

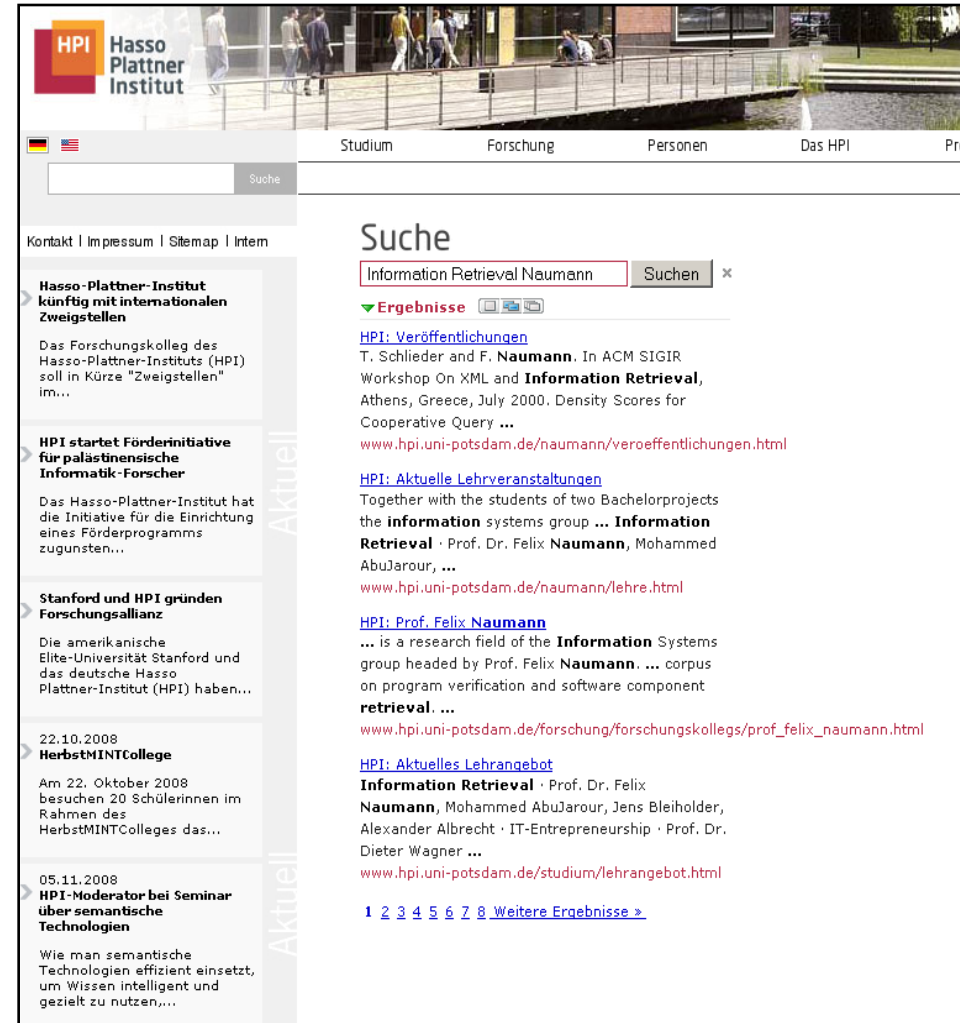
Seminar: Selected Topics in IR

Presentation of topics

Alexander Albrecht

Hasso-Plattner-Institut für Softwaresystemtechnik GmbH
FB Informationssysteme
Prof.-Dr.-Helmert-Str. 2-3, D-14482 Potsdam
Alexander.Albrecht@hpi.uni-potsdam.de



Models for Document & Query Representation



The screenshot shows the HPI website's search results page. The header includes the HPI logo and navigation links for 'Studium', 'Forschung', 'Personen', and 'Das HPI'. A search bar contains the query 'Information Retrieval Naumann'. The search results are displayed in a list format, with each entry including a title, a brief description, and a URL. The results are categorized by 'Aktuelle' (Current) and 'Ergebnisse' (Results).

