

# Zukunftsorientierte Unternehmen bauen auf Wissen

## 3. DGI-Praxistage 2013 „Viel Information ist nicht genug: semantische Technologien in wissensbasierten Unternehmen“

Marlies Ockenfeld

» Zum dritten Mal hat die DGI am 25. und 26. April in Frankfurt am Main Praxistage durchgeführt, ein Treffen, bei dem sich Vertreter der Informationsbranche gegenseitig über die tagtäglichen Herausforderungen im betrieblichen Informationswesen und vermutete Trends informieren sollen. Wer die etwa siebzig Teilnehmerinnen und Teilnehmer an der diesjährigen Tagung bei Frühlingswetter im kommunikationsfreundlichen Gästehaus der Goethe Univer-

zu erzeugen. Informationstechnische und organisatorische Anforderungen jenseits der konkreten fachlichen Inhalte bestimmen zunehmend den Berufsalltag und müssen damit auch stärkeren Eingang in die Ausbildungsgänge angehender Informationsfachleute finden. Diesen Bedarf äußerten nicht nur Studenten der Informationswissenschaft, die sich mehr informationstechnische Inhalte in ihrem Studienangebot wünschen, sondern er wurde auch im Vortrag von Ulrich Wieneke, Head of Scientific Database Services bei Roche Diagnostics, über den Einsatz semantischer Technologien in einem pharmazeutischen Unternehmen bestätigt: Eingestellt hat er in seiner Abteilung zuletzt einen Bioinformatiker, weil dieser sowohl Terminologie und Semantik des Fachs versteht als auch informationstechnische Zusammenhänge erkennt und umsetzen kann.

Stefan Zillich, freiberuflicher Information Professional, erläuterte das von einer Arbeitsgruppe der Gesellschaft für Wissensmanagement (GfWM) erarbeitete Diskussionspapier zur Wissensarbeit in modernen Arbeitswelten. Wissensarbeit umfasst Tätigkeiten, für die das bisher erworbene Wissen nicht ausreicht. Sie wird zunehmend durch freiberuflich Tätige mit hoher Fach-, Methoden- und Sozialkompetenz sowie kreativem Potential erbracht, die nicht in starre Strukturen eingebunden sind. Unternehmer, so die Empfehlung, müssen ihre Wissensarbei-

ter wertschätzen und ihnen Raum und Gelegenheit zur Generierung neuen Wissens geben. Das Diskussionspapier ist unter [www.gfwm.de/node/8](http://www.gfwm.de/node/8) verfügbar.

Die Beachtung von Urheberrecht und Copyright stellen viele Unternehmen bei der Informationsbeschaffung und -verwendung vor große Herausforderungen. Jörg Weizenboerfer von der 2010 in Amsterdam gegründeten Rights Direct wies auf ein diesbezügliches Schulungsvideo auf der Firmenwebsite hin. Rights Direct vermittelt umfassende Lizenzrechte von weltweit über 400.000 Rechteinhabern auch für geographisch verteilte Unternehmen und ihre Mitarbeiter.

Dass Effizienz und Qualität bei professioneller Übersetzungsarbeit kein Widerspruch sein müssen, zeigte Clemens Weins von der Cocomore AG. Das EU-Projekt MultilingualWebLT hat die Open-Source-Technologie Drupal ausgewählt, um den vom W3C erarbeiteten Standard IST 2.0 (Internationalization Tag Set (ITS) Version 2.0) zu implementieren. IST 2.0 ([www.w3.org/TR/its20](http://www.w3.org/TR/its20)) beschreibt, wie Texte semantisch ausgezeichnet werden sollen, damit sie sich für eine effiziente Kombination von menschlicher und maschineller Übersetzung eignen. Erprobt wird das System derzeit mit dem VDMA für die verschiedenen Sprachversionen der Website. Wer Drupal selbst nicht einsetzt, kann die Cocomore Translation Cloud nutzen.

Axel Ermert und Monika Hagedorn-



*Mitglieder der Jungen DGI und erfahrene Kollegen kamen bei den DGI-Praxistagen im Gästehaus der Goethe Universität ins Gespräch.*

sität und seinem wunderschönen Außengelände beobachtete, konnte feststellen, dass dieses Konzept offensichtlich angenommen wird und ein angeregter Austausch stattfand. An Informationen besteht kein Mangel, die Bewältigung der Publikationsflut war von Anfang an Triebfeder der Dokumentationsbewegung und von Informationsdiensten. Heute versprechen die Protagonisten so genannter semantischer Technologien neue Wege, um in automatisierter Weise nützliche Informationsströme

Saupe vom Institut für Museumsforschung in Berlin informierten über die derzeitige Erarbeitung einer deutschen Fassung des polyhierarchischen Art and Architecture Thesaurus, der sich aufgrund seiner wissenschaftlichen Genauigkeit und seines breiten Themenspektrums für das Wissensmanagement im Kulturbereich anbietet. Jeder der etwa 35.000 Begriffseinträge im Thesaurus ist mit einer eindeutigen ID versehen. Dadurch wird die direkte, automatische Verknüpfung mit anderen Wortbeständen, also die Vernetzung und Nutzung von „Linked Data“ im Web möglich. Einzelheiten über das Projekt, für das derzeit noch Übersetzer gesucht werden, finden sich unter [www.aat-deutsch.de](http://www.aat-deutsch.de).

