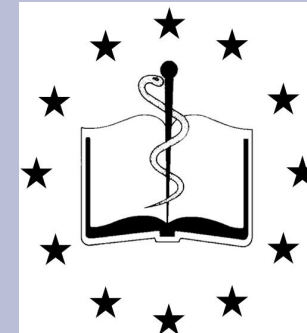


11th European Conference of Medical and Health Libraries



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

Nicolas Ariste Fairon

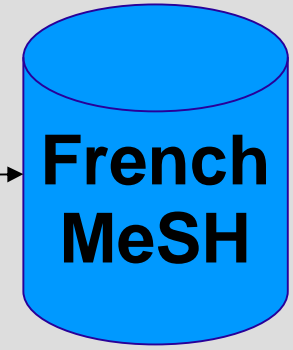
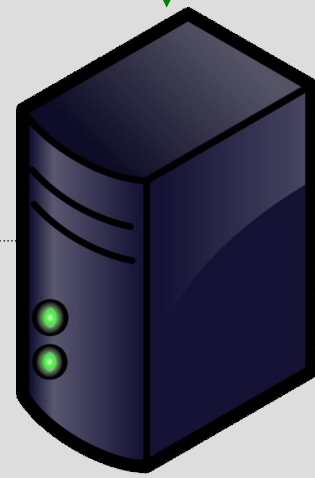
Life sciences library, University of Liege, 4000 Liège, Belgium
<nicolas.fairon@ulg.ac.be>



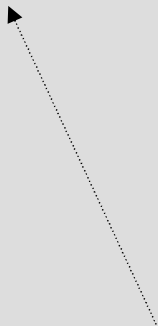
24th of June, Nicolas Fairon



Queries formulated in French natural language



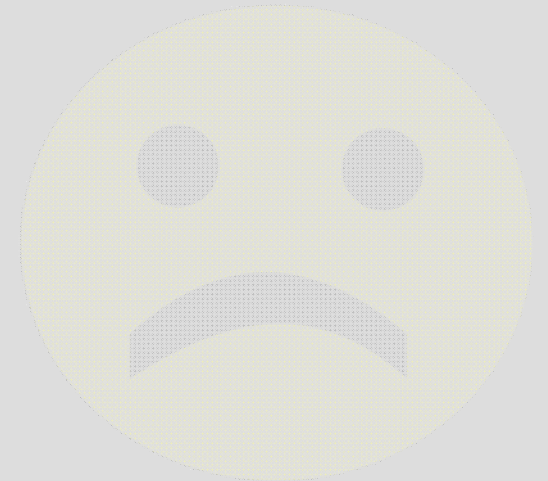
Medline search strategy



- Natural Language Processing
- Automatic extraction of concepts

The Facts

- 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...
 - Permuted index
 - MeSH translations
 - Elementary tools for natural language searching

Natural Language Approach

Analyzing the query to find relevant concepts

	Medline interfaces				
complexity	Natural language	Precision	Recall	Controlled language	Efficiency
	Torticollis	83.7%	100%	Torticollis [MeSH]	
	Congenital torticollis	40.0%	90.0%	Torticollis/cn [MeSH]	
	Smoking adverse effects	4.2%	44.1%	Smoking/ae [MeSH]	

