

# Information Retrieval: Introduction

Norbert Fuhr

# Contents

Information Retrieval Applications

What is Information Retrieval?

What is Relevance?

Representations

IR Models

Interactive Retrieval

# **Information Retrieval Applications**

Application Examples  
Facets of Search

# Web Search

Google information retrieval system  [Erweiterte Suche](#)

Suche:  Das Web  Seiten auf Deutsch  Seiten aus Deutschland

---

Web [\[+ Optionen anzeigen\]](#) Ergebnisse 1 - 10 von ungefähr 9.710.000 für **information retrieval system**. (0,28 Sekunden)

## [Information Retrieval – Wikipedia](#)

Weitere wichtige Vertreter der Frühphase des **Information Retrieval** waren Mortimer Taube, der das Uniterm-**System** entwickelte, Hans Peter Luhn, der das Modell ...

[Definition](#) - [Geschichte](#) - [Retrievalmodelle](#) - [Klassifikation von ...](#)  
[de.wikipedia.org/wiki/Information\\_Retrieval](http://de.wikipedia.org/wiki/Information_Retrieval) - [Im Cache](#)

## [\[PDF\] 1. Was versteht man unter einem Information Retrieval System?](#)

Dateiformat: PDF/Adobe Acrobat - [Schnellansicht](#)

**Information Retrieval System?** Literaturrempfehlungen a) Klassiker. SALTON,G.,  
McGILL,M.J.: Introduction to Modern **Information Retrieval**. ...  
[www.ai.wu.ac.at/~wyk/ir-vo/Folien\\_Kapitel\\_1.pdf](http://www.ai.wu.ac.at/~wyk/ir-vo/Folien_Kapitel_1.pdf)

## [Abgeschlossene Diplomarbeit: Entwicklung einer ...](#)

Die in einem Hypertext-**Information-Retrieval-System** (HIRS) verwalteten Objekte sind strukturierte Dokumente, die untereinander wiederum in Beziehung ...  
[www.is.informatik.uni-duisburg.de/dpa/sarr.html](http://www.is.informatik.uni-duisburg.de/dpa/sarr.html) - [Im Cache](#) - [Ähnlich](#)

## [Information Retrieval Systems - \[ Diese Seite übersetzen \]](#)

**Information Retrieval Systems** ... Virtual Reference Desk · University of Massachusetts Center for Intelligent **Information Retrieval** · Callan CMU IR Group ...  
[www.csc.lsu.edu/~kraft/retrieval.html](http://www.csc.lsu.edu/~kraft/retrieval.html) - [Im Cache](#) - [Ähnlich](#)

# Product Search in Online Shops

"lcd fernseher 48 zoll"

Verwandte Suchbegriffe: [lcd fernseher 46 zoll](#).

1-16 von 158 Ergebnissen

Sortieren in [Elektronik & Foto](#) ▾



## Hannspree HA191DPB 48,26cm (19 Zoll) LCD Monitor VGA, DVI (Kontrast dyn. 1000:1, HANNSPREE

**EUR 109,89**

Bestellen Sie in den nächsten **4 Stunden**, um den Artikel am Freitag, 28. September zu erhalten.

Nur noch 2 Stück auf Lager - jetzt bestellen.

Andere Angebote

**EUR 99,00 neu** (58 Angebote)

**EUR 96,34 gebraucht** (3 Angebote)

★★★★★ (2)

Kostenlose Lieferung möglich.

Computer & Zubehör: Alle 22 Artikel ansehen



## Philips 19PFL3606H/12 48 cm (19 Zoll) LCD-Fernseher, Energieeffizienzklasse B (HD-

**EUR 225,00**

Nur noch 1 Stück auf Lager - jetzt bestellen.

Andere Angebote

**EUR 195,00 neu** (3 Angebote)

★★★★★ (26)

Neuere Version vorhanden

Elektronik & Foto: Alle 132 Artikel ansehen



## Grundig 46 VLE 8160 SL117 cm (46 Zoll) 3D LED-Backlight-Fernse... Energieeffizienzklasse A+

**EUR 694,99**

Bestellen Sie in den nächsten **7 Stunden**, um den Artikel am Freitag, 28. September zu erhalten.

Andere Angebote

**EUR 694,99 neu** (14 Angebote)

**EUR 604,64 gebraucht** (2 Angebote)

★★★★★ (42)

Kostenlose Lieferung möglich.

Filme & TV: Alle Artikel ansehen

# Intranet Search

UNIVERSITÄT  
DUISBURG  
ESSEN

UDE-Suchseite Google-Web Google-Bilder

Vorlesung datenbanken default\_collection ▾ UDE Suche

Suche Ergebnisse 1 - 10 von ca. 31 für Vorlesung datenbanken. Die Suche dauerte 0,25 Sekunden.

Weiter> Sortiere nach Datum / Sortiere nach Relevanz

Universitätsbibliothek Duisburg-Essen: Schlüsselkompetenzen für ...  
... in nationalen und internationalen **Datenbanken**. Kreditierung: 0,5 ECTS. Studienleistung:  
Klausur. Termine: Mi, 09.01.2008, 14:15 - 15:45 Uhr; 1. **Vorlesung** Mi, 16.01 ...  
[www.ub.uni-duisburg-essen.de/biblio/schulung/ba.shtml](http://www.ub.uni-duisburg-essen.de/biblio/schulung/ba.shtml) - 46k

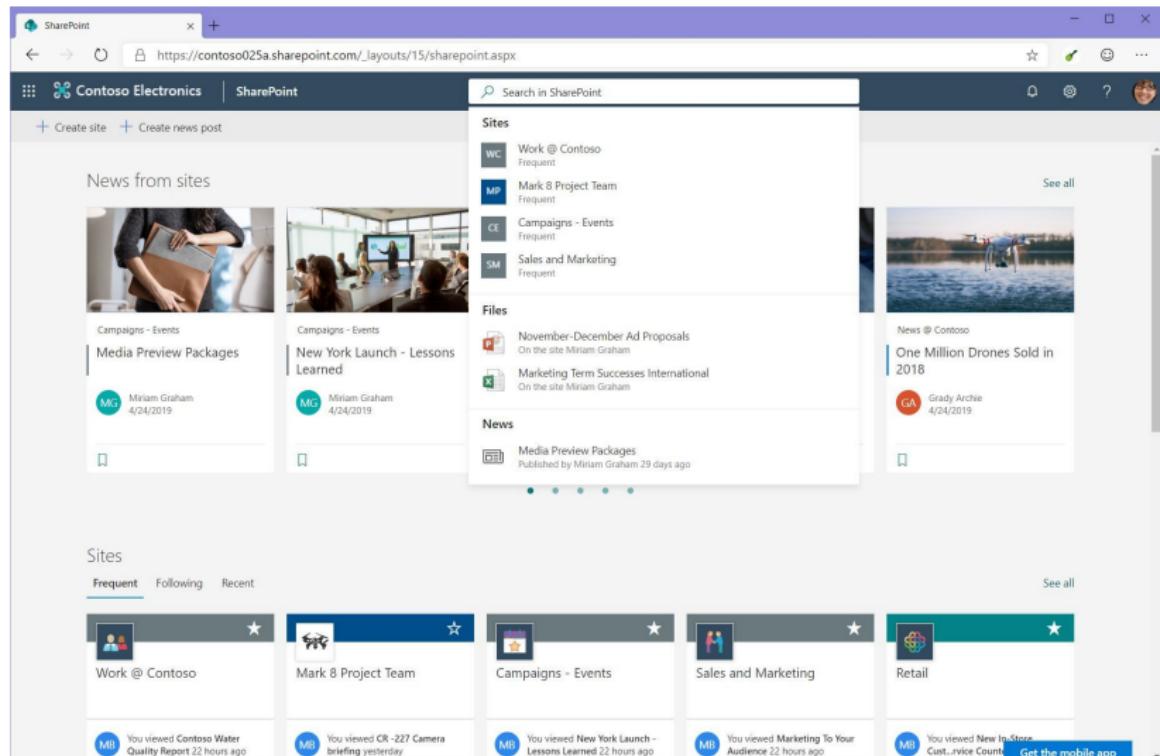
moodle uni-due: Datenverwaltungssysteme und Wissensrepräsentation  
... In der **Vorlesung** werden zunächst die Grundlagen verteilter Systeme ... Architektur  
verteilter Datenbanksysteme; Entwurf verteilter **Datenbanken**; Anfrageverarbeitung; ...  
[moodle.uni-duisburg-essen.de/course/category.php?id=75](http://moodle.uni-duisburg-essen.de/course/category.php?id=75) - 27k

moodle uni-due: Technik der Rechnernetze  
... **Vorlesung** "Netzmanagement" (2 SWS, 3 CP, WS). ... Verknüpfung mit öffentlicher  
IP-Netzverwaltung / Whois-**Datenbanken**; Pathologische Routingergebnisse. ...  
[moodle.uni-duisburg-essen.de/course/category.php?id=76](http://moodle.uni-duisburg-essen.de/course/category.php?id=76) - 55k

vorlesungen  
... **Vorlesung**, Inhalt. ... DB (**Datenbanken**), Datenbank - Grundlagen, Anwendungen relationaler  
**Datenbanken**, Anwendungen objektorientierter **Datenbanken**, Schnittstelle CAD ...  
[www.uni-due.de/ikb/vorlesungen.shtml](http://www.uni-due.de/ikb/vorlesungen.shtml) - 13k

Prof. Dr. Rüdiger Schmitt-Beck  
... Die **Vorlesung** gibt einen Überblick der wesentlichen Theoreiansätze und ... Kenntnisse  
in Programmierung - vorzugsweise in PHP und mySQL-**Datenbanken** - oder die ...  
[www.uni-due.de/politik/schmitt-beck\\_lehre.php](http://www.uni-due.de/politik/schmitt-beck_lehre.php) - 76k

# Enterprise Search



The screenshot shows the SharePoint search interface. At the top, there's a navigation bar with 'SharePoint' and a search bar labeled 'Search in SharePoint'. Below the navigation bar, there are sections for 'Sites', 'Files', and 'News'.

**Sites**

- WC Work @ Contoso (Frequent)
- MP Mark 8 Project Team (Frequent)
- CE Campaigns - Events (Frequent)
- SM Sales and Marketing (Frequent)

**Files**

- November-December Ad Proposals (On the site Miriam Graham)
- Marketing Term Successes International (On the site Miriam Graham)

**News**

- Media Preview Packages (Published by Miriam Graham 29 days ago)

**Sites**

Frequent | Following | Recent | See all

- Work @ Contoso
- Mark 8 Project Team
- Campaigns - Events
- Sales and Marketing
- Retail

Activity notifications:

- You viewed Contoso Water Quality Report 32 hours ago
- You viewed CR-227 Camera briefing yesterday
- You viewed New York Launch - Lessons Learned 22 hours ago
- You viewed Marketing To Your Audience 22 hours ago
- You viewed New In-Store Customer Experience Counter...

[Get the mobile app](#)

Figure: Sharepoint Search [docs.microsoft.com]

# Searching in Digital Libraries

 **PORTAL**  
HeBIS

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

**Search:**  The ACM Digital Library  The Guide  
  
**SEARCH**

---

**THE ACM DIGITAL LIBRARY**

 [Feedback](#)

image retrieval  
Terms used [image retrieval](#) Found 8,061 of 255,080

Sort results by relevance ▾ [Save results to a Binder](#) Refine these results with [Advanced Search](#)  
Display results expanded form ▾ Try this search in [The ACM Guide](#)

Results 1 - 20 of 8,061 Result page: 1 [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#) [>>](#)

---

**1** [Image retrieval: Ideas, influences, and trends of the new age](#)

 Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang  
April 2008 **ACM Computing Surveys (CSUR)**, Volume 40 Issue 2  
**Publisher:** ACM  
Full text available:  [PDF](#) (2.81 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

**Bibliometrics:** Downloads (6 Weeks): 482, Downloads (12 Months): 1801, Citation Count: 2

We have witnessed great interest and a wealth of promise in content-based image retrieval as an emerging technology. While the last decade laid foundation to such promise, it also paved the way for a large number of new techniques and systems, got many ...

