


Personalisierung

Ziel: Auflösen der inhärenten Ambiguität von Suche


Beispiel 1: Suche nach „IR“ könnte liefern

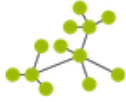
- Ingersoll-Rand Company
- Webseiten in Arabisch aus dem Iran (*.ir)
- Infrarotlicht
- Interregio (Deutsche Bahn)
- Iridium
- Israel Railways
- Information Retrieval
- ...

Ziel: Auflösen der inhärenten Ambiguität von Suche





Alle



 **deusu**

[Open-Source](#)
[Open-API](#)

[Blog](#)
[Spenden](#)

[Impressum](#)
 

Ungefä

Web

Bild

Finden

[Ir Co](#)
[www.s](#)
Conjug
subjun

427.000.000 ER

Suche im gesamten Datenbestand: 2.364.774.288 Webseiten
Es wurden insgesamt 1.202.000 Ergebnisse gefunden.

[IR – Wikip](#)
<https://de.wikip>
IR steht als Abl
Gewichtsfunktio

[Dragonir PokWiki](#)
www.pokewiki.de/Dragonir
Netzwerk Meine Werkzeuge Aktivieren Menü Navigation PokWiki Hilfen Werkzeuge Namensräume Varianten Varianten
Ansichten Mehr Suche Dragonir Aus PokWiki Wechseln zu: , Dieses Pokemon war . ? ? Dragonir Artwork von

[Pandir PokWiki](#)
www.pokewiki.de/Pandir
Netzwerk Meine Werkzeuge Aktivieren ? Neue Verlosung im PokWiki! ? Wir verlosen tolle Preise unter allen Benutzern, die
aktiv an Artikeln rund um die 7. Generation und mitwirken! Mit dabei: ein New Nintendo 3DS XL , Pokemon-Center-
Stofftiere und no

[Zur BASF-In](#)
www.basf.com/g
IR Magazin. IR
BASF-Glossar;

[IR Prognosi](#)
www.ir.com ▼
Enterprises rely
communications

[IR.on Invest](#)
www.ir-on.com ·
Die IR.on AG is
Finanzkommuni

[Infrarot - V](#)

[RSS \(beta\)](#)

[Better Call Saul - Serienhit wird um eine dritte Staffel ver](#)
www.gamestar.de/news/kino/3269459/better_call_saul.html
Saul übernimmt auch im nächsten Jahr neue Fälle: Der US-Sender AMC kündigt eine dritte Staffel Better Call Saul mit Bob
Odenkirk an. - Seite 1

[Infrarot-Verlängerung & Verstärker IR Extender & Repeater](#)
www.ir-extender.net/ge/
Kaufen höchste Qualität entfernt Control Infrarot-Verlängerung, Infrarot Verstärker und auch erhalten IR Extender, IR Repeater
aus YZC Elektronik & Technologies Limited.

[Billigflüge mit Ryanair ab 19 - Low-Cost Flugtickets - B](#)

 Universität

8-3

Zum Vergleich: Ergebnisse von 2010

ir

Web Finance

ALL RESULTS

1-10 of 63,600,000 results · [Advanced](#)

Sponsored sites

[ir camera](#) · [www.Fluke.com/IR-Camera](#)

Sponsored sites

[Ir Receiver](#)

[Buy Ir Receiver](#) You may get 8% of

Google

ir

About 452,000,000 results (0.31 seconds)

Everything

News

More

Any time

Latest

Past 2 weeks

More search tools

9:30

12:00

2:00

4:00

34.06

0.00 (0.00%)

See r

34.07

34.06

34.05

Open: 0.00

Day's Rang

Volume: NA

52 Week R

P/E Ratio: 22.40

Market Cap

Similar market cap stocks

LUX 24.49 0.00% NBL 65.55 0.00% TEPCF 21.48 0.00%

Find more stocks [Data in USD](#) · [Quoted by Comstock](#) · [Jul 7, 3](#)

[Infrared - Wikipedia, the free encyclopedia](#)

Overview · Origins of the term · Different regions in ... · Applicati

Infrared radiation (IR radiation) is electromagnetic radiation with micrometres, which equates to a frequency range between appr

[en.wikipedia.org/wiki/Infrared](#) · [Cached page](#)

[IR - Wikipedia, the free encyclopedia](#)

Ir or ir may refer to: Ir. (Ingénieur), abbreviation for the title of Er

profession of engineering; Iridium (chemical symbol Ir), the 77th

[en.wikipedia.org/wiki/Ir](#) · [Cached page](#)

[International Rectifier - The Power Management](#)

IR Launches Online Design Tool to Further Simplify and Speed

Integrated Voltage Regulators more... IR Expands Its Portfolio o

[www.irf.com](#) · [Cached page](#)

[Infrared - Wikipedia, the free encyclopedia](#)

☆

Infrared radiation (IR radiation) is electromagnetic radiation with a wavelength between 0.7 and 300 micrometres, which equates to a frequency range between ...

[en.wikipedia.org/wiki/Infrared](#) · [Cached](#) · [Similar](#)

[Information retrieval - Wikipedia, the free encyclopedia](#)

☆

Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of ...

[en.wikipedia.org/wiki/Information_retrieval](#) · [Cached](#) · [Similar](#)

[+ Show more results from en.wikipedia.org](#)

[Immersion Research 2010](#)

☆

IR Drysuits are easy to put on and take off by yourself. ... switched from paddling in my IR Comp drytop here in the Southeast to living in my Double D for ...

[www.immersionresearch.com/](#) · [Cached](#) · [Similar](#)

[ir@IIT:Main Page](#)

☆

Information Retrieval Laboratory. Research and development in information retrieval technologies.

[www.ir.iit.edu/](#) · [Cached](#) · [Similar](#)

[Welcome to Indian Railway Passenger reservation Enquiry](#)

☆

Plan you trip on Indian Railways Trains, Check Availability, Seats Availability, Reservation, Enquiry, PNR, Passenger Status, Travel by Indian Railways in ...

[www.indianrail.gov.in/](#) · [Cached](#) · [Similar](#)

[GOVERNMENT OF INDIA : INDIAN RAILWAYS](#)

☆

Indian Railways Budget 2010-11. White Paper on Indian Railways (in Hindi) (in English)

Indian Railways Vision 2020 (in Hindi) (in English). Tenders ...

[www.indianrailways.gov.in/](#) · [Cached](#) · [Similar](#)

[Welcome to Ingersoll Rand Company](#)

☆

International supplier to transportation, manufacturing, construction, and agricultural industries

Universität Trier

8-4








Ziel: Auflösen der inhärenten Ambiguität von Suche

duckduckgo.com
(nicht personalisiert)

ir

Web Bilder Videos | **Bedeutungen** Börse

Ergebnisse für ir

Iran  Iran, vollständig Islamische Republik Iran, auch als der Iran oder veraltet als Persien...	Iridium Iridium ist ein chemisches Element mit dem Symbol Ir und der Ordnungszahl 77.	Israel Railways Israel Railways ist die staatliche israelische Eisenbahngesellschaft mit	Industrielle Revolution  Als Industrielle Revolution wird die tiefgreifende und dauerhafte Umgestaltung	Industrieroboter Ein Industrieroboter ist eine universelle, programmierbare Maschine zur Handhabung,
Infanterieregiment  Das Regiment ist eine mittelgroße militärische Formation.	Information Retrieval Information Retrieval bzw.	InfrastrukturRecht Monatlich erscheinende Zeitschrift aus dem Hause C. H. Beck.	Insulinrezeptor Der Insulinrezeptor ist dasjenige Membranprotein, an welches ausgeschüttetes	International Rectifier Die International Rectifier Corporation ist ein US-amerikanischer Hersteller
Interne Revision Interne Revision ist eine vom Tagesgeschäft unabhängige, objektive Prüfungs- und...	Interregio  Der Interregio ist eine Zugart, die in mehreren europäischen Staaten	Investor Relations Investor Relations bzw. Finanzkommunikation ist ein Handlungsfeld der...	Islamrat für die  Der Islamrat für die Bundesrepublik Deutschland e.V. ist neben der größeren...	Isopren-Kautschuk Isopren-Kautschuk ist die künstliche produzierte Variante des
Infanterie  Als Infanterie bezeichnet man sich zu Fuß bewegende und kämpfende, mit Handwaffen...	Iriver  Das Internationale Unternehmen Iriver ist ein Hersteller von MP3-Playern,	Interrupt In der Informatik versteht man unter einem Interrupt eine vorübergehende	.ir .ir ist die länderspezifische Top-Level-Domain Irans.	

Germany Irgendwann

Rückmeldung senden

8-5

Ziel: Auflösen der inhärenten Ambiguität von Suche

Beispiel 2: Suche nach „Java“ sollte liefern

- Programmierwerkzeuge für einen Programmierer
- Tutorials für einen Lehrer
- FAQ-Listen für einen neuen Benutzer

} **globaler Kontext
des Benutzers**

Beispiel 3: Suche nach „restaurant“ sollte liefern

- Orte in Stuttgart während der Planung für BTW 17
- Orte in München während der Planung für VLDB 17
- Orte in Trier während der Arbeitszeit

Beispiel 4: Suche nach „Trier“ sollte liefern

- Restaurants (wenn ich vorher danach gesucht habe)
- Geschäfte, Zahnärzte, Krankenhäuser, ...

