



Introduction to Information Retrieval

Ulf Leser

Content of this Lecture

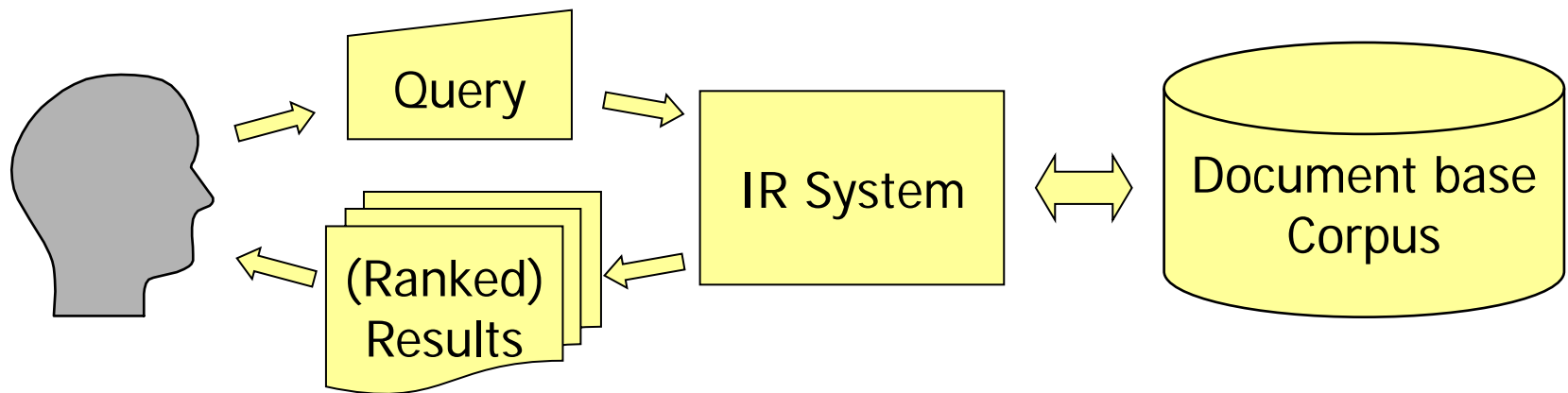
- What is Information Retrieval
- Documents
- Queries
- Related topics

Information Retrieval (aka "Search")

- Naïve: Find all **documents** containing the following **words**
- Advanced: „Leading the user to those documents that will best enable him/her to satisfy his/her **need for information**“ [Robertson 1981]
 - A user wants to know something
 - The user needs to tell the machine what he wants to know: query
 - Posing exact queries is difficult: room for interpretation
 - **Machine interprets query** to compute the (hopefully) best answer
 - Goodness of answer (relevance) depends on original intention of user, not on the query
 - Answer is always a set of docs
 - “Leading”: Sensible **ranking** of all potentially relevant docs

The Informal Problem

- Help user in **quickly** finding the **requested information** within a **given set of documents**
 - Set of documents: **Corpus**, library, collection, ...
 - Quickly: **Few queries**, **fast responses**, simple interfaces, ...
 - Requested: The “best-fitting” documents; the “most relevant” content



Difference to Database Queries

- Queries: Formal language versus **natural language**
- Result granularity: Set of **documents** versus relation as defined by query
- Exactly defined result versus loosely described **relevance**
- Result set versus **ranked result list**
- DB: Posing the **right query** is completely left to the user
- IR: Understanding the query is a **problem of the software**

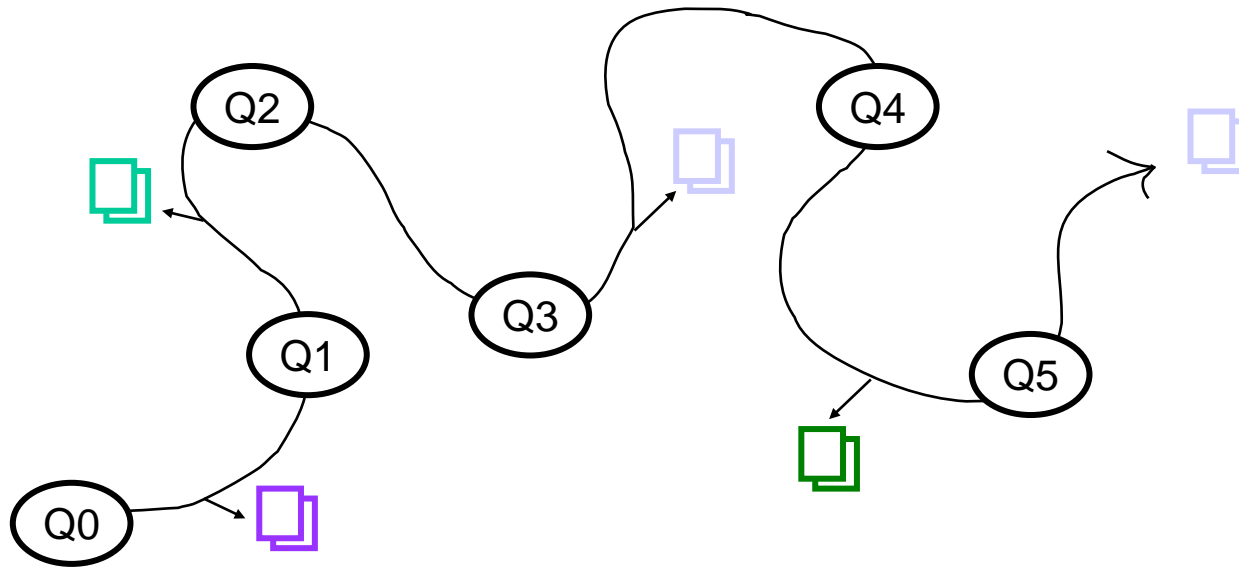
Why is it hard?