**Keywords:** Content-based image retrieval, annotation, learning, modeling, tagging

---

**2** [Content-based image retrieval: approaches and trends of the new age](#)

 Ritendra Datta, Jia Li, James Z. Wang  
November 2005 **MIR '05**: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval  
**Publisher:** ACM  
Full text available:  [PDF](#) (467.64 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

**Bibliometrics:** Downloads (6 Weeks): 108, Downloads (12 Months): 728, Citation Count: 14

# Multimedia Search: Images

Google  camera search

Q: All Images Shopping News Maps More Tools

vinho verde porto rose douro lancers madeira dao mateus green blue

Ads · Shop portuguese wine

2021 Alvarinho Classic Quinta de Wein am Limit €11.90 +€7.90 shipping €15.87 / 1l	Niepoort Batuta Tinto DOC 2020 €47.00 gute-weine.de +€5.00 shipping €62.67 / 1l	Alumia Vinho Verde 2020, Portugal, €4.99 <span>€0</span> Vinos.de +€2.99 shipping €6.65 / 1l	Smiling Donkey Douro Red 2019 €8.99 Silkes Weinkeller Free shipping €11.99 / 1l	Alentejo - Probierbox Portugal €31.86 Adriberia.de +€5.90 shipping €7.08 / 1l	Pintas Wine & Soul Douro Red 2017 €74.95 gute-weine.de Free shipping €99.93 / 1l	Macanita Macanita Tinto 2019 €12.95 gute-weine.de +€5.00 shipping €17.27 / 1l	Blo 75+ €10. Nor +€2

# Multimedia Search: Video

YouTube DE

fado lisbon

Home Explore Shorts Subscriptions Library History Your videos Watch later Liked videos

SUBSCRIPTIONS

Theoria Apophysis Browse channels

MORE FROM YOUTUBE

YouTube Premium Movies & Shows Gaming

FILTERS

About these results

 Our Favorite Fado Spot In Alfama - TascaBeat // Lisbon Travel Blog  
8.9K views • 4 years ago  
For 91 Days Travel Blog  
We're very familiar with Flamenco but we've never seen a Fado performance. We feared it's going to be dreadful, too sad and ...  
13:56

 Fria claridade - Portuguese Fado Music Live in Lisbon  
137K views • 9 years ago  
Eric Feng  
I love Portugal, especially fish, tile painting, beach without sharks and Fado. other versions of this song By Amália Rodrigues ...  
2:32

 Fado Music from Portugal - Traditional - Portuguese Music 2 Hours  
2.4M views • 4 years ago  
ZYXMusic  
Original Fado – sung sadness! That's how you could describe the typical Portuguese folk music in only two words. Fado heroes ...  
2:16:50

# (Recommender Systems)

## Finding without querying

Customers who viewed this item also viewed

P



Programming Pearls

Jon Bentley

★★★★★ 274

Paperback

\$42.46

\$9.21 shipping



Cracking the Coding Interview: 189

Programming Question...

Gayle Laakmann McDowell

★★★★★ 6,890

Paperback

#1 Best Seller in Microsoft Programming

\$24.49

\$12.16 shipping



Designing Data-Intensive Applications: The Big...

Martin Kleppmann

★★★★★ 2,779

Paperback

#1 Best Seller in Desktop Database Books

\$35.00

Get it as soon as Friday, Jun 24

\$11.47 shipping



Introduction to Algorithms, 3rd Edition

(The MIT Press)

Thomas H. Cormen

★★★★★ 1,993

Hardcover

\$75.00

Get it as soon as Friday, Jun 24

\$16.52 shipping



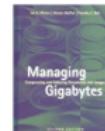
How to Solve It: A New Aspect of Mathematical Method (Princeton Science Library, 85)

G. Polya

★★★★★ 673

Paperback

\$14.70



Managing Gigabytes: Compressing and Indexing Documents a...

Ian H. Witten

★★★★★ 15

Hardcover

\$106.33

Get it Jun 29 - Jul 13

\$3.42 shipping

Usually ships within 6 to 1...

## recommendations

MovieLens helps you find movies you will like. Rate movies to build a custom taste profile, then MovieLens recommends other movies for you to watch.

### top picks

see more

based on your ratings, MovieLens recommends these movies

Band of Brothers

2001 250 min. R

★★★★★ 5

Casablanca

1942 102 min. R

★★★★★ 5

One Flew Over the Cuckoo's Nest

1975 111 min. R

★★★★★ 5

The Lives of Others

2006 137 min. R

★★★★★ 5

Sunset Boulevard

1950 116 min. R

★★★★★ 5

The Third Man

1949 104 min. R

★★★★★ 5

Pat!

1957

★★★★★ 5

### recent releases

see more

movies released in last 90 days that you haven't rated

Carrie's War

2014 105 min. R

★★★★★ 5

Felicity

2014 8

★★★★★ 5

What If

2014 111 min. R

★★★★★ 5

Frank

2014 96 min. R

★★★★★ 5

Sin City: A Dame to Kill For

2014 102 min. R

★★★★★ 5

If I Stay

2014 108 min. R

★★★★★ 5

Aren't You Curious?

2014

★★★★★ 5

# Facets of Search

## Language

### Example: Cross-lingual search in Google

Google übersetzer

Übersetzung

Übersetzte Suche

Translator Toolkit

Tools und Ressourcen

Nach Websites in anderen Sprachen suchen

Sonnenenergie

Seiten durchsuchen, die in folgender Sprache geschrieben sind:

Meine Sprache:

Übersetzen und suchen

Übersetzte Ergebnisse von englischen Webseiten

Ergebnisse 1 - 10 von ungefähr 30.200.000 für solar power

Deutsche Übersetzung

**Solarenergie - Wikipedia, die freie Enzyklopädie**  
Solarstrom ist die Erzeugung von Strom aus Sonnenlicht. Dies kann wie bei Photovoltaik (PV), als direkte oder indirekte mit solarthermischen...  
[en.wikipedia.org/wiki/Solar\\_power](http://en.wikipedia.org/wiki/Solar_power) - 168k - [Im Cache](#)

Englischer Originaltext - Englische Ergebnisse ausblenden

**Solar power - Wikipedia, the free encyclopedia**  
Solar power is the generation of electricity from sunlight. This can be direct as with photovoltaics (PV), or indirect as with concentrating solar power ...  
[en.wikipedia.org/wiki/Solar\\_power](http://en.wikipedia.org/wiki/Solar_power) - 168k - [Cached](#)

Solarenergie - Wikipedia, die freie Enzyklopädie

Solar powered elektrische Generation setzt auf Wärmeleistungsmaschinen und Photovoltaik. Solar Energy verwendet werden nur durch menschliche Genialität begrenzt...  
[en.wikipedia.org/wiki/Solar\\_energy](http://en.wikipedia.org/wiki/Solar_energy) - 247k - [Im Cache](#)

**Solar energy - Wikipedia, the free encyclopedia**  
Solar powered electrical generation relies on heat engines and photovoltaics. Solar energy's uses are limited only by human ingenuity.  
...  
[en.wikipedia.org/wiki/Solar\\_energy](http://en.wikipedia.org/wiki/Solar_energy) - 247k - [Cached](#)

Solar Power

SolarPower.org ist es, den raschen Einsatz von erneuerbaren Energien und **Sonnenenergie** gewidmet Macht in ganz Amerika. Hier finden Sie Tools, Informationen zu finden, und die Industrie...  
[www.solarpower.org/](http://www.solarpower.org/) - 27k - [Im Cache](#)

**Solar Power**  
SolarPower.org is dedicated to the rapid deployment of renewable energy and **solar power** across America. Here you'll find tools, information, and industry ...  
[www.solarpower.org/](http://www.solarpower.org/) - 27k - [Cached](#)

# Facets of Searching

## Structure

### Example: XML retrieval

Personalized Daffodil- (larsen)

2.Mobility Management for Cellular...

Mobility Management for Cellular...

- About the Article
- 1 Introduction
- 2 Mobility management
  - 3 Handoff
    - 3.1 INTER-BASE STATION HANDOFF
    - 3.2 INTERSYSTEM HANDOFF
  - 4 Roaming
    - 4.1 THE INTERCONNECTION BETWEEN THE
    - 4.2 REGISTRATION
    - 4.3 CALL DELIVERY
  - 5 Conclusion

INTERSYSTEM HANDOFF

Our description of the intersystem handoff follows IS-41[2] (GSM follows similar procedures), and we assume network-controlled handoff. Figure 3 illustrates the trunk (voice or data circuit) connection before and after the handoff. A communicating **mobile phone** user moves out of the base station served by  $MSC_1$  and enters the area covered by  $MSC_2$ . The handoff follows these steps:

- $MSC_1$  requests  $MSC_2$  to perform handoff measurement.  $MSC_2$  then selects a candidate base station,  $BS_2$ , for handoff. That is,  $MSC_2$  finds a base station that covers the **mobile phone** and has a free radio channel to cover the call.  $MSC_2$  returns the signal-quality parameter values and other information to  $MSC_1$ .
- $MSC_1$  checks if the **mobile phone** has made too many handoffs or if intersystemtrunks are not available. If so,  $MSC_1$  exits the procedure. Otherwise,  $MSC_1$  asks  $MSC_2$  to set up a voice channel. Suppose that a voice channel is available in  $BS_2$ .  $MSC_2$  asks  $MSC_1$  to start the radio link transfer.
- $MSC_1$  sends the **mobile phone** a handoff order. The **mobile phone** tries to synchronize to  $BS_2$ . After the **mobile phone** connects to  $BS_2$ ,  $MSC_2$  informs  $MSC_1$  that the handoff is successful.  $MSC_1$  then connects the call path (trunk) to  $MSC_2$  and completes the handoff

Figure 3: Before (a) and after (b) an intersystem handoff.

Detail : Shown... Exit

# Facets of Search

## Media

Example: Similarity search for images

GazoPa<sup>β</sup>  
similar image search



» www.edo2008.net  
Image  
Created: over 4 years ago  
Size : 400x316  
Post to Twitter  
Post to Facebook

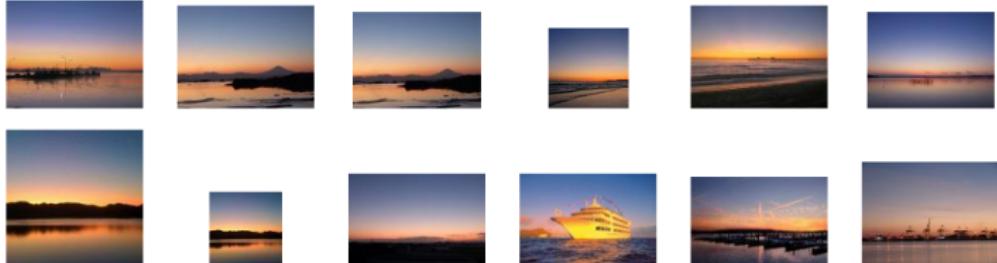
Upload Edit

Options : Standard Any size Any time Gray scale only Omit same Reset parameters Safe search is on

All Video News Sports Twitter Funny Flickr

Results 1 - 30 of 1000 for **key image**

view:



# Facets of Search

## Objects

Example: People search in 123people

The screenshot shows the 123people.de search interface. The search term 'norbert fuhr' is entered in the search bar, and the results are displayed for 'Deutschland'. A yellow sticky note overlay says 'Neu: Statistiken! Finden Sie heraus woher und wie oft gesucht wurde!'.

**Suchergebnisse** | **Statistiken**

Personensuch-Ergebnisse: Norbert Fuhr

**Norbert Fuhrs Fotos (26)**

ALLE people facebook Google LinkedIn bing incognita

Sponsored links  
mollige treffen  
gratis personen suche  
gratis auskunft

1 2 3 weiter>>

### Email Adressen (9)

- norbert.fuhr@uni-due.de
- norbert.fuhr@charite.de
- fuhr@ls6.info
- fuhr@uni-duisburg.de
- fuhr@is.informatik.uni-duisburg.de
- fuhr@lothlorien.cs.uni-dortmund.de
- fuhr@cs.uni-dortmund.de
- fuhrmann@enjoyhotels.de

1 2 weiter>>

### Telefonbuch (11)

Telefonnummern Suche nach Norbert Fuhr ergab 11 Treffer

<a href="#">Fuhr Norbert</a> Lahnsteiner Str. 21, 15366 Neuenhagen - Karte	03342 20 3
<a href="#">Fuhr Norbert Dr.</a> 12163 Berlin - Karte	030 7 93 1
<a href="#">Weiß Norbert u. Sabine Fuhr- un...</a> Mühlhauser Feld 4, 85664 Hohenlinden - Karte	08124 5 2
<a href="#">Fuhr Norbert</a> Homburger Platz 14, 98693 Ilmenau - Karte	03677 66 7
<a href="#">Fuhr Ulrike u. Norbert</a> Milanweg 4A, 44229 Dortmund - Karte	0231 73 7

# Facets of Search

## Knowledge Graph



### Lisbon

Capital of Portugal

Lisbon is Portugal's hilly, coastal capital city. From imposing São Jorge Castle, the view encompasses the old city's pastel-colored buildings, Tagus Estuary and Ponte 25 de Abril suspension bridge. Nearby, the National Azulejo Museum displays 5 centuries of decorative ceramic tiles. Just outside Lisbon is a string of Atlantic beaches, from Cascais to Estoril. — Google

**Area:** 100 km<sup>2</sup>

**Elevation:** 2 m

**Weather:** 22°C, Wind NW at 8 km/h, 73 % Humidity [weather.com](#)

**Local time:** Wednesday 10:02

**Population:** 504,718 (2016) [United Nations](#)

[Plan a trip](#)



[Item](#) [Discussion](#)

### Lisbon (Q597)

capital city of Portugal

Lisboa

« In more languages

Configure

Language	Label	Description
English	Lisbon	capital city of Portugal
German	Lissabon	Hauptstadt von Portugal
Hindi	लिस्बन	पुर्तगाल की राजधानी
Italian	Lisbona	capitale del Portogallo