Suche

Information Retrieval Naumann x

▼ **Ergebnisse**  

[HPI: Veröffentlichungen](#)
T. Schlieder and F. Naumann. In ACM SIGIR Workshop On XML and **Information Retrieval**, Athens, Greece, July 2000. Density Scores for Cooperative Query ...
www.hpi.uni-potsdam.de/naumann/veroeffentlichungen.html

[HPI: Aktuelle Lehrveranstaltungen](#)
Together with the students of two Bachelorprojects the **information** systems group ... **Information Retrieval** · Prof. Dr. Felix Naumann, Mohammed AbuJarour, ...
www.hpi.uni-potsdam.de/naumann/lehre.html

[HPI: Prof. Felix Naumann](#)
... is a research field of the **Information** Systems group headed by Prof. Felix Naumann. ... corpus on program verification and software component **retrieval** ...
www.hpi.uni-potsdam.de/forschung/forschungskollegs/prof_felix_naumann.html

[HPI: Aktuelles Lehrangebot](#)
Information Retrieval · Prof. Dr. Felix Naumann, Mohammed AbuJarour, Jens Bleiholder, Alexander Albrecht · IT-Entrepreneurship · Prof. Dr. Dieter Wagner ...
www.hpi.uni-potsdam.de/studium/lehrangebot.html

1 2 3 4 5 6 7 8 [Weitere Ergebnisse >](#)

Hasso-Plattner-Institut künftig mit internationalen Zweigstellen
Das Forschungskolleg des Hasso-Plattner-Instituts (HPI) soll in Kürze "Zweigstellen" im...

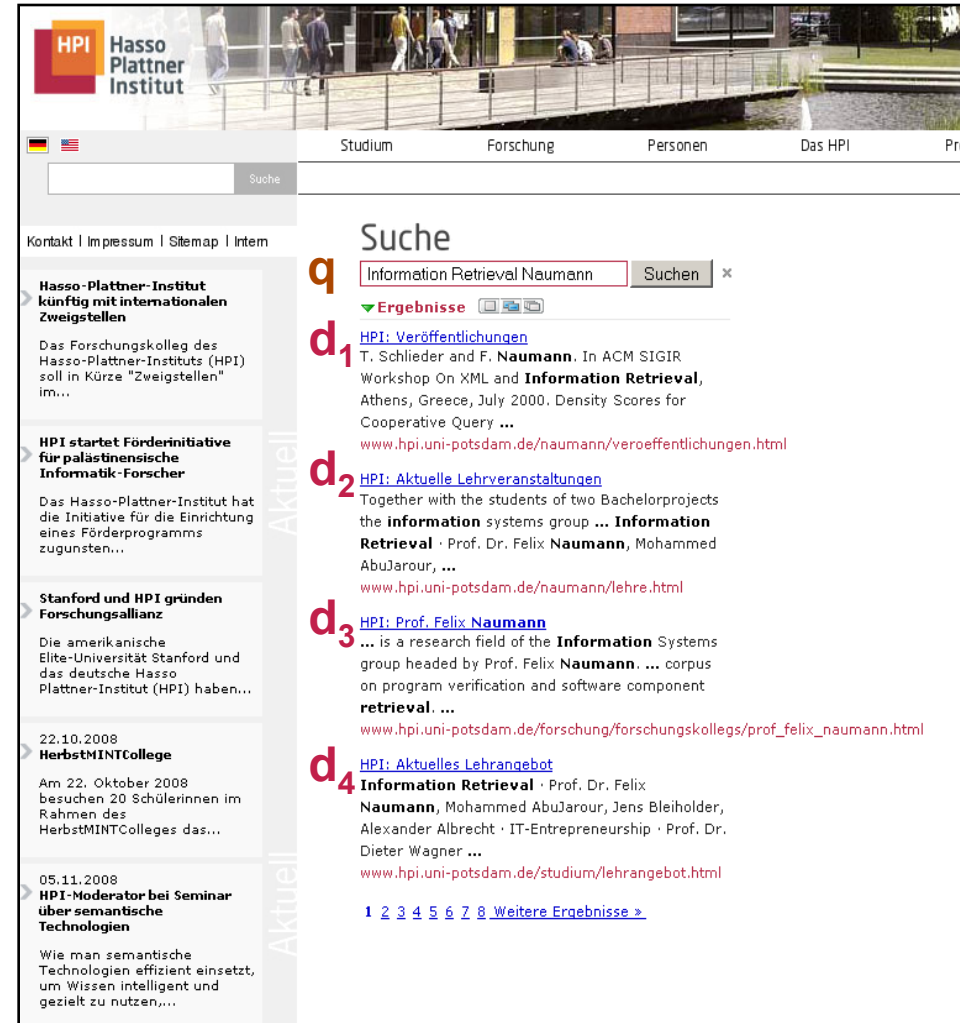
HPI startet Förderinitiative für palästinensische Informatik-Forscher
Das Hasso-Plattner-Institut hat die Initiative für die Einrichtung eines Förderprogramms zugunsten...

Stanford und HPI gründen Forschungsallianz
Die amerikanische Elite-Universität Stanford und das deutsche Hasso Plattner-Institut (HPI) haben...

