

# Integrierte Suche in heterogenen digitalen Bibliotheken eines Fachgebietes mit DAFFODIL

N. Fuhr   C.-P. Klas   S. Kriewel   **A. Schaefer**

IIIS · Universität Duisburg-Essen, Campus Duisburg

InetBib 2004



# Gliederung

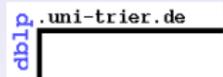
**Motivation und Konzepte**

**Verfügbare Funktionen**

**Ausblick**

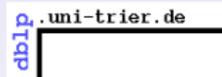
**Resümee**

# Problemstellung



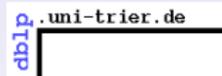
- ▶ Zur Befriedigung eines Informationsbedürfnisses: Suche in *mehreren* Digitalen Bibliotheken (DL) nötig
- ▶ Unterschiedliche
  - ▶ Funktionalitäten
  - ▶ Oberflächen
  - ▶ Anfragesprachen
- ▶ Keine strategische Unterstützung
- ▶ Keine Verknüpfungen zwischen Fakten und Daten

# Problemstellung



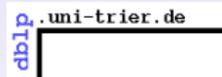
- ▶ Zur Befriedigung eines Informationsbedürfnisses: Suche in *mehreren* Digitalen Bibliotheken (DL) nötig
- ▶ Unterschiedliche
  - ▶ Funktionalitäten
  - ▶ Oberflächen
  - ▶ Anfragesprachen
- ▶ Keine strategische Unterstützung
- ▶ Keine Verknüpfungen zwischen Fakten und Daten

# Problemstellung



- ▶ Zur Befriedigung eines Informationsbedürfnisses: Suche in *mehreren* Digitalen Bibliotheken (DL) nötig
- ▶ Unterschiedliche
  - ▶ Funktionalitäten
  - ▶ Oberflächen
  - ▶ Anfragesprachen
- ▶ Keine strategische Unterstützung
- ▶ Keine Verknüpfungen zwischen Fakten und Daten

# Problemstellung



- ▶ Zur Befriedigung eines Informationsbedürfnisses: Suche in *mehreren* Digitalen Bibliotheken (DL) nötig
- ▶ Unterschiedliche
  - ▶ Funktionalitäten
  - ▶ Oberflächen
  - ▶ Anfragesprachen
- ▶ Keine strategische Unterstützung
- ▶ Keine Verknüpfungen zwischen Fakten und Daten

# DAFFODIL

## Distributed Agents For Friendly access Of Digital Libraries<sup>1</sup>

### DAFFODILS Ziele:



- ▶ vielfältige Recherchefunktionen für förderierte DL anbieten
- ▶ derzeit im Bereich **Informatik**
- ▶ leicht erweiterbar um andere Quellen einzubinden
- ▶ offenes, modulares Backend
- ▶ leicht an andere Domänen anpassbar

---

<sup>1</sup>Gefördert von der DFG im Rahmen des Schwerpunktprogrammes V3D2

# DAFFODIL

## Distributed Agents For Friendly access Of Digital Libraries<sup>2</sup>

### DAFFODILS Ziele:

- ▶ Browsing- und Suchstrategien verbinden
- ▶ Höhere Suchfunktionen anbieten
- ▶ Unterstützung von kollaborativer Recherche und Awareness
- ▶ Werkzeuge in einer benutzerfreundlichen Oberfläche integrieren
- ▶ weitgehende Drag and Drop Unterstützung
- ▶ Strategische Unterstützung



---

<sup>2</sup> Gefördert von der DFG im Rahmen des Schwerpunktprogrammes V3D2

# Verfügbare Funktionen

# Suche in Digitalen Bibliotheken

- ▶ Einheitliche  
Anfrageformulierung
- ▶ Verteilte Suche in  
heterogenen  
Digitalen Bibliotheken
- ▶ Spezifikation der  
Suchdomäne
- ▶ Ergebnisse werden  
zusammengeführt

