DAFFODIL
Grundlegende Konzepte
&
die Wrapper

Distributed Agents for User-Friendly Access of Digital Libraries
Claus-Peter Klas
Prof. Norbert Fuhr
Universität Duisburg-Essen

Motivation

- Was ist DAFFODIL ?
- Grundlegende Konzepte
- Wrapperarchitektur
  - Wrapper-Toolkit
- Demonstration von DAFFODIL

Übergeordnetes Ziel

- *System für benutzerorientiertes, effizientes und effektives Information Retrieval durch*
  - *Strategische Unterstützung,*
  - *Personalisierung und ein*
  - *Aktives System*

  zur wissenschaftlichen Literatursuche in Digitalen Bibliotheken.

Bates Matrix

strategische Unterstützung

Proaktiv / Adaptiv

	Basisaktion	Taktik	Strategem	Strategie
Keine Systemunterstützung	●	●	●	●
Aktionen anzeigen	●	●	●	?
Aktionen auf Anfrage ausführen	●	●	●	?
Beobachten & Vorschlagen	?	●	●	?
Automatische Ausführung	?	?	?	?

Ziele – Konzepte – Wrapper

The Collection of

Computer Science Bibliographies

dblp.uni-trier.de

Classification

Search

Browse

Strategic support

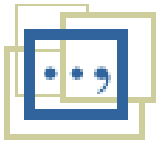
DAFFODIL

Citation

Co-Author

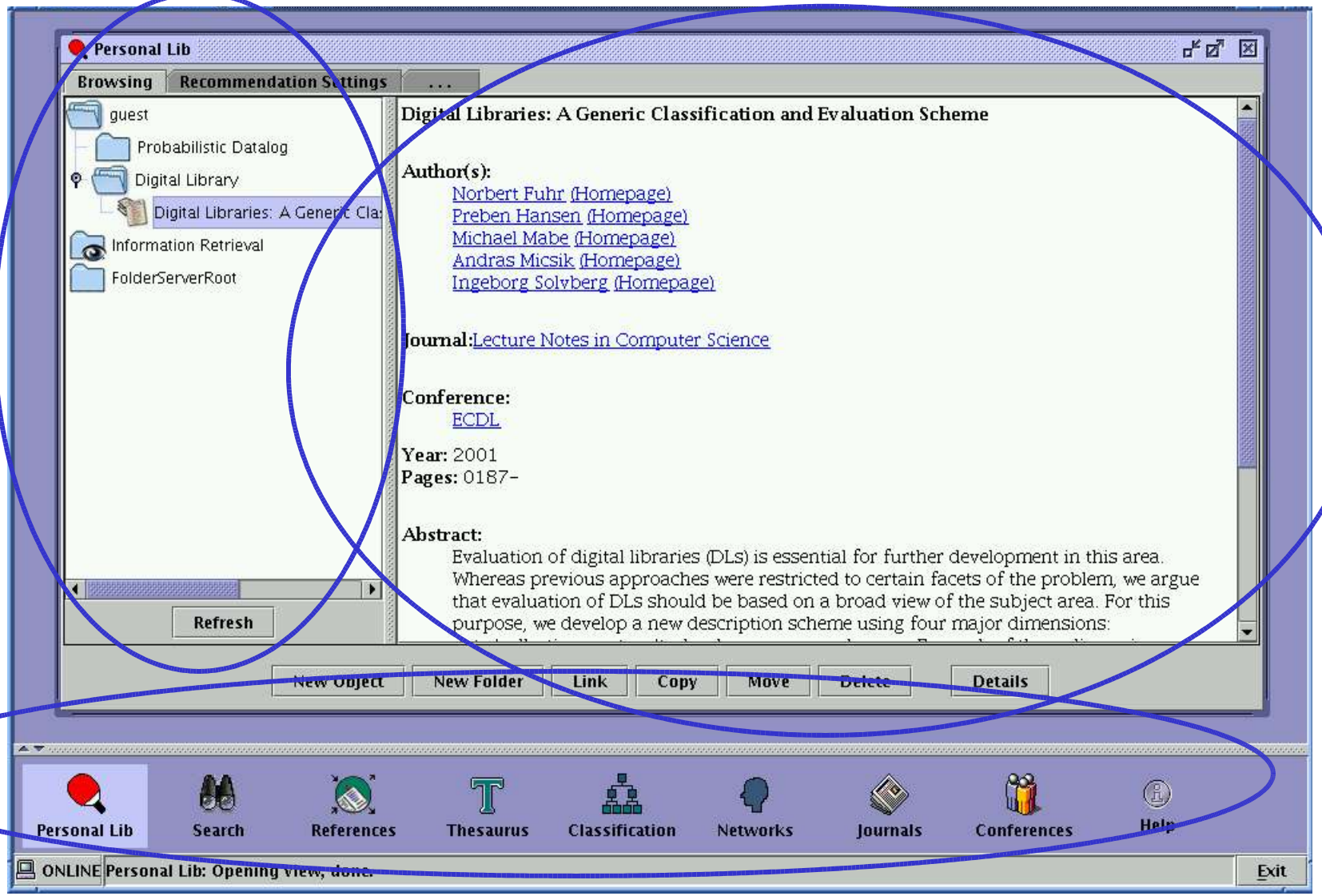
Thesaurus

HCIBIB



DIE
DIGITALE
BIBLIOTHEK

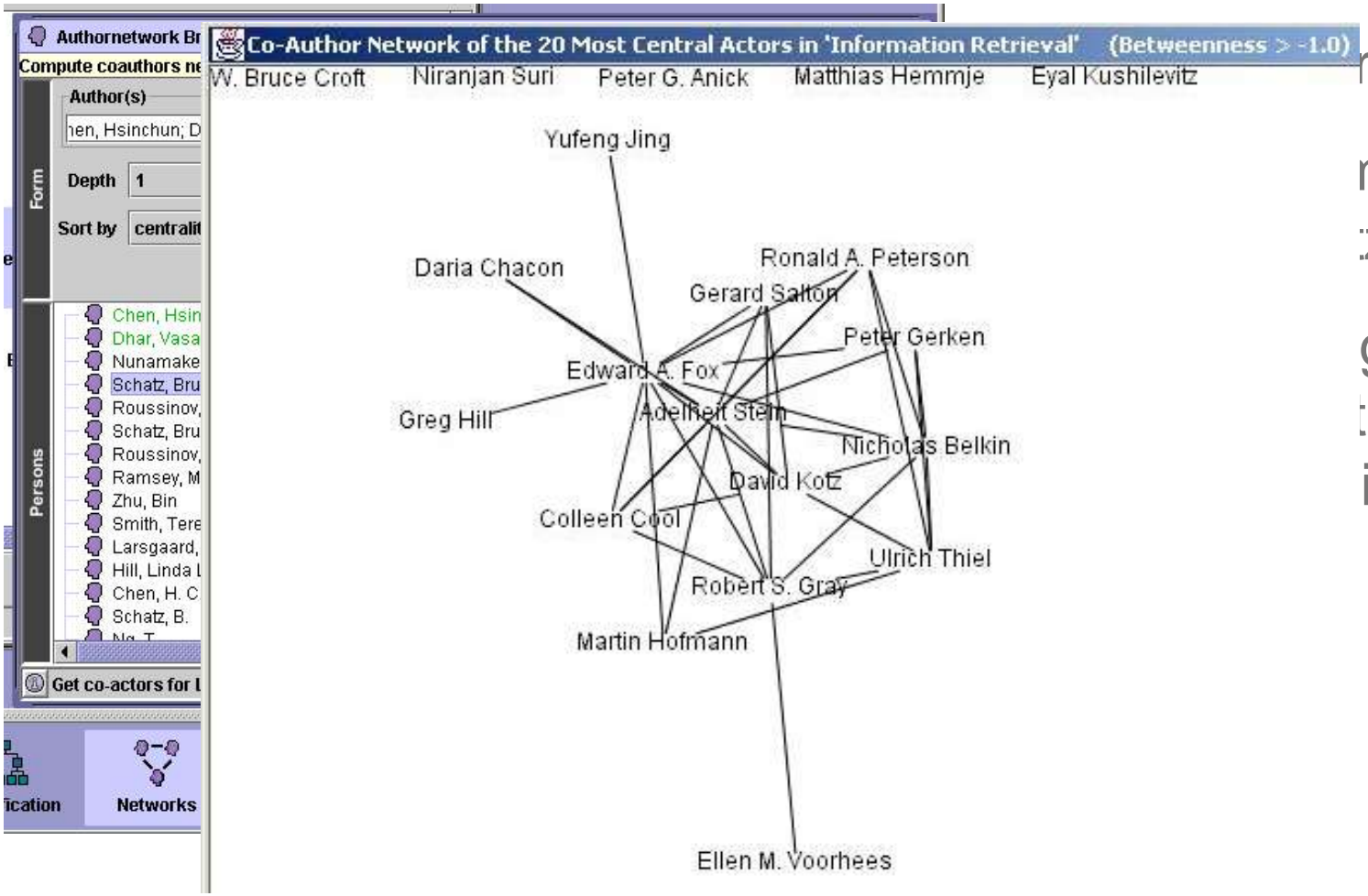
CiteSeer
Scientific Literature Digital Library