**Suchergebnisse können vom aktuellen Kontext abhängen
(der nicht konstant ist und sich mit der Zeit verändert)**

Dimensionen von Personalisierter Suche

- **Verschiedene Arten des Benutzerkontextes:**
 - **global**: Hintergrund des Benutzers, Langzeitprofil
 - **Sitzung**: Menge der Anfragen mit ähnlichen Bedürfnissen
 - **Anfrage**: verwende letzte Anfrage & folgende Aktionen/Clicks
- **Jeweils**
 - nur für Suchen
 - für alle Browseraktionen (z.B. Ausdrucken, Aufnahme in Favoriten)
 - für (andere/alle) Aktionen (z.B. Schreiben einer Mail an Hotel, Kauf eines Buchs)
- **Kontext kann an verschiedenen Stellen gesammelt und genutzt werden:**
 - Dienstanbieter vs. Webserver vs. Lokaler Rechner
- **Kontext kann auf verschiedene Weise genutzt werden:**
 - Modifiziere Anfrage
 - Verändere Ranking der Ergebnisse („reranking“)


Einfache Personalisierung: Relevance Feedback

- Sammle **Feedback** des Benutzers für Anfrageergebnisse
 - Explizites Feedback (Knopf im Interface)
 - Implizites Feedback (Klicks des Benutzers)
- Generiere **verbesserte Anfrage**
 - Füge neue Terme hinzu
 - Lösche existierende Terme
 - Ändere Gewicht von Termen

Beispiel: Einfaches Feedback mit iGoogle


2007-2013


Web [Bilder](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [E-Mail](#) [Mehr ▼](#)


 [Erweiterte Suche](#)
[Einstellungen](#)


Suche: ☒ Das Web ☐ Seiten auf Deutsch ☐ Seiten aus Deutschland


Web

[International Rectifier - The Power Management Leader](#) - [[Diese Seite übersetzen](#)]
Manufacturer of power semiconductors (MOSFET, IGBT, Diodes and Thyristors).
[Hexfet](#) - [Careers](#) - [Contact Us](#) - [Application Notes](#)
www.irf.com/indexnsw.html - [Im Cache](#) - [Ähnlich](#) - 

[Infrarotspektroskopie – Wikipedia](#)
Die IR-Spektroskopie wird zur quantitativen Bestimmung von bekannten Substanzen, ...
Spektroskopie im mittleren Infrarot – häufig nur als ...
de.wikipedia.org/wiki/Infrarotspektroskopie - [Im Cache](#) - [Ähnlich](#) - 

[Infrarotstrahlung – Wikipedia](#)
Als Infrarot wird der Bereich zwischen 780 nm und 1 mm (das sind 1.000.000 nm) ...
Umgangssprachlich wird IR-Licht oft mit Wärmestrahlung gleichgesetzt, ...
de.wikipedia.org/wiki/Infrarotstrahlung - [Im Cache](#) - [Ähnlich](#) - 
[Weitere Ergebnisse von de.wikipedia.org »](#)

[IR-Spektroskopie](#)
Auf dieser privaten Seite werden die Grundlagen der IR-Spektroskopie und die Technik der FTIR-Spektrometer erklärt.
www.ir-spektroskopie.de/ - [Im Cache](#) - [Ähnlich](#) - 

[Initiativkreis Ruhr - Home](#)
Die Repräsentanten der führenden Wirtschaftsunternehmen sowie des öffentlichen Lebens zwischen Rhein und Ruhr arbeiten seit 1989 zusammen, ...
www.i-r.de/ - [Im Cache](#) - [Ähnlich](#) - 

[Infrared - Wikipedia, the free encyclopedia](#) - [[Diese Seite übersetzen](#)]
Infrared (IR) radiation is electromagnetic radiation whose wavelength is longer than that of

Give positive or negative feedback for results

Beispiel: Einfaches Feedback mit iGoogle



Web

Ergebnisse beinhalten Ihre SearchWiki-Hinweise für **ir**. [+ Diese Hinweise weitergeben](#)

Infrarot-Strahlung **IR**




Infrarotstrahlung (**IR**-Strahlung) - auch als Wärmestrahlung bezeichnet - ist Teil der optischen Strahlung und damit Teil des elektromagnetischen Spektrums. ...

www.bis.de/de/uv/ir - [Im Cache](#) - [Ähnlich](#) -   

 1  0 - Sie sind der Erste, der dieses Ergebnis auswählt.



Infrarotspektroskopie – Wikipedia

Die **IR**-Spektroskopie wird zur quantitativen Bestimmung von bekannten Substanzen, ... Die Spektroskopie im mittleren **Infrarot** – häufig nur als ...

de.wikipedia.org/wiki/Infrarotspektroskopie - [Im Cache](#) - [Ähnlich](#) -   

Infrarotstrahlung – Wikipedia

Als **Infrarot** wird der Bereich zwischen 780 nm und 1 mm (das sind 1.000.000 nm) ... Umgangssprachlich wird **IR**-Licht oft mit Wärmestrahlung gleichgesetzt, ...

de.wikipedia.org/wiki/Infrarotstrahlung - [Im Cache](#) - [Ähnlich](#) -   

[Weitere Ergebnisse von de.wikipedia.org »](#)




IR-Spektroskopie

Auf dieser privaten Seite werden die Grundlagen der **IR**-Spektroskopie und die Technik der FTIR-Spektrometer erklärt.

www.ir-spektroskopie.de/ - [Im Cache](#) - [Ähnlich](#) -   

Infrared - Wikipedia, the free encyclopedia - [[Diese Seite übersetzen](#)]


Infrared (**IR**) radiation is electromagnetic radiation whose wavelength is longer than that of visible light (400-700 nm), but shorter than that of terahertz ...

en.wikipedia.org/wiki/Infrared - [Im Cache](#) - [Ähnlich](#) -   

AIM INFRAROT-MODULE GmbH

Die AIM **INFRAROT**-MODULE GmbH entwickelt und fertigt **Infrarot**-Detektoren und –Module für Thermografiesysteme. Das Hightech-Unternehmen ist weltweit als ein ...

www.aim-ir.de/ - [Im Cache](#) - [Ähnlich](#) -   

Sie haben Ergebnisse für diese Suche entfernt.  [Ausblenden](#)

Feedback im Google-Interface

ir Suche

Ungefähr 457.000.000 Ergebnisse (0,23 Sekunden) Erweiterte Suche

Markierte Ergebnisse für ir

★ [Infrarot-Strahlung \(IR\) - www.bfs.de/de/uv/ir](http://www.bfs.de/de/uv/ir)

[Workshop IR - LWA2010 - Lernen, Wissen, Adaptivität. 4.-6. Oktober...](#) ☆ 3 Besuche - 17. Juni
Webseite des Fachgebiets Wissensverarbeitung, Fachbereich 17, Universität Kassel
www.kde.cs.uni-kassel.de/conf/lwa10/ir - Im Cache

[Infrarotspektroskopie – Wikipedia](#) ☆
Zu [Lage der IR-Absorptionsbanden](#) springen: IR-Spektren werden dahingehend interpretiert, dass man aus der Kurve des gemessenen IR-Spektrums die ...
Varianten - Messprinzip - IR-Spektroskopie-Techniken - Siehe auch
de.wikipedia.org/wiki/Infrarotspektroskopie - Im Cache - Ähnliche

[Infrarotstrahlung – Wikipedia](#) ☆
Als Infrarotstrahlung (kurz IR-Strahlung, auch Ultrarotstrahlung) bezeichnet man in der Physik elektromagnetische Wellen im Spektralbereich zwischen ...
de.wikipedia.org/wiki/Infrarotstrahlung - Im Cache - Ähnliche
+ Weitere Ergebnisse anzeigen von de.wikipedia.org

[IR-Spektroskopie](#) ☆
Auf dieser privaten Seite werden die Grundlagen der IR-Spektroskopie und die Technik der FTIR-Spektrometer erklärt.
www.ir-spektroskopie.de/ - Im Cache - Ähnliche

[Infrarot-Strahlung \(IR\)](#) ☆
Infrarotstrahlung (IR-Strahlung) - auch als Wärmestrahlung bezeichnet - ist Teil der optischen Strahlung und damit Teil des elektromagnetischen Spektrums. ...
www.bfs.de/de/uv/ir - Im Cache - Ähnliche

[Infrared](#) - Wikipedia, the free encyclopedia ☆ [Diese Seite übersetzen]

iGoogle: Kollaboratives Feedback



Alle Such-Wiki-Einträge



19 Einträge gespeichert für: **ir**

[IR Japan | 株式会社アイ・アールジャパン](#) - [[Diese Seite übersetzen](#)]

IR活動支援のパイオニア、アイ・アールジャパンは委任状争奪戦などの有事においてはプロキシーアドバイザー (PA) として、また、平時においては買収防衛策導入支援、...

www.irjapan.net/ - vor 6 Stunden gefunden - [Im Cache](#) - [Ähnlich](#) -

2 0

[Infrarot-Strahlung \(IR\)](#) - [Wiederherstellen](#)

Infrarotstrahlung (IR-Strahlung) - auch als Wärmestrahlung bezeichnet - ist Teil der optischen Strahlung und damit Teil des elektromagnetischen Spektrums. ...

www.bfs.de/de/uv/ir - [Im Cache](#) - [Ähnlich](#) -

1 0

[Welcome to WAR LIBRARY](#) - [[Diese Seite übersetzen](#)]

welcome to war library, you can find many books about wars in this library in persian , english and arabic ,....

www.warlib.ir/ - [Im Cache](#) - [Ähnlich](#) -

1 0

[Time Warner: Time Warner Inc.: Investor Relations](#) - [[Diese Seite übersetzen](#)]

Reports & SEC Filings. Financial publications, earnings information and SEC filings, including Annual Reports; 10-K, 10-Q and 8-K reports; proxy statements; ...

ir.timewarner.com/ - vor 8 Stunden gefunden - [Im Cache](#) - [Ähnlich](#) -

1 0

[إذاعة الجمهورية الإسلامية في إيران | الصفحة الرئيسية](#)

زوروا موقع إذاعة طهران العربية على الانترنت للاطلاع على آخر الاخبار والانباء والتقارير وأهم عناوين الصحف العربية والایرانية وأحداث العالم عبر الصور.

arabic.trib.ir/ - vor 3 Stunden gefunden - [Im Cache](#) - [Ähnlich](#) -

1 0 [Kommentar](#)

[Internal Revenue Service](#) - [[Diese Seite übersetzen](#)]

US government agency responsible for tax collection and tax law enforcement. Provides downloadable income tax forms, instructions, agency publications

Implizites Feedback durch Clicks

Allgemeine Regeln zum Sammeln von impliziten Feedback:

- **Geclickte Ergebnisse** sind **relevant** für die Anfrage
 - Außer der Benutzer hat die Seite sofort wieder verlassen
- **Nicht geclickte** Ergebnisse erlauben **keine Aussage**
 - Benutzer könnte sie sofort als nichtrelevant erkannt haben (auf Basis des Snippets)
 - Benutzer könnte das Ergebnis bereits kennen (welches relevant oder nichtrelevant sein könnte)
 - Benutzer könnte das Ergebnis überhaupt nicht angesehen haben (andere Ergebnisse waren ausreichend)

<https://myactivity.google.com>



Ihre Daten verbessern Ihre Google-Erfahrung

Ihre Aktivitäten helfen dabei, Google schneller, intelligenter und nützlicher zu gestalten.