Ulrich Wienecke stellte die Wiederentdeckung der Industriesemantik dar. Von 15 Pharmafirmen setzen derzeit nur noch drei einen Schlagwortkatalog ein, um ihre Buchbestände zu erschließen. Doch in anderen Firmen-Bereichen erfahren kontrollierte Vokabular und Terminologien eine Renaissance. Glossare, Firmenvokabulare, Abkürzungsverzeichnisse, Produktsynonyme sind nicht nur für neue Mitarbeiter, sondern auch für die abteilungs- oder standortübergreifende Zusammenarbeit bedeutsam. Ein großes Desiderat zur Konkurrenzbeobachtung ist ein umfassender Firmenthesaurus, denn käufliche Thesauri enthalten nur die 5000 weltweit größten Unternehmen, nicht aber kleinere Spezialfirmen.

Auf überzeugende Weise zeigte Andreas Weber von der Spielwarenfabrik Schleich, wie der konsequente Einsatz eines semantischen Produktdatenmanagements viele unterschiedlichen Informationsanforderungen erfüllen kann. Schleich erneuert jährlich 20 Prozent seines Sortiments, liefert in über fünfzig Märkte, produziert an vielen verschiedenen Standorten weltweit und muss die gesundheitliche Unbedenklichkeit



*Ulrich Wienecke warnte vor der Buzzword-Falle beim Versuch semantische Ansätze im Unternehmen umzusetzen: Informationsfachleute sollten mit Produkten, nicht mit Schlagwörtern zu überzeugen versuchen, denn allzu oft erfolgt nach einem Hype der Absturz und man ist unversehens mit einem verbrannten Thema identifiziert.*

jedes einzelnen Spielzeugs jederzeit belegen können. Gemeinsam mit der Darmstädter Firma i-views wurde ein skalierbares System für die Verwaltung der Produktdaten entwickelt, das firmenübergreifend an allen Standorten eingesetzt wird, einfach zu verstehen und zu benutzen ist und alle Daten für die Beantwortung von Fragen der Kunden und Behörden liefert. Dazu wurde ein übergreifendes Datenmodell für die Spielzeugbranche entwickelt, das u.a. Produkte, Baugruppen, Substanzen, Rohmaterialien, Preise, Kosten, Gesetzestexte, Richtlinien, Prüfberichte umfasst. Das implementierte semantische Netz wird als Produktinformationssystem genutzt, das mit einem graphischen Navigator befragt werden kann. Die Antwort auf eine beliebige Frage ist immer ein ganz bestimmter Pfad durch dieses Netz. Vorteile sind die Erleichterung des internen Qualitätsmanagements, ein schlankes Prüfwesen und ein hohes Maß an Transparenz hinsichtlich Prozessen und Inhaltsstoffen. In seinem Beitrag „Semantic Publishing – Coping with the Overwhelm“ beklagte Jan Velterop von AQnowledge, GB die herrschende „Datarhoea“ und bemühte einmal mehr die Tatsache, dass niemand alles le-

sen könne, was tagtäglich erscheint – aber wer will und muss das schon? Einen herkömmlichen Ausweg bot die Dokumentation insbesondere in Form von Tertiärpublikationen wie Handbüchern (man denke an Gmelin und Beilstein) oder Enzyklopädiën. Die technologisch weiter entwickelte Form bezeichnet Velterop als „Nanopublications“. Hierzu soll der Autor – oder eine inhaltserschließende Instanz an seiner Seite – neu publizierte Daten und Fakten unmittelbar zum Zeitpunkt der Veröffentli-



*Einig schien man sich beim Kamingespräch zunächst darüber, dass sich Wissen nur „zwischen den Ohren“ manifestiert und dass sich außerhalb des eigenen Kopfs lediglich Daten oder Informationen austauschen und verwalten lassen. Doch gegen Ende kam die Frage auf, wie man das beurteilen und benennen sollte, was durch Data Mining, maschinelles Lernen oder automatisches Schließen in Informationssystemen erzeugt und verarbeitet wird. (v.l. Michael Dreusicke, Dr. Stefan Rehm, M. Günther Szogs, Prof. Sissi Closs, Ulrich Schmidt, Stefan Zillich)*

Zwei interne Veranstaltungen der DGI fanden am 25. April statt. Die „Junge DGI“ traf sich erstmals im Rhein-Main-Gebiet, nachmittags trat die diesjährige DGI-Mitgliederversammlung zusammen. Wesentliche Tagesordnungspunkte waren die Verabschiedung des Wirtschaftsplans für 2013 und die Namensänderung der Fachgesellschaft. Nur einzelne Stimmen, darunter die Berichterstatterin, rieten von einer Namensänderung auf dieser mit gut zwanzig Mitgliedern nur schwach besetzten Versammlung ab und plädierten zunächst für eine breitere Meinungsbildung in der Mitgliedschaft und die Sondierung möglicher Kooperationen im Vorfeld der Entscheidung. Andere, neben dem Vorstand vor allem Willi Bredemeier, sprachen sich eindringlich für die sofortige Änderung in einen griffigen Namen aus, der auch die Außendarstellung und Ansprache neuer Gruppen erleichtert. Die Mitgliederversammlung beschloss nach der Aussprache einstimmig, dass der Name der DGI unter Beibehaltung des Akronyms DGI in „Deutsche Gesellschaft für Information und Wissen“ geändert werden soll.