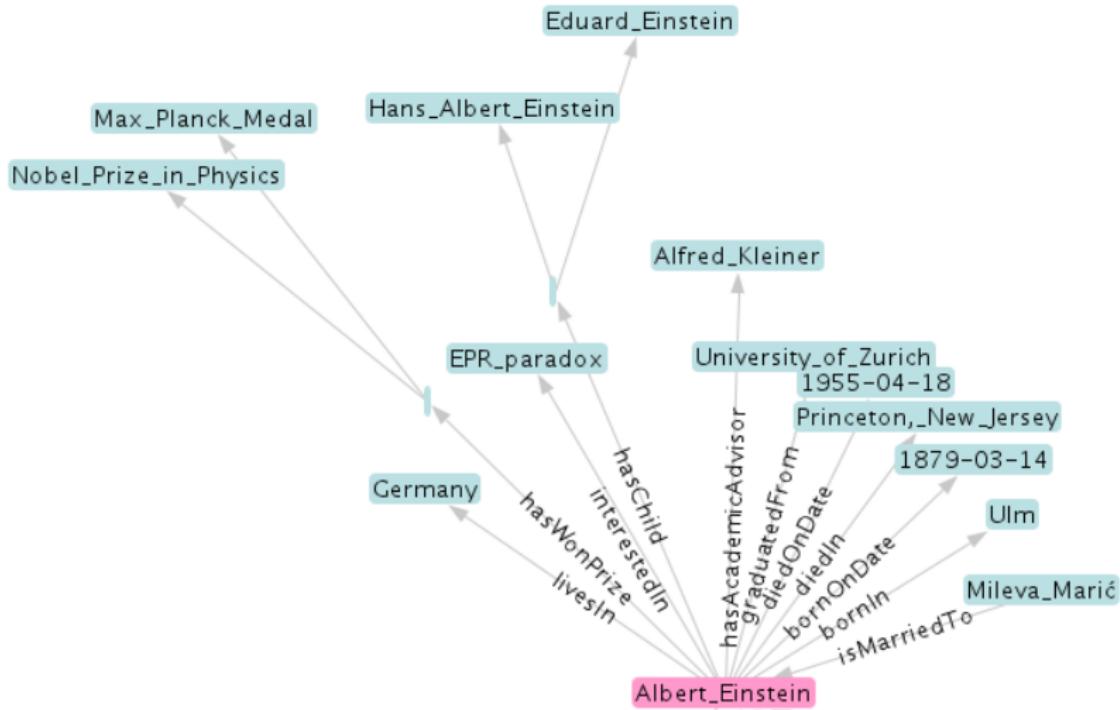
All entered languages

### Statements

instance of	 <a href="#">capital</a> + 0 references
	 <a href="#">big city</a> + 0 references
	 <a href="#">municipality of Portugal</a> + 0 references
	 <a href="#">city of Portugal</a> + 0 references
part of	 <a href="#">Lisbon Metropolitan Area</a>

# Facets of Search: Knowledge Graph Browser

YAGO [Suchanek et al. 2007]



# Facets of Search

## Question Answering

who is the current president of the USA

All Images News Maps Videos More Tools

About 1.100.000.000 results (1,00 seconds)

United States / President

**Joe Biden**



how do i delete my instagram account

All Videos Images Maps Shopping More Tools

About 2.350.000.000 results (0,55 seconds)

**How to delete your Instagram account:**

1. Go to the Instagram website and log into your account.
2. Go to the account deletion page.
3. From the drop-down menu, choose the reason you are deleting your account.
4. Enter your password again. Then click or tap the delete button.

20 Sept 2021

# Facets of Search

## Argument retrieval

The screenshot shows the args search interface. At the top, there is a logo with a green and red bird-like icon, followed by the word "args". Below the logo is a search bar containing the query "covid vaccination". To the right of the search bar is a magnifying glass icon. Further right, it says "Pro vs. con view" and "75 arguments retrieved in 552.0 ms". Below the search bar, there are navigation links: "All", "Discussions", "News", and "People".

### PRO

#### [Barely a month into a mass vaccination campaign to stop...](#)

▶ Show full argument

Barely a month into a mass **vaccination** campaign to stop the **COVID-19** pandemic, the Trump administration unexpectedly shifted gears Tuesday to speed the delivery of shots. ...  
The move came after widespread concern over a slow ...  
<https://www.allslides.com/story/us-ramps-vaccine-availability> score ▾

#### [OPINION At last, light at the end of the COVID-19 tunnel...](#)

▶ Show full argument

OPINION At last, light at the end of the **COVID-19** tunnel: At least two highly effective vaccines should begin distribution before year's end. ... Moderna this week followed Pfizer in announcing plans to have a vaccine on the ...  
<https://www.allslides.com/story/perspectives-moderna-and-pfizer-announce-covid-19-vaccines> score ▾

#### [The Covid-19 vaccine is potentially more dangerous than...](#)

▶ Show full argument

The **Covid-19** vaccine is potentially more dangerous than any other previous vaccine because it was made in a rush. In an emergency, therefore it is reasonable to doubt it. Also even if the short term consequences are more ...  
<https://www.debate.org/debates/Corona-vaccination-should-not-be-mandatory/1/> score ▾

### CON

#### [Although one cannot predict the future consequences of...](#)

▶ Show full argument

Although one cannot predict the future consequences of the vaccines - the short-term consequences have been demonstrated as beneficial to the public good in diminishing the rate of **Covid-19** transmission and **Covid-related** ...  
<https://www.debate.org/debates/Corona-vaccination-should-not-be-mandatory/1/> score ▾

#### [But the most likely outcomes and the reasons we have for...](#)

▶ Show full argument

But the most likely outcomes and the reasons we have for currently believing that the **Covid-19** vaccines will have a positive effect give little reason why people should refuse the vaccines. ... It is not more important than the ...  
<https://www.debate.org/debates/Corona-vaccination-should-not-be-mandatory/1/> score ▾

#### [Because the effects of not taking a Covid-19 vaccine have...](#)

▶ Show full argument

Because the effects of not taking a **Covid-19** vaccine have a public impact - refusing a vaccine threatens the lives of other people - personal liberty is about one's body is less important than preventing the loss of life and ...  
<https://www.debate.org/debates/Corona-vaccination-should-not-be-mandatory/1/> score ▾

# Facets of Search

## static/dynamic Content

### Example: Twitter search

The screenshot shows a Twitter search interface for the hashtag #SIGIR2022. The top navigation bar includes options like 'Entdecken' and 'Einstellungen'. The search results are filtered by 'Top' and show two tweets from the account @SIGIRConf.

**Tweet 1:** SIGIR 2022 @SIGIRConf - 20. Mai  
Call for Volunteers! Join the organizing team and other students to help running #SIGIR2022 More Info and application: bit.ly/SIGIR2022 Applications must be received by June 1st. Volunteer positions will run from Monday July 11th through Friday, July 15th.

**Tweet 2:** SIGIR 2022 @SIGIRConf - 13. Mai  
If you're planning to come to Madrid for #SIGIR2022, you should consider getting the most out of an #IberianSummer While you're in the peninsula, #SIGIR2022 is taking place in Lisbon, just a few hundred km away the following week (18-23 July)! bit.ly/2FHQq4K

**Right sidebar (Search Filter):**

- Mit Google anmelden
- Mit Apple registrieren
- Mit Telefonservice oder E-Mail-Ad...

Inkl. du dich registrierst, stimmt du den Allgemeinen Geschäftsbedingungen und Datenschutzrichtlinien sowie der Nutzung von Cookies zu.

**Suchfilter:**

- Personen: Von jedem
- Leute, denen du folgst
- Standort: Überall
- In deiner Nähe

**Trends für dich:**

- Trend in Deutschland: Iria (33.000 Tweets)
- Trend in Deutschland: Homöopathie (1.087 Tweets)
- Trend in Deutschland: Diesel (81.260 Tweets)
- Sport-Trend: #GERENIG (33.930 Tweets)
- Trend in Deutschland: Kommunikat (5.709 Tweets)
- Trend in Deutschland: Pronomen (2.685 Tweets)
- Trend in Deutschland: Büro (2.376 Tweets)
- Trend in Deutschland: Schnelltest
- Trend in Deutschland: Frauensonne
- Trend in Deutschland: Evidenz

## Facets of Search

- ▶ Language: monolingual, cross-lingual, multilingual
- ▶ Structure: atomics, fields, tree structure (e.g. XML), graph (e.g. Web)
- ▶ Media: texts, facts, images, audio, video, 3D, . . .
- ▶ Objects: products, people, companies, . . .
- ▶ Knowledge graph: Objects and relationships
- ▶ Beyond objects: answers, arguments
- ▶ static/dynamic contents (databasesstreams)

## More than Ad-Hoc search

### Some CLEF 2022 Tracks

- ▶ BioASQ: Large-scale biomedical semantic indexing and question answering
- ▶ CheckThat! Lab on Fighting the COVID-19 Infodemic and Fake News Detection
- ▶ eRisk: Early risk prediction on the Internet
- ▶ HIPE: Named Entity Recognition and Linking in Multilingual Historical Documents
- ▶ iDPP: Intelligent Disease Progression Prediction
- ▶ LifeCLEF: Biodiversity identification and prediction Challenges
- ▶ PAN Lab on Digital Text Forensics and Stylometry
- ▶ SimpleText: Automatic Simplification of Scientific Texts

# **What is Information Retrieval?**

# Definitions of Information Retrieval

## Classical definition

**IR = Content-Oriented Search**

Searching at different abstraction levels:

Syntax document as sequence of symbols

Semantics meaning of a text/media object

Pragmatics usefulness for solving my current problem

# Syntax, Semantics and Pragmatics in Image Search



Syntax colors, contours, textures

Semantics objects, properties, relationships

Pragmatics task-related

# Definitions of Information Retrieval

[Justin Zobel, SIGIR Forum 2(2017)]

Information retrieval is the study of techniques  
for **supporting human cognition** with **documents**,  
using material that is sourced from large document collections.

# Definitions of Information Retrieval

## My Favorite Definition

[**German IR SG 1992**] Information Retrieval (IR) is about *vagueness* und *uncertainty* in interactive information systems

Vagueness: user cannot give a precise specification of her information need

- ▶ vague query conditions
- ▶ iterative query formulation

Uncertainty system has uncertain knowledge about the (content of the) objects in database

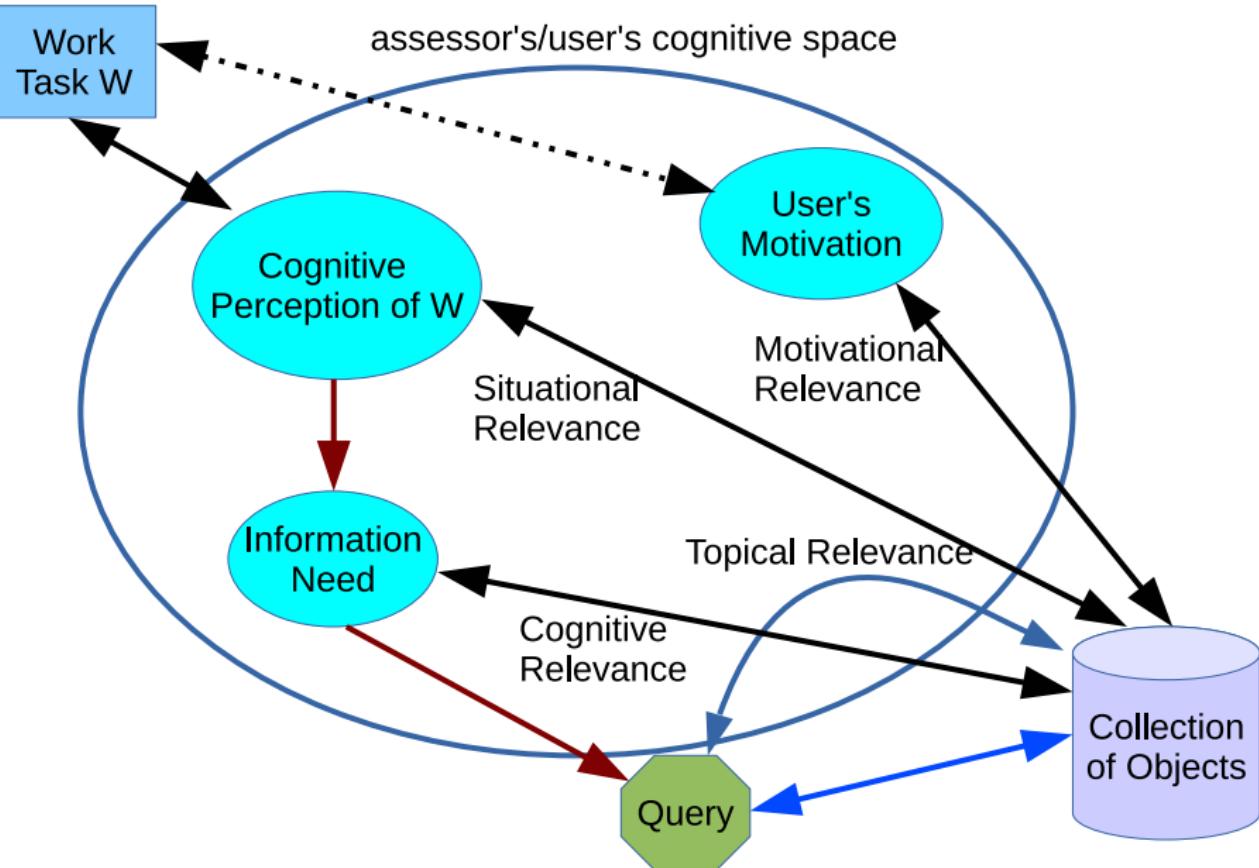
- ▶ uncertain representation  
( $\rightsquigarrow$  *wrong answers*)
- ▶ incomplete representation  
( $\rightsquigarrow$  *missing answers*)