# Suche in Digitalen Bibliotheken

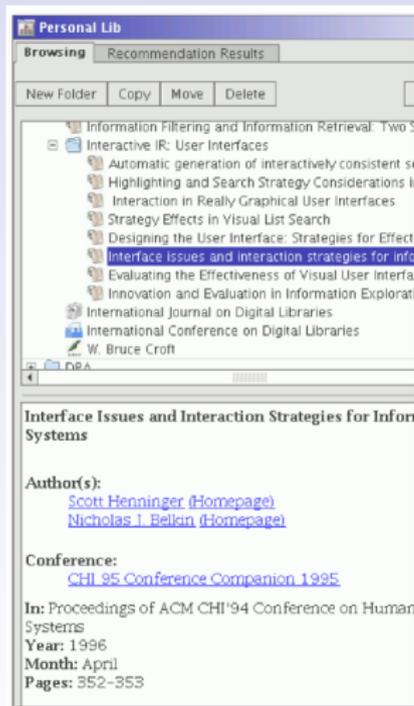
The screenshot shows a search window titled 'Search' with the query 'Title=internet AND Title=library'. It displays 50 found results, sorted by Year. The results list includes:

- 1 **Mary Minor**  
Lawfully surfing the Net: Disabling public library Internet filters to avoid more law (2004) from DBLP2.
- 2 **George DeBa; Joseph Woelfel; Eleanor Ju Rodger; Corinne Jürpensen**  
The impact of the Internet on public library use: An analysis of the current consur (2002) from DBLP2.
- 3 **Karen E. Pettigrew; Juan C. Duriance; Kenton T. Unruh**  
Facilitating community information seeking using the Internet: Findings from three (2002) from DBLP2.
- 4 **L. Adcock**  
Building a virtual music library. Towards a convergence of classification within Inte (2001) from CompuScience.
- 5 **Joyce M. Lathum**  
Positioning the Public Library in the Modern State: The Opportunity of the Children (2001) from DBLP2.
- 6 **Refnan, S. w; Obaidah, A. al-**  
Internet use and capabilities of library and information professionals at the Kuwe (2000) from CompuScience.
- 7 **John Carla Bertut**  
Public Libraries And The Internet 2000: Summary Findings And Data Tables (2000) from CiteSeer2.

At the bottom, there are buttons for 'Extract: Journals, Conferences, Authors, Keywords, MLHT' and 'Metadata: G. PDF, G. PS, G. DOC, G. PPT'. There are also 'Clear', 'Back', 'use cache', and 'Search' buttons.

- ▶ Einheitliche Anfrageformulierung
- ▶ Verteilte Suche in heterogenen Digitalen Bibliotheken
- ▶ Spezifikation der Suchdomäne
- ▶ Ergebnisse werden zusammengeführt

# Persistenz, Awareness und Annotationen



Handbibliothek ermöglicht Speichern von Objekten:

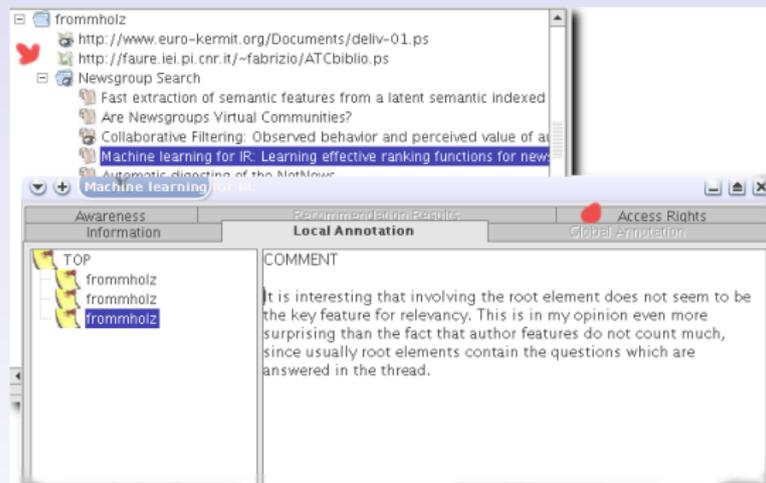
- ▶ Ergebnisse
- ▶ Anfragen
- ▶ Journale
- ▶ Suchterme
- ▶ Autoren

# Persistenz, Awareness und Annotationen

The screenshot displays a web browser interface. On the left, a file explorer shows a directory named 'frommholz' containing several files, including 'Machine learning for IR: Learning effective ranking functions for new...'. On the right, a comment window titled 'Local Annotation' is open, showing a comment text: 'It is interesting that involving the root element does not seem to be the key feature for relevancy. This is in my opinion even more surprising than the fact that author features do not count much, since usually root elements contain the questions which are answered in the thread.' A red arrow points from the comment text to the file name in the file explorer, and another red arrow points from the file name to the comment text.