22.10.2008
HerbstMINTCollege
Am 22. Oktober 2008 besuchen 20 Schülerinnen im Rahmen des HerbstMINTColleges das...

05.11.2008
HPI-Moderator bei Seminar über semantische Technologien
Wie man semantische Technologien effizient einsetzt, um Wissen intelligent und gezielt zu nutzen,...

Models for Document & Query Representation



The screenshot shows the HPI website's search results page. At the top, there is a navigation bar with links for 'Studium', 'Forschung', 'Personen', and 'Das HPI'. Below this is a search bar with the text 'Suche' and a search button. The main content area is titled 'Suche' and displays search results for the query 'Information Retrieval Naumann'. The results are numbered 1 through 4, each with a title, a brief description, and a URL. A vertical sidebar on the left contains news items, and a vertical label 'Aktuell' is positioned between the sidebar and the main content area.

Kontakt | Impressum | Sitemap | Intern

Hasso-Plattner-Institut
künftig mit internationalen
Zweigstellen

Das Forschungskolleg des Hasso-Plattner-Instituts (HPI) soll in Kürze "Zweigstellen" im...

HPI startet Förderinitiative für palästinensische Informatik-Forscher

Das Hasso-Plattner-Institut hat die Initiative für die Einrichtung eines Förderprogramms zugunsten...

Stanford und HPI gründen Forschungsallianz

Die amerikanische Elite-Universität Stanford und das deutsche Hasso Plattner-Institut (HPI) haben...

22.10.2008
HerbstMINTCollege



Am 22. Oktober 2008 besuchen 20 Schülerinnen im Rahmen des HerbstMINTColleges das...

05.11.2008
HPI-Moderator bei Seminar über semantische Technologien

Wie man semantische Technologien effizient einsetzt, um Wissen intelligent und gezielt zu nutzen,...

Suche

Information Retrieval Naumann Suchen x

▼ **Ergebnisse**  

d₁ [HPI: Veröffentlichungen](#)
T. Schlieder and F. Naumann. In ACM SIGIR Workshop On XML and **Information Retrieval**, Athens, Greece, July 2000. Density Scores for Cooperative Query ...
www.hpi.uni-potsdam.de/naumann/veroeffentlichungen.html

d₂ [HPI: Aktuelle Lehrveranstaltungen](#)
Together with the students of two Bachelorprojects the **information** systems group ... **Information Retrieval** · Prof. Dr. Felix Naumann, Mohammed AbuJarour, ...
www.hpi.uni-potsdam.de/naumann/lehre.html

d₃ [HPI: Prof. Felix Naumann](#)
... is a research field of the **Information** Systems group headed by Prof. Felix Naumann. ... corpus on program verification and software component **retrieval** ...
www.hpi.uni-potsdam.de/forschung/forschungskollegs/prof_felix_naumann.html

d₄ [HPI: Aktuelles Lehrangebot](#)
Information Retrieval · Prof. Dr. Felix Naumann, Mohammed AbuJarour, Jens Bleiholder, Alexander Albrecht · IT-Entrepreneurship · Prof. Dr. Dieter Wagner ...
www.hpi.uni-potsdam.de/studium/lehrangebot.html

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [Weitere Ergebnisse >](#)

