

Content im Kontext

Gab es auf der Infobase und den beiden parallel stattfindenden DGI-Tagungen im Mai 2001 in Frankfurt Neuigkeiten in Informationswissenschaft und -praxis? Der Trend, die grundlegende Innovation konnte nicht gefunden werden, wohl aber - Mosaiksteinen gleich - diverse interessante Details, die insgesamt, richtig zusammengesetzt, ein durchaus positives Bild der Branche zeichnen. Erstmals veranstaltete die DGI die Online-Tagung gemeinsam mit ihrer Jahrestagung. Einem breiten Angebot an Vorträgen und Messepräsentationen stand ein nur laues Besucherinteresse gegenüber, leere Gänge bei den Ausstellern und wegen Besuchermangels ausgefallene Vorträge trübten die Stimmung erheblich. Dieser Messe- und Tagungsbericht informiert - in subjektiver Auswahl - über wichtige Baustellen (am Medienbruch wird gearbeitet), alten Wein in neuen Schläuchen (Content Syndication), aktuelle Entwicklungen bei Wirtschaftsinformationen (mit Bezug auf das Jahr 1844), über Zitationsindexierung (aber nicht nur vom ISI), automatisches Indexieren (neue Projekte und Evaluation von Software) und über unsere Einsamkeit.

Auf der Baustelle: Brücken zwischen Verlagen und elektronischen Informationsdiensten

Kommt der lang ersehnte Abschied vom Medienbruch? Wachsen die Welten der bibliographischen Informationsdienste und der Volltexte im Originallayout endlich zusammen? Technisch ist dies schon lange kein Problem mehr, kann man doch entweder die Volltexte auf den Rechner des Online-Archivs laden oder vom bibliographischen Datensatz über einen Link zum Server des Verlages gelangen, wo der Text bereits elektronisch vorgehalten wird. Das große Problem ist der Preis, wohlgemerkt: der Preis eines einzigen Artikels einer Zeitschrift. Eine der Brückenbaustellen ist **Econ-**

Doc (**E**conomics **o**nline **D**ocuments). Die deutschen wirtschaftswissenschaftlichen Datenbanken sollen mit dem Online-Angebot für elektronische Volltexte verknüpft werden. Die Preisgestaltung ist in den Händen der Verlage, so dass je nach Verlag unterschiedliche Kosten pro Volltext anfallen. Ist mit diesem 2-Millionen-Mark-Projekt des BMBF ein Durchbruch zu erwarten?

Wirtschaftsinformationen: Wirtschaftswissenschaft ab 1844 - Wirtschaftsportal von heute

Monika Zarnitz berichtet über aktuelle Entwicklungen bei der Deutschen Zentralbibliothek der Wirtschaftswissenschaften (ZBW) in Kiel. Imagefördernd ist sicherlich, dass diese Bibliothek sich ein Leitbild gegeben hat ("Wir stehen für umfassende Informationsversorgung in den Wirtschaftswissenschaften durch kundenorientierte Dienstleistungen. Kompetenz, Innovation und Kooperation sind die Grundlagen unserer Arbeit"), für die Kunden wichtiger sind die Aktivitäten im Umkreis von EconDoc, DigiZeit und EconisPlus. Im **DigiZeit**-Projekt geht es darum, wichtige Fachzeitschriften jeweils ab Band 1 durchgehend online zur Verfügung zu stellen. Die ZBW ist zunächst mit fünf Zeitschriften dabei. Das älteste dieser Periodika, die "Zeitschrift für die gesamte Staatswissenschaft / Journal of Institutional and Theoretical Economics", reicht bis zum Jahr 1844 zurück. Da ECONIS (noch) nicht retrospektiv bei 1844 angelangt ist, drängt sich die Frage auf, wann die Inhaltserschließung für diese alten "Perlen" erfolgen wird.

Als Teilprojekt der Virtuellen Fachbibliothek Wirtschaftswissenschaften indexieren die Projektteilnehmer, darunter neben der ZBW auch die Universitäts- und Stadtbibliothek Köln und das HWWA, elektronische Dokumente im World Wide Web und arbeiten diese bibliographischen Datensätze - natürlich mit Link versehen - in ihre Datenbanken ein. Die ZBW nennt diese Dienstleistung **EconisPlus**.