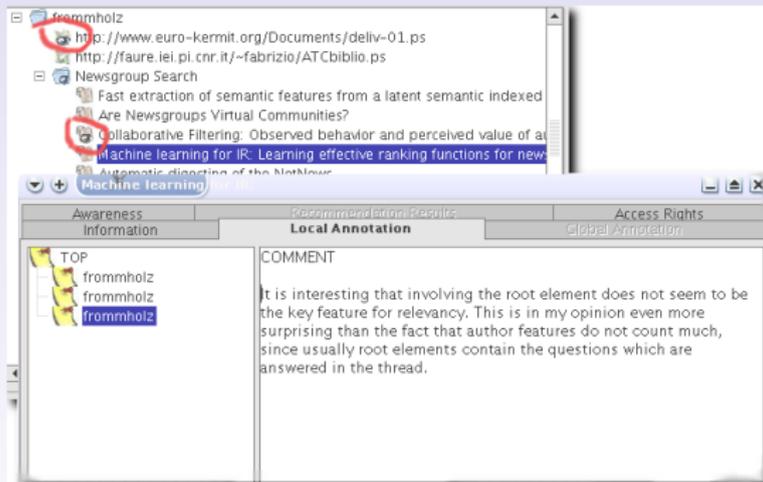
- ▶ Anmerkungen
- ▶ Diskurse
- ▶ Private Bereiche
- ▶ Gruppenordner
- ▶ Awareness auf den gespeicherten Objekten

# Persistenz, Awareness und Annotationen



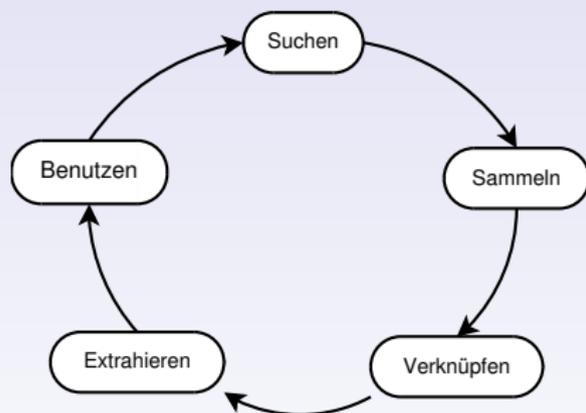
- ▶ Anmerkungen
- ▶ Diskurse
- ▶ Private Bereiche
- ▶ Gruppenordner
- ▶ Awareness auf den gespeicherten Objekten

# Persistenz, Awareness und Annotationen



- ▶ Anmerkungen
- ▶ Diskurse
- ▶ Private Bereiche
- ▶ Gruppenordner
- ▶ Awareness auf den gespeicherten Objekten

# Strategische Unterstützung



- ▶ Unterstützen des wissenschaftlichen Workflow Zyklus
- ▶ Wichtige Arbeitsschritte unterstützen
- ▶ Hochwertige Suchfunktionen
- ▶ Verknüpfung der Daten
- ▶ Persistente und zentrale Speicherung
- ▶ Einbettung in den Arbeitskontext

# Suchen nach Zitationen und Referenzen

## Referenzenwerkzeug

**References**

**Document**

- **Fuhr, Norbert**  
A decision-theoretic approach to database selection i  
(1999) from CiteSeer; Achilles; DBLP

