

Informationsintegration

Große Datenräume in Web-basierten Umgebungen

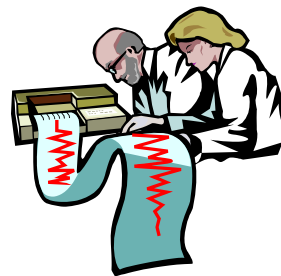
Felix Naumann

Humboldt-Universität zu Berlin

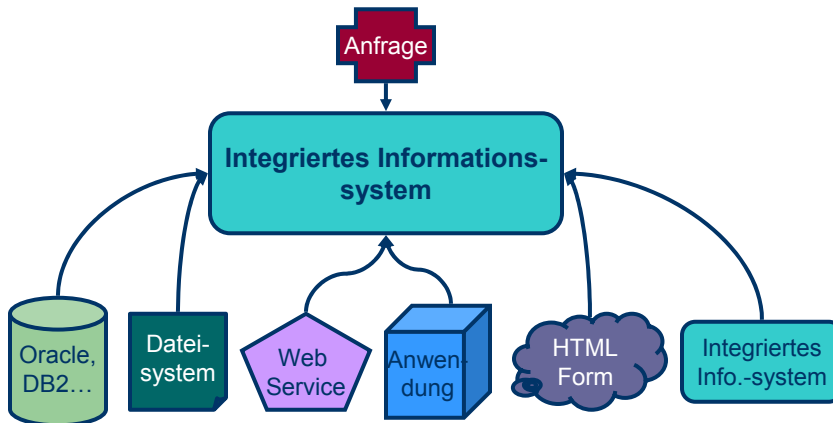
naumann@informatik.hu-berlin.de

Einige Untertitel

- Content Merging
- Objekt Fusion
- Datenintegration
- Data Amalgamation
- etwas Data Cleansing



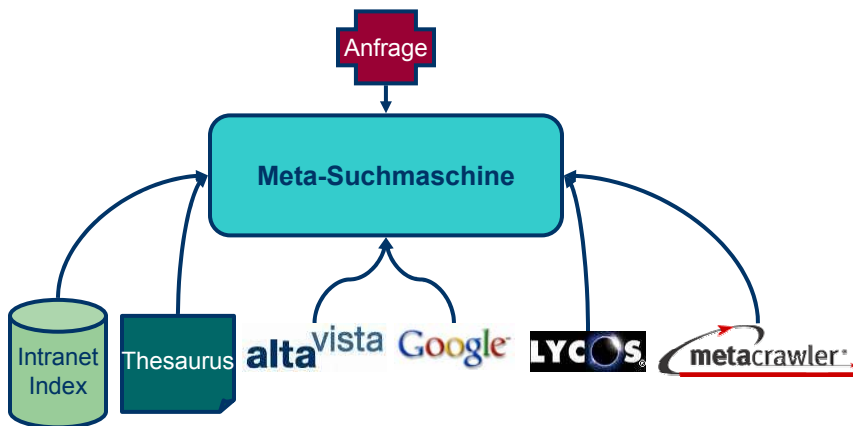
Integrierte Informationssysteme



3

Felix Naumann, HU Berlin

Integrierte Suchmaschinen

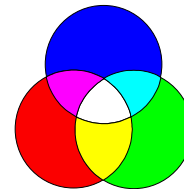


4

Felix Naumann, HU Berlin

Redundanz – pro & contra

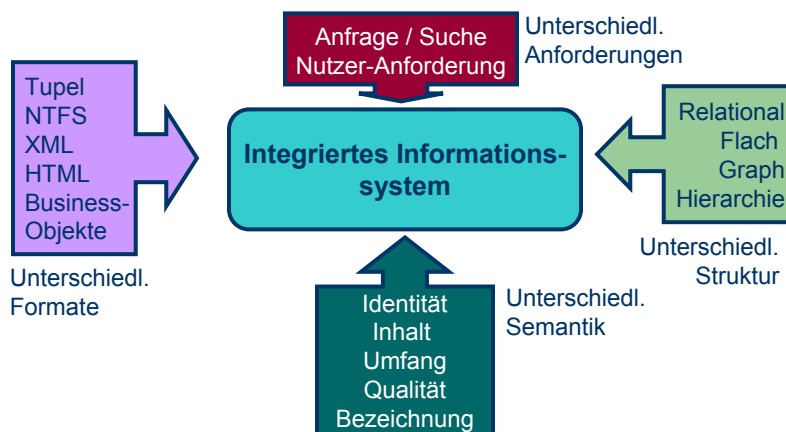
- Ist gut, denn sie bietet
 - mehr Informationen,
 - detaillierter Informationen,
 - und verifizierbare Informationen.
 - Deshalb sollten wir integrieren!
- Ist problematisch, denn
 - Redundanz herrscht nur konzeptionell.
 - Technische und strukturelle Schwierigkeiten
 - Konflikte
 - Deshalb hören Sie diesen Vortrag.