# **What is Relevance?**

## The concept of relevance

- ▶ in contrast to databases, IR system cannot decide if an answer is correct or not
- ▶ user has **information need** originating from **work task**
- ▶ **relevance**: relationship between document and information need
- ▶ judged by user/expert

## Types of Relevance



## Types of Relevance

[Cosijn & Ingwersen 2000]

- ▶ Situational Relevance: related to the perceived task
- ▶ Pertinence relevance: related to the information need
- ▶ Intellectual topicality: as judged by human observer
- ▶ Algorithmic relevance: system score comparing request/query with object

In the following: Relevance as pertinence/topicality without further distinction

## Aspects of Relevance

[Barry/Schamber 1998]

Accuracy information is accurate

Currency information is recent

Specificity a summary of the information is available

Reliability source is expert, professional, or experienced

Accessibility source is readily available when needed

Verifiability information from this source is consistent with that  
from other sources

Clarity written or spoken language is clear and well-organized

Presentation Quality presented in permanent or stable form

Presentation Preference it has interest or entertainment value

# Implementing Relevance: Information Nutrition Labels

INFORMATION NUTRITION LABEL		
Best Before: Jan 1, 2018		
Per 1000 words	Recommended Daily Allowance	
Fact	30%	60 %
Opinion	40%	20 %
Controversy	9.0	--
Emotion	6.7	1.3
Topicality	8.7	5.0
Reading Level	4.0	8.0
Technicity	2.0	--
Authority	4.3	9.0
Viralness	--	1.0
Additional substances: advertising, subscription, invective, images (2), tweets, video clips		
Traces: product placement		

**THE OFFICIAL BREITBART STORE** [SHOP NOW >](#) [▼ SHOW](#)

## TRUMP'S ATTACK ON SESSIONS OVER CLINTON PROSECUTION HIGHLIGHTS HIS OWN 'WEAK' STANCE



by ADAM SHAW | 25 Jul 2017 | 8,865

President Trump's decision Tuesday to attack Attorney General Jeff Sessions over Sessions' "position" on Hillary Clinton's various scandals only serves to highlight Trump's own hypocrisy on the issue — and is likely to fuel concerns from his base who see

**WHATSOEVER IT TAKES** with CUFT SCHILLING 9-11AM EASTERN MONDAY-FRIDAY

**SIGN UP TO GET BREITBART NEWS DELIVERED RIGHT TO YOUR INBOX**

Enter your email address

**SIGN ME UP**

**BREITBART CONNECT**

[f](#) [t](#) [g](#) [i](#) [iTunes](#)

**MOST POPULAR**

Donald Trump Continues Criticism of Jeff Sessions Amidst Replacement Rumors  
8,911 comments - 5 hours ago

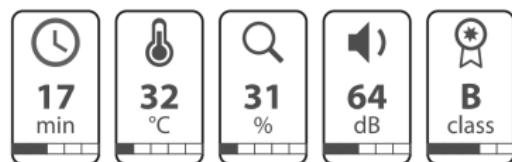
Trump's Attack on Sessions over Clinton Prosecution Highlights His Own 'Weak' Stance  
5,804 comments - 8 hours ago

INL mockup from [Fuhr et al 2018]

# Implementing Relevance: Information Nutrition Labels



[Vicentius et al. 2018]



Time, temperature, transparency, volume, credibility  
[Gollub et al. 2018]

# Implementing INL Estimation

[Fuhr et al. 2018]

- ▶ topicality
- ▶ readability
- ▶ technicality
- ▶ factuality
- ▶ virality
- ▶ emotion
- ▶ opinion
- ▶ controversy
- ▶ authority / credibility / trust

~~ automatic methods for estimating values of these aspects exist,  
but must be improved

# **Representations**

Semantic Descriptions

Free Text Search

Objects, Representations, and Descriptions

# Representations

Free text search search in document text

Semantic approach assign semantic descriptions

## Semantic Descriptions

classification schemes e.g. hierachic classification, as in libraries or product catalogs

Tagging users assign tags

Ontologies e.g. OWL: Web Ontology Language

# Classification/Ontology Example

Curlie, successor of DMOZ

**Curlie** 

About   Forum   Donate

**Collect the best websites for any topic!**

Search or browse by category

Search Curlie in English  

 <b>Arts</b> Movies, Television, Music...	 <b>Business</b> Jobs, Real Estate, Investing...	 <b>Computers</b> Internet, Software, Hardware.
 <b>Games</b> Video Games, RPGs, Gambling...	 <b>Health</b> Fitness, Medicine, Alternative...	 <b>Home</b> Family, Consumers, Cooking...
 <b>News</b> Media, Newspapers, Weather...	 <b>Recreation</b> Travel, Food, Outdoors, Humor...	 <b>Reference</b> Maps, Education, Libraries...
 <b>Regional</b> US, Canada, UK, Europe...	 <b>Science</b> Biology, Psychology, Physics...	 <b>Shopping</b> Clothing, Food, Gifts...
 <b>Society</b> People, Religion, Issues...	 <b>Sports</b> Baseball, Soccer, Basketball...	 <b>Kids and Teens</b> Arts, School Time, Teen Life...

# Free Text Search

## Problems

### Inflection

*computer – computers, fly – flies*

*go – goes – going*

### Derivation

*compute - computer - computerization - computation*

### Synonyms

*mobile – smartphone, table – bench – board – counter*

### Polysemes

*bank, head*

### Compounds

*steamboat, testbed*

### Phrases

*information retrieval – retrieval of information*

# Free Text Search

## Approaches

inflection, derivation stemming algorithms

*computer, computation, computerize* → *comput*

synonyms synonym lexicons

compounds splitting algorithms

phrases adjacency search

Most systems implement only stemming and adjacency search!

# A Document Object

## Some Inconsistencies and Misidentified Modeling Assumptions in Probabilistic Information Retrieval

WILLIAM S. COOPER  
University of California, Berkeley

---

Research in the probabilistic theory of information retrieval involves the construction of mathematical models based on statistical assumptions. One of the hazards inherent in this kind of theory construction is that the assumptions laid down may be inconsistent in unanticipated ways with the data to which they are applied. Another hazard is that the stated assumptions may not be those on which the derived modeling equations or resulting experiments are actually based. Both kinds of mistakes have been made in past research on probabilistic information retrieval. One consequence of these errors is that the statistical character of certain probabilistic IR models, including the so-called Binary Independence model, has been seriously misapprehended.

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine Systems; H.3.0 [Information Storage and Retrieval]: General; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—retrieval models

General Terms: Experimentation, Measurement, Performance, Theory

Additional Key Words and Phrases: Assumptions, bibliographic searching, consistency, document retrieval, independence, logic, modeling

---

### 1. INTRODUCTION

Probability theory provides a powerful springboard from which to launch theories of information retrieval and inductive searching. It is, of course, desirable that a formalism be logically powerful. However, such power comes at the price of a certain risk of accidental misuse and abuse. One of the hazards that an IR system designer should be aware of is that of becoming ensnared in statistical simplifying assumptions logically inconsistent with the data from which inferences must be drawn. Another danger is that the fundamental assumptions underlying a theory may be incorrectly stated, and

---

This is a revised and extended version of a paper presented at the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, October, 1991, Chicago, IL.

Author's address: School of Library and Information Studies, University of California, Berkeley, CA 94720; email: [wcooper@uclib.berkeley.edu](mailto:wcooper@uclib.berkeley.edu).

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1995 ACM 1046-8188/95/0100-0100 \$03.50

ACM Transactions on Information Systems, Vol. 13, No. 1, January 1995, Pages 100-111

## Example: document text, representation, description

### Text:

Research in the probabilistic theory of information retrieval involves the construction of mathematical models. In this kind of theory construction the assumptions laid down ...

### Stopword removal and stemming:

research probabil theory informat retriev involv construct  
mathemat model kind theory construct assume lay down

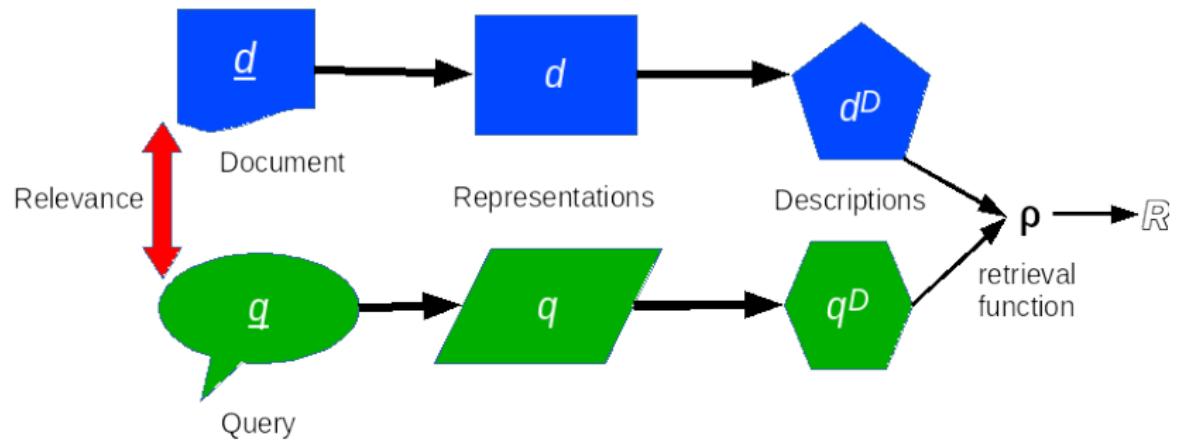
### Representation (Bag of words):

(research,1), (probabil,1), (theory,2), (informat,1), (retriev,1),  
(involv,1), (construct,2), (mathemat,1), (model,1), (kind,1),  
(assum,1), (lay,1), (down,1),

### Description:

(research,0.5), (probabil,0.5), (theory,1.0), (informat,0.5),  
(retriev,0.5), (involv,0.5), (construct,1.0), (mathemat,0.5),  
(model,0.5), (kind,0.5), (assum,0.5), (lay,0.5), (down,0.5)

# Conceptual Model



# **IR Models**

Probability Ranking Principle

Binary Independence Retrieval Model

BM25 model

Learning to Rank

# Probability Ranking Principle

defines optimum retrieval for probabilistic models:  
rank documents according to decreasing values of the

$$\text{probability of relevance } P(\text{rel}|q, d)$$

## Advantage:

PRP yields

- ▶ optimum retrieval quality
- ▶ minimum retrieval costs

## PRP Example

System computes the following probabilities of relevance  
 $P(\text{rel}|q, d)$ :

(0.9, 0.8, 0.5, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05, 0.0)

User regards the first three documents only

1. What is the expected precision? 2012
2. What is the expected recall?

1.  $p = (0.9 + 0.8 + 0.5)/3 = 0.73$

2.  $\sum_i P(\text{rel}_q, d_i) = 4, \quad r = (0.9 + 0.8 + 0.5)/4 = 0.55$

# Binary Independence Retrieval Model

[Robertson/SparckJones 1976]

represent queries and documents as sets of terms  $t_1, t_2, \dots$

$q \in Q$ : query rep.

$q^T$ : set of query terms

$d \in D$ : document rep.

$d^T$ : set of doc. terms

$$\frac{P(d|rel)}{P(d|\bar{rel})} = \frac{P(rel)}{P(\bar{rel})} \prod_{t_i \in d^T \cap q^T} \frac{p_i}{s_i} \cdot \prod_{t_i \in q^T \setminus d^T} \frac{1 - p_i}{1 - s_i}$$

$P(\bar{rel})$ : prob that random document is relevant

$p_i = P(t_i|rel)$ : prob. that  $t_i$  occurs in random relevant doc.

$s_i = P(t_i|\bar{rel})$ : prob. that  $t_i$  occurs in random nonrelevant doc.