Found: 14    << < > >>    Sort by: **Author**

**References From**

- **Abraham Bookstein.**  
Outline of a general probabilistic retrieval model  
from CiteSeer
- **Atkins, D.E., Birmingham, W.P., Durfee, E.H., Glover, E.J.,**  
Toward Inquiry-Based Education Through Interacting  
from CiteSeer
- **BOWMAN, C.M., DANZIG, P.B., MANBER, U., SCHWARTZ, A.**  
Harvest: A scalable  
from CiteSeer
- **Calkan, J.; Lu, Z.; Croft, W.**  
Searching Distributed Collections with Interference

- ▶ Welche Dokumente werden von einem Dokument zitiert?
- ▶ Wo wird das Dokument selbst referenziert?
- ▶ Aufruf durch Drag & Drop von Dokumentobjekten
- ▶ Ergebnisse weiterverwendbar

# Suchen nach Zitationen und Referenzen

## Referenzenwerkzeug

References

**Document**

- **Fuhr, Norbert**  
A decision-theoretic approach to database selection (1999) from CiteSeer; Achilles; DBLP

Found: 14    << < > >>    Sort by:

**References From**

- **Abraham Bookstein.**  
Outline of a general probabilistic retrieval model from CiteSeer
- **Atkins, D.E., Birmingham, W.P., Durfee, E.H., Glover, E.J., ...**  
Toward Inquiry-Based Education Through Interacting from CiteSeer
- **BOWMAN, C.M., DANZIG, P.B., MANBER, U., SCHWARTZ, A.**  
Harvest: A scalable from CiteSeer
- **Calkan, J.; Lu, Z.; Croft, W.**  
Searching Distributed Collections with Interference

- ▶ Welche Dokumente werden von einem Dokument zitiert?
- ▶ Wo wird das Dokument selbst referenziert?
- ▶ Aufruf durch Drag & Drop von Dokumentobjekten
- ▶ Ergebnisse weiterverwendbar

# Suchen nach Zitationen und Referenzen

## Referenzenwerkzeug

**References**

**Document**

- **Fuhr, Norbert**  
A decision-theoretic approach to database selection (1999) from CiteSeer; Achilles; DBLP

Found: 14    << < > >>    Sort by: **Author**

**References From**

- **Abraham Bookstein.**  
Outline of a general probabilistic retrieval model from CiteSeer
- **Atkins, D.E., Birmingham, W.P., Durfee, E.H., Glover, E.J., ...**  
Toward Inquiry-Based Education Through Interacting from CiteSeer
- **BOWMAN, C.M., DANZIG, P.B., MANBER, U., SCHWARTZ, A.**  
Harvest: A scalable from CiteSeer
- **Calkan, J.; Lu, Z.; Croft, W.**  
Searching Distributed Collections with Interference

- ▶ Welche Dokumente werden von einem Dokument zitiert?
- ▶ Wo wird das Dokument selbst referenziert?
- ▶ Aufruf durch Drag & Drop von Dokumentobjekten
- ▶ Ergebnisse weiterverwendbar

# Blättern in Journalen

Journal Browser

Give a term to filter journal titles:

digital

Found:5

- 1 *Innes Ferguson; Edmund Durfee*  
Artificial intelligence in digital libraries: moving from chaos to (r
- 2 *Peter Weinstein; William Birmingham*  
Creating ontological metadata for digital library content and se
- 3 *Donato Malerba; Floriana Esposito; Giovanni Semeraro; et al.*  
Adding machine learning and knowledge intensive techniques
- 4 *Atsuhiko Takasu; Norio Katayama; Masanori Sugimoto*  
A system for constructing private digital libraries through infor
- 5 *Stuart Soltysiak; I. Crabtree*  
Identifying and tracking changing interests

Extract: Journals Conferences Authors Keywords

# Autorenbeziehungen

**Author(s)**  
Catherine Plaisant

Depth 2

Sort by centrality

OK

**Persons**

- Plaisant, Catherine
- Shneiderman, Ben
- Furuta, Richard
- Bederson, Benjamin
- Marchionini, Gary
- Chalmers, Matthew
- Rose, Anne
- Rao, Ramana
- Greene, Stephan
- Grosjean, Jesse