Verbessertes implizites Feedback

Ergänze **Überwachungsfunktionen** im Browser, um Benutzerverhalten zu beobachten:

- Wie lange bleibt ein Benutzer auf einer Seite
- Wohin scrollt er, welche Bereiche sieht er wie lange an
- Mausbewegungen, z.B. über Textstellen
- Mausclicks
- Geclickte Links

⇒ Erlaubt eine bessere Schätzung der Relevanz

iGoogle: Logging von Suchen und Clicks (2010)

Webprotokoll für ralf.schenkel@gmail.com

Gesamtes Protokoll

[Web](#)
[Bilder](#)
[News](#)
[Produkte](#)
[Anzeigen](#)

Heute

11:12

Gesucht nach [google search personalization](#) -  3 Ergebnisse angezeigt

- ☆ [Google Ramps Up Personalized Search](#) - [searchengineland.com](#)
- ☆ [The Future of Google's Search Personalization - Search...](#) - [searchenginewatch.com](#)
- ☆ [Google Web History and Search Personalization](#) - [googletutor.com](#)

Context-basierte Suche auf Google (2010)



Personalisierte Suche: **ir**

Wenn möglich, passt Google Ihre Suchergebnisse anhand des Standortes und/oder Ihrer aktuellen Suchvorgänge an. Wenn Sie in Ihrem Google-Konto angemeldet sind, erhalten Sie außerdem basierend auf Ihrem Suchverlauf meist noch relevantere und nützlichere Ergebnisse.

Die folgenden Informationen wurden verwendet, um Ihre Suchergebnisse für **ir** zu verbessern:

Webprotokoll	Ein oder mehrere Elemente in Ihrem Webprotokoll wurden zur Verbesserung der Suchergebnisse verwendet. Webprotokoll verwalten Webprotokoll aus meinem Google-Konto entfernen
---------------------	---

Wenn Sie neugierig sind, können Sie sich anschauen, wie eine Suche nach **ir** [ohne Anpassungen](#) aussieht.

Der Link 'Weitere Details' auf der Suchergebnisseite kann verwendet werden, um diese Seite für etwa 30 Minuten anzuzeigen. Danach wird diese Seite nicht länger angezeigt.

[Markierte Ergebnisse für ir](#)

★ [Infrarot-Strahlung \(IR\)](#) - www.bfs.de/de/uv/ir

[Infrarotspektroskopie – Wikipedia](#) ☆

Zu [Lage der IR-Absorptionsbanden](#) springen: IR-Spektren werden dahingehend interpretiert, dass man aus der Kurve des gemessenen IR-Spektrums die ...

[Varianten](#) - [Messprinzip](#) - [IR-Spektroskopie-Techniken](#) - [Siehe auch](#)
de.wikipedia.org/wiki/Infrarotspektroskopie - [Im Cache](#) - [Ähnliche](#)

[Infrarotstrahlung – Wikipedia](#) ☆

Als Infrarotstrahlung (kurz **IR**-Strahlung, auch Ultrarotstrahlung) bezeichnet man in der Physik elektromagnetische Wellen im Spektralbereich zwischen ...

de.wikipedia.org/wiki/Infrarotstrahlung - [Im Cache](#) - [Ähnliche](#)

⊕ [Weitere Ergebnisse anzeigen von de.wikipedia.org](#)

Persönliches Google Web-Protokoll (2010)

Web History for ralf.schenkel@gmail.com

Web History

[Web](#)

[Images](#)

[News](#)

[Products](#)

[Sponsored Links](#)

[Video](#)

[Maps](#)

[Blogs](#)

[Books](#)

Pause

Remove items

Trends

Bookmarks

Show trends for: [Last 7 days](#) | [Last 30 days](#) | [Last year](#) | [All time](#)

Top queries

1. [vldb 2010](#)
2. [sigir 2010](#)
3. [icde 2010](#)
4. [dblp ralf schenkel](#)
5. [cikm 2010](#)
6. [ecir 2010](#)
7. [ralf schenkel](#)
8. [wise 2010](#)
9. [trec](#)
10. [sigmod 2010](#)

Top sites

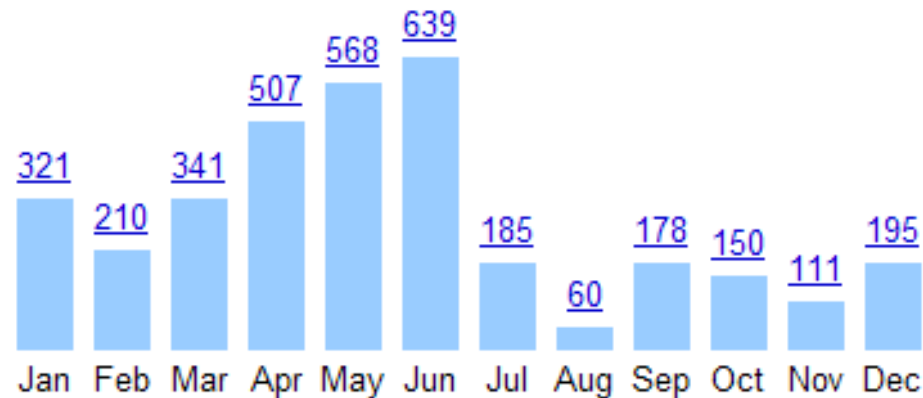
1. www.informatik.uni-trier.de
2. de.wikipedia.org
3. en.wikipedia.org
4. www.amazon.de
5. portal.acm.org
6. www.vldb2010.org
7. citeseerx.ist.psu.edu
8. www.mpi-inf.mpg.de
9. hadoop.apache.org
10. www.mail-archive.com

Top clicks

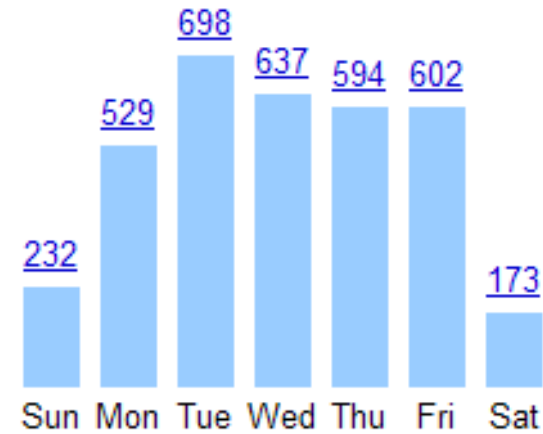
1. [DBLP: Ralf Schenkel](#)
2. [VLDB2010 Spore : Conferen](#)
3. [home \[ACM SIGIR 2010\]](#)
4. [ICDE 2010](#)
5. [ECIR 2010 | The Annual Eur](#)
6. [The 19th ACM International \(](#)
7. [The 11th International Confere](#)
8. [Text REtrieval Conference \(T](#)
9. [The 2010 ACM SIGMOD/PO](#)
10. [INEX 2010](#)

Persönliches Google Web-Protokoll (2010)

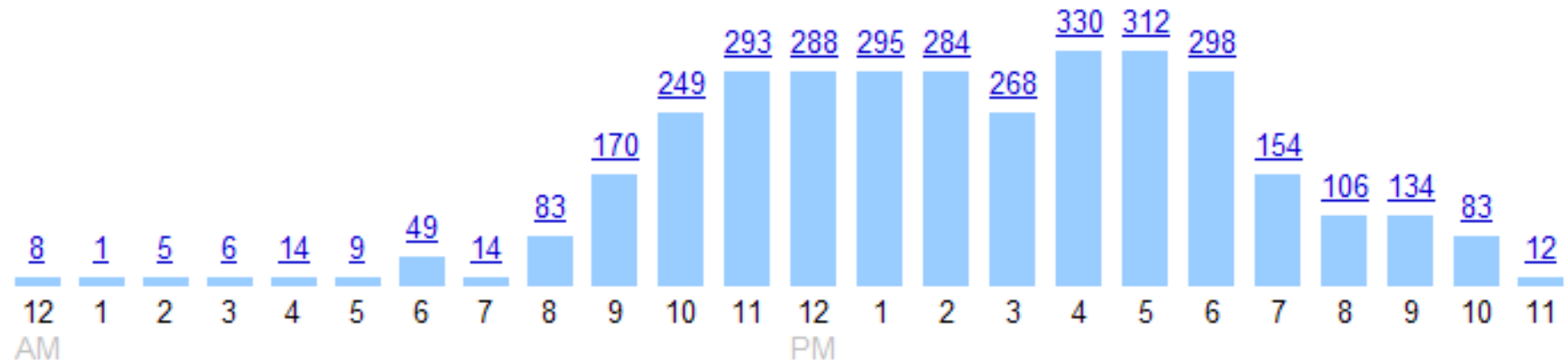
Monthly search activity



Daily search activity



Hourly search activity



Suche in persönlichen Protokollen (2010)

[Search History](#)[Search the Web](#)

Your web history is

Web History for ralf.schenkel@gmail.com

All History

[Web](#)
[Images](#)
[News](#)
[Products](#)
[Sponsored Links](#)
[Video](#)
[Maps](#)
[Blogs](#)
[Books](#)

[Pause](#)
[Remove items](#)

Trends

Bookmarks

Search results 1 - 25 of 119 for vldb - [Browse all history](#)

☆ [VLDB2010 Spore : Conference Overview](#) - [vldb2010.org](#) - Dec 23 - Viewed 25 times

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

Searched for [vldb 2010](#) - [Viewed 2 results](#)

☆ [VLDB2010 Spore : Conference Overview](#) - [www.vldb2010.org/](#) - Jun 30 - Viewed 25 times

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

☆ [VLDB2010 Spore : Accepted Papers](#) - [vldb2010.org](#) - Jun 30

VLDB 2010 , 36th International Conference on Very Large Data Bases Singapore : 13 to 17 Sept 2010, Grand Copthorne Waterfront

Searched for [vldb 2010](#) - [Viewed 1 result](#)

☆ [VLDB2010 Spore : Conference Overview](#) - [www.vldb2010.org/](#) - Jun 17 - Viewed 25 times

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

Searched for [vldb 2010](#) - [Viewed 1 result](#)

☆ [VLDB2010 Spore : Conference Overview](#) - [www.vldb2010.org/](#) - Jun 16 - Viewed 25 times

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

Searched for [vldb 2011](#) - [Viewed 1 result](#)

☆ [VLDB 2011](#) - [vldb.org](#) - Jun 14

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

Searched for [vldb 2010](#) - [Viewed 1 result](#)

☆ [VLDB2010 Spore : Conference Overview](#) - [www.vldb2010.org/](#) - Jun 14 - Viewed 25 times

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

Searched for [vldb 2010](#) - [Viewed 1 result](#)

☆ [VLDB2010 Spore : Conference Overview](#) - [www.vldb2010.org/](#) - Jun 12 - Viewed 25 times

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers

Google heute: „Meine Aktivitäten“

Google Meine Aktivitäten

Gruppierte Ansicht

Einzelansicht

Aktivitäten löschen nach


Andere Google-Aktivitäten

Aktivitätseinstellungen

Mein Konto


Hilfe

Feedback geben

 Suche


Aufgerufen [iGoogle Portal](#)

15:12 • Details

 Suche

Gesucht nach [igoogle](#)


15:12 • Details

 Suche

Aufgerufen [DeuSu.de](#) - Anonyme, werbefreie Suchmaschine aus Deutschland

15:02 • Details



 Suche

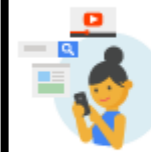
Aufgerufen [suchmaschine.com](#): Deutsche Internet Suchmaschinen-Alternative

15:02 • Details

 Suche

Gesucht nach [suchmaschine](#)

15:02 • Details



Meine Aktivitäten

Hier können Sie sehen und steuern, welche Daten bei der Verwendung von Google-Diensten erzeugt werden.

