DAFFODIL
Effective Support for Using Digital Libraries

Norbert Fuhr
University of Duisburg-Essen, Germany
Desktop system for federated digital libraries
Drawbacks of today’s DLs

In all workflow phases
- multiple access points
- multiple query forms
- poor functionality
Daffodil concepts: Focus on the user

1. Strategic information access support
2. Proactive system support
3. Digital library life cycle
4. Collaboration
1. Strategic information access support

Levels of search activities (Bates 1990):

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

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

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

1. **Strategy**: complete plan for satisfying an information need
   (e.g. subject search, browse relevant journals, find referenced articles)
Search

**Moves** – Tactics – Stratagems – Strategies
Result list

- Martin Rajman, R: Text Mining Knob (1993) from CiteSeer.
- Katherine J. Don; Michael Dewsnip; Valentin Tablan; Ian H. Witten: Text mining in a digital library (2004) from Achilles.
- Eleazar Eskin; Eugene Agichtein: Combining Text Mining and Sequence Analysis to Discover Protein Functional Regions (2004) from Achilles.
- María J. Martín-Bautista; Daniel Sánchez; Miguel Delgado; María Amparo Vila Miranda; José-María: Association Rule Extraction for Text Mining (2002) from Achilles.

Moves – Tactics – Stratagems – Strategies
Visualizing Sequential Patterns for Text Mining (translate)

Author(s)
- Harlan Foote
- Wendy Cowley
- Elizabeth Jurrus
- Pak Chung Wong
- Jim Thomas

Conference
- INFOVIS 2000

In: 2000 IEEE Symposium on Information Visualization (InfoVis 00)
Year: 2000
Month: oct
Pages: 105--114

Possible actions on this document are:
- This document has the following external links:
  - Fulltext at computer.org
  - Also take a look at Scholar Google

Moves – Tactics – Stratagems – Strategies
Chung-Hong Lee
Donald N. Kostoff Dunja Mladenic
Edilberto Magalhães Silva Edilson Ferneda Eric P. G. Martin
Hercules Antonio do Prado Hirofumi Matsuzawa Hironori Takeuchi
Hsin-Chang Yang Hsu-Chun Yu
Ian H. Witten John A. Atkinson-Aburidy Jose Ruiz-Shuklaper
Jose Palazzo Moreira de Oliveira Koichi Takeda Leandro Krug Wives
Marko Grobelnik Marko Multhaupt
Martin Rajman Mohamed Quafafou
Moshe Fresko Natasa Milic-Frayling Orly Liphstat
Rafael Berlanga Llavori Raymond Y. K. Lau Rene Tshteia
Ronen Feldman Roxana Giriu Shogo Nishida
Singamedu Srinivasan Thanaruk Theeramun Kong Tohru N
Yikun Guo Yonatan Schler
Yukitaka Kusumura Yuseop Kim Z.

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

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

<table>
<thead>
<tr>
<th>References</th>
<th>Other documents citing this one</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eugene L. Margulis, Modelling moves, Tactics, Stratagems, Strategies</td>
<td>Khalid Al-Kofahi, Alex Tyrrell</td>
</tr>
</tbody>
</table>
Evaluation of strategic support

- Evaluation with 28 subjects
- Methods: questionnaire, logging, video
- Tasks of different complexity:

![Diagram showing the relationship between knowledge, complexity, and information need. The diagram illustrates a positive correlation between complexity and information need, with knowledge and object complexity increasing along the x-axis and y-axis respectively.]
Evaluation of strategic support

- Only for very simple tasks (known item retrieval), other methods may be faster
- Complex tasks are supported very effectively
2. Proactive system support

Spell checking
Proactivity during query formulation

Semantically inconsistent queries
Proactivity during query formulation

Related Terms

- usability
- information retrieval
- cascading
- end user
- ergonomic
- in house
- interface elements
- interface engineering
- likert scale
- mental model
- post hoc analysis
- real estate
- redesign
- search engine
- search task
- user centered design
- user interface design
- user satisfaction
Usability studies

- 30 subjects
- Focus on query formulation

Results:
- Users feel rather uncertain
  - submit the same query several times,
  - stare at the keyboard
- Users ignore hints shown elsewhere on the desktop

Improvements through proactive functions:
- Fewer syntactical and semantic errors
- Better tactical decisions
- Users feel more confident
Proactivity: Search continuation

- Suggestions based on automatic analysis of current search result
- Using case-based reasoning
- Availability indicated by unobtrusive button at bottom
Search continuation

- Suggestions presented in ranked list

- Descriptive title, explanation and score bar

- Execute one or more suggestion and judge them

- Icons indicate status of suggestion (executable, used, useful)
12 participants in pilot study, each spending 40-50 minutes with the system

Case base initially contained 25 cases (generic situations) and 16 solutions (suggestions)

Results:

Search novices and casual users found suggestions on advanced tools most useful

Experienced users liked extraction of terms, authors, . . . from results

Might have used tactics on their own, but advice helped them avoid trial-and-error
24 participants, half of them supported each working on 3 tasks
- case base contained 30 cases

