



DIS12 Information Retrieval

01: Introduction

Philipp Schaer, Technische Hochschule Köln, Cologne, Germany

Version: 2019-10-04

Was ist eigentlich Retrieval?



CHEFKOCH.DE

YAHOO!

bing



Technology
Arts Sciences
TH Köln



Google



Yandex



Was ist eigentlich Retrieval?

In diesem Kurs geht es um Information **Retrieval** und wir benutzen folgende, **vorläufige Arbeitsdefinition**:

„Gegeben eine **Anfrage** und einen **Dokumentenkorpus**, finde **relevante Dokumente**.“

- **Anfrage:** Eine Anfrage ist die Beschreibung eines **Informationsbedürfnisses**, die an das IR-System geschickt wird. Kann natürlichsprachig oder formal (Anfragesprache) sein.
- **Korpus:** Eine Sammlung von durchsuchbaren Dokumenten / Ressourcen. In unserem Falle meistens Textdokumente.
- **Relevanz:** Befriedigung des Informationsbedürfnisses eines Benutzers.

Weitere Definitionen von IR

Drei Definitionen (von vielen)

- „Information retrieval (IR) is **finding material** (usually documents) of an **unstructured nature** (usually text) that **satisfies an information need** from within **large collections** (usually stored on computers).“ (Manning et al., 2008)
- „Information retrieval (IR) is a field concerned with the design, development, and evaluation of interactive systems that help users find information.“ (Arguello, 2017)
- “Information retrieval is a field concerned with the structure, analysis, organization, storage, and retrieval of information.“ (Salton, 1968)

Weitere Definitionen von IR

Drei Definitionen (von vielen)

Besser!



- „Information retrieval (IR) is **finding material** (usually documents) of an **unstructured nature** (usually text) that **satisfies an information need** from within **large collections** (usually stored on computers).“ (Manning et al., 2008)
- „Information retrieval (IR) is a field concerned with the design, development, and evaluation of interactive systems that help users find information.“ (Arguello, 2017)
- “Information retrieval is a field concerned with the structure, analysis, organization, storage, and retrieval of information.“ (Salton, 1968)

Es geht also um die Nadel im ...



Formalia – Veranstaltungen

Die Veranstaltung ist aufgeteilt in

- **Vorlesung** (Donnerstags, 13:30 Uhr, Raum 154)
 - **Laborgruppe A** (Donnerstags, 15:15 Uhr, Raum 147)
 - **Laborgruppe B** (Donnerstags, 17:00 Uhr, Raum 147)
-
- In den Übungen werden **Übungszettel bearbeitet**, die Ihnen **eine Woche zuvor** zur Verfügung gestellt werden.
 - Voraussetzung ist, dass Sie diese Zettel **zuhause** bearbeiten und vorbereiten. In den Übungen wird erwartet, dass Sie die Aufgaben **bereits bearbeitet haben!**
 - Übungen starten **ab nächster Woche!**

Formalia – Prüfungsleistungen

Es gibt folgende Prüfungsleistung

- **Abschlussklausur** (120 Minuten, max. 100 Punkte)

Unterstützend bitte ich an

- **Probeklausur Mitte des Semesters** (30 Minuten, keine Punkte)
- **Extrapunkte durch aktive Laborteilnahme** (max. 10 Punkte)
 - Vorrechnen an der Tafel / Präsentation einer Lösung / Mini-Projekt / ...
- Im gesamten Semester können Sie **100 Punkte** erreichen
- Es gilt folgende Notentabelle

<50	50	55	60	65	70	75	80	85	90	95
n.b.	4,0	3,7	3,3	3,0	2,7	2,3	2,0	1,7	1,3	1,0

Formalia – Prüfungsleistungen

Theoretisch reicht also die Abschlussklausur zum Bestehen und zum Erreichen der höchsten Punktzahl.

- **Aber:** Das rate ich natürlich niemanden!

Die Extrapunkte verfallen zum Semesterende.

- Wer in die Nachprüfung geht, startet wieder bei 0.
- Die Nachklausur bringt wieder max. 100 Punkte.
- Es gilt die gleiche Notentabelle.

Punkte hin oder her! Sie **sollten**

- regelmäßig + aktiv an den Vorlesungen und Übungen teilnehmen,
- jederzeit Fragen stellen,
- sich in den Übungen austauschen/gemeinsam arbeiten!

Vorlesungsplan WS 2019/20

10.10.19 Einführung, Boolesches R.	28.11.19 Query Expansion
17.10.19 Boolesches Retrieval	05.12.19 Evaluation
24.10.19 TF-IDF	12.12.19 Web-Crawling
31.10.19 Ranking + Vektorraum	18.12.19 Link-Analyse
07.11.19 Indexkonstruktion	25.12.19 Weihnachten (frei)
14.11.19 Probeklausur , Suchmaschinentechnologie	09.01.20 Learning to Rank (Puffer)
21.11.19 Projektwoche (frei)	16.01.20 Wiederholung + Fragestunde

Alle Vorlesungsinhalte und Übungen finden Sie im IWS-Moodle:

- <https://elearning.iws.th-koeln.de/moodle/course/view.php?id=1081>

Einbettung in das Studium

<p>Programmierung Webentwicklung Hausarbeit</p>	<p>Informationserschließung Wissensorganisation <i>Klausur oder Hausarbeit</i></p>	<p>Digitale Informationsgesellschaft Informationsethik Berufsfelderkundung (Profil2) Lehrveranstaltungsbegleitende Prüfung Präsentation</p>	<p>Informationsvisualisierung Lehrveranstaltungsbegleitende Prüfung oder mündliche Prüfung</p>	<p>Information in Unternehmen Präsentation und schriftliche Ausarbeitung</p>
<p>Programmierung Softwareentwicklung <i>Klausur</i></p>	<p>Informationserschließung Strukturierte Dokumentbeschreibung <i>Klausur oder Hausarbeit</i></p>	<p>Datenmodellierung Lehrveranstaltungsbegleitende Prüfung oder Klausur</p>	<p>Statistische Datenanalyse Projektarbeit <i>Klausur</i></p>	<p>Informationsquellen Informationsrecherche <i>Klausur oder mündliche Prüfung</i></p>
<p>Informationssysteme Content- & Dokumenten-managementsysteme Lehrveranstaltungsbegleitende Prüfung oder Klausur</p>	<p>Information Retrieval <i>Klausur</i></p>	<p>Datenbanksysteme Lehrveranstaltungsbegleitende Prüfung oder Klausur</p>	<p>Data Mining Lehrveranstaltungsbegleitende Prüfung oder Klausur</p>	<p>Informationsanalyse Projektarbeit Dokumentation</p>