What we want to do

New Medline Interface

Tools

Mesh's French Translation
Natural Language tools

Output

Search strategy
Relevant results

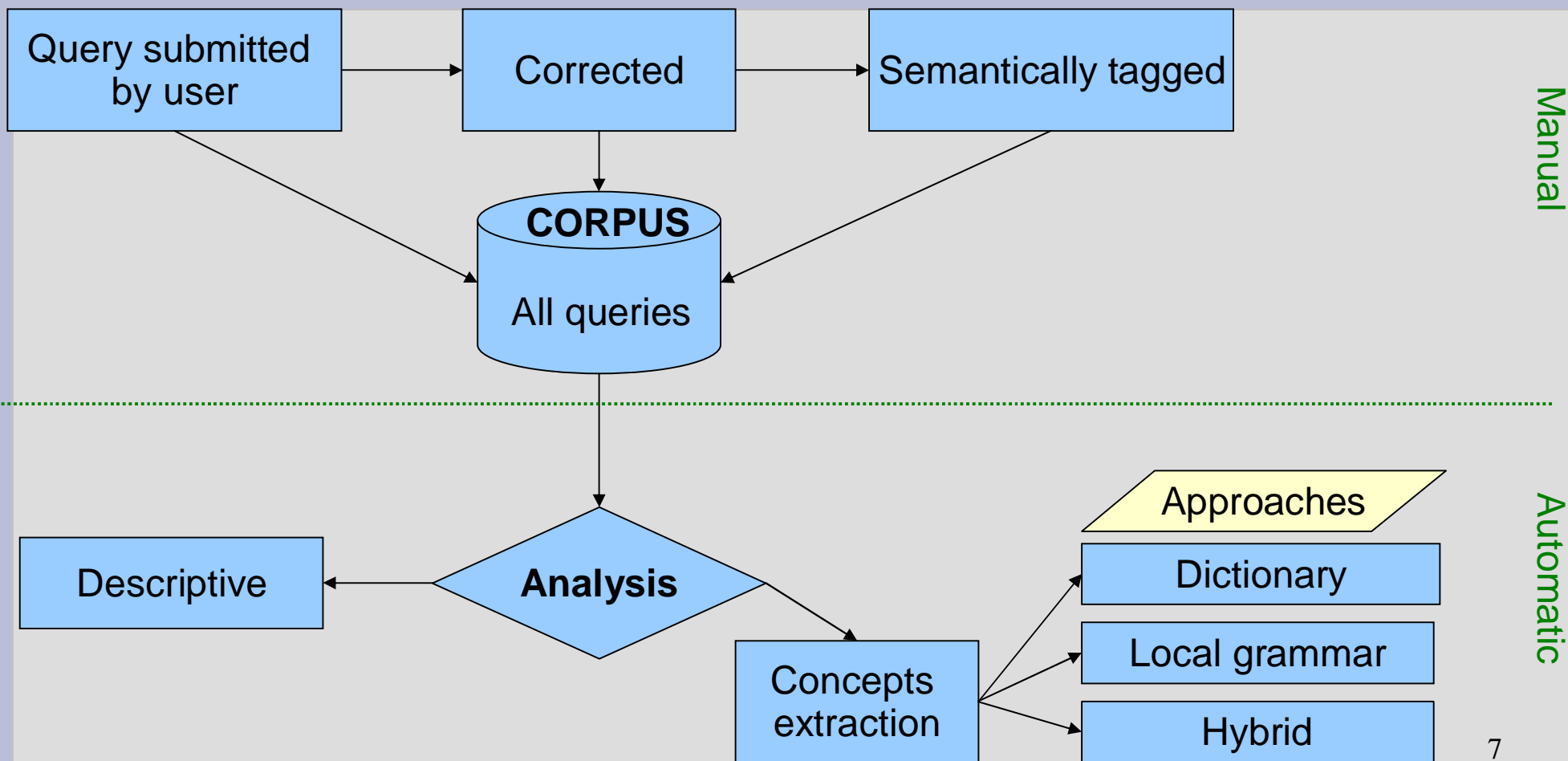
Needs

Searchable in French
Useful for occasional users

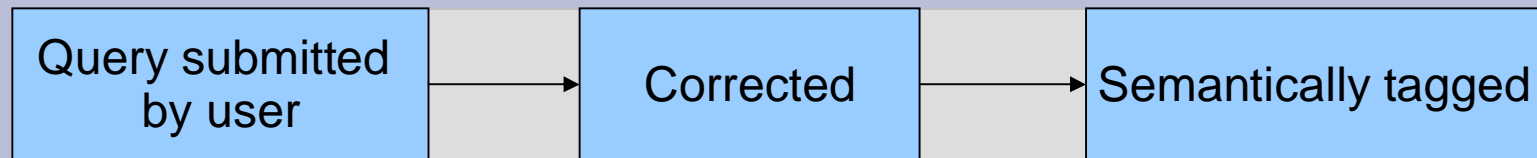
Steps

Extracting relevant concepts
Combining relevant concepts

Materials & Methods



Queries'collecting



Je cherche des articles sur le tré^étement du can^ser du sein.



