Digital Libraries and Foundations of Information Retrieval

Prof. Dr.-Ing. Ralf Schenkel

Winter Semester 2018/2019



Disclaimer

Diese Vorlesung kann Material enthalten von

Ricardo Baeza-Yates et al.

Klaus Berberich

Burkhard Freitag

Norbert Fuhr

Andreas Henrich

Michael Ley

Christopher Manning et al.

Sebastian Michel

Benno Stein

Bernd Walter

Gerhard Weikum

und natürlich von Ralf Schenkel ©



Gegenüberstellung Module alt-neu

Information Retrieval alt

- Dokument-Vorverarbeitung
- Modelle
- Anfrageverarbeitung
- Indexierung
- Evaluierung
- Websuchmaschinen

Digital Libraries alt

- Digital Libraries
- Publikationen, Verlage, ...
- Dateiformate
- Kompressionsverfahren
- Suffixbäume, Editdistanzen
- Multimedia: JPEG, MPEG

Sommer

DL & Grundlagen IR

- Digital Libraries
- Publikationen, Verlage, ...
- Modelle, Ranking, Evaluierung
- Websuchmaschinen, Metasuche
- Empfehlungen, Soziale Netze
- Suchstrategien, Textalgorithmen



Implementierung von IR-Systemen

- Dokument-Vorverarbeitung
- Anfrageverarbeitung
- Indexierung
- Kompressionsverfahren
- Suffixbäume, Editdistanzen
- Multimedia: JPEG, MPEG

Winter

Inhalte

Digitale Bibliotheken als klassische Anwendung eines Information-Retrieval-Systems

- Publikationen und Verlage in der Informatik
- Aufgaben einer Digitalen Bibliothek, typische digitale Bibliotheken
- Suchen in einer digitalen Bibliothek
 - Modelle des Information Retrieval, Bewertung von IR-Systemen
 - Suchstrategien
 - Suche über mehrere Bibliotheken hinweg
 - Suche nach Multimedia-Inhalten
- Digitaler Bibliothekar
 - Ranking von Publikationen (und von Webseiten)
 - Empfehlungen/Recommendations
 - Soziale Netze: Verschlagwortung, Suche
- Beyond Content
 - Algorithmen auf Texten: Zusammenfassung, Sentiment Analysis, Informationsextraktion

Hörer

- •Informatik 4
- Wirtschaftsinformatik11
- Digital Humanities2
- Computerlinguistik1
- Angewandte Mathematik
- Angewandte Geoinformatik 0
- ...?

Programmieren?

Praktische Inhalte?



Vorlesung

- Montag, 12:15-13:45 in H11, Beginn 22. Oktober
- Ausnahmen:
 - Keine Vorlesung 24. und 31. Dezember
 - Ersatz: Vorlesung Montag, 29. Oktober, 8:30 in H11

Übung

- Jede Woche Montag 9:00-9:45, H11, vsl. Beginn 12. November
- Übungsausgabe in Woche X: Montag nach der Vorlesung (außer heute)
- Übungsabgabe in Woche X+1: Montag vor der Vorlesung
- Übungsbesprechung in Woche X+2
- ~11 Übungsblätter, ~33 Aufgaben
- Teilweise Rechnen, teilweise Recherche, teilweise Denken
- keine offensichtlichen Plagiate
- Gruppen bis Größe 2 ok, wenn sie über das Semester konstant bleiben; es genügt eine Abgabe pro Gruppe

Prüfung

Mündliche Prüfungen

- Am Semesterende, Termin vsl. 11. Februar 2019
- Am Anfang des Sommersemesters, Termin vsl. 1. April 2019

Zulassung zur Prüfung

- 50% der Übungspunkte
- Mindestens einmal Vortragen einer Lösung einer Aufgabe in der Übung

Literatur zu digitalen Bibliotheken



Ian. H. Witten, David Bainbridge, David M. Nichols: How to Build a Digital Library, 2nd Edition. Morgan Kaufmann, Publishers, 2009, ISBN 978-0123748577



Albert Endres, Dieter W. Fellner: Digitale Bibliotheken – Informatik-Lösungen für globale Wissensmärkte. Dpunkt.verlag, 2000 (vergriffen)



William Y. Arms: Digital Libraries. The MIT Press, 2001, ISBN 978-0262511278



Michael Lesk: Practical Digital Libraries – Books, Bytes, and Bucks. Morgan Kaufmann Publishers, 1997

Literatur über Information Retrieval

- C. D. Manning, P. Raghavan H. Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008. ISBN 978-0-521-86571-5. http://informationretrieval.org/
- B. Croft, D. Metzler, T. Strohman, **Search Engines:Information Retrieval in Practice**, *Pearson Education*, 2009. http://ciir.cs.umass.edu/downloads/SEIRiP.pdf
- S. Büttcher, C. Clarke, G. Cormack, Information Retrieval: Implementing and Evaluating Search Engines, MIT Press, 2010. http://www.ir.uwaterloo.ca/book/
- S. Chakrabarti, **Mining the Web**, *Morgan Kaufmann*, 2003. https://www.cse.iitb.ac.in/~soumen/mining-the-web/
- I. A. Witten, A. Moffat, T. C. Bell, **Managing Gigabytes**, *Morgan Kaufmann*, 2nd ed., 1999. http://people.eng.unimelb.edu.au/ammoffat/mg/
- R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval: The Concepts and Technology behind Search, 2nd ed., Addison Wesley, 2011. http://www.mir2ed.org/
- D.A. Grossman, O. Frieder, Information Retrieval: Algorithms and Heuristics, Springer, 2nd ed., 2004.
- G. Kowalski, Information Retrieval Architecture and Algorithms, Springer, 2011.
- D. Lewandowski, **Suchmaschinen verstehen**, *Springer Vieweg*, 2015.

Weitere Literatur - online

 A. Henrich, Information Retrieval 1: Grundlagen, Modelle und Anwendungen, Universität Bamberg, 2008.

http://www.uni-bamberg.de/minf/ir1 buch/

- R. Belew, Finding Out About, Cambridge University Press, 2001.
 http://www-cse.ucsd.edu/~rik/foa/l2h/
- E. Greengrass, **Information Retrieval: A Survey**, University of Maryland, Baltimore, 2000.

http://www.csee.umbc.edu/cadip/readings/IR.report.120600.book.pdf

- C. J. van Rijsbergen, **Information Retrieval**, *Butterworths*, 2/1979. http://www.dcs.gla.ac.uk/Keith/Preface.html
- R. Ferber, Data Mining und Information Retrieval, FernUniversität Hagen, 2000.

http://information-retrieal.de/irb/ir.html

Einführung

Was ist eine Bibliothek?

Die Bibliothek ist eine Einrichtung, die unter archivarischen, ökonomischen und synoptischen Gesichtspunkten publizierte Information für die Benutzer sammelt, ordnet und verfügbar macht.

Gisela Ewert, Walther Umstätter: *Lehrbuch der Bibliotheksverwaltung*, Hiersemann, Stuttgart 1997, <u>ISBN 978-3-7772-9730-9</u>, S. 10.

Beispiel: Universitätsbibliothek Trier

2.388.061 Medieneinheiten (Mai 2018?), davon

- 1.349.772 Bände Monographien, 349.869 Bände Zeitschriften,
- 162.429 Mikroformen, Tonträger, Karten u.ä. sowie
- 525.991 monografische E-Dokumente (einschließlich 481.863 Nationallizenzen).
- 803 Papyri bzw. Papyrusfragmente und eine Sammlung japanischer und chinesischer Holzschnitte

Jährlicher Zugang (2017):

11.408 Bände Monographien, 3.327 Bände Zeitschriften, 5 Mikroformen und 2.581 elektronische Medien.

Stellkapazität: 61.026 Regalmeter für 1.320.000 Bände, davon in Lesesälen 720.000, im Magazin 600.000.

Benutzer (2017): 12.842 eingeschriebene (10.209 aktive) Benutzer, 220.073 Ausleihen (ohne Verlängerungen), 19.333 Vormerkungen – Zahlen sinken (~10% von 2016 auf 2017)

Laufender Etat (2017): 1.523.885 Euro für Erwerbung und Einbandkosten, 113.599 Euro für Sachmittel – **deutlich gesunken von 2016 (~20%)**; 92 Mitarbeiterstellen

Organisationsprinzipien in Bibliotheken (1)

 Sortierung der Bücher nach Signatur (eindeutiger Schlüssel, der Fachgebiet codiert)

Signaturaufbau UB Trier (1)

Alle Bücher und Zeitschriften der UB Trier durch Signaturen mit folgendem Aufbau eindeutig gekennzeichnet:

Lokalzeichen = Systemstelle / Individualteil

Lokalzeichen: 2- oder 3-stellige Zahlen (codiert Standort)

Systemstelle: Buchstabencode; wenn erforderlich durch einen "Schlüssel" erweitert (Trennzeichen Punkt)

Individualteil: Ein Buchstaben ("Fachgruppenzeichen") und eine laufende Nummer ("Numerus Currens")

Bei Zeitschriften entfällt die Systemstelle, z.B.: 55=z10790

Signaturaufbau UB Trier (2)

Beispiel:

72=GX.C steht für Systemstelle "GX – Speziele Programmiersprachen" und für Schlüssel "C – Programmiersprachen C, C++, C#"

Lokalzeichen 72 bedeutet F-Gebäude, 55 (Mathematik) Gebäude E

Fachgruppenzeichen:

- f codiert Informatik
- c codiert Mathematik

Signaturen der Wirtschaftsinformatik haben den Aufbau 50=BX/t...

Lokalzeichen UB Trier

- 01 Bibliographien,
 Biographien
- 05 Wörterbücher, ...
- 09 Buch- und Bibliothekswesen
- 11 Psychologie
- 21 Medienwissensch.
- 25 Anglistik
- ...

- 50 Geographie / Geowissenschaft
- 55 Mathematik
- 60 & 61 Rechtswissenschaft
- 72 Informatik
- 201 Sprachlehrmaterial
- 710-772 Lehrbuchsammlung, z.B.
 772 Informatik

• ...