5

Felix Naumann, HU Berlin

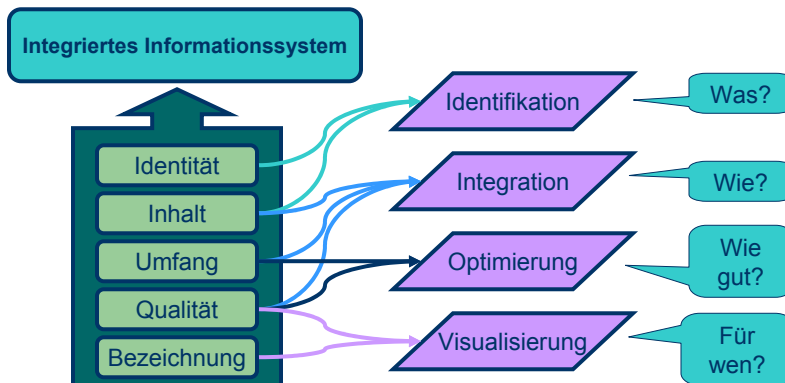
Integrierte Informationssysteme



6

Felix Naumann, HU Berlin

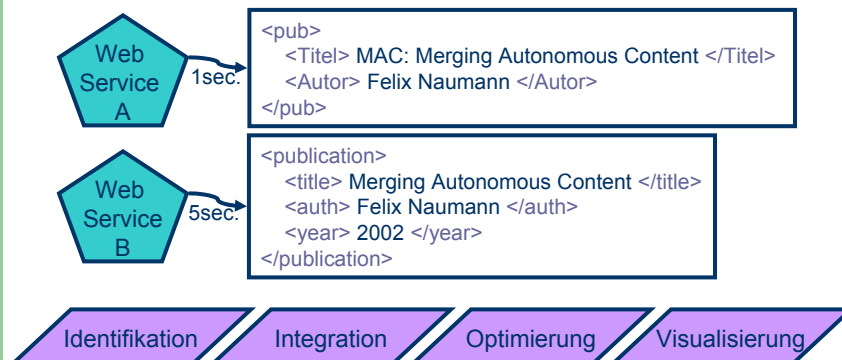
Lösung Unterschiedlicher Semantik



7

Felix Naumann, HU Berlin

Beispiel der Informationsintegration



8

Felix Naumann, HU Berlin

Objektidentifikation

Identifikation

Integration

Optimierung

Visualisierung

- Domänenspezifische Algorithmen
 - Adresdaten
 - Mikrobiologische Daten
- Manchmal gibt es IDs
 - ISBN
 - SSN
 - Personalausweisnummer
 - URL



9

Felix Naumann, HU Berlin

Adresdaten Identifizieren

Identifikation

Integration

Optimierung

Visualisierung

- Problem
 - Eine Tabelle (der Größe N), potentiell mit Duplikaten
 - Sehr große Datenmenge
 - kein quadratischer Algorithmus
 - kein Hauptspeicher-Algorithmus
- Forschung
 - Merge/Purge Technik (Hernandez, Stolfo 1998)
 - Sorted-Neighborhood Methode
 - Record Linkage (Fellegi, Sunter, 1969)
 - AssoClass (Neiling, 2003)
- Industrie
 - Trillium, Vality, ETI, et al.
 - Algorithmen sind gut gehütete Geheimnisse

10

Felix Naumann, HU Berlin

Sorted Neighborhood

Identifikation Integration Optimierung Visualisierung

- Idee
 - Daten geschickt partitionieren.
 - Nur innerhalb dieser Partitionen Duplikate suchen.
- Algorithmus
 1. Create Key:
 - Schlüssel mittels relevanter Feldern erzeugen.
 2. Sort:
 - Daten nach dem Schlüssel sortieren.
 3. Merge:
 - Fenster (der Größe w) über sortierte Tupel schieben.
 - Nur Tupel innerhalb des Fensters miteinander vergleichen.

11

Felix Naumann, HU Berlin

Sorted Neighborhood

Identifikation Integration Optimierung Visualisierung

1. Create Key

Vorname	Nachname	Adresse	ID	Schlüssel
Sal	Stolpho	123 First St.	456780	STOSAL123FRST456
Mauricio	Hernandez	321 Second Ave	123456	HERMAU321SCND123
Felix	Naumann	Hauptstr. 11	987654	NAUFEL11HPTSTR987
Sal	Stolfo	123 First Street	456789	STOSAL123FRST456

12

Felix Naumann, HU Berlin

Sorted Neighborhood

Identifikation

Integration

Optimierung

Visualisierung

2. Sort

Vorname	Nachname	Adresse	ID	Schlüssel
Mauricio	Hernandez	321 Second Ave	123456	HERMAU321SCND123
Felix	Naumann	Hauptstr. 11	987654	NAUFEL11HPTSTR987
Sal	Stolpho	123 First St.	456780	STOSAL123FRST456
Sal	Stolfo	123 First Street	456789	STOSAL123FRST456

13

Felix Naumann, HU Berlin

Sorted Neighborhood

Identifikation

Integration

Optimierung

Visualisierung

3. Merge

Vorname	Nachname	Adresse	ID	Schlüssel
Mauricio	Hernandez	321 Second Ave	123456	HERMAU321SCND123
Felix	Naumann	Hauptstr. 11	987654	NAUFEL11HPTSTR987
Sal	Stolpho	123 First St.	456780	STOSAL123FRST456
Sal	Stolfo	123 First Street	456789	STOSAL123FRST456

