy option, but in practice, because there is no automaticity about the indexing process, translating an index will probably run into many problems. If the original index is a poor or mediocre one, the quality of the translated index won't surpass it. Editing the translated index requires the skills of a professional indexer, and carrying out the necessary checks is likely to be both tedious and time-consuming. Bearing this in mind, it is far better to start from scratch.

References

Diepeveen, Caroline: Continental European indexing: then and now. The Indexer 25(2006)2, pp. 74–78.

Deutsches Institut für Normung: DIN 31630 (Teil 1). Registererstellung: Begriffe, Formale Gestaltung von gedruckten Registern.Berlin: Beuth, 1988.

Fassbender, Jochen: German indexing: some observations on typographical practice. The Indexer 25(2006)2, pp. 79–82.

The Indexer 25(2006)3, Centrepiece 2, C16.

International Standard Organization: ISO 999. Information and documentation: Guidelines for the content, organization and presentation of indexes. 1996.

>Index, Translation, Copyright, Workflow

Register, Übersetzung, Rechtsfragen, Arbeitsablauf

The Author

Pierke Bosschieter

studied library science in the 1990s, while running a catering business and dabbling in food writing. In 2004 she took up indexing to supplement her income. In October 2005 she set up her own indexing business ISB&Index. She now is indexing full time.

ISB&Index
Akkemaweg 6-8
9999 XL Stitswerd
The Netherlands
pierke@isbnindex.nl
www.isbnindex.nl

12 Roger Bennett in an email to SIdeline (listserv of the Society of Indexers)

How I reused my own index

Peter Rooney, New York, NY (USA)

When the index to a previous edition of a book is revised for use in a new edition, this is called index revision. There are pros and cons of trying to revise an index, and various measures of success. A detailed case study is given, describing the software method that was used.

Recycling eines Buchregisters

Eine Index-Revision liegt dann vor, wenn das Register zu einer vorhergehenden Auflage eines Buches für die Neuauflage überarbeitet wird. Es gibt Vor- und Nachteile beim Versuch, einen Index zu überarbeiten – sowie verschiedene Erfolgsmaßstäbe. Eine detaillierte Fallstudie wird vorgestellt und die benutzte Software-Methode beschrieben.

The question is whether it is possible or desirable to take an existing index and reuse it for a new edition of a text.

Some editors seem to think it is a simple matter to find new page numbers for the old. This used to be a job only for the most novice of indexers, at the lowest rate of pay. Certainly, the way it was done would drive any person of imagination "up the wall." The original index pages were photoenlarged by 200% or so, and the re-indexer tracked down every reference straight through from the As to the Zs, writing the new page numbers on the index copy. Then the compositor reset the index.

Still, in recent times, there are many who doubt whether index revision can be done at all. There has not been much previous discussion. One article is titled "The myth of the reusable index," but the discussion is more narrowly confined to embedded indexing (Johncocks, 2005). Another article titled "Embedded Indexing" is more optimistic, and discusses the software tools that aid this (Lamb, 2005). There was a useful exchange of views on the online discussion list INDEX-L on March 28, 2006. A balanced discussion is provided by Mulvany (2005) in a section "Revising an Index for a Revised Edition," but the discussion urges careful preparation and record-keeping by the publisher to facilitate the indexer's task, and this is often not feasible. Wellisch (1991) states that index revisions are usually not cost-effective and recommends making a new, better index by learning from the mistakes of the previous one.

The present essay will discuss index revision: when to attempt it, how to do it, and how to judge if it is successful.

What are the arguments in favor of index revision?

- it may improve an already good index: like standing on a giant's shoulders (it's the pygmy that does this, in the saying)
- it may save time and money, but this is questionable
- the editor may want it, or suggest it (though the editor can often be persuaded otherwise)
- it may rarely be used to check the work of another indexer
- it may even be used to check your own indexing achievement, if you have the opportunity to work on your previous index. That is the focus of the present essay.

If an old index can be recaptured, the advantages to the index making are these:

- the intellectual labor of distinguishing major from minor references has been done
- the references have been captured and assigned to pages
- the vocabulary is settled
- the alphabetizing has been established
- the mode of expression of headings and subheads has been suggested
- a crossreferencing structure has been created.

Further advantages are achieved if the index has been captured from a true index database:

- the text is accurate, that is, letter 'O' is distinct from digit '0', and letter 'l' is distinct from digit '1'
- main headings, subheadings, sub-subheadings, function words, page references, and crossreferences are clearly defined (instead of being squashed together)
- page references are accurate (one hopes)
- page references are in sequential order on the page (but this is not always the case for some index software)
- internal comments, if any, are passed on
- a thesaurus of the vocabulary is implicit.

With so many advantages for index revision, what possibly could be the difficulties? Here are some:

- capturing the old index is not always easy. If an index database is not available, you may be able to capture it from a wordprocessing file. If this is not available, you may be able to scan the old index. If this is not available, you may type in the old index. At each downward step, quality is lost and time is spent
- it is not always easy to find out what the previous indexer meant, particularly when trivial mentions are included and major discussions are overlooked
- outright errors in pagination may be rampant: ("10" instead of "100")
- the range problem most likely will arise, whereby old page "5" is split into "5, 6" or "5–6," or the reverse, where old "5–6" becomes either "5" or "5, 6"
- changes in the text may occur, sometimes minuscule ones (one name substituted for another, e.g., generic name "diazepam" replacing trade name "Valium")
- and worst of all, many indexes are abominable, and it is not always easy to tell the quality that you are dealing with before you are well into the work. An index can look splendid, stylistically, and be almost useless for the purpose of re-indexing. It is best to do some research and calculation before starting out. When the old index is bad – which is usually the case, sadly – you may be able to get some profit out of it (as discussed below), but it is often better to convince the editor that a new start is needed. Often, not much convincing is needed.

How is the indexer him or herself affected by the re-indexing task? To begin with, an indexer is a 2nd-hand creator, since he deals with another's text and ideas. A re-indexer is a 3rd-hand creator, since he is working with the text of the 2nd-hand creator. The work can be boring, tedious, slow, and frustrating. Errors in the old index can cancel out any gain from reusing it. It may be a case of taking ten tiny steps forward, and then being forced backward one giant step.

So is index revision worth trying? The answer to this is "measured by what standard?" It may be measured in money or time, but how does this relate to quality? If quality is the main standard, then it may well be worth doing.

How to judge success?

- if the job does not take more time or money than an original job. (It is unlikely to take less time. The corollary is never to charge less for index revision than for a job of original indexing. Find another way than cost-saving to justify the job.)
- measures of goodness, of differences, of errors – all are lacking in the present state of indexing research. Later in this essay, I'll discuss metrics.

How to recapture an old index

These are listed from best case to worst case:

- from an index database
- if this is not available, you may be able to obtain a wordprocessing file of the old index (a typesetting file, however, in Quark for example, will be quite difficult to work with)
- if a wordprocessing file is not available, you may be able to optical-scan the old index. Many errors are to be expected, so careful proofing and correction is needed
- if this is not available, it may actually be advantageous to type the old index into indexing software.

At each downward step, quality is lost and extra time must be figured into the cost.

What elements of an old index can be captured

These are listed in order from lesser to greater:

- mere terms, but of course these can be useful, particularly if they are long and difficult to type. Indexing programs may permit you to recall an indexing phrase merely by typing in the first few letters. Here are terms from a recent index revision that I was glad not to have typed:

```
acetaminophen
American Psychiatric
  Association
anticholinergic drugs
```

- not only terms, but concepts, such as a combination of mainhead plus subhead. For example, the same index had recurring structures like this:

```
Valium
  as habit-forming
  how long to stay on
  side effects of
```

  for various medications. Although the index as a whole was quite spotty and unreliable, the bound phrases were useful
- terms and concepts as above, but with reliable references. This is what we hope for when taking on an index revision job. (As a matter of terminology, any lesser degree of quality can be called "index reuse" rather than "index revision.")

Case Study

When I was called by one of my publishers (the United Nations Statistical Office in New York City) to work on a revision of a document on censuses for which I had provided the index ten years earlier, I realized I had the ideal case. I could reap most of the advantages listed above, and avoid most of the difficulties.

I also thought I could rely on my memory of the subject matter, which I found thought-provoking; as it turned out, my memory was rusty. (Among some provocative ideas: people lie about their age in all parts of the world; it's difficult to take a snapshot of populations that are constantly in movement; some countries take a day or week holiday to count the population, and in effect make the population "freeze.")
The book is called *Principles and Recommendations for Population and Housing Censuses, rev. 2*. Its intended audience is statistical offices in all countries of the world. (The U.S. Bureau of the Census under the Department of Commerce is the corresponding United States office.) The intention of the book is to guide officers in planning population and housing censuses, as well as other data-collection activities, particularly demographic and socio-economic surveys. The manual has many practical suggestions for setting up field offices, training census takers, devising questionnaires, and publishing results in a standard format so that results can be compared country to country and decade to decade. The book comes with about 125 sample tabulations. (The first one is called "Total population and population of major and minor civil divisions, by urban/rural distribution and by sex.")

Ten years ago this had been a 135-page book of about 100,000 words (not including the back matter). Now it was to be 227 pages, and 127,000 words. I compared the table of contents of the old and new editions, and determined that much of the content was new, and some of the old content had been deleted, moved, or so extensively revised that it might as well be new.

I was asked to bid, and I practically doubled the price of ten years ago. I calculated my bid based on the increase in percentage of indexable material, and a price inflation index (26% over ten years in the United States, if you want to know.)

I retrieved the data that I had stored ten years ago. You immediately wonder, when retrieving something from so long ago, whether program and data formats are the same; and whether the old media is even physically readable. Fortunately, it was retrievable.

The first few lines of the previous index looked like this:

```
Note: Reference numbers are to
parts and paragraph numbers.
"P" references refer to
Population tables in Annex I;
"H" references refer to Housing
tables in Annex II.

acceptance sampling, 1.150
accuracy, 1.145, 1.287—290
active population. See head of
 household
activity status (economic),
 2.168—208
  population, by type of dis-
    ability, urban/rural area,
    age and sex, P8.7
  See also economically active
   population; foreign-born
   population; head of house-
   hold; households
addresses of buildings, 2.317
  list of, preparing, 1.108—109
administrative divisions. See
 territorial and administrative
 divisions
adopted children, 2.129
age, 2.85, 2.87—95
  need for data on, 3.46, 3.72
  of population by single years
   of age and sex, P3.1
age at first marriage, 2.142
age-heaping, 1.266
age of mother at birth of first
 child born alive, 2.119, 2.143;
 P4.4, P4.5
```

(The last heading is not a mistake; it's a complex unitary idea of demographic statistics represented by a complex heading.)
The next step was to mark off old section breaks onto the new pages of the book text. (Some indexers reco*

mmend the reverse, but this usually doesn't work well.) This book is laid out, not in page order, but in numbered paragraphs, a great advantage because page breaks are arbitrary divisions, but paragraphs are logical divisions. I marked the old section numbers in the left margin of the book pages opposite the new section numbers, like this:

```
        I. Essential Roles of
           the Census
NEW   + 1.1  Evidence based
           decision making is ...
           effective governing
NEW   + 1.2 While the roles of
           the population and
           housing census ...
NEW   + 1.3  It is critically
           important to produce
           detailed statistics ...
           II. Definitions
1.1   = 1.4 A population census
           is the process of ...
1.2   = 1.5 Population is basic
              to the production
              ...
1.3   = 1.6 A housing census is
              the process of ...
```

For this example, I have shown the paragraph numbers and a few words from each paragraph. The left-hand column represents my marking of the new pages. Some paragraphs, having been deleted, could not be found. Some paragraphs in the new text had no corresponding old paragraph number. Some text had been so altered that it bore only a slight relationship to the old, or had been moved so far away from its original position that it could not easily be found. As shown above, I marked exact correspondences with an equal sign (=), and inexact ones with an approximation sign (~). Added paragraphs or text were marked with a plus sign (+). If a paragraph or text had been deleted, that fact was indicated with a minus sign (–).

Then I prepared a data table for use by repagination software. The table looks like this.

```
1.1//=/1.4
1.2//=/1.5
1.3//=/1.6
```

This means that old paragraph 1.1 will be renumbered as 1.4; old 1.2 will be 1.5; and 1.3 will be 1.6. Evident is the fact that new paragraphs 1.1, 1.2, and 1.3 have no old equivalent, so they will be newly indexed. Likewise, there are old paragraph numbers that have no new equivalent.

Then the repagination software goes to work. The table will translate old to new references. However, if an old reference has no corresponding new reference, the old reference will be carried over as an italicized page number (shown here, for visual effect, by underlines). The two sorts of references are interspersed.

```
acceptance sampling, 1.150
accuracy, 1.145, 1.410–413
active population. See head of
 household
activity status (economic),
 2.245–330
   population, by type of
    disability, urban/rural
    area, age and sex, P8.7
   See also economically active
    population; foreign-born
    population; head of house-
    hold; households
addresses of buildings, 2.460
   list of, preparing, 1.174–175
administrative divisions. See
 territorial and administrative
 divisions
adopted children, 2.183
age, 2.133, 2.135–95
   need for data on, 3.71, 3.99
   of population by single years
    of age and sex, P3.1
age at first marriage, 2.192
age-heaping, 1.389
age of mother at birth of first
   child born alive, 2.169,
   2.193; P4.4, P4.5
```

At this stage, the index is at least 50% accurate. That is to say, it is almost publishable, especially if you cut out the italicized refs.

Now one goes on to compare the actual data. It's a reality check. The data of the old index is sorted in page order, and placed in an Old Window on the left. The New Window on the right is initially empty. The illustration below shows a repagination program I devised, called DEAL'M. It can be used with some of the standard indexing programs, or with my own indexing system, which is called MTOPIC. To see what the screen looks like, see Figure 1.

With the new text in front of me, I start the program. I see whether a reference in the Old Window actually appears in the text. If it does, I click it, and it is moved to the New Window. Then I underline the term or phrase on the page. If a new idea comes up, or for sections that have no corresponding old section, I enter the new term directly into the New Window.

Similar methods have been described in INDEX-L discussions. It is best to look at the text, decide what you think should be indexed, and then locate it in the Old Window. (This is better than trying to match everything in the Old Window.) When finished with a page, you look at the Old Window to see if any terms were not checked. These may be trivial

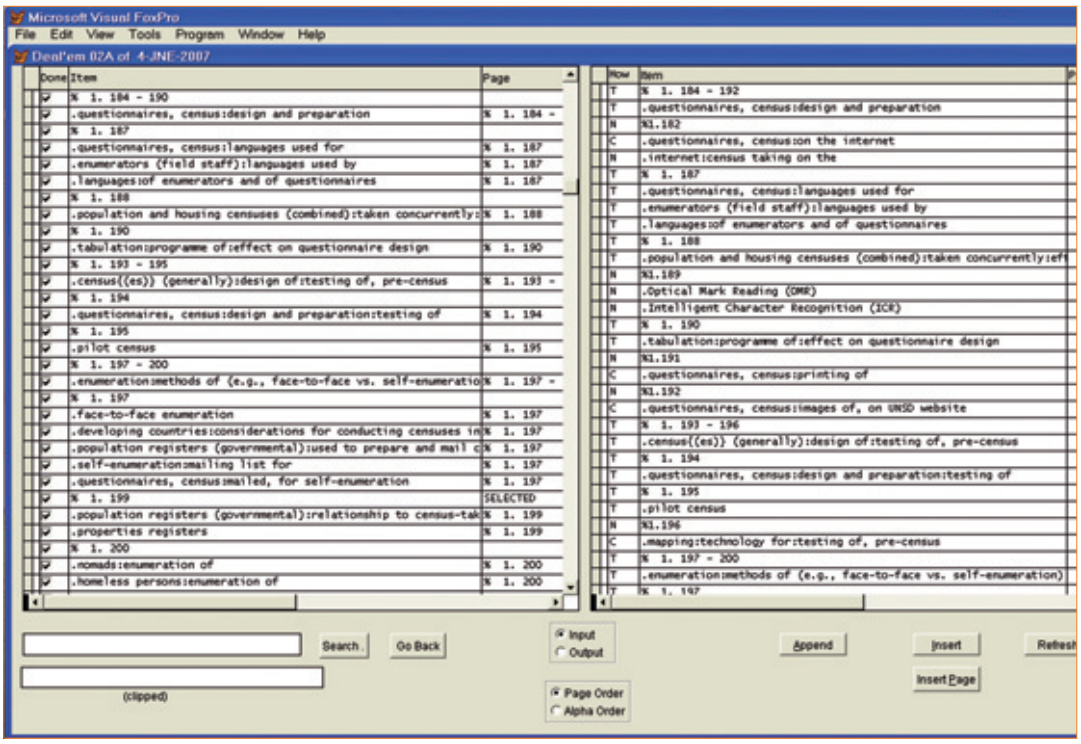


Figure 1: Screen of MTOPIC

mentions, but possibly you overlooked them and they should be included in the New Window.

This process amounts to a thorough proofing of the old data. In fact, the program can be used for that purpose. I can operate at the rate of about 200 references per hour if there are no complications; and still faster, if I trust the data sufficiently. But overall, when new data must be added, the rate is about 100 references per hour.

After this manual repagination is made, there is still the usual work of refining the wording, and checking for accuracy of cross references and mirror references.

The final index looks like this:

```
absent at time of count, 2.24,
2.36, 2.74
acceptance sampling, of batches
 of census data, 1.313
accessibility of information,
 and quality of a census, 1.237,
 3.95
accuracy, 1.410–413
  and quality of a census,
1.234
active population. See
 economically active population
activity status. See economic
activity status
addresses of buildings, 2.460
  list of, preparing, for use
    by enumerators, 1.174–175
administrative divisions. See
 territorial and administrative
 divisions
adopted children, 2.183
aerial photography, 1.138
age, 2.135–143
  direct question about,
   improving accuracy of
   answers, 2.138–139, 2.141–
   142
  estimated, 2.140
  need for data on, 3.71, 3.99
  tabulations on, P1.4-R, P2.1-
   R, P3.2-R, P4.1-R, P5.4-R,
   P1.1-A, P3.5-A, P3.6-A,
   P8.1-A
    of population by single
     years of age and sex,
     P4.1-R
age at first marriage, 2.192
aged. See elderly
age-heaping, 1.389
age of mother at birth of first
 child born alive, 2.169, 2.193
```

Just for this section alone, the index is 25% larger. New ideas have been identified, and headings have been more finely qualified or have been reworded for clarity. Many cross references have been removed (replaced by actual entries.) Some errors in the previous index have been corrected. And yet there are traces of the original index still present, and the intermediate page conversion yielded some permanent results.

Although I benefited from working with my own specialized re-indexing software, the same result could have been obtained by other methods, although more laboriously, so long as page-sorted data could be made available. The fact that this book was segmented into numbered paragraphs was another advantage, but not an overwhelming one.

Theory

An index can be likened to a map of a territory. A bad index is like a map drawn by a child or an unsophisticated person. An index to a previous edition is like an old map to a contemporary city – many streets remain the same, but some have been obliterated, new buildings have

risen, and even some new land has been developed.

If you go to Google Earth™, you can look at an area with your choice of features: place names, roads, rivers and mountains, buildings, political boundaries, and various other features. It's really several maps, overlaid on each other. A poor index is an illogical mixture of such features, inconsistently laid down, omitted, or inaccurately placed. Keep in mind the purpose of an index – it's for users. Some may want place names; some may want roads. One index needs to serve them both.

Or consider the metaphor of architecture. We can reuse the plan (layout) of an index; but we may need to rearrange the furniture, or buy new furniture. Side buildings may be added, and some rooms removed. This is similar to the case of this "Census index" where sections had been rearranged and in some cases were barely recognizable. It would be foolish to tear down a usable building just for new occupants. We remodel the building to suit the new needs.

Need for Statistics

If indexing is ever to be a science, it needs quantitative methods, and hence metrics. If we are comparing an old with a new index, the questions are:

> What is the extent of the differences?
> How much effort has been saved in reusing the old index?

I'd invite a mathematician reader to quantify the differences, and to tell us what it means to say that a text or an index is X% different from its predecessor.

The first means of comparison is purely visual, and that may be the best. People are good at pattern recognition. I can simply compare the original index to the final revised index side by side, and the differences stand out. Using Microsoft Word to compare the two indexes is an even greater visual aid.

Since we already know that the page numbers have changed, and that there are new sections, to make a fair comparison we should:

- change page numbers to placeholders, e.g., to fictional page "999"
- exclude sections from the old index that were not carried over
- exclude sections that are new.

Then we can use a file comparison program. Though there are several, I used one which also produces statistics.

The crude statistics for this job of repaginating the "Census index" are:

Conclusion

	Previous Index	New Index
Words of indexed text	100,000	127,000
References in index	1,715	2,462
Lines of index	1,818	2,290
Characters of index	51,814	65,383

I would like to say that I left the new "Census index" better than I found it. Even to claim this is somewhat embarrassing, since it suggests that the first index which I prepared was not quite adequate. A new index might have benefited from the mentality of another indexer. I may be too inclined to endorse my ten-year-ago self's effort. I do several index revision jobs on an annual basis. It is probably easier to make progress in one-year intervals, than with a ten-year hiatus.

This gets into the topic of double indexing. Surprisingly, few if any controlled studies have been made of two indexers indexing the same text. A research study should be undertaken, funded by an index publisher or a foundation. This would be followed by an analysis of which index is better, and a recommended method of merging the best elements from each into the final product.

The era of the continuous text is upon us, such as the online encyclopedias. And yet they don't have indexes, as we conceive them. Rather, they use a kind of algebraic free search. We need to develop something between the traditional index and free search; and a method of embedding indexes better than we have. By "we" I mean "we indexers."

There are perhaps more questions posed here than answers, but I hope I've persuaded you that some indexes can be reused. Certainly the economic trend in publishing suggests it is a worthwhile and necessary effort.

References and acknowledgments

Google Earth™. http://earth.google.com

INDEX-L. Discussion list hosted at the University of North Carolina (listserv.unc.edu). Messages on 28 March 2006, among Janet Perlman, John Culleton, Deborah Patton, and others.

Johncocks, Bill (2005) The myth of the reusable index. The Indexer 24(4), October 2005, pp. 213–217

Lamb, James (2005) Embedded indexing. The Indexer 24(4), October 2005, pp. 206–209

Mulvany, Nancy (2005). Indexing Books. Chicago and London: University of Chicago Press. "Revising an Index for a Revised Edition," pp. 238–241.

United Nations (1998). Principles and Recommendations for Population and Housing Censuses - Rev. 1. Statistical Papers, No. 67. Available through http://unstats.un.org/unsd

United Nations (2007). Principles and Recommendations for Population and Housing Censuses – Rev. 2. Statistical Papers, No. 67. In final stage of preparation as of September 2007.

Wellisch, Hans (1991). Indexing from A to Z. Bronx, N.Y.: The H. W. Wilson Company. "Revision of Indexes," pp. 335–337.

Book, Index, Revision, Software, Efficiency

Buch, Register, Überarbeitung, Software, Wirtschaftlichkeit

The Author

Peter Rooney

is an independent software developer and freelance indexer. He devised the first bibliographic software used for the MLA Bibliography, and has worked in fields of museum and library cataloguing. He uses his own custom indexing software to accomplish his freelance work. Rooney has been involved in the American Society of Indexers since its inception in the late 1960s.

Magnetic Reports
332 Bleecker Street, #X6
New York, NY 10014-2980
USA
magnetix@verizon.net
www.magneticreports.com

Cyborg indexing: Half-human, half-machine

Seth Maislin, Arlington, MA (USA)

The quality advantages of human indexing are not replaced by computerized or automatic indexing, but for particularly large projects human indexing is infeasible. A cost-benefit analysis is necessary to balance the practical benefits of automated indexing with the likelihood and consequences of errors created from each approach. Integrating both human and computerized approaches, here called "cyborg indexing", can minimize the frequency and impact of errors created by both approaches.

Cyborg Indexing: halb intellektuell, halb maschinell

Die qualitativen Vorteile des Indexierens durch den Menschen werden nicht durch automatisches Indexieren verdrängt, aber bei besonders großen Projekten ist das Indexieren durch den Menschen nicht durchführbar. Eine Kosten-Nutzen-Analyse ist notwendig, um den praktischen Nutzen automatischer Indexierung mit der Wahrscheinlichkeit von Fehlern und deren Konsequenzen abzuwägen, die in beiden Ansätzen entstehen können. Das Integrieren von nicht-automatischer und automatischer Indexierung, das hier „Cyborg Indexing" genannt wird, kann Häufigkeit und Auswirkung von Fehlern beider Ansätze minimieren.

The term "indexing" is used in two very different contexts but with the same basic meaning: the analysis-driven production of locator terms that can be used to help identify documentation or other content that may be of value.
In one context, humans perform the indexing.
In the other context, computers perform the indexing. For both the required prerequisites are exactly the same: awareness of the audience, knowledge of the subjects being indexed and their intended use, full access to the content, editorial analysis, word selection skills, and topical organization skills. Nevertheless, ask users of automated indexing tools to switch to human indexing, and they'll likely refuse, claiming that human indexing takes too long, costs too much, can't possibly keep up with the reality of information production, and may require interaction with a belligerently academic human being with an impractical sense of perfectionism. Meanwhile, human indexers claim that computer-based analyses produce indexes of substandard quality, insufficient usability, and profound linguistic inaccuracy, with the consequence of reinforcing the faulty belief that indexes provide little benefit beyond that possible with a simple search engine.

In this author's opinion, both sides are correct.

Cost-benefit analysis of both indexing approaches

The primary benefit of human indexing is quality. The primary benefits of computer-based indexing – which from here forward I will call "automatic indexing" – is speed and scalability. It is impossible to produce the highest-quality product in the least amount of time, and so the "best" choice of indexing approach depends very much on the scope of the project and the expectations of the audience, and little else.

When deciding how to approach the creation of an index, two questions must be answered. First, how much time is required to produce the index? Second, how forgiving is the audience regarding error? For purposes of this paper, let's define an error as "an unpleasant event as perceived by the user"; we'll explore different gradations of unpleasantness later.

Consider an example where keywords must be selected for four million documents. A computer might accomplish this analysis in a month. A human capable of indexing four documents per minute could complete the task in eight years; a team of ten indexers would need ten months. The computer has a 100-to-1 resource advantage over humans. However, let's also assume that the resulting index contains 100 times more errors (i.e., it's 100 times more likely to annoy its users). To determine which approach is best, then, a cost-benefit analysis can be performed. If a human work-hour were worth £30, and the damage caused by a single user annoyance were estimated at £3, then the automatic index is the best choice for this project, at ten-to-one.

Estimating the average cost of error is not a simple task. A single but unfortunate error in an important religious, legal, scholarly, or governmental document may have profound effects on the image of the producers. In a well-publicized story from 1999 [1], Microsoft Corporation was sued when an image of an African-American couple appeared as a search result for the word "monkey." Although this event was the result of neither error or malice – the full keyword is "monkey bars" – this widely read unpleasantness likely damaged the reputation of Microsoft and its product.

Additionally, human error differs from computer error. If a document's keywords were based on some standard kind of data, such as a company name, account number, or personal identification, errors caused by humans would be quite different in nature from errors caused by automatic indexers. Humans are more likely to create input or copying errors, whereas computers are more likely to misinterpret data. For example, a human might misspell the term "Baton Rouge," whereas a computer might misinterpret the term instead, treating this term as a person's name and not the name of a city. As another example, while humans and computers might both recognize the term "New York, New York" as a city as well as a song title, humans can better disambiguate based on context.

There is another valuable dichotomy between errors that are noticed and errors that are consequential but unnoticed. A noticeable error occurs when an index contains references to passing mentions, a common artifact of automatic indexing (in that computers have a difficult time judging the importance of an idea). With these references, the reader who attempts to locate information by using that entry will discover he has been mislead by a

frivolous inclusion. An unnoticed but consequential error is when important information does not appear in the index, such that a reader may never realizes that content unindexed nevertheless exists in the text. These errors are insidious. They confound quantification, require significant resources to correct, and individually can have disastrous effects. For example, an insurance company that fails to include mandatory language in a policy form can be forced to pay higher claims on that policy. [2]

The purpose of cost-benefit analysis is to find the least costly approach to creating a product of sufficient quality. We've seen two kinds of cost, resource costs and error costs, and it's clear that resource costs generally are higher for human indexing, while error costs often are higher for automatic indexing. However, it is certainly possible to combine human and automatic indexing into a single indexing approach, a "cyborg indexing" process, in which the weaknesses of each approach is mitigated by the strengths of the other.

Cyborg indexing: the printed book

For a single book, human indexing far surpasses automatic indexing in terms of quality [3]. Further, the cost of a single book index is relatively small as a production cost. Thus the cost savings from using an automatic indexing tool is insignificant compared to the precipitous drop in quality. Finally, the human resources required to "repair" a computer-generated index are very high, further discouraging any attempt at cost savings.

The fundamental technology behind automatic indexing tools hasn't changed in over ten years: statistical algorithms. The success of these algorithms is strongly dependent on the quantity of information being analyzed. In other words, the quality of automatic indexing increases only when there is more to index. Google, for example, uses ranking algorithms influenced by popularity among its visitors, and so infrequently results of searches demonstrate far lower quality. [4]

With book indexing, computers offer very little to improve the quality of human ingenuity. Spell-checking and auto-completion tools are perhaps the strongest artificial intelligence offerings available, despite their imperfections. Simple word-search tools can also provide assistance. Note that these tools provide an absolute minimum of analysis, in that they match (or fail to match) terms to other terms.

Even with additional knowledge supplied by humans, such as synonym lists, thesauri, taxonomy, and the like, automatic book indexing cannot succeed as long as the books are being indexed individually. This author estimates that at least several hundred separate books, on similar subjects for nearly similar audiences, would be necessary before the quality of automatic indexing might start to compete with the quality of what humans are producing on a customized, book-by-book basis.

Cyborg indexing: multiple print publications

As the sheer quantity of printed information increases, the probabilistic calculations of automatic indexing tools gain precision, increasing the resulting quality. As stated above, however, the human index is a custom product, and thus is likely to be notably better. For the casual user, the automatic index might suffice, but for readers of technical documentation, scholarly authorship, legal and government documents, etc., either the discrepancies of a computer-generated index will become apparent, or the consequences of discrepancies will be experienced.

With multiple publications, unnoticed errors are further camouflaged because they can exist in tandem with correct interpretations; for example, an index entry can have four references when it should have five. However, as the volume of documentation increases, the likelihood that more than one human indexer is required to write the index(es) increases as well. Inconsistencies among indexing teams, specifically in language and interpretation, are not uncommon, and additional resources are required to correct them.

Improvement to both the human and automated indexing processes is possible with the use of taxonomy-style tools; building the knowledge requires humans, but use of these tools can be automated. Synonym lists, for example, must be created editorially (by humans) but can be applied: entries for "unencrypting" can be automatically duplicated as entries for decoding, or else cross-references can be generated. (In fact, the automatic creation of cross-references is a customizable feature of some indexing software.) Thesauri and taxonomy, which provide both language and relationship control, can further enhance the utility of an index.

Attempts to combine automatic indexing applications with library science tools may improve the quality of the resulting index, but each tool requires the computer to interpret the knowledge and the language, adding error. In other words, not only can automatic indexing tools fail to interpret the text being indexed, but it compounds the problems by then incorrectly applying taxonomy knowledge. The resulting product is wholeheartedly better, but its fewer errors are more pronounced, serious, and difficult to repair.

Attempts to enhance human indexing with automated language tools can speed up the indexing process while increasing the possibility of more error, but human interpretation can remain the final step. When a computer attempts to do something, the process could require that a human being inspect each attempt before it is accepted, in a way similar to how spell-checking tools ask the user to accept, correct, or discard the computer's suggestions before continuing. Without final human intervention, computer-generated errors remain.

Cyborg indexing: hypertext documentation

With online documentation and databases, the biggest factor influencing the choice of human or computer indexing is the quantity of information. One must remember, however, that many Web sites contain far less information than a 200-page book; just because information is presented over the Internet does not guarantee a cost savings using automated tools. When the amount of information on a Web site is small, the arguments that favor human indexing for printed books are equally applicable here.

Other relevant environmental factors must be considered as well. For example, many Web sites have search engines. (Studies have shown Web users far prefer the search experience over browsing an online index [5], despite this additional burden.) The efficacy of a search engine can be improved by the application of keywords, taxonomy, and linguistic analysis, all of which are created by humans. Thus we can think of search as a low-grade automatic indexer, and the application of these other tools as human enhancements. Search engines that leverage the power of these "human" features are quite impressive; they fail only when the Web site extends beyond the scope of what humans write – when the content is created faster than the humans can maintain – or when the search tool itself is too difficult to use effectively. Finally, sometimes the inclusion of an accompanying index to a

searchable site has enough value to merit the expense of creating and maintaining it.

Another factor is that Web content is much more likely to change on a regular basis, perhaps even daily or hourly. The maintenance required exceeds what is possible with human indexers, and so an automatic indexing tool, including search, is often the only option. Still, there are many techniques for human intervention available. Taxonomy is the most powerful, because its features include language control, context definitions, hierarchical organizations, and manageable relationship types. As long as the taxonomy remains applicable to the site as the site evolves, the index can sustain its original quality over time. Other scalable tools involve community effort, such as Google's popularity algorithms. Perhaps the newest trend is community indexing or folksonomy (a merging of the words folks and taxonomy), also known as social classification. The production value of collaborative tagging comes with its use of a large, uncompensated community; in terms of quality, professionals are justifiably critical of their chaotic, idiosyncratic, and sometimes contradictory use of language. Even so, folksonomies can be edited by subject matter experts before being applied them to indexing. [6]

It is important to clarify that search and automatic indexing must both attempt to interpret the value of information; in this, they are almost identical tools. (Choosing which documents to write keywords for is a human intervention that can improve this process.) However, while search engines must rank these values from "most relevant" to "least relevant," automatic indexers need only identify topics that surpass some value threshold, and are therefore worth including in its index. Further, automatic indexers select language for the references to that content, and then organize that language in alphabetical, hierarchical, and/or other meaningful structures. Because automatic indexers perform much more interpretive work than search engines, there is greater potential for both error and human intervention. Finally, the index produced by an automatic indexer is a browseable tool, whereas the interfaces for search engines often provide no language assistance to readers, increasing the user's burden.

Humans and computers in harmony

Indexing is fundamentally a human endeavor because it requires three levels of interpretation that come naturally to human beings: recognition of value, application of words, and organization of ideas. And while computers can be programmed to mimic these human skills, they will never surpass what humans are capable of in this realm. Still, computers can perform other kinds of tasks that far surpass what humans can accomplish within a reasonable amount of time – which is why we live in such a computerized world today. To insist that human indexing is the only approach is folly, but to believe that computer indexing is better is wrong.

The best solution, clearly, is a complementary approach – a cyborg system – that leverages the best that human and computer systems can produce. Human and computer indexers are both capable of error, and so getting both systems to work together effectively requires careful analysis of the risks. Measuring the possibility and consequences of all indexing errors from both indexing approaches is a key variable in designing a cyborg indexing system.

References

[1] "Microsoft sued for 'racist' application." http://news.zdnet.co.uk/software/ 0,1000000121, 2072468,00.htm (among others).

[2] The Hartford Financial Services Group. Personal communication.

[3] *Mulvany*, Nancy; *Milstead*, Jessica: Indexicon, the Only Fully Automatic Indexer: A Review (1994). www.bayside-indexing.com/idxcon.htm.

[4] *Blachman*, Nancy: How Google Works. In: GoogleGuide. www.googleguide.com/google_works.html.

[5] TechCOMM 51:2 (May 2004).

[6] Folksonomy, at Wikipedia. http://en.wikipedia.org/wiki/Folksonomy.

Indexing, Automated Indexing, Cost, Quality

Inhaltliche Erschließung, Qualität, Kostenbewertung, Maschinelles Indexierungsverfahren

The Author

Seth Maislin

is a managing partner of Potomac Indexing, an adjunct instructor at three colleges in Massachusetts, and author of the online course, Writing Indexes for Books and Websites. He is an indexer, information architect, and taxonomist who has consulted for companies and institutions including Mercedes Benz, the United Nations Library, Microsoft, Elsevier, and The Hartford. Seth is a former president of the American Society of Indexers and founder of the techindexing mailing list.

Potomac Indexing
11 Quincy Street
Arlington, MA 02476-6031
USA
seth@maislin.com
http://taxonomist.tripod.com
www.potomacindexing.com

The Indexer: past, present and future

Maureen MacGlashan, Largs (UK)

The (British) Society of Indexers (SI) was established in 1957, its journal, The Indexer, following in 1958. At first little more than the SI house journal or newsletter, it developed over the years into what it now proudly calls itself: 'the international journal of indexing', published on behalf of all the indexing societies. It has grown from the 28 pages of its first issue to the 72 (plus 16-page supplement) of today, and has moved from the cutting and pasting that comprised so much of the early editorial task to the prospect of online publication in 2008. Despite predictions from time to time that there was a limit to how much could be usefully written about indexing, the flow of good material has now continued unabated for some 50 years, always topical but always representative of its time. From the beginning, The Indexer has aspired to the very best contemporary journal practice. That remains the challenge for today.

The Indexer: Vergangenheit, Gegenwart, Zukunft

Die britische Society of Indexers (SI) wurde 1957 gegründet, ihre Zeitschrift The Indexer *folgte 1958. Was am Anfang nicht viel mehr als eine verbandsinterne Zeitschrift bzw. einen Newsletter darstellte, entwickelte sich über die Jahre hinweg zu etwas, was heute mit Stolz den Untertitel „die internationale Fachzeitschrift des Indexing" trägt, herausgegeben im Namen aller Fachverbände, die sich dem Indexieren widmen. Sie wuchs von den 28 Seiten ihrer ersten Ausgabe zu den heute 72 (plus 16 Ergänzungsseiten) und entwickelte sich vom Klebeumbruch mit Papier, Schere und Kleister, der mit viel Aufwand verbunden war, hin zur möglichen Online-Publikation im Jahr 2008. Trotz zeitweiliger Voraussagen, dass das, was über Indexieren und Registermachen geschrieben werden könnte, bald erschöpft sein würde, hält der Strom guten Materials seit 50 Jahren unvermindert an, stets mit aktuellen, zeitgemäßen Themen. Seit den Anfängen strebte The Indexer die allerbesten zeitgemäßen Standards für Fachzeitschriften an. Dies bleibt auch heute die Herausforderung.*

The Society of Indexers (SI)

1957 is an important date in the history of indexing, for this was the year which saw the establishment in the United Kingdom of the Society of Indexers, so far as we are aware the first such surviving Society to see the light of day. It was greeted with a fanfare in *The Times* (then the British newspaper of record) of 8 May 1957:

> There are far too many societies. But plenty of people will readily supply a list of half a dozen that can be dispensed with in order to make room for the newly formed Society of Indexers. Here is a necessary body if ever there was one. The first of its aims should surely be that every non-fiction book should have an index … There should be plenty of material for the 'papers and notes on the subject of indexing' which the Society hopes to issue. They should make interesting reading. But without any disrespect to future authors it is safe to say that the part of the Society's publications which will be most eagerly scrutinized will be its own index.

From newsletter to international journal

Over the years the Society has indeed issued various 'papers and notes' on the subject of indexing but, for present purposes, the most relevant of these is undoubtedly *The Indexer*. Volume 1, issue 1 appeared in March 1958, in a smaller format than we use today (at 180 x 235mm approximating to today's B5) and with just 28 pages. It too opened with a fanfare, this time in the shape of a letter from the then Prime Minister, Sir Harold Macmillan. He writes:

> I am glad … to send a message to the Society for the first issue of *The Indexer*. …
> In summoning up remembrance of things past, I can recall indexes in great variety, from the humdrum but invaluable list of names and pages, to those so majestic that they occupied a whole volume to themselves … I know, too, how often an author, whether he knew it or not, was indebted to an indexer for pointing out errors, discrepancies, or repetitions that had otherwise escaped detection in the proofs.

It is perhaps not surprising that he quoted with particular relish the instruction reportedly given by Macaulay (one of the great British historians, but not of the Tory Prime Minister's political persuasion): 'Let no damned Tory index my "History"'. Like *The Times* leader of the previous year, Macmillan saw every prospect that the Society of Indexers would have the opportunity of doing 'valuable work in upholding the standard of indexing and in bringing its practitioners into touch with those who require their services'.

Fifty years on, *The Indexer* has developed from those tiny beginnings as 'the Journal of the Society of Indexers' (1958–1971), turning by 1979 into 'the Journal of The Society of Indexers and of the affiliated American & Australian & Canadian Societies'. In 1999 it became quite simply 'The International Journal of Indexing'. The British Society of Indexers retains responsibility for publication of *The Indexer* and for the appointment of the editor, but it does so on behalf of all the affiliated indexing societies who are entitled to a say in policy, and whose members figure increasingly as contributors, as guest editors and as members of the production team. The International Agreement of indexing societies can be found at www.theindexer.org/files/25-3cp2_16.pdf.

Setting the pattern

The 28 pages of that first 1958 issue of *The Indexer* had four pages of editorial material (including *The Times* leader and the Prime Minister's letter), a couple of articles offering advice on indexing techniques, a book review and two pages of 'No index – no comment', a precursor of one of today's most popular features, 'Indexes reviewed', which collects together instances of books identified by reviewers as offering shining examples of good indexes or as falling short on this count. The remaining 8 pages were devoted to the business of the Society, including the Constitution and a full list of the 150 or so Society members.

The second issue (September 1958) was rather longer, coming in at 38 pages, with just one page of editorial, but 13 pages of

Society business including the minutes of its first Annual General Meeting and a very full report of the first Society training course, held in London and attracting more than forty participants. The journal also carried articles on the purpose of indexing, on index typography, and on the report of the American Standards Association subcommittee on indexing. And it saw the first Letter to the Editor.

The third issue saw some all-too-familiar soul-searching, recording (Report of the Council for the year ended 31 March 1959) that 'The usual difficulties that attend the production of a new periodical have had to be met; these have, unfortunately, caused some delay … The cost of production has been an anxiety, but arrangements have now been made by which this will be very considerably reduced.' (What the arrangements were is not indicated.) Subscription rates at that time were $1.50 (including postage) and bound copies of volume 1 (four issues: 1958–59) were available at $5.50 (including postage). Perhaps it is not surprising that costs were a matter of concern (or rather finding a way of meeting them) if, as it would seem, it was cheaper to buy 4 bound copies than 4 single copies.

Scope and coverage

Those who knew no better (and, from my own experience as editor, still know no better) saw no future for a journal dedicated to indexing. Again and again, editors were warned: 'You'll never be able to keep it up; you'll find that by the end of another year you have completely exhausted all the possible aspects of indexing.' Perhaps, but as G Norman Knight, in reporting one of these warnings in 1969 (*Indexer* **6** 145) went on to say (in alphabetical order, you'll note):

> Events have flatly falsified that prediction, because, like chess, the variations of indexing seem inexhaustible. Since then, *The Indexer* has contained a multitude of articles, weighty or gay, on such varied topics as: archive indexing; chain indexing; children's books indexes; citation indexing; computer-produced indexes; cumulative indexing; the design of indexes; documentary indexing; encyclopaedic indexes; foreign surnames; the future of indexing; humorous indexes; legal indexing; masterpieces of indexing; medical indexing; one index or more than one; scientific and technical indexing; telephone directories; the typography of indexes. And this list is nothing like complete.

The range has remained much the same over the years but with new slants on old problems, and new responses in the light of new technology. There seems no more reason now than in 1969 to think the well will run dry. The amount of good material being presented during my time as editor has meant the journal has shot up from the norm of 56 pages or so I inherited to 72 or even 80 pages. In addition, most issues now carry a 16-page supplement, the *Centrepiece*, covering in detail a particular aspect of indexing in which the 'rules' are far from clear. The three *Centrepieces* which have appeared so far deal with the matter of non-English personal names (Aboriginal, African click languages, Chinese, Ethiopian, French, German, Dutch and Afrikaans, Italian, Spanish, Tibetan, Turkish). Other languages and other subjects are under consideration.

From time to time the editor of the day has clearly opted for a themed approach: in my view this works better in theory than in practice and seems never to have been sustained for long. What is both editorially easier and, in my view, more interesting, is to draw the theme from the material available for any given issue: juggling the articles into a cohesive order is one of the most challenging but also the most rewarding of editorial tasks.

'Themes' also tended to emerge from the appointment, from 2000, of Issue or Guest Editors. As the then editor (Christine Shuttleworth) wrote in introducing the scheme (*Indexer* 21 153)

> For each issue an individual or group, who may be members of the Society of Indexers or of one of the associated Societies in the international group, will be appointed by the Executive Editor to commission articles, but without responsibility for production. We hope that this new arrangement will further increase the international interest and appeal of the journal.

All the societies and indexing networks other than the China Society of Indexers have now taken their turn at guest editing, the German and Netherlands Networks taking responsibility for the October 2006 issue (*Indexer* **25**(2)). This last, as would be expected, had a strong focus on indexing outside the Anglo-Saxon world. The October 2005 issue (*Indexer* **24**(4)) was edited by a group of first-time attendees at the British Society of Indexers Conference in April 2004 and naturally leaned heavily towards the experiences and interests of the new indexer – this has proved a particularly popular issue. The great advantage of an issue editor approach is that it extends *The Indexer* network and opens up possibilities for all sorts of new ideas. But prospective guest editors are slow to come forward, largely because they fear what will be involved. They need not. All the production side of things is handled by the 'resident' team: all the guest editor needs to do is to persuade half a dozen contributors to write an article, and then nurse them through to submission on schedule. The invitation to guest edit is on the table, the best time scale being agreement to take on an issue two years ahead of its appearance: experience shows that it takes that sort of time to get all the bits in place.

And what matters above all in terms of the spread of articles is that there is something for everybody, something to read immediately over a cup of coffee (usually, I suspect, 'Indexes reviewed'), something for the beginner, for the specialist, something that is of no present interest to the reader, but comes into its own two or twenty years later when it is just what you need for your current indexing project. And it is particularly pleasing to get reports of the interest with which non-indexers read the journal. Next challenge: to turn interested readers into paid-up subscribers.

The Indexer: a window on sociological changes

The Indexer offers an interesting window on the changes in the indexing and the wider world. Elizabeth Wallis writes (*Indexer* **25** 229):

> I remember the overwhelming maleness of the Society's officers; now it is the exact opposite, almost exclusively female. The men all seemed like elderly clubbable gentlemen to me, and so many freemasons among them.

This indeed is exactly what all those fusty photographs from earlier issues seem to say, and the patronizing tone used toward some of the women involved in the journal would not be well-received today. (But I wonder what our successors 50 years hence will be saying about us?) Gender bias in indexing came up quite frequently in the pages of *The Indexer*, as did various job creation schemes in Britain and Canada, designed to turn the unemployed into working indexers. And I cannot resist quoting the following identification of indexing as a suitable task for the literate convict.

> Ninety-six years ago in The Nation [an American news weekly], a writer calling himself Anobium Pertinax from Johns Hopkins University set the stage with this blast which the editor titled

'New Field for Convict Labor'. He said, 'Let all convicts who can read and write be set, under competent supervision, to indexing books; and let those who cannot, receive the necessary instruction as soon as may be.'
Pertinax then stated these two claims for his scheme; 'That it will not conflict with the interests of any class of laboring persons, or at least any that has a claim to consideration,' and, 'That the kind of labor proposed is peculiarly suited to the reformatory idea, being incomparable for teaching order, patience, humility, and for thoroughly eradicating the last trace of the Old Adam in whoever pursues it'. (Drazan, *Indexer* **12** 27)

Indexers and the brave new information retrieval world

One might, of course, expect more recent issues to have more extensive coverage than earlier ones of computer-related topics and more recognition of what is going on in the closely related world of information retrieval. Certainly, every recent issue has had several items on such topics, ranging from 'Misbehaving computers' (Jermey, *Indexer* **25** 195–6) to 'Annotating document content: a knowledge management perspective' (Ciravegna and Petrelli, *Indexer* **25** 23–7) (too difficult for some of us), from 'Web indexing: extending the functionality of HTML Indexer' (Unwalla, *Indexer* **25** 238) to 'Google and beyond: information retrieval on the World Wide Web' (Northedge, *Indexer* **25** 192–5). But the fact is, discussion of the role of the computer in indexing has been on the Indexer agenda from the very beginning. Volume 2(1) (Spring 1960) carried a couple of articles on early attempts at mechanized indexing: 'Mechanized indexing of information on chemical compounds in plants' (Claridge, *Indexer* **2** 4–17) and 'Some procedures in indexing' (Holmstrom, *Indexer* **2** 17–30). Both considered the possibility of computer-aided indexing, but in the end (while adopting some features of what a computer could offer in terms of data ordering and presentation) found the 'feature card' system ('Peek-a-boo' in the United States, 'Sichtlochkarten' in Germany) best met their needs. A year later, there was a short report (*Indexer* **2** 123) on the US National Library of Medicine's similar practice.