Schwerpunkt Data Librarian

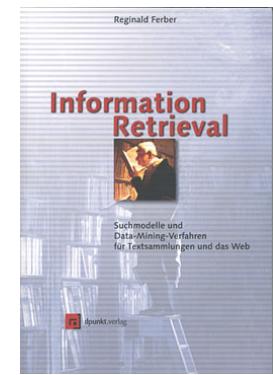
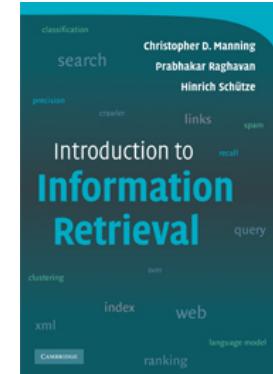
<p>Suchmaschinentechnologie</p> <ul style="list-style-type: none"> Webtechnologien <p>Lehrveranstaltungsbegleitende Prüfung Klausur oder Hausarbeit</p>	<p>Projektarbeit I</p> <p><i>Data Librarian</i></p> <ul style="list-style-type: none"> Interdisziplinäres Projekt (Profil2) 	<p>Informationsrecht & Datenschutzrecht</p> <ul style="list-style-type: none"> Wissenschaftliches Arbeiten <p><i>Klausur</i> Lehrveranstaltungsbegleitende Prüfung</p>	<p>ASC</p> <ul style="list-style-type: none"> Studienportfolio <p><i>Lernportfolio</i></p>	<p>Information Consultancy, Wissenschaftskommunikation & Wissenschaftssoziologie und -politik</p> <p><i>Data Librarian</i></p> <p>Lehrveranstaltungsbegleitende Prüfung Hausarbeit</p>
<p>Projektarbeit II</p> <p><i>Data Librarian</i></p> <p>Lehrveranstaltungsbegleitende Prüfung</p>	<p>Lehrveranstaltungsbegleitende Prüfung Projektarbeit</p>	<p>Forschungsdaten I</p> <p><i>Data Librarian</i></p> <ul style="list-style-type: none"> Lizenzmanagement, Digitales Publizieren & Open Access <p><i>Klausur</i></p>	<p>Informetrie, Bibliometrie, Szientometrie</p> <ul style="list-style-type: none"> Empirische Forschungsmethoden <p><i>Klausur</i></p>	<p>Automatische Erschließung</p> <p><i>Data Librarian</i></p> <p>Klausur oder Hausarbeit</p>
<p>Projektarbeit III</p> <p><i>Data Librarian</i></p> <p>Lehrveranstaltungsbegleitende Prüfung</p>	<p>Forschungsdaten II</p> <p><i>Data Librarian</i></p> <ul style="list-style-type: none"> Digitalisierung & Langzeitarchivierung <p><i>Hausarbeit</i> Projektarbeit</p>	<p>Seminare</p> <ul style="list-style-type: none"> Seminar zu aktuellen Themen Seminar zur Bachelorarbeit <p><i>Schriftliche Arbeit</i> Lehrveranstaltungsbegleitende Prüfung</p>	<p>Bachelorarbeit</p>	<p>Bachelorarbeit</p>

Schwerpunkt Data Analyst

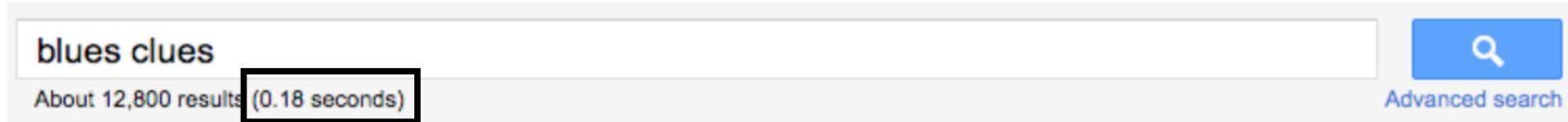
<p>Suchmaschinentechnologie</p> <ul style="list-style-type: none"> Webtechnologien <p>Lehrveranstaltungsbegleitende Prüfung Klausur oder Hausarbeit</p>	<p>Projektarbeit I</p> <p>Data Analyst</p> <ul style="list-style-type: none"> Interdisziplinäres Projekt (Profil2) 	<p>Informationsrecht & Datenschutzrecht</p> <ul style="list-style-type: none"> Wissenschaftliches Arbeiten <p>Klausur Lehrveranstaltungsbegleitende Prüfung</p>	<p>ASC</p> <ul style="list-style-type: none"> Studienportfolio <p>Lernportfolio</p>	<p>Trend-Detektion & Trendanalyse</p> <p>Data Analyst</p> <p>Fallstudie mit Präsentation</p>
<p>Projektarbeit II</p> <p>Data Analyst</p> <p>Lehrveranstaltungsbegleitende Prüfung</p>	<p>Lehrveranstaltungsbegleitende Prüfung Projektarbeit</p>	<p>Market & Business Intelligence</p> <p>Data Analyst</p> <p>mündliche Prüfung und/oder Dokumentation und/oder Projektpräsentation</p>	<p>Informetrie, Bibliometrie, Szientometrie</p> <ul style="list-style-type: none"> Empirische Forschungsmethoden <p>Klausur</p>	<p>Natural Language Processing</p> <p>Data Analyst</p> <p>Lehrveranstaltungsbegleitende Prüfung Hausarbeit</p>
<p>Projektarbeit III</p> <p>Data Analyst</p> <p>Lehrveranstaltungsbegleitende Prüfung</p>	<p>Big Data</p> <p>Data Analyst</p> <p>Projektarbeit oder Klausur</p>	<p>Seminare</p> <ul style="list-style-type: none"> Seminar zu aktuellen Themen <p>Seminar zur Bachelorarbeit</p> <p>Schriftliche Arbeit Lehrveranstaltungsbegleitende Prüfung</p>	<p>Bachelorarbeit</p>	<p>Bachelorarbeit</p>

Hinweise zum Folienmaterial

- Vorlesung und Übung basieren auf dem **Lehrbuch** „Introduction to Information Retrieval“ von Manning et al.: <http://nlp.stanford.edu/IR-book/>
- Gibt es auch als **Coursera-Video-Kurs**:
<https://class.coursera.org/nlp/lecture>
- Auch gut: „Information Retrieval“ von Reginald Ferber (Deutsch):
<http://information-retrieval.de/irb/irb.html>
- Danke auch an **Vivien Petras, Philipp Mayr und Jaime Arguello** für zusätzliches Folienmaterial (gesondert gekennzeichnet).



Was gehört noch zum Thema Retrieval?



Zwei zentrale Größen

- **Effizienz:** Wir wollen das Ergebnis noch vor Feierabend (oder noch besser: in 0,18 Sekunden)
 - **Effektivität:** Liefere nur Ergebnisse, die Informationsbedürfnisse der Nutzer befriedigen
-
- Später: Klarer Fokus auf das Thema Effektivität.
 - Allerdings: Wir werden uns auch darüber unterhalten, wie Suchmaschinen Ergebnisse möglichst schnell liefern können.

Effizienz von Suche: grep

- Lorem ipsum dolor sit amet,
- consetetur sadipscing elitr,
- sed diam nonumy eirmod tempor,
- invidunt ut labore et dolore magna
- aliquyam erat, sed diam voluptua.
- At vero eos et accusam et justo
- duo dolores et ea rebum.

Finde alle Zeilen mit „et“.