Personal Lib
Browsing
Recommendation Settings
...
guest
Probabilistic Datalog
Digital Library
Digital Libraries: A Generic Cla
Information Retrieval
FolderServerRoot
Refresh
Digital Libraries: A Generic Classification and Evaluation Scheme
Author(s):
Norbert Fuhr (Homepage)
Preben Hansen (Homepage)
Michael Mabe (Homepage)
Andras Micsik (Homepage)
Ingeborg Solvberg (Homepage)
Journal:Lecture Notes in Computer Science
Conference:
ECDL
Year: 2001
Pages: 0187-
Abstract:
Evaluation of digital libraries (DLs) is essential for further development in this area. Whereas previous approaches were restricted to certain facets of the problem, we argue that evaluation of DLs should be based on a broad view of the subject area. For this purpose, we develop a new description scheme using four major dimensions:
New Object
New Folder
Link
Copy
Move
Delete
Details
Personal Lib
Search
References
Thesaurus
Classification
Networks
Journals
Conferences
Help
ONLINE
Personal Lib: Opening view, done.
Exit

Strategische Unterstützung im Informationssuchprozess

- Citation Search
- Cross Reference Linking
- Author network based stratagems
- Journal/Conference Run

Author Networks



rke:

n

zwerk

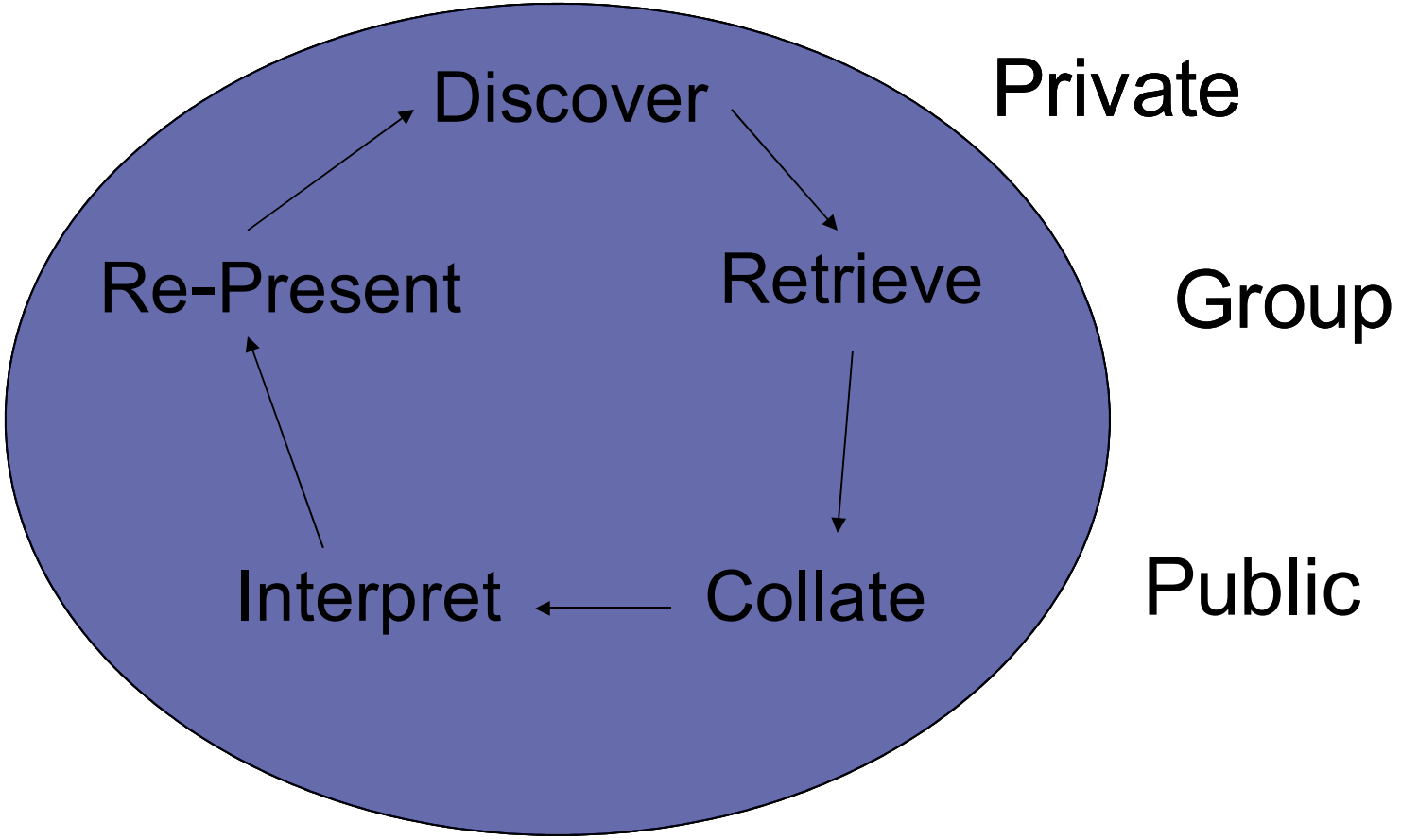
g von

en nach

ität von

Ziele – ***Konzepte*** – Wrapper

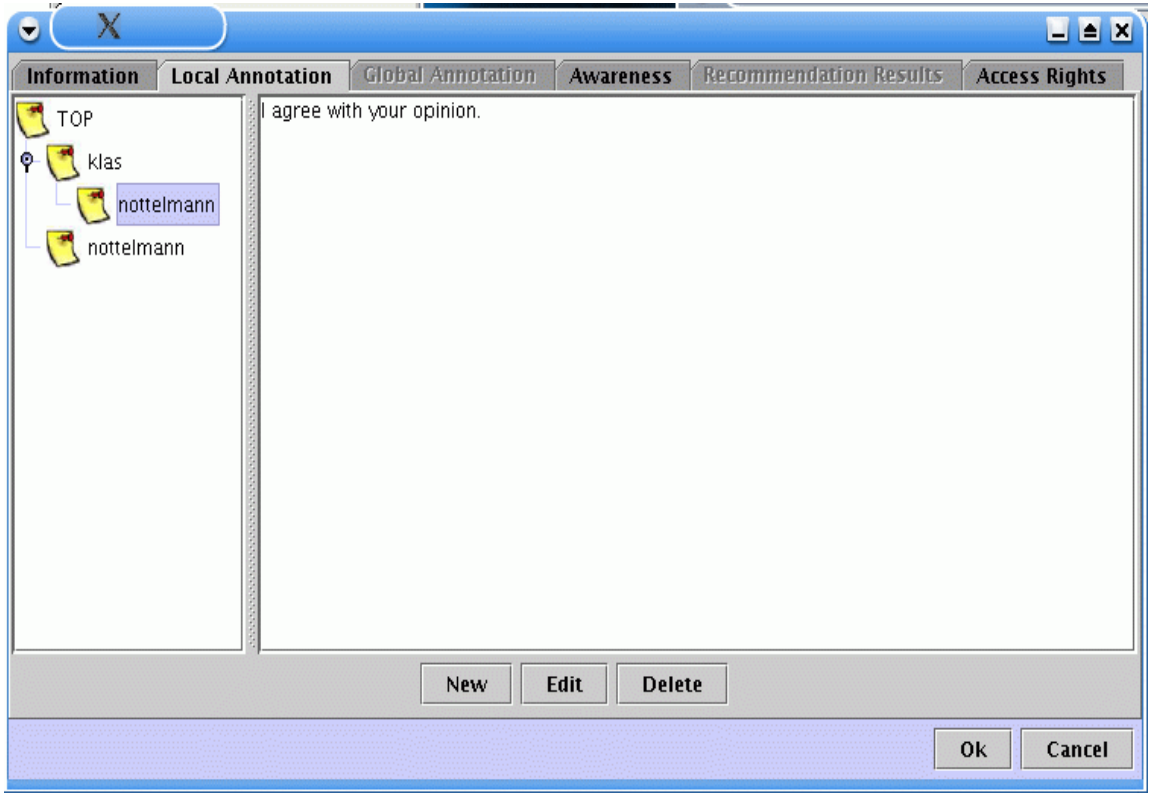
Kollaboration

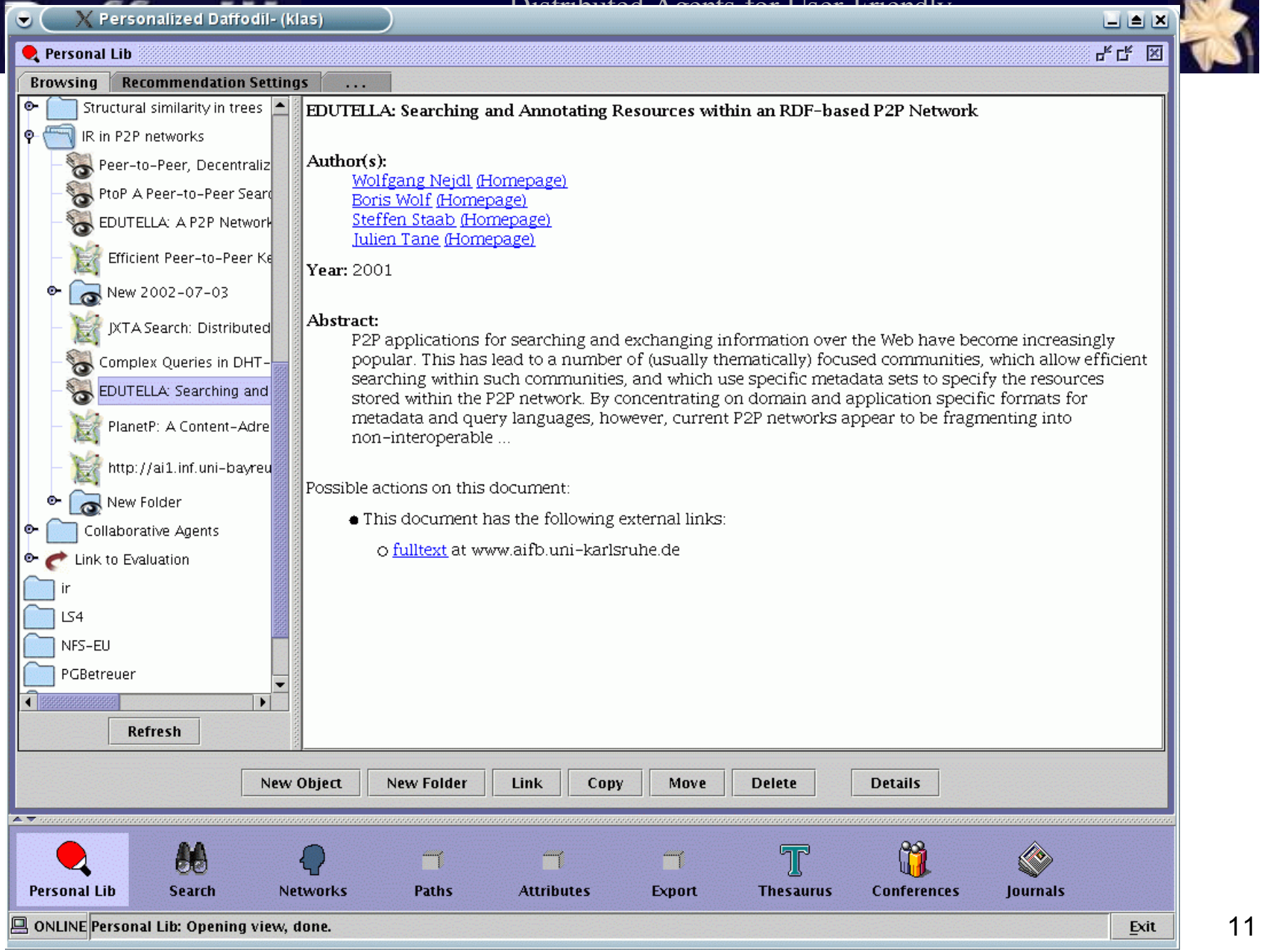


Ziele – ***Konzepte*** – Wrapper

Kollaboration durch Annotation

- Annotationen
- Diskussionen
- Inline Annotationen




Personalized Daffodil- (klas)
Personal Lib
Browsing
