DAFFODIL
A User-Oriented Approach for Accessing Federated Digital Libraries

Norbert Fuhr
University of Duisburg-Essen, Germany
Daffodil concepts

1. Strategic information access support
2. Levels of system support
3. Digital library life cycle
4. Collaboration
5. User-friendly integrated desktop
1. Strategic information access support

Levels of search activities (Bates 1990):

2. **Move**: Low-level search function
   
   (e.g. type in search term, view retrieved document)

3. **Tactic**: several moves to further a search
   
   (e.g. broaden/narrow a query)

4. **Stratagem**: set of actions on a single domain
   
   (citation database, tables of contents of journals)

5. **Strategy**: complete plan for satisfying an information need
   
   (e.g. subject search, browse relevant journals, find referenced articles)
Distributed Agents for User-Friendly Access of Digital Libraries

System architecture

- Strategic support in 4 layers
- Federated digital libraries of different types

Moves – Tactics – Stratagems – Strategies – GUI
Distributed Agents for User-Friendly Access of Digital Libraries

Moves

- Distributed search of data from different Digital Libraries & Web sources
- Provide homogenous access to data and functions
- About 10 CS sources accessed

Moves – Tactics – Stratagems – Strategies – GUI
Distributed Agents for User-Friendly Access of Digital Libraries

Resultlist

Fellner
On Minimal Graphs (Note)
(1982) from Achilles

Ketterer, Jens; Puzicha, Jan; Held, Marcus; Fischer, Martin; Buhmann, Joachim M.; Fellner, Dieter W.
On Spatial Quantization of Color Images
(1998) from Cora; Springer; DBLP

Fellner, Wolf-Dietrich
PC und Btx − Koppeling &uuml;ber einen Frontend–Prozessor
(1986) from DBLP

Fellner, Dieter W.; Schäfer, Stefan; Zens, Marco
Photorealistic Rendering in Heterogeneous Networks
(1997) from CiteSeer; DBLP

Moves – Tactics – Stratagems – Strategies – GUI
Combined detail view

Attributes

Links

internal

external

Moves – Tactics – Stratagems – Strategies – GUI
Tactics

- Combine services
- Speed up search
- Redefine query

Moves – Tactics – Stratagems – Strategies – GUI
Coauthors

Moves – Tactics – Stratagems – Strategies – GUI
Stratagems

- Cover an information domain
- Combine all retrieved information & links

Moves – Tactics – Stratagems – Strategies – GUI
Stratagems on the Desktop

- Subject Search
- Journal/Conference Run
- Citation Search
- Author networks
Citation Search: Reference Tracking

- A relevant document is dragged to the tool for citation tracking
- Documents that cite or are cited are retrieved
- Browsing, Inspection and Navigation
- Drag and Drop
  - Can be performed iteratively
**Distributed Agents for User-Friendly Access of Digital Libraries**

### Author Networks

1. Author network browse
2. Ranking a document result set using author centrality

---

**Author Networks**

**Authors:**
- Chen, Hsinchun
- Dhar, Vasant
- Nunamaker, J. F.
- Schatz, Bruce
- Roussinov, Dm
- Schatz, Bruce
- Roussinov, Dm
- Ramser, Marshall
- Zhu, Bin
- Smith, Terence
- Langsjoen, Marit
- Hill, Linda L.
- Chen, H. C.
- Schatz, B.
- Yu, Tao
- Chen, Hsinchun
- Dhar, Vasant
- Nunamaker, J. F.
- Schatz, Bruce
- Roussinov, Dm
- Schatz, Bruce
- Roussinov, Dm
- Ramser, Marshall
- Zhu, Bin
- Smith, Terence
- Langsjoen, Marit
- Hill, Linda L.
- Chen, H. C.
- Schatz, B.
- Yu, Tao

**Networks:**
- DL A
- DL B
- DL C
Strategy

Complex plan to fulfill a specific task

Moves – Tactics – Stratagems – Strategies – GUI
A) Resource selection (MIND project)

- Federated DLs: querying all DLs is too expensive $\Rightarrow$ resource selection
- Traditional approaches (e.g. CORI):
  - Rank DLs w.r.t. similarity to query
  - Similarity computed with some statistical metadata (e.g. average tf)
- Non-co-operating DLs:
  - Retrieve representative sample (QBS)
Cost-based resource selection

Basic idea:
- Retrieval produces costs $C_i(s_i)$
  $s_i$ #documents to be retrieved from $Dl_i$
- Goal: minimise costs