**Plaisant, Catherine**

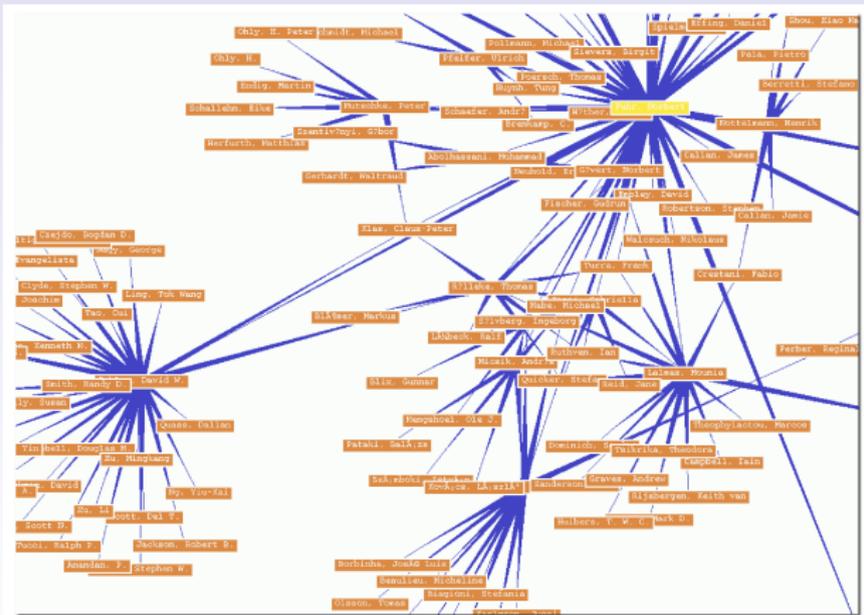
Co-Authors:

A [Ahlberg, Christopher](#) |  
 B [Barth, Randy](#) | [Bederson, Benjamin](#)  
[Rodrigo, A.](#) | [Brethauer, Dorothy](#) | [Br](#)  
 C [Campbell, Laura](#) | [Carr, David](#) | [Cl](#)  
[Matthew](#) |  
 D [Doan, Khoa](#) |  
 E [Ellis, Jason B.](#) |  
 F [Faloutsos, Christos](#) | [Fekete,](#)  
[Jean-Daniel](#) | [Fredrikson, Anna](#) | [Fu](#)  
 G [Greene, Stephan](#) | [Grosjean, Jesse](#)  
 H [Harberts, Bob](#) | [Hasco&et-Zizi,](#)  
[Mountaz](#) | [Hasegawa, Hiroaki](#) | [Hop](#)  
[Don](#) | [Horn&A;k, Kasper](#) |  
 J [Johns, Steve](#) |  
 K [Komlodi, Anita](#) | [Korthage, Robert](#)  
[Harsha P.](#) |  
 J Lee, Raymond | Li, Ha | Lindsey, F

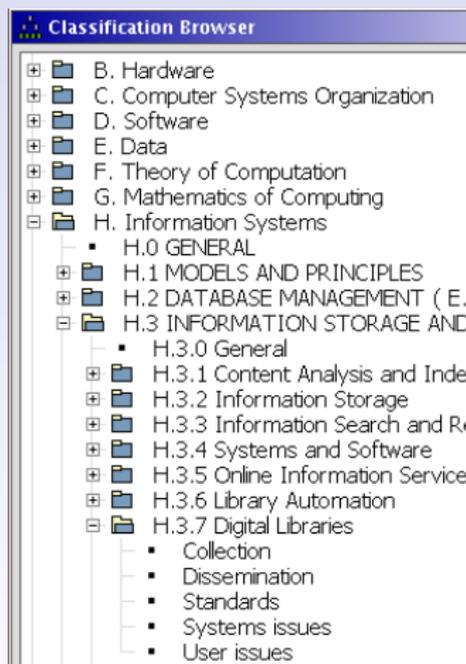
- ▶ Ausnutzung der Koautoren-Beziehungen
- ▶ Aufbau eines Beziehungsnetzes
- ▶ Browsen innerhalb des Netzwerkes

