Einführung in Information Retrieval

Gerard Salton, 1983:

"Information Retrieval beschäftigt sich mit der Repräsentation, Speicherung und Organisation von Informationen und dem Zugriff auf Informationen. Im Regelfall bestehen die gespeicherten Informationen aus Texten (Bücher, Zeitungsartikel, Briefe, Berichte und sonstige Dokumente)."

Gerald Kowalski, 1997:

"An Information Retrieval system is a system that is capable of storage, retrieval, and maintenance of information.

Information in this context can be composed of **text** (incl. numeric and date data), **images**, **audio**, **video** and **other multi-media objects**.

Although the form of an object in an IR System is diverse, the text aspect has been the only data type that lent itself to full functional processing.

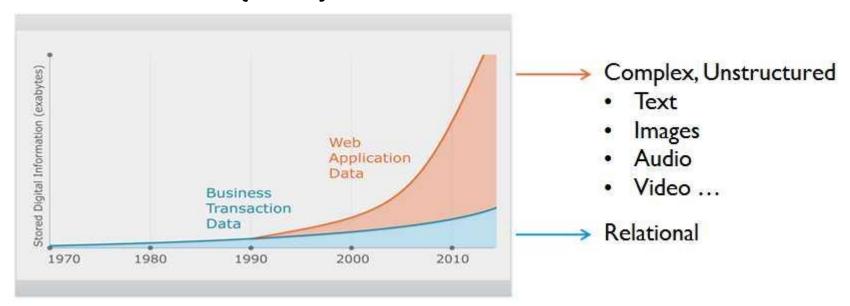
The other data types have been treated as highly informative sources, but are primarily linked for retrieval based upon search of text.

Techniques are beginning to emerge to search these other media types."

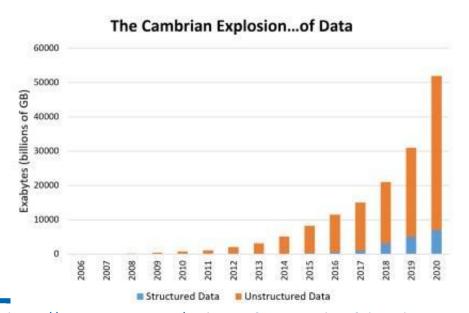
Christopher D. Manning, 2008:

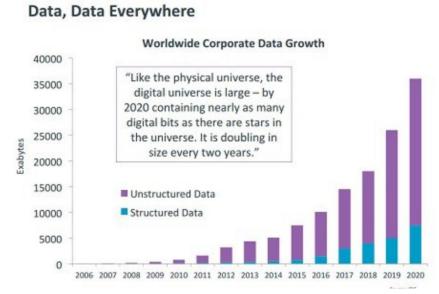
"Information Retrieval (IR) is **finding** material (usually documents) of an **unstructured nature** (usually **text**) that satisfies an information need from within large collections (usually stored on computers)."

Wachstum (un-)strukturierter Daten



http://bigdataintegration.blogspot.de/2012/02/unstructured-data-is-myth.html







GI-Fachgruppe Information Retrieval:

Gegenstand des IR: Informationssysteme in Bezug auf ihre Rolle beim Wissenstransfer vom menschlichen Wissensproduzenten zum Informations-Nachfragenden

Schwerpunkt sind Fragestellungen, die im Zusammenhang mit vagen Anfragen und unsicherem Wissen entstehen.

Vage Anfragen sind dadurch gekennzeichnet, dass die Antwort a priori nicht eindeutig definiert ist. Hierzu zählen:

- Fragen mit unscharfen Kriterien
- Fragen, die nur im Dialog **iterativ** durch Reformulierung beantwortet werden können.

Die Unsicherheit dieses Wissens resultiert meistens aus der begrenzten Repräsentation der Semantik, der Unsicherheit oder der Unvollständigkeit der gespeicherten Daten.

Ein weiterer wichtiger Punkt ist die Bewertung der Qualität der Antworten eines Informationssystems.

http://fg-retrieval.gi.de/

Retrieval-Szenarien: Adhoc-Suche

"Liefere Dokumente, die die Terme «Information» und «Retrieval» enthalten."

Unrealistisch: niemand würde solche Dokumente suchen!

Google information retrieval solche Dokumente such Alle News Bilder Bücher Videos Mehr ▼ Suchoptionen

Ungefähr 14.800.000 Ergebnisse (0,56 Sekunden)

Zweck der Anfrage? Informationsbedürfnis?

Information Retrieval – Wikipedia

https://de.wikipedia.org/wiki/Information_Retrieval ▼
Information Retrieval [ˌɪnfə-ˈmeɪʃən ɹɪˈtɹiːvəl] (IR) bzw. Informationsrückgewinnung, gelegentlich ungenau Informationsbeschaffung, ist ein Fachgebiet, ...
Anwendungsbereich - Geschichte - Grundbegriffe - Relevanz und Pertinenz

Information retrieval - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/Information_retrieval ▼ Diese Seite übersetzen Information retrieval (IR) is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches ...