Recommendation Settings
...
Structural similarity in trees
IR in P2P networks
Peer-to-Peer, Decentraliz
PtoP A Peer-to-Peer Searc
EDUTELLA: A P2P Network
Efficient Peer-to-Peer Ke
New 2002-07-03
JXTA Search: Distributed
Complex Queries in DHT-
EDUTELLA: Searching and
PlanetP: A Content-Adre
http://ai1.inf.uni-bayreu
New Folder
Collaborative Agents
Link to Evaluation
ir
LS4
NFS-EU
PGBetreuer
Refresh
EDUTELLA: Searching and Annotating Resources within an RDF-based P2P Network
Author(s):
Wolfgang Nejdl (Homepage)
Boris Wolf (Homepage)
Steffen Staab (Homepage)
Julien Tane (Homepage)
Year: 2001
Abstract:
P2P applications for searching and exchanging information over the Web have become increasingly popular. This has lead to a number of (usually thematically) focused communities, which allow efficient searching within such communities, and which use specific metadata sets to specify the resources stored within the P2P network. By concentrating on domain and application specific formats for metadata and query languages, however, current P2P networks appear to be fragmenting into non-interoperable ...
Possible actions on this document:
This document has the following external links:
fulltext at www.aifb.uni-karlsruhe.de
New Object
New Folder
Link
Copy
Move
Delete
Details
Personal Lib
Search
Networks
Paths
Attributes
Export
Thesaurus
Conferences
Journals
ONLINE
Personal Lib: Opening view, done.
Exit

Aktives System

- Bisher „nur“ passives System
- Jetzt soll das System „Aktiv“ am Suchprozess beteiligt werden durch:
- Adaptivität
- Proaktivität