Mechanized or machine indexing came up again in the form of a review (*Indexer* **3** 158–60) of *Machine indexing: progress and problems* (a collection of conference papers on the subject). Among the many problems identified (we are in 1963) is the difficulty of input into the machine.

> Assuming technical advances make it economic to input a complete book, the computer will have considerable difficulty in understanding it … One major obstacle … is cost. The only people who can afford electronic systems are the big manufacturers and universities … and the US Government … The electronic systems, of which the computer is probably the cheapest, are still uneconomic.

In Volume **7**(2) (Autumn 1970) Theodore Hines and Jessica Harris discussed ('Computer-aided production of book indexes', *Indexer* **7** 49–54) a program they had developed at the Columbia University School of Library Service:

> The identification and expression of indexable matter remain the province of the human indexer. But much of the handling of the entries, of repetitive typing, and of the styling of entries … has been shifted to the broader back and faster fingers of the computer. At 1,000 entries per minute filing time and at about 20 pages per minute formatting time, a computer can give you quite a lot … We do not think any system we or anyone else may devise can replace any professional indexers. We are convinced, however …'

Volume **8**(3) (April 1973) had several detailed accounts of the use of a computer in the production of an index. Brenda Hall, writing about the production of the index to the *Cartographic Journal* ('A computer-generated index technique', *Indexer* **8** 130–8) concludes:

> For very complex indexes, particularly those dealing with very diverse experience, derived over a long period of time, experience … has shown that at present it needs a cumbersome system of controls, and at the preparation of the input stage it certainly did not save any indexing time …. Even the punching of cards took longer than the preparation of a type-script because mistakes took longer to rectify … the program is a considerable *tour de force* … most of the remaining problems could be removed … to bring the end product closer to the standard set by the accurate excellence of the best manually produced indexes.

Michael Green (from the New York-based Lawyers Co-operative Publishing Company) described a system of indexing using tape recorders and punch cards (*Indexer* **8** 139) and Peter Thomas ('The use of KWIC to index the proceedings of a Public Inquiry', *Indexer* **8** 145). KWIC (Keyword in Context), developed in the late 1950s, was chosen as apparently obviating the problems of manual indexing (sorting all those 58,000 cards into alphabetical order!) and of the punched-card system (because of the key-punching time). Volume **13**(1) (April 1982) carried a full and careful review (Purton, *Indexer* **13** 27–31) of the available equipment and discussed the exciting possibility of receiving proofs on a floppy disk and returning the completed index in the same way.

The first advertisement for an early version of MACREX, still one of the leading dedicated indexing software programs, appeared in Volume **13**(2) (October 1982):

> COMPUTER-ASSISTED INDEXING
> Thoroughly tested and documented programs to run on microcomputers using the
> CP/M operating system
> AVAILABLE NOW
>
> **Capable**
> Close co-operation between the programmer and an indexer with ten years' experience of manual indexing has produced a package which takes account of the specialised needs of the professional indexer. Layout follows BS3700: 1976.
>
> **Easy to use**
> No technical knowledge is needed. Clear and simple instructions guide the user at all stages; there are no disasters if the 'wrong' key is pressed; uppercase and lower-case letters and all punctuation marks can be used in entries without any special codes.
>
> **Fast**
> Sorting time is negligible; corrections, additions and deletions are made in seconds; typing time can be reduced substantially; merging of identical entries with different page numbers is accomplished automatically; temporary page numbers are quickly transformed to the correct final page numbers when they are known.

'Natural' language crept in in Volume 19(1) (April 1990) (Korycinski and Newell, 'Natural-language processing and automatic indexing'), was there again in Volume 25(2) (October 2006) (Zargayouna et al., describing IndDoc, a French Natural Language Processing-based programme to assist the indexer) and featured in the lively exchanges between authors and indexers on the subject of indexing, first published in the *Times Literary Supplement* in November 2006 and reprinted in *The Indexer* **25**(3) (April 2007).

And, of course, in due course, discussion turned to possibilities for Internet publishing, with T.G. McFadden warning in October 1994 (*Indexer* **19** 81) that not only did no indexing exist for the Internet or was likely to do so in the near future, but that it was most unlikely that scholars and researchers would rush to publish their findings on the Internet as an alternative to traditional print journals – it had been grossly oversold. Well, that was 13 years ago and how times change, with journals (including *The Indexer*) falling over themselves to go electronic. We all know that if we don't, we are doomed, and no scholar now could conceive of not publishing on the Web or not having continuous on-line access to what his peers have written.

Editing *The Indexer*

The Indexer has had just eight editors in its 50 years, Harold Smith (1958–9) (resigned 'possibly in pique', suggests Monty Harrod (*Indexer* **4** 144) when his favorite feature 'No index – no comment' was discontinued following a near brush with the law), John Thornton (1959–63) (repeatedly lauded for his competence as editor), Monty Harrod (1964–78) ('somewhat strait-laced and Victorian' says Elizabeth Wallis, his assistant editor (*Indexer* **25** 229), Hazel Bell (1978–95) ('facilitator, innovator, gatekeeper, subversive' according to Janet Shuter, *Indexer* **29** 1) and then Janet Shuter and Nancy Mulvany (1996–9), Christine Shuttleworth (1999–2004) and myself (Maureen MacGlashan, since 2004). We've all had our own style, our own limitations and constraints, our own ambitions. The early editors all seem to have struggled not so much to identify suitable subject matter, but to find someone to write about it and to deliver on time. There was a repeated plea: 'If only you would write for us, we could publish on a quarterly basis'. John Thornton describes it thus (*Indexer* **4** 99–105):

> … Almost every article was prised out of the author. Many friends were approached, many promises were made … there was no possibility of publishing four issues a year. As it was, I had to solicit shorter items as fill-ups, provoke correspondence, write most of the reviews, and fill in the pages with lists of references on indexing, and similar items. … once having obtained the typescript [the editor] hangs on to it like grim death.

No doubt he was not the only editor to suffer in this way, or to wonder, as he faced an empty in-tray, whether he would have enough material to make more than a pretence of the next issue, but certainly during my time there has been no shortage of quality material. Of course, confidence that there is an inexhaustible supply of worthwhile material to write about is not enough – the challenge is to tap into it, and turn potential into reality on the page. But that's what the editor's task is all about.

John Thornton also describes the travails of production and distribution:

> We look forward to perusing issues that we have not read in manuscript, edited, proof-read, pasted-up, re-read in page form, and despatched. These duties have been lightened by the ever ready co-operation of our printers … and the assistance from my family in pasting address labels on envelopes, stamping them, and inserting the appropriate number of copies. (valedictory editorial, *Indexer* **3** 134–5)

This vision of paste-pots and stamp-licking families and friends comes up again and again over the years, and for some of the people associated with *The Indexer* in its earlier days seems to be their chief memory. Fortunately, by the time I took over, this cottage-industry approach was behind us and I have a first-rate production team to take most of the pain of getting the journal onto the printed page away from the editor. I am hugely grateful to them all for their skill and their patience.

The Indexer index

It goes without saying that *The Indexer* had to have an index, and an index of quality. For many years, the practice was to invite distinguished members of the Society, many of them Wheatley Medal winners, to undertake this task. 'Editorial policy', wrote the Chairman of the Society of Indexers (John Gordon, *Indexer* **13** 253), 'was to give each volume-indexer extensive freedom to display his/her personal virtuosity. Each compilation has its own special qualities, revealing fascinating variations of style, of technique, of decision-making: it has indeed been said with pride that these indexes are living proof that indexing is an art rather than a science.' But such virtuosity did not necessarily serve the user well, and certainly did nothing to simplify the preparation of a cumulative index, something to which the Editorial Board had been giving much thought. So, following consultation with readers and with a panel of experienced indexers, a clear-cut house style was put in place to which future *Indexer* indexers would be asked to conform. Readers could now look forward with some pride, Gordon concluded, to an era in which indexing *The Indexer* would exemplify in model fashion the application of the established principles of the indexing of periodicals.

This remained the ambition and the practice until I took over as editor, though not much progress had been made on the preparation of a cumulative index, the one attempt to prepare such an index, to mark the journal's 40th anniversary, running into the sort of problems which tend to beset a project of this size and complexity. My own surprise, on taking up the editorial role and finding myself responsible for finding *Indexer* indexers, was to discover that the index was not being prepared on a cumulative basis, with each new volume being prepared within a cumulated version with the relevant entries simply extracted for publication as the index to the volume in question. This had been my own approach for several years to various serial publications on which I worked and seemed the obvious way to go about it. As an interim solution, I decided to take on responsibility for indexing the journal myself, and as part of that task to begin work on compiling a cumulative index. To the extent that the style and coverage of the indexes over the years was indeed homogeneous, as the 1983 Editorial Board hoped, simple merger of the existing indexes should largely suffice, but the less homogeneous the material, the more likely it was that much re-indexing would be called for. This is indeed the case, with re-indexing being the more necessary the further back I go. This work of cumulation has now been completed for volumes 20–25 (1996–2007), with the index being updated with every new issue.

But in adopting this approach, the other major change I instituted was to abandon the volume by volume indexes entirely (and indeed any printed version of the index), simply posting the cumulative version to the Website. There were one or two protests in advance, but no complaints since. Doing it this way is something which has, of course, become possible only in the very recent past. The advantages seem to me very considerable: a cumulative on-line index should offer greater consistency, volume on volume, and can be more comprehensive, in particular making it possible to add or amend the index continuously as hindsight notices error, omission, or lack of user-friendliness. An on-line version offers the user the possibility of approaching it as a typical back-of-the-book index, or by a simple search for a word or concept. Cross-references in *The Indexer* index are already linked to the preferred entry, and in time the locators themselves will also be linked to the

article to which they refer, a system which will come into its own once *The Indexer* itself becomes available on-line.

An on-line index has the advantage that there are no real constraints on length. This means that it would, for example, be possible to go in for a lot of double-posting (i.e., simple duplication of entries under alternative headings) but that makes for a lot of extra work and leaves the risk of introducing inconsistency. Where I am very generous is in a) listing as many terms as possible which users might make their entry point, with appropriate, linked, cross-references; and b) multiple, cross-cutting, listing of an item (so that, for example, an article will appear by title, by author, and (once or more) by subject-matter).

And the on-line index is, of course, available to subscriber and non-subscriber alike.

The future

The changes in place for 2008 include:

1) A change to the pattern of volume numbering from a two-yearly basis to annual numbering. It took us 50 years to complete 25 volumes – it will take just 25 years to complete the next 25. The easiest of changes to make, but it should make life simpler for editor and subscriber alike.
2) A switch from twice-yearly to quarterly publication. As indicated above, this has been a long-standing aspiration, baulked at first by want of material, and then later by cost considerations, in particular the cost of postage. Postage still remains a factor and will need to be taken account of in the subscription price. Lack of material does not (see above). Institutional subscribers are said to prefer a quarterly publication. Individual subscribers have said they would prefer less more often. And from the editorial point of view, it is much easier to manage the flow of material and to give the journal a topical feel. It doesn't matter nearly so much not making it into one issue if the next is just three months away rather than six.
3) *The Indexer* has now been fully digitized with a view to putting it on-line, fully and freely accessible up to two years retrospectively. All that is now needed is to secure permissions to publish electronically, a task which we hope to complete by September 2008.
4) The aim is also to complete a fully-linked cumulative index to the whole run from 1958 to 2008.
5) A new feature on the Website is the Readers' Forum, a sort of ongoing letters to the editor page in which subscribers and potential subscribers now have a chance to air their views on journal content, presentation and policy.
6) We are moving to on-line hosting (Ingenta-Connect), giving subscribers on-line access to current issues (i.e., issues less than two years old), and non-subscribers the chance to buy single articles should they so wish.
7) The print version will continue in parallel with the on-line version, a strong preference amongst individual subscribers. Institutional subscribers, whose preferences differ, will be able to choose either or both.
8) Subscription rates have been re-structured. As in the past, members of the affiliated indexing societies will enjoy a preferential rate (£26 or EUR40 for four issues including postage), paid either as part of their society membership fee (Society of Indexers and Indexing Society of Canada) or through direct subscription to *The Indexer*. The subscription rate for individuals who are not members of an indexing society is £ 40 or EUR 60, in fact a reduction on subscription rates of past years. Institutional rates are £ 120 or EUR 180 online or print only, £ 150 or EUR 225 online *and* print.

To subscribe visit *The Indexer* Website: www.theindexer.org or contact the Subscriptions Manager: subscriptions@theindexer.org; telephone: +44(0)114 244 9561; fax: +44(0)114 244 9563.

In my first editorial, in April 2005 (*Indexer* **24**(3)) I said that I had no plans for radical change: what changes there might be would be evolutionary and consultative. Although in some ways *The Indexer* of 2008 will be very different from that of 2004, most of the changes have been on the agenda for some time, in some cases (such as going quarterly) for a very long time indeed, and have only been made possible (or, indeed, forced upon us) by the dramatic technological changes of recent years. Even as recently as 2005, the cost of digitizing *The Indexer* seemed likely to be far beyond our budget: two years later the estimate was a mere one 10th of the original figure. What some of our predecessors (remember McFadden?) thought was unimaginable has quite suddenly become the norm, the industry standard. All the changes we are making have been designed to serve our readers, to enhance the editorial and production process and to bring *The Indexer* into line with the best contemporary practice amongst scholarly journals. Continuing to do this will be our challenge for the next 50 years.

The Indexer, Journal, History, Scope, Index, United Kingdom, Society of Indexers

Zeitschrift, The Indexer, Geschichte, Körperschaft, Entwicklung, Register, Society of Indexers, Großbritannien

THE AUTHOR

Maureen MacGlashan, MA LLM

read law at Cambridge University (1957–61), served in the British Diplomatic Service from 1961–98, taking four years out from 1986–90 to help establish the Lauterpacht Research Centre for International Law in Cambridge where she began her indexing career, cutting her teeth on one of the most difficult of indexing challenges: preparing a cumulative index to 45 volumes of law reports, a task on which she is still engaged though the volumes now stretch to 135. She served as President of the Society of Indexers from 2001 to 2004 and has been editor of *The Indexer* since 2004.

16 G Main Street
Largs, Ayrshire KA30 8AB
United Kingdom
td64@dial.pipex.com

Training in indexing: the Society of Indexers' course

Ann Hudson, Chichester (England)

Good indexing needs a variety of talents and skills. Although some of these are inborn, training in indexing principles and techniques is essential for success in the indexing profession. The Society of Indexers' training course, leading to the qualification of Accredited Indexer, is based on British and International Standards and consists of four Course Units. Assessment is through four formal tests, and trainees also have to undertake Online Tutorials and a Practical Indexing Assignment. Additional support is provided for trainees, and a programme of workshops caters for both trainees and working indexers.

Fortbildung für Indexierer: Der Kurs der Society of Indexers

Gutes Indexieren erfordert verschiedene Begabungen und Fertigkeiten. Obwohl einige davon angeboren sind, ist ein Training in den Prinzipien und Techniken des Indexierens unentbehrlich für den beruflichen Erfolg als Indexer. Der Lehrgang der Society of Indexers, der zur Qualifizierung eines akkreditierten Indexers führt, basiert auf britischen und internationalen Normen und besteht aus vier Lehrgangseinheiten. Die Beurteilung geschieht durch vier formale Tests, und die Lehrgangsteilnehmer müssen außerdem Online-Tutorials durcharbeiten und einen praktischen Indexierungs-Auftrag übernehmen. Für die Lehrgangsteilnehmer steht zusätzliche Unterstützung zur Verfügung und ein Workshop-Programm ist auf die Belange der Lehrgangsteilnehmer und der beteiligten Indexer zugeschnitten.

What qualities does an indexer need?

Indexing is not, as many people suppose, a mechanical task which can equally well be done by a computer. Essential qualities for a successful indexer include the ability to read and understand complex text, skill in organizing material, pleasure and satisfaction in creating order out of chaos, and a high level of accuracy. Some people are born with such qualities; others are not, and will never make good indexers or (equally important) enjoy the work. According to Nancy Mulvany, '… the ability to analyze text objectively and accurately and to produce a conceptual map that directs readers to specific portions of the text involves a way of thinking that can only be guided and encouraged, not taught.' (1)

Indexers also require both common sense and imagination in order to create an information retrieval tool for users with whom they usually have no contact. As David Crystal has written:

'From the point of view of discourse analysis, indexing is truly a strange behaviour. It is a task where the compiler is trying to anticipate every possible query about content which future readers of a publication might have. Indexers are in effect trying to provide answers to a host of unasked questions – and this in itself is an interesting reversal of communicative priorities. Normally, we wait for a question before we try to answer it. Indexers therefore need to work as if their audience is present. But there are two snags: first, in most cases they do not know who this audience will be; second, in most cases they do not receive any feedback as to whether their judgements have been successful. From a communicative point of view, there is probably no more isolated intellectual task than indexing.' (2)

Why is formal training necessary?

Even someone with excellent inborn indexing ability needs training. Pat Booth, author of the widely used textbook *Indexing: the manual of good practice*, (3) writes:

'Doing it well requires certain aptitudes as well as general knowledge of the kind possessed by anyone with a good education and an inquiring and retentive mind. To these have to be added a knowledge of indexing techniques and the ability to apply them in different indexing contexts. This is where training (in the sense of systematic instruction and practice, aimed at achieving and maintaining proficiency) comes in.' (4)

A trainee indexer must learn the current accepted practice for indexing, encapsulated in British and International Standard guidelines, particularly International Standard BS ISO 999: 1996, *Information and documentation – Guidelines for the content, organization and presentation of indexes*. (5) Indexers need to know not only about traditional back-of-the-book indexing, but also the indexing of periodicals, newspapers, websites and databases, and they must understand the publishing process and how indexing fits into it. They need some knowledge of design and layout and familiarity with computer technology; publishers require indexes to be supplied in electronic form and often supply text for indexing as pdf files. Most indexers use dedicated indexing software such as MACREX, CINDEX or SKY; (6) this handles the drudgery (for instance, sorting into alphabetical order and ensuring consistent punctuation and spacing) and leaves the indexer free to concentrate on the more intellectual aspects of the job.

The role of the Society of Indexers in training

The Society of Indexers (SI) is the oldest indexing society in the world; founded in 1957, it celebrated its 50th anniversary in 2007. It has around 700 members, most of whom are practising freelance indexers. As the professional society for indexing in the UK, SI regards training provision as one of its most important responsibilities. Jill Halliday, former SI Director of Professional Development, has explored the meaning of 'professionalism' for the indexer:

'Education requires a framework, which can be established by a training curriculum, and allows professionals to continue to grow in knowledge and expertise during their working life. High standards of intellectual knowledge can only be gained if the required education and training is successfully completed. To achieve this, professional individuals and their professional bodies must be prepared to invest in training and qualifications.' (7)

Training in indexing has been part of SI's remit since it was founded. From 1958 SI organized regular courses of lectures for trainee indexers, and since 1973 it has been involved in training by correspondence course, at first in conjunction with an outside body, the Rapid Results College. SI's own distance learning course, 'Training in Indexing', was launched in 1988. The course is administered from SI's offices in Sheffield and the assessment process co-ordinated by SI's Training Course Co-ordinator.

The current (third) edition of the SI course, published in 2002–3, consists of four Course Units (8) which are supplied both in hard copy format and on CD-ROM, so that trainees can study in whatever way suits them best. Each Unit, which on average takes about 45–50 hours of study, contains a self-administered test so that trainees can check their progress before undertaking the formal tests (one for each Unit). The CD-ROM version also includes extra electronic self-assessment exercises, and has hyperlinks to a glossary, reading list and other relevant sections.

Anyone may purchase the Course Units, but only members of SI may take the formal tests which lead to the qualification of Accredited Indexer, entitling them to an entry in SI's 'Indexers Available' online directory, accessible via the SI website (www.indexers.org.uk) and widely used by editors seeking indexers. Accredited Indexers are also allowed to use the SI logo on their headed notepaper and other stationery. The Betty Moys Prize, funded from a legacy bequeathed to SI by the well-known legal indexer Betty Moys, is awarded annually to the best new Accredited Indexer. Once they have at least two years of working experience Accredited Indexers are encouraged to progress to SI's highest qualification, Fellowship of the Society of Indexers.

Course outline

The course content is based on British and International Standard guidelines, particularly BS ISO 999: 1996.

Unit A (Indexers, users and documents) is an introduction to the function and characteristics of indexes and to how indexers work. It covers:

- basic indexing terminology
- the functions and characteristics of indexes
- what users want from indexes
- what kind of people make good indexers
- the indexing process
- the role of authors and other document originators
- document production and categorization
- creating bibliographic references

Unit B (Choice and form of entries) gives more detailed guidance on the intellectual processes involved in indexing. It covers:

- selecting concepts for indexing
- choosing appropriate index terms
- forming headings and subheadings
- indexing proper names
- presenting locators (page numbers)
- devising cross-references

Unit C (Arrangement and presentation of indexes and thesauri) deals with the principles of index arrangement and the requirements of specialized types of indexing and thesaurus construction. It covers:

- rules for arranging index entries (filing order)
- compiling multiple sequences of indexes
- team and cumulative indexing; book and journal production
- index layout and presentation
- thesaurus use and construction

Unit D (The business of indexing) discusses good practice in establishing and running a freelance indexing business. It covers:

- starting up
- finance
- commissions and contracts
- customer relations
- legal matters
- developing the business
- the activities and services (to members and publishers) of the Society of Indexers

The assessment process

The standard required to pass the four formal tests is deliberately high, as the aim is to enable successful candidates to enter the commercial world and to practise competently. To give trainees practice in working to deadlines, there is a time limit for completing each test. The tests, each of which includes compiling a short index, are marked by a team of qualified and experienced members of SI; to maintain confidentiality and impartiality, candidates and markers are anonymous to each other. Marks are not divulged, but all candidates (whether successful or unsuccessful) receive helpful written comments.

More than 100 trainees are actively engaged in the SI training course at any one time. Whilst it is possible to complete the course within a year, most people take longer; up to five years is allowed, making it possible to combine training with other commitments. Many trainees are in full-time work and seeking to go part-time or completely freelance; others may have family commitments. Not all trainees will complete the course: some discover that indexing is much harder than they expect and fail to get through the tests, while others drop out for personal reasons or decide that indexing is not the right career choice for them after all.

No-one can become a good indexer just by studying and taking tests. Indexing skills are built up and developed through practice; every text has its particular problems, and the more experience one has the easier it is to confront these problems and find the best solution. Trainees are therefore required to undertake three Online Tutorials in which, working in a group of between six and ten people, they compile indexes to a set text (c. 30–40 pages) and compare their results under the guidance of an experienced indexer. The texts vary widely in order to give experience of a range of subject matter and indexing difficulties, such as dealing with complicated personal names or deciding what to include if space is limited. Each Tutorial runs via a dedicated Yahoo group for one month: two weeks for compiling the index, and two weeks for guided discussion, after which the tutor posts his/her version of the index to the group. An added benefit is personal contact with a named tutor and with other trainees; distance learning can be an isolating experience. Trainees are encouraged to do their three Tutorials at different stages of their progress through the Training Course, to give them constant practice in the techniques they are learning and to focus their minds on the needs of the index user, thus making them better prepared for the formal tests.

To demonstrate that they are ready to work in the commercial world, after passing the four formal tests trainees must undertake a Practical Indexing Assignment, consisting of the submission of an index to a complete book or document of their own choice. This is studied carefully by an experienced indexer, who indicates areas for improvement and gives constructive advice.

Additional support

Trainees have access to an email helpline for queries about topics in the Course Units or Standards, and an email discussion list for trainees only.
SI also runs a programme of workshops, both 'live' and online, to supplement the

training experience and also to provide Continuing Professional Development for working indexers. The popular 'Introduction to Indexing' workshop, held several times a year in various different venues throughout the UK, is aimed at people who are considering indexing as a career and want to find out more, and also at those who have just begun the training course and want extra support. It offers an overview of the indexing process; practical exercises and discussions led by an experienced tutor who is also a practising indexer; an opportunity to meet and talk with others who are starting out on a career in indexing; a chance to ask questions about what it's like to run your own freelance indexing business from home; and a set of useful handouts to take away.

Other workshop topics, for more advanced trainees and working indexers, include editing the index, user needs, embedded indexing, periodicals indexing, and specific subject areas such as law, cookery and gardening.

Future plans

The SI training course has produced over 300 Accredited Indexers since it began in 1988, and the first three Course Units also form the basis of the American Society of Indexers' recently launched Training in Indexing course. The next edition of the SI course is currently in the preliminary planning stage; while the course content, though updated, will remain basically the same, the method of delivery will take full advantage of the latest e-learning and e-assessment methods. Course materials will be accessed from the SI website, the electronic learning aids in the existing CD-ROM will be greatly expanded, and testing of some of the more mechanical indexing skills and recording of results will be carried out electronically.

Notes and References

(1) *Mulvany*, Nancy C.: Indexing books, 2nd edition. Chicago and London: University of Chicago Press, 2005

(2) *Crystal*, David: "Is there anybody there?" In: The Indexer 19(1995)3, p. 153

(3) *Booth*, Pat F.: Indexing: the manual of good practice. München: K. G. Saur, 2001

(4) *Booth*, Pat F.: "Be a peach: enhancing your work through training." In: The Indexer 23 (2002)1, pp. 18–22

(5) British Standards Institution

(6) www.macrex.com; www.indexres.com; www.sky-software.com

(7) *Halliday*, Jill: "Professionalism and the indexer." In: The Indexer 25(2007)3, pp. 167–168

(8) Available from the Society of Indexers, www.indexers.org.uk

Society of Indexers, United Kingdom, Indexer, Job Description, Training, Curriculum, Teaching Material

Society of Indexers, Großbritannien, Indexer, Berufsbild, Weiterbildung, Lehrgang, Lehrplan, Lehrmaterial

The Author

Ann Hudson, MA PGCE

has been indexing for as long as she can remember and professionally since 1981. A Fellow of the Society of Indexers, she specializes in archaeology, history of art and architecture and related subjects, indexing both books and periodicals. She is the Training Course Co-ordinator for the Society of Indexers and also a workshop leader, running workshops both 'live' and online for indexers at all levels, and also in-house for publishers.

7 Hawthorn Close
Chichester W Sussex PO19 3DZ
United Kingdom
hudsonindexing@aol.com

Indexing classes offered by the Graduate School (USDA)

Kari Kells, Olympia, WA (USA)

This article provides an overview of the two indexing courses offered by the Graduate School (USDA), which teach several hundred students from around the world every year. These two courses are the most popular indexing classes in the United States and until recently were the only practice-oriented, self-paced indexing classes in the English-speaking world.

Indexierungs-Kurse der Graduate School (USDA)

Dieser Artikel liefert einen Überblick über die beiden von der Graduate School (USDA) angebotenen Indexierungs-Kurse, die jedes Jahr mehrere Hundert Teilnehmer haben. Diese beiden Kurse sind der populärste Indexierungs-Unterricht in den Vereinigten Staaten und bis vor nicht allzu langer Zeit auch der einzige praxisorientierte, dem eigenen Lerntempo folgende Indexierungs-Unterricht in der englischsprachigen Welt.

History and overview

I want to begin by admitting that I struggled with which details to include in this article and which to exclude. I decided to err on the side of including too many details because these details provide a good view of the complexity of indexing – something that is likely to appeal to readers of this journal.

The Graduate School's indexing courses are more commonly known as the USDA indexing courses because the parent organization of the Graduate School is the U.S. Department of Agriculture (USDA). I won't explain in detail the history of why the Graduate School is connected to the Department of Agriculture because it is a tangled explanation. Simply put, this is an instance of an organization assigning responsibilities to its agencies in an unusual way. The Graduate School was established in 1921 with the goal of providing educational opportunities for government employees who wanted to gain additional job-related skills. These courses are now offered to anyone, not just governmental employees and not just students within the U.S. Although it is associated with a governmental agency, the Graduate School is self-sustaining and receives no federal funds.

First offered in 1984, these are the oldest distance-learning indexing courses in the United States. Several hundred students from around the world register for the courses each year. Since their original design, these courses have been updated by several professional indexers working as a team with experts in instructional design of distance-learning courses. These courses were the only practical, distance-learning indexing classes for many years. I was one of the many indexers in the U.S. who initially learned indexing in theory-heavy classes in Library & Information Science degree programs. We found that while we completed courses with a clear understanding of information science theories related to indexing, we had little understanding of the practical issues related to professional indexing. Completing the coursework offered by the Graduate School is one way to balance our knowledge and skills because both technical rules and artistic guidelines are addressed in these courses.

In that vein, rather than testing students solely on their knowledge of theory, instructors test and grade their practical skills. Each index submitted is unique because indexes are a form of creative writing within a prescribed format. No indexes will ever be identical because there are very few absolute rules in indexing. Instead there are only guidelines, industry standards, and style considerations. Indexes result from each indexer's understanding of the materials, their ideas about readers' needs, and their opinions about the best way to connect readers with the required information. Instructors consider and grade each index as a unique piece of work – just as writing instructors would consider and grade each student's composition as a unique work. Set criteria for analysis and grading are provided to students with their graded indexes.

Students are assigned to an instructor who is an experienced, professional indexer. This is not the case with many other indexing courses. Instructors not only grade lessons and provide detailed reviews of each index; they also act as mentors who guide students through learning how to write useful indexes, contracts, bids, and invoices.

These courses are self-paced, allowing students to submit lessons at their leisure. They have one year to complete their course and can receive an extension for a small fee. Students send their lessons to instructors using e-mail or postal mail.

Basic Indexing course

Nancy Mulvany, author of *Indexing Books* – the most popular English language indexing book – designed the current Basic Indexing course. The course currently has the following features:

- 11 lessons that focus on information provided in *Indexing Books* and the *Chicago Manual of Style*. Some lessons include additional readings written by other industry leaders.
- Practice editing indexes.
- Practice writing indexes for four documents that provide students with progressively more complex indexing situations.
- Students receive a certificate upon completion.

Lesson 1: Indexing and Computers. Readings and graded assignment introduce what indexes are and are not, the history of indexes, how indexes are used, where indexing fits into the book production cycle, and definitions of indexing terminology. In addition it includes instructions about deciding what content should be included in the index and what should not, how to index discussions that are considered indexable, and about phrasing index entries. It also has an overview of technology used in indexing including automated and semi-automated indexing, embedded indexing, dedicated indexing software, and online Indexing.

Lesson 2: Structure and Arrangement of Indexes. The second lesson provides more detail about index style and formatting options, with an emphasis on alternatives to page numbers, indexing continuous and non-continuous discussions, and notes. It also introduces "See" and "See also" cross-references and addresses phrasing and formatting options, appropriate use and accuracy of cross-references, multiple cross-references within a single entry, and completeness and directness of cross-references. Another feature of this lesson is the overview of alphabetizing and

alternative sorting methods (such as chronologic).

Lesson 3: Proofreading, Layout, and Typographic Considerations. This lesson continues to concentrate on index formats including type size, column width, page margins, indention of subheadings, and situations calling for the use of special typography. In addition, it includes techniques and checklists for editing indexes, and advice on writing notes to clients in order to resolve questions about style and format.

Lesson 4: Structure of Indexes. In this lesson, students read about the art of writing useful subheadings, the processes of indexing when working with hard copy of texts, and how to submit indexes to clients. This lesson includes an introduction to *The Chicago Manual of Style* guidelines for index format.

Lesson 5: Term Selection. This lesson provides students with their first opportunity to practice applying indexing techniques and styles. Students create, edit, and get feedback from their instructor about the first index that they write this class.

Lesson 6: Creating a Reader-Friendly Index. After reviewing instructor's feedback on their Lesson 5 index, students put their new knowledge to work by writing another index for a slightly more challenging text. In addition, readings address the usefulness of indexers having specialized subject expertise and advice on reading the text from the audience's perspective so that we create index entries that are truly useful to them. Depth of indexing is also discussed.

Lesson 7: Special Problems in Indexing. The lesson introduces some interesting challenges that indexers encounter: acronyms, abbreviations, numbers, and symbols; handling issues with personal names including names in non-English languages, suffixes (such as "Jr."), personal names and subjects having the same name (such as "London, England" and "London, Jack"), names and abbreviations of organizations (such as NATO), numerals in main headings (such as "Henry VIII"), particles in names (such as "de Gaulle, Charles"), and compound names (such as "Mies van der Rohe, Ludwig"); geographic names including names beginning with "Mount" and "Lake", place names beginning with articles (such as "Hague, The"), names with foreign articles (such as "El Paso"), and names of places beginning with "Saint".

Lesson 8: Evaluating Your Index. Writing a third index is the activity for this lesson. The text introduces more complex indexing issues than were encountered in Lesson 6. Students are asked to concentrate on editing by using a checklist provided in course materials.

Lesson 9: Editing and Marketing Your Index. This lesson provides guidance, samples and practice in two important and necessary activities for freelance indexers: finding new clients and editing indexes.

Lesson 10: The Business of Indexing. In this lesson, students read about and practice bidding on freelance indexing projects including writing contracts, knowing what questions to ask potential clients, and knowing what to expect clients to provide upon acceptance of a bid.

Lesson 11: Submitting Your Index to the Publisher. The fourth and final index in the course is the most challenging yet. Students must index this text in a week so that they can experience deadlines much like what freelance indexers experience. In addition, students practice writing a cover letter and invoice.

Advanced Indexing course

Course materials for Advanced Indexing were last updated in 2006 and currently feature:

- Practice writing seven indexes in a variety of disciplines.
- Students write essays about how they will handle indexing and business situations presented to them.
- Students receive a certificate upon completion.

This course has two very different but equally important goals: to provide students with practice handling increasingly difficult indexing issues, and also to teach students how to apply business principles to project-related situations frequently encountered by freelance indexers such as writing project bids, submitting invoices, producing indexes that meet client requirements, and handling misrepresented projects.

This course presents students with experiences that approximate experiences encountered by professional indexers. Students write indexes as well as essays about specific indexing and business situations. Students are expected to act as if these were actual business situations so they submit indexes as if they are preparing each index for an actual client. Instructors evaluate assignments imagining that they are editors at publishing houses. One of the factors assessed by instructors of this course is how well students follow editorial direction and make indexing decisions.

Lesson 1: Evaluating an index. In this lesson, twelve common myths about indexing are explored, including all published indexes are good indexes, longer indexes are better indexes, if you study an index created by a professional indexer you will know how to write an index, and anyone can write good indexes. Students also learn how to understand their audience; how to determine which pages are indexable and which are not; and how to evaluate an existing index. This lesson also describes factors taken into consideration when estimating fees, including page size, number of columns on each page, size of type, and number of pages; subject matter; density of material; non-text material (such as tables, charts, photos, and drawings); and target audience and their needs.

Lesson 2: Indexing a book about ecology. This lesson focuses on writing an index for an ecology book. Students also calculate a per-page price for this index and submit an invoice for it; submit both a per-page price bid and an hourly price bid; and write essays about how they arrived at these prices.

Lesson 3: Indexing a book about history. The main features of this lesson are writing an index for a history book; evaluating the per-page and the hourly bids submitted for Lesson 2 and comparing estimated times with actual times; and reviewing tips on estimating time and fees.

Lesson 4: Indexing a book about health. For this lesson, students index a health book following format and style specifications that they choose. In a letter to be submitted to their client, students explain the reasons for the format and style specifications that they have chosen.

Lesson 5: Indexing a technical book. In addition to indexing a technical book, students write a letter to be submitted to their client addressing how they would handle future annual revisions to this text and also explaining the procedures that they would recommend for embedding this index.

Lesson 6: Indexing a biography. Students write an index for a biographical book and adhere to a length limit for this lesson. They also write an essay comparing chronological and alphabetical arrangement of subentries. Students also create and submit a contract for the indexing project in the next lesson.

Lesson 7: Indexing a psychology book. In this lesson, students write two indexes – one of subjects and one of names – for a scholarly psychology book. They also write two essays: one evaluates the usefulness of both indexes and another evaluating their editing process.

Lesson 8: Problematical situations for indexers. Rather than writing an index for their last assignment, students write essays about how they would handle each of ten hypothetical indexing situations such as:

- An editor asks you to index the second edition of a book. The editor says that they will give you the index from the previous edition and all they want you to do is change the page numbers.
- When the published copy of a book you indexed arrives at your office, you find editorial changes in your index such as many cross-references have been eliminated and many entries that had only one or two page preferences have been eliminated.
- You receive pages for a book to index and the authors have highlighted the terms they think ought to go in the index.
- You have written a book index and now the editor has asked you to edit the index for online publication. Your client says "we just need you to add tags and codes for the HTML version, and make sure every locator has a unique entry."
- You have finished a project and sent the invoice, which has not been paid within the required thirty days. What will you do now?

References

The Chicago Manual of Style. 15th ed. University of Chicago Press, 2003. ISBN 0-226-10403-6.

Mulvany, Nancy. Indexing Books. 2nd ed. University of Chicago Press, 2005. ISBN 0-226-55276-4.

Graduate School USDA, Indexing, U.S.A., Curriculum, Teaching Material

Indexieren, Ausbildung, Studium, Hochschule, USDA, USA, Lehrplan, Lehrmaterial

The Author

Kari Kells, MLIS

Since 1998, Kari Kells has taught indexing through the Graduate School, Index West, and at several universities. In 2005, she co-authored *Inside Indexing* with Sherry Smith. More information about Kari can be found on her web site: www.indexw.com. If you have any questions about these courses, you are invited to contact Kari at kkells@indexw.com or The Graduate School Self-Paced Training Program at selfpaced@grad. usda. gov or by visiting www.grad. usda.gov.

Index West
Kari Kells
PO Box 615
Olympia, WA 98507
U.S.A.
Telephone: 360870 4384
kkells@indexw.com
www.indexw.com

Call for Papers

24. Oberhofer Kolloquium zur Praxis der Informationsvermittlung

Informationskompetenz 2.0 – Zukunft von qualifizierter Informationsvermittlung
vom 10. bis 12. April 2008 in Magdeburg

Gesucht werden Originalbeiträge in deutscher oder englischer Sprache für die Fachtagung 2008 zu folgenden Aspekten:

- Web 2.0 – Herausforderung an Bibliotheken und Informationsvermittler
- Intelligente Suchmaschinen – Generation Google: Bedürfnisse und Erwartungen
- Informationskompetenz von 8 bis 80 – Ansprüche an Aus- und Weiterbildung
- Sprachkompetenz als Grundlage der Informationskompetenz Buchstabenwelt kontra Bilderwelt
- Informationsretrieval 2.0 – Entwicklungstrends unter Web 2.0

Das Kolloquium wendet sich an alle, die sich mit der Informationsvermittlung befassen und den Fragen, die sich aus den Bedingungen von Web 2.0 ergeben, offen und konstruktiv gegenüberstehen, sie aber nicht als modische Worthülse benutzen, sondern statt dessen die Kluft zwischen Informationstechniken sowie kommunikativen und intellektuellen Fähigkeiten des Menschen verringern wollen. Wichtig ist dabei, dass man – trotz aller modernen Techniken – die Sprache und ihre Verarbeitung im Gehirn als Basis nicht aus dem Blickfeld verlieren darf. Die Verarbeitung von Bildern alleine schränkt unsere intellektuellen Möglichkeiten wesentlich ein – auch wenn es heißt: „Ein Bild sagt mehr als 1000 Worte". Uns geht es darum, dass Informationskompetenz dem Nutzer nicht „übergestülpt" werden soll, sondern dass wir gemeinsam mit dem Nutzer Möglichkeiten der besseren Informationserschließung und -nutzung finden.

Vorschläge als Kurzfassung mit maximal 400 Wörtern erbitten wir per E-Mail an: oberhof2008@dgi-info.de. Zu nennen ist die Ansprechperson für die Benachrichtigung mit vollständiger Postanschrift, Telefon- und Faxnummer sowie E-Mail-Adresse.

Termine:	Einreichung von Vorschlägen	bis 7. Januar 2008
	Benachrichtigung über Annahme	bis 15. Januar 2008
	Abgabe der Langfassung	bis 25. Februar 2008

Veranstalter: DGI Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V., Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Telefon: (069) 43 03 13, Fax: (069) 490 90 96, otterbein@dgi-info.de
VDI Verein Deutscher Ingenieure e.V. Magdeburger Bezirksverein, Arbeitskreis Information c/o Siegfried Rosemann, Stadionstraße 13, 39218 Schönebeck, Telefon: (0 39 28) 6 97 75, siegfried.rosemann@web.de

Organisationskomitee: Wolfgang Löw, Marlies Ockenfeld, Siegfried Rosemann, Dr. Luzian Weisel

Indexing software

Caroline Diepeveen, Middelburg (The Netherlands), Jochen Fassbender, Bremen, Michael Robertson, Augsburg

This article looks at the different types of software available for indexing and whether these can be used to produce the type of index that is required by international standards. As it is often claimed that indexing can be done by a computer and no human effort is needed, the authors first look into why indexing has to be a human activity and why automatic indexing cannot produce a satisfactory index. Automatic indexing will at best provide a starting point for further indexing. The authors then look into the possibility of indexing by using general-purpose software. This can produce acceptable indexes, although it may be fairly cumbersome. For the professional indexer, dedicated indexing software is therefore the preferred option. The three most widely used dedicated indexing programs are briefly reviewed. Their orientation and appearances are very different, but the features and final output are often very similar. Website indexing is a relatively new speciality within indexing. Two software programs used for website indexing, one freeware and one commercial, are discussed. The overall conclusion is that indexing is an activity that needs human intellectual input; software greatly enhances this input, but cannot replace the human input.

Indexierungssoftware

Dieser Beitrag handelt von unterschiedlichen Arten verfügbarer Software zur Erzeugung von Registern und untersucht, ob diese dazu benutzt werden können, ein Register nach internationalen Normen zu erstellen. Da oft behauptet wird, dass die Registererstellung mit einem Computer und ohne Einsatz des Menschen durchführbar sei, untersuchen die Autoren, weshalb Indexieren eine Aktivität des Menschen sein muss und weshalb eine automatische Registererstellung kein zufriedenstellendes Register hervorbringen kann. Automatische Registererstellung kann bestenfalls einen Ausgangspunkt zur weiteren Indexierung liefern. Anschließend wird die Möglichkeit der Registererstellung mit allgemein verfügbarer Software untersucht. Dies kann akzeptable Register hervorbringen, wenngleich oft nur auf mühsame Weise. Für den professionellen Indexierer stellt daher spezielle Indexing Software die bevorzugte Option dar. Die drei am meisten benutzten speziellen Indexierungsprogramme werden kurz bewertet. Ausrichtung und Aussehen dieser Programme sind sehr unterschiedlich, aber die Merkmale und Output-Optionen sind sehr ähnlich. Website Indexing ist ein relativ neues Spezialgebiet innerhalb der Disziplin des Indexierens. Zwei Programme – eine Freeware und ein kommerzielles – zur Erstellung von Registern von Websites werden erörtert. Das Fazit insgesamt ist, dass das Registermachen eine Aktivität ist, die intellektuellen Input des Menschen benötigt. Software kann den Input hervorragend verbessern, aber nicht den Anteil des Menschen daran ersetzen.

'There is nothing automatic about the index-writing process.'
Nancy Mulvany – Indexing Books (1994, p. 245)

Introduction

One of the most common questions encountered when telling people that one is a professional indexer is: 'But can't a computer do that?' The answer is that indexers do indeed use computers, but that, as with translating, the human intellect is very much needed to produce a sensible index. Selection of search terms can't be done by a computer, because context is everything. Professional indexers, therefore, use software to assist them in the indexing process. This article discusses the software tools that can be used for indexing.

After we have shown how 'automatic' indexing works, and why it won't produce a useful index, we discuss three different kinds of software tools that can be used for indexing. First, we discuss the use of software tools that most people already possess, such as word processors, spreadsheet and database software. These packages will typically be used by non-professional indexers and people who only have to produce an index occasionally, such as authors.

We then discuss the dedicated indexing software that is used by professional indexers. What is so different about it and why is it useful? Finally, we discuss software used for indexing websites.

1 Fully automatic or semi-automatic indexing software

With automatic indexing software, it is the computer that selects the terms to be included in the index, not the indexer. The term 'automatic indexing software' is somewhat misleading, as the program only generates a concordance list, which is not an index. This means that the software produces an alphabetical list of all the words in the text (excluding a few stop words) – whereas an index is a structured listing, usually alphabetical, of the key concepts of a text.

More sophisticated automatic indexing programs produce selected word lists. Still, at best they produce only the basis for an acceptable index – certainly not the final product. Examples of recent automatic indexing programs are TExtract and IndDoc.

TExtract claims on its website that it is a semi-automatic indexing program. At the same time, however, it claims that you can produce an index with TExtract within several hours. Any professional indexer can tell you, though, that it is impossible to produce an index within several hours, no matter how familiar you are with the text.

Example: In a recently edited index made by somebody else, it quickly became clear that once a decision had been made to include

a specific term in the index, each and every mention of that term was included without any regard for the context. This meant that the term 'body' – which here referred to the female body, as it was a book on women – included several mentions where 'body' had been used as 'body of literature'. Of course, the index also included many insignificant mentions of concepts that should have been left out, but a computer is unable to distinguish between significant and insignificant uses of terms.

2 General-purpose software

You can produce an index using software that is already there on your computer and that you would generally use for other purposes. Often, one's first indexes are produced using Word, for example, but it is also possible to produce indexes with desktop publishing software, spreadsheet programs, or database programs. The advantages are, of course, that you are already familiar with the software and that you won't have to invest in additional software. Word processors are most often used by non-professional indexers. However, the alphabetization creates problems when using these software packages. Word processors sometimes sort text on ASCII character coding. This means that all upper case characters will sort before lower case characters, whereas in indexes sorting should not be case-sensitive. Also, one regularly wants to take small words out of the automatic sorting. For example, one does not want the program to sort on prepositions like 'in', 'for' and 'to' in the subheading, but you may still want to retain them at the start of an entry. The same problem arises with names that start with particles that you do not want to invert, but you would not want to sort on either, e.g. 'al-' in 'al-Qaeda'. This cannot easily be achieved with word processing software. You can change things manually in the sort, of course, and if you only produce an occasional index that may be acceptable. When you have to index regularly, you will soon be looking for better solutions.

There are professional indexers who use Word or a database program for indexing, but these programs are limited in what they can produce in terms of indexing. When professional indexers use Word or Adobe FrameMaker for indexing, they would usually make so-called 'embedded indexes'. This means that the terms selected for the index will be marked with 'tags' in the text. The index is then created by generating a listing of the selected terms. The text and the index are one document.

Producing the 'tags' is rather cumbersome, and embedded indexing is generally more time-consuming than indexing with the dedicated indexing software that is discussed in the section below. In particular, the editing of the index – which in our experience takes up about one-third of the time needed to produce an index – is much more time-consuming with embedded indexing.

Recently, software tools have become available to make embedded indexing less time-consuming and to allow professional indexers to use their dedicated indexing software in conjunction with the Word embedded indexing tool. Examples of these are DEXter and WordEmbed. Both are basically collections of macros that run in Word.

3 Dedicated indexing software

The three major dedicated indexing software packages are: SKY Index, CINDEX and MACREX. All three started out as DOS programs and they all developed Windows versions in the second half of the 1990s. CINDEX is also available for Mac. Demo versions of all three programs are available without time limits.

Basically, all three packages are able to do the same things, although they differ in the way they go about it. CINDEX works with the index card metaphor, MACREX has a very simple input screen that looks like the index itself, and SKY Index uses a database or spreadsheet table metaphor (rows, columns and cells). The interfaces of these programs are very efficient; to put it in Mulvany's words (1994, p. 272): 'Perhaps one of the strongest features of this software is that the indexer can work with the index in sorted order at all times. … The structure of the index is constantly emerging and visible. The work area context is the index itself.'

All three programs allow you to work in alphabetical order, for example, according to the ISO standard (1996) and the Chicago Manual of Style (2003, pp. 773–784) or in page number order. The latter option is very useful when you have to update an index for a new edition or when several pages of text have been inserted at the last moment. Page numbers can be increased or decreased by a specified amount, and volume numbers can be added, deleted or substituted in a number of ways.

Indexes can be split or joined in various ways. There are several methods of alphabetization available (letter-by-letter and word-by-word), the sort can also be made chronological (very useful for tables

of legal cases in Continental law), or the sort can be forced in other ways (very useful when making an index of Biblical references that has to be in the order of the Bible books).

Dedicated indexing software is excellent in keeping indexes consistent, covering such formal aspects as sorting, punctuation within entries, indents, locator format, and cross-reference consistency, thus enabling the indexer to focus on the index itself.

To highlight the differences between the three most important dedicated indexing software programs, we will present three short reviews of each of them. Each 'review' is written by an experienced user. Within the scope of this article, it is not possible to review these programs in detail; however, the references will give further guidance to those who want to know more about these software tools. There is much more to be learned about each program than can be presented here, and each program has excellent manuals (well-indexed, of course!) and gives support where needed.