Lokalzeichen 72: Systemstellen

- B Einführungen, Allgemeines
- C Informatik und Gesellschaft
- D Mathematik für Inf.
- G Programmiersprachen
- H Algorithmen und Komplexität
- J Formale Sprachen und Spezifikation
- K Rechnerarchitektur und Rechnernetze
- L Ling. Datenverarbeitung

- M Betriebssysteme, Verteilte Systeme
- N Software Engineering
- P Informationssysteme
- Q Künstliche Intelligenz
- S Modellierung, Leistungsanalyse, Simulation
- U Sonstige Anwendungen
- W PCs, Workstations, Anleitungen
- X Konferenzen, Reihen, Sammelbände
- Z Zeitschriften



72=P Informationssysteme

- PA Allgemeines, Einführungen
- PC Dateisysteme
- PD Datenbanksysteme
- PR Information Retrieval
- PU Anwendungen
- PZ Sonstiges

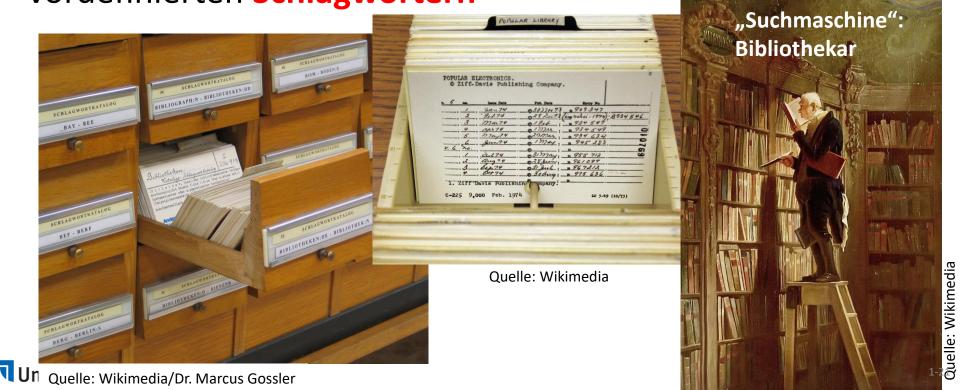
72=Q Künstliche Intelligenz

- QA Allgemeines, Einführungen
- QD Deduktionssysteme
- QE Expertensysteme, Wissensbasierte Syst.
- QL Maschinelles Lernen
- QN Neuronale Netze
- QW Wissensrepräsentation
- QZ Sonstiges

Organisationsprinzipien in Bibliotheken (2)

- Inhaltliche Erschließung der vorhandenen Dokumente
 - Formalerschließung: Autor, Titel, Verlag, etc.
 - Sacherschließung: inhaltliche Beschreibung, z.B. Schlagwörter,
 Zusammenfassung oder Klassifikation

 Einfaches Auffinden durch Suche in Index nach vordefinierten Schlagwörtern



Beispiel: Erschlossenes Dokument

Linked data management / Andreas Harth, Katja Hose, Ralf Schenkel

Harth, Andreas, [Herausgeber] |Hose, Katja, [Herausgeber] |Schenkel, Ralf, [Herausgeber] |2014 | Signatur(en): 82=GH/nh312

Verfügbar

Details Standorte Weitere Services

Titel: Linked data management / Andreas Harth, Katja Hose, Ralf Schenkel

Sonstige Beteiligte: Harth, Andreas, [Herausgeber];

Hose, Katja, [Herausgeber]; Schenkel, Ralf, [Herausgeber]

Schlagwörter: Linked Data ; Datenmanagement;

Subject Heading: Linked data;

Subject Heading: Database management;

Subject Heading: Data structures (Computer science);

Subject Heading: Semantic Web;

Subject Heading: Heterogeneous computing;

Subject Heading: COMPUTERS / Database Management / General; Subject Heading: COMPUTERS / Database Management / Data Mining;

Subject Heading: COMPUTERS / Internet / General

Beschreibung: Anm.: Auf dem Umschlag: Edited by Andreas Harth, Katja Hose, Ralf Schenkel

Verknüpfte Titel: (Emerging directions in database systems and applications)

(An A K Peters book)

Verlag: Boca Raton; London; New York: CRC Press, [2014]

Erscheinungsjahr: 2014

Format: xxvii, 547 Seiten : Illustrationen

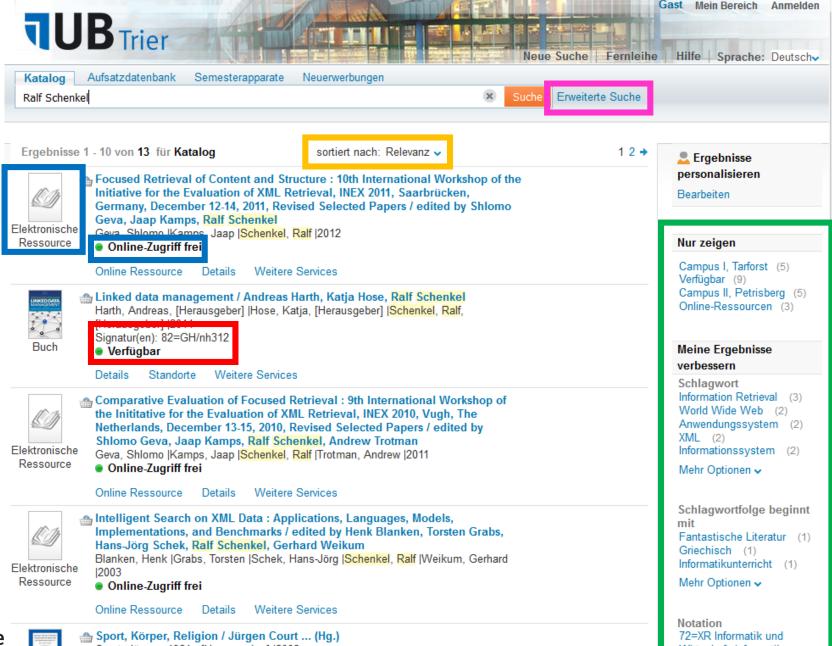
Sprache: Englisch

Identifikator: ISBN 978-1-4665-8240-8 : hardback

Notation: 82=GH Digital Humanities: Semantic Web

Werk(e): Linked data management

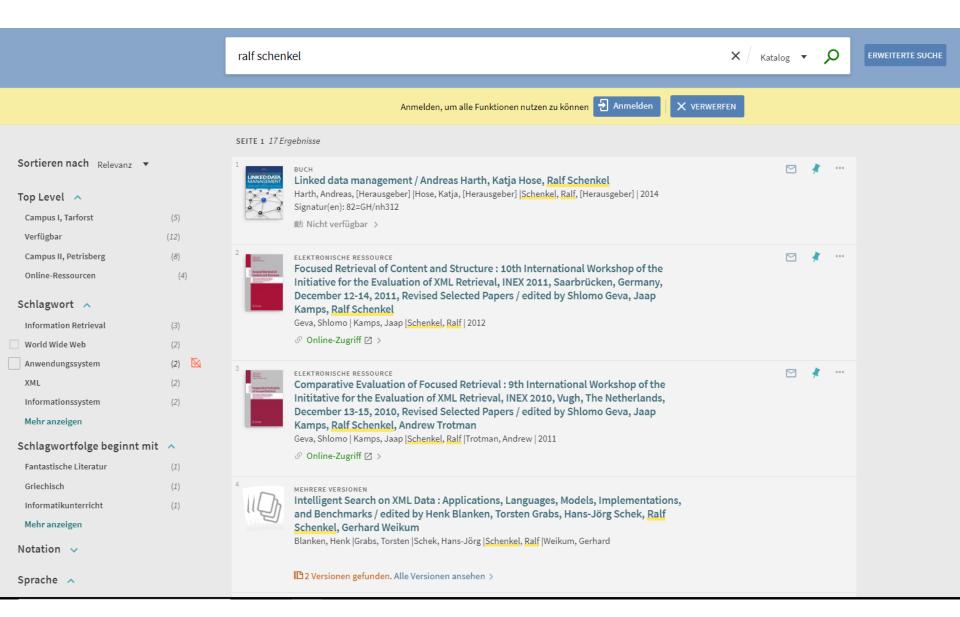
Moderne Suche UB Trier: TriCat





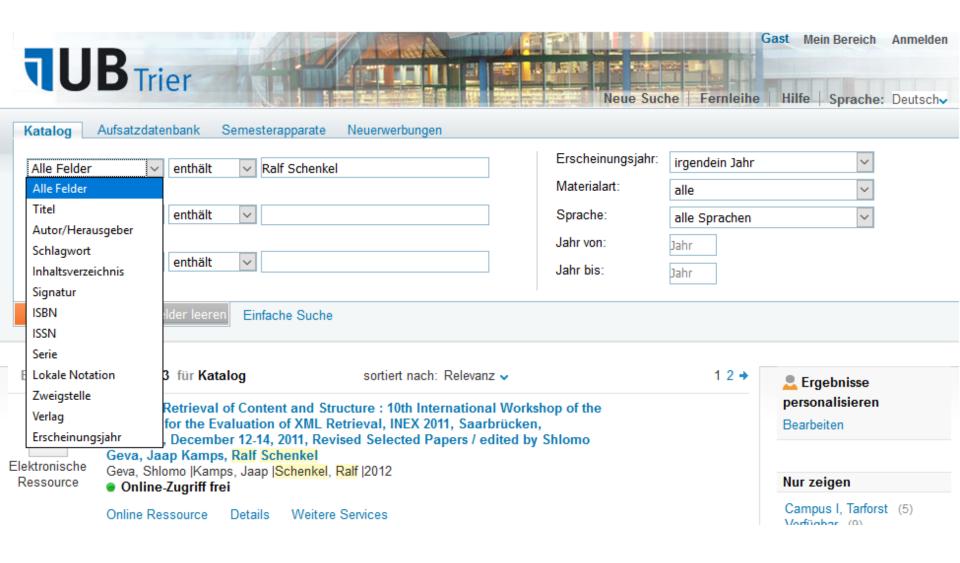


Moderne Suche UB Trier: TriCat (aktuell)





UB Trier: Erweiterte Suche



Was ist eine digitale Bibliothek?

