Mathematical Indexing and Querying

Martin Líška

<255768@mail.muni.cz>

6th October 2010

EuDML

The EUROPEAN DIGITAL MATHEMATICS LIBRARY

Contents

Motivation

- To allow users of a digital mathematics library to search the mathematical content
- Conventional search engines not applicable in the environment of a digital mathematics library
- Different rules for mathematics
  - different notation for equivalent formulae

Motivation

- To allow users of a digital mathematics library to search the mathematical content
- Conventional search engines not applicable in the environment of a digital mathematics library
- Different rules for mathematics
  - different notation for equivalent formulae

Motivation

- To allow users of a digital mathematics library to search the mathematical content
- Conventional search engines not applicable in the environment of a digital mathematics library
- Different rules for mathematics
  - different notation for equivalent formulae

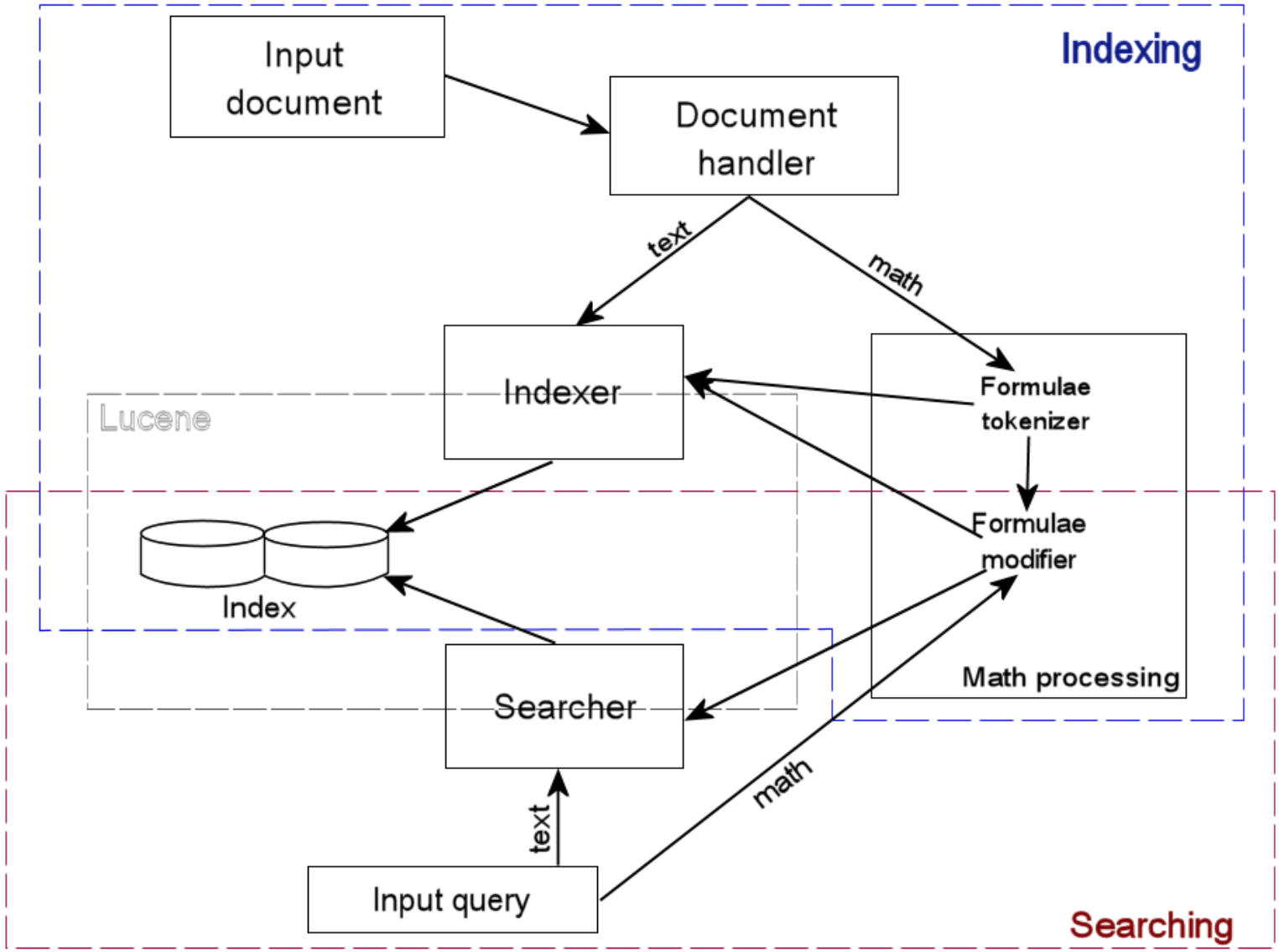
Math Indexer and Searcher – features

- Indexes xhtml, html and txt documents with MathML notation
- Searching:
  - Exact mathematical formulae
  - Similar formulae (unified variables, constants, both)
  - Subformulae
  - Mixed mathematical-textual
  - Relevancy calculation

