

# Vorwort

## Suchmaschinen im Internet - informationswissenschaftlich betrachtet

Dank der Suchmaschinen im World Wide Web ist das Thema „Information Retrieval“ in den letzten Jahren sehr populär geworden. Allerdings gilt dies nahezu ausschließlich für die Anwendung und nicht so sehr für die theoretischen und methodischen Grundlagen der Internetrecherche: Man „googelt“, man macht sich aber kaum Gedanken über die dahinterstehende Technik und noch weniger über die fundierenden Modelle und Theorien.

Lewandowskis „Web Information Retrieval“ widmet sich den Grundlagen der Suchmaschinen im Internet. Gegenstände seiner Arbeit sind erstens die bei den Web-Suchmaschinen derzeit eingesetzten sowie in der Literatur diskutierten Retrievalmethoden sowie zweitens eine Schwachstellenanalyse. Bei erkannten Defiziten (und die gibt es zahlreich) unterbreitet Lewandowski zudem Vorschläge zur Behebung. Eine Detailanalyse der Suchmaschinenalgorithmen aus mathematischer und computerwissenschaftlicher bzw. -technischer Sicht findet nicht statt, dafür versucht die Arbeit, die Erforschung der Nutzer und ihrer Informationsbedürfnisse gebührend zu berücksichtigen.

Der Autor motiviert seine Arbeit, hervorgegangen aus einer Dissertation an der Heinrich-Heine-Universität Düsseldorf, mit der großen Bedeutung, die die Internet-Suchmaschinen in den letzten Jahren erreicht haben. Die Abhandlung bietet - so Lewandowski - „die Grundlagen für das Verständnis der Funktionsweise und der Defizite von Web-Suchmaschinen“.

Als Methoden werden Literaturstudien (wissenschaftliche Fachliteratur sowie Patente) und, wo keine Angaben in der Literatur vorhanden sind, eigene empirische Untersuchungen vorgelegt. Die umfassende Erarbeitung des State of the art der heutigen Web-Suchmaschinen setzt den Zugang zu *allen* einschlägigen Dokumenten voraus. Dies ist aber nicht immer gegeben, da die Forschung und Entwicklung der Suchmaschinen zu großen Teilen bei den Suchmaschinenunternehmen (z.B. Google) selbst durchgeführt wird. Diese Unternehmen publizieren zwar durchaus im großen Umfang, man darf aber nicht erwarten, dass Such- oder Rankingalgorithmen bis ins

letzte Detail offengelegt würden. Lewandowski hat sich bemüht, so weit wie möglich zu recherchieren: Er hat graue Literatur (u.a. Vorträge) sowie die Patentliteratur gesichtet. Die verbliebenen Lücken wurden geschlossen, indem die Anwendungen analysiert wurden und von diesen auf die Grundlagen geschlossen worden ist.

Die aktuelle Literatur und die Praxis der Suchmaschinen wird mit der theoretischen Diskussion der Informationswissenschaft der letzten rund 40 Jahre konfrontiert. Dabei zeigt sich, dass ein Bezug zu den „alten“ Erkenntnissen aus Informationswissenschaft und der Praxis der (ebenso „alten“) professionellen Informationsdienste die aktuelle Diskussion befruchten kann.

Nach Lewandowski ist das Ziel jeglicher Bemühungen zum Information Retrieval, „dem Nutzer die für die Befriedigung seines Informationsbedürfnisses besten Ergebnisse zu liefern“. Entsprechend muss die Rolle des Nutzers bei der Diskussion der Suchmaschinen eine zentrale Stellung einnehmen. Information Retrieval ist nicht von den Suchmaschinenbetreibern erfunden worden (auch wenn diese das manchmal durchaus so darstellen), sondern hat eine „Vorgeschichte“. Diese Bemühungen müssen in eine gute Arbeit zum Web Information Retrieval einfließen. Lewandowski hat dies erkannt: „Die Arbeit konzentriert sich neben der Darstellung des Forschungsstandes im Bereich des Web Information Retrieval auf einem nutzerzentrierten Ansatz des Aufbaus von Suchmaschinen, der sich aus dem Retrieval in klassischen Datenbanken herleitet“.