- Properties of human languages
 - Homonyms (context): Fenster (Glas, Computer, Brief, ...), Berlin (BRD, USA, ...), Boden (Dach, Fussboden, Ende von etwas, ...)
 - Synonyms: Computer, PC, Rechner, Desktop, Laptop, Tablet, ...
- Properties of the corpus
 - Size: Corpora today may have **billions of documents**
 - Heterogeneity: **Length**, format, language, genre, **grammatical correctness**, special characters, ...
- Heterogeneous users: **Precise** queries versus **usability**
 - Lay persons: Short queries with **wide spectrum** of interpretations
 - Average web queries have 1,6 terms
 - Additional knowledge: Location, current trends, popular answers, ...
 - Professionals: Long queries trying to precisely define the intention
 - “Information broker” was/is a profession

Quickly

- Time to **execute a query**
 - Indexing, parallelization, compression, ...
- Time to **answer the request** (may involve multiple queries)
 - Understand request, find best matches
 - Success of search engines: Better results (and fast!)
 - **Process-orientation**: User feedback, query history, ...
- Information overload
 - “We are drowning in data, but starving for knowledge”
 - If the corpus is large, **ranking is a must**
 - Alternative: Result summarization (grouping on what?)
 - Different **search modes**: What’s new? What’s certain?

IR: An Iterative, Multi-Stage Process



- IR process: “Moving through many actions towards a general goal of satisfactory completion of research related to an information need.”
 - “Berry-picking” [Bates 89]

- Im Information Retrieval (IR) werden Informationssysteme in Bezug auf ihre Rolle im **Prozess des Wissenstransfers** vom menschlichen Wissensproduzenten zum Informations-Nachfragenden betrachtet. ... Fragestellungen, die im Zusammenhang mit **vagen Anfragen und unsicherem Wissen** entstehen auch solche, die nur im **Dialog iterativ** durch Reformulierung (in Abhängigkeit von den bisherigen Systemantworten) beantwortet werden können ... Die Unsicherheit resultiert meist aus der **begrenzten Repräsentation von dessen Semantik** (z.B. bei Texten oder multimedialen Dokumenten);... Aus dieser Problematik ergibt sich die Notwendigkeit zur Bewertung der **Qualität der Antworten eines Informationssystems**, wobei in einem weiteren Sinne die Effektivität des Systems in Bezug auf die Unterstützung des Benutzers bei der **Lösung seines Anwendungsproblems** beurteilt werden sollte.

Prominent Systems I: Digital Libraries

- E.g. OPAC
 - Combination of structured attributes and IR-style queries

Universitätsbibliothek der Humboldt-Universität
Digitale Bibliothek

Anmelden | Hilfe

Schnellsuche | Ressource finden | Suche in Datenbanken

Suchen | Ergebnisse

Schnellsuche

Einfach | **Erweitert**

Suche: Alle Felder | ulf leser | und | Alle Felder

Allg. Fachinform.
Zeitschriftenartikel und ...

Geisteswissenschaft
Zeitschriftenartikel und ...

Literatur
Berlin/Brandenburg Katalog

Naturwissenschaft,
Agrarwissenschaft, Technik: ...

Sozialw. und Recht
Zeitschriftenartikel und ...

Sprachwissenschaft
Zeitschriftenartikel und ...