"a collection of information that is both digitized and organized" [M. Lesk]

"a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network" [W. Y. Arms]

"a focused collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization, and maintenance" [Witten / Bainbridge / Nichols]

Mögliche Dienste einer DL

- Suche einer bestimmten Publikation
- Suche nach "passenden" Publikationen auf Basis von
 - Exakten oder inexakten Metadaten
 - Inhaltlichen Beschreibungen
- Suche nach ähnlichen Publikationen
- Suche nach "guten" Publikationen zu einem Thema
- Suche nach "passenden" Publikationen zu einem Benutzer
- Alles im lokalen Bestand oder bibliotheksübergreifend
- Erweiterte Metadaten (z.B. eingehende und ausgehende Zitate, Lesefrequenz)
- Automatische Zusammenfassung und Aufbereitung von Information aus Publikationen
 - Beschreibung einer Publikation/einer Reihe mit Schlagwörtern oder als textuelle Zusammenfassung
 - Beantwortung einer konkreten Frage
 - Zusammenstellung von Argumente für und gegen eine These

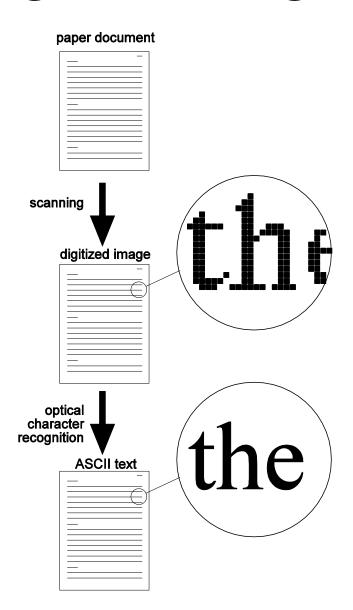
Mögliche Dienste einer DL

- Suche einer bestimmten Publikation
- Suche nach "passenden" Publikationen auf Basis von Information Retrieval
 - Exakten oder inexakten Metadaten
 - Inhaltlichen Beschreibungen
- Suche nach ähnlichen Publikationen Ähnlichkeitsmaße
- Suche nach "guten" Publikationen zu einem Thema Bewertung, Ranking
- Suche nach "passenden" Publikationen zu einem Benutzer Recommendation
- Alles im lokalen Bestand oder bibliotheksübergreifend Metasuche
- Erweiterte Metadaten (z.B. eingehende und ausgehende Zitate, Lesefrequenz)
- Automatische Zusammenfassung und Aufbereitung von Information aus Publikationen
 - Beschreibung einer Publikation/einer Reihe mit Schlagwörtern oder als textuelle Zusammenfassung – Summarization
 - Beantwortung einer konkreten Frage Siri, Cortana, Alexa
 - Zusammenstellung von Argumente für und gegen eine These

Mögliche weitere Aufgaben einer DL

- Erschließung von Dokumentbeständen
- Digitalisierung bestehender Dokumentbestände (aber auch von anderen Artefakten, insbesondere in den Geisteswissenschaften)
- Langzeitarchivierung von Dokumentbeständen

Digitalisierung von Papierdokumenten



Zwei Schritte:

- Scannen
- Schrifterkennung (OCR)
- Beispiel: Adobe Acrobat
- Probleme ...

Scanqualität...

INTRODUCTION

The Resource Description Framework (RDF)¹ was introduced in the last decade and has become a standard for exchanging data in the Web. At present, more and more data are converted to RDF. The SPARQL Protocol and RDF Query Language (SPARQL)² was officially introduced in 2008 to retrieve RDF data as easily as SQL does for relational databases. As Web data grows, the number of SPARQL Endpoints constructing SPARQL queries using HTTP also grows fast. This situation has attracted people to aggregate data from multiple SPARQL Endpoints akin to conventional distributed databases. For instance, if we search data about a drug,

s. For instance, if we search data about a drug,

wot the oft and one is a real filest of the search data.

1. INTRODUCTION

The Hesourcc Description Framework (RDF) was introduced in the last decade and has become a Standard for exchanging data in the Web. At present, more and more data are converted to RDF. The SPARQL Protocol and RDF Query Language p (SPARQL) was officially introduced in 2008 to retrieve RDF data as easily as SQL does for relatipaal databases. As Web data grows, the number of SPARQT, End- Points constriil ing-SEARQT, queries using HTTP also grows Ihis Situation has attracted people to aggregate data from multiple SPARQL Endpoints akin to conventional distributeddatabases. For instance, if we search data about a drug,

1. INTRODUCTION

The Resource Description Framework (RDF) was introduced in the last decade and has become a standard for exchanging data in the Web. At present, more and more data are converted to RDF. The SPARQL Protocol and RDF Query Language (SPARQL) was officially introduced in 2008 to retrieve RDF data as easily as SQL does for relat~ al databases. As Web data grows, the number of SPA BQT. Eudpoints constnJf!ting SPABQT. queries using HTTP also grows ~ This situation has attracted people to aggregate data from mult iple SPARQL Eudpoints akin to conventional distributed databases. For instance, if we seru:ch data about a drug,

Abby FineReader 5.0

dution el point?

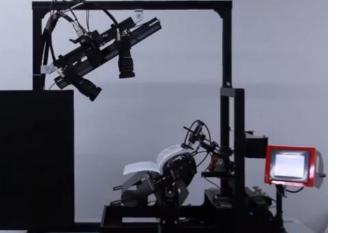
Adobe Acroba

Schritt 1: Scannen

- Auflösung?
- Farbe / Graustufen / schwarz-weiss?
- sinnvolle Parameter vom Dokumenttyp, der Schriftart, der Schriftgröße ... abhängig
 - typisch > 300dpi für saubere 10-Punkt Schrift, sonst 400 bis 600dpi

bis zu 700 Seiten/Stunde

Universität Tokyo, bis zu 250 Seiten/Minute









bis zu 2500 Seiten/Stunde

Schrifterkennung (OCR)

- Inhalt soll erkannt werden: Voraussetzung für Volltextsuche
- Prozess fehlerbehaftet
 - manuelle Korrektur
 - falls Fehler bei der Suche akzeptabel: erkannter Text zum Suchen / Bilder zum Lesen
- Notwendige Schritte:
 - Bildbereinigung: Rauschen herausfiltern, Drehung des Bildes
 - Layout-Analyse: Wo sind Textblöcke, Bilder, Tabellen ... Spaltenstruktur ...



Mit bunten Beeren und Haferstreuseln

Fettarn

Zutaten (für 24 Stücke)

Für die Streusel:

- 150 g geschmolzene Butter
 100 g Zucker
- 2 Päckchen Bourbon-Vanillezuci
- > 100 g Vollkom-Haferflocken > 180 g Mehl

Für den Teig:

> 6 Eier (M), 200 g Zucker, Salz

-) 1 Päckchen Vanillezucker
- > 160 g Joghurt, 1 TL Backpulver
- > 400 g Mehl > 1 kg gemischte Beeren
 - Für die Streusel Butter, Zucker, Vanillezucker, Haferflocken und Mehl zu

Streuseln verrühren. Kalt stellen.

- 2 Für den Rührteig Eier und Zucker mit den Schneebesen des Handrührers ca. 3 Minuten schaumig schlägen. Salz, Vanillezucker und Joghurt unterrühren. Backbulver und Mehl mischen, unter den ven. Beeren putzen, vorsichtig waschen und abtronfen lassen.
- Teig auf ein mit Backpapier belegtes Backblech (ca. 30 x 40 cm) streichen. Beeren darauf verteilen und die Streusel darüberstreuen. Im heißen Ofen ca. 30 Minuten goldbraun backen (E-Herd: 200 °C, Umluft: 180 °C, Gas: Stufe 3).

Pro Stück etwa 235 kcal, 8 g F, 36 g KH, 5 g E.

Schlank-Sterne:

Knusper, knusper - auf acht verschiedene Arten

Das Grundrezept für Streusel können Sie sich leicht merken: 1-1-2, das heißt 1 Teil Butter (z.B. 100 g), 1 Teil Zucker (100 g) und 2 Teile Mehl (200 g). • Krosser werden Streusel, wenn Sie 50 g Mehl weniger nehmen. • Marzipanstreusel: Dafür 100 g Butter, 200 g Marzipanrohmasse, 150 g Mehl und 1 Päckchen Vanillezucker zerkrümeln. • Nusstreusel: 125 g Butter, 75 g Zucker, 150 g Mehl und 75 g gemählene Haselnüsse verkneten. Statt

Pinienkerne nehmen.

• Mohnstreusel: 80 g kalte Butterstücke, 80 g Zucker, 30 g gemahle-

Haselnüssen können Sie auch Wal-

nüsse, Mandeln, Kokosnuss oder



ner Mohn und 80 g Mehl mischen.

• Mit Gewürzen: Je nach Geschmack
Streuselteig mit Gewürzen verfeinern, z.B. mit Vanille, Zimt,
Orangenschale, Zitronen- oder
Rumaroma. • Mit Kakao: Kakaopulver über den Streuselteig stäuben.

• Mit Kuvertüre: Im Wasserbad

geschmolzene Kuvertüre nehmen

Mit Sauerkirschen und zweierlei Streuseln

Gelingt leicht

Zutaten (für 12 Stücke)

- > 150 g weiche Butter
- 150 g Zucker
 Salz, 1 Eigelb (M), 250 g Mehl
- 50 g Speisestärke
 1 TL Backpulver
- > 1 gestrichener EL Kakaopulver
- > 400 g Sauerkirschen (TK oder aus dem Glas)
- > Fett für die Form
- Butter, Zucker, Salz und Eigelb mit den Schneebesen des Handrührers schaumig schlagen. Mehl, Stärke und Backpulver mischen. Unter den Teig rühren, bis Streusel entstanden sind. Eventuell 2 EL kaltes Wasser zugeben.
- 2 Eine Springform (24 cm ø) mit Backpapier auslegen, den Rand einfetten und die Hälfte der Streusel auf dem Formboden verteilen und

andrücken. Die Hälfte der restlichen Streusel mit dem Kakao verkneten. Springform und Streusel 30 Minuten kalt stellen.

