

# Proactive Support for Query formulation A case study with DAFFODIL

Claus-Peter Klas, Matthias Jordan and André Schaefer

IIIS · Universität Duisburg-Essen

Sept. 2005



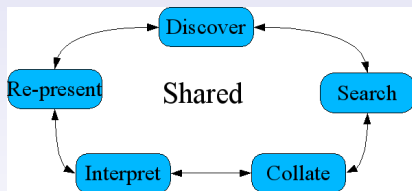
# Contents

- 1** Introduction to DAFFODIL
- 2** Motivation and Concepts for Proactive Support
- 3** Proactive Support Methods

# DAFFODIL

- is a virtual Digital Library with **strategic support** given by high-level search functions.
- is a service oriented architecture with currently over 18 different services for searching and browsing in the area of computer science.
- connects over 15 different data sources from the area of computer science
- is very flexible in including new services and GUI tools

# Goal: Support the Scientific Workflow



## ■ Scientific Workflow

- Discover relevant data sources
- Retrieve relevant data
- Collate information in own structured knowledge environment.
- Interpret gathered information
- Re-Present in new publications

# Problem Description

- Observation: Many erroneous queries are submitted!
- DAFFODIL is a meta search engine:
  - Wrong queries cost much time
  - Misleading results are harder to track down
  - Iterative query cycles are much longer
- Can proactive support improve queries and avoid frustration?

# Problem Description

## Common Query Formulation Problems

- Spelling
  - Correctly spelled?
  - Need to name all common variants (color vs. colour)?
  - Other languages?
- Wording and Vocabulary
  - Is it the right term?
  - Do I need a more specific term?
  - Synonyms?
- Semantic consistency
  - Will the parts of my query yield any result?
  - Did I specify a contradiction?
- Reuse of query parts
  - Reuse an old query: what's new?
  - Combine former query with new clauses...

# Problem Description

## Common Query Formulation Problems

- Spelling
  - Correctly spelled?
  - Need to name all common variants (color vs. colour)?
  - Other languages?
- Wording and Vocabulary
  - Is it the right term?
  - Do I need a more specific term?
  - Synonyms?
- Semantic consistency
  - Will the parts of my query yield any result?
  - Did I specify a contradiction?
- Reuse of query parts
  - Reuse an old query: what's new?
  - Combine former query with new clauses...

# Problem Description

## Common Query Formulation Problems

- Spelling
  - Correctly spelled?
  - Need to name all common variants (color vs. colour)?
  - Other languages?
- Wording and Vocabulary
  - Is it the right term?
  - Do I need a more specific term?
  - Synonyms?
- Semantic consistency
  - Will the parts of my query yield any result?
  - Did I specify a contradiction?
- Reuse of query parts
  - Reuse an old query: what's new?
  - Combine former query with new clauses...



# Problem Description

## Common Query Formulation Problems

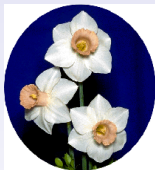
- Spelling
  - Correctly spelled?
  - Need to name all common variants (color vs. colour)?
  - Other languages?
- Wording and Vocabulary
  - Is it the right term?
  - Do I need a more specific term?
  - Synonyms?
- Semantic consistency
  - Will the parts of my query yield any result?
  - Did I specify a contradiction?
- Reuse of query parts
  - Reuse an old query: what's new?
  - Combine former query with new clauses...

# Problem Description

## Common Query Formulation Problems

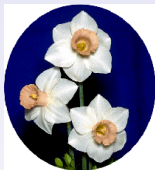
- Spelling
  - Correctly spelled?
  - Need to name all common variants (color vs. colour)?
  - Other languages?
- Wording and Vocabulary
  - Is it the right term?
  - Do I need a more specific term?
  - Synonyms?
- Semantic consistency
  - Will the parts of my query yield any result?
  - Did I specify a contradiction?
- Reuse of query parts
  - Reuse an old query: what's new?
  - Combine former query with new clauses...

# General Goals in Daffodil



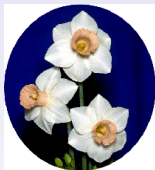
- Combine browsing and search strategies
- Offer higher level search strategies
- Strategic support

# General Goals in Daffodil



- Combine browsing and search strategies
- Offer higher level search strategies
- Strategic support

# General Goals in Daffodil



- Combine browsing and search strategies
- Offer higher level search strategies
- Strategic support

# Goals for proactive support

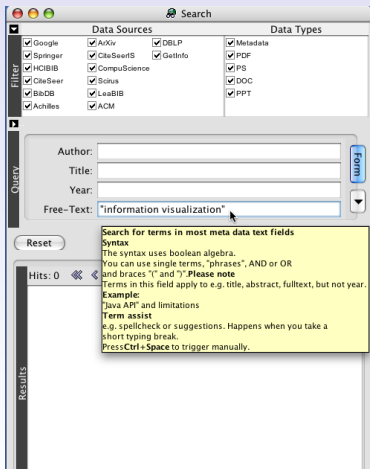
- Help users to select proper terms
- Help users to detect and correct errors easily
- Give feedback of quality of query
- Reduce the probability of erroneous queries
- Increase esteem and faith of users
- **Thus a means of tactical and strategic support**

# Approach

- 1 Test, which problems users face (User Problems)
- 2 Offer proactive support features in query formulation context
- 3 Current Solutions
- 4 Evaluate the improvements

# Search

## Easy to Use Search Interface



- Users are unsure about query syntax
- ▷ Offer precise help in context
- Many digital libraries – many query languages
- ▷ Homogenization
- Not enough - users still have problems



# Search

## Easy to Use Search Interface

The screenshot shows a search interface window titled "Search". It has a "Filter" sidebar on the left with a "Query" section containing input fields for "Author:", "Title:", "Year:", and "Free-Text:". The "Free-Text:" field contains the text "information visualization". Below the input fields is a "Reset" button. To the right of the input fields is a "Form" button. The "Data Sources" and "Data Types" sections are visible at the top, each with a list of checked items. A yellow tooltip is displayed over the "Free-Text:" field, providing instructions on search syntax and term assistance.

