

# Speicherung, Indexierung, Suche und Zugriff

- Speichern und Archivieren digitaler Dokumente
- Textsuche; Indexierung und Speicherungstechniken
- Information Retrieval
- Suchmaschinen
- Verteilte Suche

G. Görz, FAU, Informatik 8

## Relationale Datenbanken

- Systeme zum Aufbau, Verwaltung, Zugriff auf (Kollektionen von) Tabellen

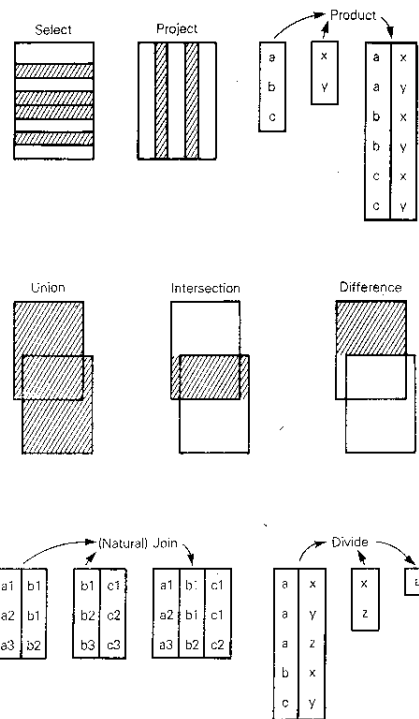
- Zeilen: Datensätze

- Spalten: „Attribute“ (Menge von gleichartigen Attributwerten)

- Primärschlüssel

- Operationen der relationalen Algebra

- Systemdienste: Transaktionen, Datenschutz, Datensicherheit (Recovery), Replikation



# Speichern und Archivieren digitaler Dokumente

- Physische Speichermedien
  - magnetische
  - optische
- Speicherorganisation und Datenverwaltung
  - Dateisystem
  - Relationale Datenbanken, DBMS
  - Objekt- und Volltext-Datenbanken
  - Multimedia-Datenbanken
  - Domument-Management-Systeme
- Speicherformate

G. Görz, FAU, Informatik 8

## Langzeitarchivierung

- Probleme
  - Lebensdauer von Speichermedien
  - Technischer Wandel
  - Authentizität: Vertrauenswürdige Archive
- Migration: Erhalt
  - des Dokumenteninhalts (Bitfolge)
  - des Zugangs zum Dokument unter weitestgehender Bewahrung der Wahrnehmung
- Daten-Migration
- Emulation: Datenträger und Wiedergabesoftware
- Gerätemuseum
- *Erhaltungs-Metadaten*

G. Görz, FAU, Informatik 8

# Suche: Was? Wo?

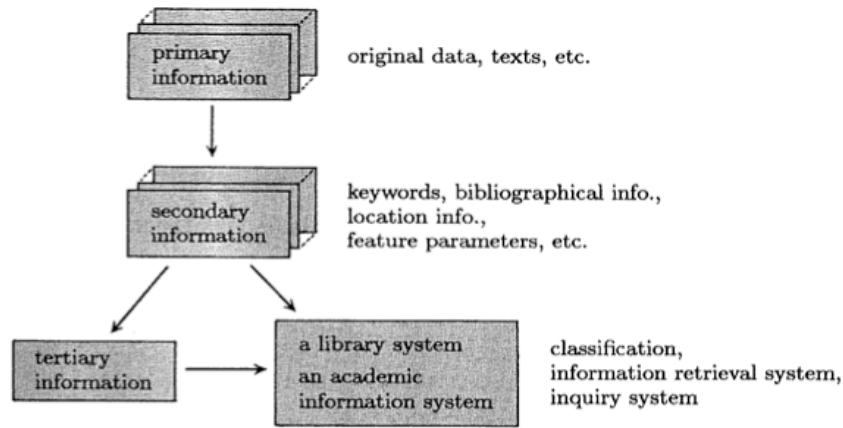


Figure 2.1 Systems for making use of various kinds of information.


G. Görz, FAU, Informatik 8

Bookmarks Location: <http://www.vd17.de/> What's Related

Internet Lookup New&Cool [www.nypl.org/library/fogelman.htm](http://www.nypl.org/library/fogelman.htm)


## Das Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts

Im Projekt *Verzeichnis der im deutschen Sprachraum erschienenen Drucke des 17. Jahrhunderts*, kurz **VD17**, wird eine retrospektive Nationalbibliographie für den Zeitraum von 1601 bis 1700 erstellt.



Die Katalogisierung begann am 1.7.1986. Die Deutsche Forschungsgemeinschaft plant eine Förderungszeit von 10 – 12 Jahren. Berücksichtigt werden alle deutschsprachigen Titel und, unabhängig von ihrer Sprache, alle in dem historischen deutschen Sprachgebiet gedruckten und verlegten Werke. Nicht verzeichnet werden Musica practica und Karten.

Die Verzeichnung der Drucke erfolgt nach Autopsie der Originale. Zur besseren Information und zur Erleichterung der Ausgabedifferenzierung sind den Titelaufnahmen digitale Images von ausgewählten Seiten (Schlüsselseiten) der Drucke beigelegt.