3 Kirschen auftauen oder abtropfen lassen. Auf dem Streuselboden verteilen. Übrige Streusel in hellen und Hunklen Streifen auf den Kirschen a. Im heißen Ofen ca. 35 Minuten packen (E-Herd: 200 °C, Umluft: 180 °C, Gas: Stufe 3).

Pro Stück etwa 255 kcal, 12 g F, 34 g KH, 3 g E. Schlank-Sterne:



Rezepte für leckere Streuselkuchen vom Blech verraten wir unter www.bildderfrau.de/streuselkuchen



Beispiel

- Mehrere Spalten
- was gehört zusammen?
- Leserichtung?

Schrifterkennung (OCR)

- Buchstabenerkennung: schwierig für "exotische"
 Schriftarten, Umlaute, Ligaturen, Sonderzeichen, math.
 Formeln ...
- Worterkennung: Lexikonbasiert, schwierig für Spezialvokabular, seltene Sprachen
- Ausgabe: HTML, RTF, Word-Format, PDF, Text-Dateien

Witten, Bainbridge, Nichols 2009

rango mai, Kahore ke poaka? kahore shaere pai ratou i tenei so, kiz tae be ringaringa bei mahi i etahi moni atu hoki ki te okiobinga tapu i te hei hoko i etahi kakahu! ke tini ta rungi. Otiis nua ahan e poka ke.o koutou mes hei hoko kakahu mo. Ne, ka moiri te ra i runga kau o te koutou.

HE TUTAKINGA TESCATA.

Ko taes tawhiti te retenga o Tewera, ka ngaro e tahi o nga whetu marifi, ku oti të hura te kichë manje. o te po, ka kitsa nga kapas ma i te marangai (to ita) ka men to takeans "has takini to see" me reira ke are logracia mai i runga i tuta totuna where, to renguire o to so; ma, ka whatekalesteeris-e to ra non mrani meaniga powi ki te kabu kura, a ka tisho me he peri tahetaha meana; rico ketta atu and te kolta, ka riagihis mai të mahanatanga ki roto ki te awaawa pal i noho si shau; ngaehe nga ran takou, kia ta'a nga kanohi ki të tiliro atu, no te mea ka puhipshia e te muri whenos, ka oho nga manu, tangi ata ka hummunga moto ngaherehere i te waiatalanga o nga manu ki te Atua, mo ratou i tiukine i te roanga o te po, mo te tukunga mai o to ra, hoi whakamahana, a ma nga hua o nga tini rekau o kai a DESCRIPTION OF THE PERSON OF T

Na, ka kises te ava kupikupiko o Waims, takazo ana ngu wai maroma i ronge i to kirikiri, a mimiai rews wie i to won not o Holdings. Offia ko

bac, ka mes toku whaksaro ka meranga i te mon te tahi tangata, a hoere mai ana tetahi i to hunraisi. tujuki pu tuju, a ka ncho ki raro.

Ko " Pai-Maori" tetahi o taua heings, he tengata pole, he mide tukopu, tu-a koumetus, he mawhata te apeko, i pafarahi te ibu, takataka ana nga kannhi, e tokil ana nga paparinga. matetoru ana nga ngutu, keokeo ana nga tukimata, a i hua awi nga wacwao, me te finana katoa. Ko nga kukultu be mea paki te wahi ripa, waho ake to korowai, pahasa rawa ki te kokowsi waho rawa he huru. È rue nga mako, ko tu kuru ponamu i te tahi a nga taringa, he pohui toma i te tahi, be heitiki maori i te kaki, e turu nga hou i tinja ki te matenga, e man assa i te ringaringa. matau te wahi mere kakano pouri. Ko " Ara Pakeko" te ingce o tetahi, he maté ma, he mahara te upoko, i hiamoemos nga karohi, i tussoa te itio, i pakonga nga paperinga, i rahirahi nga ngutu, a i humarire nga mata i to tirohangu atu, he taitamariki hoki, ko nga kakahu, he kahu ma i roto raws, he taras whakshekeheke, he kooms pakeha, he kati mangu kuruto. be potse whiti tons potse, he puts i nga wae, muro ami i te ama te tau o tona wati, he pukapuka i te ringu .-to time men i poke we to pui ke te Kihas i rea te ncheanga e te banga nei whate o toko hos wreter a to Wareton kn men ske a Poi Mosri. "Ekoro, i se taha o taun awa e to ano; to be mes pai ena pungurunguru e mau men i pellegie ai in, chara taun i te i akoe na! Na te cha te tangata repu when se, taongo renes, otha to maori i am al i o te pakehe risenga? aga wairus o ago tangats kiu ora, kin Ku mea su me waimo i aga riterga

rere uga mes entimbli, itsa mes i primgia ai e o talog topuna, a ka taagohin ngu ritengu o nga tani tangata ke."

Aru Paktha .- "I to mea kisno ahan i kito i nga tashi whalataiharo a Thoma me to kororia and o aga men i hames ai e lei, a me sign tini mahi poka ke o te pakella, e pel ana ahan ki nga ritengu o o tatou bipuna ; tena ko tenci, kan thu to maramatunga ki roto ki as, ai ka whakariha. siha ahaa ki nga met o mua ta te inea ke pichia te kino e te bah."

Pat Macri. - "He cha koia tala e mes ki te hinganga o te kino i te paid Kotaka pai to tenet, his fini nen peneties her talia kei, het neuki i tetaki men me toku korokoru, his whose i ngu mea muka, i nga mea meeri mo atietou ko ako tameriki, u ka maon té kai, kai kai, ka boki ki te takuto, and ka puta mai te taka kia mbakatika atu ki te taki, ki te tutuményahir, ki te titotiko' whiste, so Ra tahi tetahi o oku notice User.

pu ahau ki tena rituugo, horeruwa he: whatosrenga ki te wairus. Te men i hanga as tatou kitin tonei. Kan mehin tetahi ki selabi - kis whangoinga to hunga kaikere kia wieskakahuri i te hance rewalkoro—kiu ticking to panil kel he ta ratou hoko i ta riwai. -kin atawhultia te pedaru-kia tieb tirobia-kia owbiatia to turoro-kia boutu te "Taro o te Oranga" ki te longga o nobo ana i te he⇒kia buatu etahi o a tatou rawa ki nga Mikinere.

tawhite, kahore, ki bkoe isse waka-pichel tal I whisi wall washingto-HEIRS TO herby bearing 118 range."

(Her telah) are mipopa ka korire.)

(Ke to hat addition is to Karers o Mei Tirent.)

Merchan, Obotop, 20, 1842. Exced. - E pour una taku ngukau. mo aku mon i ho ete ki ega tangata. moori, hei utu riwai ; kei ki koe, he archa ki te moni taku i pouri si-Kohore, he mamue ke tako, no te men kabore ano i ek mpa mua memi i hontu ni-e run siga hersoi-na ka rico mai, e rua nga kete siwai, pel kan to abus 's non kete, disiré uta nob sal, e monti san tipa riwoi e roto. mo reira i ho nta ni situ lierenii ni. te hoatefongë e tije meni ki ngu tetigets; hacre tone din fattin, inb militi. ku tirchina e an ngh kete, ka wawahin nga ruru, katabi ka kiben nga tuongi i panga ki waenga nai o te kete, bei whakshe i te kamihi o to pakeha, ko to také tezei o te pouritanon o taka ngikuu, ki taku, bo ritonga tahae tenti, kihai su i ata titiro ki nga tangelit i pelia nibi, kia mohiotia ai o ratou ingha, otila e mea. remonen kia cukitekin hei utu molenn ahau, kia tahliukia atu tenci pokapaka ki a koo, kia kite koe i Arm Pakeko.- "Ave ! ka poten te muhi he a nga tangata, kia nichio hoki koe ki to tikanga e cho at to ngakan o to pakeha; lain tenet pukapaka naku ki roto ki tau Nuipepa kin wakarna ni emei tangala kino, kia whakatapato boki laga pakeha

Bed one. Naku.

NA TE TANGATA I TISEBANGATIA NEL R NGA KAS HORG.

[B tikn and u so, e to tangath i timihangatia, aci — e ritemes tahan ta ena. Hoberite, his terios nga kai kas when cangete, cargari, me i mobio koe ki to ki nga topito o te so-kin buste per rator ingon, ko rito, ka arahma rator. taton) te ornero o ngo iwi kates, i ki te kai whakawa. - Kat Tuar.]

rongo mai, Kahore he poaka? kahore haere pai ratou i tenei ao, kia tae he ringaringa hei mahi i etahi moni hei hoko i etahi kakahu? he tini ra o koutou mea hei hoko kakahu mo koutou.

HE TUTAKINGA TANGATA.

Ka taea tawhiti te rerenga o Tawera, ka ngaro e tahi o nga whetu maori, ka oti te hura te kaha mangu o te po, ka kitea nga kapua ma i te marangai (te ita) ka mea te tangata "ka takiri te ata" me reira ka ara kororia mai i runga i tona torona where, te rangatira o te ao; na, ka

atu hoki ki te okiokinga tapu i te rangi. Otiia aua ahau e poka ke .--Na, ka moiri te ra i runga kau o te pae, ka mea toku whakaaro ka maranga i te moe te tahi tangata, a haere mai ana tetahi i te huarahi, tutaki pu raua, a ka noho ki raro.

Ko " Pai-Maori" tetahi o taua hunga, he tangata poto, he moko tukupu, tu-a kaumatua, he mawhatu te upoko, i pararahi te ihu, takataka ana nga kanohi, e tokii ana nga paparinga, matotoru ana nga ngutu, keokeo ana nga tukimata, a i hua nui nga waena katoa. Ko nga

Beispiel: Exotische Sprache (Maori)

44 KO TE KARERE O NUI TIRENI.

Rongo mai, Kahore he poaka? kahore he ringaringa hei mahi i etahi moni hei hoko i etahi kakahu? he tini ra o koutou mea hei hoko kakahu mo koutou.