**Search for terms in most meta data text fields**  
**Syntax**  
 The syntax uses boolean algebra. You can use single terms, "phrases", AND or OR and braces "(" and ")". **Please note** Terms in this field apply to e.g. title, abstract, fulltext, but not year.  
**Example:**  
 "java API" and limitations  
**Term assist**  
 e.g. spellcheck or suggestions. Happens when you take a short typing break.  
 Press **Ctrl+Space** to trigger manually.

- Users are unsure about query syntax
- ▷ Offer precise help in context
- Many digital libraries – many query languages
- ▷ Homogenization
- Not enough - users still have problems

# Active Support

## Query Formulation

Search

Filter

Author:

Title: information retrieval AND user interface

Year:

Free-Text:

interface  
interfere

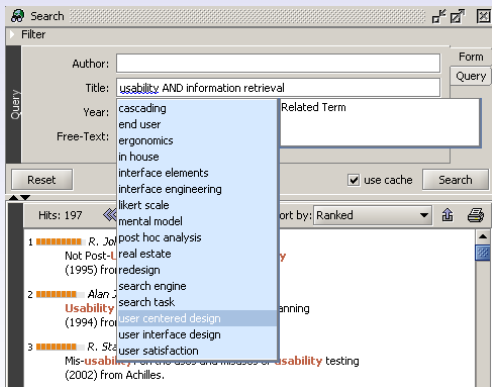
Reset  use cache Search

What do users expect?

- Auto spell check
- Mark errors
- List of possible corrections
- Completion of known proper names (Authors)

# Active Support

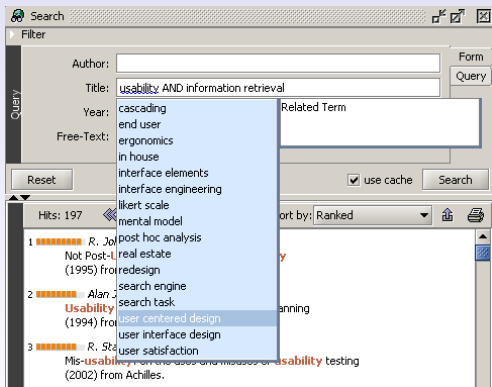
## Term Suggestions for Query Improvement



- User starts to search in ASK  
He is *insecure*.
- Synonyms and Homonyms are needed
- Offer Terms in Context

# Active Support

## Term Suggestions for Query Improvement



- User starts to search in ASK  
He is *insecure*.
- Synonyms and Homonyms are needed
- Offer Terms in Context

# Status Quo

- Current Search Engines only start to exploit the possibilities
- Most don't
- Known examples:
  - Google "Did you mean?"
  - Google Suggest
  - $AI^2RS$  [Jansen and Pooch]

# Google Suggest

## (Suggests Mistakes by Design)

Google - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.google.com/webhp?complete=1&hl=en

# Google™

## Suggest BETA

Web Images Groups News Froogle Local<sup>New!</sup> more »

information retrieval	37,300 results
information retrieval	659 results
information retrieval on the web	19,400 results
information retrieval system	22,800 results
information retrieval	4,430 results
information retrieval white papers	1,440 results

Advanced Search  
 Preferences  
 Language Tools

As you type, Google results. [Learn more](#)

## Related Work

# AI<sup>2</sup>RS

## Gives Improvement Hints

The screenshot shows a search interface with a search bar containing the query: "Leadership OR training OR ethics OR donuts". An "Agent Feedback" window is open, displaying the following text:

There may be some misspelled terms in your query. Here are some suggested corrections.

ethics

You may want to change the following query terms. They have never appeared in any result list.

donuts

The following are synonyms and context for your query terms. Is this the context you want? If not, try a different term.

leadership leading | the activity of leading his leadership inspired the team training preparation grooming | activity leading to skilled behavior

Annotations with arrows point to specific elements:

- Corrects query structure.** Points to the "Agent Feedback" button in the search bar.
- Spell check and suggestions.** Points to the word "ethics" in the feedback window.
- Terms to remove.** Points to the word "donuts" in the feedback window.
- Synonyms and context.** Points to the synonym list at the bottom of the feedback window.

# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis



# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:
    - Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1** Screen Mockups
- 2** Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3** Base Line Evaluation without Modules
- 4** Evaluation with Modules
- 5** Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1** Screen Mockups
- 2** Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3** Base Line Evaluation without Modules
- 4** Evaluation with Modules
- 5** Single User Loud Thinking Tests for Detailed Analysis

# Development and Evaluation

- 1 Screen Mockups
- 2 Implementation of Modules
  - Spell Check
  - Improvement of Query Syntax
  - Suggestions:  
Synonyms, Related Terms, Term Reuse, Author Names
  - Error Markers for Syntax and Over-Constrained Queries
- 3 Base Line Evaluation without Modules
- 4 Evaluation with Modules
- 5 Single User Loud Thinking Tests for Detailed Analysis



# Results

- Good:
  - Confidence Improves
  - Less negative remarks
  - Design does not distract
- Bad:
  - Still many wrong queries
  - Search time still high
  - People look elsewhere (keyboard)
- Overall: Right Direction

# Results

- Good:
  - Confidence Improves
  - Less negative remarks
  - Design does not distract
- Bad:
  - Still many wrong queries
  - Search time still high
  - People look elsewhere (keyboard)
- Overall: Right Direction