**VD17**  
[VD17 homepage](#)  
**Projektinformation**  
 Partner  
 Kontakte  
**Datenbankinformation**  
 Inhalte  
 Bibliographische Beschreibung  
 Schlüsselseiten  
**Recherche**  
  
 no frames

**VD17-Datenbank** [VD17-Online-Hilfe](#)


**Volltextanfrage:**


Volltext:

**Kurzinformation:**

- Schreibweisen:**  
Bitte Umlaute und ß bei der Suche auflösen: ae, oe, ue, ss
- Trunkierung:**  
% ersetzt eine beliebige Zeichenreihe (z.B. Frank%)
- Suchbeispiele:**  
Weltspiegel  
Rerum Germanicarum  
Augsburg & 1680  
Gedichte | Poemata  
Dissertat%

[Zur Maskenanfrage](#) [Lexikon](#) [Hilfextext zur VD17-Volltextanfrage](#)


 FORWISS im Projekt VD17, 01-06-99; [Hachim Haddouti](#), [Wolfgang Wöhner](#)

**VD17**  
[VD17 homepage](#)  
**Projektinformation**  
 Partner  
 Kontakte  
**Datenbankinformation**  
 Inhalte  
 Bibliographische Beschreibung  
 Schlüsselseiten  
**Recherche**  
  
 no frames

**VD17-Trefferliste**

**Treffer 1 bis 20**

- Matthiae, Christian (1622): [Oratio Panegyrica De Vita Et Obitu Johannis Schröderi, Theologi Celeberrimi Et Ecclesiae Noribergensis, Quae Est Ad D. Laurenti Superintendentis Dignissimi & meritissimi.](#)
- Crinesius, Christopherus (1626): [Disputationum Theologicarum Secunda, De Fide Catholica B. Petri, Pontificum Omnium Primi Et Maximi...](#)
- Leibniz, Gottfried Wilhelm (1664): [Specimen Quaestionum Philosophicarum Ex Jure Collectarum.](#)
- Eschenbach, Christian (1664): [Gülden Kleinod Davids/ aus dem 16. Psalm/ Zum dreyfachen Trost denen/ Die den Herrn allezeit für Augen haben; Bey... Leichbegängniß Der... Susanna/ Gebornen Leibnizin/ Deß... Josuae Hagers/ Wolverdienten Diaconi. in der Nürnbergischen Vorstad und Marckt Wehrd/ geliebten Haußfrauen.](#)
- Schwendendörffer, Bartholomäus Leonhard (1665): [Disputatio Juridica De Conditionibus.](#)
- Leibniz, Gottfried Wilhelm (1666): [Disputatio Inauguralis De Casibus Perplexis In Jure.](#)
- Leibniz, Gottfried Wilhelm (1668): [Nova Methodus Discendae Docendaeque Jurisprudentiae.](#)
- Leibniz, Gottfried Wilhelm (1668): [Ratio Corporis Juris Reconcinnandi.](#)
- Dür, Johann Conrad (1675): [Theses Philosophicae Quibus Ostenditur Quousque Fides Docentibus A Discipulis Sit Habenda.](#)
- Leibniz, Gottfried Wilhelm (1677): [Caesarini Fürsteneri- De Jure Suprematus Ac Legationis Principum Germaniae.](#)
- Leibniz, Gottfried Wilhelm -von- (1678):

**VD17**  
[VD17 homepage](#)  
**Projektinformation**  
 Partner  
 Kontakte  
**Datenbankinformation**  
 Inhalte  
 Bibliographische Beschreibung  
 Schlüsselseiten  
**Recherche**  
  
 no frames

Bilder: 1 2

Satz: H Id-Nr.: 1:009030R

Leibniz, Gottfried Wilhelm

Disputatio Inauguralis De Casibus Perplexis In Iure quam Indultu Nobiliss. & Ampliss. Ictorum Ordinis in celeberrima Universitate Norica ... Proponit M. Gottfredus Guilielmus Leibnizius ... d. 5. Mens. Nov. A. MDCLXVI

[Druck:] [Altdorf] : Hagen  
--> Altdorf: Hagen, Georg <Witwe>

1666  
36 S. 1 Ill. (Kupferst.) 4°

Ill. enthält 3-eckige Drehscheibe


Fingerprint: ast. n-eu o-l. indu 3 1666R

Gattung/Fach: Dissertation: jur.

Exemplar(e):

Id-Nr.: 1:009031Y SB Berlin  
Sign.: FI 5019

[Zurück zur Trefferanzeige](#) [Zur Volltextanfrage](#) [Zur Maskenanfrage](#) [Strukturfelder](#)

 [DFG](#) [FORWISS im Projekt VD17](#), 01-06-99; [Hachim Hadjouti](#), [Wolfgang Wehner](#)

DISPUTATIO INAUGURALIS  
DE  
CASIBUS PERPLEXIS IN JURE.

Indultu Nobiliss. & Ampliss. Ictorum  
Ordinis in celeberrima Universitate Norica  
PRO LICENTIA  
ASSUMENDI HONORES ET PRIVILEGIA DOCTORALIA  
PROPONIT  
M. GOTTFREDUS GUILIELMUS LEIBNIZIUS  
LIPSIENSIS, J. U. BACC.



## Allgemeines zur Suche (1)

- Semantische Strukturen als Voraussetzung wissensbasierter Suche
- Linking-Infrastruktur
  - Direkte Verknüpfung verschiedener Dokumente nach sachlichen Kriterien (Nachschlagewerke, Lexika, Kommentare,...)
  - Zitatenvernetzung, „reference linking“  
Research Index/CiteSeer; SFX; Springer Link
  - Kontextbezogene Navigation
- Suche mit Kombination verschiedener Ressourcen
  - Anfragen mit Einbezug ähnlicher Objekte
  - Zugriff auf zusammengesetzte Objekte
  - Zugriff mit abgeleiteten Attributen
  - Konsolidierung und Konsistenzprüfung



Q. D. B. V.  
p. l.

**D**E casibus perplexis commentaturi breviter videamus: tum *quis casus perplexus sit*, tum posito, quod datus casus sit perplexus, *quid de eo statuendum*. Illic natura, causa, subiecta, notio & vis termini; hic accidentia, effectus, prædicata, usus deniq; in ipsis rerum argumentis consistens explicatur. Et quemadmodum in pterisq; materiis separata facti & juris tractatio est, *Comr. Logus Meth. Fer. p. 1. c. 1. n. 1. Job. Althus. Dicool. lib. 1. c. 1. n. 4.* ita hic illud plus facti, hoc plus juris habere crediderim.