Results
Supported users...
- are more satisfied with the search process (p=0.067)
- are significantly more satisfied with the result
- find significantly more relevant documents
- use significantly more often Daffodil's advanced tools (unsupported users mainly concentrate on query reformulation)
3. Digital library life cycle
3. Digital library life cycle

- Discover
  - Information source selection
- Retrieve
  - Information retrieval
- Interpret
  - Results of evaluation
- Re-Present
  - Knowledge creation
- Collate
  - Organizing retrieved material
3. Digital library life cycle
3. Digital library life cycle

- Retrieve
- Interpret
- Evaluation of results
- Re-present

Relevant terms:
- Organizing retrieved material
- Knowledge creation
- Source selection
- Information retrieval
3. Digital library life cycle

- Re-present
  - Knowledge creation
- Discover
  - Information retrieval
- Retrieve
  - Information retrieval
- Collect
  - Organized preserved material
- Interpret
  - Evaluation of results
- Represent

27
3. Digital library life cycle

- discover
  - information source selection
- retrieve
  - information retrieval
- collate
  - organizing retrieved material
- interpret
  - evaluation of results
- re-present
  - knowledge creation
Collate – Personal library

Folder hierarchy (e.g. task-oriented)

Save all digital library objects
  • Document Metadata
  • Authors
  • Journals, Conferences
  • Hyperlinks
  • Query formulations
  • Fulltexts

Save own documents in the same folders
  • Papers
  • Presentations

Discover – Retrieve – Collate – Interpret – Re-Present
Text mining in a digital library

Author(s)
- Katherine J. Don
- Michael Dewsnip
- Valentin Tablan
- Ian H. Witten

Journal
- Int. J. on Digital Libraries 4

Year: 2004
Number: 1
Pages: 56--59

Possible actions on this document are:

- This document has the following external links:
  - Fulltext at dx.doi.org
  - Also take a look at Scholar Google

- This document was found on the following datasources:
  - DBLP
  - Achilles
Interpret: Annotations

- Annotation of any object in the personal library (out-of-line/inline)
- Annotations can be typed and nested
- Retrieval of annotations (POLAR: Probabilistic Object Logic for Annotation Retrieval)

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

(support generation of new information)

Export folders from personal library

Discover – Retrieve – Collate – Interpret – Re-Present
4. Collaboration

- Discover
- Information retrieval
- Information source selection
- Organizing retrieved material
- Evaluation of results
- Knowledge creation
- Re-present
- Interpret
- Collate

Personal
Group
Public
Collation and awareness

New objects

- Filed by other users
- recommendation

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

Annotation threads

Discover – Retrieve – Collate – **Interpret** – Re-Present
Collaboration: Chat-Tool

- Help
- Discussion
- Object exchange
- Ask an expert (Librarian)
- Chatter-Bot
5. Further Applications
European National Libraries
Application to Image and Music Retrieval
1. **G3 Integrates Three System Technologies** from *COMPUTER(2000)*

   Ayman Kayssi, Ralph Achkar, Mark Azar, Joseph Samaha

   - 1 Introduction
   - 10 Emerging technologies
     - 10.2 **GPS** ICs and cell **phones**
     - 5 G3 components
     - 6 Mobile site

2. **Mobility Management for Cellular Telephony Networks** from *IEEE PARALLEL DISTRIBUTED TECHNOLOGY(1996)*

   Yi-Bing Lin

   - 2 Mobility management
   - 3 Handoff
     - 3.2 **INTERSYSTEM HANDOFF**
   - 4 Roaming
     - 4.2 **REGISTRATION**
     - 4.3 **CALL DELIVERY**

**No abstract available**
XML Retrieval: INEX Interactive Track

Evaluation with 80 subjects from 10 research groups
6. Summary and Outlook

Daffodil concepts:

1. Strategic information access support
2. Proactive system support
3. Digital library life cycle
4. Collaboration

Successful evaluations for 1 and 2
ezDL: Daffodil's successor

Redesign of Daffodil:

• more modular design
• more flexible
• more user-friendly interface
• extended logging support for user experiments (e.g. eye tracking)

http://ezdl.de
ezDL application

An Information Retrieval View of Environmental Information Systems

Year: 1991
Author: Norbert Fuhr

Abstract

In the design of future environmental systems, the semantics of the data as well as the kind of queries to those systems have to be considered. Environmental data is frequently uncertain and incomplete. Vague queries and imprecise data, methods developed in information retrieval can be applied. Heterogeneous data structures can be handled with concepts from object-oriented database management systems. In multimedia information systems, the problem of full integration of the different media is yet unsolved, especially in case the information is stored in different media. We claim that the retrieval interface offered by current database management systems is not sufficient for interactive use. In addition, functions like ranking, browsing, zooming, relevance feedback and cooperative support should be provided. Finally, we emphasize the need for empirical studies for the design of future environmental information systems. Keywords: information retrieval, environmental information systems, heterogeneous data structures, multimedia data, vague queries, imprecise data, interactive information systems.
Filtering of result list

An Information Retrieval View of Environmental Info Systems

Year 1991
Author Norbert Fuhr

Abstract

In the design of future environmental systems, the semantics of the data as well as the kind of queries to those systems have to be considered.