Dirk Lewandowskis „Web Information Retrieval“ ist eine methodisch fundierte informationswissenschaftliche Bestandsaufnahme der Web-Suchmaschinen. Schon dies würde die Forschung weiterbringen, da eine solche deskriptiv-analytische Zusammenschau bislang fehlte. Darüber hinaus zeigt Lewandowski anhand dreier Beispiele (Aktualität, Qualität, Dokumentrepräsentation) auf, wie man unter Einsatz „klassischer“ informationswissenschaftlicher Erkenntnisse unter einem nutzerzentrierten Ansatz die algorithmischen Web Search Engines verbessern kann. Die Arbeit macht zudem deutlich, dass der informationswissenschaftliche Ansatz die (sonst eher informatisch dominierte) Suchmaschinendiskussion erfolgreich erweitern kann.

Lewandowski hat sich bereits seit einigen Jahren als *der* Fachmann für Suchmaschinen einen - guten - Ruf erworben. Ich erinnere stellvertretend an seine Vorträge bei den Online-Tagungen der DGI, beim Internationalen Symposium für Informationswissenschaft, bei der Tagung der Deutschen ISKO sowie durch Artikel z.B. in „IWP - Information - Wissenschaft und Praxis“ oder „Online Information

Review“. Er hat das IWP-Schwerpunktheft zum Thema Suchmaschinen herausgegeben. Und Lewandowski betreut seit 2003 die Rubrik „Suchmaschinen“ beim Branchennewsletter „Password“.

Mit dieser Monographie legt Lewandowski seine Forschungsergebnisse zu den Suchmaschinen sowie zu Modellen und Theorien des Information Retrieval in kompakter Form vor. Ich bin davon überzeugt, dass das Fach Informationswissenschaft sowie die forschenden und lehrenden Informationswissenschaftler davon werden profitieren können, und wünsche dem Buch eine weite Verbreitung sowie eine kritische Beachtung und Diskussion.

Düsseldorf, im Juni 2005

Wolfgang G. Stock

## Web Information Retrieval

Technologien zur Informationssuche im Internet

Dirk Lewandowski

Suchmaschinen haben eine große wirtschaftliche Bedeutung, sowohl für ihre Anbieter als auch für ihre Nutzer. Sie leiden jedoch unter dem von der Informatik verfolgten technikzentrierten Ansatz, der wesentliche Erkenntnisse der Informationswissenschaft und langjährige Erfahrungen aus der Praxis der professionellen Informationsdienste ausblendet. Die Nutzer „googeln“, wissen aber kaum, wie die Ergebnisse zustande kommen.

Lewandowski beschreibt das Web Information Retrieval mit seinen Besonderheiten aus informationswissenschaftlicher Sicht vor allem im Vergleich zum klassischen Information Retrieval und erklärt fundiert aber auch für Laien verständlich die Funktionsweise und Defizite von Suchmaschinen. Nach einer gründlichen Beschreibung des Standes der Technik und einer Schwachstellenanalyse stellt er einen Ansatz der Rechercheverfeinerung in mehreren Schritten vor.

Suchmaschinen müssen, so die Kernforderung des Autors, ihren Nutzern bei der Suche im Web die zur Befriedigung ihrer Informationsbedürfnisse besten Ergebnisse liefern. Für diesen Dienst am Kunden zeigt er neue Wege auf.

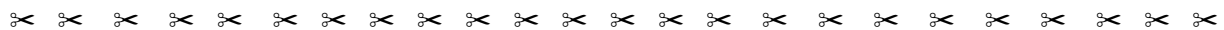
Leserkreis:

Informatiker, Informationswissenschaftler, Information Broker, Wissensmanager, Medienfachleute, Bibliothekare, Content Anbieter, Webmaster, Studenten

### Dirk Lewandowski

#### Web Information Retrieval. Technologien zur Informationssuche im Internet

Frankfurt am Main 2005, 248 Seiten mit Sachregister, DGI Schrift (Informationswissenschaft - 7)  
ISBN 3-925474-55-2, EUR 25,- (für DGI-Mitglieder EUR 20,-)



Ich bestelle \_\_\_\_\_ Exemplare der Neuerscheinung **Lewandowski: Web Information Retrieval**

zum Preis von \_\_\_\_\_ Euro plus Versandkosten.

Name, Vorname: \_\_\_\_\_

Rechnungsanschrift: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Datum, Unterschrift \_\_\_\_\_

Auslieferung September 2005, Vorbestellungen an die DGI-Geschäftsstelle, Ostbahnhofstraße 13,  
60314 Frankfurt am Main, Fax (069) 4 90 90 96, E-Mail: publikation@dgi-info.de