Die Trägerunternehmen des **FIZ Wirtschaft** arbeiten bei der Inhaltserschließung einheitlich mit dem "Standard-Thesaurus Wirtschaft", dem aktuell die noch fehlende Facette der Nachbarwissenschaften zugefügt worden ist. FIZ Wirtschaft ist ein Informationsverbund von GBI, HWWA, ifo und ZBW. Der Nutzer startet seine Recherche im Thesaurus, wobei er über die Hierarchien navigieren und dabei Terme "einsammeln" kann. Die ausgewählten Deskriptoren werden nicht - wie man vorschnell meinen könnte - in das CT-Feld der Datenbanken eingespielt, sondern ins Feld "überall". Da mit den Deskriptoren auch die jeweiligen Nicht-Deskriptoren als Synonyme mit übernommen werden, macht dies bedingt Sinn, können diese doch im Abstract oder im Titel auftauchen (und es gibt immer noch Datensätze, die keine Inhaltserschließung enthalten). Vor der Ausgabe läuft ein Relevance Ranking ab, das die Suchterme auch feldspezifisch gewichtet. Ein Eintrag im CT-Feld wiegt dabei höher als das Vorkommen etwa im Abstract. Via **EconDoc** gelangt der Nutzer im Endausbau des Portals direkt zu den Volltexten. Die FIZ Wirtschaft-Partner hängen an ihre bibliographischen Nachweise die Volltexte im ASCII- und PDF-Format. Im Laufe des Projektes werden rund 400 wirtschaftsrelevante Zeitschriften deutschsprachiger Verlage auf dem Server der GBI liegen. Die Serverlösung bei GBI wird gewählt, da viele (insbesondere kleine und mittelgroße) Verlage ihre Produkte noch gar nicht online vorhalten und damit auf die GBI-Hardware angewiesen sind. Bei Großverlagen mit ihren riesigen Lagern an Volltextdokumenten dürfte die Variante der Verlinkung auf den Verlags-server sinnvoller sein. **Sylvia Holzer** von der Münchner **GBI** betont: "Mit FIZ Wirtschaft ist sie in greifbare Nähe gerückt: die virtuelle Bibliothek, die Zeitschriftenbeiträge nicht nur differenziert und wortschatzsynchronisiert erschlossen hat, sondern sie auch direkt elektronisch auf den Schreibtisch liefert" - in der Tat, am Medienbruch wird gearbeitet. Hoffentlich ist er bald verschwunden.

Fremder Inhalt auf der Homepage: Content Syndication

Wenn ein Betreiber einer Homepage zwar gewisse Informationsinhalte (z.B. fachspezifische News) auf seiner Seite anbieten, aber nicht gleichzeitig eine Online-Redaktion beschäftigen möchte, kommt der Kauf fremder Inhalte in Betracht. In der Medienwirtschaft wird derart "Syndication" seit Jahrzehnten betrieben. "Syndication bezeichnet generell den mehrfachen Verkauf eines Medienproduktes an verschiedene Abnehmer zum Zwecke der ökonomischen Weiterverwendung", schreiben **Thomas Hess** und **Markus Anding** von der Universität Göttingen (S. 42). Gemäß Hess und Anding verläuft der Syndication-Prozess bei funktionaler Betrachtung von der Inhalteerzeugung (Akteure: Nachrichtenagenturen, Redaktionsbüros) über die Inhalteaggregation ("Aggregator") und die Inhaltspublikation (Communities, Internetportale, Intranets) zum Informationskonsum beim Rezipienten. Welchen Mehrwert schafft der Aggregator? Zunächst ist an die Übernahme der technischen Abwicklung der Bereitstellung der Inhalte zu denken, etwa an Formatangleichungen (z.B. PDF- oder HTML-Dokumente auf der Basis von XML weiterleiten). Für die Content Syndication besonders wichtig ist die themenspezifische Zusammenfassung der einzelnen Inhalte sowie die wiederum themenspezifische Publikation beim Kunden. "Publisher stehen dem Problem gegenüber, aus der Vielfalt von Inhalten verschiedener Produzenten die für ihre Zielgruppen relevanten Inhalte auszuwählen. ... Der durch einen Content Syndicator generierte Mehrwert kann in der Klassifikation von Inhalten verschiedener Produzenten entsprechend der Präferenzen der Publisher und deren Vorauswahl durch Bündelung entstehen" (S. 46). Unsere Autoren nennen als erste Fallbeispiele die Tanto AG in München sowie iSyndicate Europe.