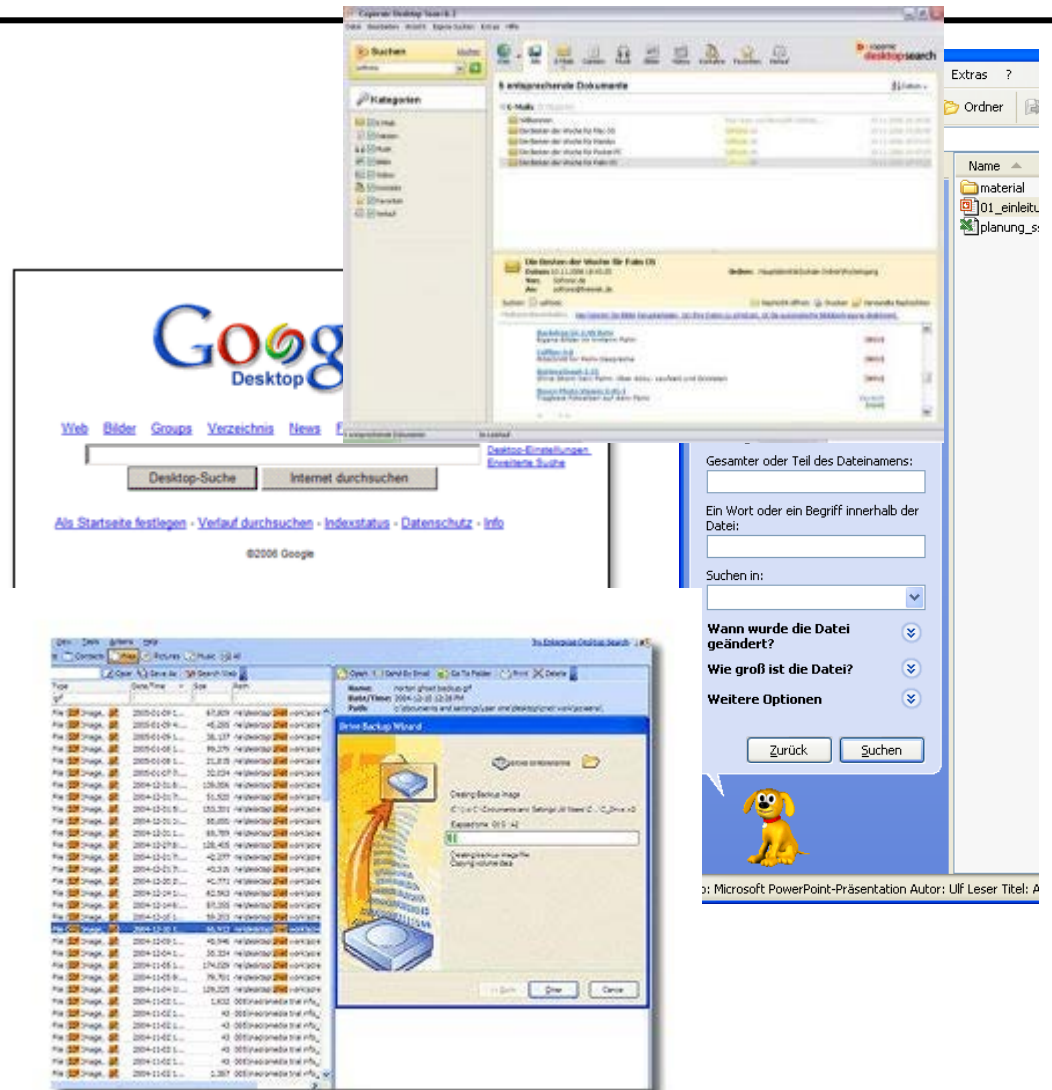
eBooks
elektronische Bücher

No.	Autor	Titel	Jahr	Quelle	Volltext?
1	Leser, Ulf	Informationsintegration :Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen	2007	KOBV Berlin-Brandenburg	☞
2	Leser, Ulf	A query language for biological networks	2005	KOBV Berlin-Brandenburg	☞
3	Leser, Ulf	Informationsintegration :Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen	2007	KOBV Berlin-Brandenburg	☞
4	Leser, Ulf [Hrsg.]	Data integration in the life sciences :third International Workshop, DILS 2006, Hinxton, UK, July 20 - 22, 2006	2006	KOBV Berlin-Brandenburg	☞
5	Leser, Ulf	Informationsintegration :Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen	2007	KOBV Berlin-Brandenburg	☞
Eintrag doppelt - siehe # 2					
6	Leser, Ulf	A query language for biological networks	2005	KOBV Berlin-Brandenburg KOBV Berlin-Brandenburg	☞
7	Leser, Ulf	Query planning in mediator based information systems	2000	KOBV Berlin-Brandenburg KOBV Berlin-Brandenburg	☞
Eintrag doppelt - siehe # 7					
8	Leser, Ulf	Query planning in mediator based information systems	2000	KOBV Berlin-Brandenburg KOBV Berlin-Brandenburg	☞
9	Heyden, Ulf	Zielgruppen des Romans	1986	Staatsbibliothek Berlin	☞
10	Heyden, Ulf	Zielgruppen des Romans :Analyse, Franz. Romanvorworte d. 19. Jh.	1986	KOBV Berlin-Brandenburg	☞

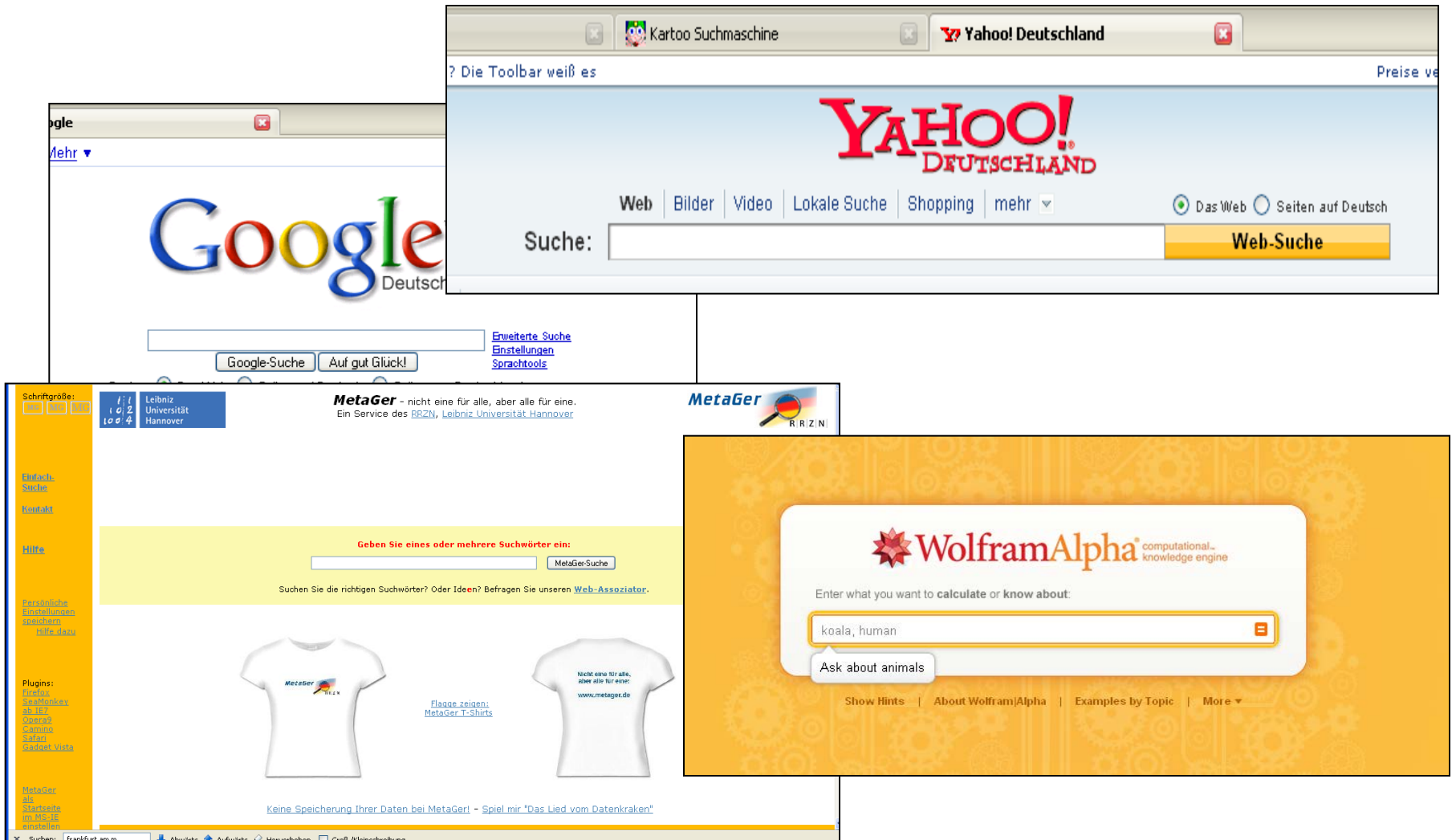
Errors?

Prominent Systems II: Desktop Search

- Much activity in 2000-2010
- Various search engines and indexing mechanisms
- Important: Search **different types of files** (txt, doc, mail, ppt, pdf, tex, odp, xls, ...)