Selection solution:
- $s=(s_1,...,s_n)$ with minimised overall costs:

$$\sum_{i=1}^{n} C_i(s_i) \quad \text{min!}$$
Estimating Costs

- Computation and communication time
- Charges for delivery
- Retrieval quality (most interesting for IR)
  - $C^+ < C^-$ for relevant (non-relevant) documents
  - Important: #relevant documents in result
    1. Mapping of indexing/retrieval weights onto probability of relevance
    2. Distribution of retrieval weights
Underlying retrieval model

- **Linear retrieval function for** $Pr(q \rightarrow d)$

$$Pr(q \leftarrow d) = \Pr(q \leftarrow c_i) \cdot \Pr(c_i \leftarrow d)$$

$c_i \in q$  condition weight  indexing weight

- **Mapping onto probability of relevance**

$$Pr(rel \mid q, d) f(Pr(q \leftarrow d))$$

$f(x) : Pr(rel \mid q \leftarrow d) x$

$$f(x) : \frac{\exp(b_0 \ b_1 \ x)}{1 \ \exp(b_0 \ b_1 \ x)}$$

- Logistic function is better approximation

![Graph showing logistic function approximation](image)
DTF-rp: Recall-precision function

- Estimate #relevant documents in $D_i$
  - Using average indexing weights of query terms

- Estimate #relevant documents in result
  - Apply recall-precision-function
    - Approximated by linear function (defined by starting point $P(0)$)
Distributed Agents for User-Friendly Access of Digital Libraries

DTF-normal: Normal distribution

- Approximate indexing weights
  - normal distribution
- Probabilities of implication $Pr(q \rightarrow d)$
  - Linear retrieval function $\updownarrow$ normal distribution
Evaluation test-bed

Collections:
- TREC-123, split up into 100 collections
- BM 25 indexing weights (normalised idf)

Queries:
- topics 51-150
  - Short: field title
  - Mid-length: field description
  - Long: all fields
Evaluation results

![Graph showing precision and recall for CORI-all, CORI-rs, DTF-rp, and DTF-normal]
Resource Selection: Summary

- Cost-based resource selection with strong theoretical background
- Different methods for estimating retrieval quality (#relevant documents in result)
  - E.g. normal distribution for indexing weights
  - Logistic mapping function outperforms linear one
- DTF in most cases better than CORI
2. Levels of system support

0. Rigid system behavior
2. Adaptive system
3. Proactive system
Adaptivity

1. Agents collect information
   - **Library agents**: content and technical aspects of connected DLs
   - **User agents**: behavior of single users/user group

2. Agents change system behavior based on collected info
   - **Library agents**: select preferred DLs,…
   - **User agents**: offer preferred functions,…
Adaptive search activities
Proactivity

- Agents act without explicit invocation
- Implemented as event-condition-action rules
  - E.g. when query result is empty, broaden query
  - E.g. when result contains multiple articles from a conference, open conference browser
3. Digital library life cycle

Discover -> Retrieve
Re-Present -> Interprete
Collate

(Paepcke 96)
Collate

Save all digital objects in a structured personal digital library

Awareness: 📚 📪 🎨 🔄 🎨 🔄 🎨 🔄 – Keep an eye on my topics

Discover – Retrieve – **Collate** – Interprete – Re-Present
Collate

Personal library:

Store Objects

- Documents
- Authors
- Journals, Conferences
- Articles, Books
- Hyperlinks, Bookmarks
- Query formulations ...

Discover – Retrieve – **Collate** – Interprete – Re-Present
Personal Library

Digital Libraries: A Generic Classification and Evaluation Scheme

Author(s):
- Norbert Fuhr (Homepage)
- Preben Hansen (Homepage)
- Michael Maiba (Homepage)
- Andreas Mocsik (Homepage)
- Ingeborg Solvberg (Homepage)

Journal: Lecture Notes in Computer Science

Conference:
- ECCL

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:
Interpret

- Annotations to DL objects
- Creation of links between DL objects

Discover – Retrieve – Collate – **Interpret** – Re-Present
Re-Present

(support generation of new information)

Export folders from personal library
4. Collaboration

Discover – Retrieve – Collate – Interpret – Re-Present

Distributed Agents for User-Friendly Access of Digital Libraries

Private Group

Public Group
Collation and awareness

- New objects (by other users/due to profile)
- Notification mechanisms (visual markers/email)

Discover – Retrieve – Collate – Interpret – Re-Present
Collaborate interpretation

- Discussion threads
- Rating of objects
Annotation threads

Discover – Retrieve – Collate – **Interpret** – Re-Present
Recommendation

Recommended digital objects through collaboration:

– What do others have for me?
– Who are the others?
– How did others search?