Environmental data is frequently uncertain and incomplete. Heterogeneous data structures as well as multimedia data have to be managed by the system. For interactive queries, the system should allow vague queries and query formulations that are independent of the specific structure of the data and its representation. For vague queries and imprecise data, methods developed in information retrieval can be applied. Heterogeneous data structures can be handled with concepts from object-oriented database management systems. In multimedia information systems, the problem of full integration of the different media is yet unsolved, especially in case the information a user searches for is stored in different media. We claim that the retrieval interface offered by current database management systems is not sufficient for interactive use. In addition, functions like ranking, browsing, zooming, relevance feedback and cooperation support should be provided. Finally, we emphasize the need for empirical studies for the design of future environmental information systems. Keywords: information retrieval, environmental information systems, heterogeneous data structures, multimedia data, vague queries, imprecise data, interactive information systems.
Sorting and Grouping of Results

- Sort by: Relevance
- Group by: Authors

1. Image Retrieval by Local Evaluation of Nonlinear Kernel Functions around Salient Points.
   - Alaa Halawani; Hans Burkhardt
   - 2004 (DBLP)

2. Retrieval Status Values in Information Retrieval Evaluation.
   - Amélie Imafouo; Xavier Tannier
   - 2005 (DBLP)

3. Evaluation of a Tool for Visualization of Information Retrieval Results.
   - Aravindan Veerasamy; Nicholas J. Belkin
   - 1996 (DBLP)

4. Here or there: preference judgments for relevance
   - Ben Carterette; Paul N. Bennett; David Maxwell Chickering; Susan T. Dumais
   - 2008 (ACM DL)

5. Semiautomatic evaluation of retrieval systems using document similarities.
   - Ben Carterette; James Allan
   - 2007 (DBLP)
Alternative Result List Based on Hypergrid
Tool Perspectives

A Probabilistic Framework for Information Modelling and Retrieval Based on User Annotations

Year: 2008
Author: Ingo Frommholz

Abstract:
Annotations are a means to make critical remarks, to explain and comment things, to add notes and give an impression of the richness of the information they can be found in digital libraries and collaboratories, for example as a building block for scientific discourse. The idea of annotations can be further refined in the product reviews, scientific databases and many "Web 2.0" applications. They can be regarded as annotations in a certain sense. Digital annotations can be (temporal) comments, markings, or other documents or document parts. Since annotations convey information which is potentially important for the question of how to exploit annotations for information retrieval, it gives a first answer to the question of how to exploit annotations for information retrieval. A survey of the "annotation universe" reveals some facets of annotations for information retrieval. Annotations (broadening the context of the annotation object) or meta level ones (saying something about annotations themselves), other objects created during the process of annotation can be interesting for retrieval.

These objects are integrated into an object-oriented model comprising digital objects such as structured or non-structured fragments. In this model, the different relationships among the various objects are reflected. From this model, a probabilistic, object-oriented logic framework called POLAR is introduced. POLAR is modelled by means of probabilistic propositions and a first-order logic. POLAR aims at specifying several annotated subparts or fragments. Queries can be posed to extract the knowledge contained in structural and content-based retrieval, i.e. document and discussion search. By applying an augmentation strategy to the propositions from subcontexts like annotations, or relevance augmentation, where retrieval status values for a context can be augmented, it is possible to derive a further probability from the first one. If efficiently implemented, POLAR allows for specifying a context-dependent retrieval strategy and allows for exploiting the large context and content-based relations in digital libraries.

Results: 44

6. A Probabilistic Framework for Information Modelling and Retrieval Based on User Annotations
Ingo Frommholz
2008 (DBLP)

7. Annotation-Based Document Retrieval with Probabilistic Logics
Ingo Frommholz, Norbert Furler
2009 (DBLP)

8. Evaluation of Relevance and Knowledge Augmentation in Discussion Forums
Ingo Frommholz, Marco Lichtenfeld
2009 (DBLP)

9. Probabilistic, object-oriented logics for annotation-based retrieval
Ingo Frommholz
2009 (DBLP)

10. Determining the Polarity of Postings for Discussion Search
Ingo Frommholz, Marco Lichtenfeld
2009 (DBLP)

Ingo Frommholz, Ray Lar son
2007 (ACM DL)
Results can be printed or copied to the clipboard.
Supporting query formulation

- Spelling correction
- Suggestions with definitions
- Searching for disambiguated concepts
- Implicit information about users used for search
- Geo location and user language
Visualizing results

- Grouped visualization of result list
- Groups predefined or chosen by users
Visualizing results

- Heterogeneous results
- Smart result items with visualization of performed actions
- Explanations and help on mouse-over
- Keyword highlighting for search or filter terms
Visualizing details

- Quick search links for forward navigation
- HON labels help users evaluating suitability of results
- Access to fulltext
- Result snippets with hits highlighted
- Saving, exporting, printing directly from result view
- Visualization of frequent terms in document as term cloud with quick search links