Prominent Systems III: Web Search Engines



Almost Any Web Site

Suchergebnis auf Amazon.de für **tiger**

1-16 von mehr als 200.000 Ergebnissen oder Vorschlägen für **"tiger"**

Ergebnisse anzeigen für

- Küche, Haushalt & Wohnen
- Bettwäsche-Sets
- Badaccessoires
- Wohn- & Kuscheldecken
- Zierkissen & -hüllen
- Teppiche
- Spielzeug
- Plüschtiere
- Party- & Scherzartikel
- Kostüme & Zubehör für Kinder
- Elektronische Haustiere
- Tierfiguren für Kinder
- Baumarkt
- Wandtattoos & Wandbilder
- Fremdsprachige Bücher
- Kinderbücher zu Löwen, Tigern & Leoparden
- Romane & Erzählungen für Kinder
- Science Fiction & Magie für Kinder
- Jugendbücher
- Soziale Themen für Kinder
- Weltreise
- Prime Video
- Prime Video Filme
- Prime Video Serien
- Prime Video
- Filme
- Serien
- Kindle-Shop
- Kinderbücher zu Löwen, Tigern & Leoparden (englischsprachig)
- Science Fiction & Magie für Kinder (englischsprachig)
- Tiere (englischsprachig)
- Kinderbücher zu Säugetieren (englischsprachig)
- Action & Abenteuer für Kinder (englischsprachig)
- Weltreise
- Alle 29 Kategorien

Filtern nach

- AmazonFresh
- fresh
- Pantry

ansich Hervorheben Groß-/Kleinschreibung Ganze Wörter Ausdruck nicht gefunden

GESPONSERT VON COBI FACTORY S.A.
COBI für Sammler von historischen Militärmotoren
Jetzt einkaufen

Ampel 24 Trampolin Ø 430 cm grün | Gartentrampolin Komplettset mit vers
★★★★★: 20
prime

Gesponsert
Handtuchstange + kostenloser Versand / Tiger Cria Chrom M Handtuchhalter, Handtuchstange
von Tiger
EUR 19,90
KOSTENFREIE Lieferung

Gesponsert
verschiedene Handtuchhalter wählbar + kostenloser Versand Handtuchhaken, Handtuchhalter, Handtuchring, Handtuchst:
von Tiger
EUR 14,90
KOSTENFREIE Lieferung

Produktkategorien

- Plüschtiere
- Fitness-Kleingeräte
- Spielzeug
- Spielzeugfiguren & Spielwetten: Bauernhof & Tiere
- Bettwäsche-Sets

Plüschtier Tiger - liegend - braun - 90 cm
von Plushfarm
EUR 20,00 + EUR 5,99 Versandkosten
Nur noch 18 Stück auf Lager - jetzt bestellen.

Suchergebnis auf Amazon.de für **tiger**

90%

Meistbesucht Frequent WBI Lehre Google News Buecher kaufen Projekte Paper Reisen MyStuff hub Berlin We

Tiger Tigerfix Klebesystem Nummer 1 für Ausstattungsserien, Metall, Chrom, 0,6
von Tiger
EUR 9,99 prime
Lieferung bis **Donnerstag, 3. Mai**
Kostenlose Lieferung möglich.
Andere Angebote
EUR 8,91 (5 gebrauchte und neue Artikel)

Onistuka Tiger Herren Onistuka Tiger Mexico 66 Low-Top
von Onistuka Tiger
ab EUR 49,00 prime
Kostenlose Lieferung möglich.
Einige Größen/Farben sind für Prime qualifiziert

Plüsch Wildtier Großkatzen Plüsch Leopard Tiger Panther Designs und Größen
von TE-Trend
ab EUR 22,95 prime
Kostenlose Lieferung möglich.
Nur noch 5 Stück auf Lager - jetzt bestellen.
Einige Farben sind für Prime qualifiziert.

Ty Beanie Babies Classic Tiggs Tiger 15 cm 33 cm Plüsch Stofftier Kuscheltier
von TY
ab EUR 4,51 prime
Kostenlose Lieferung möglich.
Einige Größen sind für Prime qualifiziert.

20 Staubsaugerbeutel geeig. Vorwerk Tiger 250 251 252 INCL. FILTER
von Leabem
EUR 12,99 prime
Kostenlose Lieferung möglich.
Nur noch 11 Stück auf Lager - jetzt bestellen.
Andere Angebote
EUR 9,99 (4 neue Artikel)

ansich Hervorheben Groß-/Kleinschreibung Ganze Wörter Ausdruck nicht gefunden

Properties of Information Retrieval (IR)

- IR is about **helping a user**
- IR is about **finding information**, not about finding data
- IR builds systems for **end users**, not for programmers
 - No SQL
 - IR (web) is used by **almost everybody**, databases are not
- IR searches **unstructured data** (e.g. text)
- **90% of all information** is presented in unstructured form
 - Claim some analysts

History

- ~300 ad. Library of Alexandria , ~700.000 „documents“
- 1450: [Bookprint](#)
- 19th century: Indices / concordance
- Probabilistic models: Maron & Kuhns (1960)
- Boolean queries: Lockheed (~1960)
- [Vector Space Model](#): Salton, Cornell (1965)
 - Faster, simpler to implement, better search results
- 80s-90s: Digital libraries, SGML, hypertext, metadata standards
- Mid 90s: The web, [web search engines](#), XML, federations
- End 90s: Personalized search engines, [recommendations](#)
- 2010-: Mobile search, user-generated cont., [social networks](#)

Content of this Lecture

- What is Information Retrieval
- Documents
- Queries
- Related topics

Document or Passage

The image displays three search results for the query "shakespeare death":

- Left: Universitätsbibliothek (University Library)** - Shows a search interface with a table of results. The table lists 10 entries, each with a number, author (Shakespeare, William), and title. The titles are related to exhibitions and biographies of Shakespeare.
- Middle: Google** - Shows search results for "shakespeare death". The top result is "Death - [Diese Seite übersetzen]" with a snippet: "THE DEATH OF SHAKESPEARE. Shakespeare died in 1616 on his birthday, ...". Other results include "Death: Shakespeare" and "The Death of William Shakespeare and his will".
- Right: WolframAlpha** - Shows the query "when did shakespeare die?" in the input field. The result is "Saturday, April 23, 1616". Below this, a table shows date formats: Julian calendar (Saturday, April 13, 1616), Julian day number (2 311 405), Jewish calendar (6 Iyar, 5376 (until sunset)), and Islamic calendar (6 Rabia II, 1025 (until sunset)). At the bottom, it shows the time difference from today (Thursday, October 21, 2010): "394 years 5 months 28 days ago" and "20583 weeks 5 days ago".