ECIR 2016 - 38th European Conference on Information ...

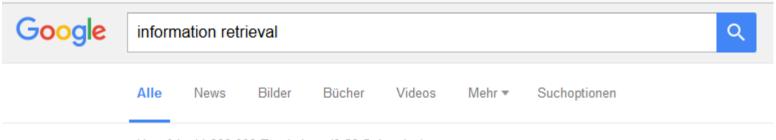
ecir2016.dei.unipd.it/ ▼ Diese Seite übersetzen

ECIR. The annual European Conference on Information Retrieval is the premier European forum for the presentation of new research results in the field of ... Sie haben diese Seite 3 Mal aufgerufen. Letzter Besuch: 07.12.15

[PDF] Information Retrieval 1 - Universität Bamberg

Retrieval-Szenarien: Adhoc-Suche

"Liefere Dokumente, die Lehrveranstaltungen über «Information» und «Retrieval» beschreiben."



Ungefähr 14.800.000 Ergebnisse (0,56 Sekunden)

Information Retrieval – Wikipedia

https://de.wikipedia.org/wiki/Information_Retrieval

Information Retrieval [ˌɪnfəˈmeɪʃən ˈtaɪˌvəɪ] (IR) bzw. Informationsrückgewinnung, gelegentlich ungen taller informationsbeschaffung, ist ein Fachgebiet, ...

* ...endungsbereich - Geschichte - Grundbegriffe - Relevanz und Pertinenz

Information retrieval - Wikipedia, the free encycle sola

ECIR 2016 - 38th European Conference on Information ecir2016.dei.unipd.it/ ▼ Diese Seite übersetzen

ECIR. The annual European Conference on L. Connation Retrieval is the premier European forum for the processing and an experience of the second secon

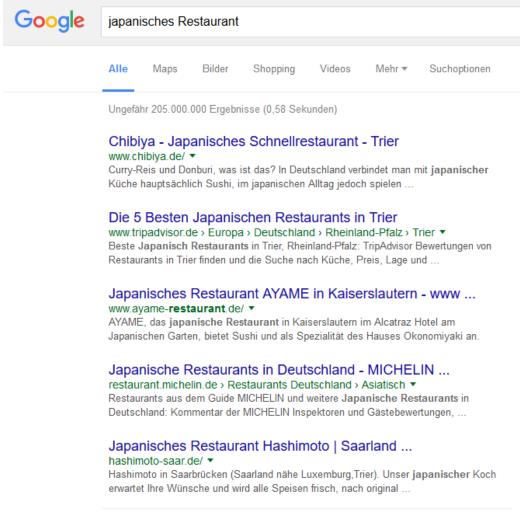
Ranking!

Evaluierung der Ergebnisqualität!

[PDF] Information Retrieval 1 - Universität Bamberg
https://www.uni-bamberg.de/fileadmin/uni/.../wiai.../henrich-ir1-1.2.pdf

Retrieval-Szenarien: Ortsabhängige Suche

"Finde japanisches Restaurant" (in der Nähe)



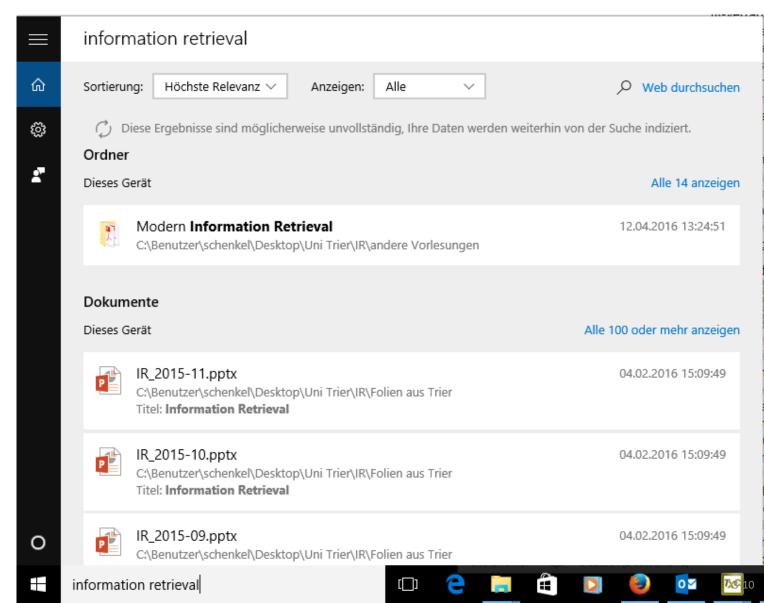
Bewertungen anderer

Benutzer



Retrieval-Szenarien: Desktop-Suche

"Finde Dokumente auf meinem Rechner über Information Retrieval"



"Wann wurde Barack Obama geboren?"