HE TUTAKINGA TANGATA

Ka taea tawhiti te rerenga o Tawera, ka ngaro e tahi o nga whetu maori, ka oti te hura te kaha mangu o te po, ka kitea nga kapua ma te marangai (*te ita*) ka mea te tangata "ka takiri te ata" me reira ka ara kororia mai i runga i tona torona whero, te rangatira o te ao; na, ka . . .

haere pai ratou i tenei ao, kia tae atu hoki ki te okiokinga tapu i te rangi. Otiia aua ahau e poka ke.—
Na, ka moni te ra i runga kau o te pae, ka mea toku whakaaro ka maranga i te moe te tahi tangata, a haere mai ana tetahi i te huarahi, tutaki pu taua, a ka noho ki raro.
Ko "Pai-Maori" tetahi o taua hunga, he tangata poto, he moko tukupu, tu a kaumatua, he mawhatu te upoko, i pararahi te ihu, takataka ana nga kanohi, e tokii ana nga paparinga, matotoru ana nga ngutu, keokeo ana nga tukimata, a hua nui nga wae-

Aspekte einer DL, die wir nicht behandeln

- Datenformate für Metadaten (Dublin Core, MARC 21, ...)
- Datenformate für Dokumente, insbesondere zur Langzeitarchivierung
- Zugriffsrechte auf Publikationen
- Sonstige rechtliche Fragen: Einscannen, Verarbeitung, Verwendung von Dokumenten
- Abrechnung und Bezahlen von Zugriffen

Beispiel einer DL: ACM Digital Library





Ralf Schenkel

Homepage at Trier University rschenkel@acm.org



- Max Planck Society for the Advancement of Science
- · Max Planck Institute for Informatics
- Saarland University
- · University of Passau



Publication years 2004-20		
Publication count	72	
Citation Count	682	
Available for download	39	
Downloads (6 Weeks)	67	
Downloads (12 Months)	670	
Downloads (cumulative)	12,552	
Average downloads per article	321.85	
Average citations per article	9.47	



http://dl.acm.org/

SEARCH

Search Author's Publications

ROLE

▶ Author only

- · Editor only
- · Other only
- All roles

AUTHOR'S COLLEAGUES
See all colleagues of this author

SUBJECT AREAS

See all subject areas

KEYWORDS

See all author supplied keywords

* CONTACT US

72 results found

Result 1 - 20 of 72

Result page: 1 2 3 4

Sort by: publication date

Export Results: bibtex | endnote | acmref | csv

ate v

Aspect-Based Similar Entity Search in Semantic Knowledge Graphs with Diversity-Awareness and Relaxation

Steffen Metzger, Ralf Schenkel, Marcin Sydow

August 2014 WI-IAT '14: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 01

Publisher: IEEE Computer Society

Bibliometrics:

Citation Count: 1

Downloads (6 Weeks): 7, Downloads (12 Months): 34, Downloads (Overall): 57

Full text available: PDF

Structured knowledge bases are an increasingly important way for storing and retrieving information. Within such knowledge bases, an important search task is finding similar entities based on one or more example entities. We present QBEES, a novel framework for defining entity similarity based only on structural features, so-called aspects, of ...

Keywords: ontology, entity similarity, semantic data, semantic similarity, information retrieval

Beispiel einer DL: ACM Digital Library

Partout: a distributed engine for efficient RDF processing

PDF Get this Article Full Text:

Authors: Luis Galárraga Télécom ParisTech, Paris, France

> Katja Hose Aalborg University, Aalborg, Denmark Ralf Schenkel University of Passau, Passau, Germany

Published in:



Proceeding

WWW '14 Companion Proceedings of the 23rd International

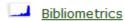
Conference on World Wide Web

Pages 267-268

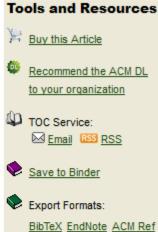
ACM New York, NY, USA @2014

table of contents ISBN: 978-1-4503-2745-9 doi>10.1145/2567948.2577302





- Downloads (6 Weeks): 2
- Downloads (12 Months): 25
- Downloads (cumulative): 139
- Citation Count: 3



Author Tags

Share:



Recent authors with related interests



Concepts in this article





















In The BigDAWG Polystore System

Michael Stonebraker wrote:

"... that spans multiple data models. This trend is fueled by the proliferation of storage engines and query languages based on the observation that a no one size fits alla. To address this shift, we ..."



Query language World Wide Web

Federated database system

Distributed computing Query optimization Relational database

Beispiel einer DL: ACM Digital Library

Abstract Authors References Cited By Index Terms Publication Reviews Comments Table of Contents

Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references.

- 1 L. Galarraga, K. Hose, and R. Schenkel. Partout: A Distributed Engine for Efficient RDF Processing. CoRR, abs/1212.5636, 2012.
- ۹
- 2 Andreas Harth , Katja Hose , Marcel Karnstedt , Axel Polleres , Kai-Uwe Sattler , Jürgen Umbrich, Data summaries for on-demand queries over linked data, Proceedings of the 19th international conference on World wide web, April 26-30, 2010, Raleigh, North Carolina, USA [doi>10.1145/1772690.1772733]
- 3 J. Huang, D. J. Abadi, and K. Ren. Scalable SPARQL Querying of Large RDF Graphs. PVLDB, 4(11):1123--1134, 2011.
- 4 Thomas Neumann, Gerhard Weikum, The RDF-3X engine for scalable management of RDF data, The VLDB Journal The International Journal on Very Large Data Bases, v.19 n.1, p.91-113, February 2010 [doi>10.1007/s00778-009-0165-y]

Abstract Authors References Cited By Index Terms Publication Reviews Comments Table of Contents

3 Citations

Alexander Schätzle, Martin Przyjaciel-Zablocki, Simon Skilevic, Georg Lausen, S2RDF: RDF querying with SP! Proceedings of the VLDB Endowment, v.9 n.10, p.804-815, June 2016

<u>Jinhyun Ahn</u>, <u>Dong-Hyuk Im</u>, <u>Hong-Gee Kim</u>, <u>SigMR: MapReduce-based SPARQL query processing by signature multi-way join</u>, <u>The Journal of Supercomputing</u>, v.71 n.10, p.3695-3725, <u>October 2015</u>

M. Tamer Özsu, A survey of RDF data management systems, Frontiers of Computer Science: Selected Publical Universities, v.10 n.3, p.418-432, June 2016

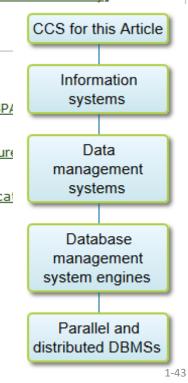
Concepts in Partout: a distributed engine for efficient RDF processing

Resource Description Framework

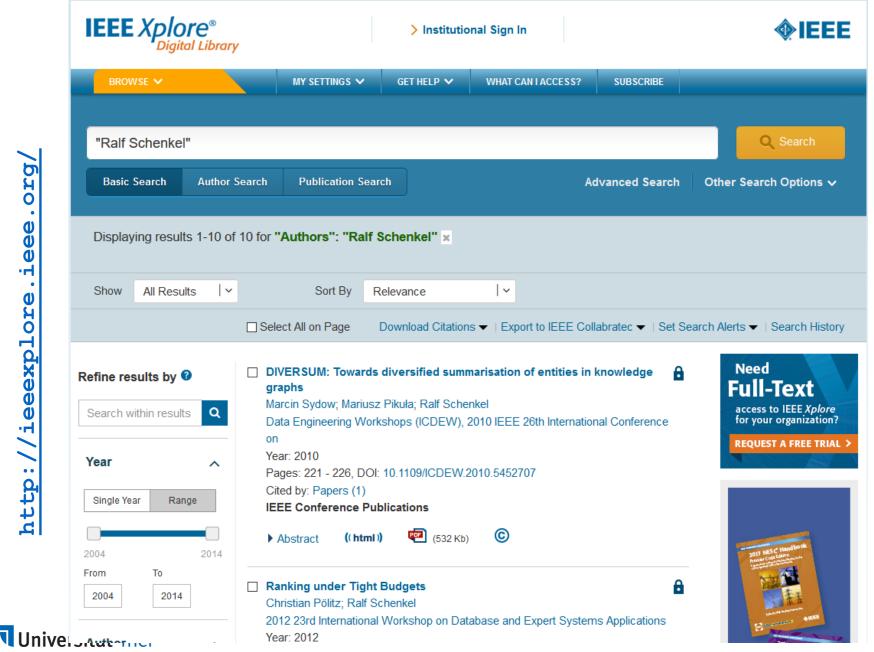
The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats. <u>more from Wikipedia</u>

Distributed computing

Distributed computing is a field of computer science that studies distributed systems. A distributed system consists of multiple autonomous computers that communicate through a computer network. The computers interact with each



Beispiel einer DL: IEEE Xplore



Beispiel einer DL: DBLP



maintained by is SCHLOSS DAGSTUHL at Universität Trier

home

browse

search about

http://www.dblp.org/



▼ Trier 1

Ralf Schenkel 🏮 🛦 🕹 약 🕻 🗩

> Home > Persons

Person information

- affiliation: University of Trier, Germany
- affiliation (former): University of Passau, Germany
- affiliation (former): Saarland University, Saarbrücken, Germany

[-] 2010 - today **©**

2018

- [j21] 📱 🖳 🥰 🕏 Sebastian Michel, Rainer Gemulla, Ralf Schenkel, Theo Härder: Editorial. Datenbank-Spektrum 18(2): 75-77 (2018)
- [c97] 🖹 🕹 🗢 🕏 Ralf Schenkel: Integrating and Exploiting Public Metadata Sources in a Bibliographic Information System. BIR@ECIR 2018: 16-21

Harvesting, Integrating, and Cleaning Metadata in a Bibliographic Information System. Grundlagen von Datenbanken 2018: 8-

■ [c95] 🗏 🕹 🕅 🤻 Mandy Neumann, Christopher Michels, Philipp Schaer, Ralf Schenkel:

Prioritizing and Scheduling Conferences for Metadata Harvesting in dblp. JCDL 2018: 45-48

■ [c94] 🖹 🕹 🤍 📽 Ralph Bergmann, Ralf Schenkel, Lorik Dumani, Stefan Ollinger:

ReCAP - Information Retrieval and Case-Based Reasoning for Robust Deliberation and Synthesis of Arguments in the Political Discourse. LWDA 2018: 49-60