Searching only
metadata

Searching tokens
within documents

Interpreting
natural text

Documents

- This lecture: [Natural language text](#)
- Might be grammatically correct (books, newspapers) or not (blogs, Twitter, spoken language)
- May have structure (title, abstract, chapters, ...) or not
- May have associated (explicit or in-text) metadata or not
- May be in different languages or even have mixed content
 - Foreign characters
- May have various formats (ASCII, PDF, DOC, XML, ...)
- May refer to other documents ([hyperlinks](#))
- Not covered
 - Semi-structured data (XML)
 - Structured data (But: Keyword search in relational databases)

IR Queries

- Users formulate queries
 - Keywords or phrases
 - Logical operations (AND, OR, NOT, ...)
 - Also other operators: “-ulf +leser”
 - Natural language questions (e.g. MS-Word help)
 - (Semi-)Structured queries (author=... AND title~ ...)
 - Voice (Siri, Alexa, ...)
- Documents as queries: Find documents similar to this one
- Query refinement based on previous results
 - Find documents matching the new query within the result set of the previous search
 - Use relevant answers from previous queries to create next query

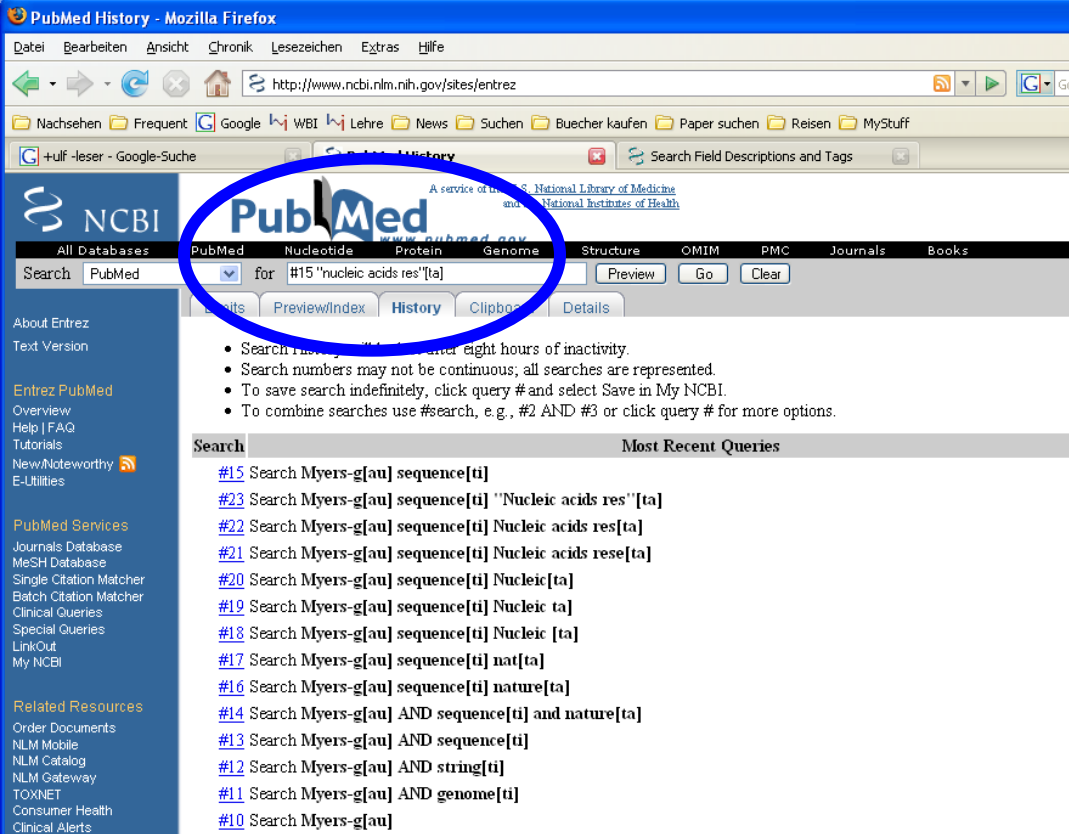
Searching with Metadata (PubMed/Medline)

The screenshot shows the PubMed website interface. The search bar contains the query "Myers-g[au] sequence[ti]". The search results show one item: "Myers GS, Parker D, Al-Hasani K, Kennan RM, Seemann T, Ren Q, Badger JH, Selengut JD, Deboy RT, Tettelin H, Boyce JD, McCarl VP, Han X, Nelson WC, Madupu R, Mohamoud Y, Holley T, Fedorova N, Khouri H, Bottomley SP, Whittington RJ, Adler B, Songer JG, Rood JJ, Paulsen IT. Genome sequence and identification of candidate vaccine antigens from the animal pathogen Dichelobacter nodosus. Nat Biotechnol. 2007 May;25(5):569-75. Epub 2007 Apr 29." The search field descriptions and tags are listed below the search results.

Search Field Descriptions and Tags

Affiliation [AD]	Issue [IP]	Place of Publication [PL]
Article Identifier [AID]	Journal Title [TA]	Publication Date [DP]
All Fields [ALL]	Language [LA]	Publication Type [PT]
Author [AU]	Last Author [LASTAU]	Secondary Source ID [SI]
Comment Corrections	Location ID [LID]	Subset [SB]
Corporate Author [CN]	MeSH Date [MHDA]	Substance Name [NM]
EC/RN Number [RN]	MeSH Major Topic [MAJR]	Text Words [TW]
Entrez Date [EDAT]	MeSH Subheadings [SH]	Title [TI]
Filter [FILTER]	MeSH Terms [MH]	Title/Abstract [TIAB]
First Author Name [1AU]	NLM Unique ID [JID]	Transliterated Title [TT]
Full Author Name [FAU]	Other Term [OT]	UID [PMID]
Full Investigator Name [FIR]	Owner	Volume [VI]
Grant Number [GR]	Pagination [PG]	
Investigator [IR]	Personal Name as Subject [PS]	
	Pharmacological Action MeSH Terms [PA]	

Query Refinement



The screenshot shows the PubMed website interface in a Mozilla Firefox browser. The address bar displays `http://www.ncbi.nlm.nih.gov/sites/entrez`. The search bar contains the query `#15 'nucleic acids res'[ta]`. Below the search bar, there are buttons for `Preview`, `Go`, and `Clear`. A blue circle highlights the search bar and the `History` button. The `History` button is currently selected, and a list of recent queries is displayed below it.