Each individual reviewer has made their own choice about which features to highlight from the program. When a particular feature is mentioned in one review and not in the others, it does not necessarily follow that this feature is missing in the other two programs.

3.1 MACREX

MACREX was developed and marketed by Hilary and Drusilla Calvert in the UK. The program has been around for more than twenty years now and it still very much has a 'DOS feeling' to it. For the history of the program, see Beare (2004). Figure 1 shows the simple startup screen that you get once you have started a new index.

Don't be fooled by the simple appearance, however, because MACREX is a very adaptable program, able to handle many different and complicated indexing situations. It is not very intuitive, though, and it has no pull-down menus. Some of its possibilities are buried in one or two layers of sub-menus and you really have to know how to get there.

The simple appearance of MACREX is going to change soon, as a new release (version 8) is planned for the autumn of 2007. New versions of MACREX always retain the features of previous versions and there is always compatibility with previous versions. Backup files of indexes made with the first versions of MACREX can always be opened in the newer versions. The new features of version 8 will be discussed briefly at the end of this section. First, we will look at the main features of the current MACREX version.

Choosing option 'I' from the startup menu brings you to the screen that allows you to add and edit entries. Again, the appearance of the screen is very simple. There are no fancy graphics or pop-up boxes for data entry. The screen very much looks like the index itself, as can be seen in Figure 2.

At the bottom of the screen is a line, and below that line you can make your new entries. People who are new to the program may find it hard to remember that you will have to choose 'F4' before you can add new entries. Once you are used to the program, you stop thinking about this. When your entry is ready, it will be inserted at the right place in your index after you have hit the 'Enter' key. When working with MACREX, you do a lot of editing at the input phase of your index. This is because the input screen looks so much like the index itself, you see right away when an entry should be phrased differently, can be merged with a similar entry or when a series of sub-entries does not look good and should be arranged differently.

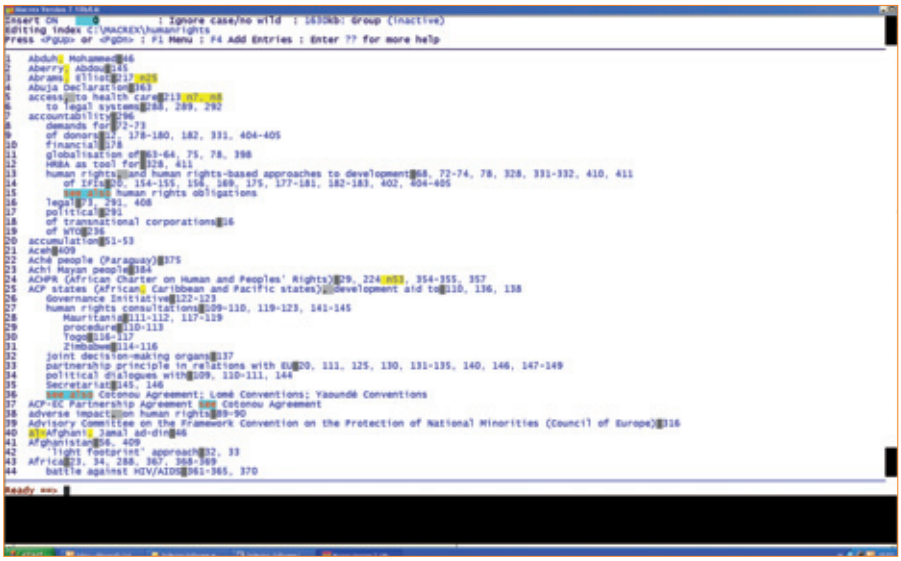


Figure 2: Index input screen

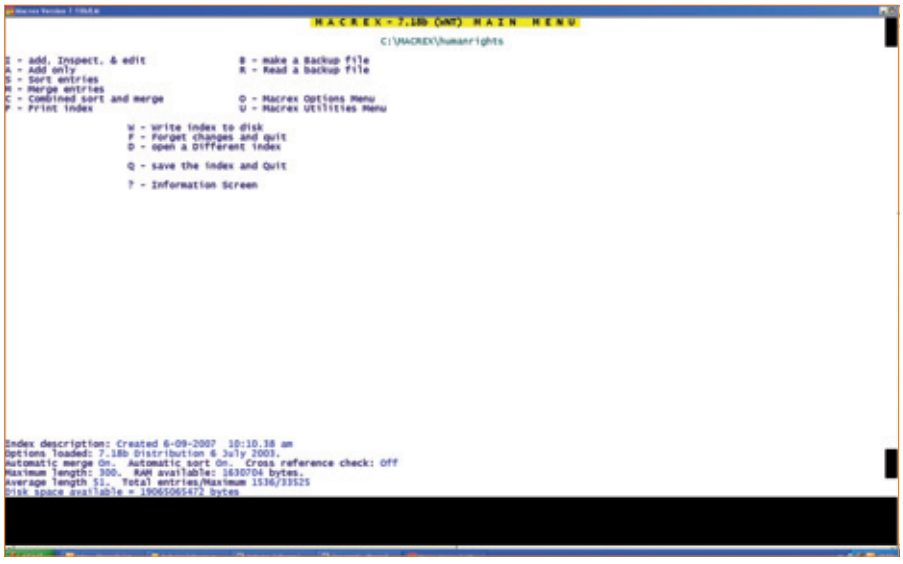


Figure 1: MACREX startup screen

Of course, you don't do all the editing as you go along, there is also editing to be done at the end. For this purpose, the 'group function' is extremely useful. Hitting 'Ctrl H' when in the 'input screen' will get you to the 'group function' screen. Figure 3 shows what the 'group function' screen looks like.

This was created by using one single search term: 'domestic'. All entries containing this term were then selected from the index. This allows you to compare different entries and to check whether your index is consistent. The numbers before the entries indicate the index entry number. Figure 3 shows that the index is not yet entirely consistent, 'domestic state parties' can probably be merged with the much simpler entry 'states' and the entry 'states' needs

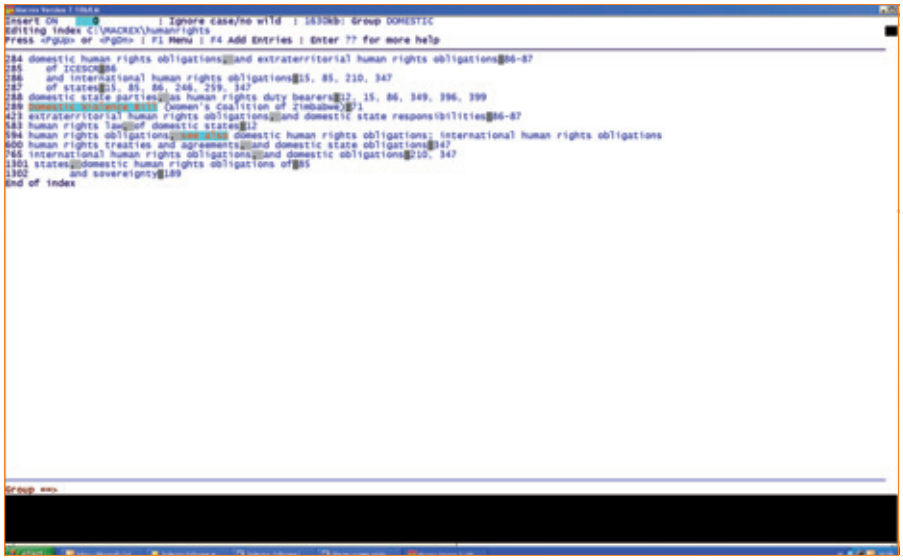


Figure 3: Screen in Group Function mode ('domestic' was used as search term)

additional page locators. This will happen automatically, though, when entry numbers '288' and '1302' are merged.

It is possible to do much more complicated searches in the group mode. You can use 'wildcards' to indicate that you want to search for a specific string of characters only at the beginning or the end of an entry. There are several other types of 'wildcard' as well, and it is also possible to do Boolean searches.

When your index is finished, the output file has to be produced. This is done by choosing the 'Print' option from the startup menu. Again, not a very intuitive choice, and producing the output file is not a very straightforward procedure. Beginners may find this procedure slightly complicated, but again, once you are used to it, it is not much of a problem. There are many layout options to choose from, although in practice one probably only works with one or two of the layout files. It is also possible to have layout files adapted for you, when publishers have very specific wishes/inhouse styles (e.g. 'OUP' stands for Oxford University Press in Figure 4). To give you an idea of the available choices, Figure 4 shows part of the layout options screen for output.

It is also possible to obtain layout files or other files adapted to your specific needs. These files are easily loaded into the program. For example, special layout files have been produced for use with Unicode fonts (when characters with diacritics are needed) and custom made files have been made for electronic indexing procedures (XML files), that tend to differ from one publisher to another.

As stated at the beginning of this section, MACREX is about to release version 8 of its program. There will be quite a few drastic changes in version 8, and we will therefore briefly describe the new features here. Version 8 will have a much more 'Windows look', with pull-down menus; however, all the version 7 keystrokes will also be retained. Navigation through the menus will be simpler, with most options available through the main menu, and there will be a context-sensitive help option available at all times. Creating the output file will be much easier, and it will be done in a single step. And there will be an 'autocomplete/table of authorities' feature. This option will also make it possible to work with a controlled vocabulary, as you can easily import lists of terms made in other indexes.

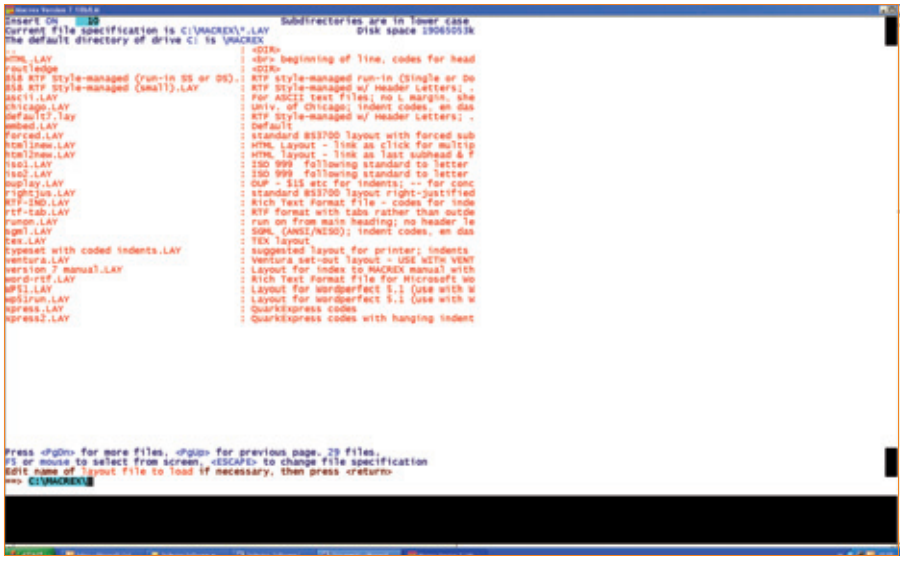


Figure 4: Screen with choices for output formats

To summarise, when starting to work with MACREX, it is not an easy program to work with. It takes time to get used to the way it works and to find out about all its different options and possibilities. Once you are past this stage, you will find that it is a very reliable and adaptable program that is a joy to work with. Version 8 will probably take away a lot of the irritation that beginners or infrequent users may experience with the program.

3.2 CINDEX

CINDEX is produced by the company Indexing Research, founded by Frances Lennie in Rochester (New York), in 1986. The program was originally produced for MS-DOS and underwent continuous development throughout the 1990s, culminating in the current Windows and Mac versions.

Entering records

In contrast to indexing programs that have a spreadsheet-type appearance, records are entered in CINDEX in a separate dialogue box with entry line fields for 'Main', 'Sub1', 'Sub2', etc., and 'Page' (Figure 5). The Page Down key then enters the record into the index and clears the box for the next record. This easily used interface means that the record that is currently being worked on is clearly legible and clearly structured (instead of being just one of the screen-wide lines displayed in a screen-long list).

The typing of index records is speeded up by the program's abbreviation and auto-complete features. The user can create abbreviations for phrases that are frequently used; when the abbreviation is typed in the entry box, the program automatically expands it. With auto-completing, the program recognizes when the first letters of a previously entered record are being typed and automatically completes the main heading or subheading line. (All of these features can of course be switched on or off as desired.)

Another much-used facility, as in other programs, is 'flipping', in which the main heading and sub-heading in a record can

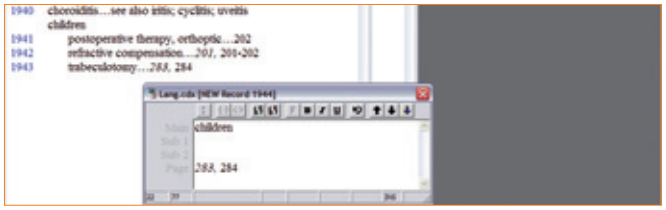


Figure 5: CINDEX entry dialogue box

be switched by pressing a screen button or a key combination. In addition, any heading or subheading from the record entered immediately previously can be copied automatically by pressing a key combination.

Hot keys can be programmed, and 'see' and 'see also' are installed by default as SHIFT F1 and SHIFT F2. The user can also set macros to speed up frequently used sequences of keystrokes.

Displaying the index

At any time during record entry or editing, the index can be displayed in various formats – sorted/unsorted, draft format, or full format with run-in or indented views. Widely flexible sorting and formatting options are available (Figure 6), and a list of subheading prefixes to be ignored during sorting is available and can be edited. Non-alphabetical sorting of subheadings (e.g., for chronological sequences) can be forced by adding hidden coding, and an index can of course also be sorted in page sequence so that additions can be made if there are changes in proof, or for a new edition of a book.

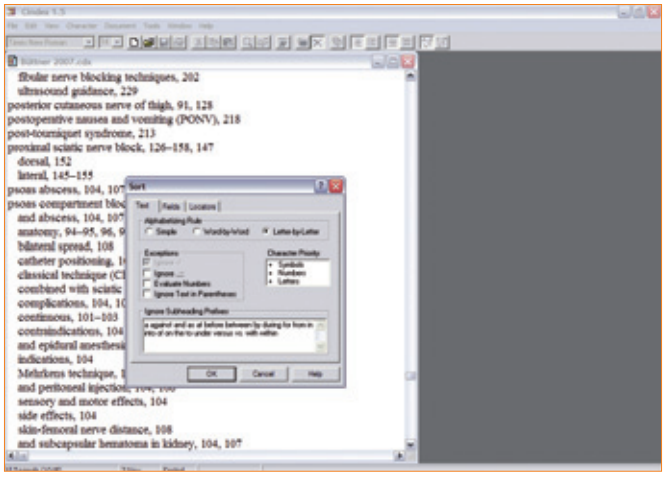


Figure 6: Options for sorting

Editing records

After the entry of records has been completed and the resulting index has been sorted alphabetically (with options for 'simple', 'word-by-word' and 'letter-by-letter'), all records that appear as single entries with a single subheading (e.g., 'knee | anatomy, 159') can be automatically joined to form single main entries ('knee, anatomy, 159').

One of the most useful features of the program during editing is the 'propagate' facility. After sorting, a main entry that needs to be corrected may have a large number of subentries. To avoid the need to edit the main heading in each individual record for all the subentries, activating the 'propagate' function (using a screen button or keyboard shortcut) copies corrections to all of the subentries.

Another important editing function is the facility to group entries when considering the best choice of phrasing or placement for conceptually related terms. Specific related entries can be identified using the 'Find' function and saved as a group, which can be viewed or printed out and edited separately – providing greater clarity and saving time. When editing of the group has been completed, a simple key combination returns the user to the full index, with all of the group edits incorporated into it.

A vital function in the program is the ability to check the validity of cross-references – a warning list is produced showing inaccurate and circular references.

The user can move at any time back and forth flexibly from entering records to editing them, sorting them in any order, and displaying them in any format – there is no need to complete an entry session before switching to a separate editing session in which different options are available.

Saving and file format

When work on an index is completed and the file is ready to go to the client, the format for the output file is selected via 'File | Save as …', with RTF (Word), Quark, and AAP options available (Figure 7).

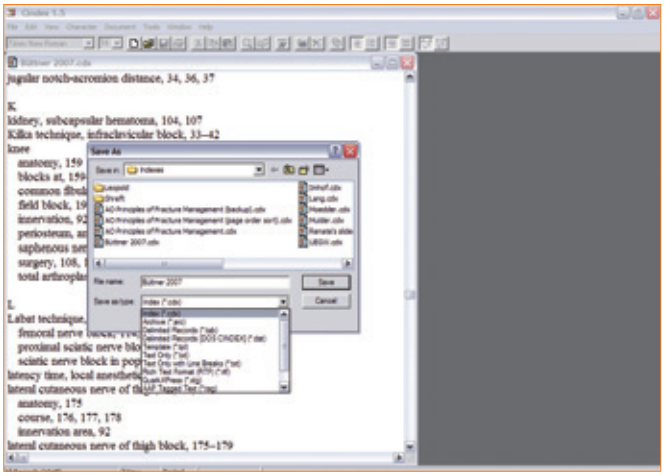


Figure 7: Options for file output

The history and development of the program is covered in Lennie (2005).

3.3 SKY Index

SKY Index (Professional Edition) was first released in 1997 and is under constant development by Kamm Schreiner of SKY Software in the USA.

User interface

The most prominent feature of SKY Index is its unique user interface. As can be seen in Figure 8, there are two main areas. The upper half is called the preview pane, which always shows the developing index in alphabetical order, including main headings, subheadings with indents, cross-references, and locators. The preview pane displays the index as WYSIWYG, including the fonts and text formatting chosen.

The lower half is the data entry grid with its spreadsheet- or database-like look and feel. Here, all the editing of index entries and individual headings takes place, and entries can be added and deleted. The data entry grid consists of columns for main heading, subheading(s), page references (and other types of locators such as volume and chapter numbers) and other details. Cross-references are entered in the page field. The order of entries in the data entry grid can be viewed in various ways – for example, as entered or in page order. This provides an efficient way of working and complements the alphabetical view in the preview pane, so that there are always two views that can be seen at the same time.

Above the preview pane are the menus, main toolbar, sort toolbar as well as acronym and macro toolbars. The main toolbar with its symbols provides more direct access to many of the individual options given in the menus. Some of these individual options and their keyboard shortcuts are the same as in other Windows programs, others are specific to SKY Index.

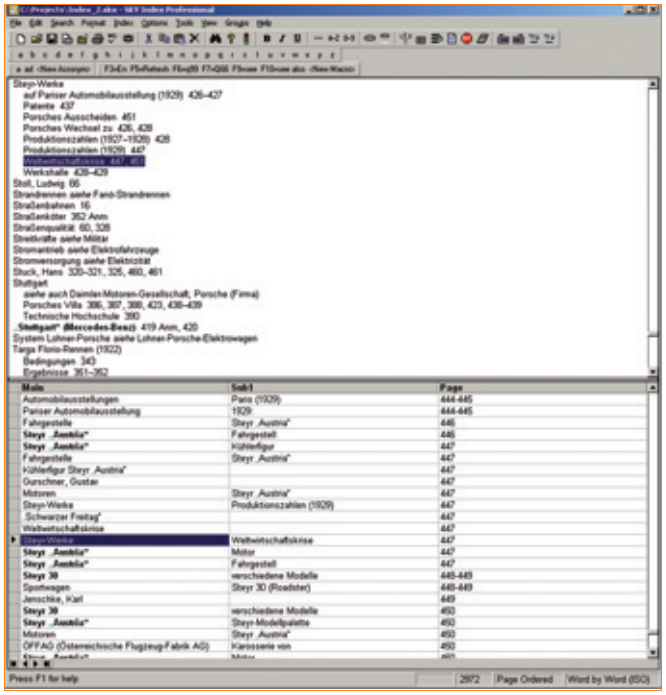


Figure 8: SKY Index main window

Getting around

Navigating in the index is very easy using the mouse or keyboard. Selecting an entry in either the preview pane or data entry grid automatically displays the same entry in the other half. Thus, one can very quickly check out an entry in its alphabetical surroundings and in another view. Moving around the index is also possible by using the sort toolbar (jumping to the corresponding first entry of the selected letter), via a browse modus, by using the scroll bars or arrow keys.

Adding and editing entries

While adding entries, the two features called AutoComplete and AutoRepeat can save a lot of typing. AutoComplete makes use of headings already entered, while AutoRepeat enables certain parts of an entry to be repeated in the next entry. On top of that, by defining acronyms the indexer can efficiently handle long character strings, and macros help to save time triggering actions which are used over and over again.

Entries and individual headings can be manipulated in a large number of ways, including duplicating, flipping, and combining, to name just a few. For example, the following entry can be duplicated and flipped as shown within a split second, while automatically adjusting prepositions to their right locations:

Autobahnen
in Bayern 73

Bayern
Autobahnen in 73

Many index entries can be treated like that at once by simply selecting any entries in the data entry grid.

Another powerful feature makes it possible to isolate (= group), view, and edit large entries with many subheadings, though this is not restricted to large entries alone. Instead, any subset of the index can be grouped and edited, including whole entries, certain words and phrases within entries, and locators. There are many ways of creating groups, which can even be defined and reused.

There is a sophisticated Find & Replace option for altering headings. In addition to that, the Propagate Edits function can alter many headings with identical text at once. Another useful option is converting cross-references into entries (double posting).

Sorting

Among the many sorting methods available are the following ones. Entries beginning with special characters can be force-sorted using Hide or Ignore text options; this formatting is displayed in colour and can easily be seen. Sorting of Roman numerals and dates is possible. Prepositions can be ignored when sorting subheadings, as in:

Autobahnen
in Bayern 73
mit Gebühren 41, 92–95
Glatteis auf 127–129

Quality control

The Error Scan tool can detect important errors, such as cross-reference inaccuracies, improperly formed entries, orphan subheadings, and too many locators behind headings. All detected errors are listed in a dialog box (Figure 9) and can be fixed via the data entry grid.

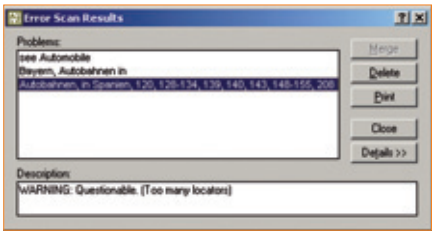


Figure 9: Error Scan Results

At any time it is possible to produce an output of the whole index or parts of it. Among the output formats available are various RTF variants, HTML, and Quark.

Overall, the program can be used in a very intuitive way, enabling even beginners to enter entries instantly, while experts benefit from the many useful features and configurations they can choose from.

A comparison of CINDEX and SKY Index has been published by Michael Wyatt (2001).

4 Website indexing software

There are different ways in which websites can be searched or made searchable. Very often one will find some sort of search engine as a tool to search a particular website, although there are other tools available as well (such as navigation menus, site maps, searchable databases, etc.). A back-of-book style index is another way of searching a website. On websites that have this feature, one would typically find an A–Z alphabet somewhere on the home page of the website. Clicking on one of the letters gives you access to the index. Two articles on website indexing itself, by Heather Hedden and Glenda Browne, can be found on pages 433 and 437 in this issue.

Specialized software for web indexing is available, but it is not very well known. Not very many professional indexers do website indexing, although this may change as more and more courses and workshops on website indexing are on offer. Professional web designers in general know very little about indexing and the software tools available. Using dedicated indexing software, it is possible to make indexes with HTML output. In combination with an HTML editor, one can also do website indexing, MACREX being the most flexible program in this respect. However, it is easier to use specialized website indexing software, and we will therefore briefly look into the two best-known website indexing software tools available: XRefHT and HTML Indexer.

4.1 XRefHT

XRefHT (commonly pronounced 'Shreft') is a freeware program developed by Tim Craven, professor of the Faculty of Information & Media Studies at the University of Western Ontario. The program can be downloaded from Tim Craven's faculty website.

The XRefHT main window (Figure 10) is relatively simple and consists of the three columns Heading, Subheading, and URL. It is possible to load these columns with the titles of HTML files, anchor names and headings <h1> to <h6> from within HTML files, and also keywords contained as meta data in the head of such files. This can be done with local files on one's computer as well as with single web pages or even a whole website. XRefHT extracts the respective data from the files selected. The URL column represents the locator field; it accommodates the file name (of HTML, PHP, and other file types) as well as subfolder and anchor name, if any.

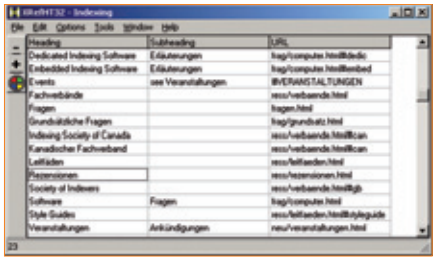


Figure 10: XRefHT main window

Once the columns have been populated, the index entries can be edited – for example, rephrasing headings and adding subheadings, as well as duplicating, inserting, or deleting entries. It is possible to sort on headings or URLs at any time in order to keep track of the developing index. It is also possible to preview the HTML output of the index in a browser window.

See and *See also* cross-references can be added in the Heading or Subheading columns by writing *'see'* or *'see also'*, followed by the target entry; the target anchor name has to be entered in the URL field. There is an alternative way of adding cross-references to an index when using XRefHT in conjunction with the thesaurus software TheW32, another freeware program available from Tim Craven.

There are a number of other features, such as the option of adding anchor names to HTML files, a spell checker, and XRefHT's own editing and web browser windows. For a freeware program, it does have a surprisingly large number of features, although the editing functionality is not as efficient as that of commercial indexing software. Also, the level of documentation is somewhat inadequate. However, there is a good description of the program by Heather Hedden (Indexing Specialities: Web Sites, pp. 61–77) and a whole book about it by James Lamb (2006).

4.2 HTML Indexer

HTML Indexer is a commercially available software tool developed by David Brown. HTML Indexer's capabilities are similar to those of XRefHT, but it has several advantages over XRefHT. A free demo version is available. The advantage of using commercial software over freeware is, of course, that one is entitled to receive technical support.

In HTML Indexer, you start an index by starting a new project and adding the website files that you want to index to your project. The entire site can be extracted at once. Figure 11 shows the 'project tree pane' that you see on the screen once you have added the website files to your project.

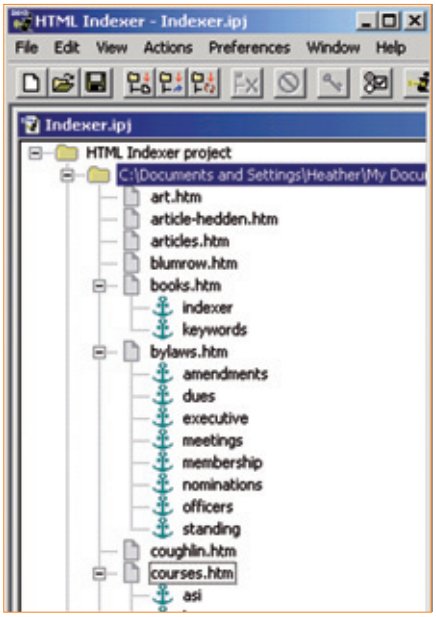


Figure 11: Project tree pane

The subfolders and the files are arranged in alphabetical order by title.

If there are subfolders, they will have plus signs next to them, which you can click once to expand the contents of the subfolder and see the list of files within.

Some of these files in turn have plus signs next to them. This is to indicate that the file has internal anchors. If you click once on the plus sign, the file expands to show the anchors within the file, as named, with an anchor icon.

Any item that is selected in the left pane by clicking once on its icon, rather than on the plus/minus sign, will then call up a list of URLs in the right pane.

It is in this left Project Tree pane that you will select the files and anchors for indexing. You can also select them from here to view in your browser when deciding how to index. *(Source: Heather Hedden, Creating Website Indexes, 2007)*

URLs of both web page file names and anchor targets are extracted into the indexing program in a single step. Subfolders are also extracted. Non-HTML files can also be indexed. These files have to be added to your project by using the 'Actions menu'. You then select files and anchors for indexing. This brings you to the 'index entry pane', illustrated in Figure 12.

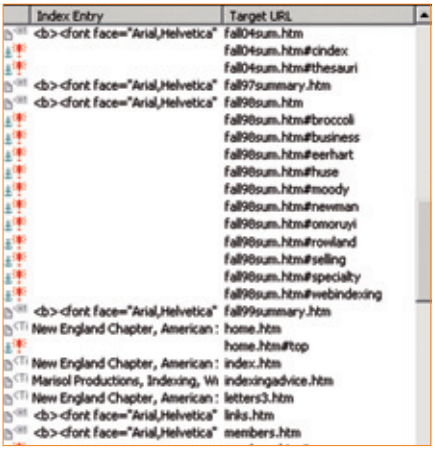


Figure 12: Index entry pane

In the right pane there are three columns. In the first column are icons indicating the source of the URL, a web page, or an anchor. The second column is for the Index Entry, which may have text that is automatically generated with something that needs to be replaced. The third column is for the Target URL, the web page name or the page name with an anchor. If there are subfolders, the folder paths will also be included in the URL.

If there is no automatically generated text for the Index Entry, which is usually the case with anchors, then a red exclamation mark appears.

There is no column for subentries. Subentries are entered as part of the main entries. *(Source: Heather Hedden, Creating Website Indexes, 2007)*

According to Heather Hedden (2006, p. 36), creating additional index entries is easier and more sophisticated than in XRefHT. And unlike XRefHT, the default text that is automatically entered is not the web page title, but the contents between the first pair of heading tags on the page. If there are no heading tags, then the page title is used. Cross-references can be added at any time. There is a special 'create cross-reference' box that allows you to make a cross-reference and also have it hyperlinked.

There are many index styles you can choose by using the 'preferences' menu. You can also use CSS (cascading style sheets). HTML Indexer comes with a default style sheet, and you can use this or create your own. Mike Unwalla (2006) has shown that, for people who know

about HTML coding, it is fairly easy to adapt HTML Indexer in various ways by including user-defined HTML code.

HTML Indexer is different from other web indexing tools in that it has a method for self-maintenance. It actually stores the index entries on the page indexed by adding a comment block, which cannot be seen by visitors to the website. There is a corresponding comment block on the index page. This ensures that no index entries are created to locations in the website that do not exist, or no longer exist.

5 Conclusion

In this article we have looked at software that is used for indexing purposes, noting that automatic or semi-automatic indexing software does not create acceptable indexes. Automatic indexing can at best provide a starting-point for further indexing. It is questionable whether this is a better starting-point than just reading the text, because only the context can help you determine whether a term is being discussed or whether it is just a passing mention that does not deserve to be included in the index.

For someone who needs to produce an index only occasionally, using general-purpose software is an acceptable solution. Each program has its drawbacks, because they were, after all, developed for other purposes. For example, these programs lack sophistication in alphabetic sorting or alternative sorting orders (page order, chronological) and are time-consuming during the editing phase, as the indexer does not have an overview of the index.

Dedicated indexing software is therefore nearly always used by professional indexers. These programs are far superior in the way they handle data entry, manipulation of data and editing of data. They are very flexible in the output formats they can produce and are adaptable to different indexing situations, such as traditional back-of the-book indexing, embedded indexing and website indexing. The three most widely used dedicated indexing programs are reviewed here. We found that their orientation and appearances are very different, but that the features (e.g. 'group function') and final output are often remarkably similar. The program of choice is therefore mainly a matter of personal taste.

For website indexing, other software solutions are available as well. And, as this is a fairly new indexing speciality, it may very well be that more and more sophisticated products will become available in this field.

We hope we have shown that indexing involves intellectual effort that is greatly enhanced by the use of software, but cannot be replaced by software (i.e., automatic indexing software).

References

Beare, Geraldine: MACREX – a History. The Indexer 24(2004)1, pp. 18–20.

Chicago Manual of Style. 15th ed. Chicago: University of Chicago Press, 2003.

CINDEX website: www.indexres.com/home.php [1st October 2007].

DEXter website: www.editorium.com/dexter.htm [1st October 2007].

Hedden, Heather: Software for HTML indexing: a comparative review. The Indexer 25(2006)1, pp.32–37.

Hedden, Heather: Creating Website Indexes. NIN workshop handout. Driebergen, August 2007.

Hedden, Heather: Indexing Specialities: Web Sites. Medford; NJ: Information Today, 2007.

HTML Indexer website: www.html-indexer.com [1st October 2007].

IndDoc web page: www-lipn.univ-paris13.fr/RCLN/outils/IndDoc/presentation.htm [1st October 2007].

ISO 999: Information and documentation – Guidelines for the content, organization and presentation of indexes. Second edition, 1996-11-15.

MACREX website: www.macrex.com [1st October 2007].

Lamb, James: Website Indexes: Visitors to Content in Two Clicks. Colchester: James Lamb, 2006.

Lennie, Frances S.: History and development of CINDEX. The Indexer 24(2005)3, pp. 135–137.

Mulvany, Nancy C.: Indexing Books. Chicago: University of Chicago Press, 1994.

Mulvany, Nancy C.: Software tools for indexing: revisited. The Indexer 21(1999)4, pp. 160–163.

Schroeder, Sandi (Ed.): Software for Indexing. Medford, NJ: Information Today, 2003.

SKY Index website: www.sky-software.com [1st October 2007].

TExtract website: www.texyz.com/textract/ [1st October 2007].

Tulic, Martin: Software for Indexing. Website article, 2004: www.anindexer.com/about/sw/swindex.html [1st October 2007].

Unwalla, Mike: Web indexing: extending the functionality of HTML Indexer. The Indexer 25(2006)2, pp. 128–130.

WordEmbed website: www.jalamb.com/wordembed.html [1st October 2007].

Wyatt, Michael: CINDEX vs Sky Index. Website article, 2001: www.aussi.org/resources/software/review.htm [1st October 2007].

XRefHT download site: http://publish.uwo.ca/~craven/freeware.htm [1st October 2007].

Index, Indexing, Software, Evaluation, Overview

Register, Indexieren, Software, Übersichtsbericht, Bewertung

The Authors

Caroline Diepeveen, MA

has university degrees in political science and international relations. She worked on various research projects and worked in development cooperation, before relocating to Scotland and taking up indexing. In 2000, she moved back to the Netherlands, where she continued her indexing career. Caroline has been an accredited indexer with the British Society of Indexers since 1998. In 2004, Caroline and others started NIN, the Netherlands Indexing Network.

Brandenburglaan 39
4333 BX Middelburg
The Netherlands
cdiep@zeelandnet.nl
www.cdiep-indexing.com

Dipl.-Dok. Jochen Fassbender

is a freelance indexer and information architect with a background in information science, based in Bremen, Germany. He is co-founder of the German Network of Indexers (Deutsches Netzwerk der Indexer – DNI) and presently its main organizer. He is also a member of the Information Architecture Institute. Since 1991 and prior to becoming a freelancer he worked in several long-term projects (indexing as well as thesaurus development) for various institutions and companies.

Indexetera – Indexing et cetera
Bremen, Germany
jf@indexetera.de
www.indexetera.de

Dr. Michael Robertson

studied English at the Universities of Stirling and York (UK). He worked in university research and science publishing for several years, and has been a freelance editor, indexer and translator, based in Augsburg (Germany), since 1991.

Editorial and Translation Services
Augsburg, Germany
editchoice@yahoo.com
www.editchoice.com

Medical indexing in the United States

Janyne Ste Marie, Green Bay, Wisconsin (USA)

This article discusses education, training, business considerations, common issues in indexing medical topics, tools used in the indexing process, and indexing standards used in the United States.

Indexieren medizinischer Inhalte in den USA

Dieser Beitrag gibt einen Überblick über Ausbildung, Training, geschäftliche Erwägungen, häufig auftretende Probleme beim Indexieren von medizinischen Inhalten, Indexierungs-Programme sowie Indexierungs-Normen, die in den Vereinigten Staaten benutzt werden.

Introduction

Medical indexers serve the portion of the publishing industry related of course, to medicine. As of this writing, there are approximately 168[1] medical publishers in the United States who publish thousands of titles each year. The subject matter of these titles covers many medical specialties, from acupuncture, dentistry, and clinical and research medicine to forensic pathology, neurology, pediatrics, psychiatry, and all points in between. Indexers who specialize in medicine must have a good science background and an ability to extract relevant information from a given passage. They must also have access to a variety of dictionaries and other references. Medical indexers work in a variety of formats, providing indexes for books, periodicals, abstracts, CD-ROMs, audio recordings, graphics, and product catalogs. According to the American Society of Indexers (ASI) Salary Survey of 2004, 14% of the indexers in the United States report medicine as a specialty. (ASI, 2004).

Education/Training

According to the ASI Salary Survey of 2004 (ASI, 2004), 33% of indexers are self-taught. 43% report employer-sponsored training, 47% trained via the USDA Basic Indexing course (a correspondence course provided by the United States Department of Agriculture), and 48% trained via mentoring relationships. Many indexers receive training in more than one venue and have reported this to the survey. Other sources of training were library school curricula, apprenticeships, video courses, and various other indexing courses taught by individual indexers.

Many indexers in the field of medicine have practical experience as nurses, laboratory or hospital personnel, or as physicians in various specialties. Others have a formal university education that includes medical coursework or a background in another field of science, such as biochemistry or engineering. Continuing education is strongly encouraged and available through many sources, most notably the ASI annual conference. It's also considered good practice to keep up with publications in specific fields of interest, just as other medical professionals do.

Business Considerations

Indexers in the U.S. are divided into two general employment categories: those who work for a corporation or publishing house and those who freelance. Merriam-Webster Online has two definitions of freelance that are relevant to this discussion:

1: a person who acts independently without being affiliated with or authorized by an organization, and 2: a person who pursues a profession without a long-term commitment to any one employer. (M-W, 2007).

Many business considerations differ according to this status. How the indexer is paid, for example, differs for the two categories. If one works in-house, one is paid a set wage, either by the hour or as a salaried employee. If one is freelance, one is paid by the project and wages are more subject to negotiation. As a freelancer, one can either be offered a project with a set fee per page or per entry, or one may be asked to submit a bid.

Payment terms are also quite variable. Employees in general, are paid monthly, twice monthly, or biweekly. Usually payment terms for freelancers are thirty days, though some publishing houses pay more frequently. Freelancers can also have problems collecting overdue payments from clients. Freelancers have very few legal protections; while an index is generally considered the intellectual property of the one who produced it and is covered loosely by copyright laws; freelance indexers are considered independent contractors. There is more than one example of a large company purchasing the assets of a smaller company and declining to purchase the freelancer debts, which has left some indexers with no payment for services rendered and no ability to collect payment.

Licensing requirements also vary. While there is no indexing license or certification currently in place, many freelancers are required to register as a business with federal and sometimes, state governments. In-house indexers are viewed as employees, while freelancers are independent contractors. Freelancers pay taxes to both jurisdictions on a quarterly basis, while in-house indexers have taxes withheld from each paycheck. Many freelancers work from home offices. In-house indexers generally work at the company location, though some telecommute via the Internet.

Common Issues in Medical Indexing

Term consistency in is a constant issue, most notably in works with multiple authors. In this type of work, each chapter is written by different authors or sets or authors. As an example, a few years ago I indexed a book titled Vascular Dementia. Among the many types of dementia referred to, were vascular ischemic dementia (VID) and ischemic vascular dementia (IVD). A query of Taber's dictionary (Taber's, 1997) showed the two to be one and the same, while a query of the PubMed online database[2] showed the greater preponderance of research using the IVD variation. Upon presenting this dilemma to my production

1 the Literary Marketplace, www.literarymarketplace.com. Accessed May, 2007.

2 PubMed database: www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed

editor, I was instructed to stick with each individual author's usage. This left me with headings at both variations, with cross-references from each to the other. This example also serves to highlight the role played by the client in index creation. As a professional indexer I follow the Golden Rule: the publishing house with the Gold makes the style guide Rules. While the above resolution wasn't my first choice, it was my client's preference. Other examples of the term consistency issue include references to Parkinson's disease/Parkinsonism, heart attack/myocardial infarction/acute ischemic cardiac event, and cancer/neoplasia/tumor. Medical procedures, causal organisms, and nearly every portion of the human anatomy can also be referred to by different terms. Since term consistency is highly desirable in most indexes, this problem must be recognized by an experienced indexer and resolved. Usually one chooses the more common, natural language usage as the main heading with cross references from the others. For example, in a book that discusses tumors, neoplasia, and cancer, I would first index each term at its occurrence in the text. At the end of the reading phase of indexing, I would determine which heading has the largest entry. I would consider the audience for the book, and then choose which term is most appropriate. Cancer or tumor would be my first choice in a book for lay readers, whereas I would choose cancer or neoplasia for research scientists. Of course as previously stated, the preference of one's client is also a big factor in this decision.

Medicine is full of specialized terminology. There are scientific binomial names, chemical names, and names of genes, to give a few examples. Some standardization must be maintained in the alphabetization of these terms or no one will know where to look for them. How does one file (1R-2R)-(-)-2-Amino-1-(4-nitrophenyl)-1,3-propandiol? Where do *c-Myc* and *v-Myc* entries go? There are many references that are helpful in this regard. The British Society of Indexers has published the Occasional Paper #3: Indexing the Medical Sciences (Society of Indexers, 2002), which is widely used here in the United States. Catalogs from companies that deal in these products, such as Merck and Sigma-Aldrich[3] as is the Medical Subject Headings database, known as MeSH[4], maintained by the National Institutes of Health.

In addition to filing order, proper formatting of entries can be quite problematic. The science of genetics is especially cumbersome; as an example, a gene name is commonly formatted in italics, as in the *Myc* examples above, or the human *SRC* gene. The protein product encoded by the gene, however, is commonly expressed in a roman or other non-italic typeface as Src protein. The determination of which form an author is referring to in a given passage is one of medical indexing's more interesting challenges. Additionally, there are different conventions to consider with respect to species. The SI publication gives these guidelines on page 49:

- human genes should be printed in uppercase italics-*BRCA1*, *CYP1A2I*
- mouse and rat genes should be printed in lower-case italics, with an initial capital letter-*Brca1*, *Cyp1A2*
- yeast genes should be printed in upper-case italics for dominant alleles, and lower-case italics for recessive alleles-*ARG2*, *arg2*
- bacterial and viral genes should be printed in lower-case italics, but may include upper-case letters-*lacZ*, *cagA* (SI, 2002).

Biochemistry also has term consistency issues. Some compounds, such as DNA, are normally indexed under the abbreviation. Other compounds, notably organic compounds, have different naming conventions that various authors use interchangeably. Examples of this include citric acid/citrate, and formic acid/formate. These forms should not be indexed separately, but rather together, with cross-references given at the alternative forms where appropriate. I must again emphasize the need for an experienced professional indexer to recognize and resolve these issues.

A general issue in publication that also affects medical indexing is repagination, also called page reflow. This often occurs in a later edition of a book due to the addition and/or deletion of material for the new edition, or if the format of the book changes, or if material is moved around while the book is in production. Browne and Jermey offer excellent advice for dealing with this issue in The Indexing Companion, pg 45:

- "Get the old book, the new book, and the old index in page number order. Obtain an electronic copy, or scan it.
- Work systematically through the old index adding the new page numbers. If chunks of pages have moved exactly to new pages, you can simply increment the page numbers. Indexing software can do this automatically, but you should start at the bigger numbers or your new page numbers will be the same as some old page numbers.
- If it is not a simple movement of old pages to new page numbers, you will have to mark the new page breaks on the old pages and change the page numbers one at a time." (Browne/Jermey, 2007)

Tools of the medical profession can also be challenging. When one has a limited familiarity with surgical instruments, how can one know if an Allis clamp is the same as an Allis forcep? Again, we are vigilant to recognize such issues and consult our references. In addition to bookshelf dictionaries, there are online discussion groups and even online catalogs from medical supply companies that provide pictures of many instruments. Merck manuals and Sigma-Aldrich catalogs, mentioned previously, are helpful here also.

Tools of the Indexing Trade

In addition to our various dictionaries and reference sources, what tools do we use in order to accomplish our task? Indexing in the United States is production work; there is always a deadline and our clients want the best index that can be produced for the lowest cost in the (usually inadequate) amount of time allotted. In the past, indexers used 3"x5" cards, writing concepts throughout the reading

3 These are available from Sigma-Aldrich, www. sigma-aldrich.com and Merck, www.merck. com respectively. Merck manuals are not free of charge. Sigma catalogs can be requested free of charge.

4 MeSH is available at this Web site: www.nlm.nih.gov/mesh/

of the manuscript, then alphabetizing, typing, and editing the entire index manually. Some indexers would use handwritten cards and type the index from those, whereas others would submit an index of numbered, handwritten cards. With the advent of computers, our work has of course, been revolutionized. Every indexer I know uses some form of dedicated indexing software. The three most common packages used in the United States are CINDEX™, MACREX™, and SKY Index™. These software packages are able to alphabetize in the desired scheme (letter-by-letter or word-by-word) at the click of a mouse button. Records can be edited by groups for completeness of and agreement between entries. For example, one can isolate all records that contain IVD or VID and compare them. The same can be done for any term or combination of terms, for example, cancer OR tumor OR carcinoma OR neoplasm. My program, CINDEX™, uses the Boolean operators AND, OR, and ONLY to accomplish this. Using any of the software packages, page numbers can be shifted in some circumstances as a group, again by the click of a mouse. There are many indexing utilities available also. For example, one utility, called EntryExpander[5], greatly eases the pain of author index creation. This utility is especially useful when many authors have collaborated on different studies over a number of years. One first enters all of the names from the bibliography in the required format, next entering the proper locators from the text as they occur. At the end, one uses this utility to create separate entries for each author at the touch of a button. This can be an incredible time-saver, as this indexer can personally attest. Another utility, a Microsoft Word plug-in called Word Embed[6], automates some of the task of embedded index creation. This is another wonderful tool, in my personal experience.

Many publishing houses now send PDF files to the indexer. PDF stands for Portable Document Format and was introduced by Adobe Systems in 1993. These files can be read with the free-of-charge application Acrobat Reader, although the purchase of a fully functional version of Acrobat gives the indexer many benefits not available in Reader, alone. The use of these files makes data entry a much easier task, allowing the indexer to copy and paste index entries. This is especially useful in the cases of cumbersome disease names, author indexes, and the names of medicines, genes, and other chemical substances. This tool also reduces typing errors by a substantial margin, with the caveat that the manuscript copy is correct in the first place. Lastly, PDF files are very useful for searching for passing mentions of terms. Many publishers fail to provide the Table of Contents for the book; I've been faced many times with the prospect of searching for mentions of a topic that seemed quite inconsequential at first reading, but actually had an entire chapter devoted to it later in the book. The quality of the index is greatly facilitated with the use of PDF files.

Standards

As has been stated by many, the wonderful thing about standards is that there are so many different standards to choose from. (Browne/Jermey, 2007) Here in the United States, the most commonly accepted standard is the Chicago Manual of Style, currently in its 15th edition. ISO 999:1996 is also a very common standard. Different publishing houses have their own internal style guides, of course. The SI publication referred to in this article is a common guideline for medical indexing practice; the American Medical Association also has a style guide, currently in its 10th edition. General index creation is discussed by many fine authors in the field, including Browne/Jermey (2007), Mulvany (2005), and Stauber (2004). For a complete discussion of general indexing theory I refer the reader to those sources.

Conclusion

Medical indexing in the U.S. is a very interesting and challenging pursuit. Workers in the field have a wide variety of backgrounds and must solve many complex problems. We work in many different fields of medicine that are considered lifetime specialties for individual practitioners, and continuing education is a requirement for success. Although some researchers are attempting to construct indexing programs to replace human indexers, no one to date has succeeded. The human brain is still required for many of the finer tasks of indexing, for example, recognizing the discussion of a topic when the formal topic name does not appear on the page. Until artificial intelligence with indexing capabilities becomes a reality, human indexers can rely on steady work in this challenging field.

References

American Medical Association Manual of Style, 10th edition. United States: Oxford University Press, 2007.

American Society of Indexers. ASI Salary Survey. Wheat Ridge, Colorado, 2004.

Blake, Doreen; *Clarke*, M; *McCarthy*, A.; *Morrison*, J.: Occasional Papers on Indexing #3: Indexing the Medical Sciences. Sheffield: Society of Indexers, 2002.

Browne, Glenda; *Jermey*, Jonathan: The Indexing Companion. Melbourne: Cambridge University Press, 2007.

Chicago Manual of Style. 15th edition. Chicago: University of Chicago Press, 2003.

Merriam-Webster Dictionary Online, www.m-w.com. Accessed May, 2007.

Mulvany, Nancy: Indexing Books. 2nd edition. Chicago: University of Chicago Press, 2005.

Stauber, Do Mi: Facing the Text: Content and Analysis in Book Indexing. Cedar Row Press, 2004.

Taber's Cyclopedic Medical Dictionary, 19th edition. Philadelphia: F. A. Davis, 1997.