Warum nennen sie nicht auch die Online-Archive? Diese Hosts bündeln doch seit jeher Inhalte unterschiedlicher Anbieter und dies mittels z.T. elaborierter Methoden der Inhaltserschließung. **GENIOS** spricht mit seinem Produkt "Content4Portals" gezielt den Markt des Content Syndication an. Derzeit bietet

GENIOS Standardthemen wie "Automobil, Transport und Verkehr", "Medien und Kommunikation" oder "EDV und Computer" aus diversen Quellen (darunter natürlich auch die verlagseigenen Zeitungen und Zeitschriften wie "Handelsblatt" und "Wirtschaftswoche") an. Eine individuelle Artikelauswahl ist möglich. "Einer der wichtigsten Faktoren für einen dauerhaft erfolgreichen Webauftritt ist - neben reinen Unternehmensinformationen und E-Commerce-Angeboten - ein auf ... Zielgruppen zugeschnittenes Informationsangebot. Die meisten Unternehmen haben weder Zeit noch Ressourcen, diese Inhalte selbst zu erstellen oder zu erschließen", schreibt **GENIOS**. Auch die **GBI** hat das neue Marktsegment entdeckt. "Intranet- und Portal-Anbindungen bieten ideale Voraussetzungen, um Benutzergruppen tagesaktuell, bedarfsgerecht und preiswert über Unternehmensnetze oder das Internet mit Informationsdienstleistungen zu versorgen", meint **Holger Lamprecht**, um sogleich zu betonen, dass **GBI** "zielgruppenorientierte Leistungsangebote" erarbeitet (S. 54). Was die Hosts hier bieten, ist so neu wohl nicht, neu ist allenfalls die Integration von aktuellen Datensätzen in die Websites ihrer Kunden.

Zitationsindexierung und die Visualisierung wissenschaftlicher Kommunikation

Zitationsindexierung meint die Abbildung von Informationsinhalten durch das Notieren von Fußnoten, die als Referenzen zu den vorgelagerten (zitierten) Texten und als Zitationen zu den nachgelagerten (zitierenden) Texten führen. Ein Quasimonopol auf diese Indexierungsmethode hat das Institute for Scientific Information (ISI) - bisher. Eine interessante Variante des Citation Indexing stellen **Gerhard M. Buurman** und **Stefan Roovers** von der Hochschule für Gestaltung und Kunst Zürich vor (S. 96), ohne übrigens einen Bezug zum ISI aufzuzeigen. Ausgangsmaterial von Buurman und Roovers ist der Volltext und seine Fußnoten, Ziel ist eine graphische Repräsentation der wissenschaftlichen Diskurse als sog. "Wissensterritorium". Ein Text wird im Anschluss an den Textwissenschaftler Teun A. van Dijk durch argu-