## BIR Example

$d$	$r(d)$	$t_1$	$t_2$	$P(R d)$	BIR
$d_1$	$R$	1	1		
$d_2$	$R$	1	1		
$d_3$	$R$	1	1	0.80	
$d_4$	$R$	1	1		
$d_5$	$N$	1	1		
$d_6$	$R$	1	0		
$d_7$	$R$	1	0		
$d_8$	$R$	1	0	0.67	
$d_9$	$R$	1	0		
$d_{10}$	$N$	1	0		
$d_{11}$	$N$	1	0		

$d$	$r(d)$	$t_1$	$t_2$	$P(R d)$	BIR
$d_{12}$	$R$	0	1		
$d_{13}$	$R$	0	1		
$d_{14}$	$R$	0	1		
$d_{15}$	$N$	0	1	0.50	
$d_{16}$	$N$	0	1		
$d_{17}$	$N$	0	1		
$d_{18}$	$R$	0	0		
$d_{19}$	$N$	0	0	0.33	
$d_{20}$	$N$	0	0		

$$p_1 = \frac{8}{12} = \frac{2}{3}$$

$$s_1 = \frac{3}{8}$$

$$p_2 = \frac{7}{12}$$

$$s_2 = \frac{4}{8} = \frac{1}{2}$$

$$P(\text{rel}) = \frac{12}{20} = \frac{3}{5}$$

## BIR Example

$d$	$r(d)$	$t_1$	$t_2$	$P(R d)$	BIR	$d$	$r(d)$	$t_1$	$t_2$	$P(R d)$	BIR
$d_1$	$R$	1	1			$d_{12}$	$R$	0	1		
$d_2$	$R$	1	1			$d_{13}$	$R$	0	1		
$d_3$	$R$	1	1	0.80	0.76	$d_{14}$	$R$	0	1		
$d_4$	$R$	1	1			$d_{15}$	$N$	0	1	0.50	0.48
$d_5$	$N$	1	1			$d_{16}$	$N$	0	1		
$d_6$	$R$	1	0			$d_{17}$	$N$	0	1		
$d_7$	$R$	1	0			$d_{18}$	$R$	0	0		
$d_8$	$R$	1	0	0.67	0.69	$d_{19}$	$N$	0	0	0.33	0.40
$d_9$	$R$	1	0			$d_{20}$	$N$	0	0		
$d_{10}$	$N$	1	0								
$d_{11}$	$N$	1	0								

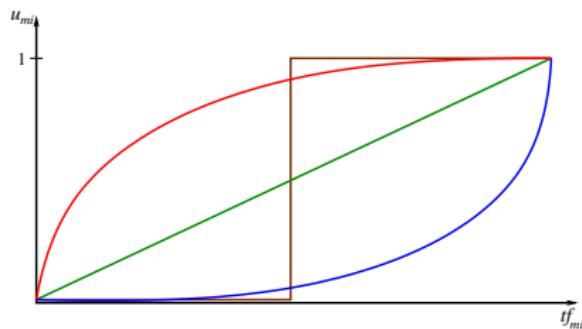
$$p_1 = \frac{8}{12} = \frac{2}{3}$$

$$s_1 = \frac{3}{8} \qquad \qquad s_2 = \frac{4}{8} = \frac{1}{2}$$

$$p_2 = \frac{7}{12}$$

$$P(\text{rel}) = \frac{12}{20} = \frac{3}{5}$$

[Robertson et al 95]  
heuristic extension of the BIR model  
from binary to weighted indexing  
(consideration of within-document frequency  $tf$ )



## tf\*idf Weighting

- ▶ originally developed for (non-probabilistic) vector space model
- ▶ set of heuristics: the weight of a term should be higher...
  1. the less frequent the term occurs in the collection  
(inverse document frequency,  $idf$ )
  2. the more often the term occurs in the document ( $tf$ )
  3. the shorter the document

## From binary to weighted Indexing

$l_m$  document length(# tokens in  $d_m$ )

$al$  average document length in  $\underline{D}$

$tf_{mi}$ : occurrence frequency of  $t_i$  in  $d_m$ .

$b$  weight of length normalization,  $0 \leq b \leq 1$

$k$  weight of occurrence frequency

length normalization:  $B = \left( (1 - b) + b \frac{l_m}{al} \right)$

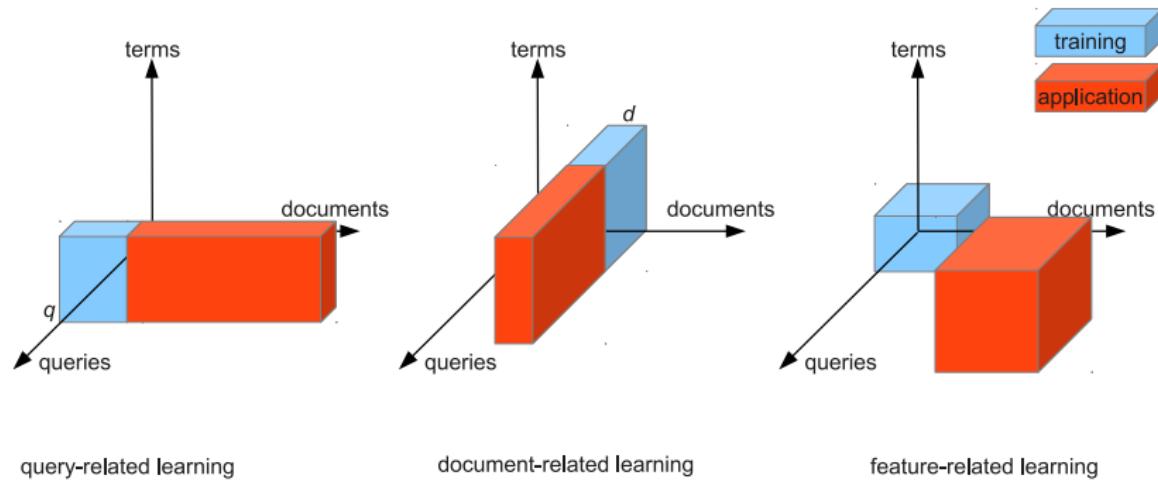
normalized within-document frequency:  $ntf_{mi} = tf_{mi}/B$

BM25 weight:  $u_{mi} = \frac{ntf_{mi}}{k + ntf_{mi}}$

$$= \frac{tf_{mi}}{k \left( (1 - b) + b \frac{l_m}{al} \right) + tf_{mi}}$$

# From Parameter Learning to Learning to Rank

[Fuhr 1992]



query-related learning

document-related learning

feature-related learning

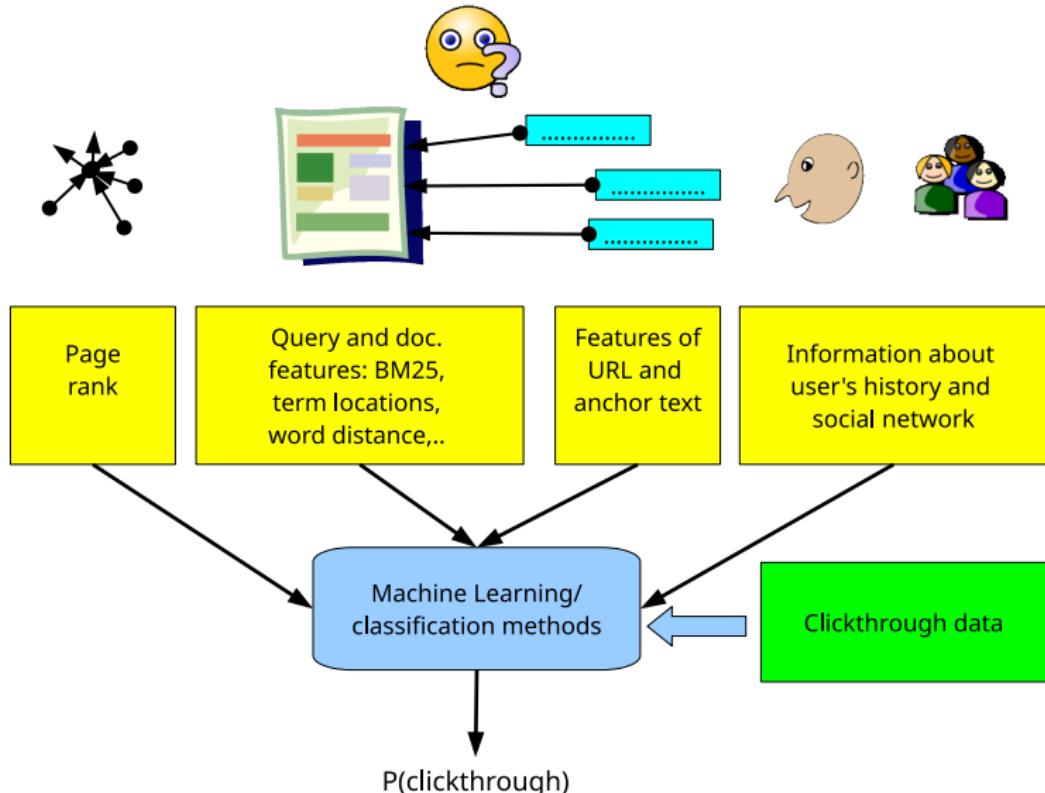
Learning approaches in IR

## Learning to rank

Instead of estimating  $P(\text{rel}|q, d)$ :

- ▶ define feature vector  $\vec{x}(q, d)$
- ▶ use training sample for learning  $g(\vec{x}(q, d)) \approx P(\text{rel}(\mathcal{R}|\vec{x}(q, d))$   
(probability that a query-document pair represented by  $\vec{x}(q, d)$  is relevant)
- ▶ apply  $g(\vec{x}(q', d'))$  to new queries  $q'$  and documents  $d'$
  
- ▶ ranking problem transformed into classification problem
- ▶ arbitrary classification methods can be used
- ▶ free choice of (complex) representation of queries and documents

# Learning to Rank for Web Searches



# Learning to Rank Approaches

[Liu 2009]

pointwise  $(q, d) \rightarrow [0, 1]$  (as above)

- ▶ training data consists of query-document pairs with relevance judgments (ordinal scale)
- ▶ learning task regarded as classification or numeric/ordinal prediction

pairwise  $(q, d, d') \rightarrow [-1, 1]$

- ▶ training data consists of  $(q, d_1, d_2)$  triples specifying a preference relation (which document is the better answer to the query)
- ▶ classification problem

listwise  $(q, d_1, d_2, d_3, \dots) \rightarrow m(d_1, d_2, d_3, \dots)$

- ▶ training data consists of queries with lists of documents specifying an ideal rank order
- ▶ algorithms aim at optimizing ranking wrt. a specific evaluation metric

# **Interactive Retrieval**

Search models

The interactive Probability Ranking Principle (IPRP)  
Approach

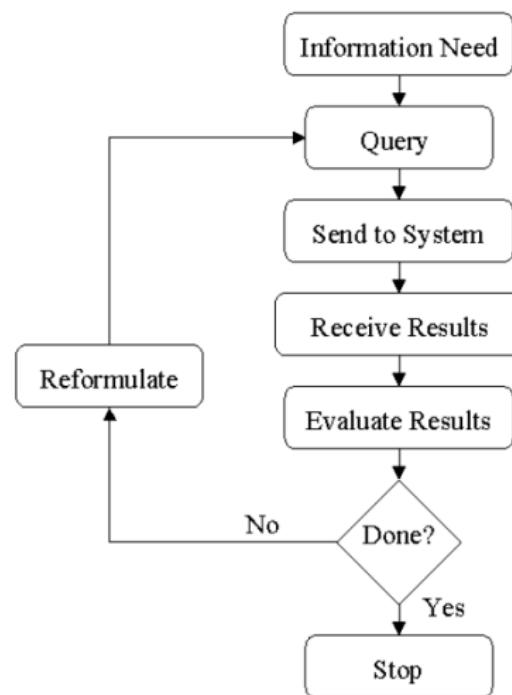
Basic Assumptions

The Model

Parameter Estimation Based on Markov Models

# Search models

## Classical search process model



## Empirical studies

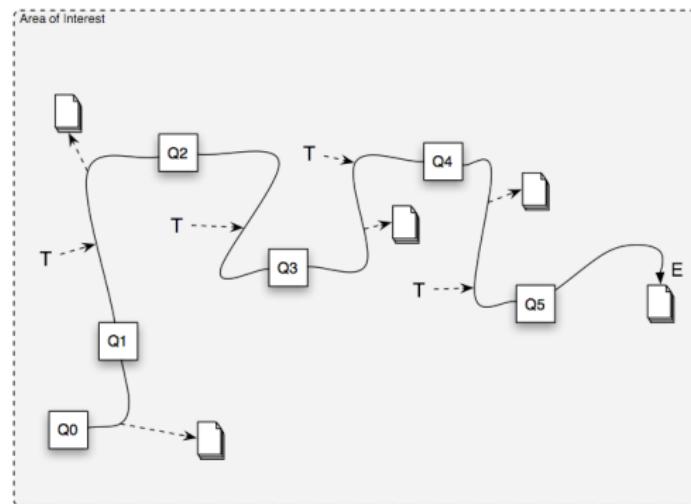
- ▶ information search consists of a sequence of connected, but different searches
- ▶ search result may trigger new searches
- ▶ only task context remains the same
- ▶ main goal of a search is accumulated learning and collection of new information while searching

# Search models

## Berry picking-Model

[Bates 90]

- ▶ continuous change of information need and queries during search
- ▶ information need cannot be satisfied by a single result set
- ▶ instead: sequence of selections and collection of pieces of information during search



## Limitations of the Probability Ranking Principle

- ▶ relevance of a document is independent of the relevance of other documents
- ▶ information need is assumed to be static
- ▶ deals with user's judgment of ranked documents only
- ▶ other tasks in interactive IR ignored

## The Interactive PRP

[Fuhr 08]

- ▶ Consider the complete interaction process
- ▶ Allow for different costs for different activities
- ▶ Allow for changes of the information need

# IPRP: IIR as Sequential Decision Making

Google search results for "interactive retrieval":

- interactive information retrieval
- hcir credentialing
- information retrieval system definition
- computer based information retrieval system

Google search results for "interactive retrieval":

All Images Videos Shopping News More Search tools

## Searches related to interactive retrieval

[interactive information retrieval](#)

[what is information retrieval](#)

[hcir credentialing](#)

[hcrl](#)

[information retrieval system definition](#)

[google scholar](#)

[computer based information retrieval system](#)

[Interactive Information Retrieval - Schloss Dagstuhl : Seminar ...](#)

<https://www.dagstuhl.de/en/program/calendar/semp/?sempn=09101> ▾

Interactive information retrieval (IIR) systems are a commodity nowadays; however, the scientific foundation of this type of system is rather limited. Information ...