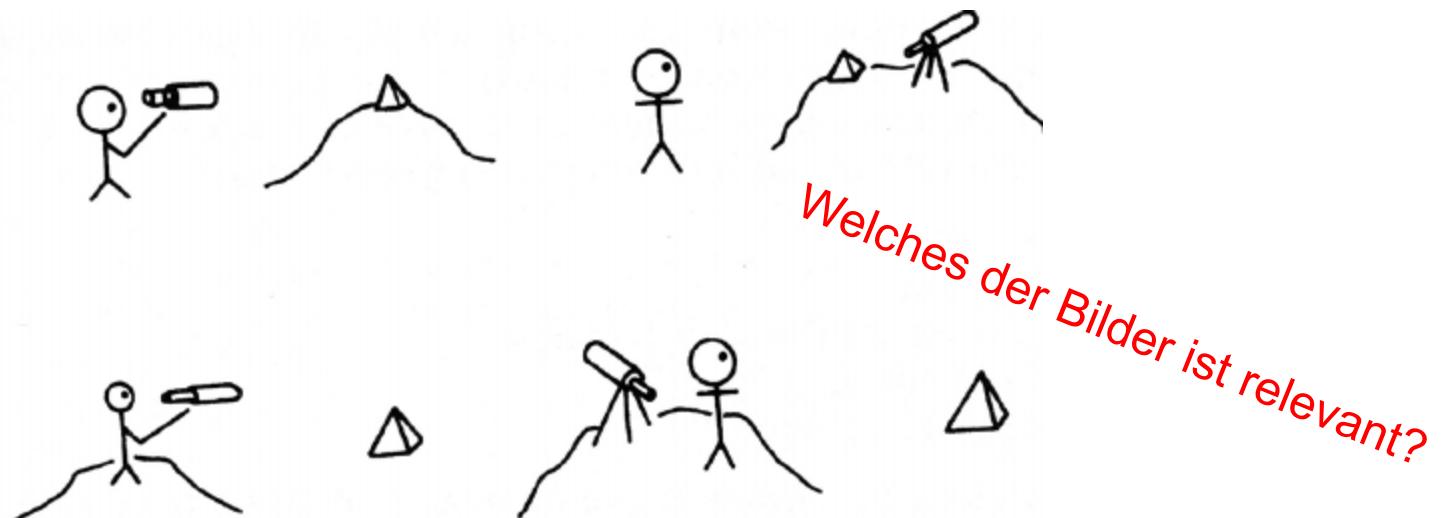
- Wie viele Schritte braucht grep?
- Ist das ein guter Weg zu suchen? Geht das nicht besser?

Effektivität: Was ist eigentlich Relevanz?

Schwierig...

„The man saw the pyramid on the hill with the telescope.“

- Viele Interpretationen dieses Satzes sind denkbar...



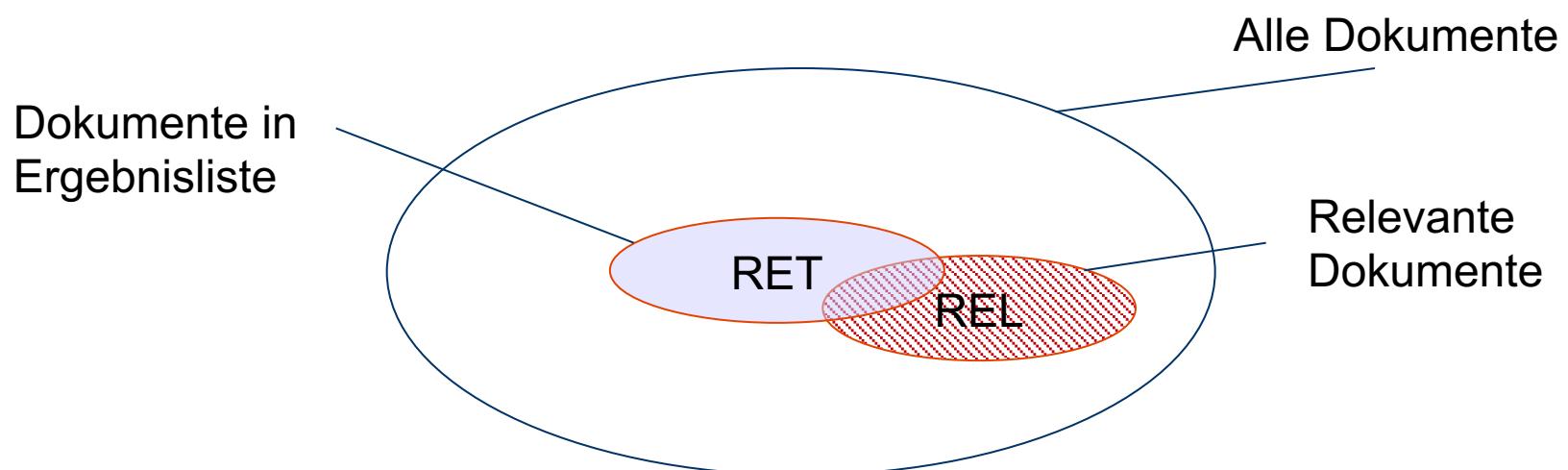
Maßzahlen für die Evaluation

- **Precision** (Treffergenauigkeit)

$$\mathcal{P} = \frac{|\text{RET} \cap \text{REL}|}{|\text{RET}|}$$

- **Recall** (Treffervollständigkeit)

$$\mathcal{R} = \frac{|\text{RET} \cap \text{REL}|}{|\text{REL}|}$$



Precision und Recall: Ein Beispiel

	Relevant	Nicht relevant
Gefunden	30	12
Nicht gefunden	14	44

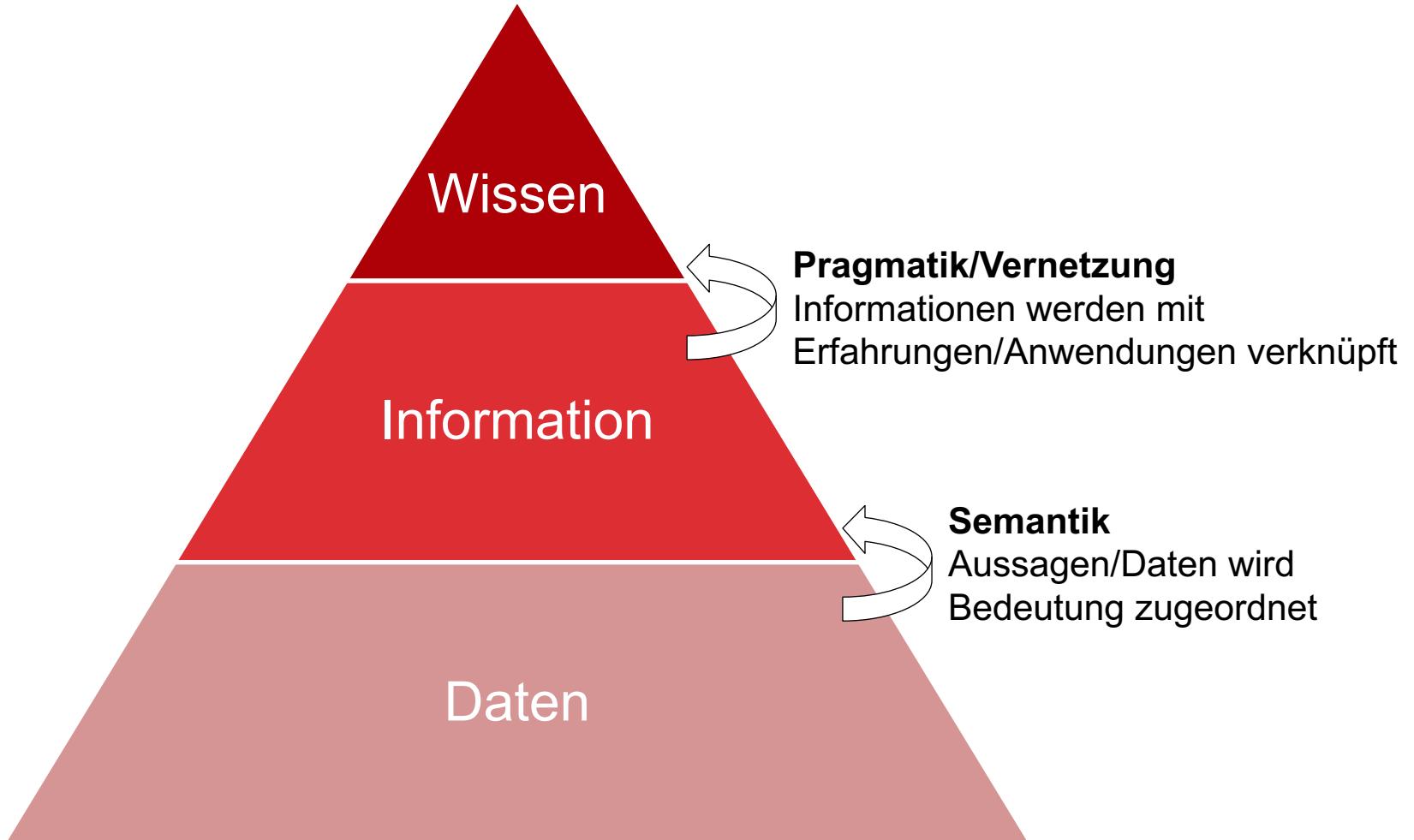
- Precision $P = 30 / (30 + 12) \approx 0,714$
- Recall $R = 30 / (30 + 14) \approx 0,681$

