Proactive Support for Query formulation A case study with DAFFODIL

Claus-Peter Klas, Matthias Jordan and André Schaefer

IIIS · Universität Duisburg-Essen

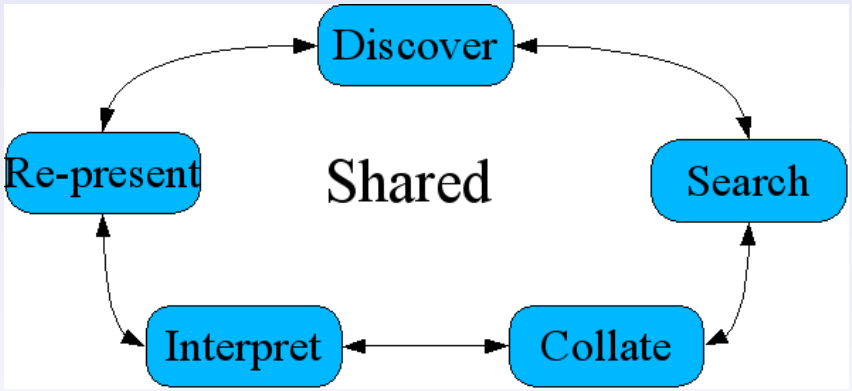
Sept. 2005

Contents

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...

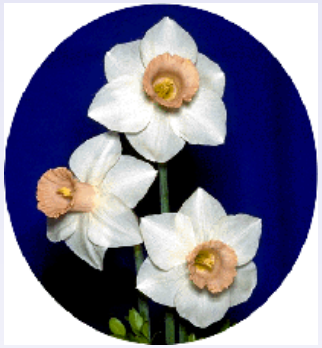
Problem Description
Common Query Formulation Problems

- 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

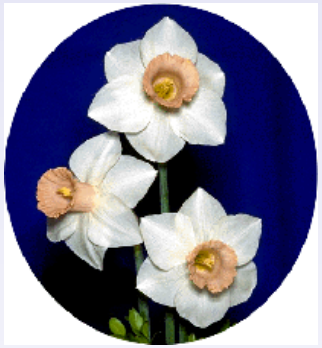
- Spelling
  - Correctly spelled?
  - Need to name all common variants (color vs. colour)?
  - Other languages?
- Wording and Vocabulary
  - Is it the right term?
  - Do I need a more specific term?
  - Synonyms?
- Semantic consistency
  - Will the parts of my query yield any result?
  - Did I specify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

General Goals in Daffodil



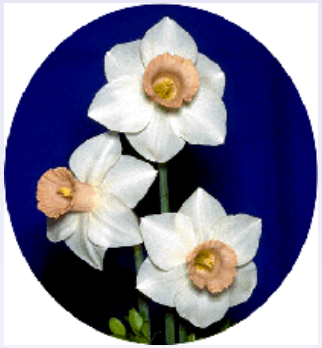
- Combine browsing and search strategies
- Offer higher level search strategies
- Strategic support