[MEINE AKTIVITÄTEN AUFRUFEN](#)

Standard RF: Rocchios Methode (1971)

- Ziel: Finde Anfrage, die kleinen Abstand zu relevanten Dokumenten und großen Abstand zu nichtrelevanten Dokumenten hat
- Berechne **Rocchio-Gewichte [1971]** für jeden Term (auch genutzt als Gewichte in der Anfrage):

$$w(t) = \alpha \cdot q(t) + \beta \frac{r_t}{R} - \gamma \frac{n_t}{N}$$

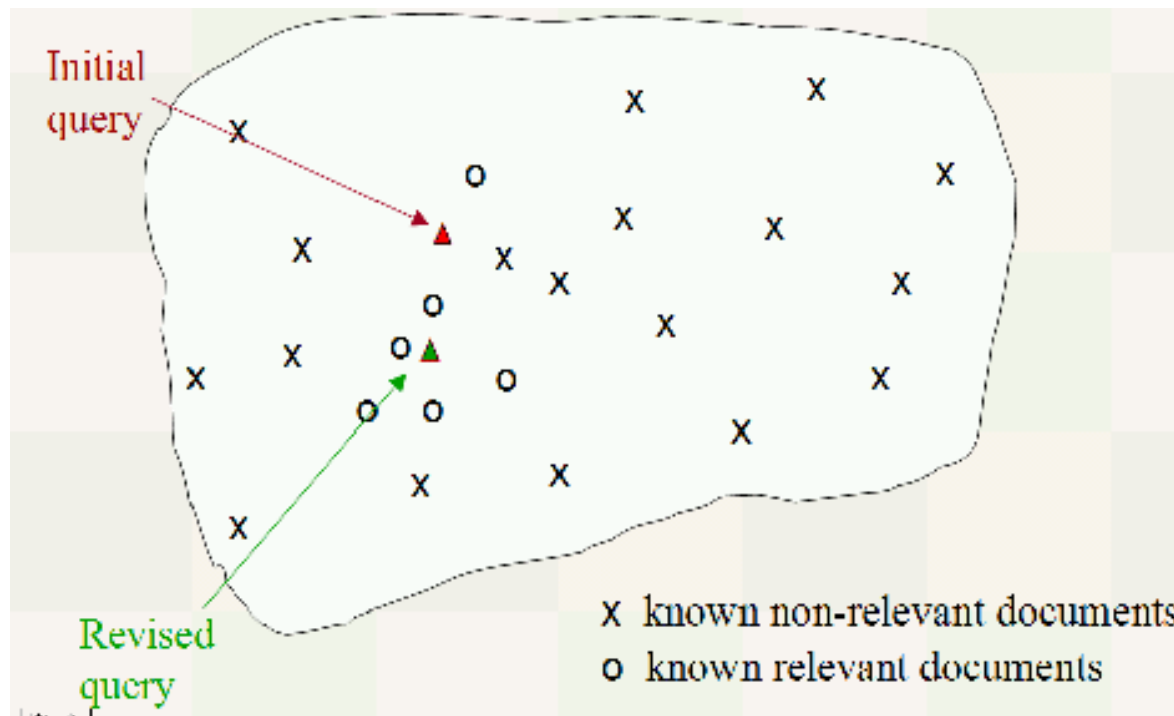
dabei sind

$q(t)$	Gewicht von Term t in der Anfrage
r_t	Anzahl relevanter Ergebnisse mit Term t
R	Anzahl relevanter Ergebnisse
n_t	Anzahl nichtrelevanter Ergebnisse mit Term t
N	Anzahl nichtrelevanter Ergebnisse

Praktisch: Wähle Terme mit positivem Gewicht oder n Terme mit höchstem Gewicht als neue (modifizierte/expandierte) Anfrage

Standard RF: Rocchios Methode (1971)

Feedback verschiebt den Anfragevektor **hin zu dem Centroid** der relevanten Dokumente und **weg vom Centroid der nichtrelevanten Dokumente**



Parameterwahl oft $\alpha > \beta > \gamma$

Einfacher Einsatz von Feedback: Promoting

- Idee: Verschiebe Ergebnisse mit **positivem Feedback** nach oben
 - **Lokal** für jeden Benutzer einzeln:
 - Speichere Feedback für jeden Benutzer (z.B. Clicks auf Ergebnisse)
 - Verschiebe Ergebnisse mit Feedback nach oben, wenn Anfrage wiederkommt (ungefähr 30% der Anfragen [Dou, WWW07])
 - Nutzt Gewohnheiten der Benutzer aus: „wetter“ -> immer wetter.de, wetter.com, ...
 - **Global** für alle Benutzer:
 - Sammle Feedback für (häufige) Anfragen
 - Verschiebe Ergebnisse mit Feedback der „meisten“ Benutzer nach oben
 - Funktioniert nicht gut für Anfragen mit unklarer Bedeutung
- ⇒ Ansatz basiert ausschließlich auf Reranking, Anfrage bleibt unverändert

Benutzerprofile

- Ziel: Konstruiere eine **Zusammenfassung der Interessen** eines Benutzers
 - Aus den bisherigen Anfragen
 - Aus den bisher zugegriffenen Seiten
 - Aus den Dokumente, den Mails, etc. (optional, z.B. bei Desktop-Suche)
- **Grundsätzlicher Ansatz:**
 - Betrachte für jede Seite p ihren Termvektor $t(p)$
 - Berechne für die Menge der angesehenen Seiten B den mittleren Termvektor $t(B)$:

$$t(B) = \frac{1}{|B|} \sum_{p \in B} t(p)$$

Persistente vs. Sitzungsprofile

- Langzeitinteressen des Benutzers können sich von seinen Interessen in der aktuellen Sitzung unterscheiden (Beispiel: Kauf eines Geschenks)
- \Rightarrow Verwalte zwei Profile: persistentes Langzeit-Profil und Sitzungsprofil
- **Sitzungsprofil:**
 - Betrachte nur die Seiten, auf die in der aktuellen Sitzung zugegriffen wurde
 - Sitzungsgrenze durch Zeit (z.B. letzte 10 Minuten) oder inhaltliche Kohärenz (Suchen/Seiten aus dem gleichen Themengebiet)
- **Persistentes Langzeit-Profil:**
 - Betrachte alle Seiten, auf die der Benutzer jemals zugegriffen hat
 - Geringeres Gewicht für alte Seiten (typisch: exponentieller Abfall)
- Profil ist Mischung von Sitzungsprofil und Langzeitprofil

Personalisierung mit Benutzerprofilen

Reranking der Suchergebnisse basierend auf Übereinstimmung mit dem Profil:

- Berechne vollständige Ergebnismenge R für die Anfrage
- Berechne für jedes Ergebnis p seine Ähnlichkeit mit dem Profilvektor (z.B. Cosinus-Ähnlichkeit)
- Sortiere Ergebnisse nach absteigender Ähnlichkeit

Problem beim Reranking: Ähnliche Ergebnisse

- Reranking kann nicht funktionieren, wenn alle Ergebnisse ähnlich sind (und nicht relevant für die Anfrage)
- Beispiel:
 - Anfrage: windows (wie die Dinger in Häusern)
 - Ergebnisse: nur über das Betriebssystem

Ziel: Diversifizierung der Anfrageergebnisse

[Microsoft Corporation](#)

Microsoft Virtualization: Download a free trial of **Windows** Server 2008 Hyper trial, tools, and downloads that work with **Windows** Essential ...

[www.microsoft.com/](#) - [Cached](#) - [Similar](#)

[Microsoft Download Center](#)

[Windows XP](#)

[Microsoft Help and Support](#)

[Windows Update](#)

[Windows Vista](#)

[Windows 7](#)

[Templates](#)

[Internet Explorer](#)

[More results from microsoft.com »](#)

[Microsoft Windows: software and services as unique as you are](#)

With **Windows** on your mobile phone, PC, or the Web, you have access to a software, services, and devices to choose from.

[Downloads](#) - [Windows XP](#) - [Windows 7](#) - [Windows Vista](#)

[www.microsoft.com/windows/](#) - [Cached](#) - [Similar](#)

[Microsoft Windows Update](#)

Latest bug fixes for Microsoft **Windows**, including fixes for some possible Do **windowsupdate.microsoft.com/** - [Similar](#)

The screenshot shows a Bing search results page for the query "windows". The search bar at the top contains the text "windows" and the Bing logo. Below the search bar, there are tabs for "Web" and "Wikipedia". The "Web" tab is selected. The results are divided into two main sections: "RELATED SEARCHES" and "ALL RESULTS".