Correcting

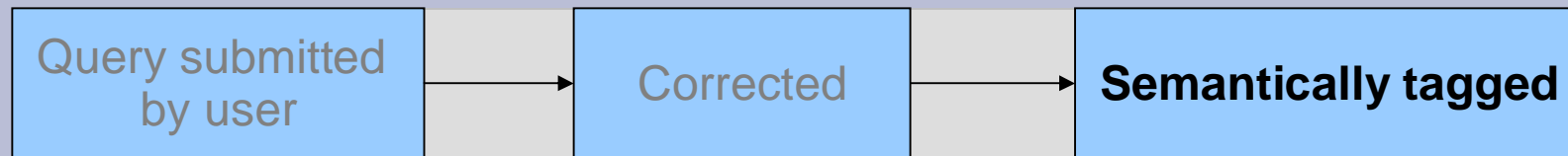
Je cherche des articles sur le traitement du cancer du sein.



Tagging

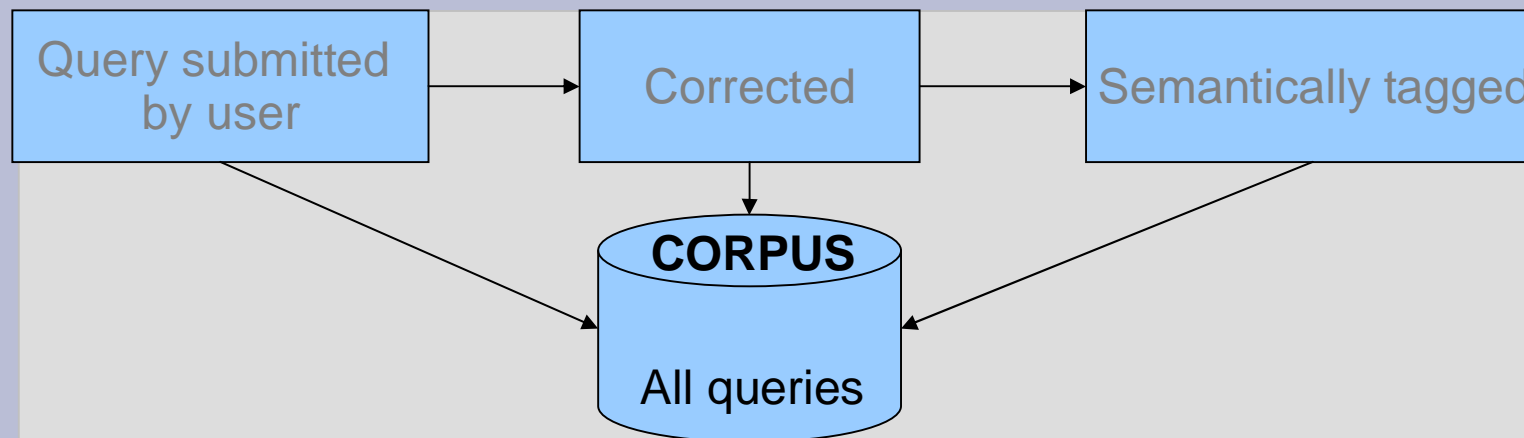
Je cherche des articles sur le {w11s*traitement*} du {w21*cancer du sein*}.