[-] Refine list

by year

showing all 131 records

refine by search term

refine by type

- Books and Theses (only)
- ✓ Journal Articles (only)
- Conference and Workshop Papers (only)
- Parts in Books or Collections (only)
- Editorship (only)
- Reference Works (only)
- Informal Publications (only)

select all | deselect all

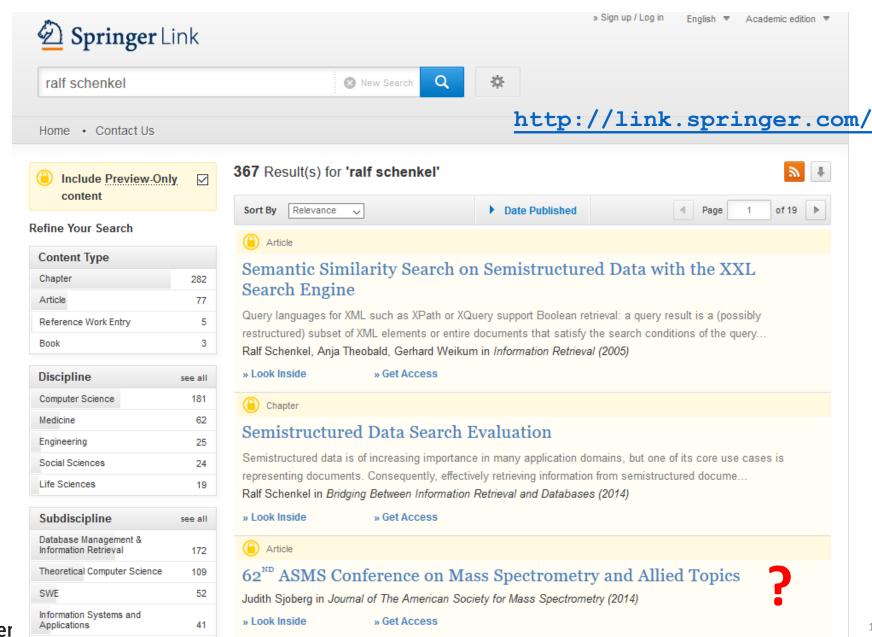
refine by coauthor

Gerhard Weikum (40) Martin Theobald (32) Katja Hose (18) Jaap Kamps (14)

Andreas Broschart (11)

1-45

Beispiel einer DL: Springer Link





Beispiel einer DL: Google Scholar



Ralf Schenkel /

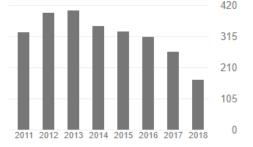
✓ FOLGEN

<u>Universitaet Trier</u>
Bestätigte E-Mail-Adresse bei uni-trier.de - <u>Startseite</u>
Information Retrieval Databases Semantic Web

https://scholar.google.com/

TITEL 🖽	:	ZITIERT VON	JAHR	
M Theobald, G Weil	luation with probabilistic guarantees kum, R Schenkel Thirtieth international conference on Very large data	292	2004	
A Schwarte, P Haas	ion techniques for federated query processing on linked data se, K Hose, R Schenkel, M Schmidt ntic Web Conference, 601-616	270	2011	
R Schenkel, T Crec	uerying over social-tagging networks elius, M Kacimi, S Michel, T Neumann, JX Parreira, 31st annual international ACM SIGIR conference on	182	2008	
M Theobald, R Scho	versatile query engine for TopX search enkel, G Weikum 31st international conference on Very large data bases	178	2005	
H Bast, D Majumda	access optimized top-k query processing r, R Schenkel, M Theobald, G Weikum 32nd international conference on Very large data bases	172	2006	
R Schenkel, A Theo	ent connection index for complex XML document collections obald, G Weikum rence on Extending Database Technology, 237-255	171	2004	
documents J Graupmann, R Sc	rch engine for unified ranked retrieval of heterogeneous XML and web chenkel, G Weikum 31st international conference on Very large data bases	126	2005	
M Theobald, H Bast	and versatile top-k query processing for semistructured data t, D Majumdar, R Schenkel, G Weikum -The International Journal on Very Large Data Bases 17 (1	120	2008	
Efficient creation	and incremental maintenance of the honi index for complex yml	119	2005	

Zitiert von	ALLE ANZEIGEN		
	Alle	Seit 2013	
Zitate	3960	1833	
h-index	32	24	
i10-index	66	43	



Koautoren BEARBEIT		RBEITEN
	Gerhard Weikum Professor of Computer Science	e, >
9	Martin Theobald Professor of Computer Science	>
*	Katja Hose Aalborg University, Denmark	>
	Jaap Kamps University of Amsterdam	>
*	Marcin Sydow Polish Academy of Sciences, a	,n >
9	Sebastian Michel Professor of Computer Science	s, >
1	Andrew Trotman Computer Science, University (of >

Beispiel einer DL: CiteSeerX





Sorted by: Citation Count

Ralf Schenkel

Homepage Not found. Submit a homepage

Affiliation Saarland University

Publications 84 H-index 13 http://citeseerx.ist.psu.edu

Publications

#Cited

105 Top-k Query Evaluation with Probabilistic Guarantees - In VLDB - 2004

- 80 An efficient and versatile query engine for TopX search In VLDB 2005
- 67 FedX: Optimization Techniques for Federated Query Processing on Linked Data - 2011
- 60 HOPI: An efficient connection index for complex XML document collections In 9th Int. Conference on Extending Database Technology (EDBT 2004
- 50 Efficient top-k querying over social-tagging networks In SIGIR 2008
- 45 The SphereSearch Engine for Unified Ranked Retrieval of Heterogeneous XML and Web Documents In VLDB 2005
- 41 lo-top-k: Index-access optimized top-k query processing In VLDB 2006
- 37 Efficient creation and incremental maintenance of the hopi index for complex xml document collections In ICDE 2005
- 32 YAWN: A Semantically Annotated Wikipedia XML Corpus -
- 30 Language-model-based ranking for queries on RDF-graphs - 2009
- 27 TopX: Efficient and versatile top-k query processing for semistructured data VLDB Journal 2008
- 21 Efficient and self-tuning incremental query expansion for top-k query processing In SIGIR 2005
- Federated Transaction Management with Snapshot Isolation In Proceedings of the 8th International Workshop on Foundations of Models and Languages for Data and Objects Transactions and Database Dynamics '99, Dagstuhl 1999

Beispiel einer DL: CiteSeerX

Top-k Query Evaluation with Probabilistic Guarantees (2004)

by Martin Theobald, Gerhard Weikum, Ralf Schenkel

Venue: In VLDB

Citations: 105 - 16 self

Save to List

Add to Collection

Correct Errors

Monitor Changes

Cached



Download Links

[www-db.stanford.edu] [www.vldb.org] [domino.mpi-inf.mpq.de]

racing and the area are

Summary

Citations

Active Bibliography

Co-citation

Clustered Documents

Version History

Abstract

Top-k queries based on ranking elements of multidimensional datasets are a fundamental building block for many kinds of information discovery. The best known general-purpose algorithm for evaluating top-k queries is Fagin's threshold algorithm (TA). Since the user's goal behind top-k queries is to identify one or a few relevant and novel data items, it is intriguing to use approximate variants of TA to reduce run-time costs. This paper introduces a family of approximate top-k algorithms based on probabilistic arguments. When scanning index lists of the underlying multidimensional data space in descending order of local scores, various forms of convolution and derived bounds are employed to predict when it is safe, with high probability, to drop candidate items and to prune the index scans. The precision and the efficiency of the developed methods are experimentally evaluated based on a large Web corpus and a structured data collection.

Keyphrases

top-k query top-k query evaluation probabilistic guarantee known general-purpose algorithm high probability various form underlying multidimensional data space local score novel data item multidimensional datasets approximate top-k algorithm probabilistic argument candidate item run-time cost many kind fundamental building block developed method approximate variant index scan index list user goal fagin threshold algorithm information discovery data collection large web corpus

BibTeX

```
@INPROCEEDINGS{TheobaldO4top-kquery,

author = {Martin Theobald and Gerhard

Weikum and Ralf Schenkel},

title = {Top-k Query Evaluation with

Probabilistic Guarantees},

booktitle = {In VLDB},

year = {2004},

pages = {648--659}
```

Share



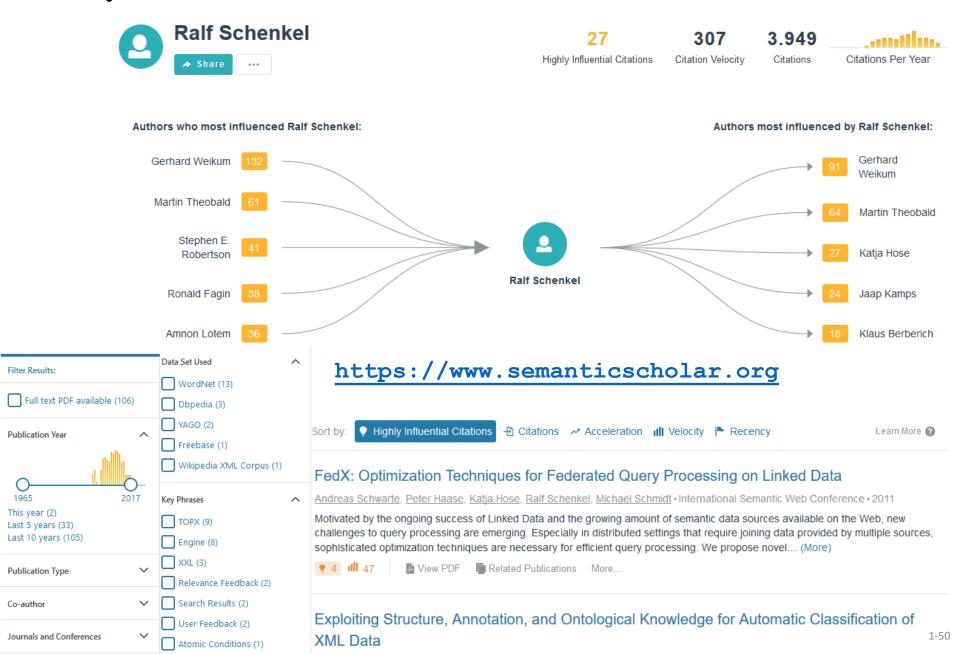








Beispiel einer DL: Semantic Scholar



Beispiel einer DL: Semantic Scholar

FedX: Optimization Techniques for Federated Query Processing on Linked Data

Andreas Schwarte, Peter Haase, Katja Hose, Ralf Schenkel, Michael Schmidt • International Semantic Web Conference • 2011



Abstract

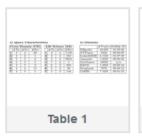
Motivated by the ongoing success of Linked Data and the growing amount of semantic data sources available on the Web, new challenges to query processing are emerging. Especially in distributed settings that require joining data provided by multiple sources, sophisticated optimization techniques are necessary for efficient query processing. We propose novel join processing and grouping techniques to minimize the number of remote requests, and develop an effective solution for source selection in the absence of preprocessed metadata. We present FedX, a practical framework that enables efficient SPARQL query processing on heterogeneous, virtually integrated Linked Data sources. In experiments, we demonstrate the practicability and efficiency of our framework on a set of real-world queries and data sources from the Linked Open Data cloud. With FedX we achieve a significant improvement in query performance over state-of-the-art federated query engines.