Models for Document & Query Representation

- Query: *Information Retrieval Naumann*

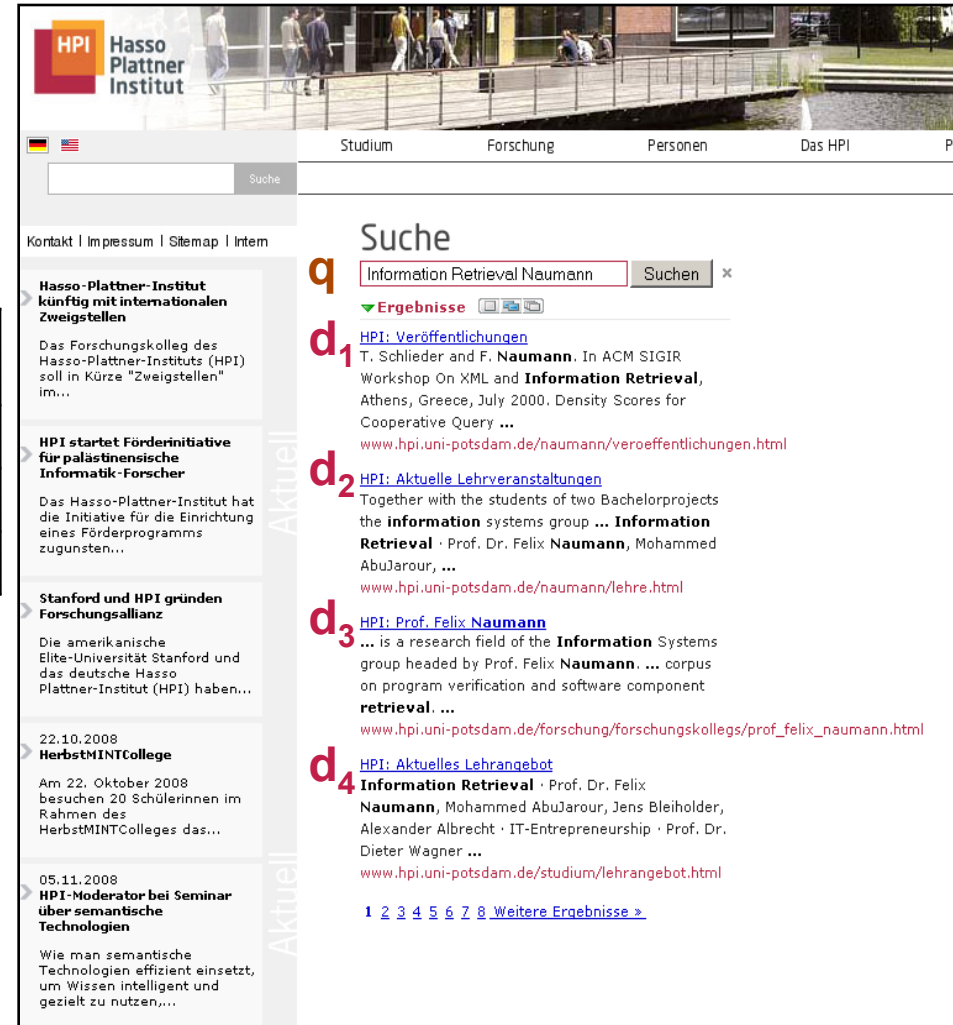
- Number of occurrences

	d_1	d_2	d_3	d_4	d_{Seminar}
Naumann	231	66	15	14	14
Information	78	8	8	9	14
Retrieval	1	4	1	2	7

- Challenges

- Ranking query results
- Importance of query terms
- Finding similar documents
- ...

- Model for document & query representation needed!



The screenshot shows the HPI website search interface. The search query is "Information Retrieval Naumann". The results are ranked and numbered as follows:

- d₁**: [HPI: Veröffentlichungen](#) - T. Schlieder and F. Naumann. In ACM SIGIR Workshop On XML and **Information Retrieval**, Athens, Greece, July 2000. Density Scores for Cooperative Query ... www.hpi.uni-potsdam.de/naumann/veroeffentlichungen.html
- d₂**: [HPI: Aktuelle Lehrveranstaltungen](#) - Together with the students of two Bachelorprojects the **information** systems group ... **Information Retrieval** · Prof. Dr. Felix Naumann, Mohammed AbuJarour, ... www.hpi.uni-potsdam.de/naumann/lehre.html
- d₃**: [HPI: Prof. Felix Naumann](#) - ... is a research field of the **Information** Systems group headed by Prof. Felix Naumann. ... corpus on program verification and software component **retrieval**. ... www.hpi.uni-potsdam.de/forschung/forschungskollegs/prof_felix_naumann.html
- d₄**: [HPI: Aktuelles Lehrangebot](#) - **Information Retrieval** · Prof. Dr. Felix Naumann, Mohammed AbuJarour, Jens Bleiholder, Alexander Albrecht · IT-Entrepreneurship · Prof. Dr. Dieter Wagner ... www.hpi.uni-potsdam.de/studium/lehrangebot.html

At the bottom of the results, there is a link: [1 2 3 4 5 6 7 8 Weitere Ergebnisse >](#)

- **(Term) Vector-Space Model**
 - *Calculate similarity between query and documents that are represented as weighted term vectors*
- **Binary Independence Model**
 - *Document d is relevant iff $P(R=1 \mid d, q) > P(R=0 \mid d, q)$*
 - *Compute $P(R \mid d, q)$ for all documents d*
- **Language Model**
 - *Compute $P(q \mid d)$ – probability that the language model of document d generates query q*

Text Classification Problem

- **Example: E-Mail Classification**
 - Document space X : E-Mails
 - Classes $C = \{\text{Spam, Private, Business, Advertising}\}$
 - Learn a classifier γ that maps E-Mails to classes, $\gamma: X \rightarrow C$
- **Learning algorithms**
 - Naive Bayes (NB)
 - Support Vector Machines (SVM)
 - Supervised Learning: A supervisor serves as a teacher directing the learning process based on a training set D of labeled documents $\langle d, c \rangle \in X \times C$ in order to produce a learned classifier γ using NB or SVM.
- Your presentation should include: Experimental results of an exemplary implementation using one of the presented approaches.