Was ist eigentlich Information?

- Schwierig...
- Häufig verwendet man den Begriff Information ganz intuitiv (und leider
 - Analog zu Potter Stewart (1964): „I know it, when I see it“
- Aufteilung in drei zentrale Begriffe:
 - Daten
 - Information
 - Wissen



Daten – Information – Wissen



Strukturierte Daten

Strukturierte Daten sind z.B. Tabellendaten

Angestellter	Boss	Gehalt
Berthold Heisterkamp	Bernd Stromberg	50000
Ulf Steinke	Bernd Stromberg	60000
Sinan Turçulu	Timo Becker	50000

- Numerische Anfragen und Exact Match sind möglich, bspw.: **Gehalt < 60000 AND Boss = Timo Becker**
- Toll, aber meistens nicht das was wir im Information Retrieval wollen → **Wir suchen in unstrukturierte Daten!**

Unstrukturierte Daten...

 WIKIPEDIA
The Free Encyclopedia

Article Discussion Read Edit View history 

Gerard Salton

From Wikipedia, the free encyclopedia

Gerard Salton (8 March 1927 in Nuremberg - 28 August 1995), also known as Gerry Salton, was a Professor of Computer Science at Cornell University. Salton was perhaps the leading computer scientist working in the field of information retrieval during his time. His group at Cornell developed the SMART Information Retrieval System, which he initiated when he was at Harvard.

Salton was born Gerhard Anton Sahlmann on March 8, 1927 in Nuremberg, Germany. He received a Bachelor's (1950) and Master's (1952) degree in mathematics from Brooklyn College, and a Ph.D. from Harvard in Applied Mathematics in 1958, the last of Howard Aiken's doctoral students, and taught there until 1965, when he joined Cornell University and co-founded its department of Computer Science.

Salton was perhaps most well known for developing the now widely used Vector Space Model for Information Retrieval^[1]. In this model, both documents and queries are represented as vectors of term counts, and the similarity between a document and a query is given by the cosine between the term vector and the document vector. In this paper, he also introduced TF-IDF, or term-frequency-inverse-document frequency, a model in which the score of a term in a document is the ratio of the number of terms in that document divided by the frequency of the number of documents in which that term occurs. (The concept of inverse document frequency, a measure of specificity, had been introduced in 1972 by Karen Sparck-Jones^[2].) Later in life, he became interested in automatic text summarization and analysis^[3], as well as automatic hypertext generation^[4]. He published over 150 research articles and 5 books during his life.

Salton was editor-in-chief of the Communications of the ACM and the Journal of the ACM, and chaired SIGIR. He was an associate editor of the ACM Transactions on Information Systems. He was an ACM Fellow (elected 1995), received an Award of Merit from the American Society for Information Science (1989), and was the first recipient of the SIGIR Award for outstanding contributions to study of information retrieval (1983) -- now called the Gerard Salton Award.

References

[edit]

1. ^ G. Salton , A. Wong , C. S. Yang, A vector space model for automatic indexing , Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975
2. ^ Spärck Jones, Karen (1972), "A statistical interpretation of term specificity and its application in retrieval" , Journal of Documentation 28 (1): 11–21, doi:10.1108/eb026526 

Das Dokument ist teilstrukturiert...

Log in / create account

Gerard Salton

From Wikipedia, the free encyclopedia

Gerard Salton (8 March 1927 in Nuremberg - 28 August 1995), also known as Gerry Salton, was a Professor of Computer Science at Cornell University. Salton was perhaps the leading computer scientist working in the field of information retrieval during his time. His group at Cornell developed the SMART Information Retrieval System, which he initiated when he was at Harvard.

Salton was born Gerhard Anton Sahlmann on March 8, 1927 in Nuremberg, Germany. He received a Bachelor's (1950) and Master's (1952) degree in mathematics from Brooklyn College, and a Ph.D. from Harvard in Applied Mathematics in 1958, the last of Howard Aiken's doctoral students, and taught there until 1965, when he joined Cornell University and co-founded its department of Computer Science.

Salton was perhaps most well known for developing the now widely used Vector Space Model for Information Retrieval^[1]. In this model, both documents and queries are represented as vectors of term counts, and the similarity between a document and a query is given by the cosine between the term vector and the document vector. In this paper, he also introduced TF-IDF, or term-frequency-inverse-document frequency, a model in which the score of a term in a document is the ratio of the number of terms in that document divided by the frequency of the number of documents in which that term occurs. (The concept of inverse document frequency, a measure of specificity, had been introduced in 1972 by Karen Sparck-Jones^[2].) Later in life, he became interested in automatic text summarization and analysis^[3], as well as automatic hypertext generation^[4]. He published over 150 research articles and 5 books during his life.

Salton was editor-in-chief of the Communications of the ACM and the Journal of the ACM, and chaired SIGIR. He was an associate editor of the ACM Transactions on Information Systems. He was an ACM Fellow (elected 1995), received an Award of Merit from the American Society for Information Science (1989), and was the first recipient of the SIGIR Award for outstanding contributions to study of information retrieval (1983) -- now called the Gerard Salton Award.

References

[edit]

1. ^ G. Salton , A. Wong , C. S. Yang, A vector space model for automatic indexing ↗, Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975
2. ^ Spärck Jones, Karen (1972), "A statistical interpretation of term specificity and its application in retrieval" ↗, Journal of Documentation 28 (1): 11–21, doi:10.1108/eb026526 ↗