I I.

Porro ipsa *Casus perplexi* notio ex duobus componitur: *Casu* & *Perplexitate*. *Casus* vocem apud Mechanicos natam vicini Geometrae primæ adhibuerunt in rem suam, quibus *stans* est ipsa figura, ( seu linearum, superficialium, corporum ad se invicem positus, ) de qua deinde *γίγνεται*, nimirum quantitas, ratio, analogia, uti apud Jctorum *de facto* ) demonstratur. Unde problemata eorum *πρόβλημα* similia sunt controversiis Jctorum distinctione expediendis, add. *Dn. Erhard. Weigel. Analys. Euclid. sect. 2. c. 12. n. 4.* & impossibilia casibus perplexis, in quibus pro solutione est advertere & demonstrare impossibilitatem, quod Algebra præstat. Manavit inde vox ad Medicos, & Theologos Moralistas

View or download:  
[doc.mmu.ac.uk/STAFF/michael/m...kr91.ps](http://doc.mmu.ac.uk/STAFF/michael/m...kr91.ps)  
[cs.ucl.ac.uk/staff/A.Hunter/kr91.ps](http://cs.ucl.ac.uk/staff/A.Hunter/kr91.ps)  
[csc.liv.ac.uk/~michael/kr91.ps](http://csc.liv.ac.uk/~michael/kr91.ps)  
 Cached: [PS.gz](#) [PS](#) [PDF](#) [Image](#) [Update](#) [Help](#)

From: [doc.mmu.ac.uk/RESEARCH\\_group\\_pub](http://doc.mmu.ac.uk/RESEARCH_group_pub) (more)  
 From: [cs.ucl.ac.uk/staff/A.Hun...papers](http://cs.ucl.ac.uk/staff/A.Hun...papers)  
 (Enter author homepages)

**Meta-Reasoning in Executable Temporal Logic (1991)** (Make Corrections) (24 citations)  
 Howard Barringer, Michael Fisher, Dov Gabbay, Anthony Hunter  
 Proceedings of Second International Conference on Principles of Knowledge Representation and Reasoning (KR)

**CiteSeer** Home/Search Context Related

(Enter summary) Rate this article: 1 2 3 4 5 (best)  
[Comment on this article](#)

**Abstract:** Temporal logic can be used as a programming language. If temporal formulae are represented in the form of an implication where the antecedent refers to the past, and the consequent refers to the present and the future, then formulae that have their antecedent satisfied can be satisfied by considering the consequent as an imperative to be obeyed. Such a language becomes a natural alternative to programming languages, such as Prolog, for temporal reasoning. Here, the approach is extended to... [\[Update\]](#)

Context of citations to this paper: [More](#)

... value of concurrently executing METATEM systems particularly for DAI applications was immediately recognised (Fisher and Barringer, 1991). Concurrent METATEM is a simple operational framework which allows societies of METATEM processes to communicate and cooperate....

... the ability to name and manipulate formulae and rules and, thus, meta level capabilities are provided as basic mechanisms within METATEM [3]. Such mechanisms provide a wide range of flexibility in developing, reconfiguring and dynamically modifying the behaviour of systems....

Cited by: [More](#)  
 Compositional Verification of Multi-Agent Systems - In Temporal Multi-Epistemic (Correct)  
 Temporal Representation - Of Ecological Knowledge (Correct)  
 Execution and Proof in a Horn-Clause Temporal Logic - Dixon, Fisher, Reynolds (1997) (Correct)

Active bibliography (related documents): [More All](#)  
 0.4: A System to Read Names and Addresses on Tax Forms - Srihari, Shin, Ramanaprasad.. (1996) (Correct)  
 0.3: METATEM: A Framework for Programming in Temporal Logic - Barringer, Fisher.. (1989) (Correct)  
 0.2: MetateM in Intensive Care - Reynolds (1997) (Correct)

Similar documents based on text: [More All](#)  
 0.2: Fair SMG and Linear Time Model Checking - Barringer, Fisher, Gough (1989) (Correct)  
 0.2: Proof Rules for Automated Compositional Verification - Through Learning Howard (Correct)

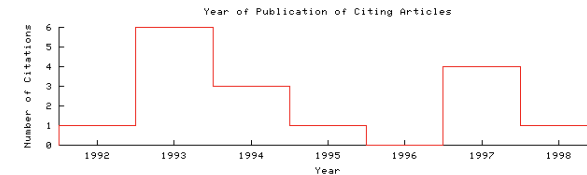
Related documents from co-citation: [More All](#)  
 17: METATEM: A Framework for Programming in Temporal Logic - Barringer, Fisher et al. - 1989  
 9: Actors: a model of concurrent computation in distributed systems (context) - Agha - 1988  
 8: From the Past to the Future: Executing Temporal Logic Programs - Fisher, Owens - 1992

BibTeX entry: [\[Update\]](#)

H. Barringer, M. Fisher, D. Gabbay, and A. Hunter. Meta-Reasoning in Executable Temporal Logic. In J. Allen, R. Fikes, and E. Sandewall, editors, Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR), Cambridge, Massachusetts, April 1991. Morgan Kaufmann.  
[http://www.doc.mmu.ac.uk/RESEARCH/LoCo/group\\_pub.html](http://www.doc.mmu.ac.uk/RESEARCH/LoCo/group_pub.html) [More](#)

```
@inproceedings{barringer91metareasoning,
  author = "H. Barringer and M. Fisher and D. Gabbay and A. Hunter",
  title = "Meta-Reasoning in Executable Temporal Logic",
  booktitle = "Proceedings of Second International Conference on Principles of Knowledge Representation and Reasoning (KR)",
  publisher = "Morgan Kaufmann publishers Inc.: San Mateo, CA, USA",
  year = "1991",
  url = "citeseer.ist.psu.edu/barringer91metareasoning.html" }
```