Medicine, Indexing, Index, Automatic Indexing, Software

Medizin, Indexierung, Register, Maschinelles Indexierungsverfahren, Software

The Author

Janyne Ste Marie, BS

is a freelance indexer and owner of the Electronic Scribe Indexing Service. She lives in Green Bay, Wisconsin, USA. She is a graduate of the University of Wisconsin, 1989 with a Bachelor of Science degree in biology and the USDA Graduate School, 2002 for her indexing certificate. She has been indexing in the medical sciences since 2002.

The Electronic Scribe Indexing Service
120 Allard Ave
Green Bay, WI 54303
USA
jstemarie@new.rr.com
www.electronicscribe.net

5 EntryExpander is available from Leverage Technologies, www.levtechinc.com/

6 WordEmbed is available from Dr. James Lamb through his Web site, http://jalamb.com/wordembed.html

Vom „point of information“ zum „point of communication“

Integration eines social networks in wiso

Der Begriff „web 2.0“ prägt zur Zeit die Medienlandschaft. Laufend liest man von neuen Portalen, Blogs, Wikis oder anderen social networks.

Als erster Datenbankanbieter wagt sich GBI-Genios innerhalb seiner Hochschulplattform wiso daran, die eigenen Datenbankinhalte mit „user generated content“ anzureichern. Damit wird das Ziel verfolgt, dem Nutzer innerhalb des „point of information“ auch einen „point of communication“ zu bieten. Der wissenschaftliche Austausch steht dabei im Vordergrund und wird über zwei Bereiche, den Lesesaal und das Forum, gesteuert.

Ausgangsbasis für die Nutzung der neuen Funktion **„Lesesaal“** ist zunächst eine Recherche. Lässt sich ein Nutzer ein Dokument anzeigen, wird ihm über einen Link die Möglichkeit eröffnet, dieses Dokument zu bewerten. Nach einer dreistufigen Skala wird die Relevanz des Dokuments für die eigene Arbeit bewertet und die Auswahl einer vordefinierten fachlichen Kategorie (z.B. Marketing oder Finanzwirtschaft) vorgenommen. Zusätzlich können ein ergänzender Kommentar verfasst sowie eigene freie Schlagworte (tags) vergeben werden. Die Informationen werden in Echtzeit an das Dokument angehängt und können von anderen Nutzern aufgerufen und angesehen werden. Dadurch können Artikel, die besonders

Abbildung 2: Bewertungsmaske

einzelne Artikel, die im Volltext oder als Referenz vorhanden sind, bewertet werden.
Die fachlichen Kategorien, die der Nutzer vergibt, spiegeln die Struktur des Lesesaals wieder. Dabei handelt es sich um eine Art virtuelle Literatursammlung, die die Suche nach relevanter Literatur für einzelne Fachgebiete erleichtern soll. Es stehen 13 verschiedene Lesesäle zur Auswahl. Innerhalb eines Lesesaals erhält ein Nutzer beispielsweise alle Dokumente, die andere Nutzer in den Bereich

Abbildung 1: Dokumentanzeige mit Kommentar

kontrovers diskutiert werden, auch direkt in der Trefferliste oder in der Dokumentanzeige erkannt werden.

Solche Bewertungsfunktionen existieren zwar auch in Online Book Shops, aber innerhalb von wiso können vor allem auch Marketing eingestellt haben. Dozenten können dadurch auch innerhalb der wiso-Datenbank ihre Literaturempfehlungen für die Studenten aussprechen und über die Bewertungsfunktion direktes Feedback erhalten.

Voraussetzung für die Nutzung der Funktionen ist eine Authentifizierung über eine E-Mail-Adresse. Das login innerhalb von „mein wiso“ wurde für diesen Zweck um die Vergabe eines nicnames erweitert, so dass die Anonymität der Nutzer gewährleistet ist.

Die zweite Komponente innerhalb des social networks ist das **„Forum“**. Mit dem Forum soll eine fachliche Diskussion unabhängig von der Literaturrecherche in Gang gesetzt werden. Dabei wird die Struktur der Lesesäle wieder aufgenommen, die ein Grundgerüst für den Austausch darstellen soll. Innerhalb der einzelnen Gruppen (z.B. Marketing) kann jeder authentifizierte Nutzer eigene Themen einstellen und diskutieren lassen. Durch die neuen Funktionen „konsumieren“ die Nutzer nicht nur die Inhalte von wiso sondern werden selbst Teil der community, in dem sie Inhalte einstellen.

Durch Selbstkontrolle der Nutzer können beispielsweise diffamierende Inhalte innerhalb des Netzwerks gefiltert werden. Dazu befindet sich bei jedem Beitrag ein Button, mit dem die Nutzer GBI-Genios per E-Mail über die Einstellung zweifelhafter Inhalte informieren können. Wenn sich die social networks in wiso etablieren und erfolgreich sind, ist die Einschaltung eines Moderators denkbar. Für GBI-Genios wäre hierfür vor allem eine Public Private Partnership interessant, um gezielt den wissenschaftlichen Austausch zu fördern.

GBI-Genios wird die Neuerungen in wiso zum 1. Januar 2008 integrieren und automatisch allen wiso-Kunden zur Verfügung stellen. Nach der Einführung in wiso soll geprüft werden, ob die Integration der Komponenten auch für andere GENIOS-Anwendungen sinnvoll ist. Besonders interessant sind die neuen Anwendungen sicherlich für geschlossene Nutzergruppen (etwa innerhalb einer GENIOS solution).

Katrin Kaiser, München

Indexing computer books: Getting started

Richard Evans, Raleigh, NC (USA)

Expanding your indexing business? Considering a new subject area, such as computer books? Beginning computer-book indexers probably have several questions: What are computer books? How do they differ from academic works? Where is the market? Who is the audience? How difficult are they? Do I have the technical expertise required to understand and index them? Are there any tips and tricks for indexing them? What are the best tools for the job? How much money can I make? What resources are available? This article describes a process for indexing computer books that strikes a balance between cost and quality, allowing the indexer to face stacks of technical, and perhaps, unfamiliar material, and index it in an orderly and logical manner.

Register für Computer-Bücher: Wie man anfängt

Sie wollen Ihr Indexing-Business erweitern? Sie erwägen ein weiteres Fachgebiet, wie z. B. Computer-Bücher? Indexer, die neu im Bereich Computer-Bücher sind, haben wahrscheinlich mehrere Fragen: Was sind Computer-Bücher? Wie unterscheiden sie sich von akademischen Werken? Wo ist der Markt für diesen Bereich? Wer ist die Leserschaft? Wie anspruchsvoll sind sie? Habe ich das nötige technische Fachwissen, um diese Bücher zu verstehen und dafür Register zu erstellen? Gibt es Tipps und Tricks für das Indexieren? Was sind die besten Programme? Wie viel kann man verdienen? Welche Ressourcen sind verfügbar? Dieser Artikel beschreibt den Prozess des Indexierens von Computer-Büchern, der einen Mittelweg zwischen Kosten und Qualität findet, was dem Indexierer erlaubt, mit oft technischem und unvertrautem Material umzugehen und es in einer planmäßigen und logischen Weise zu indexieren.

Suppose you are an established indexer, and you want to expand your business. You have two choices: Do more of the kind of work you are familiar with, or branch out into new, and perhaps challenging, subject areas. If you choose the latter, then one potential subject area is that of computer books. Depending on your familiarity with computers, or lack thereof, such an undertaking may be intimidating. If you would like to index computer books but aren't sure that you are suited for the task, I offer you this guide to getting started.

Before I begin, though, three caveats: One, answering your questions in depth would fill a whole book. However, in the space of this article, I can give you enough information to get you started. Two, my views are admittedly those of an indexer in the United States, but I believe the principles I'm discussing generalize to indexing anywhere. Three, virtually nothing I say here is cast in concrete. There will always be exceptions to the rules, but these rules provide a good place to start.

With these warnings in mind, this article answers these questions:

- What is a computer book?
- How do they differ from academic works?
- Where is the market?
- Who is the audience?
- How difficult are they?
- Do I have the technical expertise required to understand and index them?
- Are there any tips and tricks for indexing them?
- What are the best tools for the job?
- How much money can I make?
- What resources are available?

What is a computer book? Strictly speaking, a computer book is any book describing the use of computers, either hardware or software, including related devices such as digital cameras, iPods, flash drives, and the like. From a broader perspective, the guidelines given here may be applied to any technical book organized in a fashion similar to computer books. I have indexed books about introductory physics, raising German Shepherds Dogs, and constructing water gardens, all using the methods described here.

Compared to academic books, computer books have significant differences, all of which influence your approach to the index: modularity, task orientation, content emphasis, word-to-concept ratio, controlled vocabulary, and the required "weight" of the index.

Modularity. Academic works tend to be divided into chapters with page after page of uninterrupted text. Computer books are also divided into chapters, each chapter devoted to a specific topic, but then further subdivided into two, three, even four or more levels of subheadings, each subheading briefly describing the text that follows. It is not uncommon to find computer books wherein each chapter is so self-contained that it can be treated as a mini-book in its own right.

Task orientation is a key concept to writers and indexers of computer books. It means describing (and, by extension, indexing) a computer product in terms of what a reader wants to do with it. Readers of academic books approach the books because they want to **know** something. Readers of computer books approach the books because they want to **do** something. That is, they have some task to perform. Assuming the author has been diligent in discussing the product in terms of user tasks, it is the duty of the indexer to ensure that all those tasks find their way into the index.

Types of material. Computer material typically falls into two categories: user's guides and reference material. A user's guide is task-oriented, how-to information focused on telling a reader what to do with the product. Reference material describes the features of the product, such as commands, menu options, and dialog boxes, without providing much how-to instruction. Both types of information are important, but to different audiences. Reference material is important to a reader who is familiar with the product, knows the names of the features, and wants to do something with a particular feature. User's guide information is important to a reader who needs to do something, but doesn't know what the relevant feature is called or how it works. Consider, for instance, a book about Microsoft Word. There is a feature called the **Style Area**. This is an area of information that is displayed to the left of the document text that can be opened to show which styles are applied to each document element. The reference reader

may want to "open the Style Area" but the user's guide reader may simply want to "view styles in use."

User's guide information and reference material may be split across whole books, across parts of a book, or intertwined throughout a book. It is important for the indexer to recognize which is which, to understand the requirements of the two audiences, and to satisfy both.

Word-to-concept ratio. Computer texts commonly have a much smaller word-to-concept ratio than academic texts. In academic works, it is not unusual to read several pages of dense text in order to identify a single concept. Computer books are quite the opposite. For the most part, they are written succinctly, with a minimum of prose, often introducing a task with only a brief paragraph, then using a numbered list to provide the step-by-step instructions required to perform the task. Even books that are not quite so procedural are still less dense then academic books. In extreme cases, one need only look at a task-oriented section heading, and perhaps a few sentences following it, to identify the concept under discussion.

Controlled vocabulary. The computer world has a jargon, or more accurately, a collection of jargons, all its own. Once you learn the terminology for a particular area of expertise, you will find that it is used reasonably consistently across similar books. For instance, books about spreadsheets, whether the topic is Excel, Calc, Star Office/Open Office, Lotus 1-2-3, Google spreadsheets, or Quattro Pro, are sure to talk about cells, rows, columns, functions, formulas, ranges, and related terms. Highly technical works about object-oriented programming, such as C++ or Java, will universally talk about things like classes, methods, compilers, and operators.

Understanding the vocabulary is critical. The best way to learn it is to have lived it, having worked in the domain in question and learned the vocabulary first hand. Failing that, it is sometimes possible to find online glossaries to use as reference. You need to decide whether or not, in good faith, such references are adequate. For instance, I recently indexed a book about online gaming. Though I am not a gamer, I have a passing familiarity with games and found an online glossary that helped me with terms like grinding, farming, and camping, each of which has a unique meaning in the gaming world. I felt reasonably competent to index the subject and did so. On the other hand, I was once offered an 1100-page book about artificial neural networks, written for the IEEE by a dozen authors, none of whom spoke English as a first language. All the glossaries in the world would not have gotten me through that, so I politely declined.

Light vs. heavy indexes. Indexes for computer books tend to be lighter than academic indexes. A heavy index is expensive and time consuming, though not necessarily long or dense, and requires intensive analysis by the indexer and extensive rewrite during edit, while demanding more attention to task orientation and consultation on technical issues. A light index is quick and cheap to create, requires less analysis, minimal rewrites, less attention to task orientation, and little consultation on technical issues. Consider the following: Whereas an academic text may anticipate a useful life of decades, if not longer, computer books have a life expectancy of about nine to twelve months. Thus there is some justification for spending more time and effort indexing an academic work than a computer text. Your job as indexer is, to find the middle ground, somewhere between heavy and light, where cost balances with quality.

The marketplace. In the United States, there are two primary sources of computer books: Product support books written and indexed in-house at software/hardware development shops and intended for users of the shop's product; and trade books, created for the general public by freelance authors and indexers hired by publishing houses or book packagers.

Freelance indexers in the United States rarely get work from development shops because the in-house writers do their own indexing using tools like FrameMaker and its embedded indexing features. These authors index as they write, and there is never a time when they can turn their files over to an outside indexer. Instead, freelance indexers get most of their work from publishing houses, book packagers, and individual authors.

The audience for computer books varies. One dimension of variation is the casual versus the desperate reader. Casual readers are reading without a particular goal in mind. They might have bought the book to obtain some general familiarity with the product, but have no immediate need to use the product. They might intend to read the book from cover to cover and have little need for an index. Desperate readers, on the other hand, are reading to satisfy an immediate need. They might never read cover to cover. Indeed, they will probably read no farther than to satisfy their immediate goal. They rely heavily on the index, particularly its ability to guide them from their perception of the problem to the product's ability to solve the problem.

Another dimension is technical readers versus consumer readers. Technical readers are people like programmers, system administrators, and Web designers. They already know quite a bit about the product and want more in-depth knowledge. Consumers, on the other hand, may need only superficial knowledge of common functions. Again, consider a book about Microsoft Word. Consumers may want no more than to write a letter or compose a newsletter. They may need only the basics about headings, paragraphs, and columns. Technical readers (professional writers, for instance) may want all the gory details, like styles, fields, templates, and the like. Whereas the technical reader may understand terms like "uppercase" the consumer reader may think in terms of "capital letters." Where the technical reader may be comfortable with the term "fonts", the consumer may expect to find "letters" in the index. The indexer has a duty to provide a bridge from the consumer's view of the product to the terms used by the product.

Level of difficulty. The good news is that computer texts fall on a broad continuum of levels of difficulty and if you are using indexing software in the 21st century, you already know enough about computers to be comfortable, at some level, on that continuum. Your job, when starting out as a computer-book indexer, is to determine the difficulty level you feel comfortable with. At the very lowest level, there are introductory books like the Dummies series by John Wiley and Sons. Dummies books assume the reader knows little or nothing about the topic being presented. Dummies books present a superficial introduction to topics such as personal computing, podcasting, or Web design. Higher up the difficulty ladder are books like the Bible series, also by Wiley. Bible books assume a basic knowledge on the part of the reader and cover a topic in depth. Here you might find not only the basics of using a spreadsheet, but also how to program one with macros or link one to a word processor. At this level, readers may be home or office users to whom the use of the computer is tangential to their primary activities, for example, an office worker who needs to use Outlook to manage e-mail or a writer outlining a new book. To such people, "using the computer" is not a goal unto itself, merely a means unto an end. They need to know just enough to get their jobs done. Finally, at the most difficult level, there are books that really do approach the difficulty of rocket science or brain surgery, with

topics like programming with C++ or Java, network administration, data warehousing, or computer security. These books target readers to whom "using the computer" is inextricably linked to "doing a job." Such readers need to know as much as possible about the products they are using.

Required skills. How do you know you have the right skills for the job? My rule of thumb is this: If it is a book I might reasonably expect to read, then I can reasonably expect to index it. This is also known as the bookstore criteria: If it is destined to be found on the shelves of my local bookstore, I will consider indexing it.

As I said earlier, if you having even the most basic understanding of computers, there is almost certainly a level of difficulty that matches your skills. If you are comfortable with your indexing software and some operating system like Windows or Macintosh (and a few basic tools like e-mail clients, word processors, spreadsheets, and Web browsers) you are ready to index the things you are familiar with.

If your use of computers goes beyond the basics, then your prospects for indexing expand to match. Databases, financial management software, desktop publishing, and more all become grist for your indexing mill.

The high-difficulty end of the scale gets a bit tricky. Books with a high difficulty level commonly address arcane topics such as system administration, programming, network security, data warehousing, and the like. The authors tend to be experts in their fields, writing for other experts. The likelihood that your expertise approaches that of the author or intended reader is usually slim.

For example, I recently indexed a book about the C++ programming language. The book was over 600 pages long. The author was an authority on C++, had access to professional technical support, editing, and proofreading. With all those resources, he took 20 months to write the book. He described the result as "digestible to anyone with a reasonable level of experience in C++." (Wilson, 2007, p. xxiv) As the indexer, I had no direct experience with C++, minimal access to technical resources, no access to the author, editing, or proofreading, and was expected to produce an index in about two weeks.

Since 1965 I have been a computer operator, programmer, systems analyst, technical writer, and human factors engineer.[1] I have indexed nearly 500 computer books, but still consider my skills merely "adequate" when it comes to books of this level of difficulty. That said, I did the C++ programming book, and the author was happy with it.

Rating your readiness. Here's a handy scale you can use to rate your readiness when considering indexing a computer topic:

Level 1: You have no experience at all with the product. You would need immediate training in order to use it.
Level 2: You have some background. You understand the concepts, recognize the terminology, and could probably use the product with supervision.
Level 3: All of Level 2, plus you can work unsupervised, except for problem resolution.
Level 4: All of Level 3, plus you can resolve all but the most difficult problems.
Level 5: All of Level 4, plus you are an expert, able to give guidance to others.

Most indexers of computer material probably fall into levels 2 to 4, though the rating varies with the difficulty of the material. For highly technical topics like Java programming, or Web site design, I consider myself a Level 2. For more mundane topics like Microsoft Word or Outlook, I'm a Level 4.

One word of caution about Level 1: It is possible to have no experience with a particular product, Excel, for instance, but still have experience with similar products such as Quattro Pro or OpenOffice that would allow you to index Excel. However, if you have no experience with spreadsheets of any kind, you should probably turn down an Excel or Quattro Pro book.

Once you have completed a few books successfully, feel free to push the envelope into new areas.

Tips and tricks. Because computer books are so modular, it is possible to adopt an almost assembly-line procedure to indexing them. That may sound like an over simplification, and I would be the first to admit that such a procedure may not create a perfect index. For everything I'm about to say, there is more that could be said; and for every recommendation I make, there are circumstances where something else may be more appropriate. However, if you are treading new ground, this assembly-line process will give you a structure to follow while you learn. It's not a perfect method, maybe not even the best method, but it has served me well for about 500 books and, if you don't have a method of your own, it's a good place to start.

In the following discussion, I've tried to use examples based on some well known software applications. For the most part, I have settled on Microsoft Word and use, with permission, snippets from Word for Medical and Technical Writers by Peter G. Aitken and Maxine M. Okazaki.

Selecting material. Generally speaking, there are ten targets for index candidates. In order from easiest to most difficult, they are:

1. Headings
2. Figures, tables, examples
3. Acronyms and abbreviations
4. Features and their functions
5. New or special terms
6. Definitions and descriptions
7. New and difficult concepts
8. Code samples
9. Reader's synonyms
10. Reader's perspective on tasks

This is not an exhaustive list; you may well think of other categories as you get more involved with computer books. Furthermore, the categories are not rigidly defined. Quite likely you will find topics that belong to more than one category. Consider, for instance, the Ribbon, a new feature in Word 2007. The Ribbon might fall under new or special terms, definitions and descriptions, and new and difficult concepts. Don't lose any sleep over which category is most appropriate, simply recognize the Ribbon as an important term and move on.

Headings. Task orientation should be readily apparent in headings. For instance, these headings from the Aitken book:

Locking Fields
Updating Fields
Viewing Fields

readily generate index entries like these:

```
locking fields
updating fields
viewing fields
```

These entries in turn double post as:

```
fields
   locking
   updating
   viewing
```

One word of caution. If a task is too vague or generic, it does not merit a main entry[2]. For instance, this heading:

1 One who does usability tests on computers.
2 Though it is acceptable as a subentry.

Using Fields

may result in this entry:

```
fields
  using
```

but probably not this one:

```
using fields
```

Note that task orientation is an additional consideration in computer books, but certainly not the only consideration. The indexer still needs to index topics that don't necessarily involve tasks.

Some headings may not be strictly task-oriented, but nevertheless merit an entry in your index. Headers like these:

AutoCorrect Options
Character Styles
Style Area Width

do not immediately suggest a task entry. In such cases you should look at the accompanying text to determine if there is a task associated with the object. If there is no task, you may end up with something like this:

```
AutoCorrect options, description
character styles, definition
Style Area, purpose of
```

In a pinch, you can produce a credible index for a book with task-oriented headings by simply indexing the headings. In fact, I once indexed a 1,200 page book about AOL in three days by indexing only the headings. It wasn't a great index, but the client was happy to get anything in such a short time.

Figures, tables, examples. This is one of those places where I warned you there is much more to be said. With that in mind, here is a brief overview of how I handle figures, tables, and examples.

Unlike academic works, it is usually not necessary to identify figures or tables as such in computer book indexes, so when making the decision to index figures, I apply a *distance rule*: If the figure is used to illustrate a point in the surrounding discussion, I do not index it on its own. If, on the other hand, the element is separated from its surrounding discussion by additional text or multiple pages, I usually do index it separately.

For instance, suppose you have a discussion of how to indent paragraphs in Word. Suppose that discussion goes on for several pages, ending in a figure illustrating various indents. It might be a service to the readers to let them go directly to the figure instead of reading through all the text. The result might be something like this:

```
indenting paragraphs
  description, 94–97
  figure, 97
```

Tables follow the same distance rules I apply to figures. For instance, it is not uncommon for computer books to go on at some length about certain features, such as commands, menus, or dialog boxes. At the end of such discussions, you often find a table summarizing the previous discussion. In that case, the table might merit an entry of its own. Consider a multi-page discussion of all the keyboard shortcuts in Word, followed by a quick-reference table of the shortcuts discussed. Again, you may save the readers some effort by pointing them to the summary instead of forcing them to read their way up to it.

```
keyboard shortcuts
  description, 100–107
  summary table, 107
```

The same rules apply to examples.

(Note that there is no special locator typography such as 49f or 132t in computer books.)

Acronyms and abbreviations are pretty straightforward and, for the most part, can be indexed by rote (along with appropriate double postings).

Consider these entries:

```
CPU (Central Processing Unit), 114
RAM (Random Access Memory), 95,
  107, 119
WWW (World Wide Web)
   blogs, 165
   chat rooms, 171
   children, safeguarding, 214–215
   dating, 170
```

Note that these are indexed by acronym/abbreviation, followed by the spelled out version. You could just as easily do it in reverse, with the spelled out version followed by the acronym/abbreviation. Some clients have a preference, some do not. My personal preference is to index by the acronym/abbreviation.

The rules for double posting are:

If the original only takes up a couple of lines, double post it exactly:

```
Central Processing Unit (CPU), 114
```

If the original has several subentries, double post it as a cross reference:

```
World Wide Web (WWW). See WWW
  (World Wide Web)
```

If necessary, also double post under other appropriate main entries:

```
memory
  RAM (Random Access Memory), 95,
    107, 119
```

If, during your final edit, you find both versions immediately adjacent, or within a few lines of each other[3], delete the doubleposted one and keep the original.

Features and their functions are relatively simple. Typically, the author will present the name of a feature and describe what it does. For instance, in the Aitken Word book, the discussion of the AutoCorrect feature begins with:

"[AutoCorrect] controls several options that determine how and if Word automatically corrects certain errors in your document."

At the very least, this generates an entry:[4]

AutoCorrect
  description

Subsequent reading of the text may yield more subheadings, like this:

```
AutoCorrect
  capitalization errors
  description
  reversing corrections
  turning off
```

New or special terms pop up when you are indexing a book about a new version of an existing product. If you are lucky, the author lays out such terms in a section titled something like "What's New." If not, you have to pay close attention to the text to tease out such terms. In a book about the financial planning software Quicken, for instance, there may be a discussion of tags, with a mention that they were formerly known as classes in earlier versions. A book about Word 2007 introduces the concept of the Ribbon, something not found in previous versions. It is important that the index guide the reader to these terms. In a Quicken book, for example, there should be an entry for tags, but also an entry for:

```
classes. See tags
```

In a discussion of the Ribbon in a Word 2007 book, you would need to understand

3 The exact distance is a judgment call. If they fall close enough on a printed page that the reader is likely to take in both with a glance, then delete one.

4 Assuming there will be multiple subentries for AutoCorrect in the end. If there are none, there is no need for a single subentry of "description."

how the Ribbon relates to earlier versions. If that is explicitly stated, so much the better. If it is not, then the burden is on the indexer to understand the tasks related to the new function. Given that the Ribbon replaces drop-down menus and earlier toolbars, and is used to launch and organize commands, you would need to index:

```
commands
   launching.  See the Ribbon
   organizing.  See the Ribbon
drop-down menus.  See the Ribbon
toolbars.  See the Ribbon
```

Definitions and descriptions. What is the difference between a definition and a description? It's largely a matter of length. A definition is typically a one-line statement of the form: "X is…" or "X does…" A description usually goes into more detail about X.

For instance, from the Aitken book, this definition:

"[Document properties] let you display information about the document and the user."

is followed by this description:

"Some properties are determined automatically, such as the number of words in the document, the date/time it was created, and the total editing time. Other properties are specified by the user, including title, keywords, and category."

If the topic is covered in its entirety on one page, or a short range of pages, it is sufficient to simply index it as a main entry. On the other hand, if the discussion spans a longer range, or is broke up throughout the book, you should generate a list of subheadings. When there are multiple subheadings, it is helpful to point the reader to both the definition and the description where she will find the main, or basic, discussion:

```
document properties
   custom, 49
   definition, 46
   deleting, 51
   editing, 52
   linking to content, 98
```

New and difficult concepts are similar to new and special terms, but I placed them farther down the difficulty scale because although new terms are typically explicit in the text, new concepts are often only implied. Consider a graphic artist working with Photoshop. Prior to Version 3, any change made to an image was made to the entire image. There was no way to undo changes, nor any way to revert to the pre-change image, except to save a backup before making changes. In Version 3, Photoshop introduced layers, a way to isolate individual image elements, each on its own virtual layer, enabling the artist to work on individual image elements apart from the image as a whole. Thus the concept of drawing or editing images is supplemented with a new, and potentially difficult, concept of separating images into layers while working on them. If not explicitly stated, it falls to the indexer to recognize the implications of the new feature. Layers obviously affect drawing and editing, but also – and less obviously – affect things like maintaining a history list of changes, undoing changes, and backing up images.

Code samples present both good news and bad news to the indexer. The bad news is, that unless you are conversant with the coding language (C++, Java, whatever) you simply aren't qualified to parse and index individual lines of code. The good news is that you aren't expected to.

Publishers recognize that indexers, as talented as they are, are not superhuman. We are not expected to understand highly technical material at the same depth as the author or even the expert reader. Therefore, when it comes to code samples, you only need index the concept the sample illustrates. For instance, in a book about C++, you may find a discussion of things called "shims." The discussion is broken down into attribute shims, conversion shims, and composite shims. Each discussion is accompanied by a code sample, which may run on for several pages. Assuming you are not entirely in over your head, you would be expected to recognize and index the terms: shims, attribute shims, and conversion shims, but not to read the code itself and parse out specific C++ elements.

Reader's synonyms are near the most difficult end of the spectrum: indexing reader's synonyms that may not appear explicitly in the text. The text may talk about sorting, fonts, or uppercase, but the reader may be looking for alphabetizing, letters, or capitalization. It is up to the indexer to understand the reader's perspective and index appropriate synonyms.
For instance, when indexing a book about Windows, the indexer should understand that it is likely to be read by people experienced in other operating systems, such as UNIX. The Windows file management system stores files in folders; UNIX stores them in directories. Same function, different name. To accommodate the UNIX reader, the indexer working on the Windows book should understand this and create this entry in the index:

```
directories. See folders
```

Why is this so far down the list in terms of difficulty? Dr. Carol Barnum said it best in her book Usability Testing and Research: "From the moment you know enough to talk about a product…you know too much to be able to tell if the product would be usable for a human being who doesn't know what you know." (Barnum, 2002, p. xiii)

By the time we, as indexers, become familiar enough with material to index it, we have lost the ability to see that material from the perspective of someone seeing it for the first time.

Perhaps you remember a puzzle from childhood. It might be a line drawing of a jungle scene. Embedded in the lines are hidden images of jungle animals. On first viewing, it always takes time to identify the animals. However, on subsequent viewings, you can never NOT see the animals.

Indexing user synonyms is like that puzzle. The novice reader hasn't "seen the animals" yet. As the indexer, you know where the animals are hidden and can never again see the puzzle through the eyes of one seeing it for the first time. How does one cope with that problem?[5]

For synonyms that occur naturally in the language of the text, a good synonym dictionary like Roget's Thesaurus will help. For instance, when the text speaks of "copying" something, a synonym dictionary should lead you to common synonyms for "copying": duplicating or replicating. Thus your index includes these entries:

```
copying files, 105
duplicating files.  See copying
   files
replicating files.  See copying
   files
```

If you don't have a synonym dictionary handy, and can put your index into Word, the Thesaurus feature of Word will serve. Simply right-click the term in question, then select Synonyms from the resulting menu to display common synonyms.

5 One way involves something called "usability testing." People who conduct usability tests are called (in the United States) *human factors engineers* and they study for years to learn their skills. For an introduction to usability testing, see the aforementioned Barnum book in the reference section. A full discussion of usability testing is a topic for another, much longer, article.

Unfortunately, common language synonyms are the least of your problems when it comes to computer books. (There is a reason this topic is one of the most difficult to index.) There are also what I call *contextual synonyms*: words that have the same meaning depending on some specific context of use. For instance, recall my earlier example of folders and directories. No synonym dictionary will list one as a synonym for the other. They are only synonymous when used in the context of computer-based file management systems.

Proficiency in indexing contextual synonyms takes time and experience. If you are just starting out indexing computer books, be aware that such synonyms exist and learn them to the best of your ability.

Reader's perspective on tasks is tenth on the list and arguably the most difficult. I've discussed strategies for indexing tasks that are explicit in the text, and beyond that the indexer has to deal with tasks that are implicit in the mind of the reader and may have no obvious connection at all with tasks discussed in the text.

Consider a graphic designer using Adobe Photoshop. Graphics design includes a task of "transforming shapes." Assuming the text defines "transforming" as "Changing the size, shape, location or orientation of a figure," that concept leads the indexer to index this:

```
transforming shapes, 145
```

and to create alternate entries from the equivalent tasks:

```
size, changing. See transforming
  shapes
shapes, changing. See transforming
  shapes
location, changing. See transform-
  ing shapes
orientation, changing. See trans-
  forming shapes
```

Beyond that, however, there are additional alternatives that may not appear in the text but would nevertheless be helpful to the reader. For instance, there are specific types of transformations, such as rotating, scaling, shearing, skewing, and distorting. If those terms do not appear in the text, then where do they come from?

Well, it may take some digging, but there are ways to discover them.

One way is to look at other indexes for books of the same topic.[6] I sometimes go to my local bookstore and browse the indexes of books of a similar topic. If I don't have time to go to the bookstore, Amazon.com has many books that allow you to browse their indexes online.

Another way is to look for glossaries of terms for the subject. In Google (or your favorite search engine) simply search for a topical glossary such as "Photoshop glossary" or "Photoshop terminology." My results with this strategy have admittedly been spotty. I have found some good online gaming glossaries and a couple of C++ glossaries, but I did not find a Photoshop glossary that included "transforming" while I was writing this paper.

Still another resource is a good computer dictionary, though it may take some effort to find one. Once again, I was unable to find an online dictionary that included "transforming" but I did find a helpful hard copy on my bookshelf. The Microsoft Computer Dictionary offers this: "[Transforming is] to alter the position, size, shape, or nature of an object by moving it to another location (translation), turning it (rotation), changing its … coordinate system." (p. 449)

From this, the indexer can deduce these additional tasks:

```
turning a shape. See transforming
  a shape
rotating a shape. See transforming
  a shape
translating a shape. See transform-
  ing a shape
coordinate systems, changing. See
  transforming a shape
```

And so it goes. Over time you identify more and more reliable resources until eventually you have many of these terms at your fingertips.

Tools for the job. Speaking strictly of tools for back-of-the-book indexing, there are two kinds of software: dedicated and embedded indexing software.

Dedicated software is, well, dedicated. That is, all it does is help the indexer prepare an index. It is separate from, and does not interact with, the text being indexed. The indexer works from proofs that are in final pagination and creates entries and locators in a separate file. The dedicated program does what computers do best, the tedious, repetitive, mind-numbing activities of indexing. Computers can:

- Alphabetize
- Combine like entries
- Suppress duplicate headings
- Check cross references
- Check spelling
- Format the overall index

This software frees the indexer to do what human beings do best, intellectual analysis and decision making. The indexer is free to:

- Select significant material
- Recognize contextual differences
- Compose entries
- Create synonyms
- Identify implied concepts

With dedicated software, the growing index is always visible. Changes can be made on the fly. The indexer can isolate and work on groups of records, such as specific page ranges, chapters, or character strings. The format of the entire index can be changed with a few keystrokes: from indented to run-in, applying initial caps to headings, or emphasizing headings with special typography.

The output of a dedicated program is typically an RTF file, which the client imports into whatever publishing software they are using.

The three most popular dedicated indexing programs in the U.S. are CINDEX, MACREX, and SKY Index.[7] All three offer much the same features for about the same price, differing mostly in their user interfaces. CINDEX offers versions for Windows and Macintosh, MACREX for Windows and DOS, and SKY Index for Windows. All three offer trial versions and choosing one is mostly a matter of your operating system and which user interface you prefer.

The downside of dedicated software is that, because it is detached from the text, the locators do not change if the text reflows. The indexer must identify changes and update the index accordingly. The dedicated software has some limited ability to automate the changes, but the process is nevertheless mostly a manual one. This is generally not a problem among U.S. freelancers who work for publishers because their books are almost always in final pagination when the indexer gets them. However, if you find yourself indexing a book while pagination is still in flux, a book destined for translation, or a book intended to be repurposed as electronic media, you need something to automatically track and change locators as the document changes. This something is called embedding software.

6 Bear in mind that you do not know the level of expertise of the other indexer and therefore may not be able to trust his choice of terms. Take what you find with a grain of salt.

7 Contact information for these, and other programs, is on the American Society of Indexers (ASI) Web site at: http://asindexing.org/site/software.shtml.

Embedded indexing software is found as a feature of document processors such as Word, FrameMaker, and InDesign. It is intended for those indexes that must automatically repaginate when the text of the book changes. It accomplishes this by embedding special codes (tags) in the source document. The codes are tied to a specific location in the document and move with text when it moves. The index automatically changes as pagination changes. This allows not only indexing on the fly, as the book is being written, but also republishing after updates or translation without having to reindex from scratch.[8]

There are, however, serious disadvantages. First and foremost, creating entries is cumbersome, requiring the indexer to type not only the text of the entry, but surrounding tags to identify the entry to the document processor. Errors are common, and editing is especially difficult because the index is typically not visible as it grows. Instead, the indexer has to run a separate compilation step to generate the index, identify errors in the generated index, then track down and correct the embedded tags that generated those errors. The process is repeated as required. This makes creation of the original index a time-consuming process.

Also, embedded indexing software typically does not have the formatting flexibility found in dedicated software. With dedicated software one can format a see reference simply by setting a preference: adding punctuation, making it capitalized, or italicizing it. The preference then applies to all occurrences in the index. With embedding software, it is not uncommon to have to format each occurrence individually.

Nancy Mulvany, in the second edition of Indexing Books, has little good to say about embedded indexing software. The following is an excerpt (paraphrased slightly) from a review of her book that I wrote for the Society for Technical Communication (STC). It appeared in their journal Technical Communication, 54.1 (February 2007): 125–126.

> Some document processors have included indexing features since the 1980s. Since then, document processing capabilities have increased dramatically, while the indexing features have lagged farther and farther behind. Mulvany goes into some detail about the shortcomings, categorizing them as "practical problems" and "technical problems."
>
> Practical problems include managing multiple files per document; avoiding corruption of index tags when text is changed, moved, or deleted; coping with demands of single-sourcing; and, last but certainly not least, the cumbersome nature of the user interface. The indexer cannot see the index as it evolves but must take a separate compile step to generate the index. Any errors found in the compiled index must be traced to the source tags to be corrected. Recompile, correct, recompile, ad nauseam. When compared to indexing the same material with dedicated software, Mulvany estimates embedding can add 40% to the time required to edit the final index. (p. 258)
>
> Technical problems include program-imposed limitations on formatting the index, sorting entries, manipulating entries, and placing and verifying cross references. Embedding software requires more typing on the part of the indexer. Most embedding software lacks an inversion feature – a means of automatically inverting "print drivers, installing" to "installing print drivers." An indexer who wants both entries has to type both, plus any special markup required to identify each entry to the document processor.
>
> Mulvany (p. 269) sums up this sad state of affairs: "While the lack of functionality in embedded indexing software was an annoyance during the 1980s and 1990s, today it is a costly detriment to information access. The need for improved embedded indexing tools is critical."

In the U.S. virtually all freelance indexers use dedicated software, while most in-house indexers use embedded indexing software.

Which is best for you? It depends. For making your indexing activities easier, dedicated software wins easily. However, if you need the automatic update features of embedded indexing software, then that's your choice. And, of course, if your client or employer insists on a particular tool, then the Golden Rule applies: ***He who has the gold makes the rules.***

Making money. How much money can you make indexing computer books? That's difficult to answer in terms that German indexers will find useful. I'll do the best I can, based on my experience in the United States, but conversion between billing methods and currencies is left as an exercise for the reader.

That said, computer book indexers in the United States mostly bill by the page of text. Occasionally, one may bill by the project, but that typically happens when the job involves more than straight indexing: Updating a previous index or combining several indexes, for instance. Rates for indexing computer books vary from about $2.25 per page to $5.00 per page, depending on density and complexity of the material. Most of my work is in the $3.00 to $3.50 per page range.

That figure alone doesn't mean much until it is converted to an hourly rate. In her newsletter iTorque, Nancy Mulvany recommends that freelance indexers strive for at least $50 per hour for a living wage. (pp. 1–5) That covers taxes (federal, state, and Social Security), health insurance, and the overhead of running an office. Social Security takes fifteen percent. Not all states impose state income taxes, but where I work (in North Carolina) state taxes are another eight percent. Federal income tax rates vary, depending on how much you make. The more you make, the more you pay. Rates start at about ten percent and can go as high as thirty-five percent, though few indexers are likely to reach that bracket. Most fall in the 10% to 28% range.

So, starting from $50 per hour and deducting 15% for Social Security, 8% for state income tax, and 25% federal income tax, you are left with $24 per hour. From that, further deduct your operating expenses and health insurance. You quickly see that $50 per hour is not a lot.

To calculate your effective hourly rate when billing by the page[9], multiply your page rate by the number of pages per hour that you can index. If you average about twenty pages per hour (not an unreasonable figure), you would have to charge $2.50 per page to make $50 per hour. My rates on individual projects vary greatly, from a low of about $14 per hour to one or two that exceeded $100 per hour. On average, I achieve well over $50 per hour.

Resources. Here are some resources you will find useful if you decide to start indexing computer books.
The American Society of Indexers (ASI) Web site, in addition to being a good general resource, also has a number of relevant publications, including *Software for Indexing*, a comparison of popular indexing software.

8 In practice, it's not that easy. Unless the indexer of the subsequent edition is intimately familiar with the previous index, it is all too possible to unintentionally delete index tags when deleting or updating text. This leads to problems like missing cross references. Most U.S. indexers clear out the old tags and reindex from scratch.

9 When billing by the project, simply divide the total income by the hours spent on the project.

www.asindexing.org/site/asipub.shtml

Two quick introductions to indexing technical material: *Read Me First: A Style Guide for the Computer Industry* and *Indexing: A Nuts and Bolts Guide for Technical Writers.*

Sun Technical Publications. Read Me First: A Style Guide for the Computer Industry. Mountain View, CA: Sun Microsystems, Inc, 1996. ISBN 0-13-455347-0.

Ament, Kurt: Indexing: A Nuts and Bolts Guide for Technical Writers. Norwich, NY: William Andrews Publishing, 2001. ISBN 0-8155-1481-6.

Nancy Mulvany's *Indexing Books* is literally "the book" on indexing. It is also a source for evaluating the state of embedded indexing software. Be sure to get the second edition.

Mulvany, Nancy: Indexing Books. 2nd ed. Chicago, IL: University of Chicago Press, 2005. ISBN 978-0-226-55276-7.

Here are several computer dictionaries:

Downing, Douglas; *Covington*, Michael; *Covington*, Melody: Dictionary of Computer and Internet Terms. 8th ed. Hauppauge, NY: Barron's Educational Series, 2003. ISBN 0-7641-2166-9.

Microsoft Computer Dictionary: Fourth Edition. Redmond, WA: Microsoft Press, 1999. ISBN 0-7356-0615-3.

Pfaffenberger, Bryan: Webster's New World Computer Dictionary, 10th ed. Indianapolis, IN: John Wiley & Sons, 2003.

A whitepaper titled "Creating Usable Indexes: A Systematic Approach to Editing Indexes for Quality and Usability" written by yours truly is available from BrightPath Solutions at:

www.travelthepath.com/whitepapers.html

A list of indexing software and related contact information is at the American Society of Indexers (ASI) Website at:

http://asindexing.org/site/software.shtml

That about wraps it up. I've given you a broad, but not very deep, view of what it takes to get started indexing computer books. One of the drawbacks of hitting only the high points is that it makes for a bumpy ride. I hope you have found this article useful and, if you decide to begin indexing computer books, I hope you are successful. I find computer books fascinating and frankly don't know if I would be a professional indexer if I had to work on some other type of material that didn't lend itself to such a systematic approach. I occasionally accept topics outside my area of expertise, such as the role of Franklin Roosevelt in the Spanish American War, or the school system in East Asia, and I am always glad to get back to something technical.

If you recall what I said about light versus heavy indexes, the process I've described allows me to strike that balance between cost and quality, to be able to sit facing stacks of technical, and perhaps, unfamiliar, material, and index it in an orderly and logical manner. Perhaps it can do the same for you.

Acknowledgements

My sincere thanks to Peter Aitken and Maxine Okazaki for their permission to use parts of their book about Microsoft Word as examples in this paper.

Aitken, Peter G. and *Okazaki*, Maxine M: Word for Medical and Technical Writers. Chapel Hill, NC: Piedmont Medical Writers LLC, 2007. ISBN 978-1-60402-445-6.

References

Aitken, Peter G.; *Okazaki*, Maxine M: Word for Medical and Technical Writers. Chapel Hill, NC: Piedmont Medical Writers LLC, 2007. ISBN 978-1-60402-445-6.

Barnum, Carol: Usability Testing and Research. Pearson Education, Inc., 2002. ISBN 0-205-31519-4

Evans, Richard: Review of Indexing Books. Technical Communication 54(2007)1

Microsoft Computer Dictionary: Fourth Edition. Redmond, WA: Microsoft Press, 1999. ISBN 0-7356-0615-3.

Mulvany, Nancy: Indexing Books. 2nd ed. Chicago, IL: University of Chicago Press, 2005. ISBN 978-0-226-55276-7

Mulvany, Nancy: How to make money indexing books. iTorque, 2003 ISSN 1542-6467

Wilson, Matthew: Extended STL, Volume 1: Collections and Iterators. Person Education, Inc, 2007. ISBN 032-1-30550-7

Indexing, Computer, Book, Cost, Software, Usability

Register, Buch, Computer, Inhaltliche Erschließung, Informatik, Kosten, Software, Benutzerfreundlichkeit

The Author

Richard (Dick) Evans, MA, BA, AA

has over forty years experience in the computer industry. He has been, at various times, a computer operator, programmer, systems analyst, technical writer, and human factors engineer. In that last role he became interested in indexing while conducting usability tests on indexes for telecommunications manuals, though he contends that his interest goes back much farther, to his earliest computer experiences when he struggled to find critical information buried in volumes of computer manuals.

He retired from corporate life in 1992 and started Infodex Indexing Services, Inc. Now he specializes in indexing computer books for both corporate clients and publishing houses, as well as providing workshops that teach technical writers about indexing.

Dick is a past president of the American Society of Indexers (ASI) at both the chapter and national levels.

Infodex Indexing Services, Inc.
13200 Strickland Road
Raleigh, NC 27613
USA
infodex@mindspring.com
www.mindspring.com/~infodex

Book-style indexes for websites

Heather Hedden, Carlisle, MA (USA)

Site A-Z indexes are back-of-the-book style indexes with hyperlinked entries, and they are a useful method for searching for information in a website. This article examines the types of sites for which an A-Z index is most suitable and then compares website indexes with book indexes with respect to structure, locators, subentries, and cross-references. Certain modifications need to be made to adapt an index to the web. Although not complicated to create, continuous updating is an issue, and site indexes are not applied as much as they could be.

Buchregister als Vorbild für Website-Indexe

Site-Indexe entsprechen Buchregistern mit Hyperlink-Einträgen und sind eine nützliche Methode zur Informationssuche innerhalb einer Website. Dieser Beitrag untersucht die Arten von Websites, für die sich ein Site-Index am besten eignet, und vergleicht Site-Indexe mit Buchregistern bezüglich Struktur, Fundstellenangaben, Untereinträgen und Querverweisen. Bestimmte Modifikationen sind nötig, um einen Index den Erfordernissen von Websites anzupassen. Obwohl es nicht schwierig ist, einen Webindex aufzubauen, ist die fortlaufende Aktualisierung eine Kernfrage. Site-Indexe werden nicht so häufig eingesetzt, wie sie könnten.

To make information easily found in books, there has evolved the standard format of (1) a table of contents to provide a page-ordered list of the book's chapters as a kind of outline, and (2) an alphabetical index, usually at the back of the book, listing all the topics and names, and their synonyms or variant forms, that are worth mentioning, along with their corresponding page numbers. Not only has it become standard to have a table of contents and an index in a nonfiction book, the styles of indexes have become somewhat standardized as well. While there exist different formats, such as hanging versus run-on subentries, a freelance indexer has no difficulty accommodating a limited number of styles for various publishers, even in different countries.

Making information easily found in websites, however, is a very different matter. As more and more information gets put on websites, and the importance and size of websites has grown, website owners and designers have attempted various methods of making information on the sites more "findable." This fast-growing and changing medium, however, does not have the established standards that exist for printed books. Electronic data and hypertext links can be exploited for various new methods by which users can locate information. Methods include:

- navigational menus, which may include second and third level drop-down menus
- search engines, which may or may not take metadata keywords into consideration
- hierarchical taxonomies, which may or may not make use of a thesaurus or a content management system
- site maps, which might list all pages or only the top two or three levels of pages
- site indexes, which may or may not follow book index styles

Not only do different software technologies allow various search and navigation methods, but the immense variety in the kinds and sizes of websites require different methods for information searching. Finally, as we are aware from books, one method of finding information is never sufficient. Just as books have both a table of contents and an index, a website should also have two or three methods, perhaps more, for its users to locate information. A website index may often be one of these methods.

Definition of website indexes

Although technology can support more elaborate methods of searching, a site index is an option that is definitely worth considering. A site index, as we define it, is an alphabetical arrangement of topics, similar to that of a book's index, where each topic is hyperlinked to the referred web page. A site index is a method of search that is:

- familiar to users from the world of books, and thus very easy to use
- relatively low technology, relying only on HTML code, so that it can be implemented on the most simple website by people without programming/coding expertise
- more accurate in its retrieval results than site maps, navigation menus, or search engines

The term "website indexes" could have different meanings. Thus, for clarification, the browsable alphabetical back-of-the-book style index on a website is often called an "A-Z index." The implication of "A-Z" is that there is an alphabetical browse view or interface. This differs from the ordered view of a site map that looks more like a table of contents. Clarification is needed because some web designers labeled site maps as site indexes. The A-Z index is also different from an index that involves a search box. Even if words entered into the search box are matched against a human-created index, thesaurus, or controlled vocabulary, the index is not visible to the user and therefore not browsable in an A to Z list.

On websites, A-Z indexes are typically called in English (listed in order of popularity):

- Site index
- A-Z index
- Topic index
- Alphabetical index
- Browsable index

Appropriateness of site indexes

Almost all nonfiction books can and probably should have indexes. This is not the case with websites. Some are too small or too large, some change too frequently, and some have content better searched through a taxonomy or a database rather than with a traditional alphabetical index. Website size and changeability are the most important factors, since maintaining an A-Z index can be a lot of work. The type of content and type of users may also be an issue.