*Mit einem großen Dank und herzlichem lang anhaltenden Applaus verabschiedete die Mitgliederversammlung Nadja Strein, die vier Jahre lang, zuletzt als Geschäftsführerin, die DGI-Geschäftsstelle geleitet hat und stets als sachkompetente und freundliche Ansprechpartnerin zur Verfügung stand. Sie hatte am 1. April 2013 eine neue Arbeitsstelle in Frankfurt am Main angetreten.*

chung dokumentieren, indem sie als RDF-Triplett formuliert und mit Metadaten angereicht werden. Der Ansatz erinnert an die Bemühungen in den 1970/80er-Jahren, Autoren dazu zu bewegen „dokumentationsgerecht“ zu publizieren, sich also einer standardisierten Sprache ohne Synonyme und Nominalphrasen, kurzen Sätzen, die jeweils nur eine Aussage enthalten, etc., zu bedienen. Eher banal wirkte das gewählte Beispiel, wie aus strukturierten Angaben in Wikipedia automatisiert Nanopublikationen erzeugt werden könnten, denn es war eher dem Alltagswissen zuzuordnen (Einwohnerzahl von Tokio), als dass es zu hochwertigen wissenschaftlich-technischen Informationen gehörte, mit denen Informationsfachleute meistens zu tun haben. Und selbst die in Wikipedia genannte Einwohnerzahl von Tokio erweist sich als problematisches Datum, denn sie ist keine statische Größe, sondern lediglich eine gerundete Momentangabe zu einem bestimmten Zeitpunkt. Was macht man, wenn man auf seine Frage nach der Einwohnerzahl von Tokio hundert Nanopublikationen mit differierenden Angaben findet?

Mit dieser Informations-Qualität im Web der Daten befasst sich ab-

schließend Sören Auer von der Universität Leipzig, einer der führenden Köpfe in der Linked Open Data (LOD) Community. Im von der EU geförderten Integrationsprojekt LOD2 werden von 2010 bis 2014 Werkzeuge zur Erzeugung und Nutzung von Linked Data entwickelt und erprobt, etwa zur Extraktion von RDF-Triplets aus Texten, XML und SQL. Auer betonte, dass es nicht auf die absolute Datenqualität ankommt, sondern stets die Nutzbarkeit für einen ganz konkreten Anwendungsfall beurteilt werden muss. So reichen automatisch extrahierte Wikipedia-Daten in der Regel aus, um etwa Webportale mit Informationen über bekannte Persönlichkeiten und ihre Werke anzureichern, nicht aber für medizinische Ratgeber. Beim Projekt DBpedia scheitert die automatische Extraktion von Daten aus Wikipedia oft schon an der unterschiedlichen Bezeichnung der Kategorien in der Wikipedia. Daher eignet sich DBpedia wegen ihrer Inkonsistenzen auch nicht für maschinelles Schließen. Viele Freiwillige sind dabei, Konkordanzlisten zu erzeugen, um die Qualität zu verbessern (<http://mappings.dbpedia.org>). Einen stets aktuellen Überblick über die im Web bereits verfügbaren RDF-Datensets

erhält man auf der Website [stats.lod2.eu](http://stats.lod2.eu). Gute Autorenumgebungen im Sinne von streng reglementierten Eingabemasken fördern die Informationsqualität. Das in LOD2 entwickelte OntoWiki, ein semantisches Wiki zur Eingabe von Daten in hoher Qualität, wird bei Daimler erprobt. Doch fehlen vielen Fällen noch gute Explorationswerkzeuge. Es gibt also auf einem neuen technischen Niveau für Informationsfachleute erneut viel zu tun, um altbekannte Probleme zu lösen.

DGI-Präsident Stefan Gradmann griff in seinem Schlusswort die Frage auf, wie viel „Informationsschmutz“ Informationsfachleute zulassen dürfen oder müssen und kündigte an, dass dies eines der Themen für die nächste DGI-Konferenz im Frühjahr 2014 sein werde. Zunächst aber steht vom 17. bis 19. Oktober das DGI-Forum Wittenberg 2013 als Nachfolgeveranstaltung des Oberhofer Kolloquiums auf der Agenda. ■

#### ..... **Marlies Ockenfeld**

DGI  
Windmühlstraße 3  
60329 Frankfurt am Main  
[mail@dgi-info.de](mailto:mail@dgi-info.de)  
[www.dgi-info.de](http://www.dgi-info.de)  
.....