# Koautorennetz

## Visualisierung in einem Graphen



# Klassifikationen, Thesauri



## ► Browsing in Klassifikationen Repräsentation

- hierarchisch
- themenorientiert

## ► Thesaurus

- Nachschlagen verwandter Begriffe
- Hypernyme und Hyponyme, semantische Bedeutungen
- nutzt fachspezifische und webbasierte Thesauri

# Klassifikationen, Thesauri

Give a term to look up:	
library	
Broader Terms	Terms
special libraries	academic libraries
	digital libraries
	libraries
	library automation
	library mechanisation
	public libraries
	research libraries
	software libraries
	special libraries

## academic libraries

### Senses by Wordnet:

1. academic: an educator who works university
2. academic: associated with academi  
"the academic curriculum"; "acade
3. academic: hypothetical or theoretic  
expected to produce an immediate

## ► Browsing in Klassifikationen Repräsentation

- hierarchisch
- themenorientiert

## ► Thesaurus

- Nachschlagen verwandter Begriffe
- Hypernyme und Hyponyme, semantische Bedeutungen
- nutzt fachspezifische und webbasierte Thesauri

# Überblick

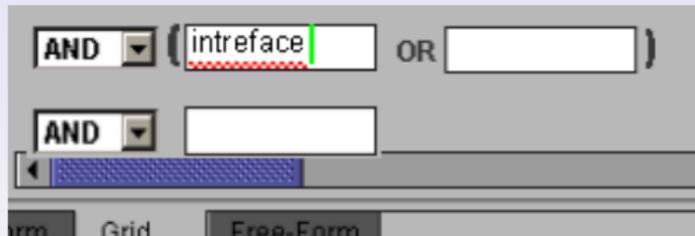
The screenshot displays a complex research information system interface with several key components:

- Search Panel (Top Left):** Includes a list of filters (ACM, Pub, GetInfo, Leabib, CompuScience, Arxiv, Springer, BiOCL) and a search form with fields for Author (R. B. Allen), Title (User models), Year, and Free-Text. A query construction section below explains the use of operators like AND, OR, and NOT.
- Personal Lib (Top Middle):** A file manager view showing a tree structure of folders and files, including 'Information Retrieval Interaction' and 'Task complexity'.
- Classification Browser (Top Right):** A hierarchical tree of classification codes, such as 'H.0 GENERAL', 'H.1 MODELS AND PRINCIPLES', and 'H.5.2 User Interfaces (D.2.2, H.5.2)'. A 'Ternsaurus Browser' is also visible below it.
- Author Networks (Bottom Middle):** A network graph showing connections between authors. The central node is 'Allen, Robert B.', with other nodes like 'Newell, Gregory S.', 'Shneiderman, Stephen', and 'Gupta, Bhupat' connected to it. The graph is sorted by centrality.
- Author List (Bottom Left):** A list of authors with search and drag capabilities.
- Graphical User Interfaces (Bottom Right):** A text box titled 'graphical user interfaces' containing a definition: '1. graphical user interfaces: a user based on graphics (icons and pictures) instead of text; uses a mouse as a keyboard as an input device.'

# Kommende Funktionen

# Aktive Unterstützung

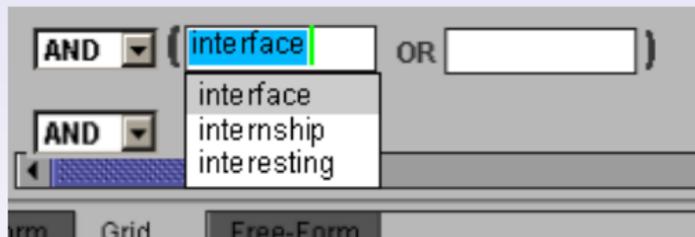
## Beispiel Anfrageformulierung



- ▶ Rechtschreibfehler bei der Eingabe von Termen
- ▶ Fehler wird angezeigt
- ▶ Liste mit möglichen Alternativen vorgelegt