Search **Most Recent Queries**

- [#15](#) Search Myers-g[au] sequence[ti]
- [#23](#) Search Myers-g[au] sequence[ti] "Nucleic acids res"[ta]
- [#22](#) Search Myers-g[au] sequence[ti] Nucleic acids res[ta]
- [#21](#) Search Myers-g[au] sequence[ti] Nucleic acids rese[ta]
- [#20](#) Search Myers-g[au] sequence[ti] Nucleic[ta]
- [#19](#) Search Myers-g[au] sequence[ti] Nucleic ta
- [#18](#) Search Myers-g[au] sequence[ti] Nucleic [ta]
- [#17](#) Search Myers-g[au] sequence[ti] nat[ta]
- [#16](#) Search Myers-g[au] sequence[ti] nature[ta]
- [#14](#) Search Myers-g[au] AND sequence[ti] and nature[ta]
- [#13](#) Search Myers-g[au] AND sequence[ti]
- [#12](#) Search Myers-g[au] AND string[ti]
- [#11](#) Search Myers-g[au] AND genome[ti]
- [#10](#) Search Myers-g[au]

Dublin Core Metadata Initiative (W3C), 1995

- identifier: ISBN/ISSN, URL/PURL, DOI, ...
- format: MIME-Typ, media type,
- type: Collection, image, text, ...
- language
- title
- subject: Keywords
- coverage: Scope of doc in space and/or time
- description: Free text
- creator: Last person manipulating the doc
- publisher:
- contributor:
- rights: Copyright, licenses, ...
- source: Other doc
- relation: To other docs
- date: Date or period

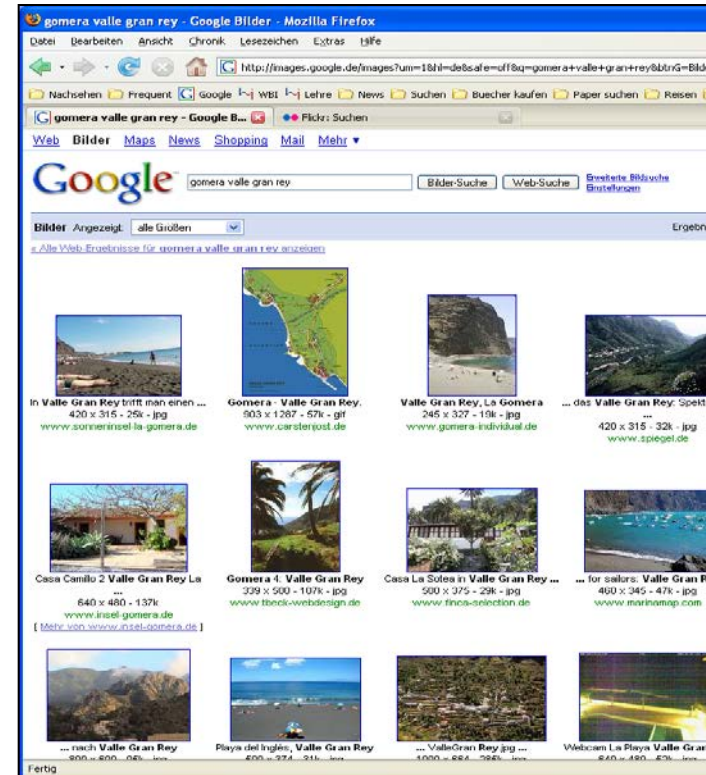
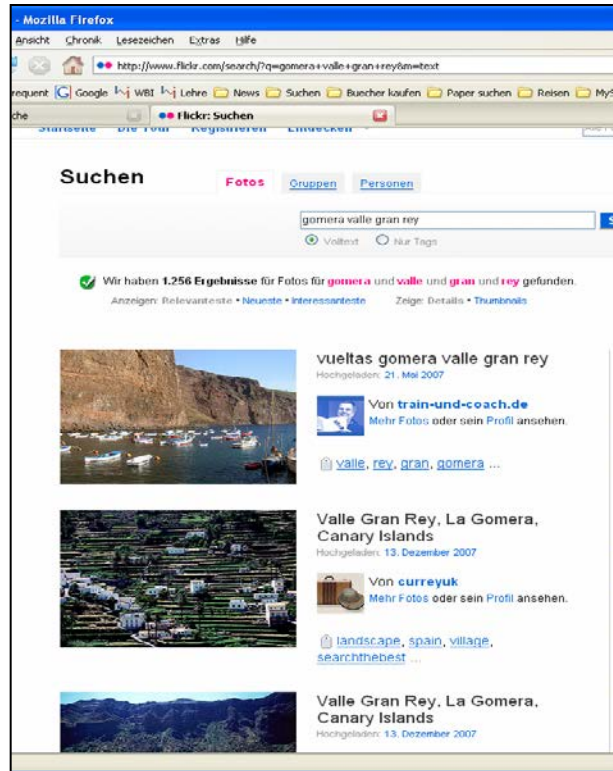
Usage in HTML

```
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Dublin Core</title>
<link rel="schema.DC" href="http://purl.org/dc/..." />
<link rel="schema.DCTERMS" href="http://purl.org/..." />
<meta name="DC.format" scheme="..." content="text/html" />
<meta name="DC.type" scheme="..." content="Text" />
<meta name="DC.publisher" content="Jimmy Whales" />
<meta name="DC.subject" content="Dublin Core Metadata" />
<meta name="DC.creator" content="Björn G. Kulms" />
<meta name="DCTERMS.license" scheme="DCTERMS.URI"
      content="http://www.gnu.org/copyleft/fdl.html" />
</head>
```