14

Felix Naumann, HU Berlin

Sorted Neighborhood

Identifikation Integration Optimierung Visualisierung

- Aufwand
 - Theoretisch:
 - $O(N) + O(N \log N) + O(wN) = O(N \log N)$ (bei $w < \log N$)
 - Praktisch:
 - Drei Läufe über die Daten auf der Festplatte
- Kommentare
 - Wahl des Schlüssels ist entscheidend
 - Mehrere Läufe mit unterschiedlichen Schlüsseln (und kleinem Fenster)
 - Parallelisierung durch Clustering
 - Entscheidung ob ein Duplikat vorliegt ist eine komplexe Berechnung
 - Erweiterung zum inkrementellen Algorithmus

15

Felix Naumann, HU Berlin

Objektidentifikation für geschachtelte Daten

Identifikation Integration Optimierung Visualisierung

```
- <author>
  <name>Bernd Amann</name>
- <publication>
  <title>XML Repository and Active Views Demonstration.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Active Views for Electronic Commerce.</title>
  <year>1999</year>
</publication>
</author>
- <author>
  <name>Sophie Cluet</name>
- <publication>
  <title>XML Repository and Active Views Demonstration.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Active Views for Electronic Commerce.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Views in a Large Scale XML Repository.</title>
  <year>2001</year>
</publication>
</author>
```

- Vergleiche <author>
 - Mit Subelementen (<publication>)?
 - Wie tief?
- Vergleiche <publication>
 - Mit parallelen Elementen (<year>)?
 - Schema, oder Daten?
- Kurz: Was ist ein Object?

16

Felix Naumann, HU Berlin

Objektidentifikation für verschiedene Strukturen

Identifikation Integration Optimierung Visualisierung

```

- <author>
  <name>Bernd Amann</name>
- <publication>
  <title>XML Repository and Active Views Demonstration.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Active Views for Electronic Commerce.</title>
  <year>1999</year>
</publication>
- <author>
  <name>Sophie Cluet</name>
- <publication>
  <title>XML Repository and Active Views Demonstration.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Active Views for Electronic Commerce.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Views in a Large Scale XML Repository.</title>
  <year>2001</year>
</publication>
</author>

- <inproceedings key="conf/vldb/AbiteboulAAACHHMMMSTV99">
  <author>Serge Abiteboul</author>
  <author>Vincent Ailleret</author>
  <author>Sébastien Ailleret</author>
  <author>Bernd Amann</author>
  <author>Sophie Cluet</author>
  <author>Brendan Hills</author>
  <author>Frédéric Hubert</author>
  <title>XML Repository and Active Views Demonstration.</title>
  <pages>742-745</pages>
  <year>1999</year>
  <booktitles>VLDB</booktitles>
  <url>db/conf/vldb/vldb99.html#AbiteboulAAACHHMMMSTV99.html</url>
  <crossref>conf/vldb/99</crossref>
  <ee>db/conf/vldb/AbiteboulAAACHHMMMSTV99.html</ee>
  <cdrom>VLDB99/P73.pdf</cdrom>
  <cite>conf/edbt/SantosAD94</cite>
  <cite>www/org/w3/dom</cite>
</inproceedings>
  
```

Felix Naumann, HU Berlin

Objektidentifikation mittels Struktur

Identifikation Integration Optimierung Visualisierung

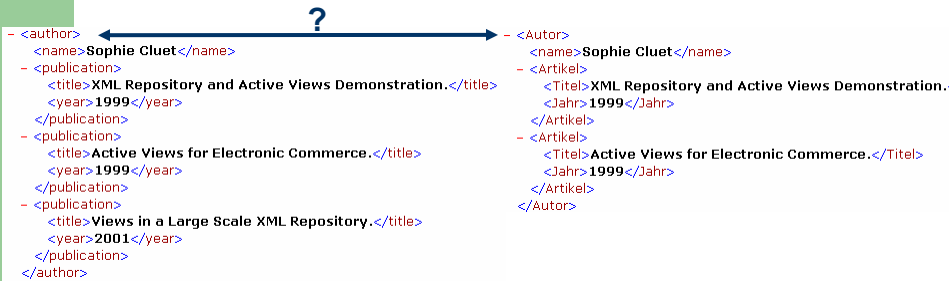
```

- <author>
  <name>Sophie Cluet</name>
- <publication>
  <title>XML Repository and Active Views Demonstration.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Active Views for Electronic Commerce.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Views in a Large Scale XML Repository.</title>
  <year>2001</year>
</publication>
</author>

- <author>
  <name>S. Kluet</name>
- <publication>
  <title>XML Repository and Active Views Demonstration.</title>
  <year>1999</year>
</publication>
- <publication>
  <title>Active Views for Electronic Commerce.</title>
  <year>1999</year>
</publication>
</author>
  
```

Umgekehrt: Strukturidentifikation mittels Objekten

Identifikation Integration Optimierung Visualisierung



19

Felix Naumann, HU Berlin

Objektidentifikation – Zusammenfassung

Identifikation Integration Optimierung Visualisierung

- Relationale Objekt Identifikation
 - Viele Algorithmen
 - No magic
- Identifikation geschachtelter Daten
 - Wahl der Granularität
 - Nutzung struktureller Information
- Ziele
 - Anpassung existierender Algorithmen
 - Entwicklung neuer Algorithmen
 - Bootstrapping mit Klassifikationstechniken
- Schnittpunkte
 - Leser: Textmining
 - Freytag: Parsergeneration