- Citations (may not include all citations):
- 110 LUSTRE: A Declarative Language for Programming Synchronous S.. (context) - Caspi, Pilaud et al. - 1987
  - 109 Using Branching Time Temporal Logic to Synthesise Synchronis.. (context) - Clarke, Emerson - 1982
  - 67 A Common Sense Theory of Time (context) - Allen, Hayes - 1985
  - 59 METATEM: A Framework for Programming in Temporal Logic - Barringer, Fisher et al. - 1989
  - 55 Analysis of MetaPrograms (context) - Hill, Lloyd - 1988
  - 44 Languages with Self Reference I: Foundations (context) - Perlis - 1985
  - 18 Temporal-Logic Theorem Proving (context) - Abadi - 1987
  - 5 Department of Computer Science (context) - Report - 1990
  - 4 Implementing a Prototype METATEM Interpreter (context) - Fisher - 1989
  - 1 Programming and Prolog (context) - Press, Nilsson et al. - 1990
  - 1 Tokio: Logic Programming Language based on Temporal Logic an.. (context) - report, Department et al. - 1986
  - 1 The Temporal Logic of Programs (context) - Intelligence, Pnueli - 1977
  - 1 Declarative Past and Imperative Future: Executable Temporal .. (context) - rd, on et al. - 1989
  - 1 Using TEQUEL (context) - Owens - 1990
  - 1 Execution and Application of Temporal Modal Logic (context) - Torsun, Manning - 1990



The graph only includes citing articles where the year of publication is known.

Documents on the same site ([http://www.doc.mmu.ac.uk/RESEARCH/LoCo/group\\_pub.html](http://www.doc.mmu.ac.uk/RESEARCH/LoCo/group_pub.html)): [More](#)  
 Fair SMG and Linear Time Model Checking - Barringer, Fisher, Gough (1989) (Correct)  
 Concurrent METATEM Processes - A Language for Distributed AI - Fisher, Barringer (1991) (Correct)  
 Executable Temporal Logic for Distributed A.I. - Fisher, Woodriddle (1993) (Correct)

[Online articles have much greater impact](#) [More about CiteSeer.IST](#) [Add search form to your site](#) [Submit documents](#) [Feedback](#)

CiteSeer.IST - Copyright Penn State and NRC

24 citations found. Retrieving documents...

Barringer, H., M. Fisher, D. Gabbay, and A. Hunter (1991). *Meta-Reasoning in Executable Temporal Logic*, in: J. Allen, R. Fikes, E. Sandewall, Proc. of the 2nd Int. Conf. on Principles of Knowledge Representation and Reasoning, KR'91.

**CiteSeer** Home/Search Document Details and Download Summary Related Articles Check

This paper is cited in the following contexts:

[Temporal Representation - Of Ecological Knowledge](#) (Correct)

...to allow modal and temporal connectives, that is based on a linear point structure of time. The linear point structure could lead us to the same problems as in the Gray's attempt. Another possibility could be the temporal logic with Since and Until and fixed point operators [Gabbay 89] Barringer et al. 91) However, as pointed by the authors the logic is based on the natural numbers as flow of time (linear structure) and is suitable for controlling process behaviour. Consequently, such logic is not suitable for our purposes as mentioned in section 1.3. The logic based on intervals proposed in ...

Howard Barringer, Michael Fisher, Dov Gabbay, and Anthony Hunter. *Meta-reasoning in executable temporal logic*. In Knowledge Representation, 1991.

[Bounded Model Search in Linear Temporal Logic and its... - Cerrito, Mayer](#) (3 citations) (Correct)

...e cts are destroyed) and classical frame conditions for all literals. A detailed description of such encodings, and the induced plan search mechanisms, would override space limits and will be given elsewhere. The application of LTL to planning has been considered in other works. For example, [3] applies the executable temporal language METATEM [2] to planning and scheduling. F. Bacchus and F. Kabanza [1] use a version of temporal logic to specify temporally extended goals as sets of acceptable sequences of states, i.e. temporal models, and de ne correct plans as those whose execution ...

H. Barringer, M. Fisher, D. Gabbay, and A. Hunter. *Meta-reasoning in executable temporal logic*. In Proc. of the Second Int. Conf. on Principles of Knowledge Representation and Reasoning, 1991.

[Using Linear Temporal Logic to Model and Solve Planning Problems - Cerrito, Mayer \(1998\)](#) (1 citation) (Correct)

...system that may be open to the user. The naturalness of semantically based methods makes it possible to conceive mixed initiative planning systems, where proofs are expected to convey understandable information. The application of LTL to planning has been considered in other works. For example, [4] apply the executable temporal language METATEM [3] to planning and scheduling. F. Bacchus and F. Kabanza [2] use a version of temporal logic to specify temporally extended goals as sets of acceptable sequences of states, i.e. temporal models, and define correct plans as those whose execution ...

H. Barringer, M. Fisher, D. Gabbay, and A. Hunter. *Meta-reasoning in executable temporal logic*. In Proc. of the Second Int. Conf. on Principles of Knowledge Representation and Reasoning, 1991.

[The Main Features of a Planner Fully Based on LTL - Mayer, Limongelli..](#) (Correct)

...For example, Stephan Biundo 1996; Koehler Treinen 1995) view planning as theorem proving in intervalbased temporal logics; Cimatti et al. 1997; Cimatti Roveri 1999) propose planning via model checking. Linear time temporal logic (LTL) has also been used (Bacchus Kabanza 1996a; 2000; Barringer et al. 1991), since it is based on a simple model of time and allows an easy and natural representation of a world that changes over time. In this work we propose the use LTL as a speci cation language for planning, keeping inside the planning as satis ability approach (Kautz Selman 1992) Di erently from ...