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

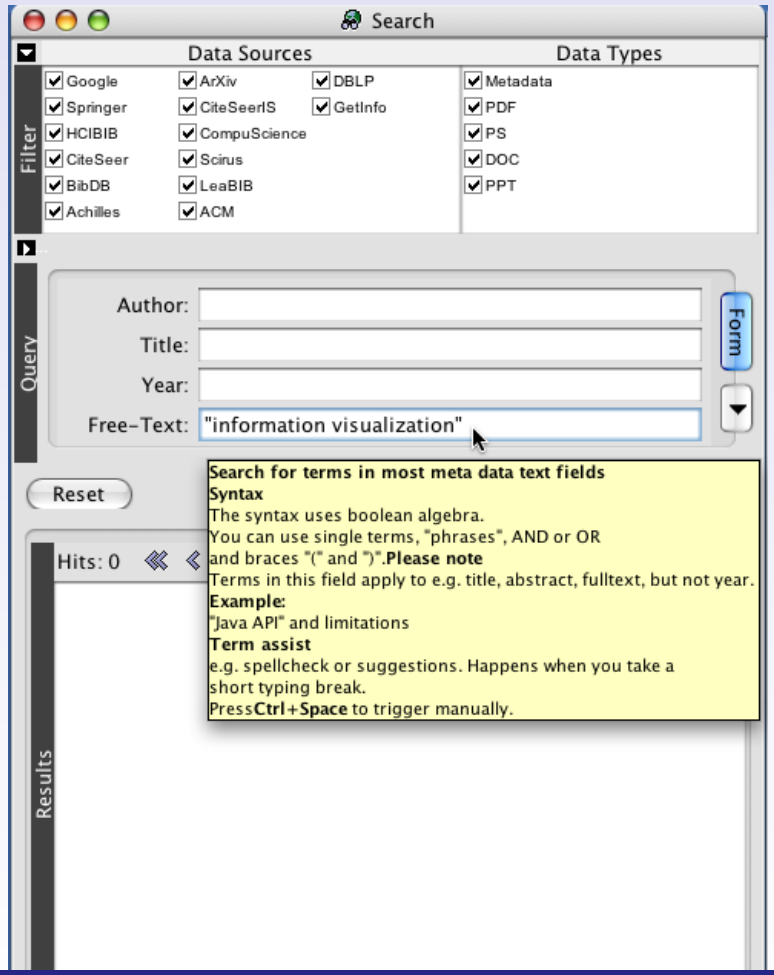
Goals for proactive support

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

Approach

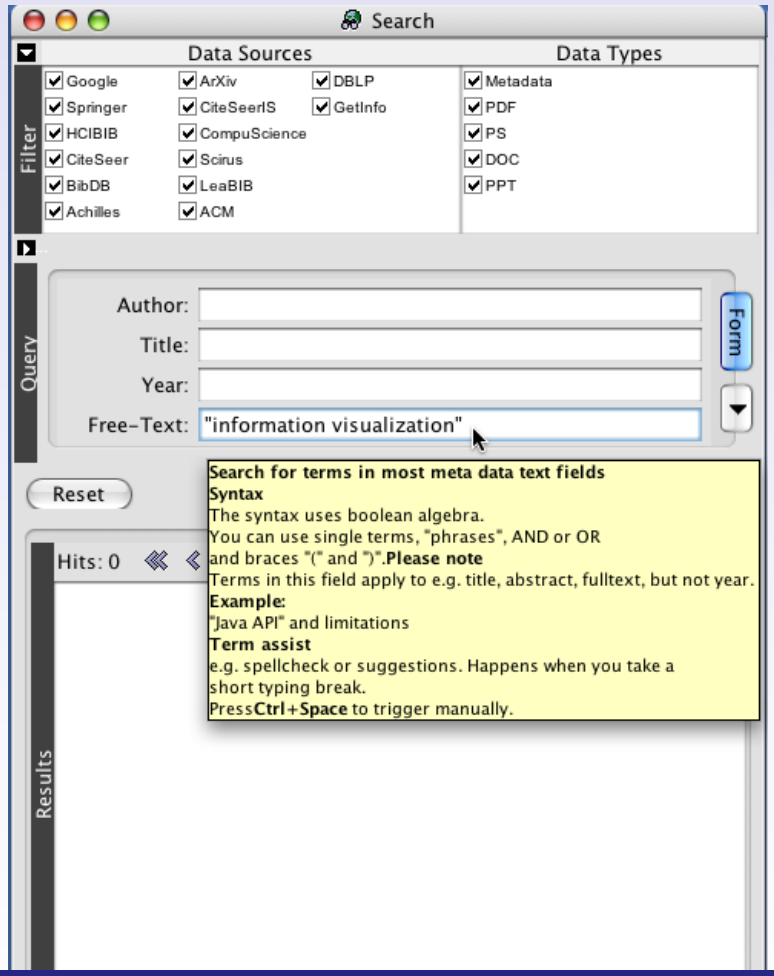
1. Test, which problems users face (User Problems)
2. Offer proactive support features in query formulation context
3. Current Solutions
4. Evaluate the improvements