mentative "Superstrukturen" dargestellt; dies sind etwa Strukturelemente wissenschaftlicher Texte wie Problemstellung, Beobachtung, Hypothese, Experiment, Resultate, Diskussion usw. Die Verwaltung der Texte geschieht durch Strukturmarker, d.h. der Autor setzt bei der Textproduktion Markierungszeichen, die ein bestimmtes Strukturelement benennen. Die Fußnoten werden nunmehr nicht dem zitierenden Text als Ganzes zugeschrieben, sondern exakt den jeweiligen Textelementen, so dass detailliert verfolgt werden kann, welche Artikel ein Autor bei seiner Problemstellung oder bei der Diskussion seiner Ergebnisse zitiert. Kern des Projekts ist die Darstellung der Zitationen als "Landkarte"; die Visualisierung lässt Texte bzw. deren Strukturelemente zu Knoten und Zitationen zu Kanten zwischen den Knoten werden. Damit gelingt es, wissenschaftliche Kommunikation bildhaft zu repräsentieren. Buurman / Roovers betonen, "die Darstellung von Texten in ihren biographischen Zusammenhängen erhellt das Wissen über den Fortgang der wissenschaftlichen Arbeit, es zeigt Dokumente mit großer Außenwirkung und es zeigt gleichermaßen 'tote Gleise'. Durch die Strukturelemente werden darüber hinaus inhaltliche Verknüpfungen oder Relationen zwischen wissenschaftlichen Ideen deutlich" (HGKZ Dokumentation Forschungsprojekt "Wissensterritorien", S. 10). M.E. liegt hier ein wichtiger Beitrag aus der Disziplin des Informationsdesign zu Problematiken sowohl der Inhaltserschließung als auch der Wissenschaftsforschung vor.

Aber auch **ISI** hat Neues zu bieten. In Vorbereitung befindet sich das Produkt "Essential Science Indicators", das von **Nancy K. Bayers** vom ISI (Philadelphia) vorgestellt wird. Es handelt sich um hochaggregierte Indikatoren für 22 Wissenschaftsdisziplinen, die aus den ISI-Zitationsdatenbanken gewonnen werden und die der Wissenschaftsforschung und Wissenschaftsevaluation Dienste erweisen sollen. Zentrale Features des Produktes sind Rangordnungen nach Autoren, Artikeln, Institutionen, Zeitschriften und Ländern. Fragen wie "Welches sind die meist-zitierten Artikel in der Immunologie?" sind genauso zu beantworten wie "Welches Land hat die höchste wissenschaftliche Wirkung in der Chemie?" Gemäß Bayers kommen als Kunden dieses Web-Produktes wissenschaftliche Einrichtungen

gen, die über ihre Leistungen und Wirkungen Rechenschaft abgeben müssen, Institutionen der Wissenschaftsevaluation sowie Forschungsministerien infrage.

Automatische Indexierung und Thesaurus - Renaissance

Ein kontrolliertes Vokabular, das durch Relationen semantisch strukturiert ist, bringt einem Unternehmen enorme Vorteile: So weiss man, welches "seine" Welt ist, und jeder Mitarbeiter hat es leicht, sich über "seinen" Ausschnitt dieser Welt zu informieren. Die Erkenntnis, dass Dokumentationsmethoden wie Thesaurus und Klassifikation für Intranets bzw. Unternehmensportale Relevanz haben, hat inzwischen die Unternehmenspraxis erreicht. So berichten **Stefanie Wasen** von der Eutelis Consult in Ratingen und **Manfred Hauer** von AGI Information Management Consultants von der Entwicklung eines Thesaurus zur Telekommunikation, der die "Welt" von Eutelis abbildet und im internem Netzwerk bei der Inhaltsbeschreibung diverser Datenbanken (von Kunden und Mitarbeitern über Bibliotheksbestände, Presseauschnitte, Projekte bis zu Links im WWW) hilft. Wie kann ein Thesaurus automatisch generiert werden? Ein Baustein von Hauers Information Center (IC) ist IC CATE (Computer Aided Term Extractor), eine Variante von AUTINDEX des Instituts der Gesellschaft zur Förderung der Angewandten Informationsforschung an der Universität des Saarlandes (iai) von **Johann Haller**. Via AUTINDEX werden rund 700 Eutelis-Datensätze automatisch indexiert und die Deskriptorkandidaten anhand ihres Gesamtgewichtungswertes in eine Rangordnung sortiert. Die oben stehenden Terme zeigen die Kernthemen der Arbeit von Eutelis, sind aber für Indexierungen viel zu breit und daher allenfalls als Topterme zu gebrauchen, "die eigentliche Musik spielt darunter" (S. 102). Letztlich wird der Thesaurus - nach einem Sprachwechsel von deutsch nach englisch - doch intellektuell erstellt und in IC INDEX verwaltet.