[\[PDF\] Interactive Information Retrieval: Context and Basic Notions](#)

[inform.nu/Articles/Vol3/3n2p57-62.pdf](http://inform.nu/Articles/Vol3/3n2p57-62.pdf) ▾

by D Robins · Cited by 66 · Related articles

This paper provides an introduction to Interactive information retrieval—the study of ... retrieval, namely, the episodic model, the stratified model, the Interactive ...

[Human-computer information retrieval - Wikipedia, the free ...](#)

[https://en.wikipedia.org/wiki/Human-computer\\_information\\_retrieval](https://en.wikipedia.org/wiki/Human-computer_information_retrieval) ▾

Human-computer information retrieval (HCIR) is the study and engineering of information ... Early work on Interactive information retrieval, such as Juergen Koenemann and Nicholas J. Belkin's 1996 study of different levels of interaction for ...

History · What is HCIR? · Goals · Techniques

[Interactive Retrieval Based on Wikipedia Concepts](#)

[arxiv.org/cs](http://arxiv.org/cs) ▾

by L Zhang · 2014 · Cited by 2 · Related articles

Dec 29, 2014 - Computer Science > Information Retrieval ... presents a new user feedback mechanism based on Wikipedia concepts for Interactive retrieval.

[Interactive retrieval of complex documents - ScienceDirect](#)

[www.sciencedirect.com/science/article/pii/030645739090104A](http://www.sciencedirect.com/science/article/pii/030645739090104A) ▾

by WB Croft · 1990 · Cited by 36 · Related articles

The query, expressed in English, is: Find a journal article where the title is about the Interactive retrieval of office documents or Interactive retrieval, at least.

[\[PDF\] Interactive Retrieval Based on Faceted Feedback - Bad Request](#)

<https://users.soe.ucsc.edu/~yiz/papers/c23-sigir10.pdf> ▾

by L Zhang · 2010 · Cited by 36 · Related articles

Jul 19, 2010 - value pair recommendation approaches and two retrieval models that ... Keywords Interactive retrieval, faceted feedback, relevance feedback..

data 2

All Images Videos Maps News Region

4,300,000 Results Date Language Region

**Dota 2 - Official Site**

The International Battle Pass is ready to start a course into the Dota summer season and the decisive battle starts this Friday just over the International tournament.

**News**

Immortal Treasure II is now available, featuring ultimate items for Dark Seer, Viper, ...

**Updates**

Underline changes: "When deleted, the item will be removed from the game." ...

**Heroes**

Play for Free | News | Store | Hemopedia | ...

**Item Explorer**

Play for Free | News | Store | Hemopedia | ...

See results only from dota2.com

**Dota 2 - Wikipedia**

https://en.wikipedia.org/w/index.php?title=Dota\_2&oldid=1000000000

Source: Valve Corporation

Developer(s): Valve Corporation

Publisher(s): Valve Corporation

Release: June 8, 2013

OS: X, Linux, ...

**Overview** **Code** **Gamemode** **Development** **Release** **Professor** >

Dota 2 is a free-to-play multi-player online battle arena video game developed and published by Valve Corporation. It is a real-time strategy game set in a fantasy universe, originally created for Blizzard Entertainment's Warcraft III: Reign of Chaos and its expansion pack, The Frozen Throne. Dota 2 is played in matches between two teams of five players each, with team members occupying and defending their own separate base on the map. Each of the ten players is responsible... See more on en.wikipedia.org. Under CC-BY-SA license.

**Data 2 Wiki**

https://dota2.gamepedia.com/

Leading resource for Dota 2 - In-depth information on heroes, items, mechanics, and patches. Full immortals item database. Updated daily.

**News about Dota 2**

long search term

\$10,000 Dota 2 Middle East Tournament Happening This Month

Gamer Hub Middle East is organizing a massive Dota 2 tournament, inviting teams across the Middle East to compete and win a whopping \$10,000 in prize money. The Dota 2 Middle East Battle pit will take place this month, where 40 tea... See more on ... 14

Dota 2: Kinguin drops their Dota 2 roster in wake of Pro Circuit changes

Kinguin has dropped its Dota 2 roster to clarify a sensitive point. With the Core 2 manager listed to clarify a sensitive point. With the Core 2 manager listed to clarify a sensitive point. There would be a new Lolti Dead in the MINCS... 25

OpenAI teaching neural networks to compete with Dota 2 professionals

Last year, OpenAI created a bot that could defeat the best Dota 2 players in the world at 1v1 matches. Now, it wants to take that a step further via a team of five neural networks called the DotaBot. These AI research company OpenAI has been... Gamersyde... 05

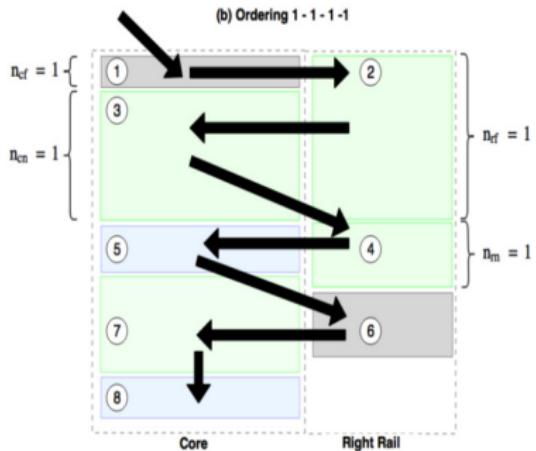
See more on Bing News

**DOTA 2 Twitter**

Twitter | @Dota2 | 🔍

The latest Tweets from DOTA 2 (@DOTA2) Dota 2 is a multi-player Action RTS game.

Weekend Sale -- Battle Level and Treasure Bundle -- Immortal Treasure II -- The 7.18 Gameplay Update is now released! blog.dota2.com/201... blog.dota2.com/news/201...



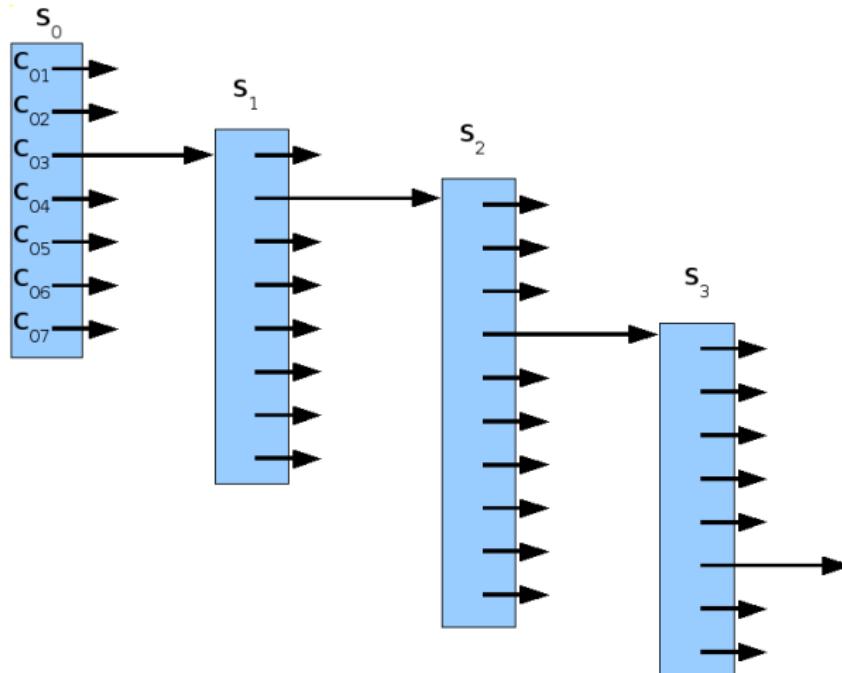
[Azzopardi et al. 18]

# The IPRP

## Basic Assumptions

- ▶ Focus on a functional level of interaction  
(usability issues disregarded here)
- ▶ System presents list of possible actions to the user
- ▶ Users evaluate possible actions in linear order
- ▶ Only positive decisions/executed actions are of benefit for a user

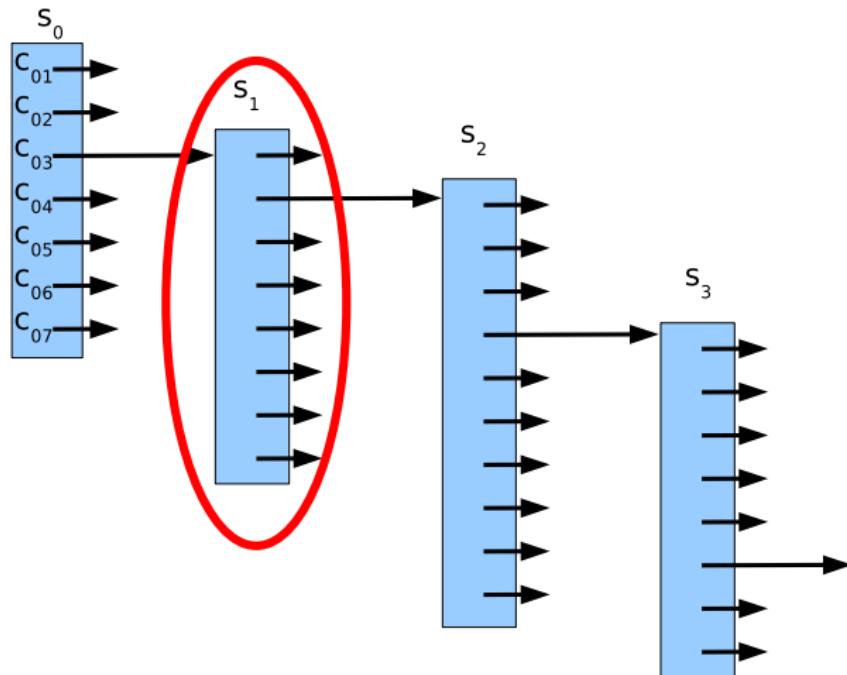
## Abstraction: Situations with Lists of Actions



## Basic idea

- ▶ A user moves from situation to situation
- ▶ In each situation  $s_i$ , the user is presented a list of possible actions  $\langle a_{i1}, a_{i2}, \dots, a_{in_i} \rangle$
- ▶ The user decides about each of these actions sequentially
- ▶ The first positive decision moves the user to a new situation  $s_j$

## Probabilistic model focusing on single situation



## Parameters for each action

$p_{ij}$  probability that user will execute  $a_{ij}$

$e_{ij} < 0$ : effort for evaluating the possible action  $a_{ij}$

$b_{ij} > 0$ : resulting benefit from positive decision

### Expected benefit of action $a_{ij}$

$$E(a_{ij}) = e_{ij} + p_{ij} b_{ij}$$

## Example for Expected Benefit

After formulating a query, a user may choose to perform the following actions with the corresponding parameter triples  $(e_{ij}, p_{ij}, b_{ij})$

1.  $(-1.0, 0.3, 8)$  add expansion term to the query
2.  $(-2.0, 0.4, 10)$  look at the first result list entry
3.  $(-10.0, 0.4, 25)$  immediately go to the first document
4.  $(-5.0, 0.3, 20)$  look at an aggregated summary of the top ranking documents

In which order should these actions be presented to the user?