Search
Easy to Use Search Interface



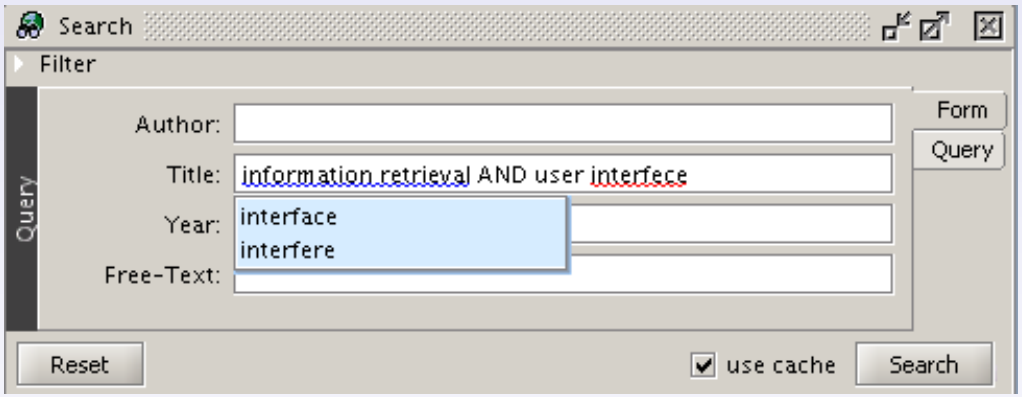
- Users are unsure about query syntax
- ▷ Offer precise help in context
- Many digital libraries – many query languages
- ▷ Homogenization
- Not enough - users still have problems