Manual tagging



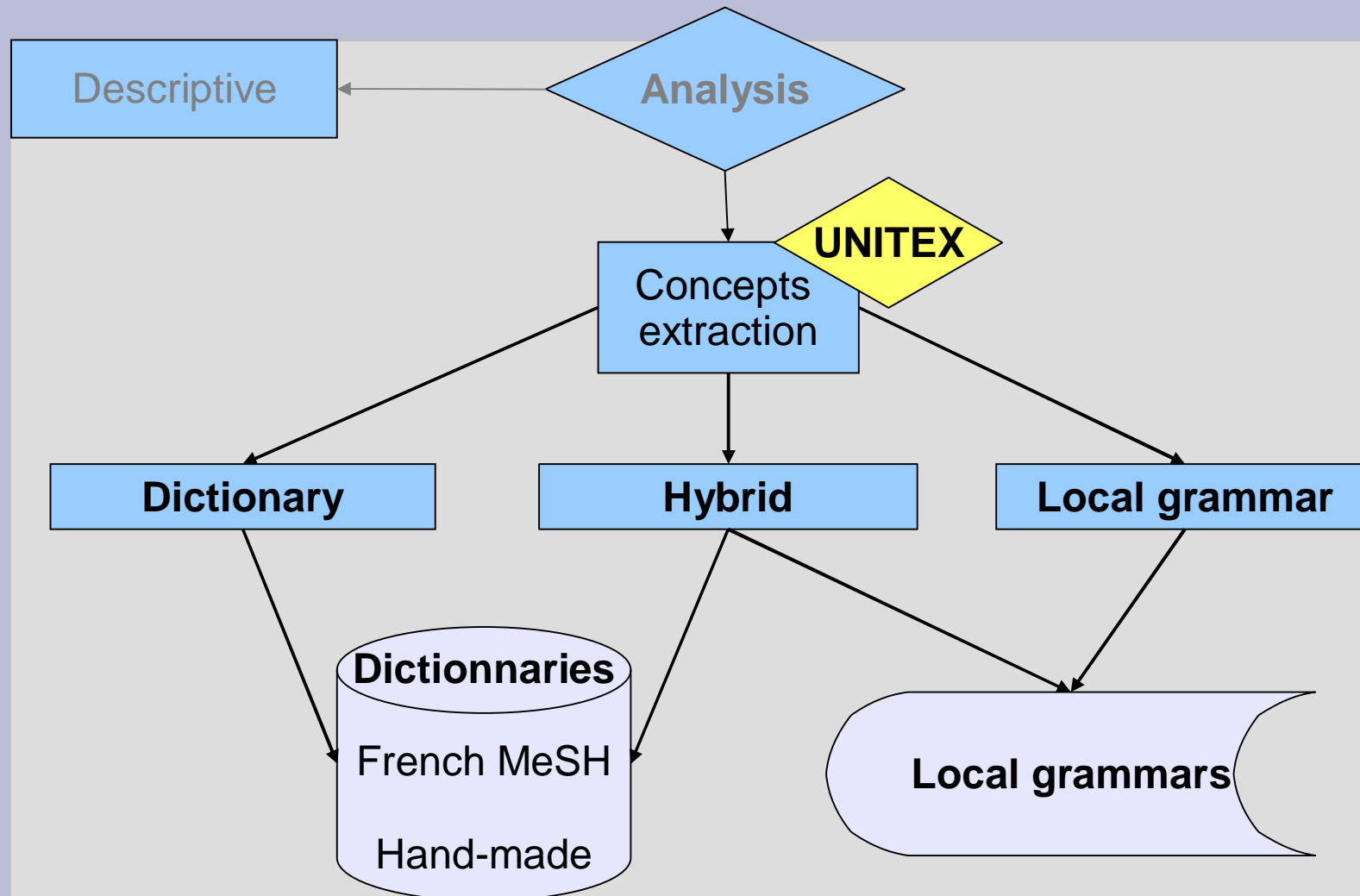
- To append semantic flags to useful concepts
- To identify and keep track of every concept
- To evaluate the efficiency of our application