Entitäten und ihre Beziehungen

Barack Hussein Obama II ... wo Obamas jüngere Halbschwester Maya geboren wurde. Dort besuchte Obama von 1967 bis 1970 die von Kapuzinern geführte St ... Familie, Kindheit ... · Studium und Beruf · Abgeordneter in Illinois · US-Senator

Barack Obama (1961) · geboren.am

geboren.am > Personen > O

10/10 ****

Wo wurde Barack Obama geboren? Obama wurde in Honolulu, Hawaii in den

Anmelden Dokumente vs. **Faktenwissen**

Barack Obama



Barack Hussein Obama II ist ein US-amerikanischer Politiker und seit dem 2 Januar 2009 der 44. Präsident der Vereinigte Staaten







Wikipedia

Twitter

Geboren: 4. Aug 1961 (Alter 54) · Honolulu, HI

Größe: 1.85 m

Ehepartner: Michelle Obama (H. 1992)

Kinder: Malia Ann Obama (Tochter) · Natasha Obama (Tochter)

Vorgänger: George W. Bush

Amt: Präsident der Vereinigten Staaten (Seit 2009)

Werke









"Wann wurde Barack Obama geboren?" – 6 Monate später



wann wurde barack obama geboren



English Anmelden

Web Bilder Videos Karten News Erkunden

212.000 ERGEBNISSE

Datum 🕶

Sprache *

Region •

Barack Obama · Geboren



4. Aug 1961



Hillary Clinton 26. Okt 1947



Donald Trump 14. Jun 1946



George W. Bush 6. Jul 1946

Barack Obama – Wikipedia

https://de.wikipedia.org/wiki/Barack_Obama >

Barack Hussein Obama II ... wo Obamas jüngere Halbschwester Maya geboren wurde. Dort besuchte Obama von 1967 bis 1970 die von Kapuzinern geführte ...

Familie, Kindheit ... · Studium und Beruf · Abgeordneter in Illinois · US-Senator

Barack Obama (1961) · geboren.am

geboren.am > Personen > O

10/10 ****

In welchem Sternzeichen wurde Obama geboren? Barack Obama wurde im westlichen

Barack Obama



Barack Hussein Obama II ist ein USamerikanischer Politiker und seit dem 20. 2009 der 44. Präsident der Vereinigten Sta



Offizielle Website

W

Wikipedia

Facebook

Geboren: 4. Aug 1961 (Alter 55) · Honolulu, HI

Größe: 1,85 m

Reinvermögen: \$12,20 Millionen USD (2016)

Ehepartner: Michelle Obama (H. 1992)

Kinder: Malia Ann Obama (Tochter) · Natasha Obama (Tochte

Eltern: Barack Hussein Obama Senior (Vater) · Ann Dunham (

Werke





"Wann wurde Barack Obama geboren?" – 2018



Politiker der Demokraten, der als erster Afroamerikaner Präsident der USA (2009–2017) war, in seiner Amtszeit u. a. "Obamacare" einführte, den Drohnenkrieg gegen Terrorführer forcierte, sich mit der NSA-

Affäre konfrontiert sah und mit dem Friedensnobelpreis 2009 geehrt wurde.

Autor: Bewertung ermittelt von geboren.am



"Wann wurde Ralf Schenkel geboren?"



wann wurde ralf schenkel geboren



English

Web Bilder Videos Karten News Erkunden

604,000 ERGEBNISSE

Datum ▼

Sprache *

Region *

Ralf Schumacher (1975) · geboren.am

geboren.am > Personen > S

5,5/10 ****

Wann hat Ralf Schumacher Geburtstag? Schumacher hat im Sommer am 30. Juni Geburtstag. Er wurde an einem Montag geboren. ... Wo wurde Ralf Schumacher geboren?

Rolf Zacher (1941) · geboren.am

geboren.am > Personen > Z

8,5/10 ****

Rolf Zacher ist ein deutscher Schauspieler und wurde 1941 geboren. ... Wann wurde Rolf Zacher geboren? ... Ralf Zeigermann, ...

Ralf Richter » Steckbrief | Promi-Geburtstage.de

www.promi-geburtstage.de/info/?id=6186 Ralf-Richter

Wann wurde Ralf Richter geboren? Ralf Richter wurde am Carl Schenkel) 1984 – Decoder ... Wo wurde Ralf Richter geboren ...