1.  $(-1.0 + 0.3 \cdot 8) = 1.4$
2.  $(-2.0 + 0.4 \cdot 10) = 2$
3.  $(-10.0 + 0.4 \cdot 25) = 0$
4.  $(-5.0 + 0.3 \cdot 20) = 1$

# IPIP Research Problems

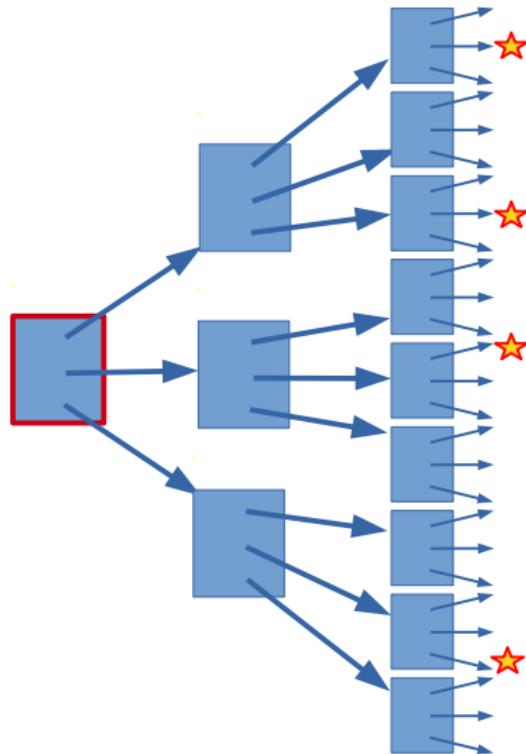
## 1. Parameter estimation

- ▶ Selection probability  $p_{ij}$ :  
focus of many IR models,  
but models for dynamic info needs required
- ▶ Effort parameter  $e_{ij}$ :  
more research needed
- ▶ Benefit  $b_{ij}$ : current research issue  
one possibility: saved effort (see below)

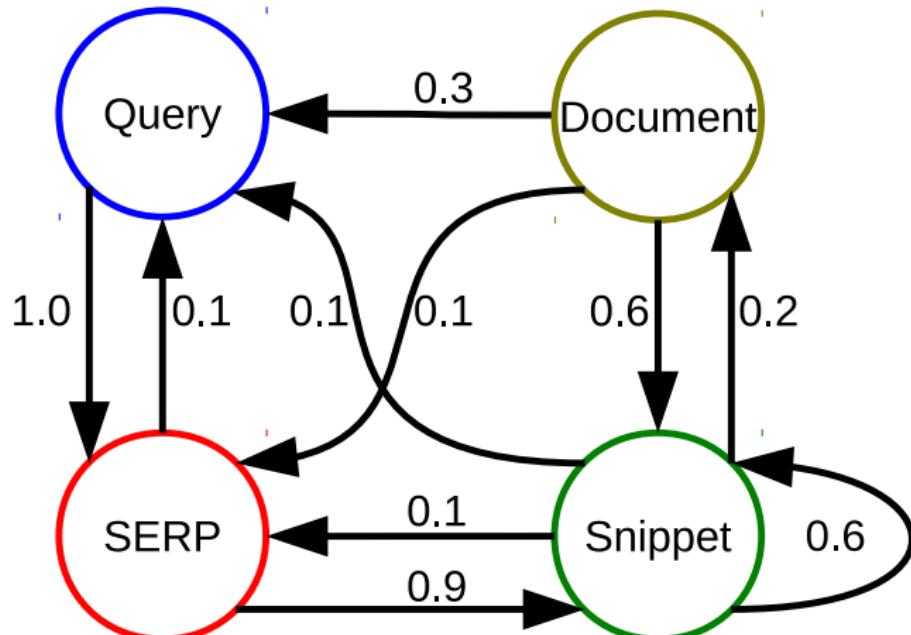
# IPRP Research Problems

## 2. Search Paths

- ▶ Choices in a situation lead to different search paths
- ▶ Search paths differ in effort for reaching a specific goal
- ▶ How to model expected effort/benefit of search paths?



## Markov models



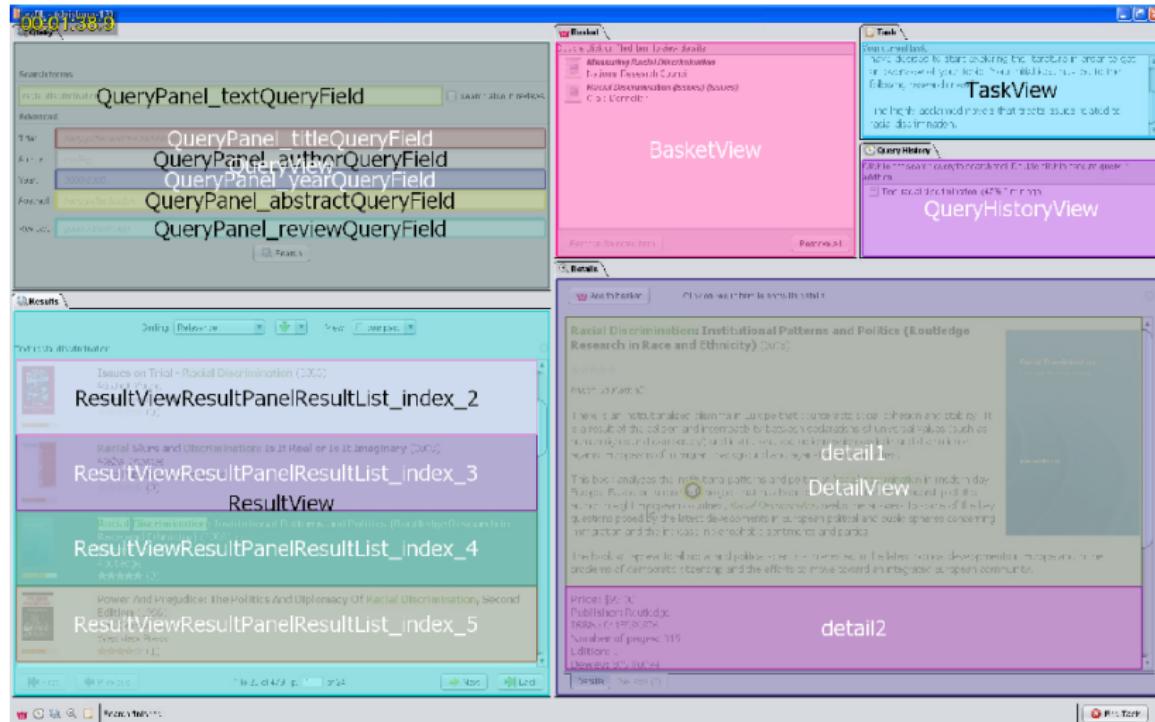
# Parameter Estimation Based on Markov Models

## User Interface

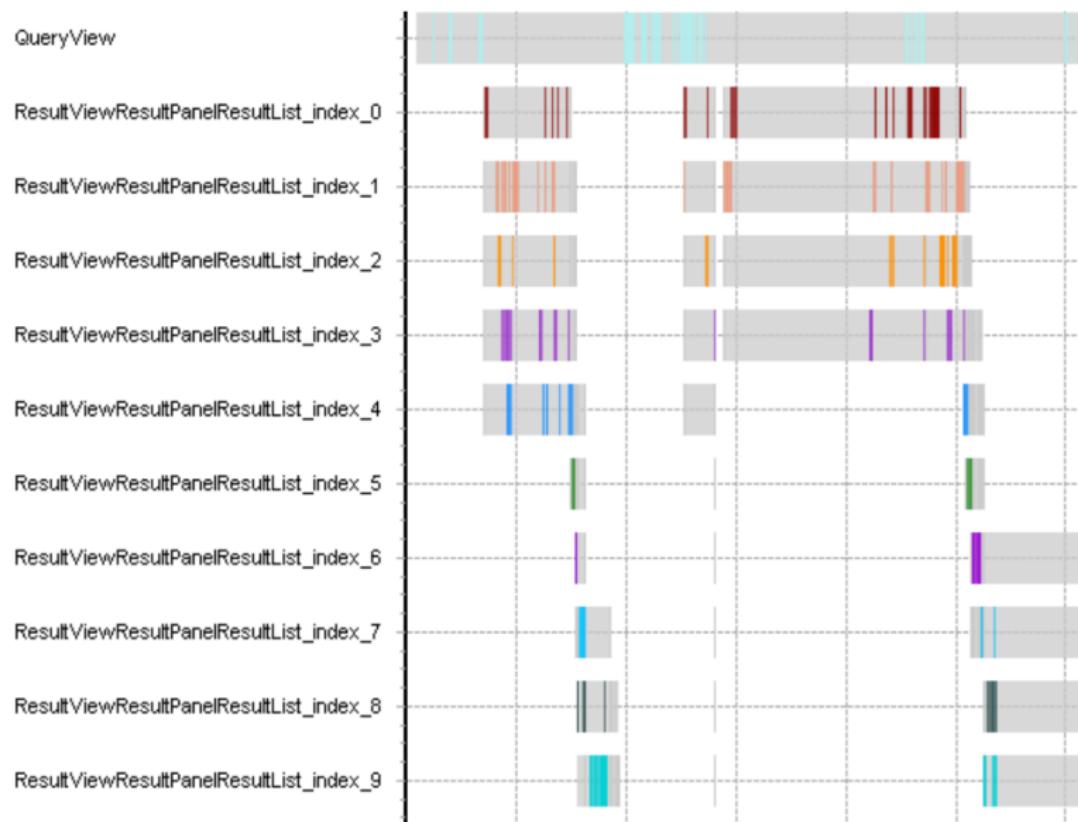
The screenshot shows a software application window divided into several panes:

- Left pane:** A search bar with "Java" typed in, followed by sections for "Author", "Title", "Subject", and "Review". Below these is a "Search" button.
- Middle-left pane:** A "Results" list containing:
  - Java, Java, Java! Object-Oriented Problem Solving (Java Series) (2002)** by Brian Goetz, Brian Goetz, Brian Goetz
  - Java Internationalization (Java Series) (2001)** by Andy Hunt, O'Reilly Media, Inc.
  - Java Cryptography (Java Series) (1997)** by Brian Goetz, Brian Goetz
  - Java 2 (2002)** by Brian Goetz, Brian Goetz
  - The Java Client: Javabeans, Volume II: jones-in-javabeans, jonesinjavabeans, jonesinjavabeans (2nd Edition) (1998)** by Brian Goetz, Brian Goetz
- Middle-right pane:** A "Details" view for the book "Java Internationalization (Java Series) (2001)". It includes a thumbnail image of the book cover, the title, author, and a short description: "For any Java programmer or developer creating software for global markets, Java Internationalization is the one-stop guide to the ins and outs of writing software that's usable around the world. It's packed with code samples and examples for translating text and dealing with locale-specific issues, such as date and time formats, currency symbols, and character sets."
- Top right pane:** A "Task Training" section with the heading "Task 1: Train a neural network to predict the outcome of the 2008 presidential election". It contains a list of steps:
  - Open History > R
  - Click the "Add new task" button
  - Click the "New Task" button
- Bottom right pane:** A preview of the book cover for "Java Internationalization" by Brian Goetz.

# Areas of Interest for Eye Tracking

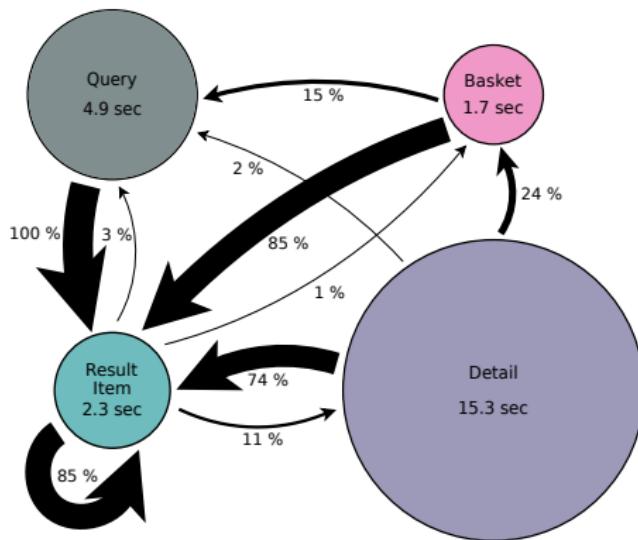


# AOI Sequence



# Markov models for estimating the IPRP parameters

[Tran & Fuhr 2012]



- ▶ Selection probability  $p_{ij}$ : transition probabilities
- ▶ Effort  $e_{ij}$ : time spent for an action
- ▶ Benefit  $b_{ij}$ : Saved time