The Corpus

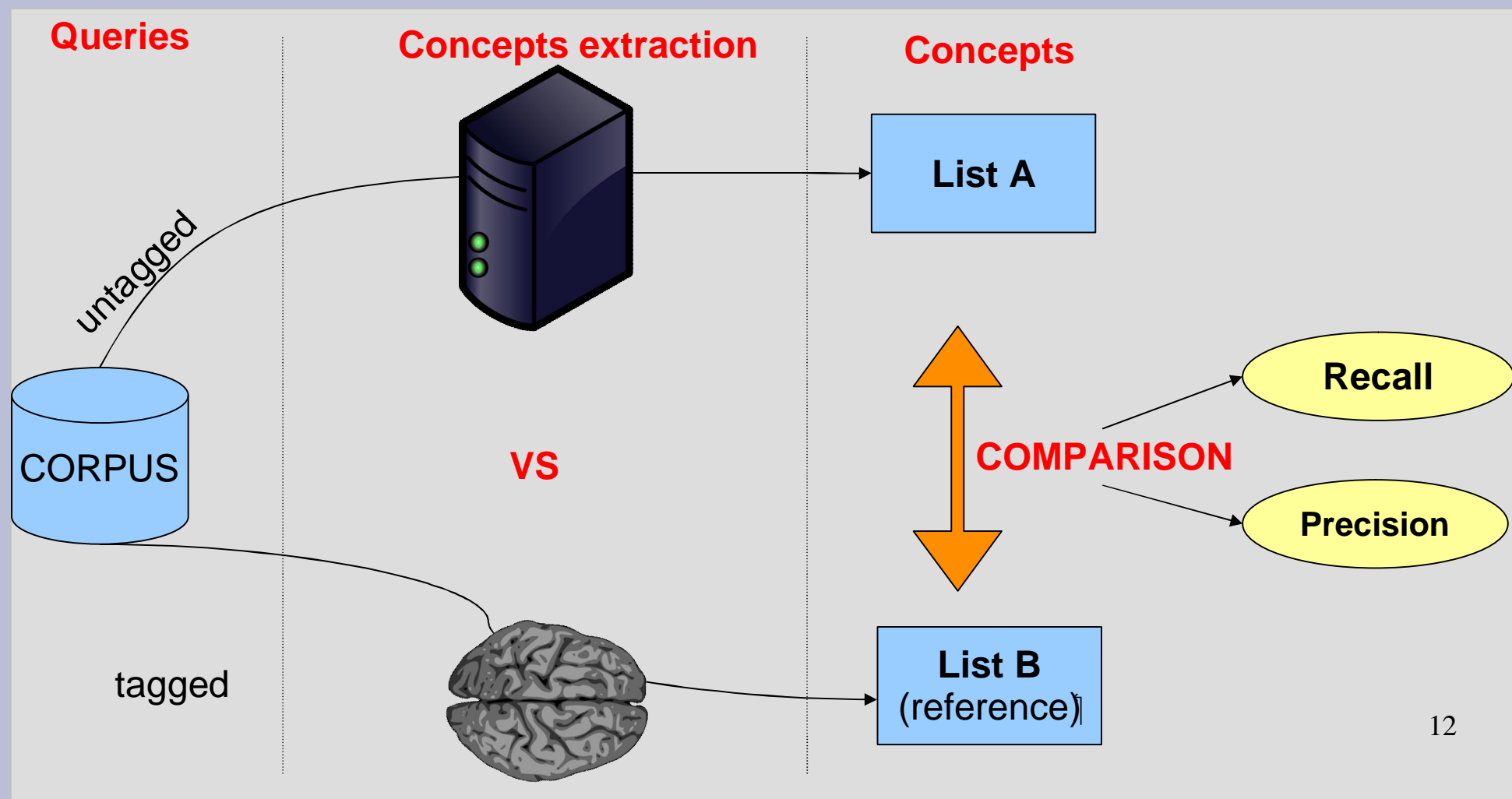


- A web application to store for each query
 - Raw, corrected, and tagged versions
 - Medline search history done by a scientific librarian
- 195 queries formulated by 68 different users
- 6 985 words