Barringer, H.; Fisher, M.; Gabbay, D.; and Hunter, A. 1991. *Meta-reasoning in executable temporal logic*. In Proc. of KR-91.

## Allgemeines zur Suche (2)

- Indexierung
  - Manuell, z.B. MedLine (Fachzeitschriften) professionelle Bearbeiter, personalintensiv, hohe Qualität
  - Automatisch, z.B. Google (z.Zt.) inhaltsblind, grosses Volumen; neue Fachbeiträge werden gefunden, da meist zuerst im Web veröffentlicht
- Suchassistentz
  - Inkrementelle Suche, schrittweise Einschränkungen
  - Linguistische Dienste, Thesauri, ...
  - „Verhandlung“

# Textzugriff und Suche (lokal)

- Lineare Textsuche
- Invertierte Dateien
- Hash-Codierung
- Weitere Hilfsmittel
- Thesauri

G. Görz, FAU, Informatik 8

# Textsuche

- Mit welchem Vokabular wird der Inhalt beschrieben? Auch: Synonyme, Indexierung, Auszeichnungen?
- Welche Verknüpfungen von Suchtermen gibt es? (Boolesche, ...)
- Welche Suchverfahren?
- Suchzeit (Eigenschaften des Speichermediums, z.B. rotierender Speicher) => Minimierung der Suchzeit!

G. Görz, FAU, Informatik 8

# Lineare Textsuche

- Zeichenweises Durchsuchen: „grep“
- Flexibilisierung durch *reguläre Ausdrücke*  
<=> *endliche Automaten*
  - . beliebiges Zeichen
  - a\* beliebige Anzahl von a (inkl. Null)
  - a+ ein oder mehrere a
  - a? kein oder ein a
  - [a-d] Zeichen a, b, c, d
  - (a) Ausdruck innerhalb Klammern
  - a|b a oder b
  - Beispiele: [a-zA-Z], (rot|gelb)?

G. Görz, FAU, Informatik 8

# Lineare Textsuche (2)

- Zeit proportional zu Zeichenkettenlänge
- Parallelisierung der Suche: Problem Verarbeitung/Kommunikation
- Effiziente Suchverfahren für Zeichenketten ohne Operatoren: Boyer-Moore, ...
  - Beispiel: „Bibliothek“
  - „B“ an Pos. x, falls an Pos. x+9 kein „k“, müssen die Positionen dazwischen nicht betrachtet werden!

G. Görz, FAU, Informatik 8

## Invertierte Dateien

- Bei grossen Datenmengen ist lineare Suche ungeeignet.
- *Index*: Erzeuge Tabelle aller Wörter mit ihrer Startposition, sortiere Wörter alphabetisch
- Aufwand: Bei  $N$  Tabelleneinträgen im Mittel  $ld N$  Schritte ( $ld 10^9 = 30$  !)
- Hierarchische Organisation durch Zerlegung grosser Indexdateien in Segmente gleicher Länge und zweistufige Suche (auch hier: Zeitverhältnis Positionierung/Lesen beim Plattenspeicher)

G. Görz, FAU, Informatik 8

## Technik der invertierten Dateien

Sekundär Info	Schlüssel Gruppe		Schlüssel	Zeiger auf Sek.Info
$l_1$	$k_{11}, k_{12}, \dots$		$k_1$	$l_{11}, l_{12}, l_{13}, \dots$
$l_2$	$k_{21}, k_{22}, \dots$	$\Rightarrow$	$k_2$	$l_{21}, l_{22}, l_{23}, \dots$
:	:		:	:
$l_n$	$k_{n1}, k_{n2}, \dots$		$k_n$	$l_{k1}, l_{k2}, l_{k3}, \dots$

Volltextsuche, KWIC-Index, ...

G. Görz, FAU, Informatik 8

## Invertierte Dateien

- Nachteile: Suche mit regulären Ausdrücken nicht möglich (Ausnahme: Wortanfänge - Präfixsuche)
- Erstellung der invertierten Dateien vor der Suche; zeit- und platzaufwendig (Overhead: 25 bis 200% - Reduktion durch Auszeichnung von Stopwörtern)
- Erhöhter Platzaufwand, wenn der Kontext eines Wortes bei der Suche relevant ist
- Bemerkung: Geschwindigkeit von Suchmaschinen (z.B. Altavista) durch Index im Arbeitsspeicher!

G. Görz, FAU, Informatik 8

## Hash-Codierung (Streuspeicherung)

- Für Indexdateien: Berechnung der Speicherposition eines Wortes aus dem Wort selbst mit „Hashfunktion“
- Buchstabenhäufigkeit in natürlichen Sprachen ist keine Gleichverteilung - soll aber für die Werte der Hashfunktion (=Speicheradressen) gelten!
- Zur Wahl der Hashfunktion: Eindeutigkeit (Kollisionsfreiheit), ohne „Speicherlöcher“ zu erzeugen (modulo...)