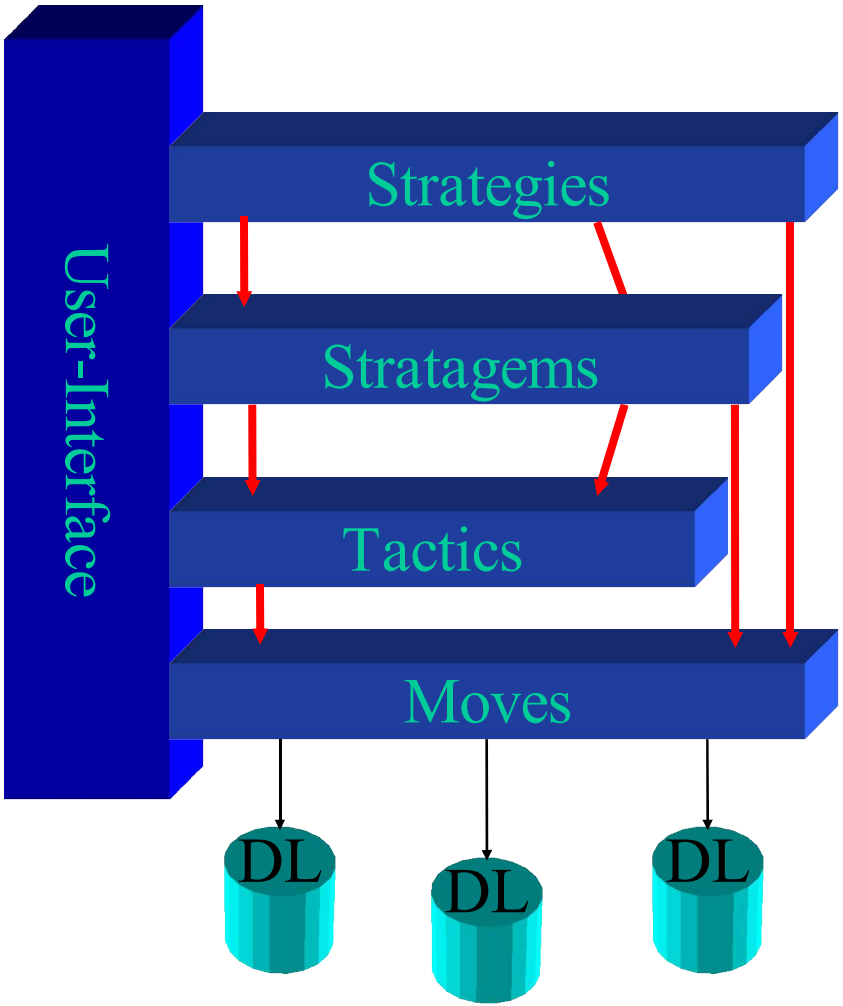
Adaptivität

- Dienste sammeln Informationen
  - Datenquellen: Inhalt und technische Aspekte
  - Nutzer: Verhalten von einzelnen Benutzern und Gruppen
- Dienste verändern Systemverhalten basierend auf den gesammelten Daten
  - Datenquellen: nur spezielle Datenquellen
  - Nutzer: Bestimmte Dienste vorschlagen