Extracting concepts

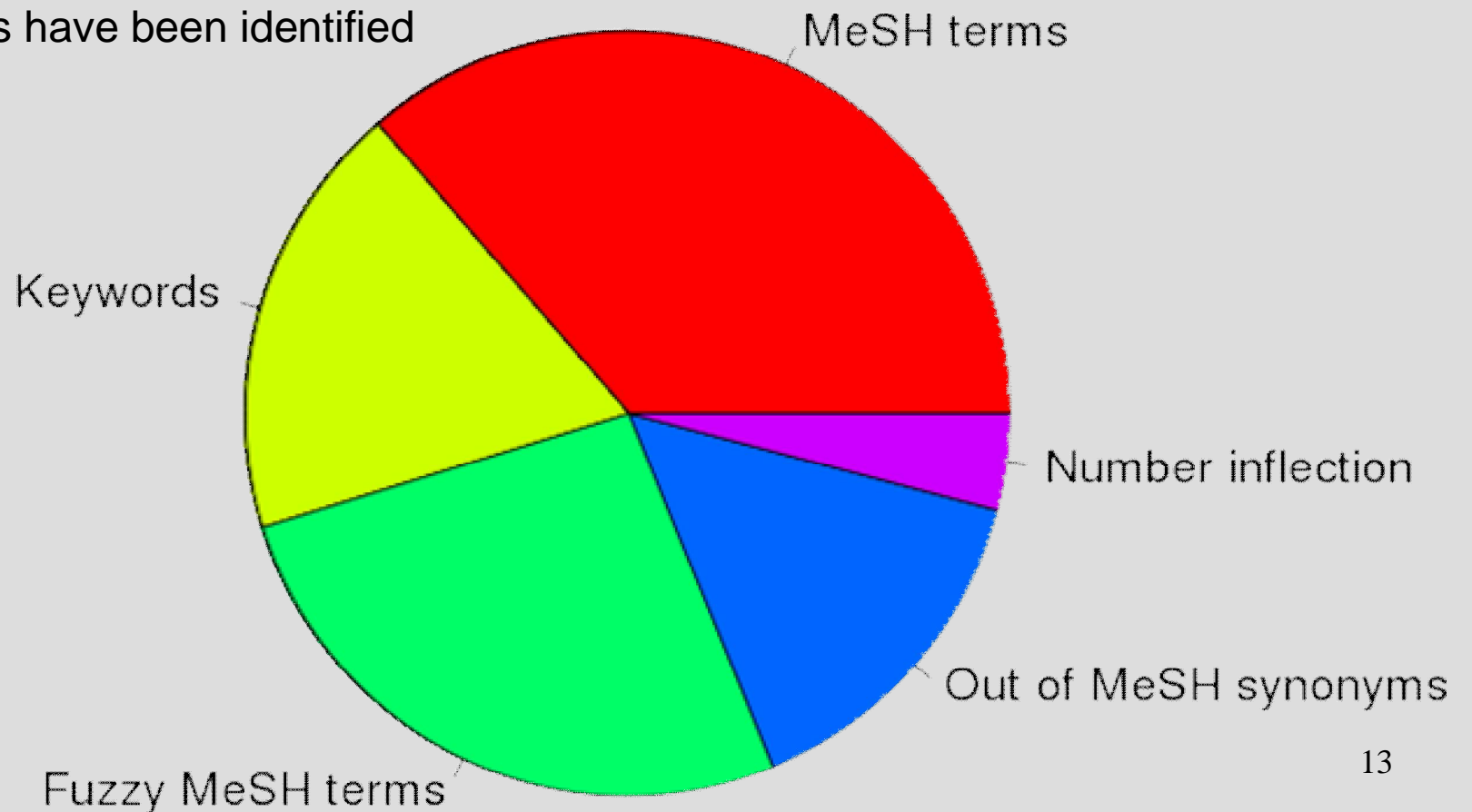


Evaluation of automatic extraction



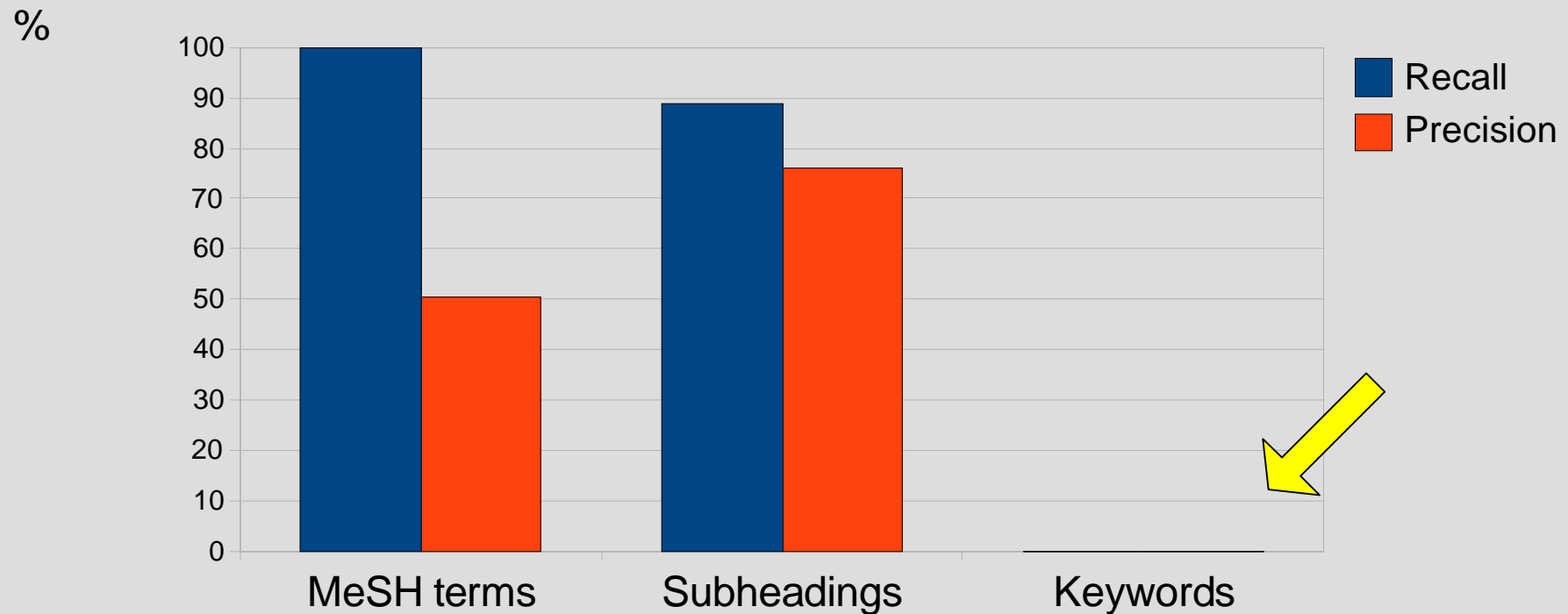
Descriptive analysis