## Expected time for reaching the basket

- ▶ effort in states  $t_q, t_r, t_d$
- ▶  $p_{XY}$ : transition probability from state  $X$  to state  $Y$
- ▶ expected times  $T_q, T_r$  and  $T_d$  for reaching the basket state

$$T_q = t_q + p_{qr} T_r$$

$$T_r = t_r + p_{rq} T_q + p_{rr} T_r + p_{rd} T_d$$

$$T_d = t_d + p_{dq} T_q + p_{dr} T_r$$

---

$$T_q = 122.7\text{s}$$

$$T_r = 117.8\text{s}$$

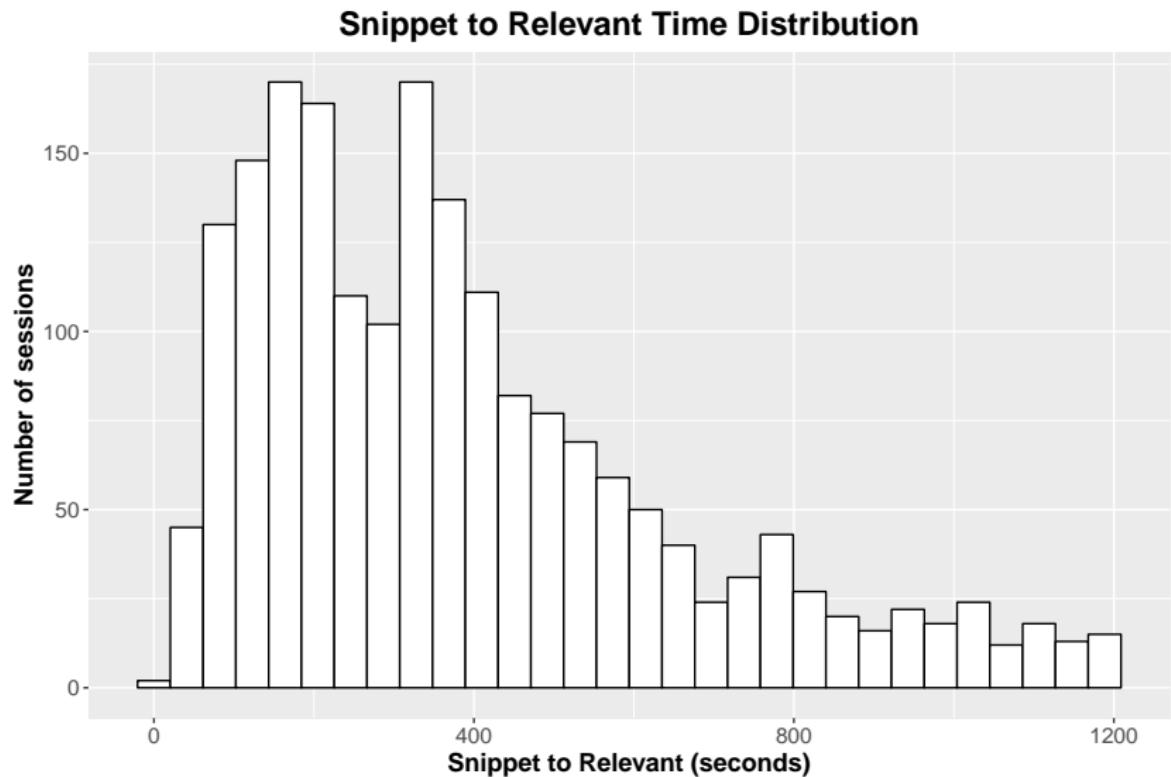
$$T_d = 104.9\text{s}$$

# Search times in an online search engine

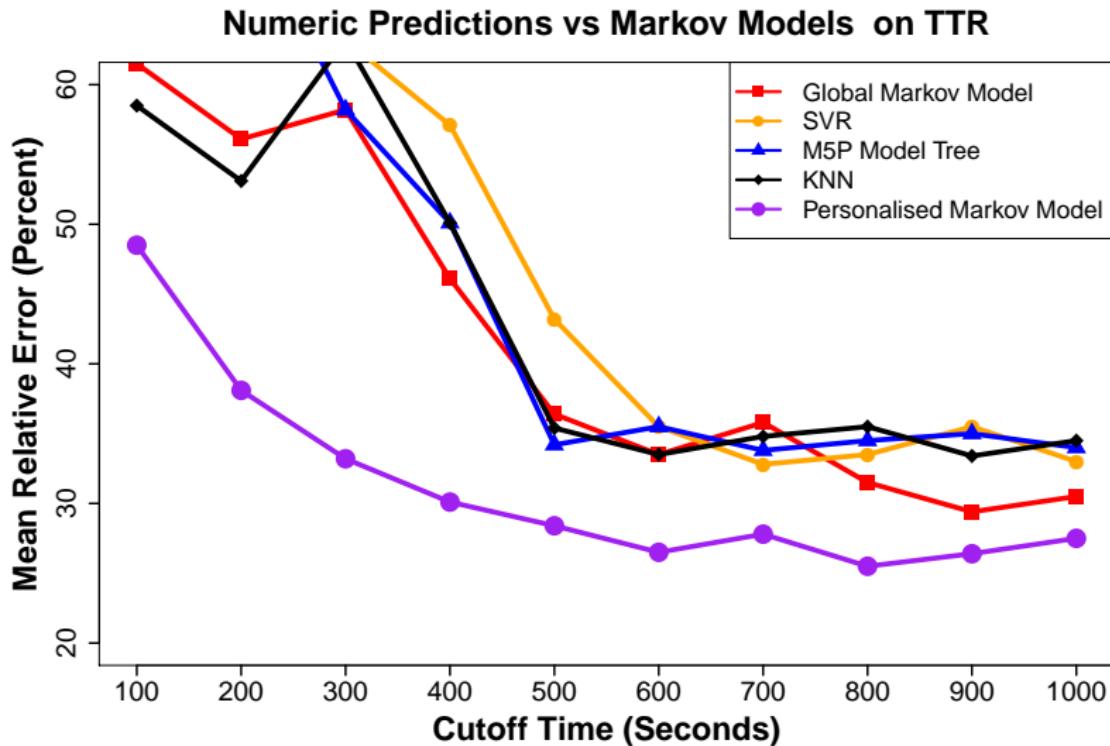
## Tran & Fuhr 2018

- ▶ Log from Sowiport social sciences search engine
- ▶ 18970 search sessions
- ▶ regard only sessions  $> 300$  sec
- ▶ 1967 sessions left

# Distribution of Times to Relevant (TTR)



## Results



## Refinements

- ▶ Document ranks
- ▶ Query reformulations
- ▶ Search phases

# Enhancing Interactive IR Context

pizzeria

## Pizzeria



## Pizzeria bei Nunzio

4,3 ★★★★★ (376) · € - Pizza

Kirchhölder Berg 7

Speisen vor Ort · Abholung vor dem Laden · Kontaktlose Lieferu

## Bistro und Pizzeria Montesino

4,5 ★★★★★ (174) · €€ - Pizza

Hagener Str. 305

Speisen vor Ort · Zum Mitnehmen · Kein Lieferdienst

## Trattoria La Conchiglia

4,4 ★★★★★ (293) · € - Restaurant

Durchstraße 22

Geschlossen · Öffnet um 17:00

„Süße Pizzeria in der Nachbarschaft.“

The screenshot shows an Amazon search results page for 'pizzeria'. At the top, there's a search bar with 'pizzeria' and a magnifying glass icon. Below it, the word 'Pizzeria' is highlighted in blue. The main content area has a header 'Lieferung an Norbert @ 44229 Dortmund' with a dropdown menu. The page is divided into several sections:

- Hallo, Norbert**: Shows a profile picture and the message 'Seit 2017'. Below it are 'Top-Links für Sie' including 'Meine Bestellungen', 'Einkommen & Preise', 'nachhaltig', and 'soziale'.
- Zuletzt angesehen**: Shows a bag of 'Bio-Schoko-Cheesecake' and a video thumbnail for 'Underground Railroad - Staffel 1'.
- Weiter ansehen**: Shows a video thumbnail for 'FLEABAG'.
- Prime-Vorteil**: Offers 'Kostenlos Lieferung GRATIS für qualifizierte Bestellungen ab 30€'.
- Video: Empfehlungen für Sie**: Shows a thumbnail for 'FLEABAG'.
- Top-Angebot**: Shows a grey orthopedic seat cushion.
- Kindle eReader**: Shows various Kindle devices.
- Auf Artikel bezogen, die Sie angesehen haben**: Shows a row of blue pop-up tents and a small image of a pizza.

# Enhancing Interactive IR

Conversational search [Ferreira et al. 2021]



u<sub>1</sub>

Q<sub>1</sub> = What is a **physician's assistant**?



u<sub>2</sub>

P<sub>1</sub> = Is a health care practitioner who practices medicine in collaboration with or under the (indirect) supervision of a physician.



u<sub>3</sub>

Q<sub>2</sub> = What are the educational requirements required to become **one**? [coreference]



u<sub>4</sub>

P<sub>2</sub> = In most cases, a physician assistant will need a master's degree from an accredited institution (two years of post-graduate education after completing a four-year degree).



u<sub>5</sub>

Q<sub>3</sub> = What is a **registered nurse**? [topic shift]



u<sub>6</sub>

P<sub>3</sub> = A registered nurse is a nurse who has graduated from a nursing program and met the requirements to obtain a nursing license.



u<sub>7</sub>

Q<sub>4</sub> = What is the difference between a **RN** and a **PA**? [context needed to decipher PA and a NP]



u<sub>8</sub>

P<sub>4</sub> = The RN model draws from the nursing tradition, including the whole person and wellness. The PA tradition draws more from a medical model.

# Conclusion

- ▶ IR is about uncertainty and vagueness in (user-oriented) information systems
- ▶ Now moving towards more ...
  - ▶ “knowledgable” objects
  - ▶ interaction + conversation
  - ▶ context: personalization + situation/work task

# References |

-  L. Azzopardi, P. Thomas, and N. Craswell.  
Measuring the utility of search engine result pages: An information foraging based measure.  
In K. Collins-Thompson, Q. Mei, B. D. Davison, Y. Liu, and E. Yilmaz, editors, *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 605–614. ACM, 2018.  
doi: 10.1145/3209978.3210027.  
URL <http://doi.acm.org/10.1145/3209978.3210027>.
-  C. L. Barry and L. Schamber.  
Users' criteria for relevance evaluation: a cross-situational comparison.  
*Information processing & management*, 34(2-3):219–236, 1998.
-  M. J. Bates.  
The design of browsing and berrypicking techniques for the online search interface.  
*Online Review*, 13(5):407–424, 1989.  
<http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>.
-  E. Cosijn and P. Ingwersen.  
Dimensions of relevance.  
*Information Processing and Management*, 36(4):533–550, July 2000.

## References II

-  R. Ferreira, M. Leite, D. Semedo, and J. Magalhães.  
Open-domain conversational search assistant with transformers.  
In *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Part I*, pages 130–145. Springer, 2021.  
doi: 10.1007/978-3-030-72113-8\_9.  
URL [https://doi.org/10.1007/978-3-030-72113-8\\_9](https://doi.org/10.1007/978-3-030-72113-8_9).
-  N. Fuhr.  
Probabilistic models in information retrieval.  
*The Computer Journal*, 35(3):243–255, 1992.
-  N. Fuhr.  
A probability ranking principle for interactive information retrieval.  
*Information Retrieval*, 11(3):251–265, 2008.  
<http://dx.doi.org/10.1007/s10791-008-9045-0>.
-  N. Fuhr, A. Giachanou, G. Grefenstette, I. Gurevych, A. Hanselowski, K. Jarvelin, R. Jones, Y. Liu, J. Mothe, W. Nejdl, I. Peters, and B. Stein.  
An information nutritional label for online documents.  
*SIGIR Forum*, 51(3):46–66, 2017.  
URL <http://sigir.org/wp-content/uploads/2018/01/p046.pdf>.

## References III

-  T. Gollub, M. Potthast, and B. Stein.  
Shaping the information nutrition label.  
In *Proceedings of the Second International Workshop on Recent Trends in News Information Retrieval co-located with 40th European Conference on Information Retrieval (ECIR 2018), Grenoble, France, March 26, 2018*, volume 2079 of *CEUR Workshop Proceedings*, pages 9–11. CEUR-WS.org, 2018.  
URL <http://ceur-ws.org/Vol-2079/paper3.pdf>.
-  V. Kevin, B. Högden, C. Schwenger, A. Şahan, N. Madan, P. Aggarwal, A. Bangaru, F. Muradov, and A. Aker.  
Information nutrition labels: A plugin for online news evaluation.  
In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 28–33, 2018.
-  T.-Y. Liu.  
Learning to rank for information retrieval.  
*Found. Trends Inf. Retr.*, 3(3):225–331, 2009.  
ISSN 1554-0669.  
doi: <http://dx.doi.org/10.1561/1500000016>.
-  S. E. Robertson.  
The probability ranking principle in IR.  
*Journal of Documentation*, 33:294–304, 1977.

## References IV

-  S. E. Robertson and K. Sparck Jones.  
Relevance weighting of search terms.  
*Journal of the American Society for Information Science*, 27:129–146, 1976.
-  S. E. Robertson, S. Walker, S. Jones, and M. M. Hancock-Beaulieu.  
Okapi at TREC-3.  
In *Proceedings of the 3rd Text Retrieval Conference (TREC-3)*, pages 109–126, Springfield, Virginia, USA, 1995. NTIS.
-  F. M. Suchanek, G. Kasneci, and G. Weikum.  
Yago: a core of semantic knowledge.  
In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 697–706. ACM, 2007.  
doi: 10.1145/1242572.1242667.  
URL <https://doi.org/10.1145/1242572.1242667>.
-  V. Tran and N. Fuhr.  
Personalised session difficulty prediction in an online academic search engine.  
In *Proceedings Theory and Practice of Digital Libraries*, pages 174–185, 2018.

## References V



V. T. Tran and N. Fuhr.

Using eye-tracking with dynamic areas of interest for analyzing interactive information retrieval.

In W. R. Hersh, J. Callan, Y. Maarek, and M. Sanderson, editors, *Proceedings of the 35th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1165–1166. ACM, 2012.

ISBN 978-1-4503-1472-5.



J. Zobel.

What we talk about when we talk about information retrieval.

*SIGIR Forum*, 51(3):18–26, 2017.

ISSN 0163-5840.

doi: 10.1145/3190580.3190584.

URL <http://doi.acm.org/10.1145/3190580.3190584>.