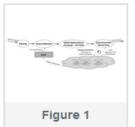
Abstract & Details Figures References Citations Statistics Related Publications

Extracted Key Phrases

FedX Linked Data Query Processing RDF Federation

6 Figures and Tables



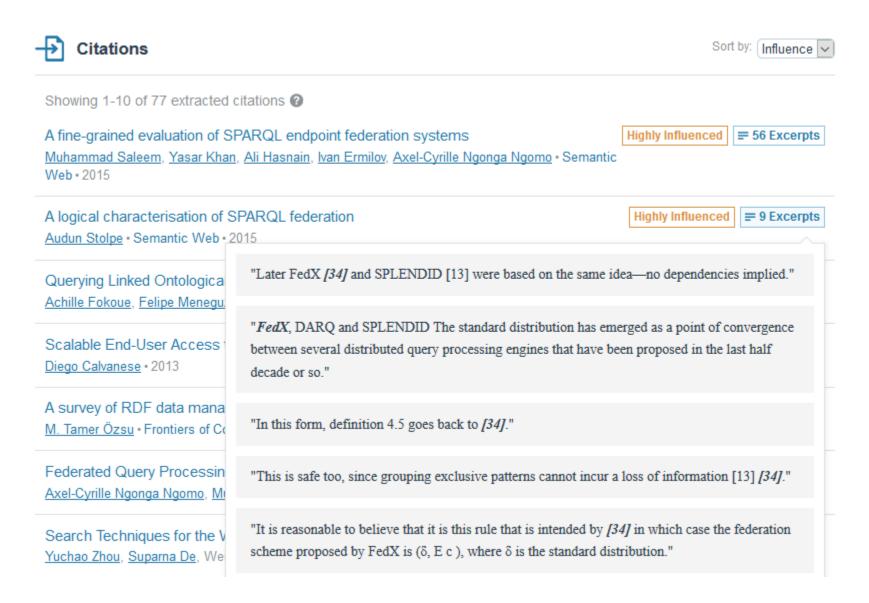








Beispiel einer DL: Semantic Scholar

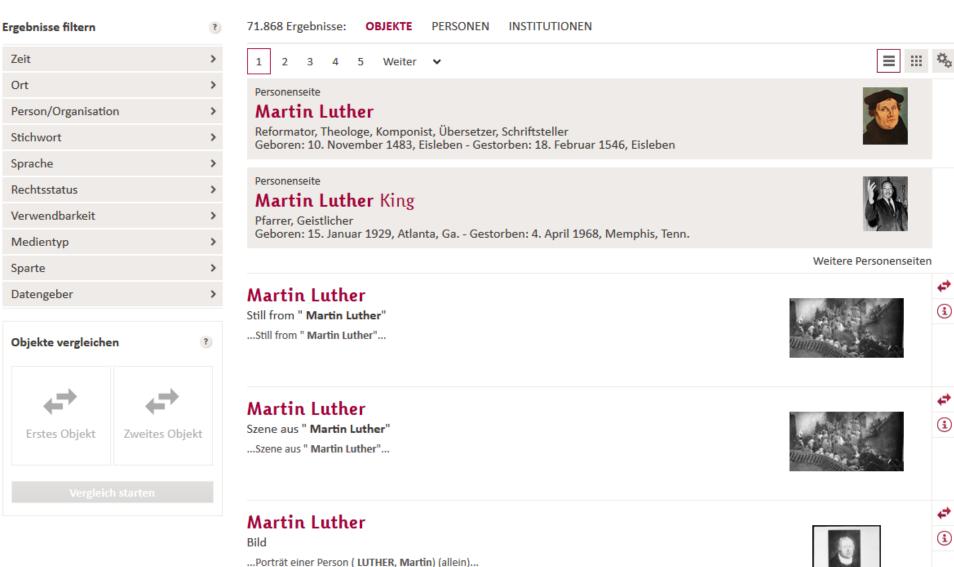




Unterschiede der DL

- Abdeckung der Publikationen
 - Fokus auf einen Verlag
 - Fokus auf "wichtige" Publikationen
 - Fokus auf online verfügbare
- Zugriffsrechte
- Volltext vs. Verweis zur Online-Publikation
- Mächtigkeit des Suchinterfaces
- Aufbereitung der Metadaten, Mehrwertdienste
 - Keyphrases
 - Zitate ein- und ausgehend
 - bibliometrische Maße
 - Einfluss von/auf Autoren und Publikationen

Deutsche Digitale Bibliothek

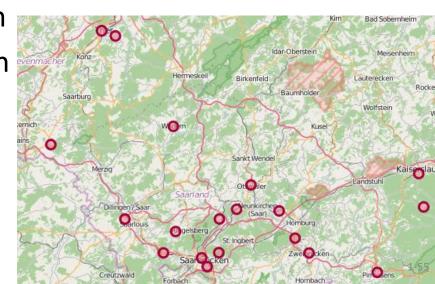


Deutsche Digitale Bibliothek

- Zentrales nationales Zugangsportal für Kultur und Wissenschaft in Deutschland
- Verlinkt die digitalen Angebote der deutschen Kultur- und Wissenschaftseinrichtungen miteinander
- Fördert Aufbau von Kooperationen und die Entwicklung und gemeinsame Nutzung von Diensten und neuartigen Werkzeugen
- Zugang zu digitalisierten Beständen und Erschließungsinformationen aus Kultur- und Wissenschaftseinrichtungen wie Bibliotheken, Archiven, Museen, Denkmalämtern, Mediatheken sowie Universitäten

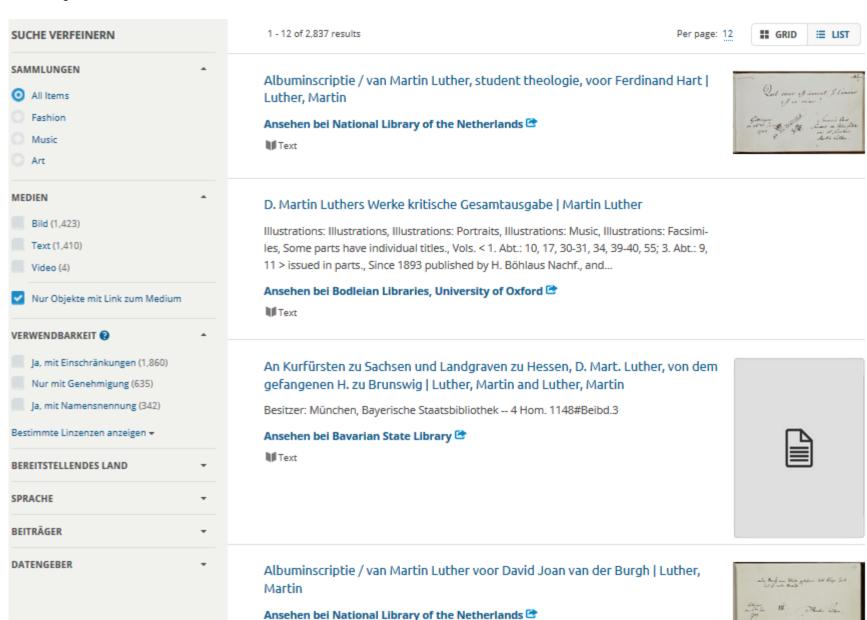
und anderen Forschungseinrichtungen

 Zugang zu den digitalen Abbildern von Büchern, Urkunden und Akten, Gemälden, Statuen, Installationen, Denkmälern bis hin zu Filmen und Musik



Europeana

http://www.europeana.eu/



III Text

Europeana

- Zugänglich machen des wissenschaftlichen und kulturellen Erbes Europas von der Vor- und Frühgeschichte bis in die Gegenwart in Form von Bild-, Text-, Ton- und Video-Dateien
- Inhalte werden von den angeschlossenen europäischen Institutionen zur Verfügung gestellt
- Keine zentrale Datenhaltung, lediglich zentraler Metadatenund Suchindex
- Eigener Metadaten- (d.h. Beschreibungs-) Standard
- Forschung zur Zugangsverbesserung als wesentliche Aufgabe
- 58,163,798 Objekte (Stand 15. Oktober 2018)

Wissenschaftliche Bibliotheken in DE (1)

- Wissenschaftliche Spezialbibliotheken: Firmen,
 Forschungseinrichtungen, Behörden, Parlamente, Museen,
 ... (Beispiel: Schloss Dagstuhl, Wadern)
- Regionalbibliotheken: 40 Landes- oder Staatsbibliotheken (früher Fürsten- und Hofbibliotheken, heute in der Trägerschaft der Länder)
- Universitätsbibliotheken: 79 Universitäten
 - Einschichtige Bibliothekssysteme: zentral organisiert, z.B. Trier
 - Zweischichtige Bibliothekssysteme: Vielzahl selbständiger Bibliotheken, an älteren Universitäten, z.B. Aachen
- Sonstige **Hochschulbibliotheken**: 130, insb. Fachhochschulen

Wissenschaftliche Bibliotheken in DE (2)