Content of this Lecture

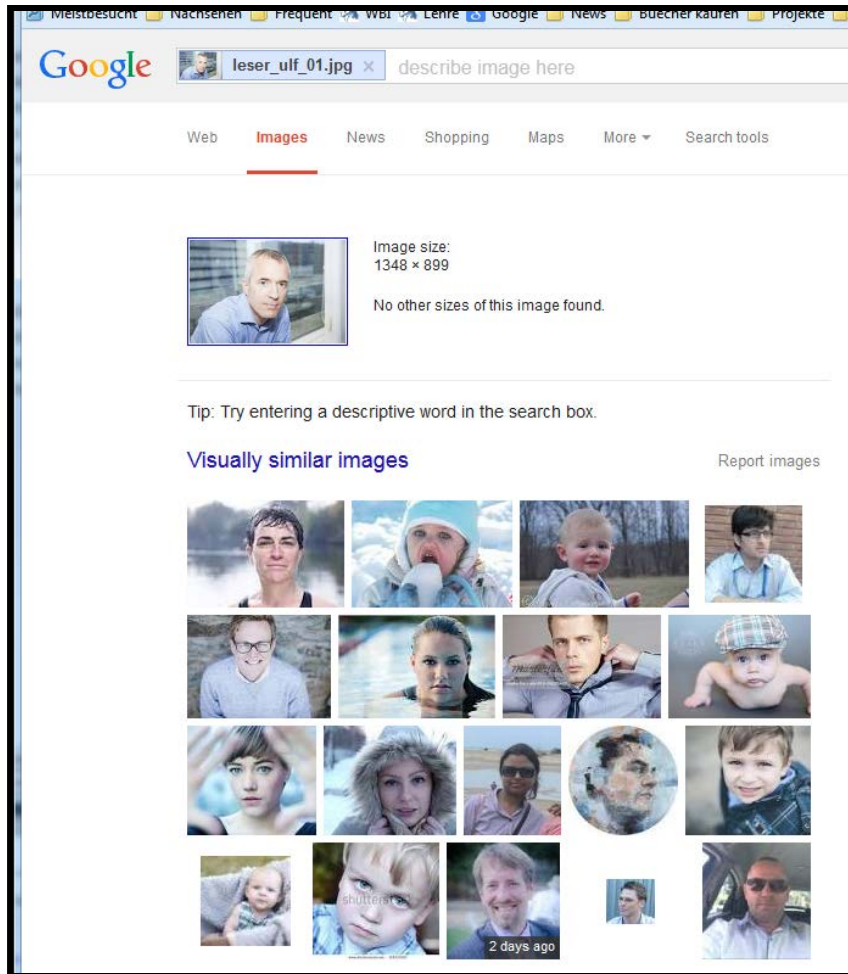
- What is Information Retrieval
- Documents
- Queries
- [Related topics](#)

Multimedia Retrieval




- Note: Neither searches within images
 - Flickr: tags ("folksonomy")
 - Google: text in neighborhood

„Search by Image“ (10/2014)



Google describe image here

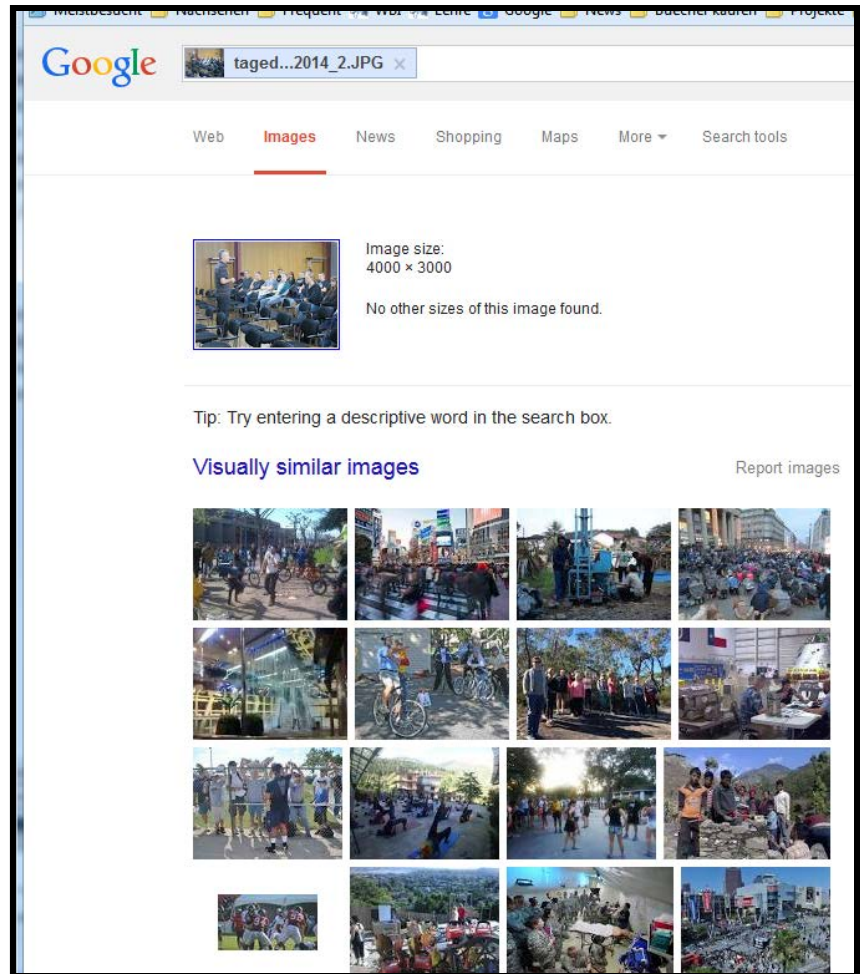
Web **Images** News Shopping Maps More Search tools

 Image size: 1348 × 899
No other sizes of this image found.

Tip: Try entering a descriptive word in the search box.


Visually similar images [Report images](#)

A grid of 20 visually similar images, including portraits of various people and a small image of a man in a white shirt with a '2 days ago' timestamp.



Google describe image here

Web **Images** News Shopping Maps More Search tools

 Image size: 4000 × 3000
No other sizes of this image found.

Tip: Try entering a descriptive word in the search box.

Visually similar images [Report images](#)


A grid of 20 visually similar images, primarily showing groups of people at outdoor events, markets, or social gatherings.

Search by Image 4/2018 – it's difficult ...