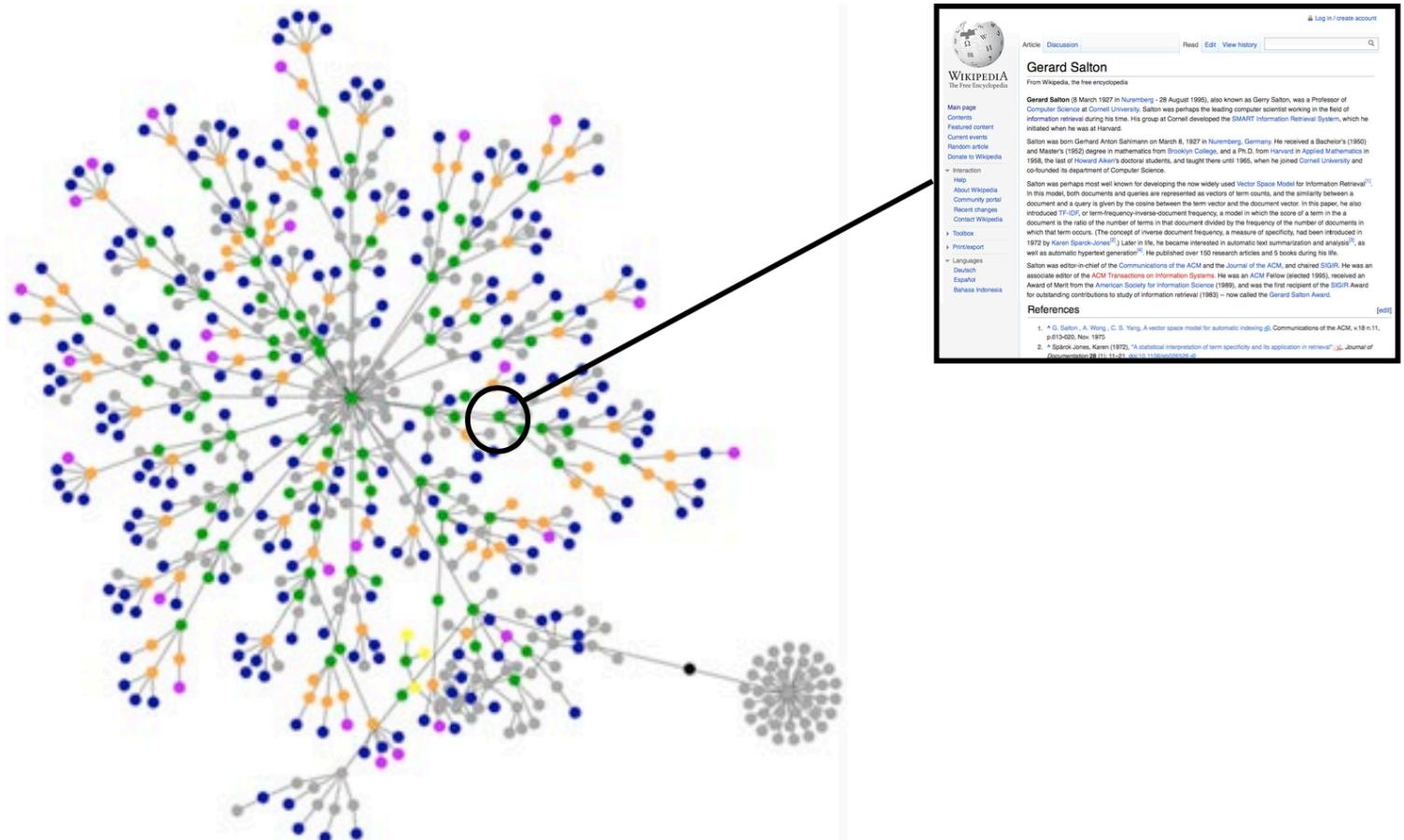
Dokumentenstrukturen

Es gibt zwar eine Struktur, aber der für uns wichtige Teil ist der Text des Artikels. Der Text hat nur **wenig bzw. gar keine Struktur**, die der Computer verstehen könnte.

Allerdings wird sich herausstellen, dass es für eine Suchmaschine gar nicht notwendig ist, den Text „zu verstehen“ um eine Anfrage z.B. nach „Gerald Salton“ mit relevanten Ergebnissen zu beantworten. Es reichen uns **unstrukturierte/semi-strukturierte Daten!**

2. ^ Spärck Jones, Karen (1972), "A statistical interpretation of term specificity and its application in retrieval" , *Journal of Documentation* 28 (1): 11–21, doi:10.1108/eb026526 

Kollektionsstrukturen



Strukturen durch Klassifikationen

 WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
Print/export

Languages
Deutsch
Español
Bahasa Indonesia

Article Discussion Read Edit View history Search

Gerard Salton

From Wikipedia, the free encyclopedia

Gerard Salton (8 March 1927 in Nuremberg – 28 August 1995), also known as Gerry Salton, was a Professor of Computer Science at Cornell University. Salton was perhaps the leading computer scientist working in the field of information retrieval during his time. His group at Cornell developed the SMART Information Retrieval System, which he initiated when he was at Harvard.

Salton was born Gerhard Anton Sahlmann on March 8, 1927 in Nuremberg, Germany. He received a Bachelor's (1950) and Master's (1952) degree in mathematics from Brooklyn College, and a Ph.D. from Harvard in Applied Mathematics in 1958, the last of Howard Aiken's doctoral students, and taught there until 1965, when he joined Cornell University and co-founded its department of Computer Science.

Salton was perhaps most well known for developing the now widely used Vector Space Model for Information Retrieval^[1]. In this model, both documents and queries are represented as vectors of term counts, and the similarity between a document and a query is given by the cosine between the term vector and the document vector. In this paper, he also introduced TF-IDF, or term-frequency-inverse-document frequency, a model in which the score of a term in a document is the ratio of the number of terms in that document divided by the frequency of the number of documents in which that term occurs. (The concept of inverse document frequency, a measure of specificity, had been introduced in 1972 by Karen Sparck-Jones^[2]). Later in life, he became interested in automatic text summarization and analysis^[3], as well as automatic hypertext generation^[4]. He published over 150 research articles and 5 books during his life.

Salton was editor-in-chief of the Communications of the ACM and the Journal of the ACM, and chaired SIGIR. He was an associate editor of the ACM Transactions on Information Systems. He was an ACM Fellow (elected 1995), received an Award of Merit from the American Society for Information Science (1989), and was the first recipient of the SIGIR Award for outstanding contributions to study of information retrieval (1983) – now called the Gerard Salton Award.