RELATED SEARCHES

- House Windows
- Home Windows
- Replacement Windows
- Residential Windows
- Windows Update
- Windows Live
- Windows Live Messenger
- Windows Media Player
- 11

SEARCH HISTORY

- windows

See all
Clear all · Turn off

ALL RESULTS 1-20 of 460,000,000 results ·

Official Windows Home Page
Microsoft **Windows** family of operating systems home page with product information, edition comparison charts, feature descriptions, downloads and more.
[www.microsoft.com/windows](#) - [Cached page](#)

Microsoft Corporation
Is your PC ready for **Windows** 7? Download the free Upgrade Advisor to find out if you can run **Windows** 7 on your PC. Learn more about **Windows** 7
[www.microsoft.com](#) - [Cached page](#)

Microsoft Windows - Wikipedia, the free encyclopedia
[Versions](#) · [History](#) · [Timeline of releases](#) · [Security](#)
Microsoft **Windows** is a series of software operating systems and graphical user interfaces produced by Microsoft. Microsoft first introduced an operating environment named **Windows** in ...
[en.wikipedia.org/wiki/Microsoft_windows](#) - [Cached page](#)

Microsoft Windows Update
The Microsoft **Windows** Update Consumer site provides critical updates, security fixes, software downloads, and Microsoft **Windows** Hardware Quality Lab (WHQL) device drivers for your ... 8-27
[windowsupdate.microsoft.com](#) - [Cached page](#)

Generierung diversifizierter Ergebnisse

- Neue Komponente im Scoring: **Unähnlichkeit** zu bisherigen Ergebnissen
- Unähnlichkeit: z.B. großer Winkel zwischen Vektoren, großer Unterschied der Sprachmodelle, etc.
- Ansatz zur Ergebnisberechnung für eine Anfrage:
 - Berechne die besten K Ergebnisse (K groß, z.B. 1000)
 - Initialisiere Ergebnismenge R mit besten Ergebnis
 - Wiederhole, bis gewünschte Ergebnisanzahl bestimmt ist:
 - Bestimme mittleren Vektor r der Ergebnisse in R
 - Wähle aus allen noch nicht gelieferten Ergebnissen das aus, dessen (gewichtete) Summe aus Anfragescore und Unähnlichkeit zu r maximal ist
- Führt zur Bestrafung wiederholter und sehr ähnlicher Ergebnisse, spült neue Aspekte der Anfrage nach oben

Beispiel: Maximale Marginale Relevanz (MMR)






Das nächste zurückzugebende Dokument d_i soll möglichst relevant zur Anfrage q sein, aber auch verschieden zu den bislang zurückgegebenen Dokumenten d_1, \dots, d_{i-1}

$$d_i := \operatorname{argmax}_{d \in D} (\lambda \operatorname{sim}(q, d) - (1 - \lambda) \max_{d_j, 1 \leq j < i} \operatorname{sim}(d_i, d_j))$$

mit Dokumentmenge D , Tuning-Parameter λ und Ähnlichkeitsmaß sim .

Beispiel:

Ergebnisliste nach Güte

	$\operatorname{sim}(q, d_1) = 0,9$
	$\operatorname{sim}(q, d_2) = 0,8$
	$\operatorname{sim}(q, d_3) = 0,7$
	$\operatorname{sim}(q, d_4) = 0,6$
	$\operatorname{sim}(q, d_5) = 0,5$

$$\xrightarrow{\operatorname{sim}(d, d') = \begin{cases} 1 & \text{falls gleiche Farbe} \\ 0 & \text{sonst} \end{cases}}$$
$$\lambda = 0,5$$

Finale Ergebnisse

	$\operatorname{mmr}(q, d_1) = 0,45$
	$\operatorname{mmr}(q, d_3) = 0,35$
	$\operatorname{mmr}(q, d_5) = 0,25$
	$\operatorname{mmr}(q, d_2) = -0,10$
	$\operatorname{mmr}(q, d_4) = -0,20$

Beispiel: diversifizierte Ergebnisse

Windows | Offizielle Website für Windows 10 und Windows-Geräte

<https://www.microsoft.com/de-de/windows> ▼

Windows | Official Site For Microsoft Windows 10 OS, laptops, PCs ...

<https://www.microsoft.com/en-us/windows/> ▼ Diese Seite übersetzen

Windows 10 Anniversary Update Features | See What's New | Microsoft

<https://www.microsoft.com/en-us/windows/features> ▼ Diese Seite übersetzen

Windows Central | News, Forums, Reviews, Help for Windows Phone

www.windowscentral.com/ ▼ Diese Seite übersetzen

Windows - Microsoft Store Deutschland

https://www.microsoftstore.com/store/msde/de_DE/.../Windows/categoryID.70039300 ▼

Microsoft Windows – Wikipedia

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

Standard-Ranking:

Sehr ähnliche Dokumente

Beispiel: diversifizierte Ergebnisse

Windows | Offizielle Website für Windows 10 und Windows-Geräte

<https://www.microsoft.com/de-de/windows> ▼

Diversifiziertes Ranking

Windows | Official Site For Microsoft Windows 10 OS, laptops, PCs ...

<https://www.microsoft.com/en-us/windows/> ▼ Diese Seite übersetzen

Windows 10 Anniversary Update Features | See What's New | Microsoft

<https://www.microsoft.com/en-us/windows/features> ▼ Diese Seite übersetzen

MI Windows and Doors | Always Built with Pre

<https://miwindows.com/> ▼ Diese Seite übersetzen

Download iCloud for Windows - Apple Support

<https://support.apple.com/en-us/HT204283> ▼ Diese Seite übersetzen

Andersen Windows & Doors

<https://www.andersenwindows.com/>

GitHub Desktop - Simple collaboration from your desktop

<https://github.com/> ▼ Diese Seite übersetzen

Windows Central | News, Forums, Reviews, Help for Windows Phone

www.windowscentral.com/ ▼ Diese Seite übersetzen

Download Slack for Windows | Slack

<https://slack.com/downloads/windows> ▼ Diese Seite übersetzen

Windows - Microsoft Store Deutschland

https://www.microsoftstore.com/store/msde/de_DE/.../Windows/category/D.70039300 ▼

Microsoft Windows – Wikipedia

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

Get started with Docker for Windows

<https://docs.docker.com/docker-for-windows/>

Weiterer Diversifizierungsansatz: Querylogs

- Nutze Information aus Anfragelogs über Sequenzen von Anfragen
- Grund für Folgeanfrage könnte sein, dass Ergebnisse der Ursprungsanfrage nicht nützlich waren
- Beispiel:
windows → house windows → vinyl windows
→ windows xp → windows vista
- Ansatz:
Um K Ergebnisse für Anfrage q zu erhalten, mische top-K/(m+1) Ergebnisse für q mit top-K/(m+1) Ergebnissen der m Anfragen, die im Log am häufigsten auf q folgen bzw. die am verschiedensten zu q sind

Searches related to: **windows**

[windows azure](#)

[house windows](#)

[windows live](#)

[windows vista](#)

[windows xp](#)

[windows 7](#)

[home windows](#)

[windows media player](#)

Personalisierung

Empfehlungen

Um was geht es bei Empfehlungen?

Input: Gesammelte Daten über das *Verhalten von Benutzern*

- Items (Bücher, DVDs, Webseiten, Publikationen,...) angesehen
 - Items (Bücher, DVDs, CDs, Publikationen,...) gekauft
 - Items (Bücher, Filme, Hotels,...) bewertet
 - Webseiten in Bookmarks aufgenommen
 - Suchen und geklickte Suchergebnisse
 - Sequenz von Aktivitäten (Surfen, Suchen, ...)
 - Mails, Dokumente etc. gelesen und geschrieben
 - Profile in Sozialen Netzen (auch Kontakte)
- ⇒ Aufbau ausführlicher *Benutzermodelle*

Um was geht es bei Empfehlungen?

Output: Items, für die sich der Benutzer **möglicherweise interessieren** könnte

- Items (Bücher, Hotels, Filme,...) zum Kaufen/Ansehen/Besuchen
- Webseiten zum Besuchen
- Verbesserte Suchergebnisse
- Mögliche Anfrageerweiterungen und -verfeinerungen
- Leute, die man in Sozialen Netzen kontaktieren könnte

Ähnliche Artikel wie die, die Sie sich angesehen haben [Mehr](#)



Ralf, machen Sie Ihre Reise nach [München](#) zu einem Abenteuer!
Andere Gäste haben ihre Reise verlängert und [Salzburg](#), [Wien](#), [Berlin](#) und [Prag](#) besucht

MEISTGELESENE ARTIKEL

- 1 Adolf Hitler**
"Ruhig abwarten!"
👁 342.423 Aufrufe
- 2 USA und China**
Kriegsgefahr im Pazifik
👁 244.881 Aufrufe
- 3 USA**
Trump entlässt kommissarische Justizministerin
👁 112.794 Aufrufe

Wird oft zusammen gekauft



Gesamtpreis: **EUR 37,50**

Beides in den Einkaufswagen

🔍 Einer der beiden Artikel ist schneller versandfertig. [Details anzeigen](#)

☒ **Dieser Artikel:** 6 Stück Franziskaner Weissbier Gläser 0,5l **EUR 16,50**

☒ 6 Stück Franziskaner Weissbier Gläser 0,3l **EUR 21,00**

Kunden, die diesen Artikel gekauft haben, kauften auch



Franziskaner Perfect Draft
Weizenbier Mehrweg (1 x
6,0 l)
★★★★★ 9
EUR 23,32 (EUR 3,89 / l)



Kesper 68215 Brez'n
Ständer, 6-armig, Höhe 35
cm
★★★★★ 29
EUR 8,99 [Prime](#)



Philips HD3620/25 Perfect
Draft Bierzapfanlage
(Metallfasser 6 Liter)
schwarz
★★★★★ 297
EUR 181,69 [Prime](#)



infactory 24 festliche
Strohsterne für den
Weihnachtsbaum
★★★★★ 63
EUR 7,45 [Prime](#)

Drei orthogonale Ansätze

Collaboratives Filtering („nächste Nachbarn“):