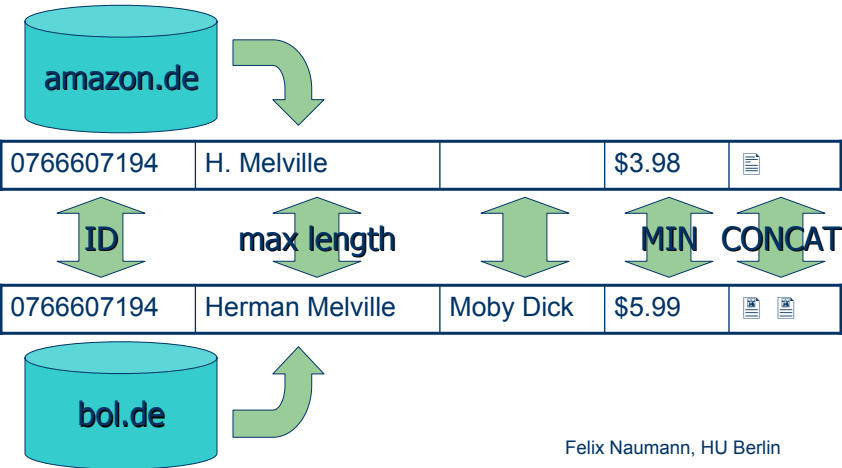
Jetzt wissen wir, WAS integriert werden soll. Bloß WIE?

20

Felix Naumann, HU Berlin

Objektintegration

Identifikation Integration Optimierung Visualisierung



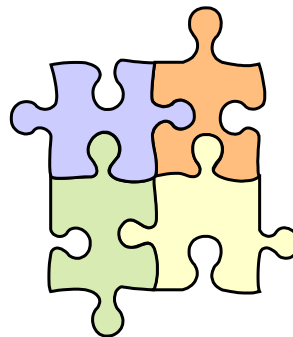
21

Felix Naumann, HU Berlin

Relationale Objektintegration

Identifikation Integration Optimierung Visualisierung

- **Union** (Vereinigung)
 - Duplikat-Eliminierung
- **Minimum union** [GL94]
 - Eliminierung sub-summierter Tupel
- **Merge union**
 - Duplikatintegration
 - Konfliktlösung



22

Felix Naumann, HU Berlin

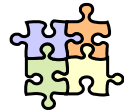
Was UNION nicht kann

Identifikation

Integration

Optimierung

Visualisierung



Quelle 1 →

a, b
a, b

⇒ a, b

Quelle 2 →

a, b
a, -

⇒ a, b
a, -

a, b
a, c

⇒ a, b
a, c

a, -, d
a, c, -

⇒ a, -, d
a, c, -

23

Felix Naumann, HU Berlin

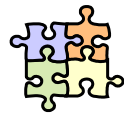
Was MINIMUM UNION nicht kann

Identifikation

Integration

Optimierung

Visualisierung



Quelle 1 →

a, b
a, b

⇒ a, b

Quelle 2 →

a, b
a, -

⇒ a, b

a, b
a, c

⇒ a, b
a, c

a, -, d
a, c, -

⇒ a, -, d
a, c, -

24

Felix Naumann, HU Berlin

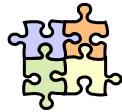
Was Duplikatintegration können sollte

Identifikation

Integration

Optimierung

Visualisierung



Quelle 1 →
Quelle 2 →

$\begin{matrix} \underline{a}, b \\ \underline{a}, b \end{matrix} \Rightarrow a, b$

$\begin{matrix} \underline{a}, b \\ \underline{a}, - \end{matrix} \Rightarrow a, b$

Bedeutung von null?

$\begin{matrix} \underline{a}, b \\ \underline{a}, c \end{matrix} \Rightarrow a, r(b,c)$

Konfliktlösungs-funktion

$\begin{matrix} \underline{a}, -, d \\ \underline{a}, c, - \end{matrix} \Rightarrow a, c, d$

25

Felix Naumann, HU Berlin

Konfliktlösung

Identifikation

Integration

Optimierung

Visualisierung

$$f(x, y) := \begin{cases} \perp & \text{if } x = \perp \text{ and } y = \perp \\ x & \text{if } y = \perp \text{ and } x \neq \perp \\ y & \text{if } x = \perp \text{ and } y \neq \perp \\ g(x, y) & \text{else} \end{cases}$$

null = unbekannt

interne Konfliktlösungsfunktion

26

Felix Naumann, HU Berlin

Konfliktlösfungsfunktionen

Identifikation

Integration

Optimierung

Visualisierung

- Numerisch:
SUM, AVG, MAX, MIN, ...
- Nicht-numerisch:
MAXLENGTH, CONCAT, AnnCONCAT, ...
- Spezialisiert:
RANDOM, COUNT, CHOOSE, FAVOR, MaxIQ, ...
- Domänen-spezifisch
...

