How do users formulate their queries?
A morpho-syntactic analysis

Nicolas Ariste Fairon

Life sciences library, University of Liege, 4000 Liège, Belgium
<nicolas.fairon@ulg.ac.be>
Queries formulated in French natural language

- Medline search strategy
- Natural Language Processing
- Automatic extraction of concepts

French MeSH
Despite the efforts, many users remain unable to perform an efficient Medline research. Why?

- Bad query formulation
- Bad knowledge of MeSH terms
- Not enough practice
- Problems with boolean operator
What exists

- Medline interfaces, with interesting features:
  - Query expansion
    - Searching MeSH and keywords
    - Automatic explosion...
  - Permutated index
  - MeSH translations
  - Elementary tools for natural language searching
Natural Language Approach

Analyzing the query to find relevant concepts

<table>
<thead>
<tr>
<th>Natural language</th>
<th>Precision</th>
<th>Recall</th>
<th>Controlled language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Torticollis</td>
<td>83.7%</td>
<td>100%</td>
<td>Torticollis [MeSH]</td>
</tr>
<tr>
<td>Congenital torticollis</td>
<td>40.0%</td>
<td>90.0%</td>
<td>Torticollis/cn [MeSH]</td>
</tr>
<tr>
<td>Smoking adverse effects</td>
<td>4.2%</td>
<td>44.1%</td>
<td>Smoking/ae [MeSH]</td>
</tr>
</tbody>
</table>

Medline interfaces
What we want to do

New Medline Interface

Tools
- Mesh's French Translation
- Natural Language tools

Output
- Search strategy
- Relevant results

Needs
- Searchable in French
- Useful for occasional users

Steps
- Extracting relevant concepts
- Combining relevant concepts
Materials & Methods

Query submitted by user → CORPUS: All queries → Corrected → Semantically tagged

Descriptive Analysis

Concepts extraction

Approaches:
- Dictionary
- Local grammar
- Hybrid
Je cherche des articles sur le traitement du cancer du sein.

Correcting

Je cherche des articles sur le traitement du cancer du sein.

Tagging

Je cherche des articles sur le {w11s*traitement*} du {w21*cancer du sein*}. 
Manual tagging

- To append semantic flags to useful concepts
- To identify and keep track of every concept
- To evaluate the efficiency of our application
The Corpus

- A web application to store for each query
  - Raw, corrected, and tagged versions
  - Medline search history done by a scientific librarian
- 195 queries formulated by 68 different users
- 6,985 words
Extracting concepts

- Descriptive
- Analysis
  - Concepts extraction
    - Dictionary
    - Hybrid
    - Local grammar
      - Dictionaries
        - French MeSH
        - Hand-made
      - Local grammars
Evaluation of automatic extraction

Introduction – Material & Methods – Results - Conclusions

Queries

Concepts extraction

Concepts

CORPUS
untagged

VS

List A

Recall

Precision

List B
(reference)

tagged
Descriptive analysis

464 concepts have been identified

- MeSH terms
- Keywords
- Fuzzy MeSH terms
- Out of MeSH synonyms
- Number inflection
Concepts' extraction: dictionary approach

- Applying MeSH dictionary to queries in order to identify them.
Concepts' extraction: Local grammar approach

- Use recognition patterns relying on queries' morphology and syntax.

![Bar chart showing recall and precision for MeSH terms, subheadings, and keywords.](chart.png)
Concepts' extraction: Hybrid approach

- Using local grammars combined with dictionaries

![Bar chart showing Recall and Precision for MeSH terms, Subheadings, and Keywords.](chart.png)
Conclusions

- Creating a new interface based on natural language processing involves
  - Concept mapping
  - Concepts combination
- Hybrid approach shows best results
  - Dictionaries
  - Local grammar
- Dictionaries' quality influes on performance
What's next?

- Disambiguation of fuzzy MeSH concepts
- Combination of the concepts with adequate boolean operators
- Made the tool available to users as a web application
Thank you for your attention

nicolas.fairon@ulg.ac.be

Open source tools used for the work and the presentation:

- OpenOffice.org
- Python
- MySQL