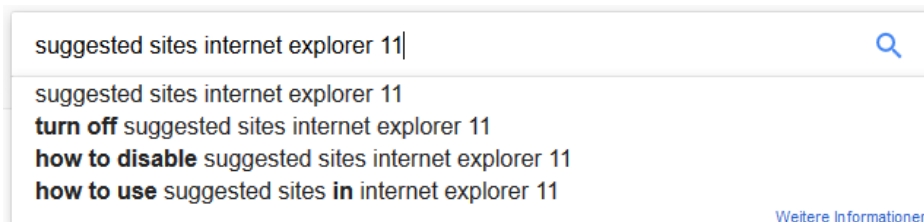
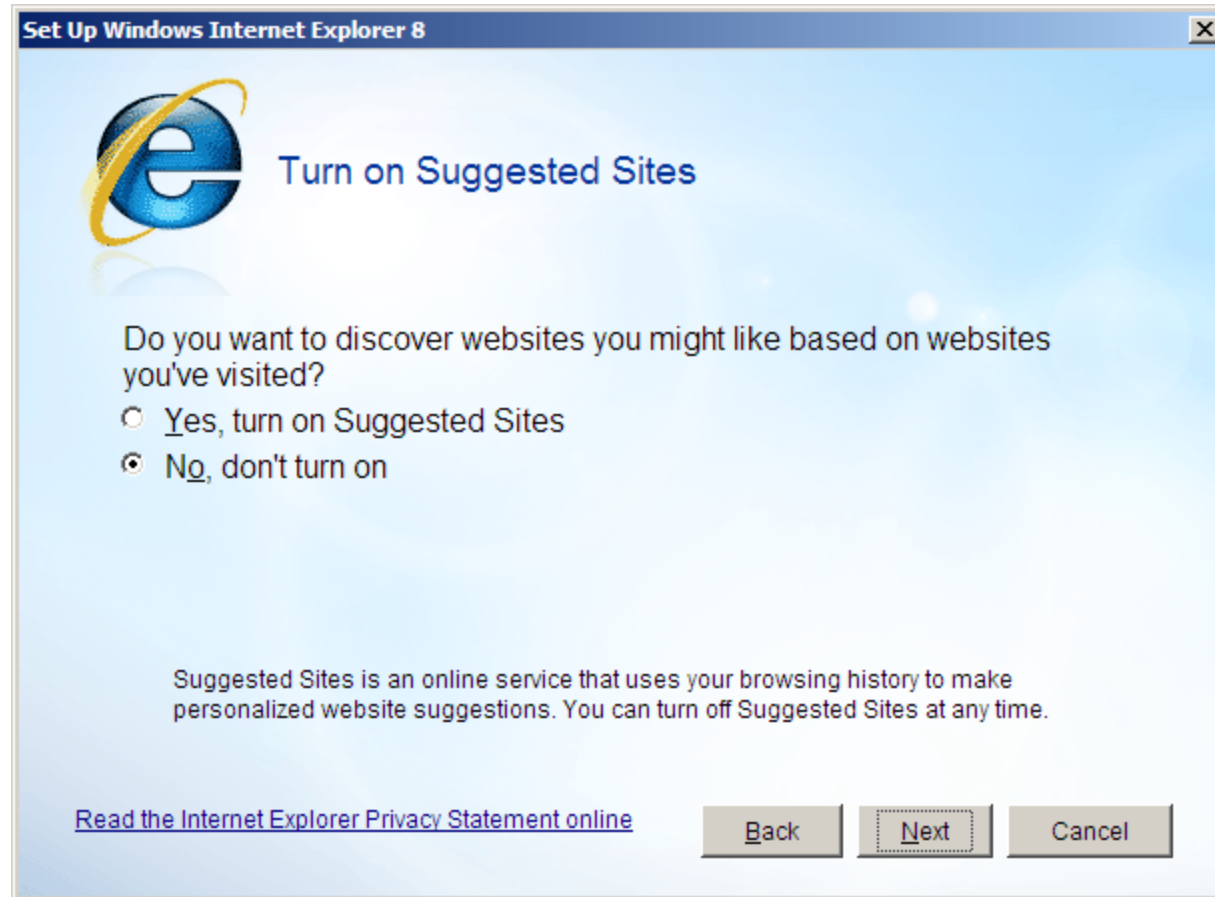
Benutzer A mag/kauft/... Item X	}	Benutzer B könnte Item X mögen
(Modell von) Benutzer B ähnlich		
zu (Modell von) Benutzer A		

Content-based Filtering (Item-Fokus)

Benutzer A mag/kauft/.. Item X	}	Benutzer A könnte Item Y mögen
Item X ähnlich zu Item Y		

[*Statischer Ansatz*: Viele Leute kaufen X]

Beispiel 1: Vorschlägen von Webseiten




Beispiel 1: Vorschlägen von Webseiten

Vorgeschlagene Sites


Personen, die Folgendes
angesehen haben

Mochten auch

 [Kompetenzzentrum Informatik
der Universität des Saarlandes](#)

 [Aktuelle News && Events — ZBI](#)

Umzug der Zentrums für Bioinformatik. Ab 01. Dezember ist das Zentrum für Bioinformatik mit Geschäftsführung sowie Prüfungsamt in einem eigenen Gebäude zu finden: Gebäude E ...
<http://www.zbi.uni-saarland.de>

 [E-Government-Akademie des IFIB - Institut für Informationsmanagement ...](#)

Die vom Institut für Informationsmanagement Bremen (ifib) organisierte E-Government-Akademie veranstaltet Seminare u Workshops für Entscheider und Gestalter. Wir schlagen die ...
<http://www.egovernment-akademie.de/academy/content/sections/>

 [Forwiss Index](#)

Develops knowledge-based systems on the basis of mathematical models. Features current and completed projects and contact information.
<http://www.forwiss.uni-passau.de>

[Weitere Vorschläge anzeigen](#)

⇒ Sehr simples collaboratives Filtering ohne explizites Benutzermodell (“wer A clickt, clickt auch B”)

Beispiel 2: Produktempfehlungen

Most Popular in Desktop Computers



[Apple iMac MB417LL/A](#)
20...
\$1,194.00



[HP Pavilion A6700F](#)
[Desktop PC](#)
\$731.00 **\$429.79**



[Datenbanksysteme: Eine Einführung](#)

von Alfons Kemper (April 22, 2009)
Auf Lager.

Preis: EUR 39,80

[62 Angebote](#) ab EUR 39,80



Gehört mir



Kein Interesse



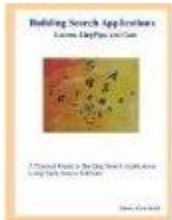
☆☆☆☆☆

Diesen Artikel bewerten

Diesen Artikel haben wir empfohlen, weil Sie **Übungsbuch Datenbanksysteme** gekauft haben.

Mehr zu entdecken

Sie haben sich angesehen:



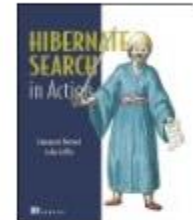
[Building Search Applications: Lucene...](#) Taschenbuch von
Manu Konchady
EUR 31,89

> [Verwandte Artikel entdecken](#)

Ihnen könnten diese Artikel gefallen:



[Lucene in Action](#) Taschenbuch von Erik Hatcher, Otis...
EUR 30,95



[Hibernate Search in Action](#) Tasche...
Bernard...
EUR 33,95

⇒ Statisch und item-basiertes Filtering (aber: Inhalt der Bücher wird nicht berücksichtigt)

Beispiel 2: Produktempfehlungen

amazon.com

More to Explore

As someone who's browsed lenses, you may be interested in checking out bestselling digital SLR lenses from Amazon.com.



[Canon EF-S 55-250mm f/4.0-5.6 IS Telephoto Zoom Lens for Canon Digital SLR Cameras](#)



[Canon EF 50mm f/1.8 II Camera Lens](#)



[Nikon 50mm f/1.8D AF Nikkor Lens for Nikon Digital SLR](#)



[Nikon 55-200mm f/4-5.6G ED IF AF-S DX VR \[Vibration](#)

Canon EF 75-300mm f/4-5.6 III Telephoto Zoom Lens for Canon SLR Cameras

Other products by [Canon](#) | [See collection](#)

4.1 out of 5 stars [See all reviews](#) ([140 customer reviews](#)) | [More about this product](#)

List Price: ~~\$199.99~~

Price: **\$155.00** & this item ships for **FREE** with **Super Saver Shipping**. [Details](#)

You Save: **\$44.99** (22%)

[Special Offers Available](#)

In Stock.

We are not able to ship this item to your default shipping address.

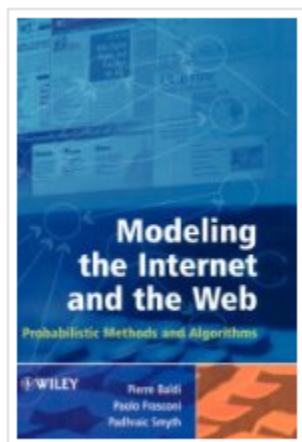
Ships from and sold by **Amazon.com**. Gift-wrap available.

[18 new](#)

[19 used](#) from **\$125.00**

[8 refurbished](#) from **\$129.90**

Beispiel 3: Buchempfehlungen



Modeling the Internet and the Web: Probabilistic Methods and Algorithms

by [Pierre Baldi](#)

Members	Reviews	Popularity	Average
10	None	439,588	

▼ Book information

Modeling the Internet and the Web: Probabilistic Methods and Algorithms

by [Pierre Baldi](#)

Wiley (2003), Hardcover, 285 pages

LibraryThing recommendations

1. [Web Metrics: Proven Methods for Measuring Web Site Success](#) by Jim Sterne
2. [Differentiated services for the Internet](#) by Kalevi Kilkki
3. [Internet Measurement: Infrastructure, Traffic and Applications](#) by Mark Crovella
4. [Designing Campus Networks](#) by Terri Quinn-Andry
5. [True Names: And the Opening of the Cyberspace Frontier](#) by Vernor Vinge
6. [Me++: The Cyborg Self and the Networked City](#) by William S. Burroughs
7. [What Just Happened: A Chronicle from the Information Age](#) by Thomas H. Davenport
8. [The Digital Sublime: Myth, Power, and Cyberspace](#) by Mark Dery
9. [24 Hours in Cyberspace: Painting on the Walls of the Internet](#) by 150 of the World's Leading Cyberspace Artists
10. [Crypto Anarchy, Cyberstates, and Pirate Utopias](#) by Peter W. Lewis

LibraryThing Recommendations

304 recommendations — page [1] | 2 | 3 | 4

1. [Machine Learning](#) by [Thomas Mitchell](#)

169 copies, 1 reviews, Average rating 4.08.

No thanks! | [Why?](#) (close why) Recommendation based on:

[Artificial Intelligence: A Modern Approach](#) by [Stuart J. Russell](#)

[Data Mining: Practical Machine Learning Tools and Techniques, Second Edition](#) (Morgan Kaufmann Series in Data Management) by [Ian H. Witten](#)

[An Introduction to Support Vector Machines and Other Kernel-based Learning Methods](#) by [Nello Cristianini](#)

[All of Statistics: A Concise Course in Statistical Inference](#) (Springer Texts in Statistics) by [Larry Wasserman](#)

Collaboratives Filtering

Wir betrachten n Benutzer U , m Items I .

Wir modellieren die *Benutzer-Item-Relation* als $n \times m$ – Matrix V :

- $V = \{0,1\}^{n \times m}$: Binäre *Einkaufsmatrix*
- $V = [\min, \max]^{n \times m}$: quantifizierte *Präferenzmatrix*

Beide sind sehr dünn besetzt, die meisten Einträge sind 0!