The screenshot shows a Google search by image interface. The search term is "informatik". The main image shown is a blurry photograph of a person in a classroom setting. The search results include a Wikipedia entry for "Informatik" and a link to "Informatik Studium: Studiengänge, Gehalt & Berufsaussichten". A sidebar on the right shows "Andere suchten auch nach" with categories like Wissenschaft, Computer, Unternehm..., Algorithmus, and Mathematik. The bottom of the page shows "Optisch ähnliche Bilder" with a grid of related images.


search by image - Google-Suche x Google-Suche x "ulf leser" - Google-Suche x Google-Ergebnis für https://w... x Girls' Day am Institut für Inform... x

Meistbesucht Frequent WBI Lehre Google News Bücher kaufen Projekte Paper Reisen MyStuff hub Berlin Wetter

Google  JPG x informatik Anmelden

Alle Bilder Maps Shopping Mehr Einstellungen Tools

Ungefähr 3 Ergebnisse (0,60 Sekunden)

 Bildgröße: 200 x 133
Dieses Bild in einer anderen Größe suchen: [Alle Größen - Klein](#)


Vermutung für dieses Bild: **Informatik**

Informatik – Wikipedia
<https://de.wikipedia.org/wiki/Informatik>
Informatik ist die „Wissenschaft von der systematischen Darstellung, Speicherung, Verarbeitung und Übertragung von Informationen, besonders der automatischen Verarbeitung mithilfe von Digitalrechnern“. Historisch hat sich die Informatik einerseits als Formalwissenschaft aus der Mathematik entwickelt, andererseits als ...

Informatik Studium: Studiengänge, Gehalt & Berufsaussichten
<https://www.studycheck.de> > [Studiengänge](#) > [Informatik & Mathematik](#) > [Informatik](#)
Ein Informatik Studium interessiert Dich? Hier findest Du eine Übersicht der Studieninhalte & der Voraussetzungen sowie Infos zum Thema Gehalt & Karriere.






Optisch ähnliche Bilder

Unangemessene Bilder melden

Informatik 

Informatik ist die „Wissenschaft von der systematischen Darstellung, Speicherung, Verarbeitung und Übertragung von Informationen, besonders der automatischen Verarbeitung mithilfe von Digitalrechnern“. [Wikipedia](#)

Andere suchten auch nach Über 10 weitere ansehen

Wissensch... Computer Unterneh... Algorithmus Mathematik

Feedback

Question Answering

- Asking for a specific bit of information
 - What was the score of Bayern München versus Stuttgart in the DFB Pokal finals in 1998?
 - How many hours of sunshine has a day in Crete in May?
 - When does the next S9 leave this station?
- Prominent until recently: IBM Watson
 - “IBM Watson is a technology platform that uses natural language processing and machine learning to reveal insights from large amounts of unstructured data” [2011]
- Hot topic for **personal assistants**
 - E.g. Amazon Echo, Apple Siri, Google Assistant, ...
- QA: Mixture of **statistical NLP**, **Machine learning** and **IR**



Historic Texts



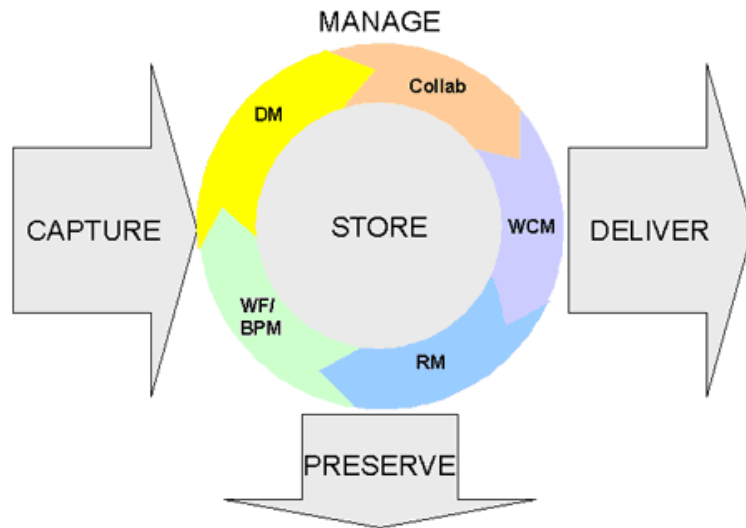
- Sachsenspiegel, ~1250
 - "Swerlenrecht künnen wil•d~ volge dis buches lere.alrest sul wi mer ken, daz ..."
- Multiple representations
 - Facsimile
 - Digitalization / diplomacy
 - How well can the facsimile be reproduced from the dig. form?
 - Differences in individual writers (proliferating errors)
 - Different translations
 - Different editions

Other Buzzwords

- Document management systems (DMS)
 - Large [commercial market](#), links to OCR, workflow systems, etc.
 - Many legal issues (compliance, reporting, archival, ...)
 - Essentially all companies run some form of a DMS
 - Every DMS includes an IR system
- Knowledge management
 - “More sophisticated” DMS with [semantic searching](#)
 - Ontologies, thesauri, topic maps, ...
 - [Social aspects](#): Incentives, communities, enterprise standards, ...
- Digital libraries
 - Somewhat [broader](#) and less technical
 - Includes social aspects, [archiving](#), multimedia, ...

Enterprise Content Management

- „The technologies used to capture, manage, store, deliver, and preserve **information** to **support business processes**“



Quelle: AIIM International

- Authorization and authentication
- Business process management and **document flow**
- **Compliance**: legal requirements
 - Record management
 - Pharma, Finance, ...
- Collaboration and sharing
 - Inter and intra organizations
 - Transactions, locks, ...
- **Publishing**: What, when, where
 - Web, catalogues, mail push, ...
- ...

Technique versus Content

- IR is about techniques for searching a **given doc collection**
- **Creating doc collections** is a business: **Content provider**
 - Selection/filtering: classified business news, new patents, ...
 - Augmentation: Annotation with metadata, summarization, linking of additional data, ...
- **Examples**
 - **Medline**: >5000 Journals, >28M citations, >700K added per year
 - Thompson Reuter
 - Impact factors: which journals count how much?
 - Web catalogues ala Yahoo
 - “Pressespiegel”, web monitoring

Self Assessment

- Give a definition of „Information retrieval“
- How is information retrieval different from database query evaluation?
- What are means to shorten the number of queries necessary to fulfil an information request?
- What is the difference between classical IR and Question Answering?
- What are possible types of answers to a IR query?