Ähnliche Suchen

Wann Wurde Die Bundeswehr Gegründet

Wann Wurde Microsoft Gegründet

Wann Wurde Die DDR Gegründet

Wann Wurde Die EU Gegründet

Wann Wurde Jesus Geboren Datum

Wann Wurde der Euro Eingeführt

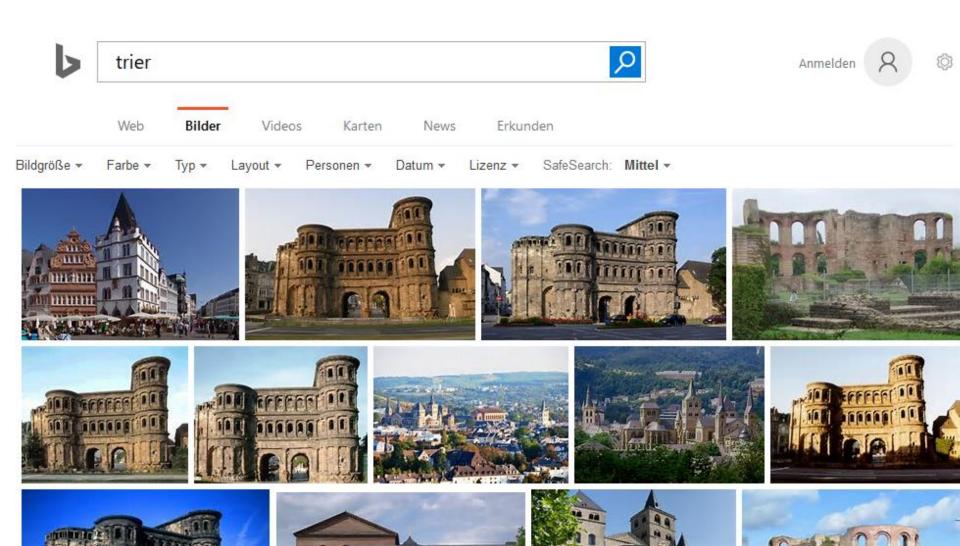
Wann Wurde Jesus Gekreuzigt

Wann Wurde Jesus Geboren

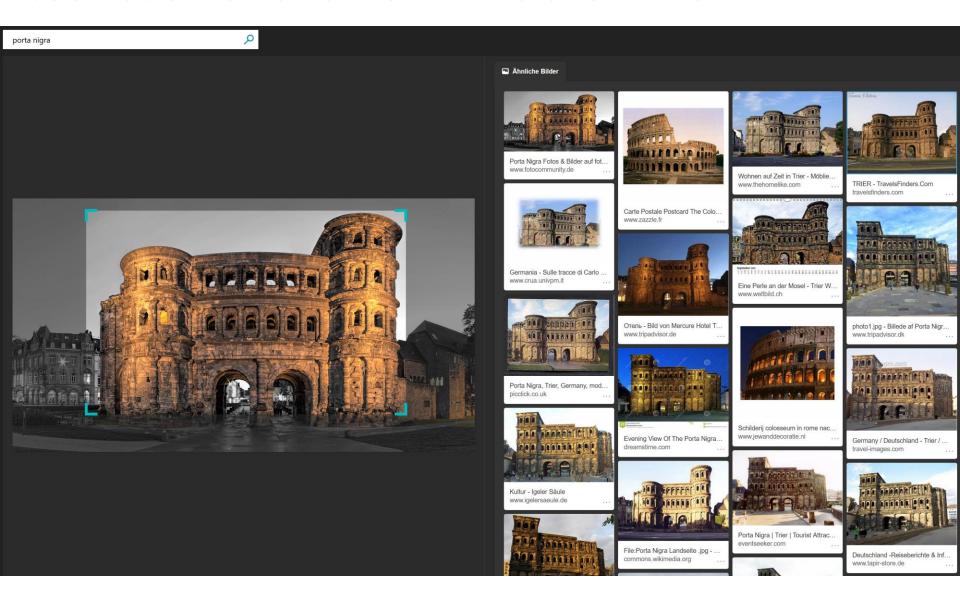
Such-Hilfen für den Benutzer

Ähnliche Suchvorgänge für **wann wurde ralf schenkel ge...**Wann Wurde Die Bundeswehr Gegrü... Wann Wurde Die EU Gegründet

"Finde (schöne, nützliche, ...) Bilder über Trier"







Ungefähr 62 Ergebnisse (0,67 Sekunden)



Bildgröße: 200 × 200

Dieses Bild in einer anderen Größe suchen: Alle Größen - Mittel

Vermutung für dieses Bild: gregoire bautier

Seiten mit übereinstimmenden Bildern

Marie Nossereau - Bilder, News, Infos aus dem Web

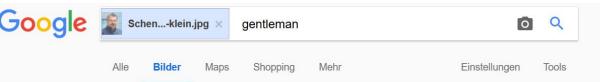


www.vebidoo.de/marie+nossereau ▼
473 × 472 - Freunde und Netzwerk. Laurence Bonnet. In den eigenen Kreisen:
190. Jobs und Berufe: chef de groupe in TBWA\. Greg Bautier. account director. Ausbildung:

Greg Bautier - Google+



https://plus.google.com/1144409320704... ▼ Diese Seite übersetzen 473 × 472 - Greg Bautier - Megalo & Co. - New York - Paris - Waterford, irlande - Toulouse - gregoire hautier changed his profile photo. Shared



Ungefähr 2 Ergebnisse (0,60 Sekunden)



Bildgröße: 1691 × 1906

Keine anderen Größen für dieses Bild gefunden.

Vermutung für dieses Bild: gentleman

Gentleman (Musiker) - Wikipedia

https://de.wikipedia.org/wiki/Gentleman_(Musiker) •

Gentleman (* 19. April 1975 in Osnabrück; bürgerlich Tilmann Otto) ist ein deutscher Reggae-Musiker. Inhaltsverzeichnis. 1 Jugend und Familie; 2 Karriere.