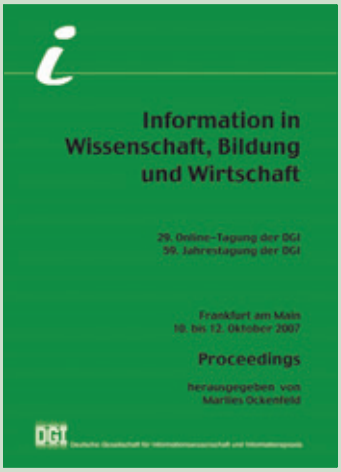
Extremely small websites do not require indexes, and indeed a very small index

Information in Wissenschaft, Bildung und Wirtschaft

herausgegeben von Marlies Ockenfeld

Das Motto der DGI-Online-Tagung Information in Wissenschaft, Bildung und Wirtschaft ist der Rahmen für ein vielseitiges Programm, das Information Professionals aus vielen unterschiedlichen Aufgaben- und Einsatzfeldern interessante Einblicke in informationswissenschaftliche Forschung, praktische Anwendungen, das Berufsfeld und geplante Projekte bietet: Informationsportale, Informationsarchitektur, Patentometrie, Such- und Antwortmaschinen, Web 2.0, Dokumentation im Gesundheitswesen sowie E-Journals und Open Access, um nur einige zu nennen. Neben den Beiträgen, die aufgrund des Call for Papers eingereicht und nach einer Begutachtung durch das Programmkomitee als Vortrag angenommen worden sind, wurden insbesondere für die Auftaktveranstaltung und die beiden ersten Sitzungen über Trends, Herausforderungen und Perspektiven zusätzliche Fachleute als Rednerinnen und Redner eingeladen. Die begutachteten Beiträge sind als Volltexte enthalten. Von den eingeladenen Vorträgen gibt es Zusammenfassungen. Insgesamt enthält der Tagungsband 31 Langfassungen und elf teilweise ausführliche Zusammenfassungen.

Leserkreis
Informationswissenschaftler, Information Broker, Wissensmanager, Medienfachleute, Bibliothekare, Content Anbieter, Verlagsmitarbeiter, Indexer, Fachangestellte für Medien- und Informationsdienste, Informationsassistenten, Studenten



Marlies Ockenfeld (Hrsg.)
Information in Wissenschaft, Bildung und Wirtschaft. Proceedings.
Frankfurt am Main 2007, 350 Seiten, DGI-Tagungen – Band 9
ISBN 978-3-925474-60-6, EUR 50,– (für DGI-Mitglieder EUR 40,–)

Bestellungen an die DGI-Geschäftsstelle, Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Fax (069) 4 90 90 96, publikation@dgi-info.de

does not work well. A minimum would be about 25 pages, but this also depends on how much information the pages contain. In any website, there will be some pages not suitable for indexing. Extremely large sites also pose a problem. For websites or intranets of over 1000 pages, not only is it a lot of work, but also by the time an index is complete, the site will likely have changed. This does not mean that large sites should not include indexes. Rather than indexing the entire site, individual indexes can be applied to individual subsections.

A major issue in website indexing is that, unlike printed materials, websites and intranets can change quite frequently. Content within pages changes and new pages are added and old ones deleted from time to time. Websites that change especially frequently, such as a website dedicated to a special event, should be avoided in indexing. A website with changing sections could still be indexed, but the indexer should omit pages known to be temporary and refer to pages whose content changes by general topic of the pages only and not to specifics.

A website rich in content and with a variety of content is better suited for indexing. Some sites may have a sufficient number of pages, but insufficient indexable content. These would include pages displaying mostly images (for graphic effect rather than as content), components of online games, or short directory entries. Sites that contain a lot of content but all of a similar type, such as a catalog of products or a directory of names or organizations, are also not so suitable for indexes. In some cases, all the information can easily be arranged in categories. A directory-type site might require an alphabetical list of names to look up, but this is not what we would consider a true index. Most sites that sell products do not need indexes if most of the pages comprise a listing of products. Customers tend to look up products by category, not by alphabetical names.

Websites that tend to get repeat visitors make especially good candidates for indexes. They are analogous to reference books, that researchers repeatedly go back to and where an index is especially appreciated. Of course, indexes can be useful on websites that receive many one-time visitors, but often the objective of such websites is to draw in the visitor to explore the site, rather than come in, get information, and then leave. Sites that have high repeat visitors, and thus are good candidates for indexes, include company intranets visited by employees, educational institutions by students, organizations by members, and municipal sites by residents.

Structure of a site A-Z index

The A-Z index is often just a single page within a website. But if the index is very long, it may be broken up into multiple pages, such as one for entries that start with each letter of the alphabet or for a range of letters of the alphabet. When an index page is long, occasional "Back to Top" or "Top of page" jump links are also useful additions to the page. These may be inserted after each letter section or after multiple letter sections if the sections are short.

Compared with a printed book index, which is usually in two columns to save space, there is no need to save space on a scrollable web page, so a site A-Z index can be left in a single column. Furthermore, since at least some scrolling is expected, it's actually preferable to have the website index in a single column, since it is less practical for the user to have to scroll up and down between two or more columns on the same page.

Whether a website index is on one page or divided into multiple pages, the standard method of navigating the entire index, which is too large to fit on one screen, is to present the letters of the alphabet in the initial display and have these letters hyperlinked to the section of the index that begins with that letter of the alphabet. The letters of the alphabet may be displayed only at the top of the index or index page, or also at the bottom of pages and in between letters, or in a separate frame that is always visible.

There is also a choice of style as to whether all the letters of the alphabet are displayed, including those for which there are no index entries, for example Q and X, or to only list letters for which there are index entries. In the Boston-IA topic index example (Figure 1), the letters with no entries (Q and Y) are simply omitted. Alternatively, if there are no index entries for a letter, then the letter is simply not hyperlinked and should display in a different color, or be "grayed out".

Hyperlinked Locators

In printed indexes, entries or subentries are followed by one or more page numbers (locators), referring the reader to the page of the book with the topic of the index entry. In a website index, there are typically no page numbers, rather the text of the entry or subentry itself is hyperlinked to the page within the website where the topic is discussed.

A complication with website indexes is that a hyperlink can go to only one destination. So, when relying on the

Figure 1: Example of an excerpt of a website index.

hyperlink of the text of the entry as the means of jumping to the source content, there cannot be multiple locators, in contrast to a series of page numbers following an index entry when the topic is found on multiple pages. Having hyperlinked page numbers following the entry is a possibility, but the page numbers themselves are rather meaningless in a website. Thus, it is rare to see them in website indexes. There are ways, though, to get around this problem in website indexes:

- Create additional subentries
- Reword the subentries
- Create sub-subentries
- Decide that one mention of the topic does not provide original information, so omit it from the index

As it turns out, the issue of how to handle multiple undifferentiated locators in hypertext is a greater problem in theory than it is in practice. Many websites are written more concisely than books and information is not repeated as often. While the linear nature of print documentation requires a concept to be introduced in an introduction, then discussed in detail, then summarized, web content, by contrast, tends to get to the point. If indexing a book converted to HTML, however, the problem of multiple locators is greater.

Not only do print indexes have multiple page numbers after many entries, but they also have page ranges, such as: pp. 67–70. Page ranges also cannot be indicated in hypertext, but in practice this is not much of an issue in website indexes. If several pages of a website all discuss the same topic, then they are likely to be sub-pages of an intermediate page introducing that section of the website. The index entry then can link to the intermediate page. Also, websites have no restriction on page length, so if there is a lengthy discussion on one topic, it may be covered in a single, lengthy page. The general design of websites is to create one web page for each topic, no matter the length of the page.

Thus, the nature of traditional websites makes them relatively easy to index with a single locator/link per entry. Websites that are collections of articles, however, are more challenging. But the challenges are the same as in print when creating a cumulative index of periodical articles.

Hypertext locators also have an advantage over page numbers. Instead of pointing to an entire page, index entries can be more precise by being linked to specific points (sections, paragraphs, etc.) within a page by inserting HTML anchors to link to within the page. The named anchors may already be present in the web pages, but often the indexer may feel the need to create a few more named anchors. Adding named anchors is the only additional task the website indexer has beyond writing the index itself. In most websites, where the pages are not unusually long, creating index entries that point to an entire page, and not a more specific anchor point, is sufficient.

Not every entry in an index is hypertext and linked. If there is a main entry with multiple subentries, the main entry typically is not linked to any pages, but serves as a gathering point for the subentries. This is the same style as in book indexing. In the Boston-IA topic index, the entry "adaptive technologies" is not hyperlinked, since it has no reference of its own, but is rather the gathering term for two subentries. It should be clear which entries are hyperlinked to content and which are not by use of different color/use of underline for the hypertext.

Orphan subentries

In book indexing, a single subentry under a main entry (also called an "orphan" subentry) is considered bad style. If there were a general discussion of indexes on page 4, and a more specific discussion of website indexes on page 6, rather than an index entry of the following:

```
indexes, 4
  on websites, 6
```

in a book index, it would be entered as:

```
indexes, 4, 6
```

Since multiple locators are not supported in a hyperlinked website index, the index entry should be structured as:

```
indexes
  on websites
```

Here the main entry word "indexes" would be hyperlinked to the general discussion on one page, and the subentry phrase "on websites" would be hyperlinked to the specific discussion of website indexes on a different page.

Cross-references

In a website index, not only are the entries hyperlinked, but so are the cross-references of *See* or *See also*. In website indexes, the term following the word *See* is hypertext, and the term before it is not hypertext. In the Boston-IA topic index excerpt there is a cross-reference, where the hypertext is underlined:

```
accessibility, web. See web
  accessibility
```

Clicking on web accessibility jumps to this entry lower down in the index.

See cross-references are used in print indexes, as an alternative to "double-posting," typically when synonymous main entries have subentries so it would waste space to have all the subentries repeated under each double post. On websites, there is no concern for wasting space, so the *See* cross-references are less often used for this purpose, and equivalent variants, tend to be entered, unless the list of subentries is very long.

Another use of a ***See*** reference is to educate the user as to the preferred term. If this is the case, then such a *See* reference would also be used in a website index, but, if there are no subentries, there is no need to make the user jump somewhere else within the index before going to the source text. The *See* reference can name the preferred term, but the preferred term can link directly to the source text, rather than to its position within the index.

See also cross-references are just as helpful in website indexes as they are in print indexes, for guiding users to related terms. Whether, the link jumps to the referred term within the index or directly

Das Buchregister

Methodische Grundlagen und praktische Anwendung

von Robert Fugmann

Angesichts der automatisierten Aufbereitung von Buchinhalten in Suchmaschinen und Portalen gewinnen Sach- und Fachbücher wieder mehr Aufmerksamkeit. Sie leiden jedoch häufig unter fehlenden oder mangelhaften Registern. Dort wo es sie gibt, handelt es sich nicht selten um schlichte automatisiert erzeugte Textwortextraktionen.
Fugmann beschreibt die Fallstricke und Herausforderungen beim Registermachen und warnt vor einer Fehleinschätzung des Indexierungsproblems. Mit dem Standpunkt des Physikers Ludwig Boltzmann „Nichts ist praktischer als Theorie" gibt er eine theoretische Einführung in die Problematik und weist auf Anfängerfehler hin, die man bei der Indexierung von Büchern vermeiden sollte. Gleichzeitig zeigt er am Beispiel des vorgelegten Buchs, wie sich informative Register anfertigen lassen.
Buchregister von Sach- und Fachbüchern müssen, so die Kernforderung des Autors, ihren Lesern ein Instrumentarium bieten, mit dem sie auf Entdeckungsreise gehen können, und nicht zuvor das Buch gelesen haben müssen, um einschlägige Inhalte aus der Erinnerung heraus wieder zu finden. Für diesen Dienst am Leser zeigt er neue Wege und Werkzeuge auf.

Leserkreis
Autoren, Indexierer, Content Anbieter, Verlagsmitarbeiter, Technische Dokumentare, Information Broker, Informationswissenschaftler, Studenten

Robert Fugmann
Das Buchregister. Methodische Grundlagen und praktische Anwendung
Frankfurt am Main 2006, 136 Seiten mit Sachregistern, DGI Schrift (Informationswissenschaft – 10)
ISBN 3-925474-59-5, EUR 25,– (für DGI-Mitglieder EUR 20,–)

Bestellungen an die DGI-Geschäftsstelle, Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Fax: (069) 4 90 90 96, publikation@dgi-info.de

to the source text depends also on the presence of subentries. If there are no subentries under the referred entry, then there is no need to make the user click through another entry in the index.

A well-designed website index will use a combination of intra-index and direct-to-page links, as appropriate, for the ***See*** and ***See also*** references. The following table may also be used as a guide.

Table 1:

Main Entries	Subentries	
	None or Few	Many
terms are equal	no cross-references; repeat terms	perhaps cross-reference linked to a preferred term
one term is preferred	cross-references linked to text	cross-reference linked to preferred term

A-Z indexes on the Web

On English language websites, whether in the United States, Canada, the United Kingdom, or Australia, most of the websites with A-Z indexes tend to be either universities or government agencies. Called either "A-Z index" or "site index", more often than not, these indexes unfortunately lack subentries and sufficient double-posting or cross-references. At a quick glance, it may be difficult to discern whether they are really indexes or merely alphabetically sorted web page titles. True book-style indexes can be found, but they are not as common. Collected examples of good A-Z indexes are listed on the Web Indexing Special Interest Group site: www.web-indexing.org/web-index-examples.htm.

On German sites, "A-Z Register," "Site Index," "A-Z Index" are also sometimes used. As with English sites, "Site index" is sometimes used to designate a non-alphabetical site map. The "A-Z" indexes are usually without subentries, double-posts, or cross-references. True back-of-the-book style indexes on German websites are thus still extremely rare. One example is the Deutsche Zentralbibliothek für Medizin www.zbmed.de/siteindex.html. In this example, the subentries are not indented as in books, but follow an arrow symbol, and there are also some double-posts in this index. A good example of a German website index with indented subentries as in a book is that on the site of the Deutsches Netzwerk der Indexer www.d-indexer.org/si.html.

In conclusion, despite the development of new search and navigation technologies for websites, the traditional back-of-the-book style index still serves users and certain kinds of websites very well and could still be implemented more than it is. A site index needs to be continually updated, however, as web pages are added to the site. While there is no standard accepted style for website A-Z indexes, certain modifications should be made to a book-style index to make it more usable on a website. Converting a book index to HTML, however, has the additional problems of dealing with multiple locators. By taking advantage of hypertext, a site A-Z index is even easier to use than a printed index. With the aid of website indexing software [listed in Glenda Browne's article], indexers who write indexes for books can also write indexes for websites.

References

Hedden, Heather. Indexing Specialties: Websites. Medford NJ: Information Today, 2007.

"Best Practices for Style and Formatting of A-Z Browsable Site Indexes" on the Web Indexing SIG. www.web-indexing.org/practices.htm. [26 August 2007].

Website, Index, Software

Elektronischer Dienst, Sachkatalogisierung, Index, Website

The Author

Heather Hedden, MA

an indexer and taxonomist was manager of the Web Indexing Special Interest Group of the American Society of Indexers (2005–2006) and president of the New England Chapter of the American Society of Indexers (2006).

Hedden Information Management
98 East Riding Drive
Carlisle, MA 01741
USA
heather@hedden.net
www.hedden-information.com

Changes in website indexing

Glenda Browne, Blaxland (Australia)

Website indexing became important in the 1990s, as indexers, librarians and web managers experimented with different approaches for making the information they were providing on the Internet more accessible. The tools for creating A to Z indexes have changed over time, from simple HTML coding to HTML Indexer and other specialist software. New indexes and software products have been created, but many website indexes have also been lost. This article considers the reasons for creating website indexes and includes examples of website indexes and other access tools which were created in the last 15 years, but which are no longer available. It also provides an assessment of some of the reasons for the changes.

Wandel beim Indexieren von Websites

Website-Indexing wurde in den 1990er Jahren wichtig, als Indexierer, Bibliothekare und Web-Manager mit verschiedenen Ansätzen experimentierten, einen besseren Zugang zu Informationen zu schaffen, die sie über das Internet anboten. Die Tools, mit denen Register erstellt werden, reichten von simpler HTML-Codierung bis zu HTML Indexer und anderer spezieller Software. Neue Indexe und Software-Produkte entstanden, aber viele Website-Register sind auch wieder verschwunden. Es werden die Gründe für die Erstellung von Website-Indexen dargestellt und Beispiele von Website-Indexen und anderen Zugangsoptionen erläutert, die in den letzten 15 Jahren entstanden, die aber nicht mehr zur Verfügung stehen. Es werden einige Vermutungen über die Gründe für diese Veränderungen angestellt.

Website indexing concepts developed as indexers, librarians and web managers experimented with different approaches for making the information they were providing on the Internet more accessible. These approaches included: A to Z indexes; displaying the overall structure of the site (information architecture); site maps; and search facilities. Search facilities were sometimes enhanced by the creation of subject metadata ("catalogue cards"), which could be organised in different facets, or displayed visually as well as textually.

The tools for creating A to Z indexes have changed over time. Initially, indexers used simple HTML coding to create indexes. Features such as indents and turnaround (wraparound) lines caused difficulties. The development of HTML Indexer was a major breakthrough as it provided a simple, effective way to create indexes.

Alongside the creation of new indexes and new software products, many website indexes have been lost. The reasons for this include changes to the whole company (eg, takeovers) and changes in search policy (eg, provision of a search engine instead of an index). The need for constant maintenance of website indexes when content changes has been a major challenge.

This article considers the reasons for creating website indexes, but does not explain in detail how this should be done – this is covered in our book Website indexing (Browne and Jermey, 2004) and in the article by Heather Hedden on page 433 in this issue. The article includes examples of website indexes and other access tools which were cited in the first or second editions of the book (www.webindexing.biz/Webbook2Ed/linkspage.htm), but which are no longer on the web at the same URL, and provides an assessment of some of the reasons for the changes, and commentary on the trends we have observed over the last decade.

Why create an index?

Search engines such as Google provide quick access to much useful information, leading some people to wonder whether any other pathways are needed. People who create website indexes are aware that there are different types of information requests, and different types of information users, and that people are best served by having a range of options available to them. See articles by Brown, Leise (2002) and Bates (1989, 1998) on user needs.

Comparing A to Z indexes with search engine results

An interesting experiment involves comparing a web-based index with the results that a Google search on the same material provides. You can do this by using the 'site:' search facility in Google, through which you can limit a Google search to a single site or part of a site. For example, typing 'site:library.lanl.gov/libinfo/news GeoRef' into the Google search box will find all occurrences of the name 'GeoRef' within the site library.lanl.gov/libinfo/news. This site contains the Los Alamos National Laboratory Research Library Newsletter, which has an index, so we can compare results of an index search with results of a Google search.

If you look up 'GeoRef' in the A to Z index (library.lanl.gov/libinfo/news/newsindx.htm), you find a crafted index heading with six subheadings, clearly showing the different information covered in each section, and the date of publication:

GeoRef
- added to FlashPoint (7/04)
- available through CSA interface (9/02)
- CD version available (6/95)
- funding opportunities section (9/98)
- web version available (1/98)
- web version trial access (7/97)

When you do a Google search, on the other hand, you find a large number of hits, each with a title and brief summary. It is more time consuming to look through these results, but they do provide additional hits that may be of interest. Thus the two information access methods provide different results, each of which might be useful to a specific user at a specific time.

You can do another comparison by searching for 'RDF' (or any other topic of interest) in the W3C site index at www.w3.org/Consortium/siteindex#R, and then doing a Google search of 'site:www.w3.org RDF' (thousands of hits to browse through!).

Usability

Website indexes fulfil many of the usability heuristics (general design principles) mentioned by Jakob Nielsen (2005). For example, they offer:

- Recognition rather than recall (they are browsable, and don't require the user to consider spelling and word formatting such as hyphenation)
- Match between system and the real world (they are a familiar tool for book users)
- User control and freedom (they are easy, quick and direct, and users can move through the index as they wish)
- Help users recognise, diagnose and recover from errors (indexes provide cross-references to guide users from their entry terms to other terms, and glosses (parenthetical qualifiers) to explain the meaning or use of a term)

Provide help and documentation (indexes often provide introductory notes which contain information about the index and its role within the site).

Changes in website indexing

There have been many changes to website indexing in general, and to individual website indexes, over the last decade. This section examines some of the indexes that changed since the publication of the second edition of Website indexing (Browne and Jermey 2004).

Index removal due to company changes

Website indexes suffer when companies merge or cease certain activities. The Case-in-Point index (previously at www.acxiom. com/caseinpoint/cip-ix-home.htm), which was discussed in Beyond book indexing (1999), and was a place-winner in the AusSI (now ANZSI) Web Indexing Award, no longer exists. This is because the website, the journal, and the index have all been closed down.

The PeopleSoft Products (www.people soft.com/corp/en/indi ces/prod_index.jsp) index appears to have been lost when PeopleSoft merged with Oracle on June 1, 2005. All PeopleSoft.com content is now on Oracle.com.

Index removal due to policy changes

The website index to the University of Texas Policies and Procedures Manual was a trailblazer (Fetters 1998). By the time we wrote the second edition of our book, however, it had been replaced with a search engine enhanced by metadata.

The Murdoch University Handbook for 2005 had an online index (www.comm. murdoch.edu.au/handbook/search/hand book_in dex.html), but I can no longer find an index on the site. Similarly, the A to Z indexes in French and English at the Statistics Canada site seem to have been replaced with a site map, and simple and advanced search.

The Yale Undergraduate Regulations Index used to have an index created using HTML/Prep. When you hovered over an entry it showed 'tips' that provided your context within the index (see an archived copy at http://web. archive.org/web/ 20041205153126/http:// www.yale.edu/ycpo/undregs/pages/ in dexpage.html). There appears to be no index to the 2006–2007 Undergraduate regulations at http://www.yale.edu/yale college/publications/uregs/index.html.

Index changes due to website changes

The index to the ANZSI website (www.aussi.org, soon to be changed to www.anzsi.org) has not been updated during the website redesign process, although it may be re-introduced at a later date. Changes to the website creation software including use of a database rather than individual page creation make it more difficult (but not impossible) to use HTML Indexer to create the index.

The British Society of Indexers (SI) redesigned their website. The index is

Neue Erlösmodelle für Zeitungsverlage.

Sandra Huber. – Boizenburg: vwh, 2007. 173 Seiten. (Medienwirtschaft). 27,90 Euro, ISBN 978-3-9802643-9-6.

Wie schon der Titel des Buches suggeriert, geht es hier um das Eingemachte. Wie können wir mit Medien Geld verdienen. Bei der allgemeinen Medien-Euphorie scheint dies eher trivial zu klingen. Denken wir aber an den harten Wettbewerb, so die vielen Flops, dann kommt uns Sandra Hubers „Neue Erlösmodelle für Zeitungsverlage" wie gerufen. In diesem Buch steht das Medium Zeitung zwar im Mittelpunkt, aber auch Bibliotheken, Hosts und andere Informationseinrichtungen werden von diesem Buch profitieren, das eigentlich dem Thema Diversifizierungsstrategie einen wichtigen Platz einräumt. Vor diesem Hintergrund lassen sich Hubers Vorschläge – mit den entsprechenden Anpassungen – ebenso gut auch für den BID-Bereich anwenden. Was ist nun das Besondere an diesem Buch?

Ausgangspunkt ist die These, dass die Zeitungswirtschaft einen tiefgreifenden Wandel durchläuft. Die Gründe dafür sind offensichtlich: Verändertes Leseverhalten und multimediale Alternativen. Dadurch bedingt, gerät das klassische Anzeigen- und Vertriebsgeschäft ins Wanken. Sandra Hubers These lautet deshalb: „Starke Zeitungsmarken hüten nicht nur die Tradition, sondern beschreiten auch neue Wege", wie es im Klappentext heißt.

Und hier liegt die Stärke des Buches: Ausgehend von einer Analyse der aktuellen Situation im Zeitungsmarkt (Kapitel 1) und der Beschreibung von „Zeitungen als Wirtschaftsgüter" (Kapitel 2), beschreibt die Autorin in drei weiteren Kapiteln, wie sich der Zeitungsmarkt den neuen Herausforderungen stellt beziehungsweise stellen sollte.

Das dritte Kapitel stellt „Neue Erlösmodelle für Zeitungen" vor. In den fünf Unterkapiteln geht es

- um „Erlösmodelle für Zeitungen im Internet",
- um die „Kernkompetenz: Content", wo die Autorin thematisiert: Synergieeffekte zwischen Print und Internet, Erlösquellen (Online-Archiv, E-Paper, Weblogs), Synergien von Print und Mobile (z.B. iPod), verlagsnahe Zusatzprodukte (z.B. CD-ROM, DVD), verlagsferne Produkte (Markendehnung), Crossmedia und schließlich weitere Erlösquellen wie Zeitung als Veranstalter, Vorteilsclubs, Webauktionen.

Im vierten Kapitel „Analyse der neuen Erlösquellen" unterzieht die Autorin die im dritten Kapitel vorgestellten Potenziale einer SWOT-Analyse und einer Bewertung.

„Die Zukunft der Zeitung" bildet den Abschluss dieser hochinteressanten Arbeit. Es folgen noch ein Fazit, Literaturverzeichnis und ein Sachregister.

Sandra Hubers Arbeit zeichnet sich durch eine sehr hohe Praxisnähe aus, durch anschauliche Beispiele und einen sehr lesefreundlichen Stil. Dieses ausgezeichnete Buch aus dem neu gegründeten Fachverlag für Medientechnik und -wirtschaft des Ingenieurs für Druckereitechnik und Informationswissenschaftlers Werner Hülsbusch (vwh) sei allen im BID-Bereich Tätigen sehr empfohlen.

Wolfgang Ratzek

now available directly from the home page through provision of an alpha bar (links for all letters of the alphabet to the appropriate part of the index) as well as on the index page at www.indexers.org.uk/index.php?id=217. With a new indexer, a new software package (XRefHT32) has been used.

Long-lived indexes

There are also a number of long-lived indexes on the web. One of the keys to longevity seems to be to have a website indexer who is closely connected to the management of the website; another is to have a site or subsite that changes little or not at all.

The index to the Los Alamos National Laboratory Research Library Newsletter was an early example of the use of HTML Indexer (lib-www.lanl.gov/libinfo/news/newsindx.htm). It was mentioned in the first edition of our book in 2001 and was still going strong in 2005. The index is still online, but the monthly newsletter has been replaced with a blog, so the index is complete and is not being updated.

The index to the book Orders of Magnitude (www.informationuniverse.com/ordersmag/orders.htm) has been online for over a decade. As a book index doesn't need maintenance, it is relatively easy to keep such an index available.

Penrith City has for many years provided indexes to its council and library services. These have recently changed URLs, but are still available online:

- A–Z of Council Services at www.penrithcity.nsw.gov.au/index.asp?id=960
- A–Z of Library Services at www.penrithcity.nsw.gov.au/index.asp?id=563
- Quick Index, which covers the whole site, at www.penrithcity.nsw.gov.au/index.asp?id=634.

Montague Institute Review dynamically creates an A to Z index from database-stored metadata at www.montaguelab.com/Public/indexes.htm (go to http://www.montague.com then select 'Index'). They say:

'…Even on a relatively modest Web site, standard keyword searching can return too many matches to be meaningful. Furthermore, visitors need cross references and definitions to understand specialized terms, such as 'upstream knowledge management' or 'authority files'. Finally, maintaining indexes and tables of contents for a monthly publication is too time-consuming and error-prone without a standard list of terms in database format.'

Alternatives to website indexing

There have been interesting experiments with access methods apart from A to Z indexes.

Visualisation

Most of the information visualisation examples that were mentioned in Website indexing no longer appear on the web. For example, Inxight now uses text rather than a star tree to display their site map (www.inxight.com/map) and the Antarctica examples are no longer there, although the company itself is (www.antarctica.net). The PubMed example has gone, but you can read an article about it at www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=12556244. Web-wide visualisation tools such as kartoo.com remain.

Classification

In Website indexing (Browne and Jermey 2004) we wrote that few of the projects listed by OCLC as being classification schemes for web resources (orc.rsch.oclc.org: 6109/classification) were still current. That website is no longer available. The topic of the use of classification schemes on the web is addressed by Vizine-Goetz (2002).

The main survivor in this area is BUBL LINK (Bulletin Board for Libraries, www.bubl.ac.uk). It groups selected sites according to the Dewey Decimal Classification. Users start with the ten main classes, then browse step-by-step through the classification. The site also provides an A to Z index and search box on home page. Classification schemes do not seem likely to have wide potential on the web because they are time-consuming to implement, and hard to keep up-to-date.

Faceted classifications on the web

Faceted classifications provide effective online searching because they allow users to select the facets (general topic areas) in which they want to limit their search. For example, on a site about wine you can choose to search first by type of wine, or country of origin, or price, depending on your needs at the time (Browne 2003).

Most of the examples of faceted classification from our book have gone, but most of the major programs remain. Those that have disappeared include:

- Annotated Wordnet (www.siderean.com/wordnet17.jsp)
- CMS Faceted Product Directory (www.cmsreview.com/timelines/ShopFeatureDirectory.html)
- Meta Matters (dcanzorg.ozstaging.com/mb.aspx)
- Online proceedings of the DC- 2002 Dublin Core conference (www.siderean.com/dc2002.jsp).

Those that remain include:

- Epicurious (www.epicurious.com/recipes/find/advanced)
- Tower records (www.towerrecords.com). Select Advanced on the home page next to search box. Then the left hand column also has 'Browse by', at which you can search according to specific facets and their values (eg, Price: 'under $10').

Faceted search at Langemarks Cafe (www.langemark.com/taxonomy_search/blog) has been replaced by a cloud map at www.langemark.com/tagadelic/chunk/1?PHPSESSID=7078c8337f0bba91181735d0c8a14fca.

A new implementation of faceted classification is seen in the NCSU libraries catalogue. You enter terms in the Catalog search box at www.lib.ncsu.edu/search collection, and can gradually narrow your search, while seeing the numbers of hits retrieved.

Topic maps

Topic maps (Browne 2002a) do not seem to have had a major impact on the web, although they are used in some company intranets. The Diffuse site (www.diffuse.org) provided an important example of the way topic maps could be implemented on the web, but it is no longer available on the web.

Subject gateways

There are a number of subject gateways on the web, which lead to quality, selected information resources on a specific topic or, occasionally, on all topics. These depend on human activity and can only deal with a small proportion of the available information.

In October 2003 Infomine stopped taking suggestions for sites to include as they had been plagued by bulk inappropriate commercial submissions. Site submission is now active again at infomine.ucr.edu/contact/suggest.shtml.

Zeal used to take suggestions of non-commercial sites at no charge for inclusion in their directory (www.zeal.com/users/non_profit.jhtml?rpc=49). Its owner LookSmart is now focussing on other objectives and Zeal.com has been shut down. They recommend the use of the online bookmarking service Furl (www.Furl.net).

The AVEL subject gateway (Australasian engineering & IT resources, avel.edu.au/docs.html) operated from 1999–2005. The site has now been decommissioned, and the history of the project can be found at avel.library.uq.edu.au.

New methods of indexing

As well as needing individually-created indexes, the web is being gradually organised through the process of social bookmarking, in which people tag sites of interest to them using keywords, and then share these keywords with other users.

One popular site is Delicious (del.icio.us) – here you can look at your own tags and those of other people as lists or clouds (with topics that have more content presented in a larger font). You can follow links by topic and tag creator, and can add people to your network, thus easily sharing website discoveries. The lists of tags are somewhat like an index, but lack features such as cross-references and subheadings, and often lack consistency. For example, at del.icio.us/url/e6f93dcf17dfd5f1a25e0b8770203ff2 you can see the different ways in which people have tagged the Website Indexing SIG site, depending on their points of view.

A similar approach is taken at sites such as Citeulike (www.citeulike.org/tag/indexing) and Technorati (www.technorati.com/tag).

Software changes

A number of software packages have been written to aid in the construction of website indexes. These are all tools for human construction of indexes, and do not aim to create fully automatic indexes, although some create a default index that can be used as a starting point. The significant change in this area has been the creation of two new programs.

For information on HTML/Prep, which can be used to transform print indexes for use on the web, as well as to create website indexes from the start, see Browne (2002b) and www.levtechinc. com. For information on HTML Indexer, which embeds metadata in webpages to make easily updatable website indexes, see www.html-indexer.com, Browne (1999) and Unwalla (2006). HTML Indexer has a free demo version which makes functional indexes, except that the projects can't be saved for later editing (although the indexes themselves can be edited).

XRefHT32 ('shreft') is a free, open source program that can be used to create website indexes. It can be incorporated with a thesaurus created using TheW32, a free thesaurus creation program by Timothy Craven (publish.uwo.ca/~craven/freeware.htm). You can see examples on the web at www.ulb.ac.be/ecoles/eiulb/liens/liensM.htm and publish.uwo.ca/~craven/craven.htm. For more information see Hedden (2005) and Lamb (2006).

Basedex is the latest program, and has just been mentioned on the Webindexing mailing list in March 2007. It can be used to create large-scale indexes through the merging of separate indexes. See the discussion at the Web Indexing Yahoo mailing list – tech.groups.yahoo.com/group/web-indexing – and the site at blazingrails.basedex.com.

Conclusion

Website indexing has been constantly developing since the early days of the web. There have been a range of approaches, supported by software products to make the job easier, although still requiring human input for the best quality results. Although search engines provide one good approach, website indexes can provide complementary approaches to access to important information.

For more information see the ASI Web Indexing SIG (www.webindexing.org/web-index-examples.htm) and 'Indexing resources on the WWW' by Stephenson (2005).

References

Bates, Marcia J. 1989: The design of browsing and berrypicking techniques for the online search interface. www.gseis.ucla.edu/faculty/bates/berrypicking.html.

Bates, Marcia J. 1998: Indexing and access for digital libraries and the Internet: human, database and domain factors. In: Journal of the American Society for Information Science 49(1998), pp. 1185–1205, www.gseis.ucla.edu/faculty/bates/articles/indexdlib.html.

Brown, David M.: Why create an index. www.web-indexing.org/article-brown.htm.

Browne, Glenda 1999: Web site indexing. www.webindexing.biz/Articles/WebSiteIndexing.htm. First published in Online Currents 14(1999)10.

Browne, Glenda 2002a, Topic maps. www.webindexing.biz/articles/TopicMaps.htm. First published in Online Currents 17(2002)3.

Browne, Glenda 2002b, HTML/Prep: transforming indexes for the web, www.webindexing.biz/articles/HTMLPrep.htm. First published in Online Currents 17(2002)7.

Browne, Glenda 2003: Faceted classification. www.webindexing.biz/articles/FacetedClassification.htm. First published in Online Currents 18(2003)9.

Browne, Glenda; *Jermey*, Jonathan 2004. Website indexing: enhancing access to information within websites. 2nd edn. Adelaide: Auslib Press 2004. Now for sale as an e-book through RMIT Informit (www.informit.com.au), as a PDF at www.webindexing.biz and as a PDF or print-on-demand book from Lulu (www.lulu.com).

Fetters, Linda 1998: A book-style index for the Web: the University of Texas Policies and Procedures Web site. In: The Indexer, 21(1998)2.

Hedden, Heather 2005: HTML indexing freeware: XRefHT32. www.hedden-information.com/XrefHT32_Review_KW2005_10-12.pdf. First published in Key Words, 13(2005)4.

Hedden, Heather 2007: Indexing specialties: web sites Medford, NJ, Information Today, for the American Society of Indexers.

Kamel Boulos, Maged N 2003: The use of interactive graphical maps for browsing medical/health Internet information resources. In: International journal of health geographics 2(2003)1, doi: 10.1186/1476-072X-2-1.

Lamb, James 2006. Website Indexes: visitors to content in two clicks, or website indexing with XRefHT32 freeware (available as a print-on-demand book from www.lulu.com).

Leise, Fred 2002. Improving Usability with a Website Index. www.web-indexing.org/article-leise.htm. Originally published July 15, 2002 on Boxes and arrows (www.boxesandarrows.com).

Nielsen, Jakob 2005: Ten usability heuristics. www.useit.com/papers/heuristic/heuristic_list.html.

Rowland, Marilyn; *Brenner*, Diane 1999: Beyond Book Indexing: How to Get Started in Web Indexing, Embedded Indexing, and Other Computer-Based Media. Medford, NJ, Information Today, for the American Society of Indexers.

Stephenson, Mary Sue 2005: Indexing resources on the WWW. www.slais.ubc.ca/resources/indexing/database1.htm.

Unwalla, Michael 2006: Web indexing: extending the functionality of HTML Indexer. www.techscribe.co.uk/ta/web-indexing.htm. First published in The Indexer 25(2006)2.

Vizine-Goetz, Diane 2002: Classification schemes for internet resources revisited. staff.oclc.org/~vizine/JIC/v5n42002/ClassificationSchemes Revisited.doc. Manuscript of paper published in Journal of Internet Cataloging 5(2002)4.

Website, Subject Indexing, Trend

Elektronischer Dienst, Website, Sachkatalogisierung, Index, Entwicklungstendenz

The Author

Glenda Browne, BSc, MSc, DipIM-Lib

has been an indexer of books, journals, databases, online help and websites since 1988. With her partner, Jon Jermey, she has written two books – Website indexing (Auslib Press, 2nd ed, 2004, www.webindexing.biz) and The Indexing Companion (Cambridge University Press, 2007). Glenda is the editor of the Around the World section of The Indexer journal, which involves gathering information from all of the indexing societies. She has taught indexing to indexers and editors, through ANZSI, the NSW Society of Editors, Macquarie University and the University of NSW. She has been involved on ANSZI state and national committees, including roles as newsletter editor, Sydney conference administrator and NSW Branch President (current).

PO Box 307
Blaxland NSW 2774, Australia
webindexing@optusnet.com.aus
www.webindexing.biz

Evolution towards ISO 25964: an international standard with guidelines for thesauri and other types of controlled vocabulary

Stella G. Dextre Clarke, Oxfordshire (UK)

The history of the development of ISO 2788: Documentation – Guidelines for the establishment and development of monolingual thesauri and ISO 5964: Documentation – Guidelines for the establishment and development of multilingual thesauri *is briefly described. In 2001 work began on development of BS 8723:* Structured Vocabularies for Information Retrieval – Guide, *a five-part standard designed to update the international standards, with particular emphasis on enabling interoperability. Procedures have recently begun to adopt BS 8723 as an international standard.*

Der Weg zur ISO 25964: eine internationale Norm mit Richtlinien für Thesauri und andere kontrollierte Vokabulare

Die geschichtliche Entwicklung von ISO 2788: Documentation – Guidelines for the establishment and development of monolingual thesauri *und ISO 5964:* Documentation – Guidelines for the establishment and development of multilingual thesauri *wird kurz beschrieben. 2001 begann die Arbeit an der Entwicklung von BS 8723: Structured Vocabularies for Information Retrieval – Guide, einer fünfteiligen Norm, die konzipiert wurde, um die internationalen Normen zu aktualisieren, insbesondere mit Hinblick auf Interoperabilität. Der Verfahrensablauf zur Übernahme von BS 8723 als internationale Norm hat begonnen.*

1 Introduction

A controlled vocabulary is essential for the efficient construction of many indexes, whether printed or electronic. For building electronic indexes especially, a vocabulary such as a thesaurus needs to be integrated into the desktop indexing environment. This is most easily achieved if the vocabulary conforms to widely accepted standards. But the long established international standards for thesauri have been falling behind needs in today's networked environment. This article will track the history of these standards, and describe current efforts to bring them up to date.

2 Evolution of the existing standards ISO 2788 and ISO 5964

Roberts (1984) traces the pre-history of the information retrieval thesaurus back to C N Mooers in 1947, closely but independently followed by C L Bernier and E J Crane. Both he and Krooks and Lancaster (1993) recognise the thesaurus developed in 1959 for the E I Du Pont Nemours and Co., Inc., as the first practical manifestation. Intensive innovation over the next few years led to joint publication by the Engineers Joint Council and the US Department of Defense in 1967 of the influential *Thesaurus of Engineering and Scientific Terms (TEST).* According to Aitchison and Dextre Clarke (2004) TEST embodied most of the standard features visible in alphabetically organised thesauri for the next thirty years. The first edition of ISO 2788 *Documentation – Guidelines for the establishment and development of monolingual thesauri* followed in 1974, largely built on the same rules and conventions set out in Appendix 1 of TEST.

ISO 2788 was last updated in 1986. It was designed mainly for postcoordinate information systems, while acknowledging that precoordinate indexing and retrieval applications might find it useful too. Understandably, its context was the era before the advent of personal computers, when the thesaurus was a massive printed tome, or set of volumes. In those times, a set of optical coincidence cards for in-house use was considered a sophisticated information retrieval system. Only the publishing giants offered online bibliographic databases, and an electronic online thesaurus was a rarity, cumbersome to use. Usually terms had to be selected manually before keying into any electronic systems that were available. The principles of ISO 2788 have always been highly applicable to electronic retrieval systems, but the expression of those principles was tied to the paper-based era.

ISO 5964 *Documentation – Guidelines for the establishment and development of multilingual thesauri* was published in 1985, and has not subsequently been updated. It relies on the same principles as ISO 2788, applying them to a multilingual situation. It suffers from the same paper-based orientation.

Both ISO 2788 and ISO 5964 have been adopted as national standards in many countries, for example France, Germany, Spain and the United Kingdom. In the UK they are known respectively as BS 5723 and BS 6723.

3 Needs at the end of the 20th century

Today few users have patience with the idea of looking up a printed thesaurus before they search for information. Even in the 1980s, the only users who could be relied upon to understand the relevance of a thesaurus were trained information professionals. Today, information aids need to work intuitively behind the scenes (or rather, behind the computer screens) if they are to gain widespread acceptance. There is still a role for trained professionals to consult a thesaurus (preferably a well-designed electronic one) in the process of indexing, now often called meta-tagging. But only a minority

of end-users have the patience to look up a thesaurus, in the event that this is available through an advanced interface. Plainly the presentation and sometimes the mode of application of the thesaurus need to change.

As we move into the 21st century, networking has placed previously unimaginable assemblies of resources at our fingertips. The end-user is not satisfied with convenient, effective access to one database. Rather he or she wants to access a variety of different resources at one pass through a common interface, although each of those resources may have been indexed or classified with a different vocabulary. "Interoperability" is the great goal.

In academia, for example, the catalogues of multiple universities constitute a marvellous shared resource, but some have been classified with the *Dewey Decimal Classification*, others with the *Universal Decimal Classification*, some use the *Library of Congress Subject Headings,* or perhaps the *Medical Subject Headings* of the U.S. National Library of Medicine, or the *INSPEC Thesaurus*, etc. Just as the speakers of Spanish and Greek may have difficulty understanding each other, so the incompatibilities between indexing languages impede cross-database searches. The demand, therefore, is for mapping between vocabularies, including vocabularies of very different types and structures. This will enable a search to be automatically converted (in theory, at any rate) from one vocabulary to another. The same problem and solution apply, even more so, when the indexing vocabularies use different natural languages.

A different sort of interoperability is needed at the data exchange level. A thesaurus or classification scheme is typically built and maintained with purpose-built software. Periodically, vocabulary data need to be exported from this and imported into downstream applications for indexing and/or searching, such as content management systems, electronic publishing systems, search engines, etc. A common data format is needed for the import and export process. Further demands are made if the vocabulary database is used for live querying – perhaps for electronic thesaurus look-up or for conversion of search statements from one vocabulary to another, in the course of a search of information resources. In this case an agreed protocol is required. The absence of any format or protocol recommendations in ISO 2788 and ISO 5964 has long impeded data exchange, in fact it has probably inhibited widespread take-up of thesauri.

4 Initiatives in the UK and USA

At around the start of the new century, both ISO 2788 and ISO 5964 came up for their regular review, by the committees of each of the participating standards bodies. The BSI committee IDT/2/2 felt that revision was essential, both to update the standards generally and to address the interoperability needs. After the long lapse in activity on updating the standards, however, it was hard to establish enough contacts to assemble a viable international working group under the aegis of ISO. The committee felt it more practicable to proceed in two stages: first, work on the corresponding British standards; then, on approval, offer these up for adoption by the international community. The fact that the British and International standards were identical made this strategy easy to implement. A drafting committee, known as IDT/2/2/1, was established in 2001.

Professional colleagues in the USA faced a similar scenario. Their corresponding national standard, known as Z39.19, was first published in 1974, and had evolved in parallel with ISO 2788. At the end of the 20th century, the current edition was ANSI/NISO Z39.19 *Guidelines for the construction, format, and management of monolingual thesauri*, dated 1993. A workshop was organized by NISO (National Information Standards Organization) in November 1999, to investigate "the desirability and feasibility of developing a standard for electronic thesauri". Among its conclusions were that a new standard "should provide for a broader group of controlled vocabularies than those that fit the standard definition of 'thesaurus'", and that, "the primary concern is with shareability (interoperability), rather than with construction or display." (See full report at www.niso.org/news/events_workshops/thes99rprt.html). As an outcome of this meeting, an advisory group was assembled to oversee revision of Z39.19.

The committees in UK and USA were very much aware of each other's efforts, and have maintained communication over the years of hard work that followed. It is no coincidence that the goal of interoperability has pervaded thinking on all sides.

5 Outline of BS 8723

5.1 Overview

From the start, BSI's IDT/2/2/1 took a decision to bring the existing standards for monolingual and multilingual thesauri together into one framework. Thus BS 5723 and BS 6723 would be united in a new BS 8723 entitled **Structured vocabularies for information retrieval – Guide**. (Similarly, it is hoped that ISO 2788 and ISO 5964 will eventually be united in the new ISO 25964. This will avoid some of the duplication and repetition found between the existing standards.) At the same time the scope of the combined standard should be radically extended to meet the interoperability needs described above. BS 8723 should therefore be presented as a five-part standard. The content and status of each of the parts will now be described.

5.2 BS 8723 Part 1: Definitions, symbols and abbreviations

The definitions and conventions set out in this short document apply to all the parts of BS 8723. Part 1 was published in November 2005, at the same time as Part 2.

5.3 BS 8723 Part 2: Thesauri

This document covers and updates the entire scope of BS 5723 (ISO 2788), as well as adding much new material. It applies to both monolingual and multilingual thesauri, although it does not cover the special requirements of multilingual situations. While the principles enshrined in ISO 2788 are unchanged, they have been restated in today's parlance, including many new examples. The most noticeable changes include:

- clearer guidance on applying facet analysis to thesauri
- some changes to the 'rules' for compound terms
- more guidance on managing thesaurus development and maintenance
- functional specification for software to manage thesauri
- a general expectation that people will use the thesaurus electronically, as well as in print.

5.4 BS 8723 Part 3: Vocabularies other than thesauri

This part of the standard breaks entirely new ground. There were two main objectives:

- to provide good practice guidance where needed for emerging types of vocabulary;
- to facilitate interoperability between different vocabulary types, by describing the similarities and differences between them.

For the first objective, users need *normative* advice. For the second, *informative* description is more helpful. Accommodating mixed objectives was not always straightforward, as may be seen from the following background to the decision.

In the experience of the members of the drafting committee, confusion about the

different types of controlled vocabulary has been widespread. Not only are the names 'ontology' or 'taxonomy' commonly misused to brand any sort of vocabulary as exciting and futuristic, but also the structural differences are misunderstood. For example, the captions of a classification scheme are sometimes treated as though they were preferred terms in a thesaurus, ignoring the role of notation. And the pre-coordination found in subject heading schemes and some taxonomies is an endless source of confusion. One could argue that well-known schemes such as the *Dewey Decimal Classification Scheme (DDC)*, the *Universal Decimal Scheme (UDC)* or the *Library of Congress Subject Headings (LCSH)* are already standards in their own right, needing no higher authority. But the committee took the view that it would be helpful to compare and contrast the essential elements of each of these vocabulary types. The aim was not to send the publishers of the UDC or the LCSH back to the drawing board, but to provide enough description of these types of tool so that an Information Technology developer would be able to implement them interoperably with other vocabularies.

As to the term 'taxonomy', this has been borrowed from its roots in the classification of plants and animals, to apply to all manner of things. For some users it is a switching tool for the knowledge systems in their organization; for others it is a directory on the web, still others use it as a more fashionable name for their existing thesaurus, etc. The drafting committee took the view that there is a need for good practice guidance in developing vocabulary tools to support browsing of websites and portals.

In the case of ontologies, however, it was felt that the details of a normative standard are best left to the artificial intelligence community, supported by a growing band of Semantic Web enthusiasts. The advice provided by BS8723-3 should be kept to the minimum required in the context of interoperability with other vocabularies.

After much debate of the above issues and consideration of a broad range of vocabulary types, six main clauses were prepared, oriented as shown in Table 1. Almost all the chapters have some normative advice, but in no case is the treatment as deep or exhaustive as are the guidelines for thesauri in Part 2.

Part 3 was issued as a "Draft for Public Comment" in the spring of 2007. Publication of the approved standard is expected in November 2007.