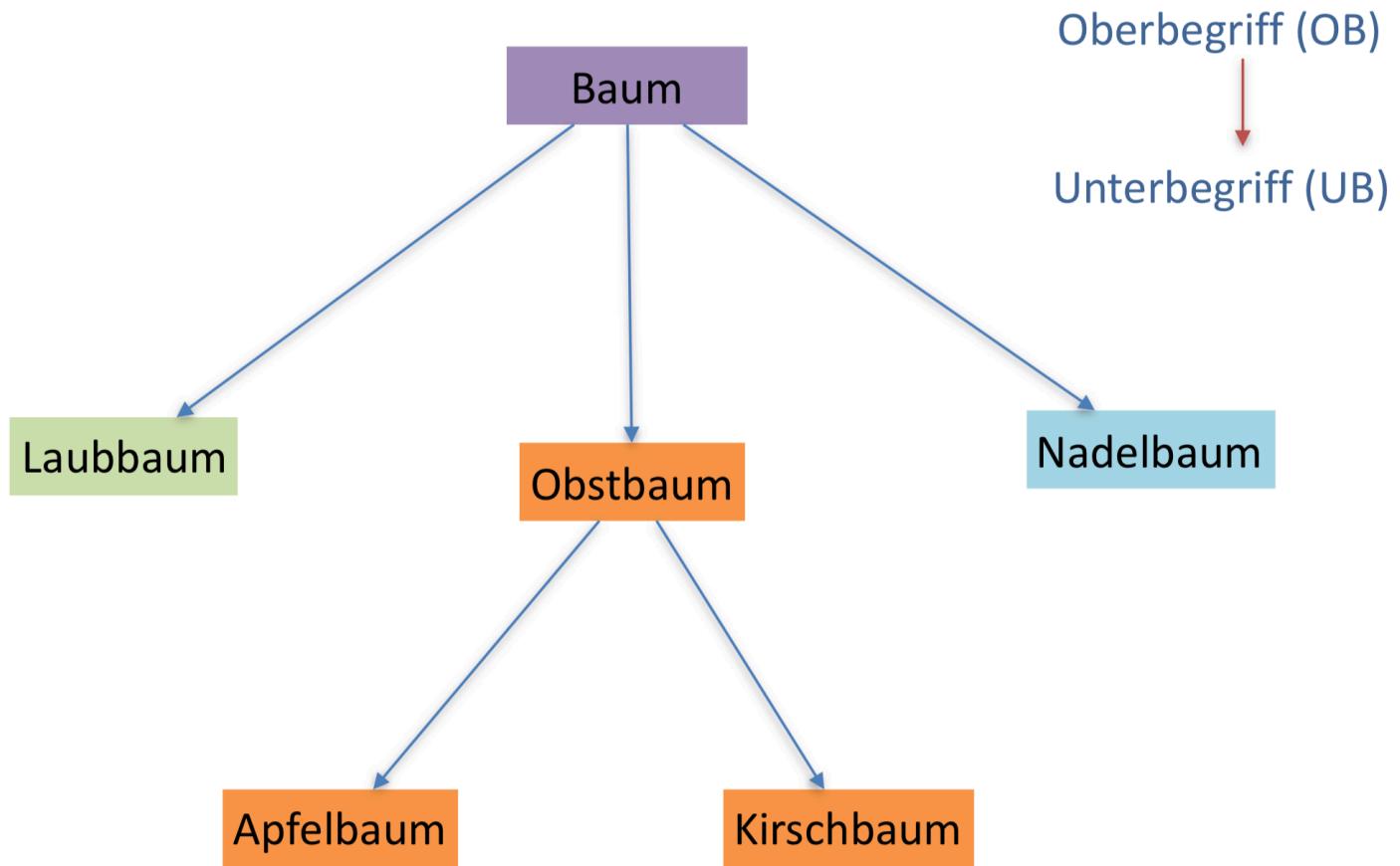
References

[edit]

1. ^ G. Salton , A. Wong , C. S. Yang, A vector space model for automatic indexing , Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975
2. ^ Sparck Jones, Karen (1972). "A statistical interpretation of term specificity and its application in retrieval" , Journal of Documentation 28 (1): 11–21, doi:10.1108/eb026526

Categories: 1927 births | 1995 deaths | American computer scientists | Computer pioneers | Harvard University alumni | Harvard University faculty | Cornell University faculty | Fellows of the Association for Computing Machinery | Guggenheim Fellows

Weitere Strukturen...



IR für Webseiten

Anfrage

Ergebnis

facebook and productivity

- ▶ [Study: Facebook use cuts productivity at work - Computerworld](#)  

www.computerworld.com › Internet › Web 2.0 and Web Apps - Cached
Jul 22, 2009 – A Nucleus Research study found that Facebook work in the workplace is cutting employee productivity.
- ▶ [Pulling the Plug on Facebook, Productivity/Time Management Article ...](#)  

www.inc.com › Leadership and Managing › Human Resources - Cached
Pulling the Plug on Facebook, Productivity/Time Management Article - All that friending and superpoking wastes a lot of time at the office -- and could be ...
- ▶ [Twitter and Facebook: The New Tools of Productivity or Distraction ...](#)  

www.briansolis.com/.../twitter-and-facebook-the-new-tools-of-prod... - Cached
Mar 26, 2010 – RT Twitter and Facebook: Tools of Productivity or Distraction RT
@PRSAcolo: Twitter & Facebook: New tools of productivity or ...
- ▶ [Twitter, Facebook Can Improve Work Productivity | PCWorld Business ...](#)  

www.pcworld.com/.../twitter_facebook_can_improve_work_produc... - Cached
Apr 2, 2009 – Reach Older Users on Facebook and Twitter · The Web's Best Productivity Sites. According to a study by the Australian University, ...
- ▶ [Is Facebook Killing Your Employees' Productivity? | WebProNews](#)  

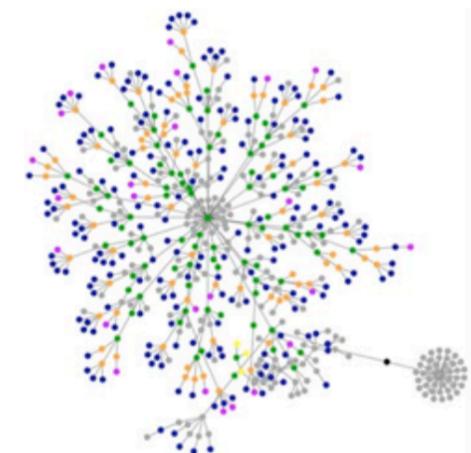
www.webpronews.com/is-facebook-killing-your-employees-produc... - Cached
Jul 21, 2009 – On the heels of a study indicating that social media can significantly impact a brand's bottom line positively, another one has come out ...
- ▶ [Productivity Strategies | Facebook](#)  

www.facebook.com/beproductive - Cached
Productivity Strategies - To learn more about the Productive Today "Content Collaborative" faculty, click the "Info" tab or this direct link: | Facebook.
- ▶ [Butt Out IT! Facebook "Productivity Loss" Is No Concern of Yours](#)  

blogs.gartner.com/.../butt-out-it-facebook-productivity-loss-is-no-co... - Cached
Facebook "Productivity Loss" Is No Concern of Yours. by Brian Prentice | November 23, 2008 | 10 Comments. Like my colleague Anthony Bradley, I also speak to ...
- ▶ [Productivity Levels Plummet After Yale Student Makes Facebook Look ...](#)  

www.betabeat.com/.../yale-student-makes-facebook-look-like-excel... - Cached
5 days ago – Productivity Levels Plummet After Yale Student Makes Facebook Look Like Excel. By Rebecca Panovka 7/28 6:11pm ...

Korpus



Webseiten

IR für digitale Bibliotheken

Anfrage

SEARCH

Ergebnis

- 1 [Effective teaching practices using free Google services: conference tutorial](#)
Paul Gestwicki, Brian McNely
 October 2010 **Journal of Computing Sciences in Colleges**, Volume 26 Issue 1
 Publisher: Consortium for Computing Sciences in Colleges
 Full text available:  [Pdf](#) (22.76 KB)
Bibliometrics: Downloads (6 Weeks): 2, Downloads (12 Months): 48, Downloads (Overall): 48,

In this 90-minute tutorial, we will share our experiences using free Web services from Google to teach effectiveness. Participants will engage with these services as part of the tutorial. We have used and studied these technologies, ...

- 2 [Model-Based Engineering of Software: Three Productivity Perspectives](#)
Shawn A. Bohner, Sriram Mohan
 October 2009 **SEW '09: Proceedings of the 2009 33rd Annual IEEE Software Engineering Workshop**
 Publisher: IEEE Computer Society
 Full text available:  [Publisher Site](#)

Bibliometrics: Downloads (6 Weeks): n/a, Downloads (12 Months): n/a, Downloads (Overall): n/a, Citation Count: 0

Evolving software products is a tricky business, especially when the domain is complex and changing rapidly. Like other fields of engineering, software engineering productivity advances have come about largely through abstraction reuse, process, and ...