Gentleman - Wikipedia

https://de.wikipedia.org/wiki/Gentleman ▼

Der Begriff Gentleman bezeichnet einen aufgrund seiner Geburt, seines Charakters oder anderer Umstände in irgendeiner Form sozial herausgehobenen Mann ...

Optisch ähnliche Bilder



Unangemessene Bilder melden

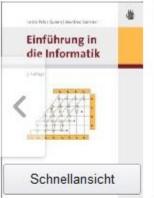
Gentleman Der Begriff Gentleman bezeichnet einen aufgrund seiner Geburt, seines Charakters oder anderer Umstände in irgendeiner Form sozial herausgehobenen Mann und ist ins Deutsche am ehesten zu übersetzen mit Ehrenmann oder Kavalier. Wikipedia Andere suchten auch nach Esq. Esq. Esquire Ritter Vicomte Markgraf

Feedback

Retrieval-Szenarien: Empfehlungen

Finde "interessante Dinge"

Inspiriert von Ihren Stöber-Trends Mehr









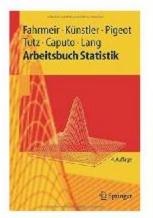


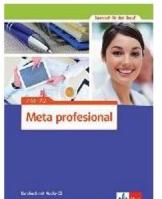


Mehr Top-Empfehlungen für Sie













Retrieval-Szenarien: Empfehlungen





















Retrieval-Szenarien: Recherche

"Erstelle eine Ausarbeitung über die Rolle von McDonalds in der europäischen Esskultur"



Retrieval-Szenarien: Recherche



Herausforderungen an IR-Systeme

- Speicherung und effizienter Zugriff auf riesige Datenmengen
- effiziente und effektive Suche
- komplexe Suchanfragen (Queries)
- Bewertung und Vergleich von Anfragen und Suchergebnissen
- (visuelle) Aufbereitung von Suchergebnissen, Navigation, Benutzerführung
- einfaches Textverstehen
- automatische Textsynthese

Begriffsbildung

Suche in Dokumentkollektionen kann auf verschiedenen Abstraktionsstufen stattfinden. Vergleiche hierzu die Ebenen der Semiotik:

Syntax

Ein Dokument wird als Folge von Symbolen betrachtet. Beispiele:

- Zeichenkette in Texten
- Histogramm oder Kontur in Bildern

Semantik

Ein Dokument wird auf der Ebene seiner Bedeutung betrachtet. Semantik hat immer etwas mit Interpretation zu tun.

Pragmatik

Ein Dokument wird hinsichtlich seines Verwendungszusammenhangs betrachtet. Beispiele:

- Enthält ein Dokument eine Lösung meines Problems?
- Was ist die Absicht des Autors des Textes?

Daten, Information, Wissen



Beispiel: Datenbank als Sammlung von Werten Beispiel: Interpretation der Werte in einer Datenbank Transformation von Information, um das benötigte Wissen zu erhalten.

Wissen ist die Information, die in einer konkreten Situation zur Lösung eines Problems benötigt wird.

Achtung: Manche Autoren wenden auch Information und Wissen in vertauschten Rollen an!

Vergleich Daten- und Information Retrieval

	Daten-Retrieval	Text-IR
Matching	exakt	partieller Match, bester Match
Inferenz	Deduktion	Induktion
Modell	deterministisch	probabilistisch
Klassifikation	monothetisch	polithetisch
Anfragesprache	formal	natürlich
Fragespezifikation	vollständig	unvollständig
gesuchte Objekte	Fragespezifikation erfüllend	relevante
Reaktion auf Datenfehler	empfindlich	robust

[Rijsbergen 1979, Fuhr 2004]