27

Felix Naumann, HU Berlin

Implementierung von Integration

Identifikation

Integration

Optimierung

Visualisierung

- Relationales Modell
 - Relationale Algebra
 - SQL
 - C++ Programm
 - Als Erweiterung einer föderierten Datenbank
- XML Datenmodell
 - XML Query Algebra
 - XQuery
 - Java Programm

28

Felix Naumann, HU Berlin

Integration - Zusammenfassung

Identifikation Integration Optimierung Visualisierung

- “Wahre” Integration auf Datenwert-Niveau
- Vereinen ist nicht Integrieren
 - Redundanzen
 - Widersprüche
 - Komplemente
- Ziele
 - Merge-Operatoren für relationale und XML Daten
 - Konfliktlösung
 - Implementierung
- Schnittpunkte
 - Freytag: DB Theorie

**Jetzt wissen wir, WIE integriert werden soll.
Bloß WIE GUT schaffen wir das?**

29

Felix Naumann, HU Berlin

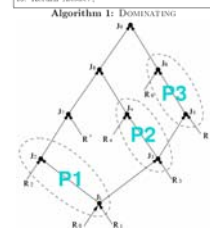
Optimierung - Ziele

Identifikation Integration Optimierung Visualisierung

1. Integrationsalgorithmen
 - Stand-alone Algorithmen zur schnellen Ausführung von Integrationsoperationen
2. Anfrageoptimierung
 - Strategien zur optimalen Ausführung komplexer Anfragen mit Integration
3. Optimierungsziel
 - Paradigmenwechsel von *Zeit* zu *Inhalt*

```

Input: Query Q, sources S = {s1, s2, ..., sn};
       costs [c1, c2, ..., cn]; limit L
Output: Result for Q
1: Answer ← {}
2: R ← S;                               [remaining sources]
3: U ← ∅;                                 [used ones]
4: while U < (L - C) and R is not empty do
5:   greedy ← greedySequenceCoverage(R);
6:   single ← singleLogicalCoverage(R);
7:   if greedy > single then
8:     Next ← maxGreedySequence(R);
9:   else
10:    Next ← maxSingle(R);
11:   R ← R - {Next};
12:   Emitting Q at Next;
13:   if Next is available then
14:     Collect result into Answer;
15:     U ← U ∪ {Next};
16: Return Answer;
    
```



30

Felix Naumann, HU Berlin

Anfrageoptimierung

Identifikation Integration Optimierung Visualisierung

- Das Problem
 - Gegeben: Anfrage
 - Gegeben: Kostenmodell
 - Gesucht: Kostenoptimaler Ausführungsplan, der die Anfrage beantwortet
- In klassischen Datenbanken
 - Anfrage = SQL
 - Kosten = Ausführungszeit
 - Anfrageplan = Baum

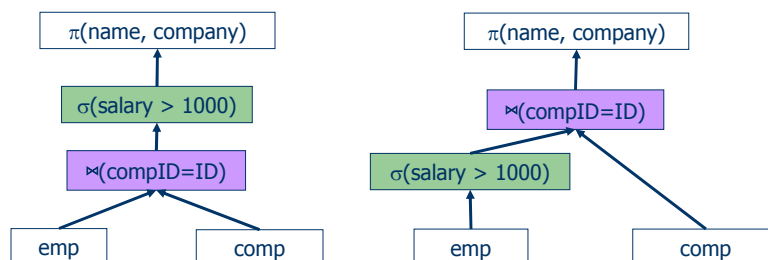
31

Felix Naumann, HU Berlin

Anfrageoptimierung - Beispiel

Identifikation Integration Optimierung Visualisierung

```
SELECT name, company FROM emp, comp
WHERE emp.compID = comp.ID AND emp.salary > 1000
```



32

Felix Naumann, HU Berlin

Anfrageoptimierung mit Merging

Identifikation

Integration

Optimierung

Visualisierung

- Integration hat Vorteile
 - Redundanz \Rightarrow Mehr Alternativen
- Integration verursacht Schwierigkeiten:
 - Pro Relation mehr als eine Quelle
 - Neuer *Merge* Operator mit unklaren Kosten
 - Gleich: Paradigmenwechsel der Optimierung
 - Jetzt: Beispiel