G. Görz, FAU, Informatik 8

## Hash-Codierung (Streuspeicherung)

- Beispiel (Lesk)
  - Hashfunktion: Summe der Zahlenwerte der Buchstaben in einem Wort
  - $h(now) = 52$ ,  $h(is) = 28$ ,  $h(the) = 33$ ,  $h(time) = 47, \dots$
  - Beobachtung: Maximum 80 bei 16 Wörtern!
  - Neue Hashfunktion  $h'$ : Rest modulo 19 (Primzahl!)
  - $h'(now) = 14$ ,  $h'(is) = 9$ ,  $h'(the) = 14$ ,  $h'(time) = 9$
  - Problem: Kollision für *now* und *the* !
  - Es kann unbesetzte Zahlen in der Folge geben
  - Bessere Hashfunktionen..., Kollisionsketten

G. Görz, FAU, Informatik 8

## Weitere Hilfsmittel

- *Tries*: Buchstabenbäume (<=> invertierte Dateien)
  - leicht zu ändern (Einfügen, Löschen)
  - schnelle Suche; Präfixsuche!
  - Nachteil: Speicher-Overhead
- *Signaturdateien*: Jeder Datensatz wird durch einen kürzeren ersetzt, der den Inhalt codiert. Codierung via Hashfunktion; lineare Suche

G. Görz, FAU, Informatik 8

## Weitere Hilfsmittel

- Linguistisch unterstützte Suche
  - Robustheit (Tippfehler) durch „Phonetische Suche“: Soundex-Algorithmus (1916)  
Ähnlich klingende Wörter werden auf gleiche Zeichenketten eines reduzierten Alphabets (Vokalelimination, Konsonantenreduktion) abgebildet (Codierung: Anfangsbuchst. + 3 Konsonantencodes; {bfpv}=1, {cgjkqsxz}=2, {dt}=3, {l}=4, {mn}=5, {r}=6, {-}=0)
  - Lemmatisierung - Wortstämme und Suffixe
  - Flexionsmorphologie
  - Kombination mit Thesaurus
  - Nominalphrasen-Analyse

G. Görz, FAU, Informatik 8

## Thesauri

- Motivation aus „Information Retrieval“-Sicht:
  - Auslassungsfehler (Recall) - ein Begriff in verschiedenen Wörtern
  - Inklusionsfehler (Precision) - Ambiguität
- Lösungsansatz: Thesaurus
  - z.B. Roget (für das Englische): 1000 Begriffswörter mit Äquivalenten; Nummerierungssystem (Deutsch: Dornseiff; online: wortschatz.uni-leipzig.de)
  - WordNet
  - Fachthesauri: Medizin, Technik,...

G. Görz, FAU, Informatik 8

## Thesaurus

**Thesaurus:** Systematik von Termini (Wörtern)

ROGET: Hervorhebung von *Synonymie* vs. Begriffsumfang (enger/ weiter)

Table 2.6 Roget's system

class	section	given code			
1. abstract relations	existence	1 8	(2) communication of ideas	nature of ideas communicated	516-524
	relation	9-24		models of communication	525-549
	quantity	25-57		means of communicating ideas	550-599
	order	58-83			
	number	84-105			
	time	106-139			
	change	140-152			
	causation	153-179			
	space in general	180-191			
	dimensions	192-239			
2. space	form	240-263	(2) social volition	general social volition	737-759
	motion	264-315		special social volition	760-767
	matter in general	316-320		conditional social volition	768-774
	inorganic matter	321-356		possessive relations	775-819
3. matter	organic matter	357-449	6. emotion, religion, and morality	general	820-826
				personal emotion	827-887
4. intellect (the exercise of the mind) (1) formation of ideas	general	450-454		interpersonal emotion	888-921
	precursory conditions and operations	455-466		morality	922-975
	materials for reasoning	467-475		religion	976-1000
	reasoning processes	476-479			
	results of reasoning	480-504			
	extension of thought	505-513			
	creative thought	514-515			

G. Görz, FAU, Informatik 8

## Indexierung im Information Retrieval

- Zu entscheiden, wie gut („nahe“) ein Dokument zu einer Anfrage paßt
- Güte der Passung wird durch den Vergleich von Worthäufigkeiten geschätzt
- Ansatz von Salton (ca. 1970):
  - jedes Dokument wird durch einen multidimensionalen Vektor repräsentiert
  - die Ähnlichkeit von zwei Dokumenten wird durch den Winkel zwischen ihren Vektoren gemessen

G. Görz, FAU, Informatik 8

## „Recall“ vs. „Precision“

- **Recall:** Prozentsatz der relevanten gefundenen Texte im Vergleich mit der Gesamtzahl der enthaltenen relevanten Texte
  - z.B.: Enthalte eine DB 100 Texte, die relevant bzgl. einer bestimmten Anfrage sind. Findet die Anfrage 75 Texte, von denen 50 für den Benutzer relevant sind, ist der Recall:  $50/100 = 50\%$
- **Precision:** Prozentsatz der relevanten Texte, die im Ergebnis einer Anfrage enthalten sind
  - z.B.: Hat die Anfrage 75 Texte gefunden, von denen 50 für den Benutzer relevant sind, ist die Precision:  $50/75 = 66.6\%$

G. Görz, FAU, Informatik 8

## Automatische Indexierung

- Web-Suchdienste: Konstruktion von Indexdateien
  - „Web Crawler“ muss entscheiden, welche Seiten zu indexieren sind,
  - Duplikate eliminieren,
  - einen kurzen Index-Datensatz für jede Seite erstellen,
  - Die auf der Seite gefundenen Termini in seine invertierten Dateien integrieren
- Suche im Index
  - Suchmaschine muss Benutzeranfrage in Suchkommando konvertieren,
  - mit den invertierten Dateien vergleichen,
  - die Ergebnisse bewerten und zurückgeben

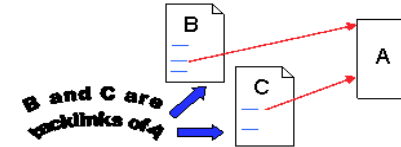
G. Görz, FAU, Informatik 8

# Suchmaschinen: Kriterien

- Dokumentenformate
- Multilingualität
- Metadaten
- Thematische Fokussierung; Thesauri
- Anfragesprache
  - reguläre Ausdrücke
  - Datenbankabfragesprache (SQL)
  - natürliche Sprache
- Bewertungsverfahren
- Ergebnisdarstellung, Markierung