464 concepts have been identified



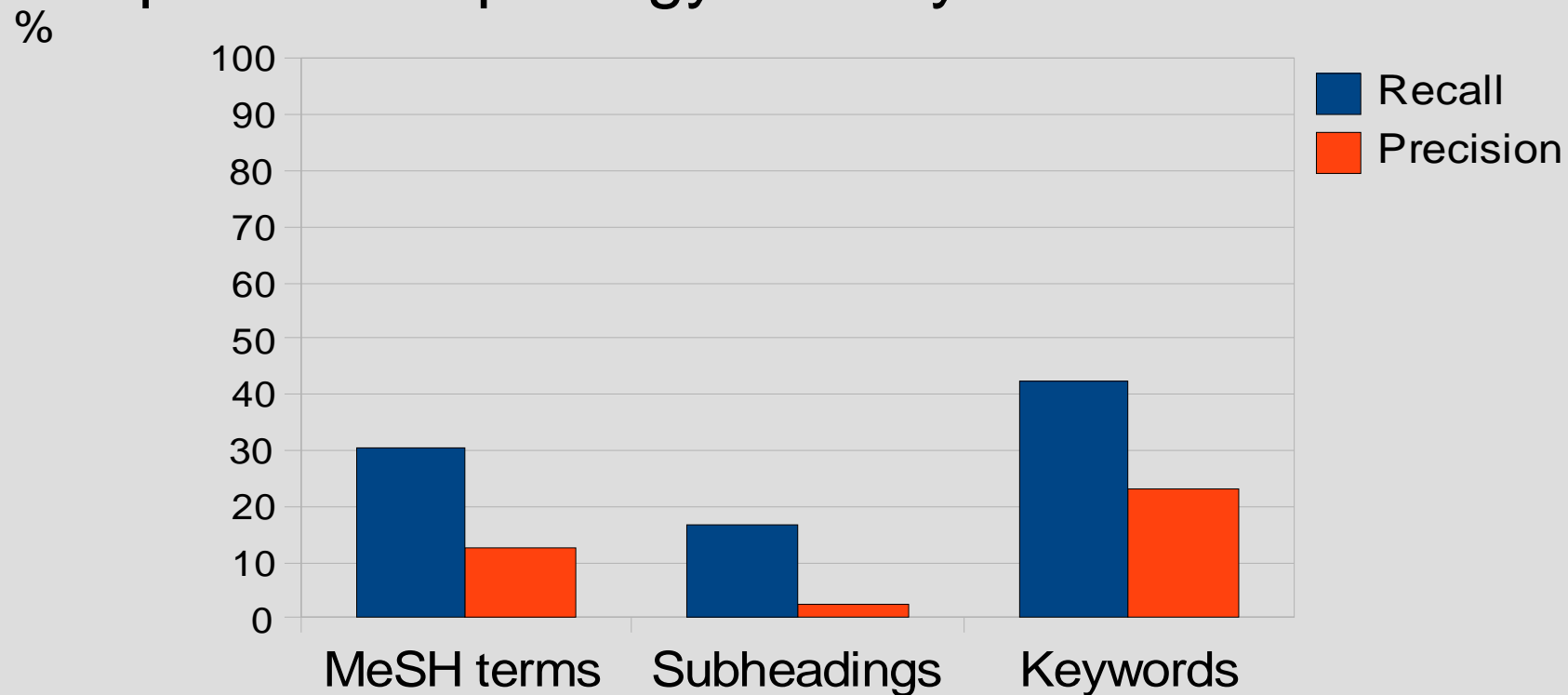
Concepts' extraction: dictionary approach

- Applying MeSH dictionary to queries in order to identify them.