Search
Easy to Use Search Interface



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

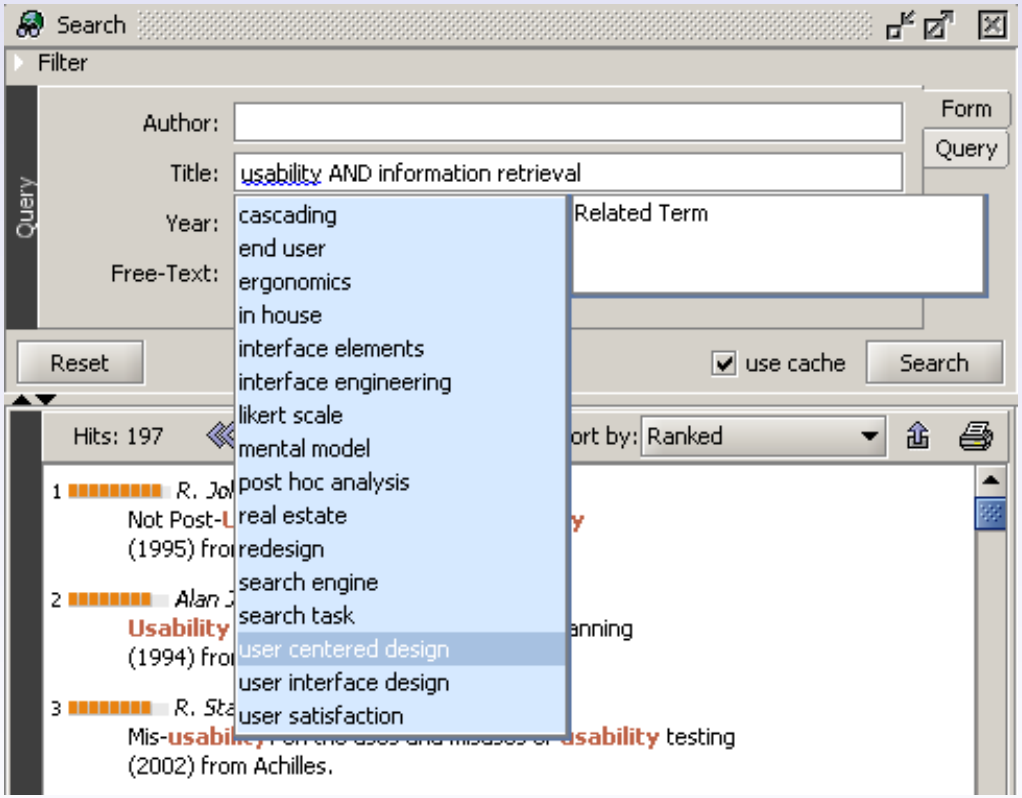
Active Support
Query Formulation



What do users expect?

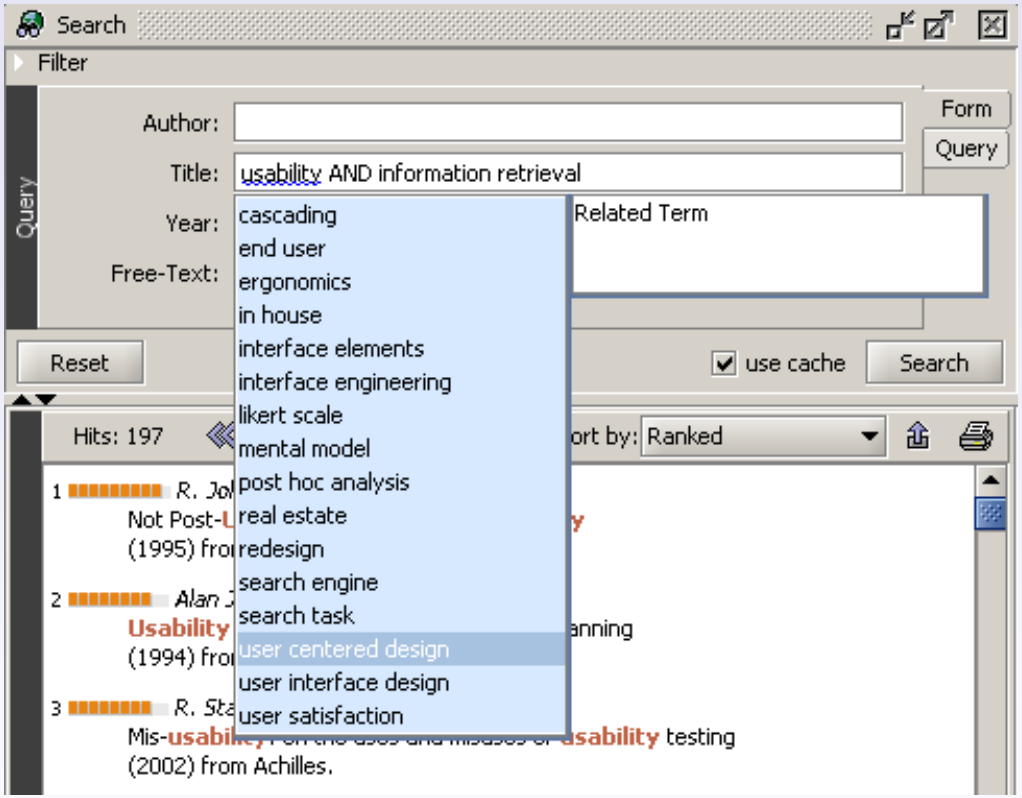
- Auto spell check
- Mark errors
- List of possible corrections
- Completion of known proper names (Authors)