33

Felix Naumann, HU Berlin

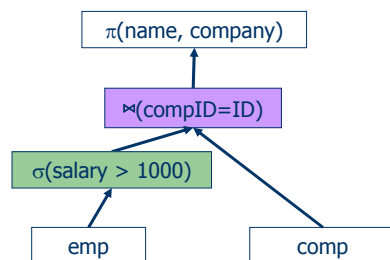
Anfrageoptimierung mit Merging

Identifikation

Integration

Optimierung

Visualisierung

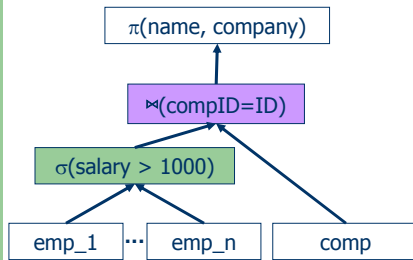


34

Felix Naumann, HU Berlin

Anfrageoptimierung mit Merging

Identifikation Integration Optimierung Visualisierung

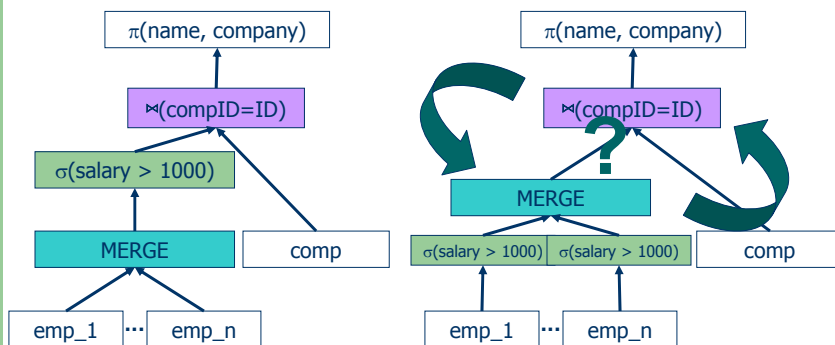


35

Felix Naumann, HU Berlin

Anfrageoptimierung mit Merging

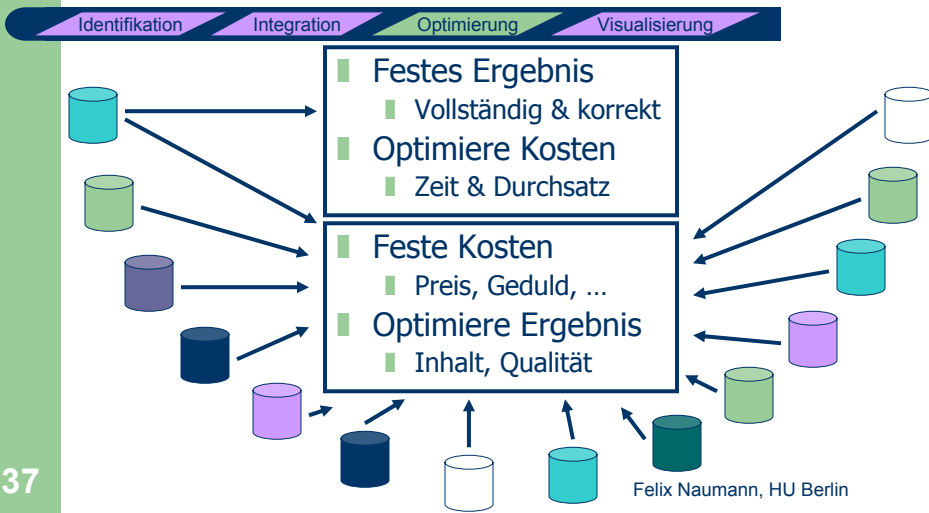
Identifikation Integration Optimierung Visualisierung



36

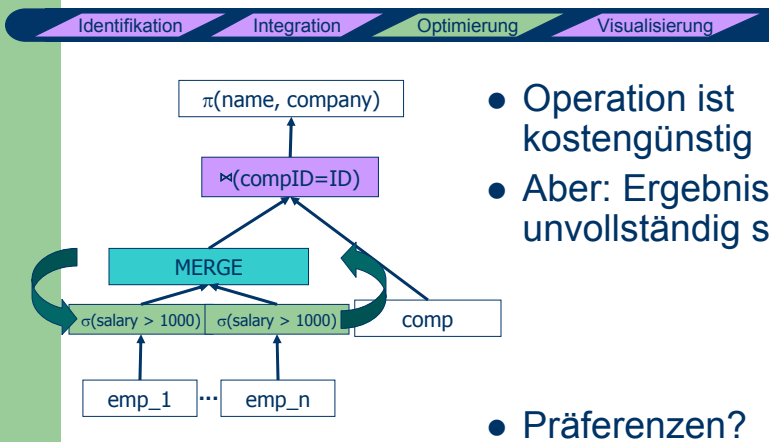
Felix Naumann, HU Berlin

Optimierung – Ein Paradigmenwechsel



37

Optimierung mit neuen Zielen



38

Felix Naumann, HU Berlin

Optimierung - Zusammenfassung

Identifikation

Integration

Optimierung

Visualisierung

- Neue Operatoren
- Paradigmenwechsel der Optimierung
 - Nutzeranforderungen
- Ziele
 - Multi-dimensionale Optimierung
 - SQL, XQuery
 - Optimierung in DBMS
- Schnittpunkte
 - Reinefeld: Parallelisierung
 - Leser: Anfrageplanung

**Jetzt wissen wir, WIE wir schnell und gut integrieren.
Aber FÜR WEN tun das?**

39

Felix Naumann, HU Berlin

Visualisierung – Ziele

Identifikation

Integration

Optimierung

Visualisierung

- Nutzerfreundliche Darstellung
- Gemischte (integrierte) Ergebnisse
- Drill-Down
- Kein kleinster gemeinsamer Nenner!

40

Felix Naumann, HU Berlin

Visualisierung – Metasuchmaschinen

Identifikation

Integration

Optimierung

Visualisierung

- MetaGer
- Mamma
- MetaCrawler
 - Google: Title, Summary, Descr., Category, URL, Size
 - Fast: Title, Summary, Descr., URL, Size
 - Inktomi: Title, Summary, URL

41

Felix Naumann, HU Berlin

The screenshot shows a Netscape browser window displaying the MetaCrawler search results for the query "humboldt-universitaet". The browser's address bar shows the search query. The page header includes the MetaCrawler logo and navigation links. The search results are listed as follows:

- Humboldt-Universität zu Berlin**
English version, **Humboldt-Universität zu Berlin**. Studium, Forschung, Angebote Zugang ... **Humboldt-Universität zu Berlin**.
<http://www.hu-berlin.de/> (Google, Fast, Inktomi, LookSmart Reviewed Sites) | [More like this](#)
- Universitätsbibliothek der HU Berlin**
Willkommen auf der Hauptseite der Universitätsbibliothek der **Humboldt-Universität zu Berlin**: Kataloge, Dienstleistungen, Webinformationen. ... <http://www.uib.hu-berlin.de/> (Google, Fast, LookSmart Reviewed Sites) | [More like this](#)
- Wirtschaftswissenschaftliche Fakultät der Humboldt-Universität Berlin (Deutschland)** <http://www.wwi.hu-berlin.de/> (Google, Fast, Inktomi) | [More like this](#)
- Landwirtschaftlich-Gärtnerische Fakultät der HU zu Berlin**
Die Landwirtschaftlich-Gärtnerische Fakultät der **Humboldt-Universität zu Berlin** benutzt für ihr Webangebot die Frametechnologie. ... <http://www.agr.hu-berlin.de/> (Google, Fast, Inktomi) | [More like this](#)
- edoc - Dokumenten- und Publikationsserver der Humboldt...**
edoc - der Dokumenten- und Publikationsserver ist ein Service für alle Angehörigen der **Humboldt-Universität zu Berlin** zum elektronischen Publizieren ihrer ... <http://dochoost.hu-berlin.de/> (Google, Fast, Inktomi) | [More like this](#)
- Humboldt-Universität zu Berlin - Juristische Fakultät**
Das Gebäude der Juristischen Fakultät, **Humboldt-Universität zu Berlin**, Welcome Studium L&F Fakultät Studenten Alumni Online Service. FEHLER. ... <http://www.rewi.hu-berlin.de/> (Google, Fast) | [More like this](#)
- Institut für Mathematik**
... Institut für Mathematik Mathematisch-Naturwissenschaftliche Fakultät II **Humboldt-Universität zu Berlin**, Uni Homepage. Allgemeines. Unser Institut. ... <http://www.mathematik.hu-berlin.de/> (Google, Fast) | [More like this](#)
- Humboldt-Universität zu Berlin, Institut für ...**
... Institut für Bibliothekswissenschaft der **Humboldt-Universität zu Berlin** Geschäftsführender Direktor Prof. Dr. Konrad Umlauf. ... <http://www.bib.hu-berlin.de/> (Google, Fast) | [More like this](#)

Metager, Suche nach: Humboldt Universität -Netscape

Metager, Suche nach: Humboldt Universität | Mamma meta search | MetaCrawler@ Results | Search Query = h...

Für detaillierte Anfragen empfehlen wir Ihnen die direkte Benutzung dieser Suchdienste.

Yahoo.de:	10 Treffer
Forschungsportal:	20 Treffer
Thema: Humboldt:	20 Treffer
Qualifunk-Tickets:	0 Treffer
Telonea:	20 Treffer
campus-research.de:	20 Treffer
Metage:	10 Treffer
Qualifunk.ch:	20 Treffer
Qualifunk.ch:	0 Treffer
7-Oni.ins:	12 Treffer
Wisch:	2 Treffer
VIPZZ:	10 Treffer
VIPZZ:	10 Treffer
Gesamtanzahl:	166 Treffer

HINWEIS: Sie haben SEHR viele Ergebnisse erhalten. Möglicherweise ist es sinnvoll:

- Ihre Suchanfrage zu verfeinern, indem Sie weitere oder speziellere/treffendere Suchworte eingeben, oder
- die Ergebnismenge zu verringern, indem Sie "Ausschlussworte" vorgeben. Klicken Sie hierzu die [MetaGer-Tips](#) an, und lesen dort ggf. die Ziffer 3. oder
- Sie fassen Ergebnisse zusammen, indem Sie auf der MetaGer-Startseite anklicken: "Ausgabe ... clustern" und ggf. zusätzlich "nur Kompakt-Darstellung ausgeben".

Enthalten die Suchergebnisse wirklich das, was Sie suchten?
-> Klicken Sie auf "QCheck" (QuickCheck, Schnellprüfung) links vor dem jeweiligen Treffer ...

Volltreffer: Humboldt Universität

[QCheck:](#)

http://www.berlinfotos.de/humboldt_universitaet.htm

- (gefunden von [Tricus](#)) **HUMBOLDT UNIVERSITÄT** Berlin Mitte Bezirk: Mitte ist wurde von 1748 bis 1766 von J. Boumann nach Entwürfen von Georg Wenzelslaus von Knobelsdorff als Palais für Prinz Heinrich erbaut ist 1810 wird dieses Palais für die von Wilhelm von Humboldt gegründete

Volltreffer: Humboldt Universität in Berlin

[QCheck:](#)

http://www.keichel.com/ausflug/berlin/humboldt_universitaet.html

- (gefunden von [Tricus](#)) Sehenswürdigkeiten in und um Berlin **HUMBOLDT UNIVERSITÄT** in Berlin Die **HUMBOLDT UNIVERSITÄT** in Berlin wurde 1748 1766 als Palais für Prinz Heinrich, einen Bruder Friedrichs des Großen gebaut. 1809 wurde das Gebäude der Universität übergeben, die 1949

Datenübertragung von switch.admt.com...