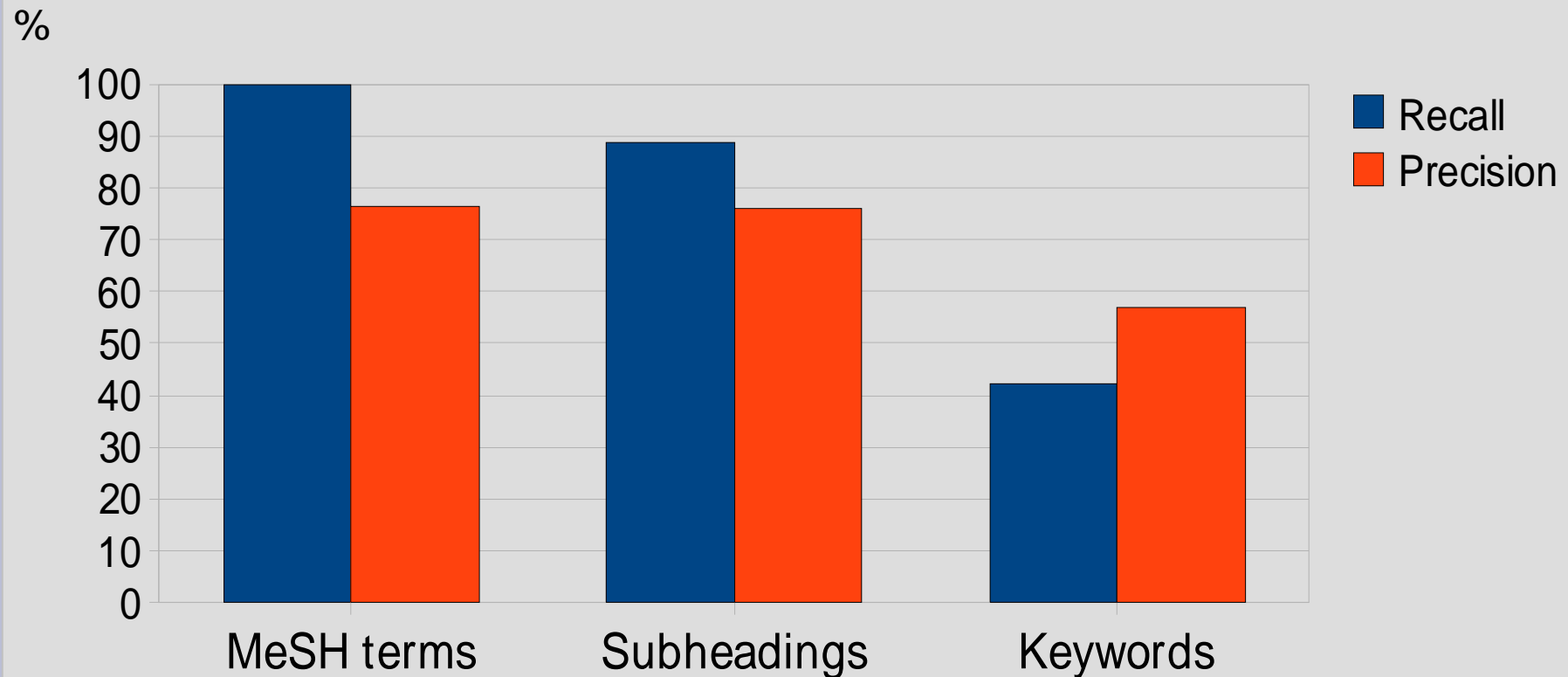
Concepts' extraction: Local grammar approach

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



Concepts' extraction: Hybrid approach

- Using local grammars combined with dictionaries



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 influences on performance

What's next?

- Disambiguation of fuzzy MeSH concepts
- Combination of the concepts with adequate booleans 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 :



python™



OpenOffice.org