Keywords: Agent-Based Software Systems, Model-Driven Architecture, Model-Driven Development, Model-Based Software Development, Model-Based Software Engineering

Korpus



Wissenschaftliche
Publikationen

IR für den Desktop

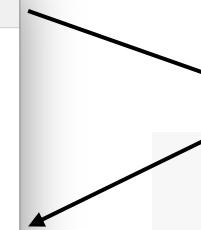
Anfrage

Ergebnis

The screenshot shows a desktop search interface with the query "facebook productivity". The results are categorized into three sections: MAIL & NACHRICHTEN, TABELLEN, and DOKUMENTE.

- MAIL & NACHRICHTEN:**
 - [SIG-IRList] CFP - SCLeM 2017: Th...
 - ACM TechNews, Wednesday, Janu...
 - diana.pacheco@gesis.org wants to...
 - ACM CareerNews Alert for Tuesday...
 - ACM MemberNet - November 19, 2...
- TABELLEN:**
 - GESIS Searchlog 2011-2014.xlsx
 - GESIS Searchlog 2011-2014.xlsx —...
 - xenu.xlsx — [SOFISwiki] Link Check...
- DOKUMENTE:**
 - top1000.csv
 - totalLog.csv
 - 2016-05-sorted.csv
 - 2016-06-sorted.csv
 - 2016-05-lilLog.csv

On the right side of the search results, there is a snippet of an email from Yadollah Yaghoobzadeh, dated March 6, 2017, about the SCLeM 2017 workshop. It includes a call for papers and submission details.



Korpus



Dateien auf
meinem Computer

Anfragen sind nur eine Krücke

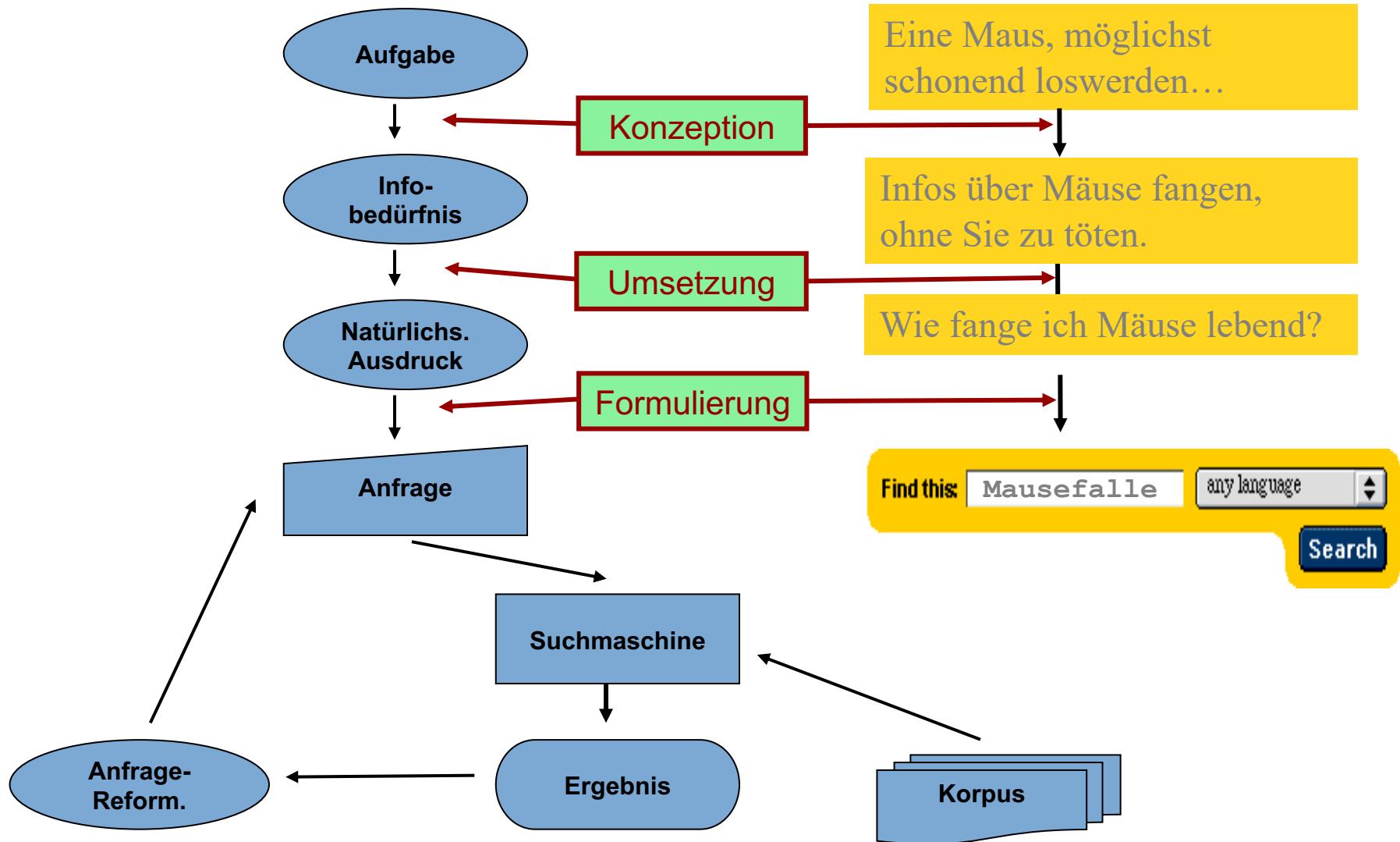
Eine Suchanfrage ist nur eine verkürzte Beschreibung des **Informationsbedürfnisses** des Nutzers.

- Mehrdeutig und eigentlich nur für den Nutzer selbst verständlich
- Eine Anfrage und ein Informationsbedürfnis setzt eine konkrete Aufgabe voraus
- Der Nutzer befindet sich in einem konkreten Kontext, den wir nicht kennen

AOL-Logdateien – User 1515830

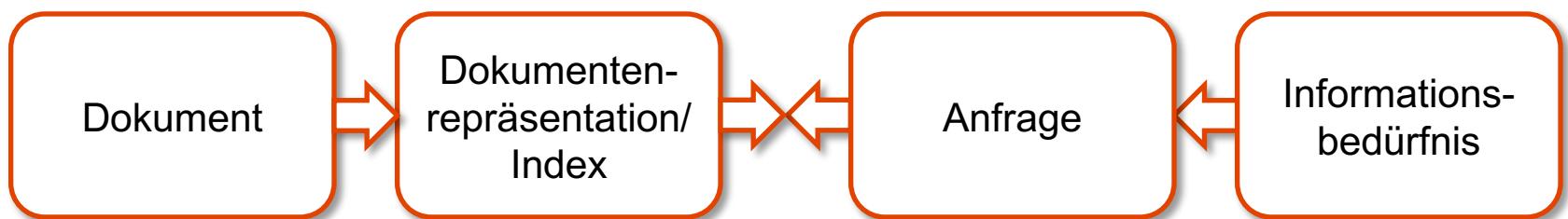
*chai tea calories
calories in bananas
aftermath of incest
how to tell your family you're a victim of incest
pottery barn
curtains
surgical help for depression
oakland raiders comforter set
can you adopt after a suicide attempt
who is not allowed to adopt
i hate men
medication to enhance female desire
jobs in denver colorado
teaching positions in denver colorado
how long will the swelling last after my tummy tuck
divorce laws in ohio
free remote keyloggers
baked macaroni and cheese with sour cream
how to deal with anger
teaching jobs with the denver school system
marriage counseling tips
anti psychotic drugs*

Anfragen sind nur eine Krücke



Klassisches Information Retrieval-Modell

- Das klassische **Ad-Hoc-Retrieval** basiert auf Abgleich von
 - Dokumenttermen (Document Representation) und
 - Anfragetermen (Query).
- Im klassischen Information Retrieval-Modell sind das Informationsbedürfnis als auch die Anfrage starr und verändern sich nicht.