Discover – Retrieve – Collate – Interpret – Re-Present
Distributed Agents for User-Friendly Access of Digital Libraries

5. User-friendly integrated desktop

- Integrate distributed services and software agents in consistent manner
- Provide tool chains to enable users to combine services
- Ensure flexible workflow with free choice of starting point
- Hide complexity

Moves – Tactics – Stratagems – Strategies – GUI
Distributed Agents for User-Friendly Access of Digital Libraries

Desktop

Moves – Tactics – Stratagems – Strategies – GUI
Conclusions

Daffodil concepts
1. Strategic information access support
2. Levels of system support
3. Digital library life cycle
4. Collaboration
5. User-friendly integrated desktop

Try Daffodil: http://www.daffodil.de
Evaluation results (2)

Mid queries, optimum results

<table>
<thead>
<tr>
<th></th>
<th>CORI-all</th>
<th>CORI-rs</th>
<th>DTF-rp</th>
<th>DTF-sample</th>
<th>DTF-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.3160 / +0.0%</td>
<td>0.3240 / +2.5%</td>
<td>0.4440 / +40.5%</td>
<td>0.3840 / +21.5%</td>
<td>0.4480 / +41.8%</td>
</tr>
<tr>
<td>10</td>
<td>0.2930 / +0.0%</td>
<td>0.2970 / +1.4%</td>
<td>0.4190 / +43.0%</td>
<td>0.3580 / -22.2%</td>
<td>0.4220 / +44.0%</td>
</tr>
<tr>
<td>15</td>
<td>0.2853 / +0.0%</td>
<td>0.2713 / -4.9%</td>
<td>0.3987 / +39.7%</td>
<td>0.3327 / +16.6%</td>
<td>0.4087 / +43.3%</td>
</tr>
<tr>
<td>20</td>
<td>0.2670 / +0.0%</td>
<td>0.2570 / -3.7%</td>
<td>0.3830 / +43.4%</td>
<td>0.3160 / +18.4%</td>
<td>0.3980 / +49.1%</td>
</tr>
<tr>
<td>30</td>
<td>0.2430 / +0.0%</td>
<td>0.2310 / -4.9%</td>
<td>0.3627 / +49.3%</td>
<td>0.2900 / +19.3%</td>
<td>0.3747 / +54.2%</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.0253 / +0.0%</td>
<td>0.0220 / -13.0%</td>
<td>0.0605 / +139.1%</td>
<td>0.0297 / +17.4%</td>
<td>0.0703 / +177.9%</td>
</tr>
</tbody>
</table>

Long queries, optimum results

<table>
<thead>
<tr>
<th></th>
<th>CORI-all</th>
<th>CORI-rs</th>
<th>DTF-rp</th>
<th>DTF-sample</th>
<th>DTF-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.5660 / +0.0%</td>
<td>0.5060 / -10.6%</td>
<td>0.5880 / +3.9%</td>
<td>0.5200 / -8.1%</td>
<td>0.5780 / +2.1%</td>
</tr>
<tr>
<td>10</td>
<td>0.5340 / +0.0%</td>
<td>0.4990 / -6.6%</td>
<td>0.5680 / +6.4%</td>
<td>0.5130 / -3.9%</td>
<td>0.5690 / +6.6%</td>
</tr>
<tr>
<td>15</td>
<td>0.5093 / +0.0%</td>
<td>0.4780 / -6.1%</td>
<td>0.5440 / +6.8%</td>
<td>0.4927 / -3.3%</td>
<td>0.5587 / +9.7%</td>
</tr>
<tr>
<td>20</td>
<td>0.4940 / +0.0%</td>
<td>0.4510 / -8.7%</td>
<td>0.5400 / +9.3%</td>
<td>0.4605 / -6.8%</td>
<td>0.5450 / +10.3%</td>
</tr>
<tr>
<td>30</td>
<td>0.4683 / +0.0%</td>
<td>0.4260 / -9.0%</td>
<td>0.5233 / +11.7%</td>
<td>0.4260 / -9.0%</td>
<td>0.5323 / +13.7%</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.0763 / +0.0%</td>
<td>0.0622 / -18.5%</td>
<td>0.1111 / +45.6%</td>
<td>0.0506 / -33.7%</td>
<td>0.1230 / +61.2%</td>
</tr>
</tbody>
</table>
Evaluation results (3)