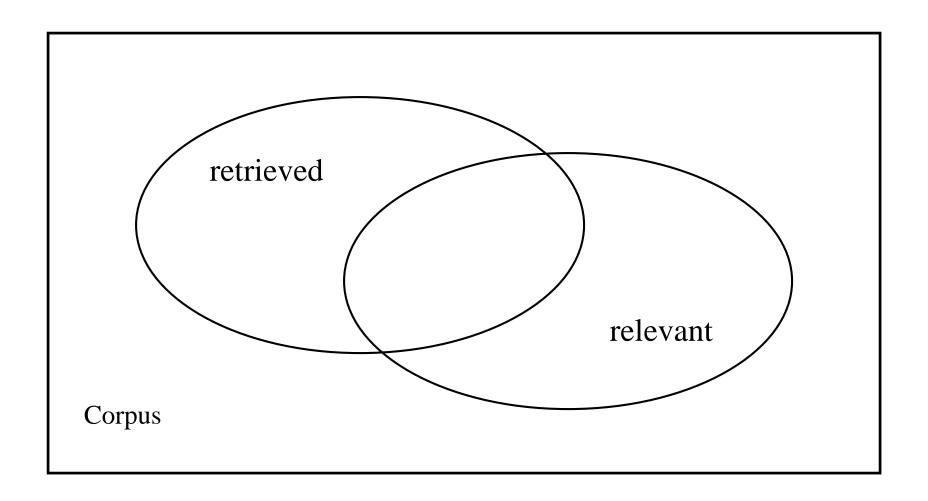
Ziel: Retrieval von relevantem Wissen

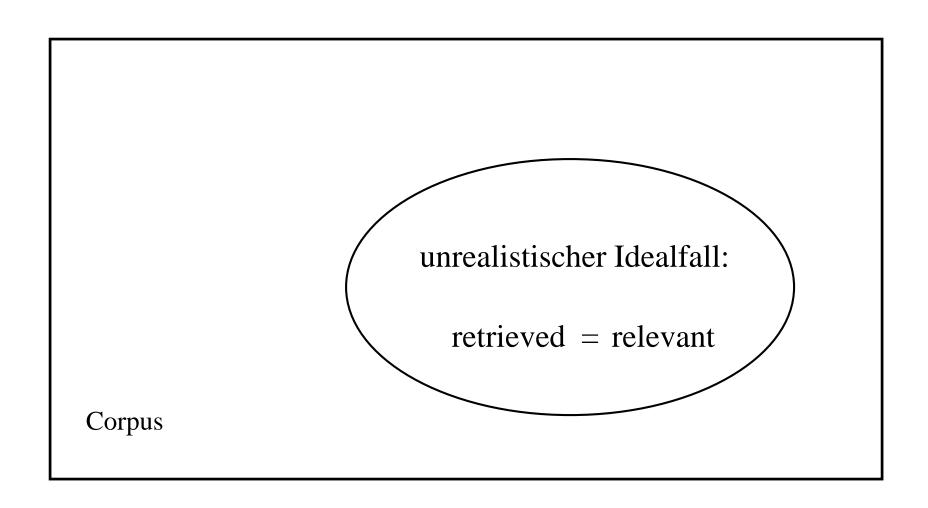
Relevanz ist abhängig

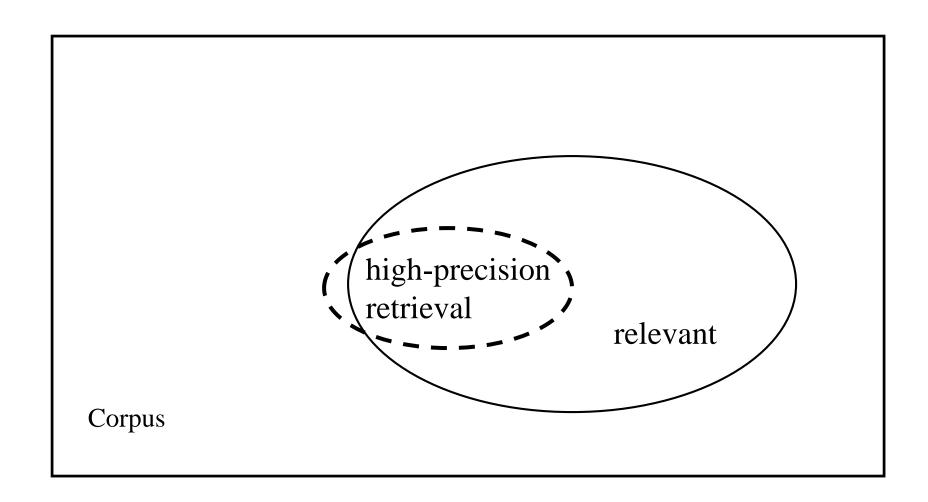
- vom aktuellen Wissen des Benutzers
- vom aktuellen Problem des Benutzers
- von der subjektiven Erwartung des Benutzers

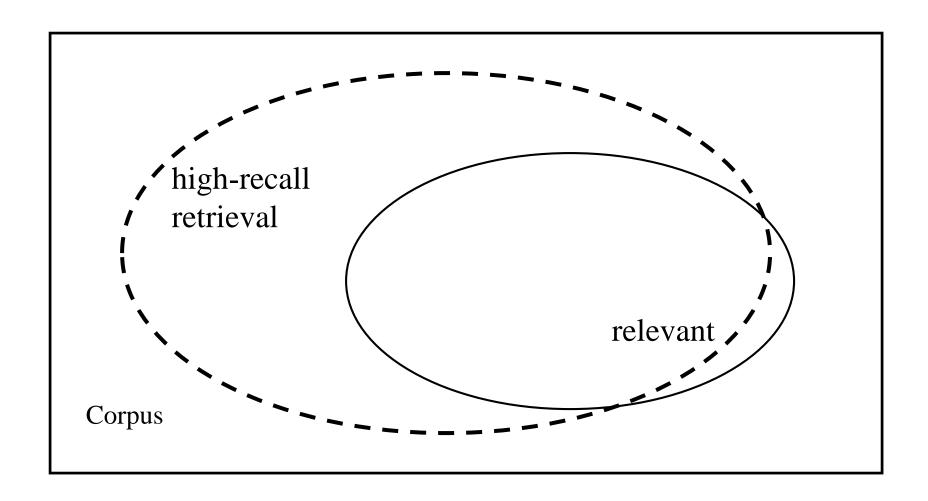


- Suchmaschinen- bzw. IR-System-Problematik
- Nicht alle Dokumente, die der Benutzer für relevant erachten würde, werden aufgefunden.
- Nicht alle aufgefundenen Dokumente werden vom Benutzer für relevant gehalten.