(Librarything@2010: 1,000,000 Benutzer, 52 Millionen Bücher, weniger als 200 Bücher für die meisten Benutzer $\Rightarrow 0,0004\%$ Einträge, die nicht 0 sind)

„Bedeutung“: v_{ik} als „Stimme“ von Benutzer i für Item k

Das Empfehlungsproblem

Inputs:

- Menge von Stimmen von Benutzer u für Items I_u
- Menge von Stimmen anderer Benutzer

Ziel: Sage die Stimmen von u für Items voraus in $I \setminus I_u$
(um die Items mit den höchsten Stimmen bestimmen zu können)

⇒ Das führt zu einem Skalierbarkeitsproblem, da $|I|$ groß ist!

Vorhersagen von Stimmen

Initiale Kalibrierung, um Vorlieben (Bias) zu entfernen:

$$v_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{ij} \qquad v_{ij}^* = v_{ij} - v_i$$

Sage die Stimme von Benutzer u für Item j voraus als gewichteten Durchschnitt über die Stimmen aller anderen Benutzer:

$$\hat{v}_{uj} = v_u + \frac{1}{C} \sum_{i=1}^n w_{ui} \cdot v_{ij}^* \qquad C = \sum_{i=1}^n |w_{ui}|$$

Ähnlichkeit der Benutzer u und i

Schätzen der Ähnlichkeit von Benutzern

- Ähnlichkeit basierend auf Korrelation

$$w_{ai} = \frac{1}{C_2} \sum_{j \in I_a \cap I_i} (v_{aj} - v_a)(v_{ij} - v_i)$$

$$C_2 = \left(\sum_{j \in I_a \cap I_i} (v_{aj} - v_a)^2 \sum_{j \in I_a \cap I_i} (v_{ij} - v_i)^2 \right)^{1/2}$$

Ergebnisse sind unbrauchbar,
wenn Überlappung der Item-
Mengen der beiden Benutzer
klein ist

- Vektorähnlichkeit (Kosinus):

$$w_{ai} = \sum_{j \in I} \frac{v_{aj}}{\sqrt{\sum_{k \in I_a} v_{ak}^2}} \frac{v_{ij}}{\sqrt{\sum_{k \in I_i} v_{ik}^2}}$$

Das Problem ist aber immer noch hochdimensional (Zahl der Benutzer und Zahl der Items ist sehr groß)

Reduziere Dimensionalität: SVD

Ersetze V durch **Approximation von** V mit Rank K mittels Singulärwertzerlegung SVD:

$$V = A \times S \times B^T$$

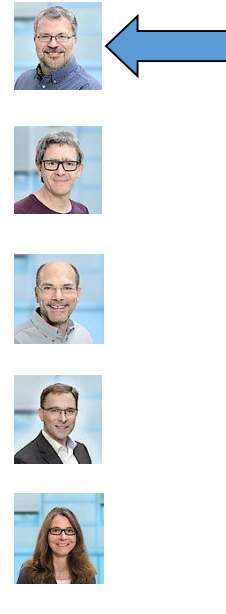
A: **Benutzer-Konzept-Ähnlichkeit**smatrix ($n \times r$)

S: Diagonalmatrix von **Singulärwerten** (mit r Einträgen $\neq 0$, wobei $r = \text{rank}(V)$), entspricht **Themen**

B^T : **Konzept-Item-Ähnlichkeit** ($r \times m$)

Zusätzlich **Beschränkung auf k größte Singulärwerte**, um die Dimensionalität weiter zu reduzieren

Beispiel: SVD (mit erfundenen Daten)



$$V = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.707 & 0 & -0.544 & 0 & 0.707 \\ 0.5 & 0 & -0.707 & 0 & -0.5 \\ 0.5 & 0 & 0.707 & 0 & -0.5 \\ 0 & 0.788 & 0 & -0.615 & 0 \\ 0 & 0.615 & 0 & 0.788 & 0 \end{pmatrix} \times \begin{pmatrix} 2.414 & 0 & 0 & 0 & 0 \\ 0 & 2.136 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.662 & 0 \\ 0 & 0 & 0 & 0 & 0.414 \end{pmatrix} \times \begin{pmatrix} 0.5 & 0.5 & 0.707 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.369 & 0.657 & 0.657 \\ -0.707 & 0.707 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.929 & 0.261 & 0.261 \\ 0.5 & 0.5 & -0.707 & 0 & 0 & 0 \end{pmatrix}$$

A

S

B^T

Beispiel: SVD

					
---	---	---	---	---	---

$$V = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$







$$\approx \begin{pmatrix} 0.707 & 0 \\ 0.5 & 0 \\ 0.5 & 0 \\ 0 & 0.788 \\ 0 & 0.615 \end{pmatrix} \times \begin{pmatrix} 2.414 & 0 \\ 0 & 2.136 \end{pmatrix} \times \begin{pmatrix} 0.5 & 0.5 & 0.707 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.369 & 0.657 & 0.657 \end{pmatrix} = \begin{pmatrix} 0.854 & 0.854 & 1.207 & 0 & 0 & 0 \\ 0.604 & 0.604 & 0.854 & 0 & 0 & 0 \\ 0.604 & 0.604 & 0.854 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.621 & 1.106 & 1.106 \\ 0 & 0 & 0 & 0.485 & 0.864 & 0.864 \end{pmatrix}$$

A
S
B^T

Empfehlungen mit SVD

- Sage Stimmen voraus auf A, nicht auf V
⇒ Bestimme Schätzung v'_{uj} für jedes Thema j
- Erweitere die Stimmenschätzung von Themen auf Items

$$v_{ui} = \sum_{j=1}^k \left(v'_{uj} \cdot S_{jj} \cdot B_{ji} \right)$$

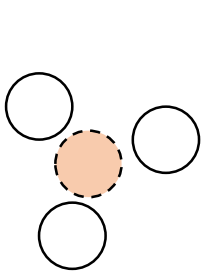
Neues Problem: Aktualisierung der SVD bei Änderung der Daten

SVD erzeugt implizite Clusterung der Items

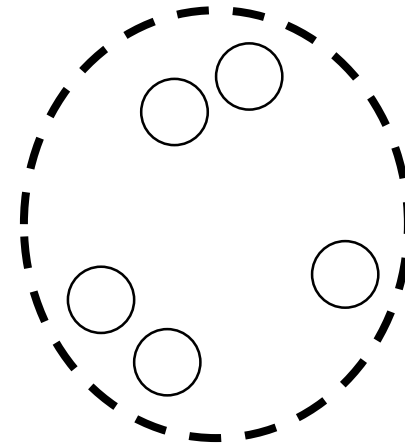
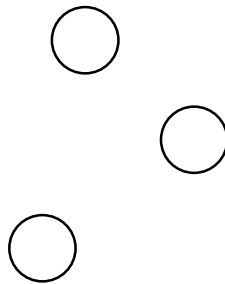
Hintergrund: Clusteranalyse

Ziel:

Gruppierung von Datenpunkten in (a priori unbekannte) Zahl von **Clustern** (Gruppen, Haufen), so dass jeder Datenpunkt ähnlicher zu anderen Punkten innerhalb seines Clusters ist als zu anderen Datenpunkten



Centroid (Zentrum)
als Mittelwert der
Datenpunkte



Anzahl der Cluster hängt von Verfahren
und seinen Parametern ab; könnte auch
nur ein Cluster sein

Reduziere Dimensionalität: Clustering

- Reduziere Anzahl der Benutzer durch Vorberechnung von ***K Clustern ähnlicher Benutzer***
- Repräsentiere jeden Cluster P durch sein ***Centroid*** $c(P)$:

$$c(P)_i = \frac{1}{|P|} \sum_{u \in P} v_{ui}$$

- Für die Vorhersage:
 - Ordne jeden Benutzer zu seinem Cluster zu
 - Berechne kollaboratives Ergebnis für Cluster anstelle von Benutzern
- **Mögliches Problem:**
Benutzer könnten zu mehreren Clustern gehören (Wissenschaftler vs. Harry-Potter-Fans)

Kollaboratives Filtern ist teuer

- Aktionen der Benutzer sind hochgradig dynamisch
 - Schwierig, Ähnlichkeiten vorauszuberechnen und zu aktualisieren
 - Beste Empfehlungen basieren oft auf dem gerade gekauften Item
- Eine Empfehlung benötigt $O(n+m)$ Zeit:
 - Muss alle Benutzer und ihre Items scannen
 - Die meisten Benutzer haben $\leq C1$ Items
 - Wenige Benutzer ($\leq C2$) haben $>C1$ Items
 - Kosten beschränkt durch $(n-C2) \cdot C1 + C2 \cdot m = O(n+m)$
 - n, m groß
- Empfehlungen müssen in Echtzeit berechnet werden ($\leq 200\text{ms}$)

Content-Based Filtering

Beobachtung:

Beziehung zwischen Items (z.B. Korrelation von Käufen) ist viel weniger dynamisch als die Beziehung zwischen Benutzern

- Information von gestern ist heute immer noch einigermaßen genau
- Tolerierbar, wenn neue Items nicht sofort empfohlen werden

Sage Stimme von Benutzer u für Item j voraus als gewichtetes Mittel über die Stimmen von Benutzer u für andere Items:

$$\hat{v}_{uj} = v_u + \frac{1}{C} \sum_{i=1}^m w_{ji} \cdot v_{ui}^*$$

\uparrow
similarity of items j and i

$$C = \sum_{i=1}^m |w_{ji}|$$

Schätzen der Item-Item-Ähnlichkeit

Auf Basis der Ähnlichkeit von inhaltlichen Beschreibungen oder der Vektoren der Stimmen (ähnlich wie Benutzerähnlichkeit)

Beispiel: Cosinusähnlichkeit der Stimmvektoren

$$w_{ji} = \sum_{u \in U} \frac{v_{uj}}{\sqrt{\sum_{k \in U} v_{kj}^2}} \frac{v_{ui}}{\sqrt{\sum_{k \in U} v_{ki}^2}}$$

Berechnung teuer ($O(m^2n)$), aber offline

Berechnung der Vorhersagen ist billig ($O(m)$ wenn nur eine konstante Zahl von Items berücksichtigt wird)

Verwende Suche für Empfehlungen

Wir nehmen an, dass wir *Features* von Items identifizieren können (Genre, Schauspieler, Regisseur, Schlüsselwörter, ...)