# Aktive Unterstützung

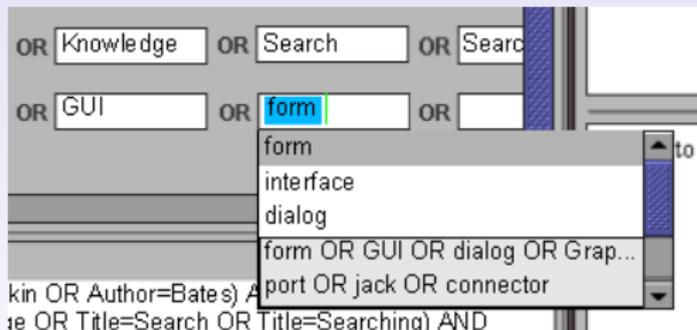
## Beispiel Anfrageformulierung



- ▶ Rechtschreibfehler bei der Eingabe von Termen
- ▶ Fehler wird angezeigt
- ▶ Liste mit möglichen Alternativen vorgelegt

# Aktive Unterstützung

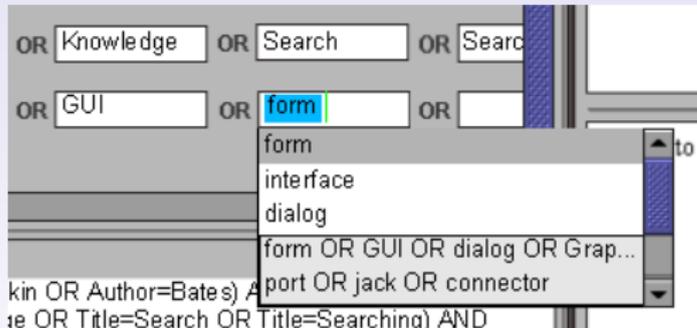
## Terme zu Anfrageerweiterung



- ▶ Beim Suchen ist der Benutzer unsicher (ASK-Hypothese)
- ▶ Synonyme und Homographe machen Sprache mehrdeutig
- ▶ DAFFODIL legt Listen mit synonymen Begriffen vor, um die Anfrage zu erweitern

# Aktive Unterstützung

## Terme zu Anfrageerweiterung



- ▶ Beim Suchen ist der Benutzer unsicher (ASK-Hypothese)
- ▶ Synonyme und Homographe machen Sprache mehrdeutig
- ▶ DAFFODIL legt Listen mit synonymen Begriffen vor, um die Anfrage zu erweitern

# Aktive Unterstützung

## Anfragen ohne Treffer

The screenshot shows a search interface with the following elements:

- Form** section:
  - Author:  OR
  - AND  OR
  - AND
  - AND
  - Title:
- Query** section:
  - Form | Grid | Free-Form
  - Author=Salton AND Author=Knuth
  - Clear
  - Estimated number of results: 0
- A yellow tooltip box states: "No common publication of these authors is known in DaFodil so you might end up without any results."
- A red triangle icon points to the "Estimated number of results: 0" text.
- A toolbar at the bottom contains various icons for navigation and search.

- ▶ Anfragen ohne Resultate: Over Constrained
- ▶ Solche Queries zu verarbeiten kostet wertvolle Wartezeit
- ▶ Verdacht auf resultatlose Anfrage wird angezeigt

# Aktive Unterstützung

## Anfragen ohne Treffer

Author: ( Salton OR )

AND Knuth OR )

AND

AND

Title:

Form Grid Free-Form

Author=Salton AND Author=Knuth

No common publication of these authors is known in Daffodil so you might end up without any results.

Clear Search

Estimated number of results: 0

- ▶ Anfragen ohne Resultate: Over Constrained
- ▶ Solche Queries zu verarbeiten kostet wertvolle Wartezeit
- ▶ Verdacht auf resultatlose Anfrage wird angezeigt

# Aktive Unterstützung

## Veröffentlichungszeiträume

The screenshot shows a search form with a vertical label 'Form' on the left. The form contains the following elements:

- A 'Year:' label followed by a text input field containing '1987-1994', an 'OR' operator, and an empty text input field.
- An 'AND' dropdown menu followed by an empty text input field.
- The word 'AND' centered below the previous row.
- An 'Author:' label followed by a text input field containing 'Fuhr', an 'OR' operator, and an empty text input field.
- An 'AND' dropdown menu followed by an empty text input field.