$$recall \equiv \frac{|retrieved \cap relevant|}{|relevant|}$$
 $precision \equiv \frac{|retrieved \cap relevant|}{|retrieved|}$



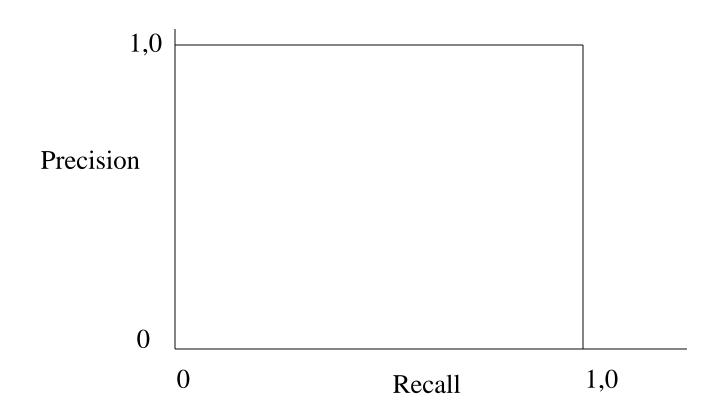
Precision

- erfordert nur die Analyse des Retrieval-Resultats
- kann vom Endbenutzer eingeschätzt werden
- ist ein subjektives Maß

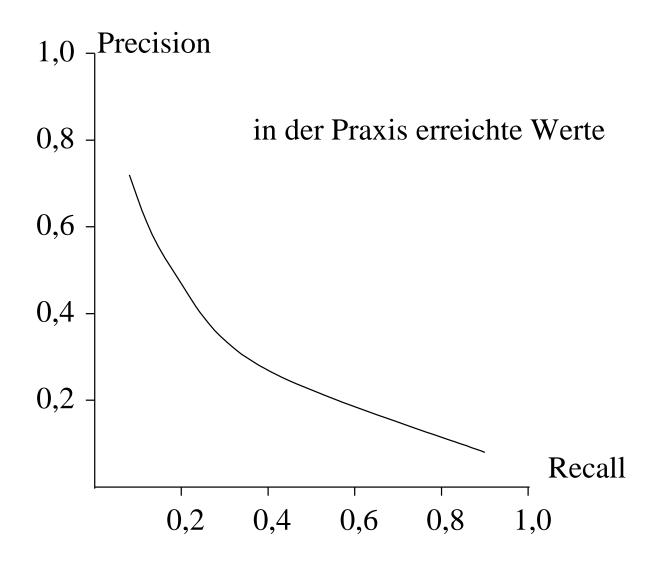
Recall

- erfordert die Analyse der gesamten Dokumentenbasis
- ist dem Endbenutzer nicht zugänglich
- ist ein subjektives Maß