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

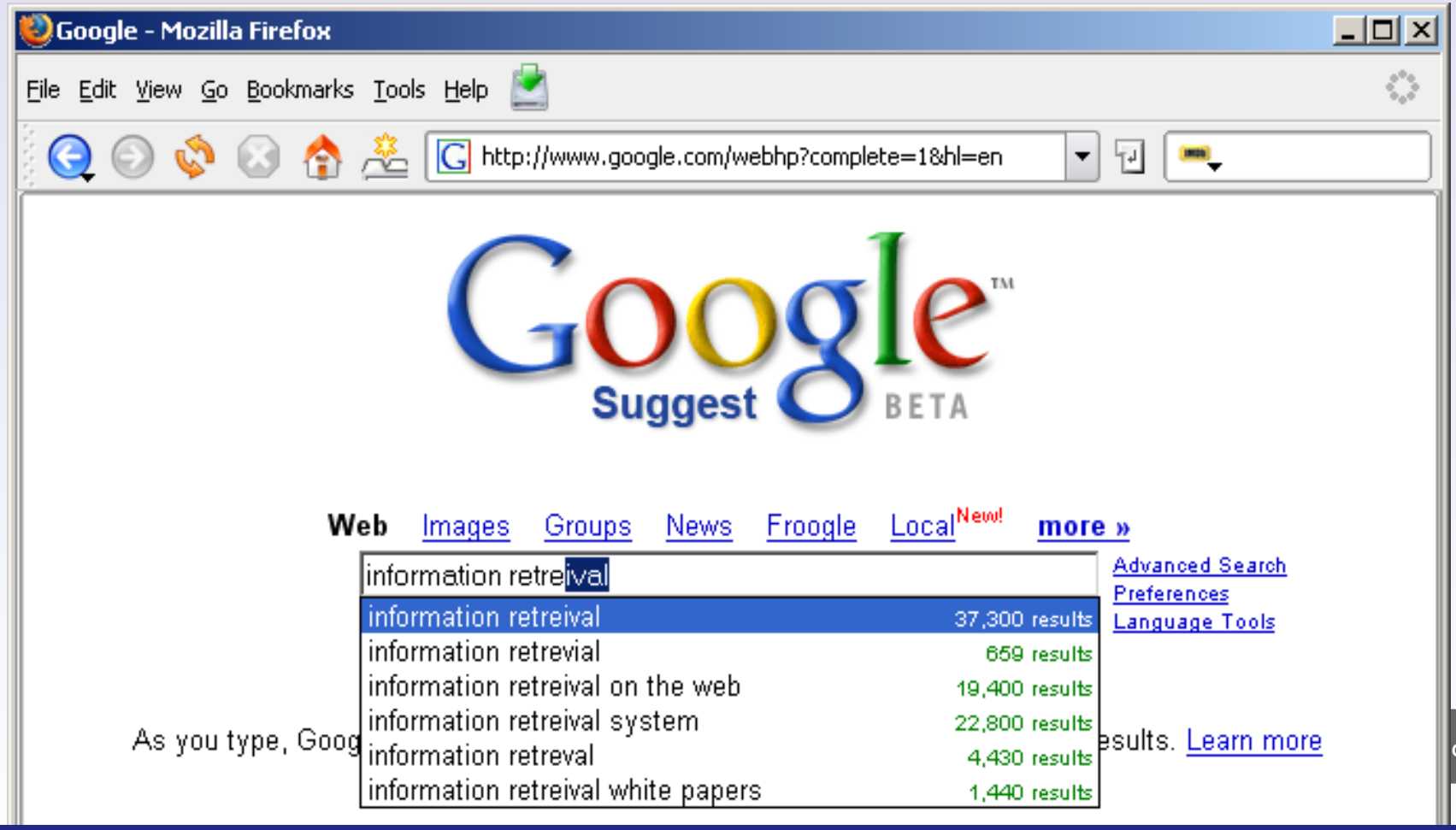


- User starts to search in ASK
  He is *insecure*.
- Synonyms and Homonyms are needed
- Offer Terms in Context