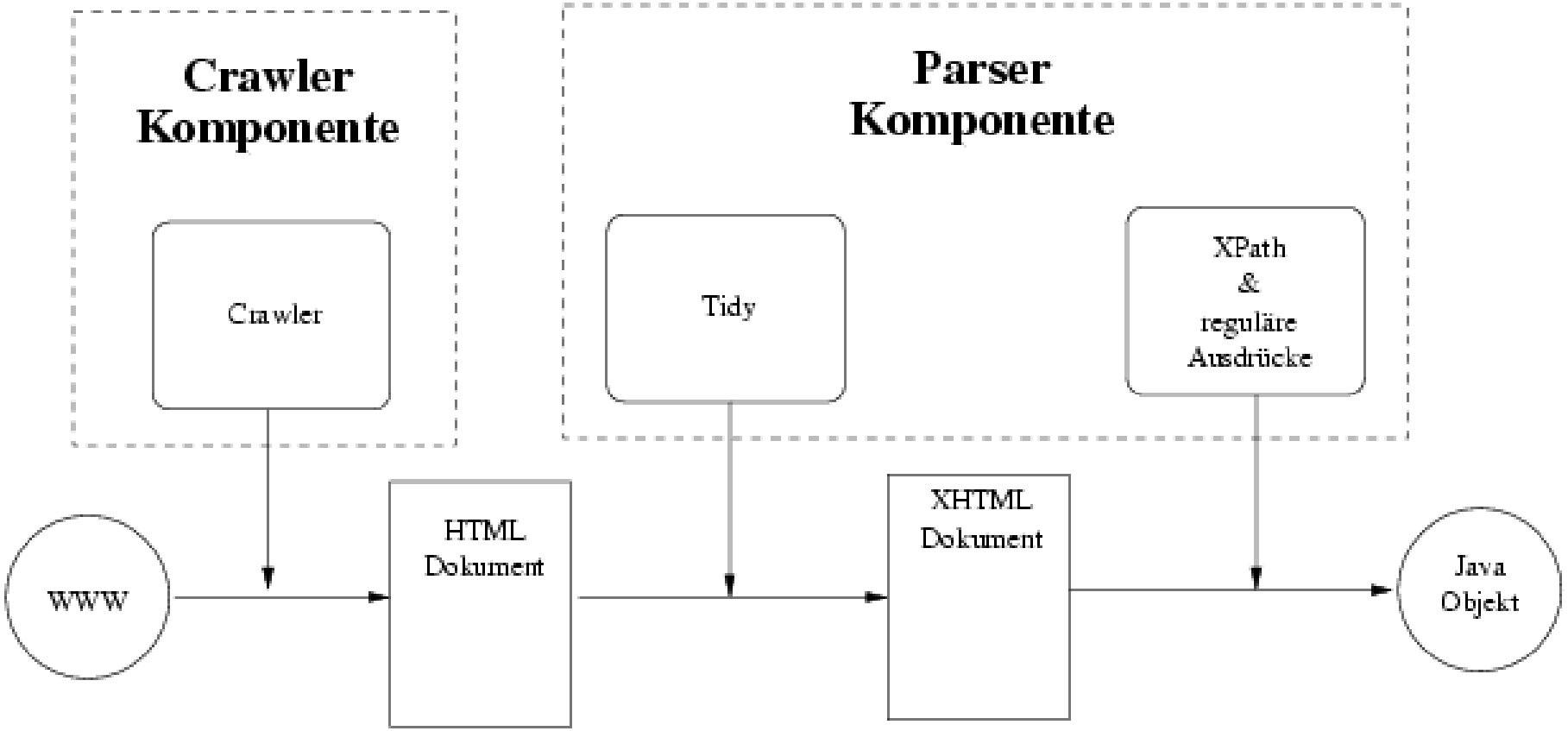
Proaktivität

- Dienste agieren ohne explizite Aufforderung
- Implementiert als ereignisbasierte Regeln
  - Falls Anfrageergebnis leer, Erweitere die Anfrage
  - Falls Resultatliste mehrere Dokumente einer Konferenz beinhaltet, öffne den Konferenz Browser

Agentenbasierte Architektur



Wrapper-Toolkit



WTK: Crawler

- Methoden zum Heranholen von Information
  - HTTP Request Methoden: Get & Post
  - Laden aus Dateisystem
  - Direkt (Datenbank)
  - Encodings

WTK: Parser
Tidy

- HTML -> XHTML
  - Replace Methode
    - Vereinfachung von Syntax
    - Filtern von Elementen
      - Entitäten (&)
      - Sonderzeichen

WTK: Parser
XPath

- Auffinden von Informationen durch XPath
- Beliebige XPath Ausdrücke (Xerces)

```
<extract key=„title“>
    <text xpath=„//title/b“/>
</extract>
```

WTK: Parser
Reguläre Ausdrücke

- XPath reicht nicht aus -> RegExp
  - <split delimiter=„.“>
  - <substitute pattern="},.*$" replacewith="">
  - Beliebig erweiterbar !

WTK: Ausgabe

- Iterationen liefern Java Vector Klasse
- Parsen liefert Java Hashtable Klasse
- Objekte können auch verschachtelt sein!
- Dokumente und Einstellungen\peter \Eigene Dateien\bibdb_metadata.xml
- Dokumente und Einstellungen\peter \Eigene Dateien\acm_search2.xml

WTK: Nice to have

- Semiautomatische Erstellung von XPath Ausdrücken
- Grafische Oberfläche zur Generierung von Konfigurationsdateien
- Automatische Neugenerierung bei Layoutänderungen (DPA)

Zusammenfassung

- Präsentation Daffodil
- Konzepte zu
  - Strategische Unterstützung
  - Personalisierung
  - Aktives System
- Wrapper-Toolkit

www.daffodil.de