AUTINDEX spielt auch eine Rolle beim derzeit laufenden Projekt BINDE, das **FIZ Technik** in Kooperation mit **IEE / INSPEC** und dem Saarbrückener **iai** durchführt. Es geht um die automatische Indexierung von Abstracts technischer Artikel in englischer und deutscher Sprache unter Zuhilfenahme eines bestehenden Thesaurus. Seit AIR/PHYS von Gerhard Lustig und Gerhard Knorz, das rund ein Jahrzehnt beim Input der Physik-Datenbank des FIZ Karlsruhe im Einsatz war, liegt nunmehr wieder ein groß angelegter Versuch vor, thesaurusgestützt automatisch zu indexieren. Das Projekt wird Anfang 2002 beendet sein.

Retrievaltests bei automatischer Indexierung

Nehmen wir an, ein Unternehmen entschied sich für sein Intranet für eine Variante automatischer Indexierung. Welches Softwareprodukt wäre das Beste? **Kristin Rapke** von der Fachhochschule Potsdam, eine der EBSCO-Preisträger/innen, hat dieses Problem für die Pressedatenbank von **Gruner + Jahr** gelöst. Zur Diskussion stehen zwei Produkte, Autonomy und DocCat von IBM. Als "Eichdatenbank" steht die mittels

intellektueller Inhaltserschließung erstellte Digitale Dokumentation (DigDok) zur Verfügung. Rapkes Retrievaltests arbeiten mit dem relativen Recall und mit der Precision. Bei kleinen Ergebnismengen (bis zu fünf Treffer) stimmen die Treffer stark überein, bei mittelgroßen Mengen (zwischen sechs und 50 Treffern) weichen die Hits schon stärker voneinander ab, um bei großen Mengen (mehr als 50 Treffer) erheblich zu schwanken. Der relative Recall des intellektuellen Systems DigDok erreicht 74%, die Systeme automatischer Indexierung erbringen ebenfalls 74% (DocCat) bzw. nur 56% (Autonomy), die Präzision beträgt bei DigDok 73% und bei den beiden Testkandidaten 57% (DocCat) bzw. 32% (Autonomy). Die qualitative Analyse der Testläufe zeigt bei DocCat große Probleme bei Personennamen, die viel zu häufig als Schlagworte zugeteilt werden (z.B. bei dreiseitigen Artikeln im Schnitt über 19 Namen). Autonomy schneidet auf der ganzen Linie schlecht ab. "Das zunächst einfach zu handhabende System stellt sich bei genauerer Betrachtung als äußerst kompliziert heraus. Die Suchstrategie erscheint beliebig und die Treffermenge zufällig" (S. 338). So ergibt sich aus dem Retrievaltest, dass Autonomy bei der intendierten Anwendung nicht einsetzbar ist. Das Vorgehen von Rapke dürfte verallgemeinerbar sein: Vor der Kaufentscheidung für ein System automatischer Retrievals sollten Retrievaltests (nach relativem Recall und nach Precision) auf der Basis bestehender Dokumente und echter Nutzerfragen - möglichst im Vergleich zu bewährten (ggf. noch intellektuell gepflegten) Systemen - durchgeführt werden.

Perspektiven der Informationswissenschaft und -praxis aus Verbandssicht

Die Deutsche Gesellschaft für Informationswissenschaft und -praxis (DGI) und der Hochschulverband Informationswissenschaft (HI) versuchen in einer gemeinsamen Sektion, Ziele und Perspektiven ihrer beiden Verbände zu diskutieren. Aus DGI-Sicht scheint es vorwiegend darum gegangen zu sein, den HI und seine erfolgreichen Symposien in die DGI zu integrieren (**Robert Funk**: DGI und HI haben dasselbe Ziel. Wenn der HI