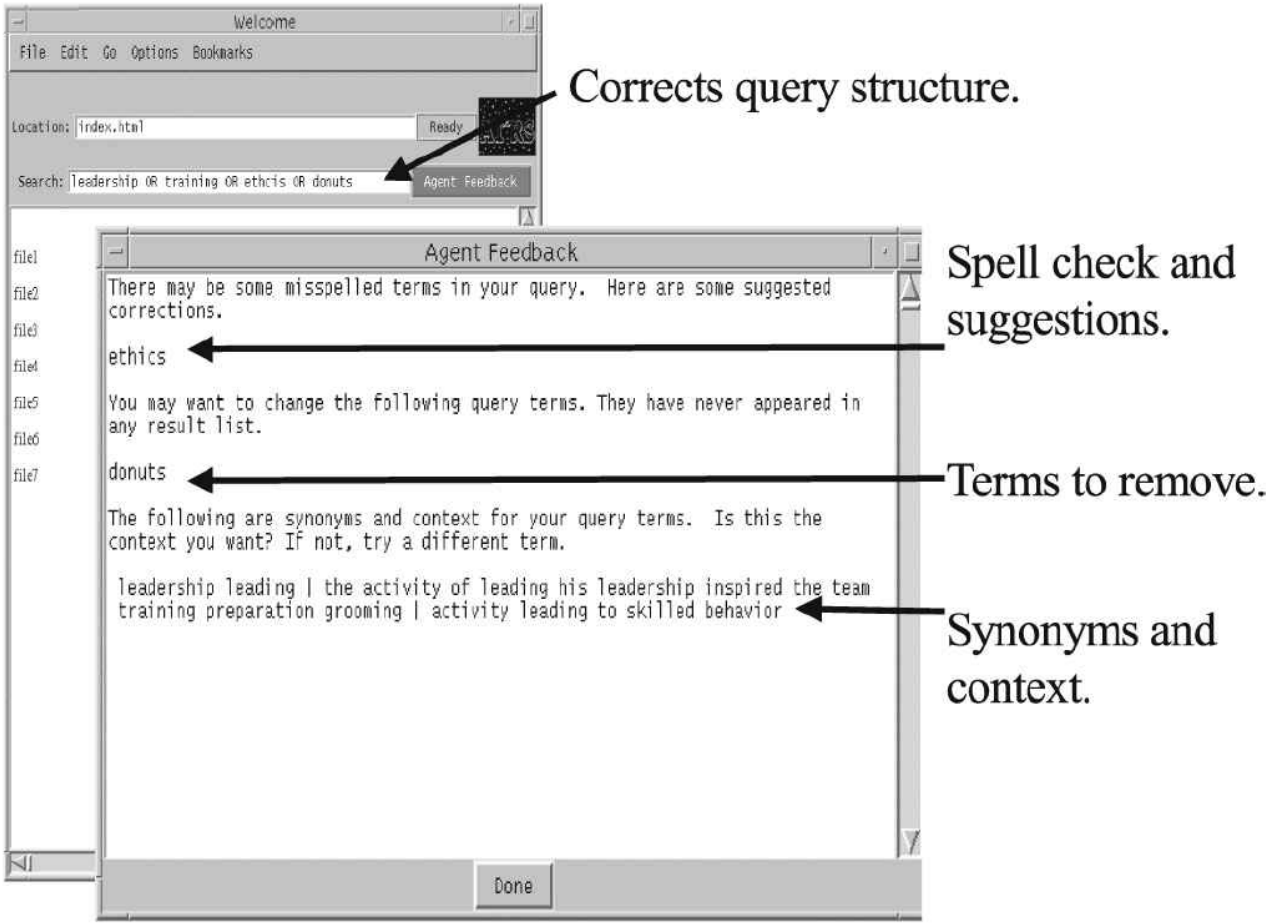
Status Quo

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

Google Suggest
(Suggests Mistakes by Design)



AI^2RS
Gives Improvement Hints



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