Idealfall

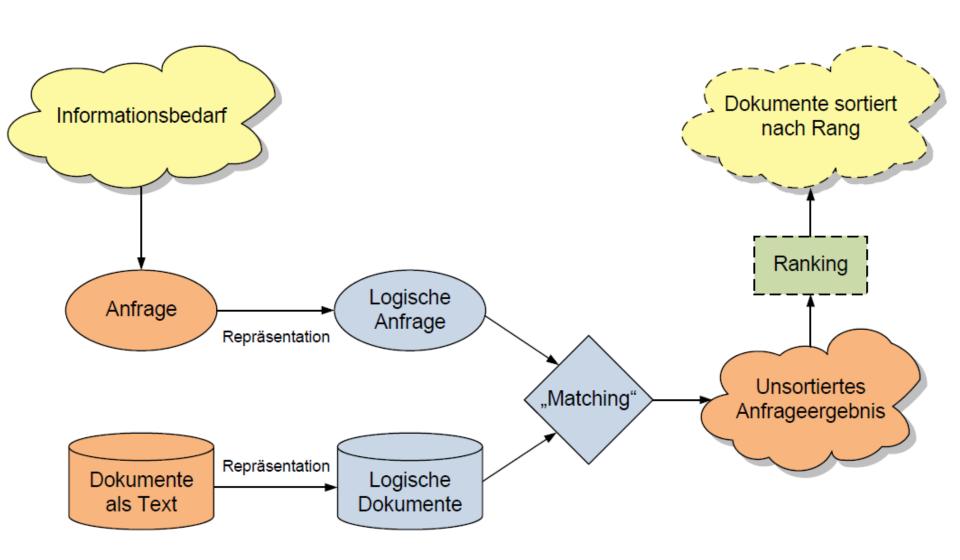




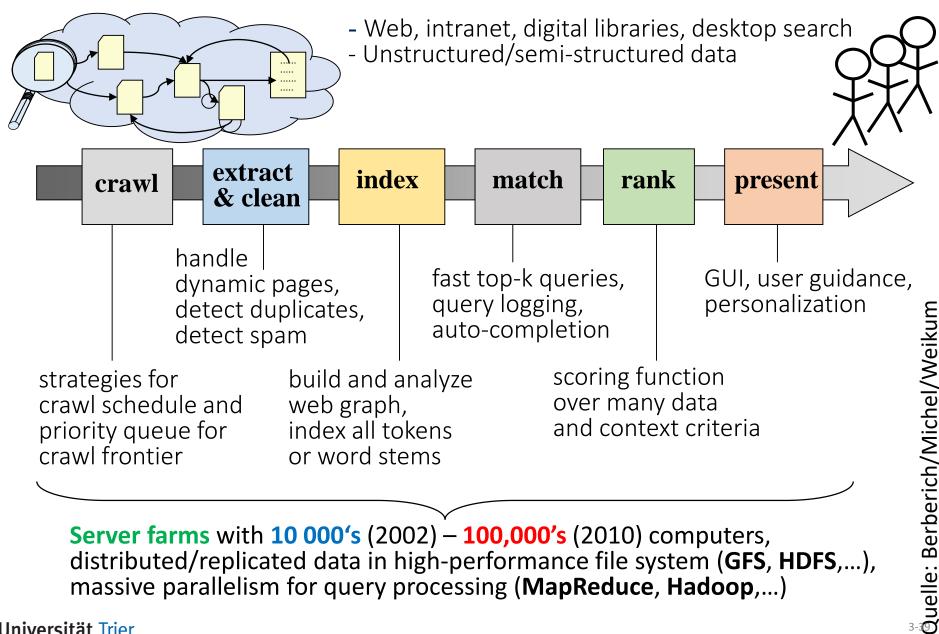




Abstrakter Information Retrieval Prozess



Systemorientierte Prozess-Sicht



Server farms with 10 000's (2002) - 100,000's (2010) computers, distributed/replicated data in high-performance file system (GFS, HDFS,...), massive parallelism for query processing (MapReduce, Hadoop,...)

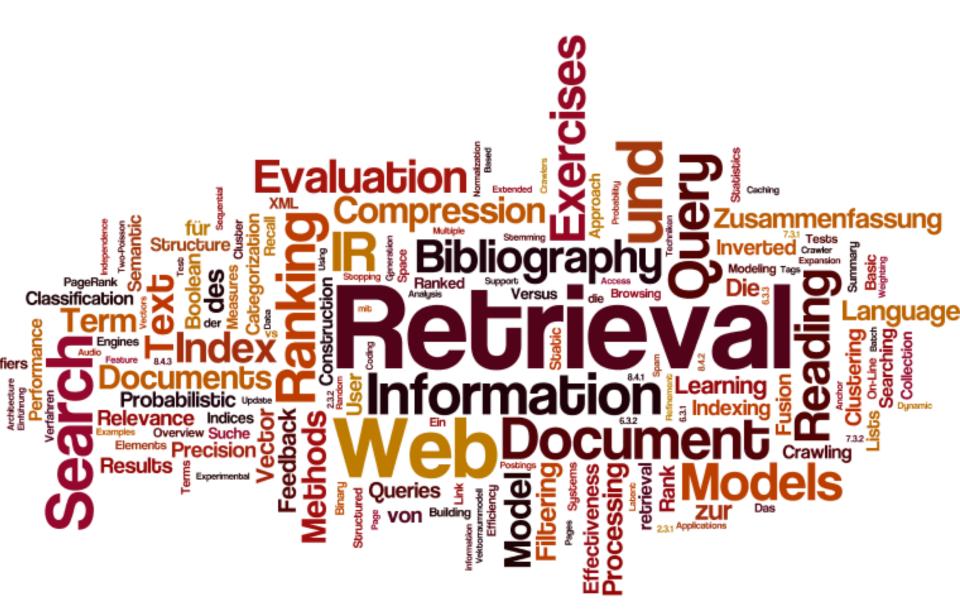
Methoden und Techniken des IR

- Modellierung von Dokumenten und Text DL+IR
- (approximatives) String-Matching IIR
- Textvorverarbeitung und Indexing DL+IR IIR
- Benutzerinteraktion und Visualisierung DL+IR
- Benutzermodellierung und Personalisierung DL+IR
- Relevanzanalyse DL+IR
- verteilte und Peer-to-Peer Softwaretechnik IIR
- Kategorisierung, Klassifikation Data Mining
- Natural Lange Processing (NLP) (Computer-) Linguistik DL+IR
- Web-Technologie DL+IR
- Datenstrukturen, effiziente Symbolverarbeitung IIR

Universität Trier

3-40

Information Retrieval hat viele Aspekte



Universitat irier