mit der DGI kooperiert, "dann können Sie eine Geschäftsstelle nutzen!"). **Christa Womser-Hacker** von der Universität Hildesheim sieht dies differenzierter. Die DGI ist ihrer Meinung nach vor allem an der Informationspraxis orientiert und der HI vorrangig an der Theorie und nur sekundär - wenn überhaupt - an der konkreten Praxis. Der HI hat natürlich Praxisbezug innerhalb der Hochschulen; Womser-Hacker erinnert an die Kommunikation zwischen den einschlägigen Universitäten und Fachhochschulen sowie an den möglichen Austausch von Unterrichtsmodulen. Dies ist jedoch nicht die Praxis der Informationspraktiker im Berufsleben. **Ralf Dirk Hennings** von der FH Potsdam vermisst eine eindeutige Identität der Informationsspezialisten, die ja zusätzlich auch Experten anderer Gebiete sein können. Ein Archivar, der Geschichte studiert hat, sieht sich nach Hennings als Historiker, ein Dokumentar mit Informatikstudium als Informatiker, obwohl sie sich beide als Informationsspezialisten sehen sollten. Die DGI ist - so **Gerhard Knorz** - im Gegensatz zu einem "starken" Verband wie der Gesellschaft für Informatik nicht in der Lage, wissenschaftspolitisch Einfluss zu nehmen.

Inhaltlich kommt ein umfangreicher Katalog an Anregungen an beide Verbände heraus:

1. Themen besetzen!
Tragfähige Konzepte erarbeiten!
2. Aktive Mitglieder schaffen!
3. Charismatische Persönlichkeiten hervorbringen!
4. Politisch Einfluss nehmen!
Politikberatung durchführen!
5. In der Öffentlichkeit präsent sein!
6. Viele neue Mitglieder aufnehmen!

Es zeigt sich erschreckend deutlich, dass keine der Anregungen derzeit realisiert ist. Fehlt es denn an konkreten Themen? Oder fehlt es an den Persönlichkeiten?

Fazit: Wir haben etwas zu bieten. Aber: Warum sind wir so einsam?

Der kleine Markt der elektronischen Informationsdienstleistungen, wie er sich auf der INFOBASE und den DGI-Tagungen präsentiert, zelebriert eine reine Insiderveranstaltung, die von der Öffent-

lichkeit kaum und von Nachbarmärkten nur wenig wahrgenommen wird. Da die Branche übersichtlich ist und zudem die "revolutionären" Neuerungen rar sind, nutzen offenbar viele Information Professionals diese Messe und diese Tagungen nicht für ihre current awareness. Im Vergleich zum Vorjahr gibt es auf der Messe einen Besucherrückgang von 15% zu berichten, bei der Tagung fallen ganze Sessions wegen Teilnehmer mangels aus oder gestalten sich zu "Privatissima" für wenige Zuhörer - und das zu einer Zeit, in der die Informationswirtschaft zum Motor der gesamtwirtschaftlichen Entwicklung wird. Offenbar ist die selbstgewählte Isolation der INFOBASE nicht (mehr) tragfähig: Die elektronischen Informationsdienste stehen doch in engstem Kontext zur Verlagswirtschaft mit ihrem Content, der Softwarebranche mit der benötigten Informationstechnik sowie der Internetwirtschaft mit ihren Suchwerkzeugen und Portalen. Da die benachbarten Branchen nicht zu uns kommen, müssen wir den Kontakt suchen und zu ihnen gehen. Es ist durchaus angebracht, (wieder) darüber nachzudenken, ob die Aussteller der INFOBASE als geschlossene Ganzheit bei anderen Messen besser aufgehoben sind, bei der CeBIT, der Buchmesse oder der Internet World. Tut sich nichts, so werden wir - trotz vorzeigbarer Erfolge - einsam bleiben. ■

Wolfgang G. Stock

Literatur

Ralph Schmidt (Hrsg.): Information Research & Content Management. Orientierung, Ordnung und Organisation im Wissensmarkt. 23. Online-Tagung der DGI und 53. Jahrestagung der Deutschen Gesellschaft für Informationswissenschaft und Informationspraxis e.V. DGI. Proceedings. - Frankfurt: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis, 2001.