- Nationalbibliothek: In Deutschland gab es bis 2006 keine mit der Library of Congress oder der British Library vergleichbare Nationalbibliothek
- Die Deutsche Bibliothek, seit 2006 Deutsche Nationalbibliothek http://www.d-nb.de/:
 - Deutsche Bücherei, Leipzig, seit 1912, dt. Nationalbibliothek, bis 1969 Pflichtexemplarbibliothek von BRD und DDR
 - Deutsche Bibliothek, Frankf./Main, 1946 gegründet, seit 1974 dt. ISBN-Zentrale, Deutsche Bibliographie

Zusammenschluss nach Wiedervereinigung

		9	O	
GND			GND	
Link zu diesem Datensatz	http://d-nb.info/gnd/12372094X		Link zu diesem Datensatz	http://d-nb.info/gnd/1054050813
Name Schooled Polf		Person		
Publikationen von Autoren mit diesem Namen	3 Publikationen 1. News [electronic resource]		Akademischer Grad	DiplInf. Prof. DrIng.
	In: Datenbank-Spektrum 20.6.2016: 1-4	Land	Deutschland (XA-DE)	
		Beruf(e)	Informatiker Hochschullehrer	
		Weitere Angaben	Vertretungsprofessur Universität Passau, Lehrs	
Beteiligt an	1. QBEES: query-by-example entity search in semantic knowledge graphs based on maximal i	Beziehungen zu Organisationen	Universität Passau Universität Trier	
		Тур	Person (piz)	
			1-59	

Wissenschaftliche Bibliotheken in DE (3)

- Zentrale Fachbibliotheken (alle in der Leibniz-Gemeinschaft):
 - überregionalen Literatur- und Informationsversorgung für den hochspezialisierten Bedarf von Wissenschaft und Forschung
 - Möglichst vollständige Abdeckung der Literatur im Fachgebiet
 - In DE:
 - Technische Informationsbibliothek, Hannover (Finanzierung: 30% Bund, 70% Länder)
 - Zentralbibliothek für Medizin, Köln (30B/70L)
 - Zentralbibliothek für Wirtschaftswissenschaften, Kiel (50B/50L)
- Fachinformationsdienste (DFG-Förderung):
 - Derzeit (2018) 38 teilweise sehr spezifische Fächer abgedeckt, Beispiele: Kriminologie (TÜ), Pharmazie (BS), Darstellende Kunst (F), Afrikastudien (F), Benelux (MS), Anglo-American Culture (B)
 - Ergänzt die zentralen Fachbibliotheken
 - möglichst schneller und direkter Zugriff auf Spezialliteratur und forschungsrelevante Informationen => digitale Dienste
 - Getragen von leistungsstarken Bibliotheken und Forschungseinrichtungen

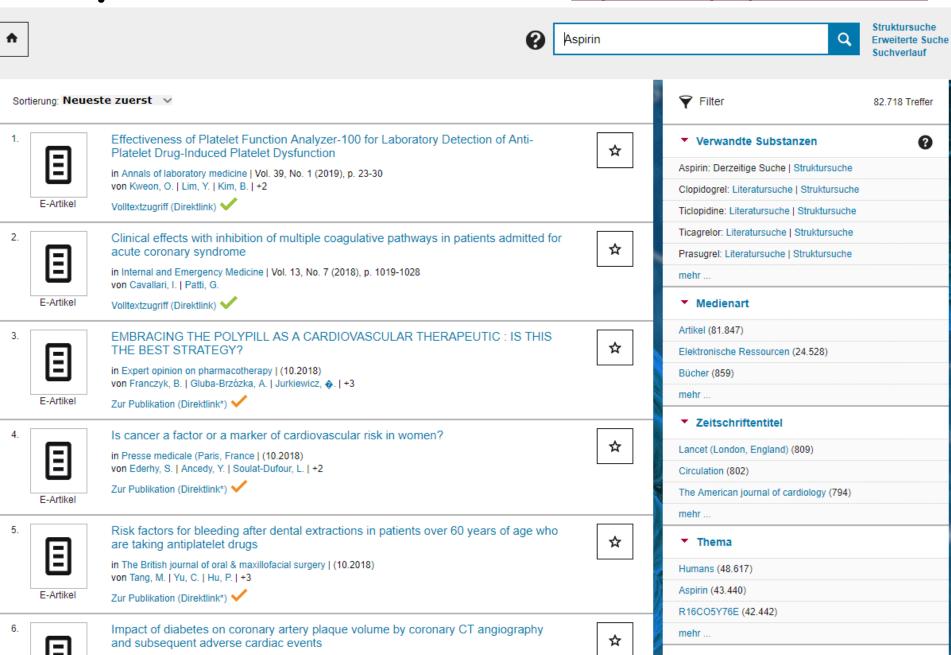
Beispiel: FID Pharmazie

- http://pharmazie.fid-lizenzen.de/
- Ziele:
 - überregionale Verfügbarkeit von elektronischen Fachzeitschriften und E-Books
 - nutzerzentrierter Ausbau einer erweiterbaren und personalisierbaren Informationsinfrastruktur
 - erhöhte Flexibilität und Passgenauigkeit des Dienstangebots
 - Umsetzung einer E-Only-Policy
 - Digitalisierung und Langzeitarchivierung von Quellen
 - direkte Einbindung von Fachdatenbanken
 - Retrieval-Dienste im Rahmen eines neuen Discovery-Systems als Keimzelle einer virtuellen Forschungsumgebung
 - Zusammenarbeit mit der Fachcommunity



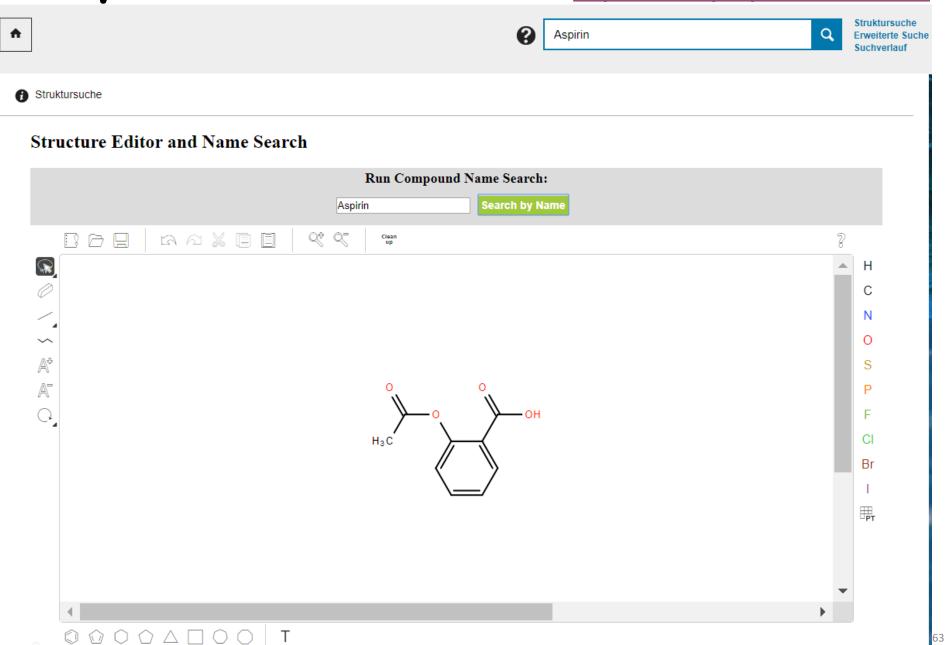
Beispiel: FID Pharmazie

https://www.pubpharm.de/vufind/



Beispiel: FID Pharmazie

https://www.pubpharm.de/vufind/



Wissenschaftliche Bibliotheken

Bestand

2015

Wissenschaftliche Bibliotheken	Fachhochschul-/ Hochschulbiblio- theken	National-/Zentrale Fachbibliotheken	Regional- bibliotheken	Universitäts- bibliotheken
Buchbestand	18.274.110	48.881.661	21.835.149	166.976.387
Bestand sonst. Druckwerke	4.906.757	8.999.481	1.942.557	33.667.032
Bestand nicht- elektronischer Materialien (analoge AV-Materialien; Tonträger, Dias, Filme, Videos, Kunstdrucke, Bilder etc.)	691.165	18.492.341	1.597.114	27.438.102
Bestand Handschriften	9.894	919.631	410.244	1.151.633
Digitaler Bestand ohne digitale Zeitschriften	10.708.044	4.543.654	541.215	12.059.249
Bestand gedruckter Zeitschriften-Abonnements	40.217	132.045	54.076	182.646
Bestand digitaler Zeitschriften-Abonnements	1.519.635	116.664	188.076	2.687.298

Wissenschaftliche Bibliotheken verfügen über einen Bestand von über 387 Millionen Bücher und Zeitschriften, analog und digital. Dazu kommen rund 2,5 Millionen Handschriften und Nachlässe.

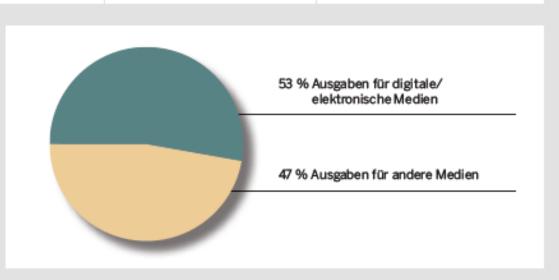
Wissenschaftliche Bibliotheken

Digitale Medien

2015

Ausgaben für Erwerbung	Ausgaben insgesamt Mio €	Ausgaben für digitale/ elektronische Medien Mio €	Anteil von den Gesamt- ausgaben für Erwerbung %
Fachhochschul-/ Hochschulbibliotheken	40,65	17,95	44
National-/Zentrale Fachbibliotheken	34,82	7,59	22
Regionalbibliotheken	14,70	1,71	12
Universitätsbibliotheken	237,28	147,41	62

Erstmals werden mehr als die Hälfte der Erwerbungsausgaben in Wissenschaftlichen Bibliotheken für digitale und elektronische Medien aufgewendet.



Vorteile einer DL (nach Arms, 2001)

- DL brings the library to the user
- Computer Power is used for searching and browsing
- Information can be shared
- Information is easier to keep current
- Information is always available
 - Öffnungszeiten
 - Material wird nicht ausgeliehen
- New forms of information become possible