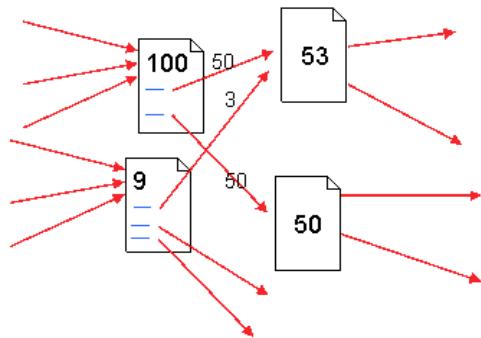
# Bewertungsverfahren: Google

## PageRank: A Citation Importance Ranking

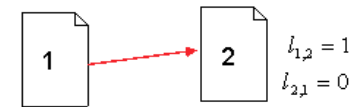


- Number of backlinks (~citations)
  - Approximation of importance/quality
  - Citation analysis literature
    - › Citation indexes
  - But, the web has extreme variation in importance/quality

## Idealized PageRank Calculation



## Idealized Model



$$n_i = \sum_{j=1, i \neq j}^N l_{i,j} \quad \text{number of outgoing links on page } i \text{ (includes multiple links to same page)}$$

$$W_j = \sum_{i=1, i \neq j}^N (l_{i,j} \frac{W_i}{n_i}) \quad \text{PageRank of page } j$$

## Actual PageRank Model

$$W_j = (1-d) + d \sum_{i=1, i \neq j}^N l_{i,j} \frac{W_i}{n_i}$$

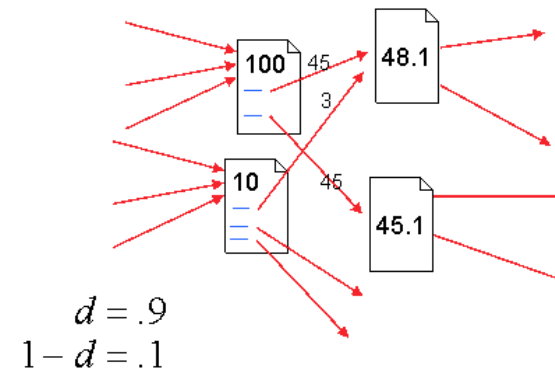
- Calculated using iterative method
- Makes graph fully connected
- $d$  parameter approximates “random surfer” path lengths

11

© Page

G. Görz, FAU, Informatik 8

## Actual PageRank Calculation



10

© Page

G. Görz, FAU, Informatik 8

## Linguistische Dienste

- Zur Erinnerung:
  - Robustheit (Tippfehler) durch „Phonetische Suche“: Soundex-Algorithmus (1916)
  - Lemmatisierung - Wortstämme und Suffixe
  - Flexionsmorphologie
  - Kombination mit Thesaurus / **Wissensrepräsentation**
  - (Nominal-) Phrasen-Analyse
  - **Multilingualität**
- Beispiele für linguistische Suchunterstützung:
  - OSIRIS: Katalogrecherche
  - SISIS-Elektra
  - IAGO (Berkeley)

G. Görz, FAU, Informatik 8


## OSIRIS

- Lexikonprozess
  - Lexika: Deutsch (400.000 Stämme), Englisch (50.000 Stämme), Namen
  - Fehlerkorrektur
- Parserprozess
  - Suffixanalyse: Lexikon Endungen Deutsch, Endungen Englisch
  - Grammatik (NP/PP): Deutsch, Englisch
  - Funktionswörter: Deutsch, Englisch
- Retrievalprozess

G. Görz, FAU, Informatik 8

Suche Literatur zum Thema:

digitale Bibliotheken

 Speicherbibliotheken - digitale Bibliotheken : Wissen verteilen und bewahren / hrsg. von Elisabeth Böllmann. - 1998

 Informatik. Wirtschaftsinformatik

 Allgemeines. Gesamtschriftum der Informatik

Serien, Hochschulschriften

 Buch- und Bibliothekswesen

 Bibliotheksverwaltung

Online-Katalog

 Bibliothekare und Bibliotheken

Einzeluntersuchungen und -probleme


 Bibliothekswesen allgemein

Einzelne Bibliothekare (hier auch personenbezogene Festschriften)

 Aufstellung alph. nach Namen

 Bibliotheksverwaltung

Internet

 Buchwesen. Buchgeschichte

Buchgattungen und -formen

 Sonstiges

Universitätsbibliothek

[Autor/Stichwort](#)

[Thema](#)

[Autoren](#)

[Zeitschriften](#)


[Neuerwerbungen](#)


[Literaturliste](#)


[Benutzerkonto](#)


Suche Literatur zum Thema:


digitale bibliotheken

 Internet in öffentlichen Bibliotheken : up (to) date! / Deutsches Bibliotheksinstitut. [Hrsg. von Marion Sommerfeld ...]. - 1999

 Document delivery: elektronisch verfügbare naturwissenschaftlich-technische Datenbanken in Europa und den USA : Darstellung und Vergleich unter Leistungs- und Kostengesichtspunkten / Steffi Engel. [Universitäts- und Landesbibliothek Sachsen-Anhalt]. - 1999

 Weiter auf dem Weg zur virtuellen Bibliothek! : Praxis, Projekte, Perspektiven / 2. InetBib-Tagung der Universitätsbibliothek Dortmund und der Fachhochschule Potsdam Fachbereich Archiv-Bibliothek-Dokumentation vom 10. - 1997

 Bibliotheken ans Netz! : öffentliche Bibliotheken als Partner moderner Informationsverbände / Deutsches Bibliotheksinstitut. [Hrsg. von: Uta Kaminsky ...]. - 1997

 Internet in öffentlichen Bibliotheken : Referate und Materialien aus einem Fortbildungsseminar des Deutschen Bibliotheksinstituts / [Seminarleitung und Red.: Marion Sommerfeld]. - 1997

...