Mamma meta search -Netscape

Metager, Suche nach: Humboldt Universität | Mamma meta search | MetaCrawler@ Results | Search Query = h...

humboldt

Web News Images Audio Kids **Power Search**

Search Tips | FAQ

mamma's meta-search results - about these results

Search Results 1-15 of 72 for **humboldt** 1 2 3 4 5 (Next)

- Humboldt State University Front Page** - ... About Humboldt. Admissions. Alumni. Athletics. Calendars. Departments. Employment. ... Humboldt State University is part of the California State University system. ...
[Google, Lookmart-Directory, Teoma, MSN] <http://www.humboldt.edu>
- County of Humboldt - Site Map** ... Humboldt County Government ... About Humboldt County ... About Humboldt County. Doing business with the County. Child Support Services...
[Google, Teoma, MSN, Business.com] <http://www.co.humboldt.ca.us>
- Humboldt County Guide: HumGuide / Index** - A guide to web sites in the redwood country of Humboldt County, California, including jobs, real estate and tourism information.
[Teoma, Open-Directory, MSN] <http://www.humguide.com>
- Humboldt University of Berlin - Main website** - Links to information on study programs, research, and departments of the Humboldt Universität zu Berlin. Limited English-language information.
[Google, Open-Directory] <http://www.hu-berlin.de/index.html>
- IC Hotels Group - Holiday Inn Berlin-Humboldt Park Hotel** - Use this official reservations site to book a hotel room at the Holiday Inn in Berlin-Humboldt Park, Germany. Offers details about the Priority Club rewards program.
[Lookmart, Lookmart-Directory] http://www.ihotelsgroup.com/MSH/hd/BERWD7_JATAn+99504403 [Sponsored Source Listing]
- Humboldt, IA Hotels - Expedia.com** - Make hotel reservations in Humboldt, Iowa. Check for availability and special rates, see a list of amenities, and book a room.
[Lookmart, Lookmart-Directory] <http://click.atdmt.com/avenue/af/go/kemtep/013hotdest/181510Xhumbol> [Sponsored Source Listing]
- fellowships, awards, scholar, Alexander von Humboldt Foundation** - ... - The Alexander von Humboldt Foundation grants research fellowships and research awards to highly qualified scholars and scientists of all nationalities not ...
[Google, MSN] <http://www.avh.de/en>
- Humboldt Hotels - Hotel Locators.com** - Offers reservations at hotels in Humboldt, IA. Book online and get discounts at hotels such as Super 8.
[Lookmart, Lookmart-Directory] <http://www.hotellocators.com/offers/US/IA/Humboldt/index.htm?ref> [Sponsored Source Listing]
- Humboldt County Convention & Visitors Bureau** - Come see the world's tallest trees! Eureka! You've discovered the official website for the Humboldt County Convention & Visitors Bureau... now discover...
[Teoma, Business.com] <http://www.redwoodvisitor.org>

Datenübertragung von switch.admt.com...

HiQIQ Meta Search Engine

Identifikation

Integration

Optimierung

Visualisierung

<http://www.icdt.org/> (14 [Altavista] > [Northern Light] > [Fast] > [Hotbot] > [Google])

ICDT Home Page [Altavista]
ICDT Home Page [Northern Light]
ICDT Home Page [Fast]
ICDT Home Page [Hotbot]
ICDT Home Page [Google]

five search engines found this page

the descriptions differ

description Download your FREE evaluation copy at www.auscomp.com. . [Altavista]
Islamic Centre for Development of Trade. Islamic Centre for Development of Trade web site was built to facilitate your business and trade with OIC. Trade / [Northern Light]
BusinessOpportunities|EconomicOperators|TradeGuides|Events&Fairs|Indicators|Statistics|Publications
Islamic Centre for Development of Trade has built this web site to facilitate your business and trade with OIC member states. For Members and NFP [Fast]
BusinessOpportunities|EconomicOperators|TradeGuides|Events&Fairs|Indicators|Statistics|Publications
Islamic Centre for Development of Trade has built this web site to facilitate your business and trade with OIC member states. For Members and NFP. [Hotbot]
data conflict ~~BusinessOpportunities|EconomicOperators|TradeGuides|Events&Fairs|Indicators|Statistics|Publications~~
Islamic Centre for Development of Trade has built this web site to ... [Google]

date 07/19/1999 [Altavista] 07/19/1999 [Northern Light] 1/18/2000 [Hotbot] size 8K [Altavista] 8k [Google] language English [Altavista] category Non-profit site [Northern Light] rating 94% [Northern Light]

different attributes from different engines

45

Felix Naumann, HU Berlin

Metadaten der Integration

Identifikation

Integration

Optimierung

Visualisierung

- Herkunft
 - Data lineage
 - Durch Operatoren hindurch
- Transformationen
 - Aggregation / Integration
- Konflikte
- Datenqualität
- Ranking

46

Felix Naumann, HU Berlin

Visualisierung – Zusammenfassung

Identifikation Integration Optimierung Visualisierung

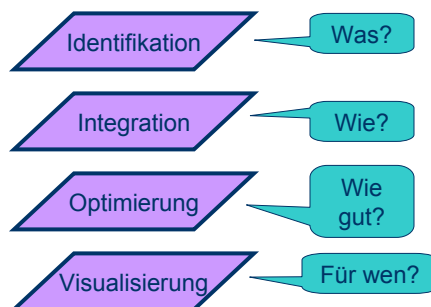
- Zeige mehr, nicht weniger
- Bsp. Metasuchmaschinen
- Metadaten
- Ziele
 - Spezifikation notwendiger Metadaten
 - Automatisierte Erfassung
 - Metadaten-Verwaltung
 - „Visualisierungskonzept“

Jetzt kennen wir alle notwendigen Komponenten zur effektiven Informationsintegration.

47

Felix Naumann, HU Berlin

Informationsintegration - Zusammenfassung



www.informatik.hu-berlin.de/macnaumann@informatik.hu-berlin.de

48

Felix Naumann, HU Berlin