- Identifiziere *häufige/charakteristische Features* für die Items eines Benutzers
- Verwende *Suche* nach diesen Features und empfehle die Ergebnisse

Probleme:

- Skaliert nicht gut, wenn der Benutzer viele Items bewertet bzw. gekauft hat
- Liefert oft keine guten Empfehlungen
(Beispiel: viele Tatort-Filme angesehen, aber nicht jeder Tatort ist gleich gut)

Evaluierung von Empfehlungen

Ziel:

Bestimme, welcher von mehreren Empfehlungsalgorithmen die besten Empfehlungen generiert.

Notwendige Komponenten eines solchen *Benchmarks*:

- Menge von (Benutzer,Item,Bewertung)-Tupel zum *Trainieren* (Vorab-Eingabe für den Algorithmus)
- Menge von (Benutzer,Item,Bewertung)-Tupel zum *Testen* (Algorithmus erhält nur Benutzer und Item und muss die Bewertung vorhersagen)
 - Möglich als Offline-Experiment (mit festgelegten Bewertungen) oder Live-Experiment mit Benutzern
- Metriken für *quantitative Bewertung der Ergebnisqualität*

Eigenschaften von Testkollektionen

- Können *synthetisch* oder aus *realer Anwendung* sein
- Features der Anwendungsdomäne
 - *Neuheit* vs. *Qualität* als Fokus der Empfehlungen
 - *Wichtigkeit von* true/false positives/negatives
 - *Granularität* der echten Benutzerpräferenzen (vs. Ratings)
- Inhärente Features der Daten
 - *Implizite* oder *explizite* Ratings
 - *Skala* & *Dimensionen* der Ratings (z.B. 1-5, „sehr gut“-„schlecht“)
 - *Historie* der Ratings (Zeitstempel) und Empfehlungen
- Beispiel-Features
 - *Dichte* der Bewertungsmenge (insgesamt & für Testbenutzer)
 - *Größe* der Daten

Offline-Evaluation vs. Benutzerexperimente

- **Offline-Evaluation:** **Vergleiche** die vorhergesagte Bewertung mit der tatsächlichen Bewertung durch den Benutzer
 - Wenig Aufwand, kann automatisch gemacht werden
 - Kann verwendet werden, um eine Sequenz von Bewertungen zu evaluieren (Zeitstempel)
 - Aber: Es können nur die wenigen existierenden Bewertungen für die Evaluation verwendet werden
- **Live-Experimente mit Benutzern:** **Frage** den Benutzer nach der Meinung oder **beobachte** Verhalten
 - Verständnis, ob und warum Benutzer Empfehlungen, Interfaces, Systeme mögen oder nicht mögen

Metriken für die Evaluation

Verbreitet: Messe die **Genauigkeit** (Accuracy) der Vorhersagen, indem der **Fehler** zwischen Vorhersage und tatsächlicher Bewertung gemessen wird

- mean absolute error (MAE)

$$|\bar{E}_{MAE}| = \frac{\sum_{i=1}^N |p_i - r_i|}{N}$$

p_i : Vorhersage
 r_i : Bewertung
 N : # Messpunkte

- Root mean square error (RMSE; Betonung großer Fehler)

$$|\bar{E}_{RMSE}| = \sqrt{\frac{\sum_{i=1}^N (p_i - r_i)^2}{N}}$$

- Precision/Recall, Metriken für die Genauigkeit von Ranglisten (z.B. Kendall's Tau), ...

Weitere Dimensionen für die Evaluation

- **Coverage (Abdeckung):**

- Anteil der Items, für die Empfehlungen berechnet werden können
- Wie viele Items werden tatsächlich vorgeschlagen

- **Lernrate:**

- Wie schnell wird die Qualität der Empfehlungen besser, wenn mehr Trainingsdaten verfügbar sind

- **Novelty (Neuheit):**

- Schwerpunkt auf Items, die unbekannt für den Benutzer sind, aber innerhalb seiner Interessen liegen (z.B. ein neuer Film des Lieblingsregisseurs)

- **Serendipity („glücklicher Zufall“):**

- Überraschende Empfehlungen (z.B. der neue Film eines neuen Regisseurs, der den Geschmack des Benutzers trifft)

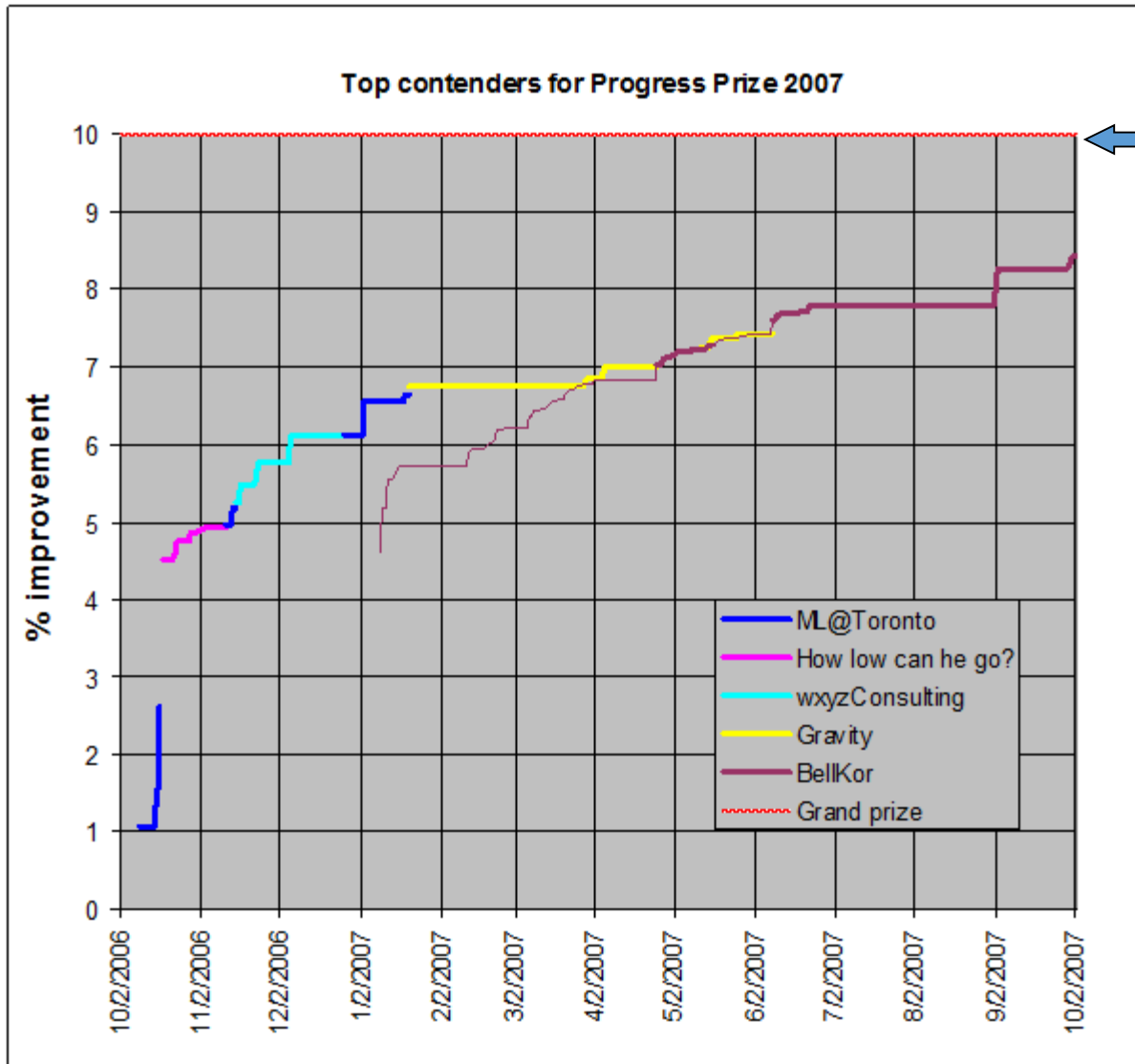
- **Confidence (Vertrauen)**

- des Algorithmus in seine Empfehlungen

Benchmarks: Netflix Prize

- <http://www.netflixprize.com>
- Durch ein Online-Filmportal ausgeschrieben
- stellt (anonymisierte) Trainingsdaten bereit (480.000 Benutzer, 18.000 Filme, 10^6 Bewertungen auf einer Skala von 1 bis 5)
- Ziel: Verbesserung des Empfehlungsalgorithmus des Portals (RMSE: 0.9514)
- Hohe Belohnung, um den Benchmark attraktiv zu machen: **1.000.000\$** für die ersten **10% improvement** in RMSE auf den Testdaten (1,4 Millionen Benutzer-Film-Paare), bis dahin **50.000\$** für den besten Fortschritt jedes Jahr

Netflix: Verbesserungen über die Zeit



10% Verbesserung erreicht
am 26. Juli 2009

Probleme der Personalisierung

- Ansatz fokussiert auf **Maximierung der kurzfristigen Benutzerzufriedenheit**
 - Mehr Suchen, mehr Clicks, mehr Einkäufe, mehr Werbung, mehr Profit
- Teil der möglichen Ergebnisse wird ausgeblendet, weil das System sie für **nicht relevant für diesen Benutzer** hält
 - Eingeschränkte Sicht auf die Welt
- Problematik der „**Filter-Bubble**“
 - Man sieht nur Dinge, die seinen eigenen Präferenzen entsprechen
 - Im Prinzip ok für Restaurants, Musik, Filme
 - Schlecht für Nachrichten, Postings, etc.

Referenzen

- P. Baldi et al: Modeling the Internet and the Web, chapter 8
- J.B. Schafer et al.: E-Commerce recommendation applications, Journal of Data Mining and Knowledge Discovery 5,115-153, 2001
- B.M. Sarwar et al.: Analysis of recommender algorithms for e-commerce, 2nd ACM Conf. On Electronic Commerce, 2000
- G. Linden et al.: Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing 7(1), 2003
- R.M. Bell and Y. Koren: Scalable Collaborative Filtering with Jointly Derived neighborhood Interpolation Weights, ICDM Conference, 2007
- J. Herlocker et al.: Evaluating Collaborative Filtering Recommender Systems, ACM Transactions on Information Systems 22, 2004
- K. Sugiyama et al.: Adaptive web search based on user profile constructed without any effort from users, WWW Conference, 2004
- P. Brusilovsky et al. (eds.): The Adaptive Web, Lecture Notes in Computer Science 4321, 2007
- Z. Dou et al.: A Large-scale Evaluation and Analysis of Personalized Search Strategies, WWW Conference, 2007.
- F. Radinski et al.: Improving Personalized Web Search using Result Diversification, SIGIR Conference, 2006.