Mid queries, 10 DLs selected

<table>
<thead>
<tr>
<th></th>
<th>CORI-all</th>
<th>CORI-rs</th>
<th>DTF-rp</th>
<th>DTF-sample</th>
<th>DTF-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.3160 / +0.0%</td>
<td>0.3240 / +2.5%</td>
<td>0.3900 / +23.4%</td>
<td>0.3860 / +22.2%</td>
<td>0.3940 / +24.7%</td>
</tr>
<tr>
<td>10</td>
<td>0.2930 / +0.0%</td>
<td>0.2970 / +1.4%</td>
<td>0.3710 / +26.6%</td>
<td>0.3700 / +26.3%</td>
<td>0.3730 / +27.3%</td>
</tr>
<tr>
<td>15</td>
<td>0.2853 / +0.0%</td>
<td>0.2713 / -4.9%</td>
<td>0.3500 / +22.7%</td>
<td>0.3453 / +21.0%</td>
<td>0.3533 / +23.8%</td>
</tr>
<tr>
<td>20</td>
<td>0.2670 / +0.0%</td>
<td>0.2570 / -3.7%</td>
<td>0.3360 / +25.8%</td>
<td>0.3210 / +20.2%</td>
<td>0.3365 / +26.0%</td>
</tr>
<tr>
<td>30</td>
<td>0.2430 / +0.0%</td>
<td>0.2310 / -4.9%</td>
<td>0.3103 / +27.7%</td>
<td>0.2933 / +20.7%</td>
<td>0.3107 / +27.9%</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.0253 / +0.0%</td>
<td>0.0220 / -13.0%</td>
<td>0.0372 / +47.0%</td>
<td>0.0307 / +21.3%</td>
<td>0.0354 / +39.9%</td>
</tr>
</tbody>
</table>

Long queries, 10 DLs selected

<table>
<thead>
<tr>
<th></th>
<th>CORI-all</th>
<th>CORI-rs</th>
<th>DTF-rp</th>
<th>DTF-sample</th>
<th>DTF-normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.5660 / +0.0%</td>
<td>0.5060 / -10.6%</td>
<td>0.5420 / -4.2%</td>
<td>0.5320 / -6.0%</td>
<td>0.5480 / -3.2%</td>
</tr>
<tr>
<td>10</td>
<td>0.5340 / +0.0%</td>
<td>0.4990 / -6.6%</td>
<td>0.5140 / -3.7%</td>
<td>0.5220 / -2.2%</td>
<td>0.5300 / -0.7%</td>
</tr>
<tr>
<td>15</td>
<td>0.5093 / +0.0%</td>
<td>0.4780 / -6.1%</td>
<td>0.5007 / -1.7%</td>
<td>0.4967 / -2.5%</td>
<td>0.5033 / -1.2%</td>
</tr>
<tr>
<td>20</td>
<td>0.4940 / +0.0%</td>
<td>0.4510 / -8.7%</td>
<td>0.4815 / -2.5%</td>
<td>0.4795 / -4.8%</td>
<td>0.4845 / -1.9%</td>
</tr>
<tr>
<td>30</td>
<td>0.4683 / +0.0%</td>
<td>0.4260 / -9.0%</td>
<td>0.4477 / -4.4%</td>
<td>0.4323 / -7.7%</td>
<td>0.4490 / -4.1%</td>
</tr>
<tr>
<td>Avg.</td>
<td>0.0763 / +0.0%</td>
<td>0.0622 / -18.5%</td>
<td>0.0671 / -12.1%</td>
<td>0.0518 / -32.1%</td>
<td>0.0617 / -19.1%</td>
</tr>
</tbody>
</table>
Recommendation: Path model

User leaves traces; DAFFODIL monitors
- User actions
- Attributes of documents or digital objects viewed (within context)

3 different ways to guide the user
- System-given paths (static)
- User-defined paths (macro)
- System-adapted paths (dynamic)
  - learned from one user
  - or all users
Path model

- User action
- Current position in path
- Possible paths to go
Adaptivity: Architecture
Distributed Agents for User-Friendly Access of Digital Libraries

Search

- Search Domain Specification
- Filter Settings and Query Composition
- Integrated Result Lists with Draggable Objects
- Navigation
- Detail Inspection

Daffodil: Distributed Agents for User-Friendly Access of Digital Libraries

Author(s):
- Norbert Gürt (Homepage)
- Norbert Fahl (Homepage)
- Claus-Peter Kra (Homepage)

Journal: Lecture Notes in Computer Science

Conference:
- ECDL

Year: 2000
Pages: 032

Abstract:
The Internet makes searching for literature in Digital Libraries (DLs) feasible. However, often a user has to contact several DLs to satisfy a given information need. This leads to usability problems due to the heterogeneity of the DLs. One aspect is that the information structures of the systems differ. In fact, relevant information may be spread across several DLs. The other aspect of heterogeneity is differing browsing and searching functionality, of course presented to the user through different user interfaces and query languages.

Possible actions on this document:
- This document has the following external links:
  - fulltext at link.springer.de