Table 1. Clauses of BS 8723-3

Vocabulary type	General orientation
Classification schemes	mostly informative
Business classification schemes for records management	mostly informative
Taxonomies	mostly normative
Subject heading schemes	mostly informative
Ontologies	informative
Authority lists	mostly informative

5.5 BS 8723 Part 4: Interoperability between vocabularies

Part 4 builds on the groundwork laid in Part 3. It addresses mapping between vocabularies, where a mapping is defined as a "statement of the relationships between the terms, notations or concepts of one vocabulary and those of another". It treats multilingual thesauri as a special case of mapping between vocabularies. Thus it takes in all of the scope of BS 6723 (ISO 5964), while adding much more.

Part 4 starts at quite a basic level, with three models for interoperability:

- structural unity
- non-equivalent pairs
- backbone model

The first is used in a multilingual thesaurus. It requires the same structure of concepts, hierarchical and associative relationships in each language version of the thesaurus. The second and third are illustrated in Figures 1 and 2.

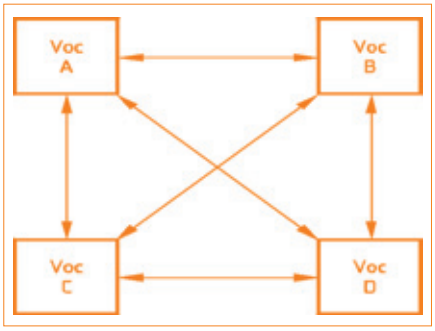


Figure 1: Non-equivalent pairs model, as applied to four vocabularies

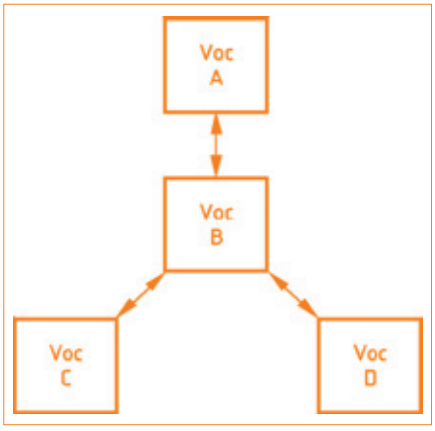


Figure 2: Backbone model, as applied to four vocabularies

Although these models are not new (see, for example, Horsnell 1975), this is the first time they and other basics of mapping have been brought together in a standard.

An important difference between the vocabulary types is that concepts are represented by different elements, for example by notations in a classification scheme and by preferred terms in a thesaurus. While it may seem obvious that mappings should be made between these elements, misunderstandings are frequently observed in practice. Hopefully the clarification in Part 3 together with the mapping advice in Part 4 will inspire better practice in future.

For mappings between structurally identical vocabularies, as in the case of a multilingual thesaurus, most of the advice previously in ISO 5964 has been retained.

Part 4 was issued as a "Draft for Public Comment" in the spring of 2007. Publication of the approved standard is expected in November 2007.

5.6 BS 8723 Part 5: Exchange formats and protocols for interoperability

Part 5 has brought new and difficult challenges. An early draft reviewed a number of existing formats and protocols (such as MARC21, Zthes, SKOS, the ADL protocol, etc.) and set out criteria for evaluating which would be the most suitable in various circumstances. In October 2006 a group of experts was invited to a workshop to consider this draft. They rejected much of it, and recommended instead the development of one standard format, based on XML (eXtensible Markup Language). It should concentrate on thesauri rather than the other types of vocabulary, they said, and should be based on a data model.

Ten months later, at the time of writing, the data model written in UML (Universal Modelling Language) and derived XML schema are almost complete. BS 8723-2 and BS 8723-4 provide not just for simple thesauri but also for sophisticated ones with a range of optional features. Modelling and seeking agreement on all these features for Part 5 has consumed a great deal of time and voluntary effort.

The results, including testing with a range of sample data, are visible on a development website at http://porism.tdmweb.co.uk/BS8723/Documentation/Home.html (likely to move to http://schemas.bs8723.org/ in future). When the standard is eventually published, it is hoped that this same website will support the namespace required for the XML schema.

While the volunteer working party has tested the model and schema to the extent possible with limited resources, it is now felt that extended testing among the potential user community would be advisable. Furthermore, consensus among the advisory group of experts has been hard to win. Some user feedback has suggested that even if the new format is issued as a standard, the pre-existing formats such as SKOS and Zthes will continue to be needed for specific applications. In the light of these uncertainties, it is now proposed to issue the results as a "Draft for Development" rather than a full British Standard. This fast-track route will put all the work into the public domain for extended evaluation and feedback, hopefully also in November 2007.

6 Preparations for ISO 25964

In early 2007, with the bulk of the BS 8723 development work well in hand, it was time to think about moving towards an international standard. The International Organization for Standards (ISO) has well established procedures for considering whether to adopt a national standard as a future ISO standard. In this case, the proposal to adopt BS 8723 was submitted to the committees of all the national standards bodies participating in ISO 2788 and ISO 5964. These are coordinated by the technical subcommittee ISO TC46/SC9. The proposal was accepted in August and at least nine countries have agreed to participate in Project ISO NP 25964: Canada, Finland, France, Germany, New Zealand, Sweden, Ukraine, United Kingdom and United States of America. Participation of the USA offers an exciting prospect for eventually bringing the comparable ANSI/NISO standard Z39.19 into the same fold as the ISO standard. Colleagues from China, Denmark and Spain have informally expressed interest, and the hope is to involve all who can spare the time.

ISO NP 25964 will be based on the published parts of BS 8723, which will hopefully include the Draft for Development for Part 5. This approach should facilitate the international testing work appropriate for the exchange format.

7 Conclusion

The work on BS 8723 has laid a strong foundation on which to base an international standard for the interoperable controlled vocabularies we need in the years ahead. All professional colleagues who have an interest in the outcome are advised to contact the ISO TC46/SC9 representatives in their national standards body. The international working group will welcome participants offering a positive contribution.

Acknowledgements

The work described in this article was undertaken by BSI committee IDT/2/2/1, whose members are Stella Dextre Clarke (Convenor), Leonard Will, Alan Gilchrist and Ron Davies. Acknowledgements are due also to Nicholas Cochard, who led development of the UML model and XML schema for Part 5 of the standard.

References

Thesaurus of Engineering and Scientific Terms (TEST). New York: Engineers Joint Council and US Department of Defense; 1967.

Aitchison, J.; *Dextre,* Clarke S.: The thesaurus: a historical viewpoint, with a look to the future. Cataloging & Classification Quarterly. 37(2004), pp. 5–21.

British Standards Institution; BS 5723:1987 British Standard Guide to Establishment and Development of Monolingual Thesauri. London: British Standards Institution; 1987.

British Standards Institution. BS 6723:1985 British Standard Guide to Establishment and Development of Multilingual Thesauri. London: British Standards Institution; 1985.

British Standards Institution; BS 8723-1:2005 Structured Vocabularies for Information Retrieval – Guide – Definitions, Symbols and Abbreviations. London: British Standards Institution; 2005.

British Standards Institution; BS 8723-2:2005 Structured Vocabularies for Information Retrieval – Guide – Thesauri. London: British Standards Institution; 2005.

Horsnell, V.: The Intermediate Lexicon: an aid to international co-operation. Aslib Proceedings. 27(1975), pp. 57–66.

International Organization for Standardization; ISO 2788-1986. Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri. 2nd ed. Geneva: International Organization for Standardization; 1986.

International Organization for Standardization. ISO 5964-1985. Documentation – Guidelines for the Establishment and Development of Multilingual Thesauri. Geneva: International Organization for Standardization; 1985.

Krooks, D. A.; *Lancaster*, F. W.: The evolution of guidelines for thesaurus construction. Libri. 43(1993), pp. 326–342.

National Information Standards Organization. ANSI/NISO Z39.19-1993. Guidelines for the Construction, Format and Management of Monolingual Thesauri. Bethesda, Maryland: NISO Press; 1993.

Roberts, N.: The pre-history of the information retrieval thesaurus. Journal of Documentation. 40(1984), pp. 271–285.

Standard, Standardisation; international, Thesaurus, Vocabulary

Norm, Normung; international, Thesaurus, Wortschatz

The Author

Stella Dextre Clarke, MSc, FCLIP

is an independent consultant specializing in the design and implementation of knowledge structures, including thesauri, classification schemes and taxonomies. She is the Convenor of BSI committee IDT/2/2/1 and Project Leader of ISO NP 25964. She is a Fellow of the UK Chartered Institute of Library and Information Professionals and 2006 winner of the Tony Kent Strix Award for outstanding achievement in information retrieval.

Information Consultant
Luke House, West Hendred,
Wantage, Oxon, OX12 8RR
United Kingdom
sdclarke@lukehouse.demon.co.uk

Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39.19 for the 21st century

Emily Gallup Fayen, Boston, MA (USA)

In 2003, NISO began work on revising Z39.19. The new standard extended the scope from documents to content objects and extended the coverage from thesauri to all types of controlled vocabularies. It also moved beyond a mostly paper environment to formats appropriate for electronic content. This paper describes that work and the resulting new standard.

Richtlinien für Aufbau, Format und Management von einsprachigen kontrollierten Vokabularen: Eine Revision der ANSI/NISO Z39.19 für das 21. Jahrhundert
Die National Information Standards Organization (NISO) begann im Jahre 2003 mit der Überarbeitung der Norm Z39.19. Die neue Norm erweiterte den Geltungsbereich von Dokumenten auf Inhaltsobjekte und die Abdeckung von Thesauri auf alle Arten kontrollierter Vokabulare. Sie bezieht sich zudem über das reine Papierumfeld hinaus auf geeignete Formate für elektronische Inhalte. Dieser Artikel beschreibt diese Arbeit und die daraus resultierende neue Norm.

Introduction

Thirty years after the introduction of ANSI/NISO Z39.19, *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*, it was still the Standard most frequently requested for download from the NISO site. The strong interest in this standard provided evidence of the importance it held in the information community. Z39.19 was the primary source for guidance in the construction, format, and maintenance of this special type of controlled vocabulary.

In the intervening years, however, vast changes had occurred in the information industry and in the way people expect to store and retrieve information. These have resulted from very rapid changes in computer and information processing technology and the global rise of the Internet. Today, the expanding use of information databases in all aspects of business and commerce, government, and education, and the need to discover millions of sites on the Internet, means that there are thousands of applications in which controlled vocabularies of various types provide better ways to manage large amounts of content while at the same time making it easier for users to find the information they need.

Further, in the mid-1970s when Z39.19 was introduced, work with thesauri and controlled vocabularies was limited to use by highly trained professionals. With widespread use of powerful search engines available to anyone, it became clear that Z39.19 needed to be revamped, both to make its guidelines accessible to anyone and also to make it adaptable to contemporary display technology.

Background

The first edition of this standard, published in 1974, was prepared by Subcommittee 25 on Thesaurus Rules and Conventions of American National Standards Committee Z39 on Standardization in the Field of Library Work, Documentation, and Related Publishing Practices (later known as Library and Information Sciences and Related Publishing Practices). The subcommittee drew heavily on standards of practice developed by the Engineers Joint Council, the Committee on Scientific and Technical Information of the Federal Council for Science and Technology, and UNESCO.

At the time Z39.19 was first proposed, terminology was needed both in the indexing and searching various collections of documents. This terminology was generally selected from a thesaurus listing appropriate terms. The collections of documents being indexed might be printed resources such as journal articles, technical reports, newspaper articles, etc. As new information storage and retrieval systems have emerged, the concept of *document* has been extended to include materials such as patents, chemical structures, maps, music, videos, museum artifacts, and many other types of materials that are not traditional documents. Furthermore, the display methods described in the Standard were almost entirely for various sorts of printed products. In today's online world, other methods of organization and display had to be taken into account.

In 1998, the standard was reviewed and reaffirmed. At this time, however, the review confirmed a need for a fresh look at the standard, updating it for use in the rapidly evolving electronic information environment.

In response, NISO organized a national Workshop on Electronic Thesauri, held November 4–5, 1999, to investigate the desirability and feasibility of developing a standard for electronic thesauri. The workshop was co-sponsored with the American Psychological Association (APA), the American Society of Indexers (ASI), and the Association for Library Collections and Technical Services (ALCTS), a division of the American Library Association. The project to revise Z39.19 grew out of the recommendations developed by consensus at the Workshop.

The Workshop identified a number of limitations of the existing Standard:

- **Difficult for non-lexicographers to understand.** Many potential users

who expressed interest in the Standard had no background in library science or related fields and thus found the concepts difficult to apply to their particular applications even though many recognized the need to do so.

- **Focused on construction and maintenance.** The Standard assumed knowledge of the underlying principles of information science that promoted the use of controlled vocabularies.

- **Limited to document indexing applications.** Although the original context for controlled vocabularies was for indexing and retrieval of documents, in the intervening years it became highly desirable to apply the underlying discipline to many different types of materials including Web sites.

- **Limited to post-coordinate retrieval.** The Standard assumed that the controlled vocabularies within its scope were to be used in post-coordinate retrieval systems. This assumption limited its applicability to other types of retrieval, including browse and navigation systems.

- **Limited to printed products.** The display formats for the controlled vocabularies that were recommended included only printed presentations of the controlled vocabularies. Because of the date the work was first conceived (and even at the time of its last revision in 1993) virtually no controlled vocabularies were being used in a Web-enabled environment.

- **Outdated technology.** Finally, although the principles presented in the original standard were still relevant, many of the examples were based on outdated technology. These needed to be updated to make the Standard relevant to contemporary users.

Relying on feedback from the community and extensive internal discussions, NISO launched an initiative to revise Z39.19. The work was made possible by generous support from The H.W. Wilson Company, The Getty Foundation, and the National Library of Medicine. With an aim toward achieving the best possible result, and making sure the major stakeholders were involved, NISO assembled an advisory group to guide the work. The Thesaurus Advisory Group, or TAG as it was known, consisted of members from many segments of the information industry. The members were:

Vivian Bliss	Microsoft
Carol Brent	ProQuest
John Dickert	U.S. Department of Defense, DTIC
Lynn El-Hoshy	Library of Congress
Emily Gallup Fayen	Chair, MuseGlobal, Inc.
Patricia Harpring	Getty Foundation
Stephen Hearn	American Library Association
Marjorie Hlava	Access Innovations, Inc.
Sabine Kuhn	Chemical Abstracts Service
Pat Kuhr	The H.W. Wilson Company
Diane McKerlie	DMA Consulting
Peter Morville	Semantic Studios
Stuart Nelson	National Library of Medicine
Diane Vizine-Goetz	OCLC, Inc.
Marcia Lei Zeng	Special Libraries Association

Cynthia Hodgson (NISO) and Emily Fayen, MuseGlobal, Inc. and NISO SDC Liaison, prepared the revision.

NISO's Goal for the revised Z39.19 Standard

In February 2003 NISO conducted a survey to learn more about how Z39.19 was being used. The survey results showed that respondents wanted several things from the new Standard:

- The revised standard must provide a better, more inclusive way to represent content; that is, the standard should be applicable to a broader array of materials than documents.
- The revision must take into account a changing audience as well as a vastly different information environment.
- The revision must take into account the need for interoperability and sharing across applications.

About the Revised Standard

The first **four sections** of the revised Z39.19 serve as background and contextual information for the Standard which follows. These sections are:

- Introduction,
- Scope
- Referenced Standards, and
- Definitions, Abbreviations, and Acronyms.

Section 5, presents a comprehensive discussion on controlled vocabularies, their purpose, concepts, principles, and structure and begins the main body of the Standard.

Section 6 presents rules and guidelines for term choice, scope, and form.

Section 7 discusses the special problems presented by "compound terms", i.e. terms that include more than a single word.

Section 8 describes the various types of relationships that may be defined between and/or among terms in a controlled vocabulary.

Section 9 presents various factors that must be taken into consideration in displaying controlled vocabularies – in print products, online, and in conjunction with various Internet sites.

Section 10 discusses many of the issues arising from concerns about interoperability.

Finally, **Section 11** presents guidelines and principles involved in construction, testing, maintenance, and management systems for controlled vocabularies.

In addition, the revised Standard includes six Appendices to provide more information and to guide the reader to the parts of the Standard that may be most relevant to a specific area of interest. These are:

- Appendix A Summary of Standard Requirements / Recommendations
- Appendix B Comparison of Vocabulary Types
- Appendix C Characteristics and Uses of Controlled Vocabulary Display Options
- Appendix D Methods for Achieving Interoperability
- Appendix E Sample Candidate Term Forms
- Appendix F: References

Finally, there is a Bibliography, a Glossary and a full index to the Standard.

The Scope of the Revision

As a first step, the Advisory Group discussed several ways in which the scope of the standard would be broadened and changed to meet the changing needs of implementers.

- **Expand the scope beyond *thesaurus* to include *controlled vocabularies.*** This change in scope reflected the need to make the Standard applicable to controlled vocabularies other than those that had been so extensively used in the indexing of documents by the various abstracting and indexing (A&I) services.

- **Make the standard more accessible to users.** The original standard was developed by lexicographers for lexicographers. It assumed that readers were familiar with the underlying concepts and principles of vocabulary control. This is no longer true for the greatly expanded audience for the standard that is resulting in the frequent requests for download from the NISO web site.

- **Explain important concepts.** Because many potential users of this Standard did not have a background in library science or information science, it was important to explain the important concepts so that users would understand the reasons behind the rules and guidelines.

- **Explain principles of vocabulary control.** As above, many new users of the standard were unfamiliar with the basic principles of vocabulary control. Consequently, it was important to explain these ideas without getting too technical and to provide good examples to illustrate the points being made.

- **Include the electronic information environment.** At the time Z39.19 was first proposed, little information was available in electronic form. Thirty years later, many content resources are available in electronic form and many controlled vocabularies are available in electronic form. Consequently, the revision had to include information on display formats for print, electronic, and web-enabled environments.

- **Include additional user access methods.** The original version of Z39.19 assumed that the predominant search mode would be post-coordinated searching using Boolean operators. In today's information environment, the Standard must also provide for controlled vocabularies that are suitable for browsing and navigation as well as keyword searching.

- **Expand beyond the abstracting and indexing (A&I) applications.** As the information industry has grown and matured, other applications have realized that the formalism and discipline inherent in controlled vocabularies would be useful in their applications as well.

- **Include web applications.** As information resources and the tools to manage them have moved inexorably to the Internet, it became imperative to include web implementations, both of controlled vocabularies and their target databases, within the scope of the revision.

Extended to cover more than documents

A major change in the revised Z39.19 was to extend the concept of a *document*. The scope of the revised standard was broadened by use of the term *content object* in place of *document*. A *content object* is any information-bearing entity. It can exist in virtually any physical or electronic form. *Content objects* may be contained in databases or archives or other information repositories or they may simply be one or more web sites on the Internet. Further, the metadata describing a *content object* is also itself a *content object*.

Extended to include different types of Controlled Vocabularies

Similarly, for this revision, the notion of a *thesaurus* was expanded to include other types of *controlled vocabularies* such as *lists, synonym rings*, and *taxonomies*.

- **Lists**

A *list* is a simple group of terms

Example:

Alabama
Alaska
Arkansas
California
Colorado
...

Lists are frequently used in Web site pick lists and pull-down menus.

- ***Synonym Rings***

A *synonym ring* is a list of synonyms or near-synonyms that are used interchangeably for retrieval purposes.

Example:
Speech disorders
Speech defects
Speech, disorders of
Defective speech

Synonym rings are frequently used to enhance retrieval in systems where the content is not indexed or where the indexing vocabulary is not controlled.

- **Taxonomies**

A *taxonomy* is a set of preferred terms, all connected by a hierarchy or polyhierarchy.

Example:
chemistry
organic chemistry
polymer chemistry
nylon

Taxonomies are widely used in classification schemes and for web navigation systems.

- ***Thesauri***

A *thesaurus* is a controlled vocabulary with multiple types of relationships.

Example:
rice
UF paddy
BT **cereals**
BT **plant products**
NT **brown rice**
RT **rice straw**

where:
UF = Used For
BT = Broader Term
NT = Narrower Term
RT = Related Term
bold type face = a preferred term

Three *relationship types* are permitted in a thesaurus. These are:

- Use/Used for – indicates the preferred term
- Hierarchy – indicates broader and narrower terms
- Associative – indicates other types of relationships among terms

Enhanced Display formats

Being able to make controlled vocabularies available in electronic format and on the Web allows users much greater flexibility to move around the collections of terms. Today's web-enabled controlled vocabularies take advantage of navigation tools such as GO TO, Browse, and hyperlinks to additional types of displays and specific information such as Scope Notes, History Notes, Tree Structures, and so forth.

Interoperability

As the number of information resources and controlled vocabularies used to index them has grown, the need for tools that will enable cross-database and cross-system access has grown in importance. A lot of work and research has been done to address the many problems. However, there are no general solutions to the problem. The revised standard identifies the critical issues so that controlled vocabulary designers and users will be aware of the potential problems.

A special case of *interoperability* involves both indexing and searching content across multiple languages.

Construction, testing, maintenance, and management systems for controlled vocabularies

A final section was added to present rules and guidelines for the construction, testing, and maintenance of controlled vocabularies. Management systems for controlled vocabularies are also included in the discussion.

Conclusion

Clearly, this revision of Z39.19 – important as it is – takes only another small step into the future. It goes beyond the concept of a "document" and extends the guidelines to controlled vocabularies in general, not just thesauri. Much more work remains to be done to ensure that ANSI/NISO Z39.19 and the British standards still under development are compatible with each other and with the ISO work in the same area. Much more work also will be required to address the complex issues introduced by multilingual controlled vocabularies and other issues of interoperability. This committee decided that in 2003-2004 when most of the work on this version was done, not enough was known to advance the standard into these areas and that that work would have to wait for the next revision.

References

ANSI/NISO Z39.19 – 2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, www.niso.org/standards/index.html

Fayen, Emily: A new standard for controlled vocabularies. The Indexer, 24(2004)2, p. 62.

Karinch, Maryann: NISO Releases New Standard for Vocabulary Control, November 17, 2005. www.niso.org/news/releases/pr-vocab-11-05.html

Standardization, USA, ANSI/NISO Z39.19, thesaurus

Normung, Vereinigte Staaten von Amerika, ANSI/NISO Z39.19, Thesaurus

The Author

Emily Gallup Fayen

is Vice President, Digital Content and Access at MuseGlobal, Inc, a developer of metasearch and content integration software. From 2003 to 2005, Ms. Fayen chaired the ANSI/NISO Committee charged with revising ANSI/NISO Z39.19 (1993), Guidelines for the Construction, Format, and Management of Monolingual Thesauri. She wrote her master's thesis on controlled vocabularies and later participated in some of the early work to standardize thesauri.

VP, Digital Content and Access
MuseGlobal, Inc.
334 Beacon Street #10
Boston, MA 02116
USA
emily@museglobal.com

Informationstheorie: Der Jahrhundertbluff

Eine zeitkritische Betrachtung (Teil 1)

Robert Fugmann, Idstein

In ihrer „Mathematical Theory of Communication" beschrieben Shannon und Weaver zur Mitte des vergangenen Jahrhunderts die Technik einer möglichst ungestörten und wirtschaftlichen Nachrichten-Übertragung. Die Einbeziehung der Deutung (Interpretation und Semantik) und der Nutzung der Nachrichten (Pragmatik) blieben der späteren Entwicklung überlassen. Ohne dass es zu dieser Fortentwicklung gekommen wäre, wurde der Geltungsbereich der Theorie jedoch bald auf den gesamten Kommunikationsprozess ausgedehnt. Dies geschah dadurch, dass diese Theorie in „Information Theory" umbenannt wurde, mancherlei Widersprüchen aus der Fachwelt zum Trotz. Noch immer wurde kein Unterschied zwischen Nachricht und Information gemacht, und einer jeglichen Nachricht und jeglichem Signal wurde eine neu definierte Art von „Informationsmenge" zugewiesen. Dieser rein statistische Begriff ist weit entfernt von der ureigentlichen Bedeutung des Wortes „Information". Was eine Nachricht bedeutet und ob sie für den Empfänger verständlich, interessant und nützlich ist, bleibt in dieser Theorie außer Betracht. Die Ursachen, der Verlauf und die Folgen dieser Verirrung werden untersucht und kritisiert. Wenn und so lange auch die Informatik einen solchen „Informations"-Begriff zu ihrer Grundlage hat, entbehrt sie jeglicher Kompetenz auf dem Gebiet dessen, was traditionell und umgangsprachig unter Information verstanden wird. Durch die ungerechtfertigte Beanspruchung und Durchsetzung von Zuständigkeit für das Gesamtgebiet der Information *hat die „Informations"-Theorie weitverbreitet großen Schaden verursacht. Dies gilt bei aller Anerkennung der großen Fortschritte in der Informatik bei der reinen Technik der Datenverarbeitung, die aufgrund dieser Theorie ebenfalls erzielt worden sind.*
Teil 1 behandelt die index-relevanten Aspekte der „Informationstheorie".

Information theory: the bluff of the century
In the middle of the last century, Shannon and Weaver in their Mathematical Theory of Communication *described the technique of an optimally undisturbed and economical transfer of messages through data transmission channels. The inclusion of the interpretation of messages and of their semantics and pragmatics was postponed to a later point in time. Even though the latter development was not subsequently carried out, however, the scope of the theory was soon extended and it was postulated that it was valid for the entire communication process. This was done by renaming the theory as "information theory", despite various objections from the information profession. A distinction between message and information (to be construed as interpreted message) was still not made. In addition, a newly defined "amount of information" was assigned to each signal – a purely statistical entity that is defined by the frequency with which a message under consideration is encountered. The meaning of a message – whether it is of interest and whether it is understandable and useful for the receiver – is irrelevant in this theory. This type of information concept is far from what is traditionally and colloquially understood by the word "information". The origin, spread and consequences of this aberration are investigated and criticized. Through its restriction to the mechanics and statistics of information transfer, computer science based on information theory loses the competence to deal with what is conventionally understood by "information" in information science. Through its unjustified claim to jurisdiction and competence to the entire information field, this type of computer science based on information theory has caused a great deal of damage. While fully recognizing the substantial progress that has been achieved in the field of computerized data processing, management should realize that information science is substantially different from computer science and that successful use of it requires a correspondingly different approach.*

1 Einführung

Das Sammeln und Wiederauffindbarmachen von publiziertem Wissen gewinnt auf allen Gebieten immer größere Bedeutung. Je mehr Wissen erarbeitet worden ist, desto nützlicher ist der gesicherte Zugriff zu diesem Wissensschatz. Weitverbreitet stellt sich die Aufgabe, aus diesem großen und fortgesetzt weiterwachsenden Schatz von Erfahrungen dadurch Nutzen zu ziehen, dass diese Erfahrungen auf Anfrage zur Wiederverwertung wieder aufgefunden werden können. Angesichts dieses Bedarfs werden oftmals die Nutzer von Informationsdienstleistungen auf informationswissenschaftlichem Gebiet tätig, nun aber als Neulinge auf einem Gebiet, auf dem sie meistens keine Ausbildung genossen haben und keine Erfahrungen gesammelt haben.

So erblicken viele branchenfremde Neulinge das Wesen der „modernen" Informationsbereitstellung ausschließlich darin, in einem computerisierten Speicher diejenigen Dokumententexte wiederzufinden, von denen ihnen bekannt ist (oder von denen sie vermuten), dass in ihnen ganz bestimmte Wörter vorkommen. Mit einer Philosophie der Begriffe, der Textinterpretation und Indexierung sind sie niemals in Berührung gekommen, und entsprechend gravierend sind die Fehler, welche ihnen bei der Gestaltung ihrer Informationssysteme unterlaufen.

Im Bewusstsein ihrer Kenntnisse im Umgang mit Computern nehmen viele branchenfremde Neulinge auch Abstand davon, sich die Kenntnisse anzueignen, welche auf dem Gebiet der Informationsbereitstellung seit Jahrhunderten in den Bibliotheken gesammelt worden sind[1]. Es herrscht die Meinung vor, dass die Lösung der dortigen Probleme allein in hochentwickelter Informationstechnologie liegt, dass auf jegliche informationsphilosophisch fundierte Grundlage ver-

1 Hahn (2003) „The sociologists got us started, but we quickly developed our own body of research in the second half of the last century. " (Es sind also nicht die Bibliothekare und nicht die Informationsexperten, von denen man gelernt hat.)

zichtet werden könne und dass es für die Wiederauffindung von publiziertem oder betriebsinternem Wissen genüge, wenn Texte Wort für Wort und Bilder Punkt für Punkt eingespeichert werden. Die sogenannte „Informations"-Theorie ist ein prominentes Beispiel für die Überfremdung der Informationsprofession durch branchenfremdes Gedankengut, hier aus der Nachrichtentechnik stammend, wie in diesem Aufsatz dargelegt werden soll.

Die Neulinge bringen auch Fachbegriffe in Gestalt derjenigen Definitionen mit, wie sie ihnen auf ihren Fachgebieten vertraut sind, zuweilen ohne Rücksicht darauf, ob diese Definitionen auf dem Informationsgebiet noch sinnvoll sind (vgl. Abschnitt 7.1). Die Neulinge vermissen auf dem Informationsgebiet auch manches, was ihnen z.B. auf ihrem angestammten Gebiet vertraut ist, insbesondere auf ihrem vertrauten naturwissenschaftlich-technischen Gebiet. Sie glauben, solchen vermeintlichen Mängeln abhelfen zu müssen und abhelfen zu können, in Verkennung des eigenartigen Wesens der Informationsbereitstellung. Beispiele hierzu sind der vermeintliche Mangel an Reproduzierbarkeit („consistency", vgl. Abschnitt 7.5) und an Objektivität (vgl. Abschnitt 3) und die vermeintliche Notwendigkeit des Messbarmachens von Information.

Zunächst wird in Erinnerung gerufen, was von einer guten Theorie erwartet werden darf. Sodann unternehmen wir einen philosophischen Ausflug in die Welt der Begriffe und untersuchen die Verirrungen, in die man leicht gerät, wenn man seine Arbeit auf einer verfehlten, von branchenfremder Perspektive beherrschten Theorie stützt. Zum Schluss wird dargelegt, wie vielerlei Mängel der beschriebenen Art in der sogenannten „Informations"-Theorie anzutreffen sind und welcher Schaden der wissenschaftlichen Gemeinschaft zugefügt wird, wenn diese Theorie außerhalb der Nachrichtentechnik, in der sie sich und für die sich entwickelt hatte, angewandt wird.

2 Wesen und Nutzen von Theorie

Was dürfen wir von einer guten Theorie erwarten? Sie erklärt beobachtete Phänomene und sagt Zukünftiges voraus. Dann braucht man das Zukünftige nicht erst mühsam experimentell zu erforschen, auf Wegen, die zuweilen prohibitiv zeitraubend und kostspielig wären. Einer guten Theorie verdanken wir es, dass wir nicht viele Brücken bauen müssen, bis endlich eine Konstruktion gelungen ist, die den Belastungen gewachsen ist und bezahlbar ist. Wir brauchen nicht viele Flugzeuge zu Probe zu bauen, bis endlich eines gefunden ist, das wirklich fliegt. Wir wissen auch, dass es kein Perpetuum mobile geben kann, also keine Maschine, die sich selbst antreibt, und dass alle Anstrengungen in dieser Richtung Verschwendung von Zeit und Arbeitskraft sind. Wir unterscheiden auch die determinierten Prozesse von den indeterminierten Prozessen und wissen, dass die letztgenannten sich einer jeden befriedigenden Programmierung widersetzen und geben uns diesbezüglich keinen Illusionen hin[2]. Gute Theorie gibt rechtzeitig zu erkennen, wenn man im Begriff ist, sich in eine Sackgasse zu begeben, längst bevor dem empiristisch veranlagten Menschen das tote Ende der Sackgasse deutlich wird und damit dann auch die Notwendigkeit einer zeitraubenden, kostspieligen und entmutigenden Umkehr.

Es liegt im Naturell vieler Menschen, dass sie nicht die Mühe des angestrengten Denkens auf sich nehmen wollen und dass sie mehr vom Beobachten und Probieren halten. Sie wollen nur Fakten zur Kenntnis nehmen. Mit einem solchen Naturell findet man sicherlich viele Tätigkeitsfelder. Aber Kreativität ist jedenfalls nicht die Stärke dieser Menschen. Wenn sie auf Gebieten eingesetzt werden, wo gerade dieses Talent gefordert ist, dann werden sie nicht nur kaum Fortschritte herbeiführen, sondern sie können sogar noch zum Hindernis werden. Auf dem Informationsgebiet können sie den sprichwörtlichen Flurschaden anrichten (Schumacher 2000[3]), wenn sie dort ihren Standpunkt durchsetzen können und nach der Maxime arbeiten „Probieren geht über Studieren".

3 Der Begriff

Die reibungslose, missverständnisfreie zwischenmenschliche Verständigung über die Gegenstände gemeinsamen Interesses ist seit alters her ein fundamentales Anliegen der Menschen. Eine solche Kommunikation ist die Voraussetzung für die nutzbringende und kreative Wiederverwertung von Erfahrungen, die andernorts gemacht worden sind. Schon in der Philosophie der Antike nimmt dieses Anliegen im Gebiet der Hermeneutik einen breiten Raum ein und findet auch in der Gegenwart großes Interesse (vgl. z.B. Riggs 1997).

Die Menschen machen sich unbewusst immer ein Bild vom Wesen der Gegenstände ihres Interesses. Ein jeder Gegenstand hat sehr viele Eigenschaften und Merkmale, von denen man die wesentlich erscheinenden Merkmale in Betracht zieht und die unwesentlichen Merkmale außer acht lässt, je nach der eigenen Interessenlage und Perspektive.

Beim Kauf eines Tisches beispielsweise beachtet man vielleicht die Größe, das Material, aus welchem er hergestellt ist, die Zahl der Tischbeine, den Preis. Unwesentlich hingegen dürfte für die meisten Menschen der Name der Person sein, die ihn ausgeliefert hat, von welchem Holzhändler das Material stammt und mit welchem Lastwagen es geliefert worden ist, in welcher Schreinerei die Tischplatte gehobelt worden ist, usw. Es gibt zahllose Merkmale von einem jeden Gegenstand, aber es sind immer die wesentlichen Merkmale, die der Mensch unbewusst und mehr oder minder fest mit dem Gegenstand seines Interesses assoziiert. So ist z.B. für denjenigen, der am Tisch sitzt und daran arbeitet, das Gewicht des Tisches unwesentlich, nicht jedoch für denjenigen, der berufsmäßig Möbel transportieren muss.

Die Bedeutung eines Wortes spiegelt sich immer am präzisesten in der Definition des Gegenstandes wider, welche mit einem sprachlichen Ausdruck gemeint ist. Von einer Flüssigkeit mit dem Namen Benzol werden sich die Fachleute verschiedener Gebiete recht unterschiedliche Definitionen bilden, je nachdem, welche Merkmale dieses Gegenstandes für sie wesentlich oder nebensächlich sind. Müssten die Chemiker mit der Benzol-Definition des Biologen arbeiten, in welcher vielleicht ausschließlich die Wirkungen auf das Blut von Warmblütlern genannt sind, dann käme es zu keiner nützlichen Verwendung dieses Stoffes für die Herstellung wichtiger anderer Stoffe in der Chemie, denn hierfür ist die Kenntnis der molekularen Struktur dieses Stoffes notwendig. Müsste umgekehrt der Biologe mit der Definition des Chemikers arbeiten, in welcher vielleicht ausschließlich die Sechsringstruktur der Kohlenstoffatome und die eigenartigen Bindungsverhältnisse zwischen denselben im Vordergrund stehen, dann wäre in der Biologie ebenfalls jeglicher Fortschritt blockiert.

Nur dadurch, dass der Mensch (das für ihn!) Wesentliche vom Unwesentlichen, von den Nebensächlichkeiten trennt, findet er sich in seiner Welt zurecht, mögen diese Entscheidungen auch hochgradig subjektiv sein, aus der Sicht seines Fachgebiets und auch aus der persönlichen Perspektive. Diese Subjektivität der Betrachtungsweise ist charakteristisch für jegliches Denken und Handeln des Men-

2 Als „indeterminiert" gelten alle Prozesse, die in einer praktisch unendlich großen Mannigfaltigkeit von Varianten ablaufen können, ohne dass es voraussehbar ist, welche dieser Varianten zum Zuge kommen wird oder zum Zuge gekommen sein könnte. Beispiele sind das freie Formulieren eines natursprachigen Textes, demzufolge auch das Fremdsprachenübersetzen, das Abstract-Formulieren, das sachkundige, traditionelle Indexieren.

3 „Großen Flurschaden richtet die Werbewirtschaft an, die Werbebotschaften mit Information verwechselt".

schen. Den rein naturwissenschaftlich-technisch eingestellten Menschen hingegen mag dieser Mangel an Objektivität, dem er auf informationswissenschaftlichem Gebiet begegnet, ein Dorn im Auge sein.

So verstehen wir in der klassischen Weise unter einem

Begriff

die Summe der jeweils wesentlich erscheinenden Merkmale eines Gegenstandes,

dasjenige also, was mit einer sprachlichen Ausdrucksweise gemeint ist.

So ist es nicht verwunderlich, dass z.B. in der Kommunikations- und Nachrichtentechnik andere Begriffe in anderen Begriffsdefinitionen gebraucht werden als in der Informationswissenschaft. Auf die bedenkenlose und verfehlte Transplantation des „Informations"-Begriffs aus der Nachrichtentechnik in das informationswissenschaftliche Gebiet, wie wir ihr bei der sogenannten „Informations"-Theorie begegnen, werden wir noch ausführlich zu sprechen kommen (vgl. Abschnitt 8 im Teil 2).

Voraussetzung für eine funktionierende zwischenmenschliche Kommunikation ist es, dass die Menschen vereinbaren, was sie mit einem Wort ihrer Umgangs- oder Fachsprache meinen oder auch mit einem Bild, z.B. mit einem Verkehrszeichen, und dass sie sich an diese Vereinbarungen halten. In definitionsartigen Beschreibungen werden dann diejenigen Merkmale genannt, welche man aus seiner fachlichen oder persönlichen Perspektive an dem betreffenden Gegenstand für wesentlich hält.

Unter der

Definition eines Begriffes

verstehen wir in der klassischen Weise die *Nennung* all dieser jeweils wesentlich erscheinenden Merkmale.

Diese Definitionen lenken die Aufmerksamkeit auf die essentiellen Merkmale der Gegenstände ihres Gebiets und verhindern die Ablenkung und Verirrung auf Nebensächlichkeiten und auf die damit verbundene Vergeudung von Zeit und Arbeitskraft und auf die damit ebenfalls verbundene Behinderung und Lähmung des Fortschritts. Importiert man auf seinem Gebiet die Begriffsdefinitionen fremder Gebiete, so öffnet man Tür und Tor für die Fruchtlosigkeit der Nebensächlichkeitenforschung (vgl. Abschnitt 7.5). Sachdienliche Definitionen der Begriffe auf einem Fachgebiet sind das Rückgrat von jeglicher guter Theorie, d.h. von einer Theorie mit Erklärungswert und Voraussagekraft. Wir werden sogleich einigen Beispielen von solchen verfehlten Begriffsimporten begegnen und werden sodann an der „Informations"-Theorie ein Beispiel von solchen Verirrungen eingehend betrachten.

So ist zum Beispiel in der informationswissenschaftlichen Forschung der Unterschied zwischen Wort und Begriff weitgehend verloren gegangen[4]. „Begriff" wird oftmals nur noch als eine gehobene Bezeichnung für ein Textwort angesehen, und die Verwechselung von Wort mit Begriff wird den Software-Benutzern regelrecht eingehämmert. Wenn dort vom „Such-Begriff" die Rede ist, welcher zur Volltextsuche eingegeben werden soll, dann ist in Wirklichkeit ein Such-Wort gemeint.

4 Die Ausdrucksweise für Begriffe

Damit wir über die Begriffe unseres Interesses untereinander in Gedankenaustausch treten können, müssen wir ihnen in unserer natürlichen Fach- und Umgangssprache Ausdruck verleihen, verbal oder bildhaft. Hierzu gehört es, dass man den Gegenständen seines Interesses bestimmte Zeichen zuordnet und dass möglichst weitgehendes Einvernehmen über die Bedeutung dieser Zeichen und Symbole herrscht.

Zum Zeitpunkt des Ursprungs eines Begriffes steht für dessen Benennung nur die definitionsartige, nichtlexikalisch- umschreibende Ausdrucksweise zur Verfügung (vgl. Fugmann (1999, S.30), sieht man einmal von den (lexikalischen) Eigennamen für Individualbegriffe ab[5]). Dies geschieht unter Zuhilfenahme des verfügbaren Wortschatzes und der Grammatik der Fach- oder Umgangssprache. Man spricht beispielsweise vom „Verlust der Bodenhaftung eines Kraftwagens bei Überflutung der Fahrbahn und bei überhöhter Geschwindigkeit" oder von der „krankhaften Scheu vor dem Betreten einer jegliche Art von Wasserfahrzeug".

Erst wenn immer öfter über einen solchen Begriff kommuniziert werden muss, wird die Sprachschöpfung aktiv. Es wird ein Lexikalikum (Fugmann (1999, S. 29) eingeführt, zum Beispiel „Wasserglätte" oder „Thalassophobie", damit man sich nun in prägnanter Form ausdrücken kann. Aber auch nach der Einführung eines Lexikalikums bleibt die nichtlexikalische Ausdrucksweise weiter in Gebrauch. Entweder ist man sich nicht sicher, ob das Lexikalikum wirklich genau dasjenige ausdrückt, was man im Sinn hat, oder man hat die lexikalische Ausdrucksweise nicht mehr in Erinnerung oder man ist ihr noch gar nicht begegnet. Sucht man nach Behandlungsmethoden für Thalassophobie oder nach aquaplaning-verhindernden Reifenprofilen, dann ist es für den Suchenden gänzlich nebensächlich, ob diese Begriffe in den Dokumenten seines Interesses in der lexikalischen oder nichtlexikalischen Ausdrucksweise beschrieben sind.

Es ist ein gravierender Irrtum zu glauben,
dass man sich bei der Informationsrecherche nur auf die Suche nach lexikalischen Ausdrucksweisen zu begeben brauche
und dass die nichtlexikalischen Ausdrucksweisen keinerlei Beachtung verdienten.

Wenn heutzutage in der Fachliteratur von „Recherche" die Rede ist, dann ist oftmals stillschweigend überhaupt nur noch die Recherche im „interpretationslos" erstellten Volltextspeicher gemeint. Dieser irrige Standpunkt erfreut sich gegenwärtig großer Verbreitung, weil er einen gravierenden, unüberwindlichen Mangel der modernen Volltext- oder Freitextspeicherung zu verschleiern gestattet, den Mangel nämlich, dass die nichtlexikalischen Ausdrucksweisen in einem solchen Speicher praktisch unauffindbar sind. Dies gilt zumindest bei der Entdeckungsrecherche (vgl. Abschnitt 5). Dem kann nur durch die Abkehr von der nichtinterpretierenden Volltext- oder Freitextspeicherung begegnet werden und nur durch die klassisch-sachkundige, intellektuelle Interpretation von Text und Bild und zur ebensolchen Indexierung.

5 Erinnerungsrecherche versus Entdeckungsrecherche

Eine hochrangige Aufgabe von Informationswissenschaft und -praxis besteht darin, publiziertes Wissen durch die Recherche im Bedarfsfall (wieder-) auffindbar zu machen. Bei jeglichem derartigem Suchen macht es einen großen Unterschied aus, ob man Einzelheiten des Gesuchten bereits genau kennt oder ob dies nicht der Fall ist.

4 Es finden sich sogar Warnungen in der Literatur, einen solchen Unterschied zu machen, mit der Begründung, dass dies allzu kompliziert und verwirrend sei.

5 Unter einem Individualbegriff wird hier nach v. Freytag-Löringhoff (S. 27) ein Begriff verstanden, zu welchem es keinen sinnvollen, spezifischeren (d.h. merkmalsreicheren) Unterbegriff mehr gibt, zumindest nicht auf dem betreffenden Fachgebiet. Beispiele sind Friedrich Schiller, Donau, Zugspitze, Nordpol, usw. Den Gegensatz bildet der Allgemeinbegriff, welcher stets mindestens einen spezifischeren, merkmalsreicheren Begriff unter sich hat. Beispiele sind Münzen, Metalle, Lebewesen, Bauwerke, usw.

Bei der simplen Erinnerungsrecherche („question of recall“, „known item search“, Bernier 1960) erinnert man sich von einer früheren Lektüre her an irgend ein Dokument-Detail der wesentlichen oder auch der unwesentlichen Art, vielleicht an ein Erscheinungsdatum, an eine bestimmte Redewendung. Es kann sich auch um einen Zahlenwert handeln. Wenn man sich erinnert, dass in einer Studie zur Qualität der Internet-Recherchen festgestellt wurde, dass „66%“ der im Speicher enthaltenen Antworten nicht gefunden wurden, dann kann man mit der Suchbedingung „66%“ durchaus erfolgreich suchen[6]. Von einem Buch kann Größe, Farbe, Verlag usw. in Erinnerung geblieben sein, wonach man sich in den Regalen auf die Suche begeben kann, wenn sie nicht allzu groß sind. Für solche bescheidenen Anforderungen würde es genügen, die Bücher in einer Bibliothek der Größe nach (oder gar nur nach ihrem Gewicht!) zu ordnen. Auch ein in Erinnerung gebliebener Autorenname kann eine nützliche Ausweich- und Ersatz-Suchbedingung sein für ein Dokument bestimmten Inhalts. Bei dieser Sachlage befindet man sich ersatzweise auf der Suche nach einem gut bekannten Detail von möglicherweise ganz nebensächlicher Art.

Die heuristisch eigentlich ertragreichen Recherchen jedoch sind diejenigen, bei denen man Dokumente auffindet, denen man noch nicht begegnet ist, deren verbale Details oder sonstige formale Details nicht mehr in Erinnerung sind und bei denen die Vermutungen nach sprachlichen Details ins Uferlose führen würden (Entdeckungsrecherchen, „question of discovery“, „unknown item search“, Bernier 1960). Hier befindet man sich auf der

Suche nach einem *Begriff,*
gleichgültig wie dieser auch immer verbal ausgedrückt sein mag.

Weitverbreitet ist in der gegenwärtigen Praxis die Unterscheidung zwischen der simplen Erinnerungsrecherche und der anspruchsvolleren Entdeckungsrecherche verloren gegangen. Dies hat weitreichende nachteilige Folgen: Ist man mit dem Ergebnis einer Erinnerungsrecherche in einem Speicher von Volltexten leidlich zufrieden, so wird leicht irrtümlich angenommen oder behauptet, dass dies nun bei allen Recherchen der Fall sei, zumindest in Bälde[7], dies unter stillschweigender Einbeziehung auch der Entdeckungsrecherchen. Man macht ja keinen Unterschied zwischen beiden Recherchentypen. Der Forschungsleiter in einem Unternehmen ist leicht durch eine Erinnerungsrecherche zu beindrucken, wenn ihm bei der Vorführung eines solchen Speichers all seine Publikationen auf den Bildschirm gebracht werden, dies zumeist ohne Erwähnung der Mängel und Grenzen dieser Primitivform von Recherche. Dann fällt es ihm leicht, der unternehmensinternen Einführung eines solchen Dokumentationssystems zuzustimmen, zumal da ein solches billig zu gestalten und zu unterhalten ist.

Die Erinnerungsrecherche kommt mit Werkzeug einfachster Art aus und ohne jegliche Art von Informationsphilosophie, insbesondere ohne jegliche Interpretation von Text und Bild. Man begibt sich einfach nur auf die Suche nach einer klar definierten Zeichenfolge, von der man schon im Voraus weiß (oder mit großer Sicherheit vermutet), dass sie in dem gesuchten Dokument vorkommt. Bei der Entdeckungsrecherche hingegen verfügt man nicht über diese Kenntnisse. Das Gesuchte wird in dem „interpretationslos“ angelegten Speicher von Volltexten (oder von Teilen derselben) in einer Art und Weise angetroffen, welche meistens

unvorhersehbar ist, weil nichtlexikalisch, umschreibend
und
nur in Andeutungen besteht, für den Fachmann jedoch klar erkennbar,
und
hochgradig vieldeutig ist.

Dies führt zwangsläufig zu hohem Informationsverlust und zu viel Ballast in den Suchergebnissen in einem Volltextspeicher, und dies bis hin zur gänzlichen Unbrauchbarkeit der Recherchenergebnisse. Ein solcher Speicher ist für Entdeckungsrecherchen ungeeignet.