- ▶ Suche nach Publikationen bestimmter Autoren
- ▶ zeigt wahrscheinlichen Veröffentlichungszeitraum an
- ▶ Benutzer kann den Autor nun besser zeitlich einordnen

# Aktive Unterstützung

## Veröffentlichungszeiträume

The screenshot shows a search form with a vertical label 'Form' on the left. It contains two main search sections. The first section is for 'Year', with a text input field containing '1987-1994' and an 'OR' operator followed by an empty text input field. Below this is an 'AND' operator in a dropdown menu followed by an empty text input field. The second section is for 'Author', with a text input field containing 'Fuhr' and an 'OR' operator followed by an empty text input field. Below this is another 'AND' operator in a dropdown menu followed by an empty text input field.

- ▶ Suche nach Publikationen bestimmter Autoren
- ▶ zeigt wahrscheinlichen Veröffentlichungszeitraum an
- ▶ Benutzer kann den Autor nun besser zeitlich einordnen

# Aktive Unterstützung

## Anfrage Historie

Author: (Sal OR )

AND

AND

Title:

AND

Free-Text:

not

Mon, 28 Oct

**Used terms:**  
Salton Belkin Bates "Information Retrieval"  
Knowledge Search Searching Interface form dialog

**Found documents included:**  
Salton, G.: "Automatic information retrieval"  
Bates, M.: "A Search Interface for Public Libraries"

**Actions in the context of this search:**  
Before: Move documents in the PLib  
After: Author Network for Fuhr, N.

- ▶ Benutzer möchte alte Suche wiederholen oder variieren
- ▶ Liste alter Queries wird vorgelegt
- ▶ Liste ist mit Kontext angereichert um das Wiederfinden zu erleichtern

# Resümee

- ▶ Modularer Client
- ▶ Prototypische Realisierung leicht übertragbar
- ▶ Fachgebietsunabhängig
- ▶ Netzwerk von Diensten, Wrappern und Services
- ▶ Integriertes Suchen und Browsen
- ▶ Unterstützung von Taktiken und Suchstrategien
- ▶ Durchgehende Kombinierbarkeit von Werkzeugen
- ▶ Personalisierung und Gruppenunterstützung

## Resümee

- ▶ Modularer Client
- ▶ Prototypische Realisierung leicht übertragbar
- ▶ Fachgebietsunabhängig
- ▶ Netzwerk von Diensten, Wrappern und Services
- ▶ Integriertes Suchen und Browsen
- ▶ Unterstützung von Taktiken und Suchstrategien
- ▶ Durchgehende Kombinierbarkeit von Werkzeugen
- ▶ Personalisierung und Gruppenunterstützung

# Resümee

- ▶ Modularer Client
- ▶ Prototypische Realisierung leicht übertragbar
- ▶ Fachgebietsunabhängig
- ▶ Netzwerk von Diensten, Wrappern und Services
- ▶ Integriertes Suchen und Browsen
- ▶ Unterstützung von Taktiken und Suchstrategien
- ▶ Durchgehende Kombinierbarkeit von Werkzeugen
- ▶ Personalisierung und Gruppenunterstützung

# Resümee

- ▶ Flexibel und leicht erweiterbar
- ▶ laufendes System
- ▶ evaluiert
- ▶ über 100 regelmäßige Nutzer
- ▶ über das Stadium des reinen Forschungsprototypen hinaus

Probieren Sie DAFFODIL aus:

`http://www.daffodil.de`



The screenshot shows a web browser window titled "Daffodil". The header area is grey and contains the text "UNIVERSITÄT D U I S B U R G E S S E N". Below the header, there are two logos: on the left, the logo for "IZ Sozialwissenschaften, Bonn" (a stylized 'I' and 'Z' in red and black), and on the right, the logo for "Uni-Dortmund" (a blue and white grid pattern). Below the logos are two input fields: "Login Name" with the text "kriewel" and "Password" with "\*\*\*\*\*". At the bottom of the form are two buttons: "login" and "cancel".