 Internet

(Alle Buecher)

Universitätsbibliothek

[Autor/Stichwort](#)

[Thema](#)

[Autoren](#)

[Zeitschriften](#)


[Neuerwerbungen](#)


[Literaturliste](#)


[Benutzerkonto](#)


Suche Literatur zum Thema:


digitale bibliotheken

 Speicherbibliotheken - digitale Bibliotheken : Wissen verteilen und bewahren / hrsg. von Elisabeth Böllmann. - 1998

 Bibliotheksbau: Auswahldokumentation 1994/95 : Neubau, Ummutzung, Sanierung von Bibliotheksgebäuden / [Hrsg. von der Baukommission des Deutschen Bibliotheksinstituts. Bearb. von Ute Stephan unter Mitarb. von Rebekka Macaluso]. - 1997

 Ouvrages et volumes : architecture et bibliothèques / par Anne-Marie Bertrand et Anne Kupiec, avec la collab. de Joseph Belmont .... - 1997

 Bibliotheksbauten in der Praxis : Erfahrungen und Bewertungen ; [Gerhard Liebers gewidmet zur Vollendung des 80. Lebensjahres am 23. Mai 1994] / hrsg. von Roswitha Poll .... - 1994

 Bibliotheksbau : Ummutzung bestehender Gebäude für öffentliche und wissenschaftliche Bibliotheken : [die in diesem Band veröffentlichten Referate wurden im Rahmen der Fortbildungsveranstaltung "Ummutzung Bestehender Gebäude für öffentliche und wissenschaftliche Biblioth / Deutsches Bibliotheksinstitut. [Red.: Ute Kissling ; Ute Stephan]. - 1993

...

 Einzeluntersuchungen und -probleme

(Alle Buecher)

Universitätsbibliothek

[Autor/Stichwort](#)

[Thema](#)

[Autoren](#)

[Zeitschriften](#)

[Neuerwerbungen](#)

[Literaturliste](#)

[Benutzerkonto](#)

Suche Literatur zum Thema:

digitale bibliotheken

Bibliographien zum Gesamtbereich

Lexika und Nachschlagewerke

Allgemeines

Einzeluntersuchungen (Geheimschriften, Kuzschrift, u.a)

Geschichte der Schrift insgesamt

Einzelne Schriften

Lateinische Schriften

Druckschriften

Schönschreiber. Schreiblehrer. Schriftkünstler

Schreibmaterial. Schreibstoffe allgemein

Papier

Sonstiges Schreibmaterial

Schreibgeräte und Schreibmittel

Übungsbücher u.ä.

 Handschriftenkunde

 Buchwesen. Buchgeschichte

 Buchkunst. Einband. Buchpflege. Bibliophilie

 Graphisches Gewerbe. Drucktechnik. Buchbinderei

Universitätsbibliothek

[Autor/Stichwort](#)

[Thema](#)

[Autoren](#)

[Zeitschriften](#)

[Neuerwerbungen](#)

[Literaturliste](#)

[Benutzerkonto](#)

## Zu Indexierung und Suche von Multimedialen Objekten

- Hypermedia = Hypertext + Multimedia
- Bild-, Video-, Audiobibliotheken:
  - Speicherung (Komprimierung)
  - Analyse und Indexierung
  - Katalogisierung
  - Suche
  - s. Digitalisierung, Kap. 4
- hier: Anmerkungen zu Audio und Video

G. Görz, FAU, Informatik 8

## Indexierung von Bildern

- Manuelle Klassifikation; spezielle Thesauren und Metadaten-Schemata
- Bildverarbeitungsverfahren (z.B. IBM QBIC, Berkeley Blobworld)
  - Farbe (Histogramm)
  - Textur: Granularität, Kontrast, Richtung
  - Gestalt: Größe, Zirkularität, Exzentrizität, Orientierung
- => Ähnlichkeitsmaß für automatische Klassifikation

G. Görz, FAU, Informatik 8

## Audio

- Audioformate: Kompressionsproblem  
mp3: orientiert an menschlicher Hörkurve
- Automat. Indexierung (?) vs. Metadaten
- Automatische Spracherkennung:  
*Segmentierung und Klassifikation*
  - Aufnahmequalität
  - Sprecherunabhängigkeit; -adaptivität
  - gelesene vs. spontane Sprache
  - Wortschatz; „keyword spotting“
  - Kontextproblem (Semantik, Pragmatik)

G. Görz, FAU, Informatik 8

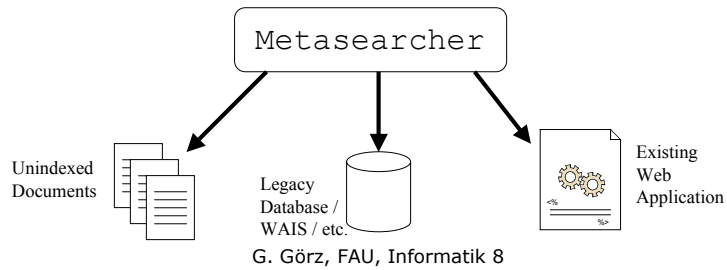
## Video: Bewegungssegmentierung

- Erkennung von Schnitten (Schwelle für Rate der Pixel-Änderung)
- (Stand-) Bildklassifikation; spezielle Erkenner: Gesichter, etc.
- gestaltorientierte Bewegungsverfolgung
- Kombination mit Audio-Klassifikation und (ggf.) Information in der „Austastlücke“ (<=> Videotext)
- Problem: sehr viele Schnitte in modernen Filmen (u.U. 4000 / 2h)

G. Görz, FAU, Informatik 8

# What is "Metasearching"?

- Given many document **sources** and a **query**, a metasearcher:
  - Finds the good sources for the query
  - Evaluates the query at these sources
  - Merges the results from these sources



© Lagoze

# Aggregators