Überwunden können diese Probleme nur durch sachverständige Interpretation mit anschließender Indexierung[8] von Text und Bild, also nur durch die aufwendige Mitwirkung des sachverständigen Menschen schon bei der Einspeicherung. Hierin liegt der Gegensatz zu einer Art von Einspeicherung, welche nur dem Zweck der Erinnerungsrecherche zu dienen braucht. Wer keinen Unterschied zwischen Erinnerungsrecherche und Entdeckungsrecherche macht, verkennt den großen Unterschied in der Verarbeitung von Dokumenten zu deren Wiederauffindbarkeit. Die sogenannte „Informations“-Theorie verzichtet auf jegliche Text-Interpretation (Semantik und Pragmatik), wie in Abschnitt 8 in Teil 2 dargelegt wird. Dadurch propagiert diese Theorie eine Art von Textverarbeitung, die ausschließlich der Erinnerungsrecherche dient (und selbst dieser nur mit Einschränkungen).

Die Erinnerungsrecherche erfreut sich großer Beliebtheit und weiter Verbreitung. Durch sie und durch die großen Fortschritte in der Kommunikations- und Nachrichtentechnik ist der große Internetspeicher einem großen Publikum zugänglich geworden. Für viele Fragen aus dem Alltag oder aus dem eigenen Fachgebiet bekommt man auf diesem Weg brauchbare Antworten. Wenig bekannt und von den Anbietern gern verschwiegen sind die Mängel und Grenzen dieser Suchmethode.

6 Die Interpretation von Text und Bild

Alles, was der Mensch an Nachrichten oder sonstigen Signalen empfängt, unterliegt unbewusst und sofort der Interpretation, zumindest dem Versuch einer Interpretation. Dies gilt für einen Text, ein Bild oder eine sonstige Nachricht. Man begnügt sich nicht mit der Feststellung, ein rotes Licht gesehen zu haben, sondern man zieht die Schlussfolgerung für sein eigenes Verhalten, je nachdem, ob das rote Licht von einer Verkehrsampel kommt, vom Bremslicht eines Kraftwagens oder vom eigenen Armaturenbrett. Was nicht interpretierbar ist, z.B. ein Wort oder ein Satz in einer unbekannten Sprache, bleibt unverwertet. Nur durch Interpretation und Deutung kann Nachricht zu Information werden (Rechenberg 2000, S. 302).

In einem natur- oder fachsprachigen Volltextspeicher begegnet man einem Wort in den unterschiedlichsten Bedeutungen. Die Vieldeutigkeit der Wörter ist dort nicht bereinigt, und die eventuelle Nebensächlichkeit eines Worts am Platze seines Auftreten ist noch nicht geklärt. So kommt es zu Ballast-Fundstellen.

Es fehlt dort auch die Übersetzung der nichtlexikalischen Ausdrucksweisen in eindeutige lexikalische Ausdrucksweisen, die dann als Suchbedingung brauchbar wären, ebenfalls eine Leistung des sachverständigen Indexers. In den natursprachigen Speichern findet man auf diesem Weg nur diejenigen Texte, in denen man zufällig eine vom Textautor gewählte Ausdrucksweise beim Suchen verwendet hat, dort ausgewählt aus einer Unendlichkeit von denkbaren Ausdrucksweisen und daher meistens unvorhersehbar. Dies führt zur Lückenhaftigkeit des Ergebnisses der mechanisierten Suche in einem solchen nichtinterpretierten Speicher.

6 Als Folge der Unterlassung von jeglicher Interpretation in einem solchen Speicher findet man dann als Ballast aber z.B. auch, dass in einem Kreis von Ingenieuren die Befragten 40-66 Prozent ihrer Arbeitszeit mit Literaturarbeit zugebracht hatten.

7 Hier begegnet man dem von Bates (1998) kritisierten Standpunkt: „Only a little more research and we will have it solved“.

8 Unter „Indexieren“ wird hier das Erkennen der wiederauffindbar zu machenden Essenz eines Dokuments verstanden und das anschließende Darstellen dieser Essenz in ausreichend gut voraussehbarer (rekonstruierbarer) und genügend wiedergabetreuer Form (vgl. FID/Classification Research (1981), p. 96; Fugmann (1979), p. 14; Fugmann (1985), p. 126; 1999, S. 216).

Wollte man den Ablauf eines indeterminierten Prozesses mechanisieren, hier also die Interpretation und Indexierung von Text und Bild, dann müssten die unbegrenzt vielen denkbaren Ausdrucksweisen der nichtlexikalischen Art im Speicher in Betracht gezogen werden. An einem solchen Programm müssten unbegrenzt viele Programmierer arbeiten, sie würden unbegrenzt viel Zeit verbrauchen und würden unbegrenzt hohe Kosten verursachen. Die Unprogrammierbarkeit eines solchen Vorhabens teilt diese Art von Natursprachenverarbeitung mit vielen anderen Prozessen der gleichfalls indeterminierten Art[9].Die Zuversicht, mit welcher gegenwärtig in der Künstlichen Intelligenz an der automatisierten Text- und Bildinterpretation gearbeitet wird, d.h. an der Programmierung des Unprogrammierbaren, erinnert an die Zuversichtlichkeit, mit der die mittelalterlichen Goldmacher am Werk gewesen sind (vgl. die Kritik von Shpackov, (1992)) und bis heute noch immer die Perpetuum-mobile-Konstrukteure[10]. In den großen Volltextspeichern sind Zehntausende von Antworten auf eine gestellte Frage keine Seltenheit, ohne dass man unter Hunderten von geprüften Antworten auch nur einen einzigen Treffer findet. Außerdem wird in solchen Speichern meistens nur ein Drittel und noch weniger von dem Wissen gefunden, das als Antwort auf eine gestellten Frage im Speicher enthalten ist[11].

All die Schulen in Kreisen der Künstlichen Intelligenz, welche die Mechanisierbarkeit der Textinterpretation behaupten, z.B. auch die Mechanisierbarkeit von Indexieren und Abstrahieren, sollten vorrangig daran arbeiten, das nun schon über fünfzig Jahre alte Versprechen der vollmechanisierten Fremdsprachenübersetzung zu erfüllen, bevor sie mit immer neuen unerfüllbaren Versprechungen weitere Verwirrungen stiften.

7 Verirrungen bei der Gestaltung einer Informations-Philosophie

Für die Aufgabe des Einspeicherns und (Wieder-) Findens sachdienlicher Information stehen in der Neuzeit als Werkzeuge Computerprogramme zur Verfügung, welche, wenn sie sinnvoll und sachkundig eingesetzt werden, wesentliche Erleichterungen und Fortschritte im Zugriff zu Informationen der verschiedensten Art darstellen. Sie bieten sich auch demjenigen an, der als Neuling auf dem Gebiet der Informationsbereitstellung tätig wird, dilettantisch oder mit dem Anspruch der Professionalität.

Den meisten Neulingen erscheint diese Aufgabe trügerisch einfach (Bates 1998)[12], weil sie nur die Erinnerungsrecherche vor Augen haben. Sie erblicken die hier zu lösende Aufgabe auf diesem Gebiet also lediglich darin, Texte zu finden, in welchen eine vorgegebene natursprachige Zeichenfolge (erinnert oder gemutmaßt) vorkommt. In dieser Simplifikation wird ihnen diese Aufgabe auch von den meisten Software-Anbietern dargestellt.

Die Neulinge, welche sich aus ihren angestammten Fachgebieten dem Informationsgebiet zuwenden, z.B. aus Medizin, Naturwissenschaft und Technik, aber auch aus näherliegenden Gebieten, wie Informatik und Sprachwissenschaft, haben wenig Neigung gezeigt, sich mit den philosophischen Grundlagen der Informationsbereitstellung zu beschäftigen. (vgl. Fußnote 1). So wird nun manches, was in dieser Profession längst bekannt ist[13] und schon seit langem praktiziert wird, wenn auch unter anderem Namen, erst aufgrund eigener gesammelter Erfahrung neu entdeckt. Manch einer Einsicht steht die (Wieder-) Entdeckung durch die Branchenfremdlinge sogar erst noch bevor, z.B. die Mangelhaftigkeit des Arbeitens mit einem lediglich „controlled vocabulary“ und die Außerachtlassung der Regel von Cutter (Regel vom engsten Schlagwort, s.a. Fugmann, 1999, S. 118, „verbindliches Indexieren“).

Aufgrund von Unkenntnis und mangelnder Erfahrung kommt es immer wieder zu Missgriffen bei der Gestaltung einer Philosophie von Informationssystemen. Welchen Typs diese Missgriffe sind und welche Folgen hiermit verbunden sind, soll im folgenden an einigen Beispielen verdeutlicht werden. Sie reichen von der nachlässigen oder absichtlichen Begriffsverdrehung bis hin zum scheinbar kommunikationsdienlichen Import von Begriffsbildungen aus Nachbargebieten oder gar aus weit entfernten Fremdgebieten. Nach diesen Betrachtungen wird es leichter fallen, die Denkfehler zu erkennen, die sich in der „Informations“-Theorie eingenistet haben (vgl. Abschnitt 8).

7.1 Verirrungen bei der Festlegung der Fachbegriffe

Es liegt im Wesen von Natursprache, dass die dortigen Wörter und Ausdrucksweisen verschiedene Bedeutungen haben, je nachdem, in welchem Zusammenhang die Wörter auftreten. Auch werden die natur- und fachsprachigen Ausdrücke immer wieder in neuen Bedeutungen verwendet, die kein Programmierer voraussehen kann. Es obliegt der Interpretation durch den verständigen Empfänger einer Nachricht als Leser oder Zuhörer, die jeweils gemeinte Bedeutung zu erkennen (vgl. Abschnitt 6). So hat das Wort „Licht“ im Lauf der Zeit schon viele verschiedene Bedeutungen angenommen, je nach dem Zusammenhang, in welchem es verwendet worden ist. Für das Wort „Auge“ lassen sich über hundert verschiedene Bedeutungen zusammentragen. Aber auch für viele vermeintlich so klare Fachausdrücke gibt es eine Vielzahl von unterschiedlichen Bedeutungen.

Der Mensch verfügt aber nicht über grenzenlose Freiheiten, ein Wort der Umgangs- oder Fachsprache in jedem beliebigen Sinn zu verwenden. Wird eine solche Grenze überschritten, dann sind gravierende Missverständnisse unvermeidbar, bis hin zur Lüge und Verleumdung. Der Verfasser ist Zeuge einer Anschuldigung geworden, in welcher ein Gartenbesitzer von seinem Nachbarn als „Kannibale“ bezeichnet worden ist, weil er einen Ahornbaum in seinem Garten auf ein Maß gestutzt hatte, das dem Nachbarn missfiel. Der Nachbar war der ehrlichen Meinung, dass unter Kannibalismus das übermäßige, den Fortbestand eines Baumes gefährdende Zurückschneiden zu verstehen sei.

In einem fachlichen Streitgespräch wurde behauptet, dass ein Computer Lebenserscheinungen zeigen und Durst empfinden könne (Dörner 2000). Der Gesprächspartner (Weizenbaum 2000) bestritt diese Möglichkeit. Die Kontroverse klärte sich erst, als Dörner veranlasst wurde zu offenbaren, was er unter „Leben“ und unter „Durst“ verstand: „Ob ich einen Bedarf an Blutzucker habe oder eine Maschine an elektrischer Energie ist letztlich gleich. Es handelt sich jeweils um das Bedürfnis nach bestimmten Stoffen, um die Lebensvorgänge aufrecht zu erhalten“. Hier gilt eine Maschine also als eine Art von Lebewesen, und wenn sie ausge-

9 Nur dort, wo keine Interpretation erforderlich ist, weil die Texte frei von Mehrdeutigkeiten sind, wo weiterhin keine sprachlichen Ellipsen aufzufüllen sind, wo auch die Ausdrucksweise voraussehbar ist und wo alle Begriffe lexikalisiert sind (was z.B. bei Wetterberichten der Fall ist), lassen sich hier Erfolge erzielen. In der weit überwiegenden Mehrzahl der Fälle herrschen diese vereinfachenden Voraussetzungen nicht. Dort haben die vollautomatischen Übersetzungen lediglich Unterhaltungswert (vgl. das von Rechenberg (2000, S. 324) zitierte Beispiel).

10 Wer die (zumindest baldige) vollautomatische Interpretierbarkeit von Text und Bild behauptet (und damit zugleich die baldige oder gar sofortige Entbehrlichkeit jeglicher intellektueller Interpretation, dies zugunsten des Absatzes seiner Computerprogramme), propagiert damit die Primitivisierung der Informationsbereitstellung und damit auf längere Sicht den Ruin jeglicher brauchbarer Inhaltserschließung und Informationsbereitstellung auch für die Entdeckungsrecherche.

11 Z.B. Green, David (1999): 66% des Web nicht gefunden. Password 2/99, 16.

12 „Information retrieval has looked deceptively simple to generations of newcomers to the field“.

13 Z.B. „Ontologie“, „Metadata“, „Vererbung“ von Begriffsmerkmalen in einer Abstraktionshierarchie, usw.

schaltet ist, dann wird sie als „durstig“ angesehen.

Handelt es sich darum, die Leistung von Schülern zu beurteilen, dann ist es abwegig, die Subjektivität in der herkömmlichen Art der Beurteilung zu beklagen und stattdessen die Definition von Leistung aus der Physik zu übernehmen, welche lautet: „Leistung ist Kraft mal Weg pro Zeit“. Die Eltern bekämen die Zeugnisse ihrer Kinder in Watt ausgehändigt, vielleicht nicht ohne den Hinweis, dass nun endlich die wünschenswerte Objektivität in der Benotung der Schüler erreicht sei. Es würde auch ganz der Art entsprechen, mit welcher Shannon auf den Rat von John von Neumann seine „Informations“-Theorie mithilfe des Fachausdruckes „Entropie“ gegen Kritik abgeschirmt hatte, wenn man Leistung als „Erg pro Zeit“ definierte und in die Form eines Integrals kleidete mit der Begründung: „Niemand weiß, was Erg in Wirklichkeit ist. Dadurch ist man in der Diskussion immer im Vorteil“ (vgl. Abschnitt 8.4).

Wenn eine vielversprechende, dem konventionellen Sprachgebrauch stillschweigend widersprechende und daher trügerische Bezeichnung für einen Gegenstand gewählt wird, mit der Absicht der Irreführung, dann kann man von Etikettenschwindel sprechen. Einer Begriffsverdrehung von solchem Ausmaß begegnet man in der sogenannten „Informations“-Theorie bei der missbräuchlichen Verwendung des Fachausdruckes „Entropie“ (siehe unten). Um keine Unklarheit offen zu lassen, ist in dieser Abhandlung das Wort Information immer in Anführungsstriche gesetzt, wenn es in dem Sinn gebraucht wird, wie es in dieser Theorie geschieht.

Fast ebenso abwegig, aber trotzdem weitverbreitet (weil leicht und schnell durchzuführen) ist die Messung der Leistung eines Fachmannes daran, wie oft er zitiert wird, dies ohne jede Unterscheidung, ob positiv oder negativ zitiert und zitiert von wem und warum und ohne Beachtung der komplexen Zitatenpsychologie, welche hier zur Wirkung kommt und zu welcher es bereits eine umfangreiche eigene Literatur gibt (z.B. Warner 2003, Kaube 2005).

In einer solchen Art von Verirrung befindet sich die sogenannte „Informations“-Theorie, wenn sie den Anspruch erhebt, außerhalb der Kommunikations- und Nachrichtentechnik anwendbar zu sein und wenn sie den „objektiven Informations-Begriff“ von dort in das informationswissenschaftliche Gebiet transplantiert.

14 Nach dem Wahlspruch: „Was ich nicht weiß, macht mich nicht heiß“.

Hier handelt es sich nämlich um eine Art von „Information“, welche mit all dem, was man konventionell und gutgläubig unter diesem Wort versteht (siehe oben), nichts zu tun hat, sondern um eine weltfremde, buchstaben- und wörterstatistische Konstruktion, geprägt ausschließlich von der Nebensächlichkeit der Wahrscheinlichkeit ihres Vorkommens (vgl. Abschnitt 8) und propagiert vor allem wegen ihrer Messbarkeit.

Häufig begegnet man in der Informationsprofession der Definition vom Begriff als der „Bedeutung eines Wortes“, einem Standpunkt, wie er in der Sprachwissenschaft geläufig ist. Das bedeutet aber, dass einem Begriff erst dann Existenz zugesprochen wird, wenn sich für ihn eine lexikalische Ausdrucksweise eingebürgert hat. Alle nichtlexikalischen, definitionsartigen Ausdrucksweisen werden dann nicht als Repräsentationen eines Begriffes anerkannt. Für einen Fragesteller handelt es sich jedoch um eine Nebensächlichkeit, ob das Gesuchte in lexikalischer oder nichtlexikalischer Form ausgedrückt ist. Beim Suchen nach Unfällen durch Wasserglätte ist immer damit zu rechnen, dass das Wort „Wasserglätte“ in einem einschlägigen Artikel gar nicht vorkommt, und dass vielmehr dieser Gegenstand in völlig ausreichender Deutlichkeit in umschriebener Form präsentiert wird (vgl. Abschnitt 4). So ist die Ausdrucksweise immer zufallsabhängig und unvorhersehbar und ist eine Nebensächlichkeit für den Fragesteller.

Hier begegnet man einem zwar sprachwissenschaftlich fundierten, aber informationswissenschaftlich verfehlten Standpunkt. Die Aufgabe scheint aus dieser Perspektive ja gar nicht darin zu bestehen, auch die nichtlexikalischen Ausdrücke auffindbar zu machen oder aufzufinden. Die hierdurch eintretenden Informationsverluste gelten aus dieser Sicht also gar nicht als Verlust, und es werden keine ernsthaften Anstrengungen unternommen, diesen zu mindern.

7.2 Der Informationsbegriff

Nach all diesen vorbereitenden Betrachtungen zu dem Thema, welche Missgriffe in der Welt der Begriffe unterlaufen können, wenden wir uns dem Informationsbegriff zu, dem Fundament der gesamten Informationsprofession. Wir begegnen einem breiten Spektrum von Informationsbegriffen unterschiedlicher Herkunft und unterschiedlicher Eignung für das Informationsgebiet, zum Teil auch entstanden aus Unkenntnis und aus dem Streben nach Originalität aufseiten ihrer Schöpfer.

Aus der Literatur des Informationsgebiets lassen sich leicht hundert verschiedene Definitionen des Informationsbegriffs zusammentragen. Wir vergleichen einige Varianten davon in Bezug auf ihre Sinnhaftigkeit und Fruchtbarkeit für das Gebiet von Informationswissenschaft und -praxis. Am negativen Ende der Brauchbarkeits-Skala werden wir demjenigen Informationsbegriff begegnen, wie er der „Informations“-Theorie zugrunde liegt und aus der Nachrichtentechnik importiert wurde. Für einen Rückblick auf die Geschichte des Informationsbegriffes und auf die umfangreiche Literatur zu dieser Theorie empfehlen sich Rechenberg 2003 (im dortigen Abschnitt: „Kurze Geschichte des Informationsbegriffes“, S. 317-318) und Kuhlen (2004: Kapitel A1: „Information“, S. 9-17). Wir können unsere Argumentation großenteils allein auf diese wenigen Quellen stützen.

Was traditionell und umgangssprachig unter

Information
verstanden wird,
ist *Auskunft, Belehrung, Unterrichtung, Aufklärung,*

wie man in vielen Nachschlagewerken wie z.B. in Brockhaus, Duden, Wahrig nachlesen kann. Hiergegen dürfte wohl kaum etwas einzuwenden sein. Dies ist es auch, was man traditionell und konventionell als unbefangener Nutzer eines Informationssystems erwartet, wenn man ein solches System mit der Bitte um Dienstleistung in Anspruch nimmt.

7.3 Die Immaterialität der Information

Zunächst setzen wir uns mit der irrigen, wenn auch in der zeitgenössischen Informationsforschung und betriebswirtschaftlichen Perspektive weit verbreiteten Auffassung auseinander, dass es sich bei der Information um einen Gegenstand der herkömmlichen materiellen Art handele, um einen Gegenstand auch, welcher in ganz der üblichen Weise angeboten, gekauft und verkauft werden könne, den Regeln des freien Marktes entsprechend.

Nachfolgend sind einige Eigenschaften des Phänomens der Information aufgezählt, wodurch sie sich fundamental von jeder Art von materiellem Gegenstand unterscheidet.

1. Mangel an Information wird nicht gespürt, nur geahnt[14]. Hat man hingegen keinen Kühlschrank, dann spürt man den Mangel sofort.
2. Information kann nicht besichtigt und nicht, wie ein materieller Gegenstand, zur Probe in Benutzung genommen werden und kann nicht bei Nichtgefallen oder Nichteignung wieder zurückgegeben werden.

3. Niedrige Qualität von Information kann darin bestehen, dass sie mit viel Ballast bereitgestellt wird. Es macht einen großen Unterschied aus, ob man eine nützliche Information zusammen mit zehn wertlosen anderen Hinweisen geliefert bekommt oder ob man sich diese eine Information unter zehntausend wertlosen anderen Mitteilungen erst heraussuchen muss. Es kommt hingegen nicht vor, dass man bei der Bestellung eines Kühlschrankes gleichzeitig viele andere Gegenstände mitgeliefert bekommt, aus denen man sich das Gewünschte erst noch heraussuchen muss.
4. Niedrige Qualität von Information besteht auch in Lückenhaftigkeit, was bis zur Wertlosigkeit oder sogar Gefährlichkeit gehen kann, z.B. wenn eine Korrektur oder Ergänzung zu einer Äußerung unaufgefunden oder unauffindbar bleibt. Lückenhafte Information ist schwer als solche zu erkennen, sie kann kaum reklamiert werden. Einen Kraftwagen hingegen, dem beim Kauf wichtige Teile fehlen, wird man gar nicht erst annehmen.
5. Viele Menschen denken irrtümlich, sie erführen von ihrem Informationssystem alles, was sie wissen müssen und behaupten die Entbehrlichkeit von noch vollständiger arbeitenden (und auch entsprechend teureren) Informationssystemen. Es wird auch behauptet, man brauche (generell) nicht alles zu wissen, was im Speicher an Einschlägigem gesammelt ist. Hierbei wird übersehen, dass in mangelhaften Informationssystemen die Selektion der Antworten auf eine Fragestellung dem Zufall überlassen ist. Ein solcher Fragesteller würde seine Meinung sofort ändern, wenn er all dasjenige gezeigt bekäme, was ihm zunächst (und dies sogar mit seiner ausdrücklichen und leichtfertigen Billigung) bei seiner billigen Recherche entgangen ist. (Hiermit soll nicht bestritten werden, dass man zuweilen auch mit lückenhafter Informationsbereitstellung gut bedient sein kann.)
   In dieser Unkenntnis und in diesem voreiligen Urteil liegt ein großes Hindernis für dauerhaft leistungsfähige Informationssysteme. Schon Sokrates wusste: Niemand kann wissen, was und wie viel er nicht weiß. Der Verfasser kennt manch einen Fall, in welchem jemand, der glaubte, er sei aus einem (hochgradig lückenhaft arbeitenden) Informationssystem gut bedient worden, aus allen Wolken fiel, als er andernorts in einem leistungsfähigeren Informationssystem recherchierte, gezwungen von seinem Dienstvorgesetzten.
6. Ein Primitivsystem zur Informationsbereitstellung arbeitet anfänglich trügerisch zufriedenstellend, weil es sich im anfänglichen Kleinstadium auf *Gedächtnisleistung* stützen kann und weil man dann allein mit Erinnerungsrecherchen auszukommen glaubt. Mit fortschreitender Größe und Inanspruchnahme entwickelt sich ein solches System leicht in den Zustand von Unbrauchbarkeit wegen der ständigen Zunahme von Ballast und Informationsverlust. Eine Reparatur ist dann praktisch unmöglich, denn sie würde die totale Neubearbeitung des gesamten, bisher angesammelten Dokumentenbestandes erfordern.
   Ist hingegen ein materieller Gegenstand unbrauchbar geworden, so lässt er sich reparieren oder ersetzen, ohne dass nun die gesamte Arbeit, die an diesem Gegenstand und mit demselben bisher geleistet worden ist, mit einem verbesserten oder ersetzten Gegenstand wiederholt werden muss. Außerdem entwickelt sich ein jeder materieller Gegenstand keineswegs mit derjenigen Zwangsläufigkeit in den Zustand der praktischen Unbrauchbarkeit, wie es bei den billigen Primitivsystemen zur Informationsbereitstellung der Fall ist, wenn sie, immer größer werdend, immer häufiger mit Entdeckungsrecherchen beauftragt werden.
7. In knappen Zeiten kann die Produktion von materiellen Gegenständen gedrosselt oder eingestellt werden. Die Produktion kann beim Ansteigen der Nachfrage meistens leicht wieder aufgenommen werden, und zwar ohne dass dies nachteilige Auswirkungen auf die bisherige Produktion hat. Stellt man hingegen vorübergehend die Inhaltserschließung von Zeitschriften und Büchern ein, dann entsteht eine Lücke, die erstens die gesamte bisherige Arbeit (durch die nun entstehende Lückenhaftigkeit) entwertet und zweitens kaum wieder geschlossen werden kann. Dies würde nämlich die Einarbeitung und vorübergehende Beschäftigung von zusätzlichem Personal erfordern.
8. Information kann beliebig weiter gegeben werden, ohne dass hierbei Spuren hinterlassen werden. Information breitet sich selbsttätig in nicht vorhersehbarer und in nicht rekonstruierbarer Weise aus, dies ganz im Gegensatz zu einem materiellen Gut.
9. Beim Spender von Information entsteht bei der Weitergabe keine Lücke. Hat man hingegen sein Auto abgegeben, dann steht es einem nicht mehr zur Verfügung.
10. Information wird nicht durch Nutzung aufgebraucht. Hat man hingegen seine Briketts verfeuert, dann ist der Keller leer.
11. Wiederholte Lieferung ein und derselben Information bedeutet keine Bereicherung, sondern zumeist sogar Belästigung. Schafft man sich weitere Kühlschränke an, dann erweitert man hingegen sofort spürbar seine Haushaltsausrüstung.
12. Ist ein und die selbe Information auch im Besitz eines anderen Nutzers, so kann dies wertmindernd sein, weil dies den Verlust eines Wettbewerbsvorteils bedeuten kann. Hingegen tritt für mich keinerlei Wertminderung ein, wenn das Modell meines Kühlschrankes oder Kraftwagens auch in vielen anderen Haushalten in Benutzung ist. (Ein Informationsbesitz aufseiten vieler anderer Menschen kann aber auch wertsteigernd sein, wenn gemeinschaftliches Know-how die Grundlage einer vorteilhaften Gemeinschaftsarbeit und Arbeitsteilung ist und wenn dieser Weg letztlich doch zur Festigung einer – nunmehr gemeinschaftlichen – Marktposition führt.)
13. Information ist gedeutete Nachricht (vgl. z.B. Rechenberg 2000, S. 302). Deutung und Interpretation einer Nachricht können sich bei der Weitergabe derselben fortgesetzt ändern. Entsprechend wandelt sich die mit der Nachricht verbundene Information. Es ist kaum rekonstruierbar, an welcher Stelle und zu welchem Zeitpunkt dies in der Weitergabekette geschehen ist. Selbst der Ursprung einer in Umlauf befindlichen Information ist kaum rekonstruierbar. All dies ist bei materiellen Gütern nicht der Fall.
14. Der Empfang von Information kann leicht bestritten werden, und es kann leicht behauptet werden, dass man das diesbezügliche Wissen selbst aufgefunden habe. So kommt es notorisch zur Unterbewertung von Informationsdienstleistungen.
15. Information kann nicht beliebig gestapelt werden, ohne dass Wertminderung oder gar Wertverlust dadurch eintritt, dass die Wiederauffindbarkeit des Gespeicherten fortgesetzt abnimmt. Ein großer Vorrat an Information erfordert nämlich zu seiner Erschließung für Suchgänge grundlegend andere Erschließungsmethoden als ein kleiner Vorrat (Mangel an Stapelbarkeit, Mangel an „scalability"). Hat man fortgesetzt Information gespeichert, ohne die Erschließungsmethode zu verbessern (was nur sehr selten prakti-

ziert werden kann), dann leidet hierunter die Brauchbarkeit des gesamten Informationsspeichers, wie er sich im Laufe der Zeit entwickelt. Es kommt hingegen nicht vor, dass der Wert der Kühlschränke, die ein Hersteller bereits geliefert hat, dadurch leidet, dass die Produktion unverändert fortgesetzt wird. Auch braucht kein Hersteller seine bereits ausgelieferten Kühlschränke nachzubearbeiten, wenn er eine Verbesserung einführt.

Betrachtet man Information als einen
rein physikalischen Gegenstand,
mit dem man in der herkömmlichen Art
Handel treiben kann
und von dem man glaubt, ihn nach Art
und Menge messen zu können,
wie es in der „Informations"-Theorie
geschieht,
dann verkennt man fundamental das
Wesen vom Phänomen der Information.

7.4 Verirrungen beim Informationsbegriff

Wir knüpfen hier an den oben zitierten Informationsbegriff an, wie er aufseiten eines Auskunftsuchenden dominieren dürfte, im Sinne von Auskunft, Belehrung, Unterrichtung, Aufklärung. Dieser Betrachtungsweise kommen die nachfolgend zitierten Definitionen. Am nächsten:

„Information is an attribute of the receiver's knowledge and interpretation of the signal, not of the sender's, not of some omniscient observer's nor of the signal itself" (Fairthorne (1954)).

Information ist jegliche Nachricht, aus welcher der Empfänger etwas lernen kann. Eine Trivialität kann nicht als Information gelten „as we do not learn anything from it" (Shrejder 1965, zitiert bei Belkin 1978), p. 72).

Information ist gedeutete Nachricht (Rechenberg (2000, S. 302).

Information: ist „Message that proves to be of interest to a recipient" (Fugmann (1985), p. 126 und 1993, S. XI).

Charakteristisch für diese Definitionen ist, dass in ihnen das hohe Maß an Subjektivität zum Ausdruck kommt, welches durch die feste Einbindung des Empfängers gegeben ist.

Der Mensch als Empfänger bereitgestellten Wissens kommt auch bei Yovits (1981) ins Spiel. Hier ist Information „data of value in decision making". Aber Vieles nimmt man dankbar und mit Interesse zur Kenntnis, ohne einen Anlass zu haben, darüber eine Entscheidung zu fällen. Beispiele hierfür sind: was Freunde im Urlaub erlebt haben oder am Arbeitsplatz, Glücks- und Unglücksfälle in fremden Ländern, allerlei Fortschritte in Wissenschaft und Technik ohne Auswirkungen auf das eigene Privatleben, das Verfolgen eines sportlichen Wettkampfes. Nur selten wird man eine Tageszeitung oder eine populär-wissenschaftliche Zeitschrift mit der Absicht aus der Hand gelegt haben, Entscheidungen nach der Lektüre zu treffen.

Die Mitwirkung des Empfängers ist auch in solchen Informations-Definitionen inbegriffen, welche auf der Verringerung von Ungewissheit basieren (vgl. Neveling, Wersig 1975; Beling, Port, Strohl-Goebel 2006, S. 38).

Aber bezüglich vieler Dinge hat man zu keinem Zeitpunkt Ungewissheit empfunden, und trotzdem nimmt man manch eine diesbezügliche Nachricht mit Interesse auf und empfindet sie als echte Information. Kein Arzt und kein Patient hat jemals zu alten Zeiten Ungewissheit darüber empfunden, ob und wann das Penicillin entdeckt werden wird, und trotzdem ist die Nachricht darüber sicherlich auf großes Interesse gestoßen, zumindest bei einem speziellen Kreis von Menschen. Man verspürt auch keinerlei Ungewissheit darüber, ob ein guter Bekannter noch lebt, mit dem man korrespondiert, und trotzdem freut man sich über ein Lebenszeichen von ihm.

Auch wird durch den Empfang von Information Ungewissheit oftmals sogar vergrößert. Erfährt man, dass der gebuchte Flug gestrichen ist, dann möchte man wissen, welcher Ersatzflug zur Verfügung steht und zu welcher Abflugzeit und mit welcher Ankunftszeit am Ziel. Zitiert ein Fachkollege in seinem Vortrag eine wichtige Feststellung, so möchte man vielleicht die bibliographische Fundstelle erfahren und noch mehr über die weiteren Zusammenhänge. Verringerung von Ungewissheit ist also ungeeignet als Grundlage einer Definition des Informationsbegriffs auf informationswissenschaftlichem Gebiet.

Zuweilen wird als Information auch jegliche Nachricht angesehen, die etwas Neues zur Kenntnis bringt. Aber nicht alles, was neu ist, ist auch Information, z.B. die Familiennachrichten in der Tageszeitung einer fremden Stadt. Umgekehrt kann etwas Bekanntes interessant sein, wenn es nur in Vergessenheit geraten war.

Als Information wird auch „everything which changes the state of the mind" aufgefasst. Aber auch durch bloßes Vergessen oder durch bloße Gemütsbewegung tritt eine Änderung im state of the mind ein, ohne dass von Zufluss von Information im oben erwähnten landläufigen Sinn die Rede sein kann.

Als Information ist auch schon jeglicher materielle Gegenstand betrachtet worden, von dem man etwas lernen kann (Buckland 1991), also nicht nur jegliche derartige Nachricht[15].

Aber das gilt dann auch für jeden Baum (z.B. durch seine Jahresringe), für jeden Stein und für jeden Himmelskörper (z.B. durch seine Zusammensetzung). Der Informationsbegriff ist hier also nicht auf Nachrichten (messages) eingeschränkt, was auf dem Informationsgebiet der Fall sein sollte.

Ein noch anderer „Informations"-Begriff, wie er in der Physik verbreitet ist, ist ebenfalls ungeeignet für die Informationswissenschaft: „Die Einheit der Information ist ein bit" (Moore-Hummel 1976, S. 195). Aber dann müssten zwei Bücher der gleichen Art oder zwei Zeitungen die doppelte „Information" enthalten.

Information im landläufigen, konventionellen Sinn als Wissenszuwachs aufseiten eines Empfängers aufgefasst, ist ein inhärent subjektiver, nichtquantifizierbarer Begriff und nicht messbar. „Ihre Messung in Bit oder Bit/Zeichen ist daher Unfug" (Rechenberg 2003, S. 321), wenn die Messung auf dem Informationsgebiet praktiziert wird und nicht auf die Kommunikations- und Nachrichtentechnik beschränkt bleibt. Schließlich ist es ja die Information im landläufigen, konventionellen Sinn, welche von einem Informationssuchenden als eine Leistung des Informationsfachmannes erwartet wird, und nicht eine buchstabenstatistische Konstruktion.

Ungeeignet für das Gebiet von Informationswissenschaft und -praxis ist auch ein Informationsbegriff, bei welchem nicht nur die Mitwirkung des Empfängers einer Nachricht außer acht gelassen wird, sondern auch all dasjenige, was eine Nachricht bedeutet und ob sie überhaupt für den Empfänger verständlich ist. Einem solchen „Informations"-Begriff begegnet man (und entgegen dem Rat von Rechenberg 2003, S. 326) auch in der „Informations"-Theorie, wo wir diese Auffassung einer näheren Betrachtung unterziehen werden. Eine Zusammenstellung neueren Datums von mehreren Informationsbegriffen und der

15 „One does not normally think of trees as information but trees are informative at least in two ways ... If lumber and firewood can be informative, one hesitates to state categorically of any object that it could not, in any circumstances be information or evidence". Es wird aber auch eingeräumt: „... whether a particular object, document, data or a fact is going to be informative depends on the circumstances ..."

diesbezüglichen Literatur findet man z.B. auch bei Kuhlen (2004) und bei Bonitz (1990).

Es sind in der Literatur auch viele Definitionen des Informationsbegriffs anzutreffen, denen man nicht näherungsweise Eignung für das Informationsgebiet zuerkennen kann. Beispiele sind:

Information is a substitute for time, space, capital, and labor.

Information is the equivalence class of signals with the same meaning.

Mit solchen fachfremden Definitionen werden Nebensächlichkeiten eingeschleppt und betont, Nebensächlichkeiten, die aber auf anderen Gebieten durchaus zur Essenz der Gegenstände gerechnet werden könnten. Orientiert sich ein Denkansatz an Nebensächlichkeiten, dann gibt er zu erkennen, dass entweder eine fachfremde Perspektive oder Laienhaftigkeit am Werke gewesen ist. Kein Käufer eines Autos wird seine Entscheidung von der Helligkeit der Innenbeleuchtung abhängig machen oder von der Größe des Handschuhfaches.

7.5 Die Nebensächlichkeiten-Forschung

Wenn sich auch die Forschung auf Nebensächlichkeiten konzentriert, dann liefert sie Ergebnisse, für die sich hernach niemand interessiert. Es kann zu einer Flut von methodischen Eintagsfliegen kommen und zur Abkehr von den eigentlich brennenden Fragen des eigenen Fachgebiets. Es kann sogar zum Ruin des eigenen Fachgebiets führen, wenn es den Nebensächlichkeitenforschern gelingt, dem Management vorzuspiegeln, dass sie es sind, welche mit ihren „revolutionären" Ansätzen den Weg in die Zukunft bahnen.

Nebensächlichkeitenforschung absorbiert einen Teil der immer nur knapp vorhandenen Ressourcen an Geld, Arbeitskraft und Mitarbeiter-Motivation. Sie setzt sich falsche Ziele und produziert Weisheiten, für welche sich in der Praxis, für die man schließlich arbeitet, niemand interessiert, und der Fortschritt stagniert. Trotz großer Anstrengungen tritt die Wissenschaft auf der Stelle.

Es ist für einen Literatursuchenden beispielsweise ausgesprochen nebensächlich, ob er den Gegenstand seines Interesses in der lexikalischen oder nichtlexikalischen Form beschrieben vorfindet. In der weitverbreiteten Verkennung dieser Tatsache und in der Verirrung auf sprachwissenschaftliches Gebiet (wo die Schreibweise wesentlich ist) oder nachrichtentechnisches Gebiet (wo die Bedeutung einer Nachricht als unwesentlich angesehen wird und wo nur die Häufigkeit einer Zeichenfolge eine Rolle spielt) liegt eine wesentliche Forschrittsbarriere.

Die Nebensächlichkeitenforschung, in Gang gebracht und unterhalten durch das Eindringen von informationswissenschaftlichen Neulingen aus der Kommunikations- und Nachrichtentechnik und Informatik, nimmt in der zeitgenössischen informationswissenschaftlichen Literatur einen breiten Raum ein. Dann interessiert es z.B., wie oft der Doppelpunkt in den Überschriften von Aufsätzen vorkommt („Titular Colonitis"), welche Buchstaben in welchen Sprachen am häufigsten vorkommen, wer wen wie oft aus welchen Ländern zitiert hat (ohne zwischen den zustimmenden und den ablehnenden Zitaten zu unterscheiden), welche Wörter am häufigsten in Nachbarschafts-Vergesellschaftung vorkommen, wie viele Co-Autoren es bei den Aufsätzen aus unterschiedlichen Ländern gibt. Es interessiert dort auch, wie lang durchschnittlich ein Wort oder ein Satz in den Aufsätzen eines Autors ist, usw.

So hat z.B. auch die Ansicht weite Verbreitung gefunden, dass die (gut messbare) Reproduzierbarkeit eines Prozesses zugleich auch ein Kriterium für die Zuverlässigkeit bei der Abwicklung des Prozesses darstelle, manch einem Einwand von professioneller Seite zum Trotz. Dann müsste aber eine automatisch durchgeführte Übersetzung, Indexierung, Zusammenfassung von höchster Qualität sein, und die intellektuell durchgeführten Prozesse müssten mangelhaft sein, in Anbetracht ihrer mangelhaften Reproduzierbarkeit. Das Gegenteil davon ist jedoch der Fall, denn die Vollautomatik ist in diesen Fällen meistens bis zur Unbrauchbarkeit schlecht und nur als Notbehelf diskutabel.

Auch wenn es bei der Gewichtung von Suchergebnissen nach dem Grad ihrer Relevanz für eine Fragestellung eine Rolle spielt, in welcher zufälligen Reihenfolge die einzelnen Antworten durch den Suchmechanismus erzielt worden sind, dann hat hier die Nebensächlichkeitenforschung ihre Spuren hinterlassen.

Erstaunlich wenig Interesse findet dagegen die z.B. brennende Kernfrage, wie man die Qualität der Informationsbereitstellung bezüglich Vollständigkeit und Genauigkeit verbessern kann.

Es ist das Überhandnehmen der Nebensächlichkeitenforschung, welches Goldmann (1987) zu der berechtigten Frage veranlasst hat: „Information Science Research: Where is the Meat?" Auch Crowley (1999) beklagt die Weltfremdheit von großen Teilen der zeitgenössischen informationswissenschaftlichen Forschung.

Ein Musterbeispiel für nebensächlichkeitenorientierte Begriffsbildung in der Informationswissenschaft ist der aus Informatik und Nachrichten- und Kommunikationstechnik stammende „Informations"-Begriff, wie er der „Informations"-Theorie zugrunde liegt. Diese Theorie wird im nächsten Abschnitt in Teil 2 näher betrachtet. Im Rahmen dieses Artikels können aus der Fülle der Arbeiten, die zu dieser Theorie im letzten Halbjahrhundert erschienen sind, nur wenige in die Betrachtungen einbezogen werden. Eine Grundlage hierfür bilden insbesondere die überblickschaffenden Arbeiten von Bar-Hillel (1964), Shaw und Davis (1983), Rechenberg (2000 und 2003), R.R. Kline (2004).

Die detaillierte kritische Betrachtung der „Informationstheorie" selbst erfolgt in Teil 2.

Literatur

Bar-Hillel, Jehoshua (1964): Language and Information. Reading, Massachusetts. Palo Alto. London: Addison Wesley Publishing Company Inc.

Bates, Marcia J. (1998): Indexing and Access for Digital Libraries and the Internet: Human, Database, and Domain Factors. Journal of the American Society for Information Science 49(1998)13, pp. 1185-1205.

Beling, Gerd; *Port*, Peter; *Strohl-Goebel*, Hildburg (2006): Terminologie der Information und Dokumentation. Frankfurt am Main: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis. ISBN 3-925474-58-7.

Belkin, N. J. (1978): Information Concepts for Information Science. Journal of Documentation 34(1978)1, pp. 55-85.

Bernier, Charles L. (1960): Correlative Indexes VI: Serendipity, Suggestiveness, and Display. American Documentation 11, pp. 277-287.

Bonitz, Manfred (1990): Wissen – Information – Informatik. Informatik Berlin 37 (4), 132-135.

Buckland, Michael (1991): Information as Thing. Journal of the American Society for Information Science 42(1991)5, pp. 351-360.

Crowley, Bill (1999): The Control and Direction of Professional Education. Journal of the American Society for Information Science 50(1999)11, pp. 1127-1135.

Davis, Charles H.: siehe *Shaw*, Debora.

Dörner, Dietrich (2000): Kann ein Rechner Durst haben? FOCUS Nr. 22, S. 182.

Fairthorne, Robert (1954): The Theory of Communication. ASLIB Proceedings 6, 255-267, zitiert bei Buckland (1991, S. 352).

FID/Classification Research (1981): FID/CR News. Inernational Classification 8, p. 96.

Freytag-Löringhoff, *von*, Bruno: Logik – Ihr System und ihr Verhältnis zur Logistik. 4. Auflage. Stuttgart: Kohlhammer Verlag.

Fugmann, Robert (1979): Toward a Theory of Information Supply and Indexing. International Classification 6(1979)1, pp. 3-15.

Fugmann, R. (1985): The Five-Axiom Theory of Indexing and Information Supply. Journal of the American Society for Information Science 36(1985)2, pp. 116-129.

Fugmann, R. (1999): Inhaltserschließung durch Indexieren: Prinzipien und Praxis, S. 9. Frankfurt am Main: Deutsche Gesellschaft für Dokumentation, 1999. ISBN 3-925474-38-2,

Goldmann, Nahum (1987): Information Science Research: Where is the Meat? Journal of the American Society for Information Science 38, pp. 215-216.

Green, David (1999): 66% des Web nicht gefunden. Zitiert in PASSWORD 2/99, S.16.

Hahn, Trudi Bellardo (2003): What Has Information Science Contributed to the World? Bulletin of the American Society for Information Science 29(2003)4, pp. 2-3.

Hummel, Dietrich O. (1976), siehe *Moore*, Walter J. (1976).

Kaube, Jürgen (2005): Zitierst Du mich, zitier ich Dich. Frankfurter Allgemeine Sonntagszeitung, 16. Oktober 2005, Nr. 41, S.78.

Kline, Ronald R. (2004): What is Information Theory a Theory of? In: Rayward, Boyd W.; Bowden, Mary Ellen, editors (2004): The History and Heritage of Scientific and Technological Information Systems. Chemical Heritage Foundation, Proceedings of the 2002 Conference. ASIST Monograph Series. Information Today Inc. Medford, New Jersey.

Kuhlen, Rainer (2004): Information. Kapitel: Jenseits der Informationstheorie. Grundlagen der praktischen Information und Dokumentation. KG Saur, München 2004, S. 3-17.

Moore, Walter J.; *Hummel*, Dietrich O. (1976): PC - Physikalische Chemie. Walter de Gruyter. Berlin, New York. ISBN 3-11-002127-7.

Neveling, Ulrich; *Wersig*, Gernot (1975): Terminologie der Information und Dokumentation, für das Komitee Terminologie und Sprachfragen der Deutschen Gesellschaft für Dokumentation Frankfurt/Main (1975). Verlag Dokumentation München ISBN 3-7940-3625-5.

Port, Peter (2006), siehe *Beling*.

Rechenberg, Peter (2000): Was ist Informatik? München, Wien: Carl Hanser Verlag, 3. Auflage, ISBN 3-446-21319-8.

Rechenberg, Peter (2003): Zum Informationsbegriff der Informationstheorie. Informatik Spektrum, 14. Oktober 2003, S. 317-326.

Riggs, Fred (1997): Onomantics and Terminology - Part IV: Neologisms, Neoterisms, Metaterms, Phrases, and Pleonisms Holly. Knowledge Organization 24(1997)1, pp. 8-17.

Schumacher, Dieter (2000): Wissensmanagement: Des Kaisers neue Kleider. PASSWORD Oktober, S. 2-4.

Shaw, Debora; *Davis*, Charles H. (1983): Entropy and Information: A Multidisciplinary overview. Journal of the American Society for Information Science 34(1983)1, pp. 67-74.

Shpackov, A.A. (1992): The Nature and Boundaries of Information Science(s). Journal of the American Society for Information Science 43, 678-680.

Shrejder, J. (1965): On the semantic characteristics of information. Information Storage and Retrieval 2(1965), pp. 221-233, p. 227.

Shrejder, J., zitiert bei *Belkin*, N.J. (1978, p. 72): Information Concepts for Information Science. Journal of Documentation 34, pp. 55-85.

Strohl-Goebel, Hildburg (2006), siehe *Beling*.

Warner, Julian (2003): Citation Analysis and Research Assessment in the United Kingdom. Bulletin of the American Society for Information Science and Technology 30, No, 1, Oct./Nov. pp. 26-27.

Weizenbaum, Joseph (2000): Kann ein Rechner Durst haben? FOCUS Nr. 22, S. 182.

Wersig, Gernot, siehe *Neveling*, Ulrich.

Wiener, Norbert, siehe *Bar-Hillel* (1964), p. 288.

Yovits, J. (1981): Information Flow and Analysis: *Theory*, Simulation, and Experiment I. Basic Theoretical and Conceptual Development. Journal of the American Society for Information Science 32(1981)3, pp. 187-202; pp. 203-210.

Information, Informationswissenschaft, Informatik, Kommunikationswissenschaft, Mathematik, Begriff, Definition, Geschichte

Information, Computer Science, Information Science, Concept, Communication Science, Mathematics, Definition, History

DER AUTOR

Dr. Robert Fugmann

Vormals Leiter der zentralen wissenschaftlichen Dokumentationsabteilung der Hoechst AG, Lehrbeauftragter und Gastprofessor an mehreren Hochschulen im In- und Ausland.

Alte Poststraße 13
65510 Idstein

NACHRICHTEN

140 Jahre Chemieforschung vollständig digitalisiert

FIZ CHEMIE Berlin hat das Chemische Zentralblatt vollständig digitalisiert. Aus 40 Metern Buch sind zwei Terabyte Daten entstanden. Mit moderner Softwaretechnologie kann der komplette Inhalt des bedeutenden Referatedienstes nun im Volltext durchsucht werden.

Von 1830 bis 1969 war das Chemische Zentralblatt, ursprünglich als „Pharmaceutisches Central Blatt“ gegründet und später mehrfach umbenannt, das wichtigste Nachschlagewerk der Chemie. Für den Informationsdienst fassten wissenschaftliche Fachredakteure den Forschungsfortschritt in der Chemie in Kurzreferaten zusammen. Als Grundlage dafür wurden internationale Fachpublikationen ausgewertet. In den Referaten tauchen so gut wie alle namhaften Chemiker des 19. und 20. Jahrhunderts auf. Neben den wertvollen fachlichen Informationen lassen sich aus den Referaten auch historische Rückschlüsse auf gesellschaftliche, politische und wirtschaftliche Entwicklungen der letzten beiden Jahrhunderte ableiten.
FIZ CHEMIE bietet den Datensatz zur Integration in Intranets an und entwickelt zudem eine Datenbank, die online im Web verfügbar sein wird. Für den Einbau in Intranets kann der Datensatz komplett oder in Jahrgängen bezogen werden. Die Freischaltung der Datenbank ist für Frühjahr 2008 geplant.