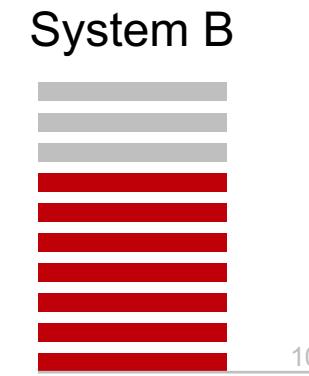
Retrieval als Suchaufgabe

The screenshot shows a search bar at the top with the query "facebook and productivity". Below the search bar is a list of ten search results, each with a title, a snippet of text, and a link. A blue arrow points downwards from the search bar towards the results.

- Study: Facebook use cuts productivity at work - Computerworld** [🔗](#)
www.computerworld.com › Internet › Web 2.0 and Web Apps - Cached
Jul 22, 2009 – A Nucleus Research study found that Facebook work in the workplace is cutting employee productivity.
- Pulling the Plug on Facebook, Productivity/Time Management Article ...** [🔗](#)
www.inc.com › Leadership and Managing › Human Resources - Cached
Pulling the Plug on Facebook, Productivity/Time Management Article - All that friending and superposting wastes a lot of time at the office -- and could be ...
- Twitter and Facebook: The New Tools of Productivity or Distraction ...** [🔗](#)
www.bransolis.com/...twitter-and-facebook-the-new-tools-of-prod... - Cached
Mar 26, 2010 – RT Twitter and Facebook: Tools of Productivity or Distraction RT @PRSAcoto: Twitter & Facebook: New tools of productivity or ...
- Twitter, Facebook Can Improve Work Productivity | PCWorld Business ...** [🔗](#)
www.pcworld.com/...twitter_facebook_can_improve_work_produc... - Cached
Apr 2, 2009 – Reach Older Users on Facebook and Twitter · The Web's Best Productivity Sites. According to a study by the Australian University, ...
- Is Facebook Killing Your Employees' Productivity? | WebProNews** [🔗](#)
www.webpronews.com/is-facebook-killing-your-employees-produc... - Cached
Jul 21, 2009 – On the heels of a study indicating that social media can significantly impact a brand's bottom line positively, another one has come out ...
- Productivity Strategies | Facebook** [🔗](#)
www.facebook.com/beproductive - Cached
Productivity Strategies - To learn more about the Productive Today "Content Collaborative" faculty, click the "Info" tab or this direct link | Facebook.
- Butt Out IT! Facebook "Productivity Loss" Is No Concern of Yours** [🔗](#)
blogs.gartner.com/...butt-out-it-facebook-productivity-loss-is-no-concern-of-yours... - Cached
Facebook "Productivity Loss" Is No Concern of Yours, by Brian Prentice | November 23, 2008 | 10 Comments. Like my colleague Anthony Bradley, I also speak to ...
- Productivity Levels Plummet After Yale Student Makes Facebook Look ...** [🔗](#)
www.betabeat.com/...yale-student-makes-facebook-look-like-excel... - Cached
5 days ago – Productivity Levels Plummet After Yale Student Makes Facebook Look Like Excel. By Rebecca Panovka 7/28 6:11pm ...

- **Ausgabe:** Ein **Ranking** von Dokumenten, in absteigender Reihenfolge Ihrer geschätzten Relevanz (macht es einfacher!).
- **Annahme:** Der Benutzer schaut sich die **ersten paar Dokumente** an und ist zufrieden, wenn er etwas passendes gefunden hat.

Ranking – Welches ist besser?



- Relevante Treffer sind rot markiert.

Warum ist IR eine schwierige Aufgabe?

Information Retrieval ist ein Prozess mit **Unsicherheiten**...

- Benutzer **wissen nicht was sie eigentlich wollen**
- Benutzer wissen nicht, wie sie das was sie suchen **ausdrücken sollen**
- Computer können Nutzer keine **Kontextinformationen** entlocken, wie es z.B. ein menschlicher Bibliothekar könnte
- Computer verstehen **keine natürliche Sprache**
- Suchmaschinen müssen **erraten, was relevant ist**
- Suchmaschinen müssen erraten, wann ein Benutzer **zufrieden** ist
- ...

Suchanfragen, Relevanz, Ranking...

Dies alles macht Information Retrieval so **schwer** – und **faszinierend**!

- Eigentlich müssten wir viel mehr „verstehen“, was der Nutzer will.

Im IR geht es aber nicht darum Text zu verstehen, sondern darum „zu bestimmen“/„zu berechnen“/„vorherzusagen“ ob ein Dokument zu einer Anfrage relevant ist – oder nicht.

- Das ist in vielen Fällen **viel einfacher**, als „verstehen“.
- Es ermöglicht uns, das Problem an **Computer abzugeben**, z.B. in Form von Suchmaschinen!
- Computer arbeiten mit **Modellen**, um Informationen zu finden, z.B. dem **booleschen Modell...**
- **Reicht aber oft genug aus!**

Hausaufgabe für nächste Woche

- **Bearbeiten Sie den Übungszettel 1 zuhause und bringen Sie Ihre Lösungen/Bearbeitungen ausgedruckt mit in die Labore.**
- Bereiten Sie sich darauf vor, die Aufgaben **an der Tafel oder am Laptop interaktiv zu lösen/vorzustellen**, in einem Peer-Review mitzumachen, oder oder...
- Die Lösungen werden dann im Plenum diskutiert, erweitert und in den Kontext gesetzt.

IR4: Information Retrieval Laborsitzung 10.10.2019

Philip Scher, Technische Hochschule Köln

Bearbeiten Sie die Aufgaben zuhause. In den Laborterminen erhalten wir dann mit den Ergebnissen Ihrer Vorarbeiten. Z.B. stellt jeder aus Ihren Reihen die vorbereitete Lösung vor. Das Lösung kann aber später im Plenum diskutiert, erweitert und in Kontext gesetzt werden. Auch Peer-Reviews sind möglich.

Die Präsentation einer Lösung an der Tafel, ein Kurzerfall, ein Vorerechnen, usw. wird von jedem von Ihnen mindestens einmal im Semester erwartet. Es gilt die Punkteregelung, wie Sie in der Vorlesung vorgestellt wurde.

Aufgabe 1: Textlektüre und Einsortierung des Gelehrten
Lesen Sie das Kapitel „Textliche Relevanz“ der Lehrbuchvorlesung von Dr. phil. Peer Ferber.

Beantworten Sie die folgenden Fragen in kleinen Mini-Essays von jeweils ca. 200 Wörtern (insgesamt ca. 1,5 DIN A4-Seiten) auf der Grundlage von Ferbers Text und des Inhalts der heutigen Vorlesung.

1. Was für einen Unterschied im Modell beschreibt die Abbildung 2.3 im Gegensatz zum vorgelegten Ad-Hoc-Modell aus dem Vorlesung?
2. Gehen Sie die Begriffe Daten, Information und Wissen voneinander ab.
3. Diskutieren Sie die Bedeutung der Begriffe Datenspeicher, Information Retrieval und Wissensorganisation vor dem Hintergrund der Begriffserklärungen.
4. Was sind die Haupt-Unterschied zwischen den Stufen in einer Datenbank und einem Retrieval-System?

Bringen Sie Ihre Mini-Essays bitte ausgedruckt mit in die Übung vom 17.10.2019.