Ansprechpartner: Stephan *Heineke*, FIZ CHEMIE Berlin, Postfach 12 03 37, 10593 Berlin, Tel.: (030) 399 77-225, Fax: (030) 399 77-132, heineke@chemistry.de

Förderpolitik sorgt für 30 Jahre bewegtes Leben

FIZ Karlsruhe feierte Ende Oktober seinen 30. Geburtstag mit einem Festakt im Schloss Bruchsal

Vera Münch, Hildesheim

Ungewöhnliche Töne klangen Ende Oktober durch die Prunksäle von Schloss Bruchsal nördlich von Karlsruhe. Es waren aber nicht nur die Klangbeispiele, mit denen der Dirigent Christian Gansch seinen Festvortrag unterlegte, und auch nicht die Melodien der faszinierenden Musikautomaten, die im Museum im Schloss zur Feier des Tages angespielt werden durften. Vielmehr sorgten die unverblümten Grußworte für Töne, wie man sie auf einer runden Geburtstagsfeier eher nicht erwartet. Sie waren durchsetzt von politischen Erklärungen und Forderungen.

Fotos: FIZ Karlsruhe

Die Frau an der Spitze: Sabine Brünger-Weilandt führt FIZ Karlsruhe seit 2003.

„Es war zeitweise nicht so selbstverständlich, 30 zu werden", überraschte Sabine Brünger-Weilandt die Festgäste gleich zu Beginn ihrer Begrüßung mit einem sehr offenen Wort. In den drei Jahrzehnten seit seiner Gründung 1977 sei FIZ Karlsruhe zeitweise durch harte Zeiten gegangen, berichtete die Geschäftsführerin, die seit 2003 an der Spitze des Unternehmens steht. Dann lieferte sie die Erklärung: „Auf politischer Ebene hat das schwierige, aber erfolgreiche Agieren von FIZ Karlsruhe zu der Frage geführt, 'Was nutzt uns FIZ Karlsruhe?'". Erst in der Woche vor der Geburtstagsfeier sei in dieser existenziellen Frage ein positiver Abschluss erzielt worden. Die Länder haben ihren Anteil an der Förderung von FIZ Karlsruhe von 15 auf 25 Prozent aufgestockt.

Die Überraschung stand vielen Festgästen und Mitarbeitern ins Gesicht geschrieben. Obwohl FIZ Karlsruhe nicht zum ersten Mal in seiner 30jährigen Geschichte vor einer ungesicherten Zukunft als gemeinnützige Serviceeinrichtung stand, hatten die wenigsten Kenntnis davon, wie sehr der Verteilungskampf um die Höhe der Förderquoten von Bund und Ländern FIZ Karlsruhe in den letzten Monaten erneut bedroht hatte.

Brünger-Weilandt kündigt E-Science als zweites Standbein an

Von der Unverzichtbarkeit der Serviceeinrichtung zur Unterstützung der Forschung überzeugt haben Brünger-Weilandt und ihr Führungsteam die Politik durch eine deutliche Ausrichtung des Angebotes auf die neuen Zukunftsmärkte und eine konsequente betriebswirtschaftliche Unternehmensführung. „FIZ Karlsruhe hat sich entschieden, E-Science als zweites Standbein aufzubauen", gab die Geschäftsführerin bekannt. Mit dieser Entscheidung sei der Schritt von der Vision zur Strategie getan. Zielgruppen sind zunächst Universitäten und Forschungsorganisationen. Später soll das E-Science-Angebot auf alle Kundengruppen erweitert werden. „Vernetzung des Wissens ist seit 30 Jahren Kernkompetenz von FIZ Karlsruhe. Kompetenzen, Know-how und Tools fließen in unsere E-Science-Aktivitäten ein".

KnowEsis, eSciDoc und Fedora-Kompetenz

Die aktuellen E-Science-Aktivitäten von FIZ Karlsruhe umfassen die Produktlinie KnowEsis, das Projekt eSciDoc und das Fedora Center of Excellence. KnowEsis stellt Lösungen und Services zum Management heterogener Informationen bereit (siehe Kasten). Im Projekt eSciDoc, das vom BMBF gefördert wird, entwickelt FIZ Karlsruhe gemeinsam mit der Max-Planck-Gesellschaft (MPG) eine integrierte Informations-, Kommunikations- und Publikationsplattform für netzbasiertes wissenschaftliches Arbeiten in der MPG. FIZ Karlsruhe übernimmt in der Kooperation zentrale Entwicklungs- und IT-Dienstleistungen sowie die Gewährleistung für langfristige Verfügbarkeit und Langzeitarchivierung. (www.escidoc-project.de). Die Softwaretools rund um eSciDoc werden, soweit möglich, auf der Basis von Open-Source-Produkten aufgebaut und sollen wieder als Open Source verbreitet werden. Das Gesamtsystem wird FIZ Karlsruhe auch anderen Wissenschaftsorganisationen als Dienstleistung anbieten. Im Fedora Center of Excellence gibt FIZ Karlsruhe Erfahrungen und Kompetenzen aus der Entwicklungsarbeit mit der „Fedora repository architecture" in Form von Schulungen, Beratungen und Workshops weiter. Fedora gilt derzeit als sehr gute Lösung für das institutionelle Archiv (Repository), das in Unternehmen und Organisationen als Zentrum der E-Science-Struktur aufgebaut werden muss.

Der Arbeitsplatz der Zukunft braucht kreatives Wissensmanagement

Seit knapp hundert Tagen ist Ministerialrat Hermann Riehl vom Bundesforschungsministerium erst als neuer Aufsichtsratsvorsitzender von FIZ Karlsruhe im Amt – im Vergleich zu den 30 Lebensjahren der Einrichtung eine denkbar kurze, allerdings ganz wichtige Zeitspanne, wie aus dem Bericht von Brünger-Weilandt herauszuhören war. Riehl brachte in seinem Grußwort zum Ausdruck, welchen Stellenwert er einer zeitgemäßen wissenschaftlichen Informationsversorgung beimisst. „Im Zuge der internationalen digitalen Vernetzung ist die Informationsversorgung zentraler Aspekt und Maßstab für eine funktionierende Infrastruktur. Der Arbeitsplatz der Zukunft ist ohne dynamisches, kreatives Wissensmanagement nicht vorstellbar". Wissen-

Meilensteine in der Firmengeschichte

1977: Unterzeichnung des Gesellschaftsvertrags durch Bund und Bundesländer. Eintrag in das Handelsregister als „Fachinformationszentrum Energie, Physik, Mathematik GmbH“

1978: Der INKA Online Service startet

1979: Eröffnung eines Büros in Bonn

1983: FIZ Karlsruhe und ACS unterzeichnen den Kooperationsvertrag zur Gründung von STN International, The Scientific and Technical Information Network

1984: MPG, FhG, DPG, VDI und GI werden neue Gesellschafter
Die ersten Datenbanken sind auf STN International online

1987: Änderung des Firmennamens in Fachinformationszentrum Karlsruhe GmbH; Aufnahme der DMV als neuen Gesellschafter;
Geschäftsführer Dr. Werner Rittberger wird mit dem Bundesverdienstkreuz geehrt

1992: Prof. Dr.-Ing. Georg Friedrich Schultheiß übernimmt die Geschäftsführung

1993: die Länder Sachsen, Sachsen-Anhalt und Thüringen werden neue Gesellschafter

1997: Markteinführung des automatischen Volltextvermittlungsdienstes FIZ AutoDoc

2001: Gründung der FIZ Karlsruhe, Inc. in Princeton, New Jersey, USA
FIZ Karlsruhe wird mit dem Total E-Quality Prädikat „Frauenfreundlicher Betrieb“ ausgezeichnet

2003: Sabine Brünger-Weilandt übernimmt die Geschäftsführung

2004: STN International feiert 20jähriges Jubiläum
Unterzeichnung eines „Letter of Intent“ mit FIZ CHEMIE Berlin zum Aufbau eines Kompetenzzentrums
Vertragsunterzeichnung mit der MPG zur Realisierung des gemeinsamen Projekts eSciDoc, gefördert vom BMBF im Rahmen der nationalen E-Science-Initiative

2005: Funktionsorientierte Organisationsstruktur wird eingeführt

2006: Informatik-Portal io-port.net geht online

2007: Jubiläum 30 Jahre FIZ Karlsruhe

schaft und Wirtschaft, so Riehl, seien hier auf starke Partner wie das FIZ Karlsruhe angewiesen, die mit versierter Technik und dem erforderlichen Know-how diese Schnittstelle abdeckten. „Der kürzlich gefasste Beschluss von Bund und Ländern, die gemeinsame Finanzierung des FIZ Karlsruhe auch künftig auf eine solide Grundlage zu stellen, ist sicherlich Indiz für die besondere Anerkennung und Wertschätzung des Geleisteten seitens der öffentlichen Hand“.

Aufsichtsratsvorsitzender Ministerialrat Hermann Riehl: „Der Arbeitsplatz der Zukunft ist ohne dynamisches, kreatives Wissensmanagement nicht mehr vorstellbar“.

WGL-Präsident Prof. Dr. Ernst Th. Rietschel: „Es ist irrwitzig, wenn eine Dienstleistungsgesellschaft Service derart gering schätzt“.

Länder erhöhen Förderanteil und stellen Forderungen

Auch Ministerialdirektor Julian Würtenberger vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg ging in seinem Grußwort sehr offen auf die Verhandlungen zwischen Bund und Ländern ein: „FIZ Karlsruhe ist ein wichtiger Arbeitgeber in der Region Karlsruhe (...) Ich freue mich deshalb besonders, dass die Zukunft von FIZ Karlsruhe als Einrichtung in der Wissensgemeinschaft Gottfried Wilhelm Leibniz in harten, aber fairen Verhandlungen zwischen Bund und Ländern gesichert werden konnte. Dies ist allerdings nur durch eine erhebliche Aufstockung des Länderanteils an der Finanzierung von bisher 15 auf künftig 25 Prozent möglich“. Die Aufstockung verknüpfte Würtenberger mit deutlichen Forderungen: „Damit verbunden sind auch höhere Erwartungen der Länder an FIZ Karlsruhe. Dies gilt besonders im Hinblick auf attraktive Angebote für Hochschulen und Wissenschaftseinrichtungen in den Ländern“. In seinen weiteren Ausführungen erklärte er, Web 2.0 und Social Software verhießen seiner Ansicht nach eine neue Qualität der Nutzung des Internets und in diesem Umfeld könne es für einen Dienstleister wie FIZ Karlsruhe schwierig werden, sich mit innovativen Dienstleistungen zu profilieren. „Die Länder werden den Erfolg von FIZ Karlsruhe daran messen, inwieweit es FIZ gelingt, das Zukunftsthema E-Science mitzugestalten“, bezog Würtenberger Position.

Aus Service wird wissenschaftliche Dienstleistung

Seit der Gründung der Leibniz-Gemeinschaft (WGL) im Jahr 1995 ist FIZ Karlsruhe Mitglied des Zusammenschlusses außeruniversitärer deutscher Wissenschaftseinrichtungen. Ihr Präsident, Professor Dr. Dr. h.c. Ernst Rietschel, erinnerte in seiner Ansprache daran, dass die Bemühungen um den Aufbau und die Bereitstellung einer zeitgemäßen Informationsversorgung mindestens genau so alt sind, wie FIZ Karlsruhe selbst. „Bereits die Geburtsstunde von FIZ Karlsruhe vor 30 Jahren war der erste Versuch, eine flächendeckende Infrastruktur für Information und Dokumentation in der Bundesrepublik zu schaffen“. Rietschel lenkte den Blick auf die rasante technische Entwicklung der letzten drei Jahrzehnte, die gro-

KnowEsis für E-Science
Zur neuen E-Science-Produktlinie von FIZ Karlsruhe gehören derzeit KnowEsis Consulting, KnowEsis Repository und Know Esis Mapping Services. KnowEsis Consulting bietet Beratung zu den Bereichen Requirements-Analyse, Pflichtenheft, Content Models und Einführung, Projektsteuerung, Customization und passgenaue Softwareentwicklung. KnowEsis Repository ist eine Open-Source Archivsoftware. Sie stellt Werkzeuge zur Erfassung, Speicherung und Weiterverbreitung von digitalen Objekten zur Verfügung und kann in Forschungseinrichtungen, Universitäten und Bibliotheken als „Institutional Repository“ eingesetzt werden. KnowEsis Mapping Services bieten Unterstützung bei der Vereinheitlichung von Daten, z.B. zur Vereinfachung interner Arbeitsabläufe sowie zur Datenanalyse und -visualisierung durch Ausgleich semantischer Differenzen

ßen Einfluss auf die Entwicklung von FIZ Karlsruhe hatte. „Damals forschte und studierte der akademische Normalbürger noch ohne Computer, Internet oder Online-Datenbanken". Die Leibniz-Einrichtungen, erklärte Rietschel, sorgten nicht nur für Wissensgenerierung, sondern über Museen, Bibliotheken, Tagungsstätten und natürlich die Fachinformationszentren auch für den Transfer des Wissens zurück in die Gesellschaft, zurück in die Politik, die allgemeine Öffentlichkeit oder wiederum die Wissenschaft. FIZ Karlsruhe gehöre als Serviceeinrichtung zu einer großen Minderheit bei Leibniz. Derzeit werden 19 Einrichtungen als Serviceeinrichtungen klassifiziert. Rietschel sieht darin ein Problem: „In Deutschland ist der Begriff Service oft nicht positiv besetzt, impliziert eine gewisse Zweitklassigkeit (...) Es ist irrwitzig, wenn eine Dienstleistungsgesellschaft Service derart gering einschätzt". Die Leibniz-Gemeinschaft will dieses Problem jetzt wenigstens terminologisch lösen: „Wir haben uns entschlossen, künftig nicht mehr von Service zu sprechen, sondern von wissenschaftlicher Dienstleistung", verkündete Rietschel in Bruchsal.

Anerkennung der Partner: Dr. Kazuhiko Onuma, Geschäftsführer der Japan Association for International Chemical Information JAICI gratulierte und dankte.

Das Orchester als Erfolgsmodell für Unternehmensführung

Nach den fordernden Grußworten freuten sich die Geburtstagsgäste auf einen entspannenden Festvortrag. Aber auch daraus wurde nichts. Mit seinem Transfer der präzisen Arbeitsabläufe und Kommunikationsstrukturen eines Orchesters auf die Führung eines Unternehmens fesselte Christian Gansch das Publikum in höchster Spannung. Der erfolgreiche österreichische Dirigent und Musikproduzent leitete aus dem Zusammenspiel eines Orchesters Regeln und Beispiele für eine erfolgreiche Unternehmensführung ab. Anhand beziehungsreicher Beispiele zeigte er in seinem Vortrag „Kontinuität durch Erneuerung – Das Orchester als Erfolgsmodell" auf, dass die Übertragung orchestraler Lösungsstrategien für ein Unternehmen sehr inspirierend und sinnvoll sein kann – weil ein Unternehmen wie ein Orchester ist. Gansch untermalte, nein, untermauerte, seine Thesen mit kurzen Einspielungen orchestraler Glanzleistungen, die er zuvor so analysiert hatte, dass man die Kraft und Wirkung des gekonnten Dirigierens heraushören konnte.

Der Festvortrag machte drei Dinge eindringlich deutlich:

1. Ein Orchester funktioniert nur dann, wenn es einen Dirigenten hat, der das gemeinsame Ziel klar vorgibt, die Musiker gefühlvoll dort hin lenkt, aber bei Bedarf das Ziel auch gegen Widerstände durchsetzt.

2. Jeder – sowohl die Musiker, als auch der Dirigent – muss sich an der richtigen Stelle zurücknehmen und den Solisten hundertprozentig unterstützen, wenn es dem großen Ganzen dient.

3. Ein wirklich großes Konzert gelingt nur dann, wenn Dirigent und Musiker sich gegenseitig vertrauen und sich gemeinsam auch in ergebnisoffene Situationen wagen.

Festvortragsredner Christian Gansch: „Ein Unternehmen ist wie ein Orchester".

Der große alte Herr: 15 Jahre führte Dr. Werner Rittberger FIZ Karlsruhe von der Gründung bis zu seinem Ruhestand.

Am Ende des Vortrages gab es im Festsaal des Schloss' Bruchsal niemanden, der aus den Thesen und Analogien, die Gansch aufstellte, nicht auch für sich selbst überraschende Lehren und Erkenntnisse gezogen hätte. Über das Orchester als Erfolgsmodell und die Partitur als Leitbild wurde beim festlichen Dinner-Buffet noch lange diskutiert.

FIZ Karlsruhe, Förderpolitik, Dienstleistung, Tagung, Informationswirtschaft

FIZ Karlsruhe, Policy, Funding, Service, Economy

Die Autorin

Vera Münch

Jahrgang 1958, ist freie Journalistin und PR-Beraterin mit Schwerpunkt Wissenschaft und Forschung. Seit vielen Jahren beschäftigt sie sich mit elektronischer Information und Kommunikation (Naturwissenschaften, Technik, Patente, Wirtschaftsinformationen) sowie Informatik und Software-Themen.

PR+TEXTE
Leinkampstr. 3
31141 Hildesheim
Telefon: (0 51 21) 8 26 13
Telefax: (0 51 21) 8 26 14
vera.muench@t-online.de

Information Wissenschaft & Praxis

Jahresregister 2007

Originalbeiträge

Alvarez, A.: Länderübergreifende Projekte – Ratschläge aus der Praxis. Erfahrungen aus Lateinamerika 367
Aschoff, F.; Rausch von Taubenberg, E.: Usability von Webportalen und Web-Suchmaschinen im Vergleich 141
Baier, Ch.: s. *Weichselgartner*, E. 173
Ball, R.: Wissenschaftskommunikation im Wandel. Die Verwendung von Fragezeichen im Titel von wissenschaftlichen Zeitungsbeiträgen in der Medizin, den Lebenswissenschaften und in der Physik von 1966 bis 2005 371
Beger, G.: DGI 2007 1
- : DGI-Online-Tagung: für alle die Information ernst nehmen! 257
- : Einladung zur Kommunikation mit der DGI 65
Bekavac, B.; *Herget*, J.; *Hierl*, S.; *Öttl*, S.: Visualisierungskomponenten bei webbasierten Suchmaschinen: Methoden, Kriterien und ein Marktüberblick 149
Bertram, J.: I would do it again 389
Böth, N.: s. Lüstorff, J. 207
Bosschieter, P.: Translate the index or index the translation? 391
Brandt, P.: Dienstleistungen für die Weiterbildung – das „Informationszentrum Weiterbildung" des Deutschen Instituts für Erwachsenenbildung 203
Braicheski, S.: s. Lande, D. 277
Braschler, M.; *Herget*, J.; *Pfister*, J.; *Schäuble*, P.; *Steinbach*, M.; *Stuker*, J.: Kommunikation im Internet. Die Suchfunktion im Praxistest 159
Browne, G.: Changes in website indexing 437
Brückner, C. : s. *Lüstorff*, J. 207
Burckhardt, D.: Historische Rezensionen online: Eine thematische Suchmaschine von Clio-online 169
Busch, D.: s. *Lande*, D. 277
Clement, W.: Informationsvermittlung im digitalen Zeitalter. Vortrag zur Eröffnung der DGI-Online-Tagung 2006 101
Dextre Clarke, St. G.: Evolution towards ISO 25964: an international standard with guidelines for thesauri and other types of controlled vocabulary 441
Diehl, S. : s. *Lüstorff*, J. 207
Diepeveen, C.; *Fassbender*, J.; *Robertson*, M.: Indexing software 413
Effenberger, C.; *Hartmann*, B.; *Streib*, S.; *Gloystein*, Heide: Das interne Marketing von Competitive Intelligence in Großunternehmen. Erfolgreicher Wettbewerbsbeitrag der Hochschule Darmstadt auf der Jahrestagung 2006 des DCIF 83
Ender, J. : s. *Lüstorff*, J. 207
Evans, R.: Indexing computer books: getting started 425
False, J.: s. *Finke*, S. 35
Fank, M.; *Riecke*, W.: WebIntelligence. Die Bedeutung des Kunden im Internet am Beispiel der Ford Werke Deutschland GmbH 355
Fassbender, J.: Editorial Indexieren – Indexing 385
- : s. *Diepeveen*, C. 413
Fayen, E.G.: Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39.19 for the 21st century 445
Felgner, C.: s. *Friedrich*, A. 302
Fels, G.. Elektronische Medien in der Chemie. Chemieinformation gestern, heute, morgen 51
Ferber, R.: s. *Koschinsky*, G. 7
Finke, S.; *False*, J.; *Grosch*, K.; *Wuttke*, H.-D.: Die akademische Weiterbildungslandschaft in Thüringen. Bildungsportal Thüringen – Plattform für akademische Weiterbildung und E-Learning in Thüringen 35
Friedrich, A.; *Felgner*, C.; *Holenstein*, K.; *Schaller*, S.: Bibliothekarische Ausbildung in Leipzig. Zukunftsperspektivenn beim Praktikertreffen 2007 des Studiengangs Bibliotheks- und Informationswissenschaft 302
Fugmann, R.: Informationstheorie: Der Jahrhundertbluff. Eine zeitkritische Betrachtung (Teil 1) 449
Geiß, D.: Aus der Praxis der Patentinformation. Teil 1: Neue Entwicklungen bei den Internetprovidern 123
- : Aus der Praxis der Patentinformation. Teil 2: Übersicht über die Entwicklung der elektronischen Medien bei den Patentblöcken 230
- : Gewerbliche Schutzrechte. Rationelle Nutzung ihrer Informations- und Rechtsfunktion in Wirtschaft und Wissenschaft. Bericht über das 29. Kolloquium der Technischen Hochschule Ilmenau über Patentinformation und gewerblichen Rechtsschutz 376
- : Die Zukunft der europäischen Patentbibliotheken – Partner für Innovationen! PATLIB Kongress des Europäischen Patentamts vom 14. bis 16. Mai 2007 in Sevilla, Spanien 293
Gering, E.: Zur Terminologie der Information und Dokumentation. Diskussionsbeitrag 47
Geske, V.; *Schmidt*, R.: Stadtbücherei Gerlingen kooperiert mit der Hochschule der Medien Stuttgart. Bachelor-Studiengang für Bibliotheksmanagement etabliert sich mit Projektarbeit 301
Gloystein, H.: s. *Effenberger*, C. 83
Götzel, A. : s. *Lüstorff*, J. 207
Gossen, M.: „Projekt Messe Leipzig 2007" – ein Rückblick 226
- : Das Projekt Messe Leipzig 2007. Studierende stellen ihr Department vor 95
Grosch, K.: s. *Finke*, S. 35
Hartmann, B.: s. *Effenberger*, C. 83
Hedden, H.: Book-style indexes for websites 433
Herb, U.: Open Access: Soziologische Aspekte 239
Herget, J.: s. *Bekavac*, B. 149
- : s. *Braschler*, M. 159
Hierl, S.: s. *Bekavac*, B. 149
Höchstötter, N.: Suchverhalten im Web – Erhebung, Analyse und Möglichkeiten 135
Höhfeld, St.; *Kwiatkowski*, M.: Empfehlungssysteme aus informationswissenschaftlicher Sicht – State of the Art 265
- :s. *Kwiatkowski*, M. 69
Hoffarth, M. : s. *Lüstorff*, J. 207
Holenstein, K.: s. *Friedrich*, A. 302
Hudson, A.: Training in indexing: the Society of Indexers' course 407
Huebenthal, I. : s. *Lüstorff*, J. 207
Huter, M.: Hype und Wirklichkeit an der Universität. „eUniversity – Update Bologna": Tagung zur Zukunft der Universitäten – Bonn, November 2006 38
Jantos, D.; *Pathmann*, A.: Aufbereitungs- und Interpretationsmethoden für lückenhafte Informationen 361
- : Die automatisierte und intelligente Competitive-Intelligence-Lösung osivo 341
Jörs, B.: Business Information Engineering – eine Anforderung auf dem Weg zum professionellen CI 333
Kells, K.: Indexing classes offered by the Graduate School (USDA) 410
Klaus, H.G.: Senior Experten Service – Eine Dienstleistung aus der „Silver Economy" 215
- : Senioren-Expertise-Netz (SENEX) der DGI nimmt Gestalt an 193

Ko, Y.M.; *Ratzek*, W.: Der Weg der Republik Korea vom Schwellenland zum Global Player: Science – Technology – Information 89
König, G.: Online Educa 2006: Beim 12. Mal die Teilnehmerzahl 2000 geknackt 119
Koschinsky, G.; *Ferber*, R.: Kommunikationswege beim E-Learning – eine empirische Untersuchung an der Hochschule Darmstadt 7
Kuhlen, R.: Vortrag vor dem Rechtsausschuss des Deutschen Bundestags am 20. November 2006 97
Kwiatkowski, M.; Höhfeld, St.: Thematisches Aufspüren von Web-Dokumenten – Eine kritische Betrachtung von Focused Crawling-Strategien 69
- : s. *Höhfeld*, M. 265
Lande, D.; *Braichevski*, S.; *Busch*, D.: Informationsflüsse im Internet 277
Lewandowski, D.: Editorial 129
- : Nachweis deutschsprachiger bibliotheks- und informationswissenschaftlicher Aufsätze in Google Scholar 165
Lüstorff, J.; *Böth*, N.; *Brückner*, C.; *Diehl*, S.; *Ender*, J.; *Götzel*, A.; *Hoffarth*, I.; *Huebenthal*, I.; *Mokline*, J.; *Neumann*, Ch.; *Rudolph*, D.; *Sprenger*, K.; *Stein*, J.; *Stricker*, A.; *Unger*, T.; *Weißenborn*, I.; *Wissel*, V.; *Yaqoubi*, E.: Informationswirtinnen und Informationswirte im Beruf. Ergebnisse einer Befragung der Alumni des Darmstädter Fachbereichs Informations- und Wissensmanagement (IuW) 207
MacGlashan, M.: The Indexer: past, present and future 402
Maislin, S.: Cyborg indexing: Half-human, half-machine 399
Marie, J. St.: Medical indexing in the United States 421
Markof, W.: Wirtschaftsinformationen in der Elektronischen Bibliothek – der Ansatz von 3M 285
Meier, B.; *Otto*, Ch.: Informationsspezialisten made in Darmstadt: Der neue Studiengang „Information Science and Engineering/Informationswissenschaft“ 15
Michaeli, R.: Competitive Intelligence als Lehrveranstaltung an Hochschulen 337
-: Editorial: Schwerpunkt Competitive Intelligence 321
Michel, K.-U.: s. *Reinhold,* D. 327
Mokline, J. : s. *Lüstorff*, J. 207
Münch, V.: CeBIT 2007: Fachinformation findet zu wenig Interesse. Sonderschau Infoltelligence lockt kaum Besucher an. EU vergibt European ICT Preise an Retrievalprodukte 228
- : Förderpolitik sorgt für 30 Jahre bewegtes Leben. FIZ Karlsruhe feiert Ende Oktober seinen 30ten Geburtstag mit einem Festakt im Schloss Bruchsal 459
- : Peter Genth gibt die Zügel aus der Hand. Der letzte große Pionier verlässt die Bühne der Informationswirtschaft mit einer Veranstaltung zu ihrer Zukunft 305
- : Strategien zurück in die Zukunft. Bericht von der Kongressmesse Online Information 2006, London, 28.–30. November 113
Neumann, Ch. : s. *Lüstorff*, J. 207
Ockenfeld, M.: Open Innovation – die Informationswissenschaft zeigt sich quicklebendig 308
Öttl, S.: s. *Bekavac*, B. 149
Oßwald, A.: Bangalore Commitment: Workshop on Electronic Publishing and Open Access. Developing Country Perspective: Bangalore, Indien, 2. bis 3. November 2006 183
Pathmann, A.: s. *Jantos*, D. 361
Petrucela, N.: „Find out what works“. Auswahl und Anwendung von Usability-Tests für das Redesign eines Medienportals 29
Pfister, J.: s. *Braschler*, M. 159
Pförtner, C.: *Workshop* „Semantische Lösungen für die Informationsintegration und Informationssuche – Praxisbeispiele aus Industrie und Hochschule“ 57
Pieper, D.; *Wolf*, S.: BASE – Eine Suchmaschine für OAI-Quellen und wissenschaftliche Webseiten 179
Pott, B.: Gestaltung von Blended-Learning-Angeboten für fachlich heterogene Gruppen. Erfahrungen mit einem Kurs für Berufstätige mit Hochschulabschluss 287
Ratzek, W.: Nationaler IT-Gipfel im Hasso-Plattner-Institut Potsdam. IT-Branche fordert mehr IT für alle 107
- :s. Ko, Y.M. 89
Rausch von Traubenberg, E.: s. *Aschoff*, F. 141
Reichmann, G.: Evaluation Informationswissenschaftlicher Lehrveranstaltungen: Eine Längsschnittuntersuchung 21
Reinhold; D.; *Michel*, K.-U.: Wissensmanagement vs. Competitive Intelligence. Die Notwendigkeit einer Symbiose beider Disziplinen für die Schaffung von strategischen Wettbewerbsvorteilen 327
Riecke, W.: s. *Fank*, W. 355
Robertson, M.: s. *Diepeveen*, C. 413
Rösch, H.: Entwicklungsstand und Qualitätsmanagement digitaler Auskunft in Bibliotheken 197
Rooney, P.: How I reused my own index 394
Rudolph, D. : s. *Lüstorff*, J. 207
Schaller, S.: s. *Friedrich*, A. 302
Schäuble, P.: s. *Braschler*, M. 159
Schmidt, R.: s. *Geske*, V. 301
Scholz, S.W.: s. *Wagner*, R. 347
Schröder, A.: Erfahrungen als LIS-Studentin in London 27
Schumacher, M.: Content is King – Content Management in Fachverlagen per Online-Software 222
Sprenger, K. : s. *Lüstorff*, J. 207
Stein, J. : s. *Lüstorff*, J. 207
Steinbach, M.: s. *Braschler*, M. 159
Stock, W.G.: Themenentdeckung und -verfolgung und ihr Einsatz bei Informationsdiensten für Nachrichten 41
Streib, S.: s. *Effenberger*, C. 83
Stricker, A. : s. *Lüstorff*, J. 207
Stucker, J.: s. *Braschler*, M. 159
Unger, T. : s. *Lüstorff*, J. 207
Wagner, R.; *Scholz*, S.W.: CI im Einsatz: Technologien zur Informationssuche, -bewertung und -aufbereitung 347
Weber, A.: „HeilFASTen“ – Entschlacken mit Leistungsgewinn. Neue Möglichkeiten für Bibliothekskataloge durch den Einsatz von Suchmaschinentechnologie 225
Weichselgartner, E.; *Baier*, Ch.: Sechs Jahre PsychSpider: Aus der Praxis des Betriebs einer Psychologie-Suchmaschine für freie Web-Inhalte 173
Weisel, L.: Eine ungehaltenen Rede auf die ausgeschlagene Ehrung für Professor Dr. Robert Funk 105
Weißenborn, I. : s. *Lüstorff*, J. 207
Wissel, V. : s. *Lüstorff*, J. 207
Wolf, S.: s. *Pieper*, S. 179
Wuttke, H.D.: s. *Finke*, S. 35
Yaqoubi, E. : s. *Lüstorff*, J. 207
Zimmermann, T.: FaMI-Portal.de – eine Communityplattform für Fachangestellte für Medien- und Informationsdienste 300

Personalien

Helmut Arntz, erster Präsident der DGD/DGI † (*Dahlberg*, I.) 313
Dr. Peter Budinger wird 70 Jahre (*Ockenfeld*, M.) 247
Ursula Deriu übernimmt Leitung von FIZ Technik (*Ockenfeld*, M.) 313
Werner Flach gestorben (*Ockenfeld*, M.) 315
Kaula-Goldmedaille für Robert Fugmann (*Ockenfeld*, M.) 104
In Memoriam. Nachruf auf Ministerialrat a.D. Dr. Heinz Lechmann (*Henrichs*, N.) 381
Dr. Werner Rittberger 80 Jahre (*Ockenfeld*, M.) 247
In Memoriam Dr. Udo Schützsack (*Laux*, W.) 185

Informationen

24. Oberhofer Kolloquium – Call for Papers 412
97. Deutscher Bibliothekartag – Call for Papers 273
Collaborative Catalog Enrichment 311
Digital Market Place mit 25 Prozent mehr Fläche 260
E-Kurs zum Selbststudium: „Archiv und Film“ der HAW Hamburg (*Kübler*, H.-D.) 312
Ehrungen auf der DGI-Jahrestagung 2006 (*Bein*, A.) 106
Erfolgreiches Scheitern: Vergessen wir den Zweiten Korb – starten wir zum dritten durch (*Kuhlen*, R.) 286
Herbstlehrgang 2007 für Informationsassistenten (*Ockenfeld*, M.) 271, 339
Mehrwert Auslandserfahrung (*Lang*, U.) 245
Professionelle Registererstellung. Das DNI informierte auf der Buchmesse 2006 (*Fassbender*, J.) 6
Vom „point of information“ zum „point of communication“ (*Kaiser*, K.) 424
Der Wettbewerb schläft nie! 379

Nachrichten

63 Millionen redaktionell überprüfte Patentdokumente 262
140 Jahre Chemieforschung vollständig digitalisiert 458
AKEP Award 2007 für ‚StudentConsult' an Elsevier / Urban & Fischer Verlag 322
BMWi-Forschungsprojekt EXPLAIN geht in die Entwicklung 68
Broadcasting-Service für Austausch und Vermarktung von Bildungsinhalten 262
Buchhandlungen und Neue Medien – Studie zu Kundenwünschen 5
Deutsche Akademie für Sprache und Dichtung: Diskussion über „Die Sprache der Bahn“ 4
Deutsche Konkurrenz für Google Earth 322
DIMDI veröffentlich endgültige Fassung der ICD-10-GM Version 2007 5
DPMA: Änderungen Internationale Klassifikation von Waren und Dienstleistungen 4

Erfolgreiche Open Source-Software aus Deutschland – Alkacon OpenCMS 7 verfügbar 261
HdM-Studierende entwickeln B.I.T.-Wiki 248
Leitfaden zum Einsatz von Web 2.0 im Online-Handel 261
Multimedia-Inhalte in Factiva-Produkten gezielt recherchierbar 261
Neue Plattform Web-Information-Retrieval.de 386
ODOK'07: Informationskonzepte für die Zukunft 248
Die PATINFO verstärkt ihren internationalen Charakter 248
Referat Fachinformation aufgelöst – Personalkarussell im BMBF 4
Studie Datensicherheit von RFID-Systemen gratis erhältlich 263
VFI-Förderungspreise 2006 vergeben 68
ZPID-Monitor 2004: Ausführlicher Ergebnisbericht 5

Leserbrief

Bauer, G.: Das „elektronische Zeitalter der Chemieinformation ist eingeläutet worden" – und sonst nichts? 132

Rezensionen

Alby, T.: Web 2.0. Konzepte, Anwendungen, Technologien. München, Wien: Carl Hanser, 2007. 241 S. ISBN 3-446-40931-9 (*Ratzek*, W.) 187
Fugmann, R.: Das Buchregister. Methodische Grundlagen und praktische Anwendung. Frankfurt am Main: 2006. 136 S. (Reihe Informationswissenschaften der DGI. Bd.10) ISBN 10-3-92547-59-5 bzw. ISBN 13978-3-925474-59-0 (*Bertram,* J.) 186
Huber, S.: Neue Erlösmodelle für Zeitungsverlage. Boizenburg; vwh, 2007. 173 S. ISBN 978-3-9802643-9-6 (*Ratzek*, W.) 438
Klumpp, D.; Kubicek, H.; Rossnagel, A.; Schulz, W. (Hrsg.): Medien, Ordnung und Innovation. Berlin, Heidelberg, New York: Springer, 2006. 414 S. ISBN 3-540-29157-1 / 978-3-540-29157-2 (*Ratzek*, W.) 187
Krömer, J.; Sen, E.: No Copy – Die Welt der digitalen Raubkopie. Berlin: Tropen, 2006. 304 S. ISBN 3-932170-82-2-15 (*Weber*, S.) 250
Lowe, R.; Marriott, S.: Enterprise: Entrepreneurship and Innovation. Concepts, Contexts and Commercialization. Amsterdam, Boston, London: Butterworth-Heinemann, 2006. 444 S. ISBN 0-7506-6920-9 (*Ratzek*, W.) 249
Mujan, D.: Informationsmanagement in Lernenden Organisationen. Berlin: Logos Verlag, 2006. 181 S. ISBN 978-3-8325-1339-9 (*Weber*, S.) 316
Park, H.-S.: Pons Lernen und Üben Koreanisch. Der direkte Weg zur Sprache mit Audio-CD. Stuttgart u.a.: Ernst Klett Sprachen GmbH, 2006. ISBN 3-12-560743-4, ISBN 13978-12-560743-9 (*Menger*, M.-E.) 299
Pöhm, M.: Präsentieren Sie noch oder faszinieren Sie schon? Der Irrtum PowerPoint. Heidelberg: mgv Verlag Redline, 2006. 286 S ISBN 978-3-636-06265-9 (*Ratzek*, W.) 252
Radtke, A.; Charlier, M.: Barrierefreies Webdesign. Attraktive Websites zugänglich gestalten. München: Addison-Wesley, 2006. 252 S. ISBN 978-8273-2379-8, 3-8273-2379-7 (*Schweibenz*, W.) 318
Redmond-Neal, A.; Hlava, M.M.K. (Hrsg.): ASIS&T Thesaurus of Information Science, Technology, and Librarianship. Medford, New Jersey: Information Today. XIII, 255 S. ISBN 1-57387-243-1 und ISBN 1-57387-244-X (*Fassbender*, J.) 59
Spitta, T.: Informationswirtschaft. Eine Einführung. Berlin, Heidelberg, New York: Springer, 2006. 169 S. ISBN 978-3-540-29635-5 (*Ratzek*, W.) 382
Stock, W.G.: Information Retrieval –Informationen suchen und finden. München: Oldenbourg Verlag, 2007. IX,600 S. ISBN 3-486-58172-4 (*Ferber*, R.) 318
Thom, N.; Ritz, A.: Public Management. Innovative Konzepte zur Führung im öffentlichen Sektor. 3., überarb. U. erw. Aufl. Wiesbaden: Gabler, 2006. 451 S. ISBN 3-8349-0096-6 (*Ratzek*, W.) 249
Weilenmann, A.-K.: Fachspezifische Internetrecherche. Für Bibliothekare, Informationsspezialisten und Wissenschaftler. München: K.G. Saur, 2006. 205 S. (= Bibliothekspraxis Bd. 38) ISBN 3-589-11723-X (*Katzmayr*, M.) 317

Gegründet von H.-K. Soeken †
unter dem Titel Nachrichten für Dokumentation (NfD)
Herausgegeben von der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis e.V. (DGI)
Präsidentin: Prof. Dr. Gabriele Beger
Hanauer Landstraße 151-153
D-60314 Frankfurt am Main
Telefon: (0 69) 43 03 13
Telefax: (0 69) 4 90 90 96
mail@dgi-info.de
www.dgi-info.de
Mitteilungsblatt des Normenausschusses Bibliotheks- und Dokumentationswesen im DIN Deutsches Institut für Normung e.V., der Fachgruppe Dokumentation im Deutschen Museumsbund und der Arbeitsgemeinschaft der Spezialbibliotheken (ASpB)

Redaktionsbeirat
Klaus-Peter Böttger, Mülheim an der Ruhr (Berufsfragen Information und Bibliothek)
Dr. Sabine Graumann, München (Informationswirtschaft)
Prof. Dr. Hans-Christoph Hobohm, Potsdam (Management von Informationseinrichtungen)
Prof. Dr. Rainer Kuhlen, Konstanz (Informationswissenschaft)
Dr. Dirk Lewandowski, Hamburg (Suchmaschinen, Internet)
Prof. Dr. Wolfgang Ratzek, Stuttgart (Informationspraxis)
Prof. Dr. Ralph Schmidt, Hamburg (Newcomer Report, Medien)

Redaktion
Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.
Marlies Ockenfeld (verantwortlich)
Viktoriaplatz 8, 64293 Darmstadt
Telefon: (0 61 51) 86 98 12
Telefax: (0 61 51) 86 97 85
ockenfeld@dgi-info.de
Daniel Ockenfeld
(Redaktionsassistenz)

Gastherausgeber
Jochen Fassbender

Verlag
Dinges & Frick GmbH
Greifstraße 4
65199 Wiesbaden
Postfach 1564
65005 Wiesbaden
Telefon: (06 11) 9 31 09 41
Telefax: (06 11) 9 31 09 43
Bankverbindung:
Wiesbadener Volksbank
BLZ 510 900 00, Kto-Nr. 714 22 26
Postbank Frankfurt
BLZ 500 100 60, Kto.-Nr. 267 204-606

Objektleitung
Erwin König,
koenig@dgi-info.de

Anzeigenservice
Ursula Hensel Anzeigenservice
Hermann-Schuster-Straße 39
65510 Hünstetten-Wallbach
Telefon: (0 61 26) 57 08 82
Telefax: (0 61 26) 58 16 47
ursula.hensel@t-online.de
Rocco Mischok
Verlag Dinges & Frick GmbH
Greifstraße 4
65199Wiesbaden
Telefon: (06 11) 3 96 99-60
Telefax: (06 11) 3 96 99-30
r.mischok@dinges-frick.de

Gestaltung
Anne Karg-Brandt, Hohenstein

Druck
Dinges & Frick GmbH
Greifstraße 4
65199 Wiesbaden
Postfach 2009
65010 Wiesbaden
Telefon: (06 11) 3 96 99-0
Telefax: (06 11) 3 96 99-30
Leonardo: (06 11) 93 20 79
Twist: (06 11) 9 10 23 78
df@dinges-frick.de

Erscheinungsweise/ Bezugspreise
Sieben Hefte jährlich (Doppelausgabe September/Oktober)
Jahresabonnement EUR 169,–
Schüler/Studenten EUR 125,–
Einzelheft EUR 30,–
inkl. Versandkosten/Porto.
Das Abonnement gilt für mindestens ein Jahr und kann danach bis sechs Wochen zum Ende des Bezugszeitraums gekündigt werden.

Redaktionsschluss für
Heft 1/2008 1. Dezember 2007
Heft 2/2008 1. Februar 2008
Heft 3/2008 10. März 2008

2008

21. bis 23. Januar *Berlin*	APE 2008 Academic Publishing in Europe Quality & Publishing	Arnoud *de Kemp*, E-Mail:info@ape2008.eu, www.ape2008.eu
28. bis 30. Januar *Zadar, Kroatien*	BOBCATSSS 2008 – Providing Access to Information for Everyone	BOBCATSSS 2008, Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Postfach 60 06 08, 14406 Potsdam, info@bobcatsss2008.org, www.bobcatsss2008.org
29. bis 31. Januar *Karlsruhe*	LEARNTEC 2008: Lernen 2.0 – Wege in die Zukunft – 16. Internationaler Kongress und Fachmesse für Bildungs- und Informationstechnologie	Frank *Pflugfelder*, Karlsruher Messe- und Kongress-GmbH (KMK), Festplatz 9, 76137 Karlsruhe, Telefon: (07 21) 3720-5145, frank.pflugfelder@kmkg.de, www.learntec.de
20. bis 22. Februar *Konstanz*	11. Deutsche ISKO-Konferenz „Wissensspeicher in digitalen Räumen: Nachhaltigkeit, Verfügbarkeit, semantische Interoperabilität“	Jörn *Sieglerschmidt*, joern.sieglerschmidt@uni-konstanz.de, www.bonn.iz-soz.de/wiss-org/
4. bis 9. März *Hannover*	CeBIT 2008	Kontakt: Deutsche Messe AG, Hannover, Messegelände, 30521 Hannover, Tel.: (0511) 89-0, Fax: (0511) 89-36694, www.messe.de
9. bis 11. April *Würzburg*	10. inetbib-Tagung Inetbib 2.0	Michael *Schaarwächter*, E-Mail: michael.schaarwaechter@ub.uni-dortmund.de, www.inetbib.de
10. bis 12. April *Magdeburg*	24. Oberhofer Kolloquium zur Praxis der Informationsvermittlung – Informationskompetenz 2.0. Zukunft von qualifizierter Informationsvermittlung	Silvia *Otterbein*, Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V., Hanauer Landstraße 151-153, 60314 Frankfurt am Main, Telefon: (069) 43 03 13, Fax: (069) 490 90 96, otterbein@dgi-info.de
3, bis 6. Juni *Mannheim*	97. Deutscher Bibliothekartag Wissen bewegen. Bibliotheken in der Informationsgesellschaft	Per *Knudsen*, Universitätsbibliothek Mannheim, Ortskomitee für den 97. Bibliothekartag, Schloss, Ostflügel, 68131 Mannheim, bibliothekartag2008@bib.uni-mannheim.de, www.bibliothekartag2008.de
29. Mai bis 11. Juni *Düsseldorf*	print media messe – drupa	Messe Düsseldorf GmbH, Postfach 10 10 06, 40001 Düsseldorf, Telefon: (02 11) 45 60-01, Fax: (02 11) 45 60-668, www.drupa.de
20. bis 24. Juli *Singapur*	SIGIR ‘08 31st Annual International ACM SIGIR Conference	International Conference Services BV, PO Box 83005, 1080 AA Amsterdam, The Netherlands, Tel. +31 20 6793218, Fax +31 20 6758236, sigir2007-reg@ics-online.nl, www.sigir2007.org
4. bis 7. August *Shanghai, China*	XVIII World Congress of the International Federation of Translators: Translation and Cultural Diversity	2008 FIT World Congress Secretariat, c/o Translators Association of China, 24 Baiwanzhuang Street, Xicheng District, Beijing 100037, China, Fax: +86 10 6899 0247, fit2008papers@gmail.com
10. bis 15. August *Québec, Kanada*	World Library & Information Congress – 74rd IFLA General Conference & Council Libraries without borders: Navigating towards global understanding	WLIC Conference Secretariat, Congrex Holland BV, Telefon +31 20 5040 201, Fax: +31 20 5040 225, E-Mail: wlic2007@congrex.nl, www.ifla.org
29. Aug. bis 3. Sept. *Berlin*	Internationale Funkausstellung	Messe Berlin GmbH, Messedamm 22, 14055 Berlin, Telefon: (030) 3038-0, Fax: (030) 3038-2325, ifa@messe-berlin.de, www.messe-berlin.de
3. bis 5. September *Graz, Österreich*	Triple I: I-Know, I-.Media, I-Semantic	Mag. Anita *Wutte*, Know-Center GmbH, Inffeldgasse 21a, A-8010 Graz, Telefon: +43 316 873 9251, Fax: +43 316 873 9254, awutte@know-center.at, www.triple-i.info
15. bis 18. September *Erfurt*	78. Deutscher Archivtag Bestandserhaltung analoger und digitaler Unterlagen	Thilo *Bauer* M.A., VdA – Verband deutscher Archivarinnen und Archivare e. V., – Geschäftsstelle –, Wörthstraße 3, 36037 Fulda, Telefon: (06 61) 29 109 72, Fax: (06 61) 29 109 74, info@vda.archiv.net, www.archivtag.de/
15. bis 19. Oktober *Frankfurt am Main*	Frankfurter Buchmesse Ehrengast Türkei	Dr. Juergen *Boos*, Ausstellungs- und Messe GmbH, Buchmesse Frankfurt, Reineckstraße 3, 60313 Frankfurt am Main, Telefon: (069) 2102-0, Fax: (069) 2102-227/-277, info@book-fair.com
21. bis 24. Oktober *München*	SYSTEMS 2008	Messe München GmbH, Messegelände, 81823 München, Telefon: (089) 949117-18, Fax: (089) 94 9117-19, info@systems.de, www.systems.de
21. bis 25. Oktober *Köln*	ORGATEC Office & Object	Koelnmesse GmbH, Messeplatz 1, 50679 Köln, Telefon: (02 21) 821-0, mailto:info@orgatec.de, www.orgatec.de
2. bis 4. Dezember *London, England*	Online Information 2008	VNU Exhibitions Europe, 32-34 Broadwick Street, London, W1A 2HG, UK, lorna.candy@vnuexhibitions.co.uk, www.online-information.co.uk/

Missing Link
Westerstrasse 114-116 | 28199 Bremen
Tel.: [0421] 504348 | Fax: [0421] 504316

Erwerbungspartner, mit denen Sie rechnen können

www.missing-link.de
info@missing-link.de

Internationale Versandbuchhandlung