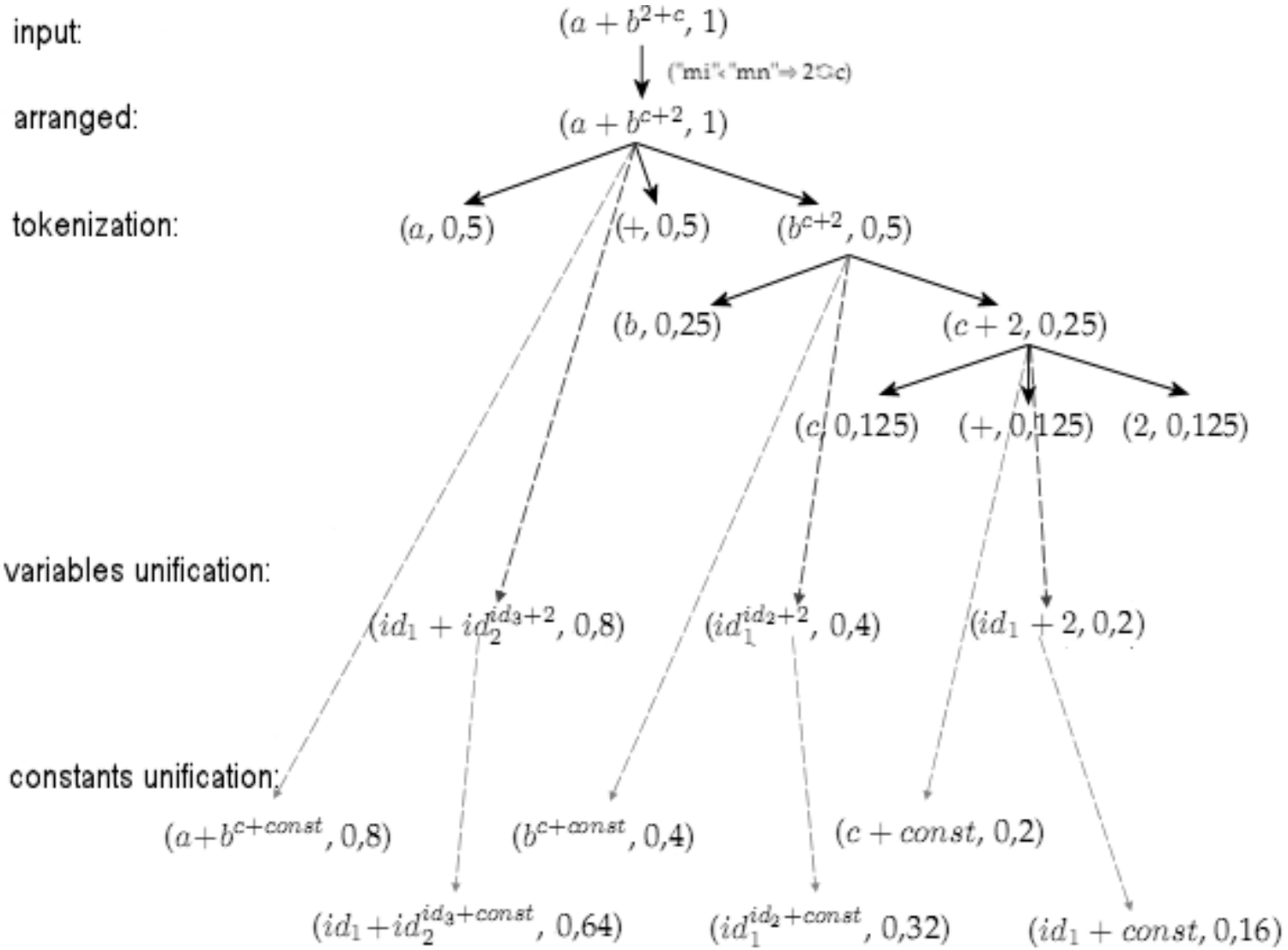
Math Indexer and Searcher – features

- Indexes xhtml, html and txt documents with MathML notation
- Searching:
  - Exact mathematical formulae
  - Similar formulae (unified variables, constants, both)
  - Subformulae
  - Mixed mathematical-textual
  - Relevancy calculation

Math Indexer and Searcher – design



Formula processing example



Testing

- Almost 10 000 documents – arXMLiv with MathML
  - Over 1 million input formulae
  - Produced almost 16 million
- Scalability test prooved linear character of the system depending on the number of documents
- Simple demo web interface: WebMIaS

Testing

- Almost 10 000 documents – arXMLiv with MathML
  - Over 1 million input formulae
  - Produced almost 16 million
- Scalability test prooved linear character of the system depending on the number of documents
- Simple demo web interface: WebMIaS

Testing

- Almost 10 000 documents – arXMLiv with MathML
  - Over 1 million input formulae
  - Produced almost 16 million
- Scalability test prooved linear character of the system depending on the number of documents
- Simple demo web interface: WebMIaS

Testing

- Almost 10 000 documents – arXMLiv with MathML
  - Over 1 million input formulae
  - Produced almost 16 million
- Scalability test prooved linear character of the system depending on the number of documents
- Simple demo web interface: WebMIaS

Conclusion

- Further testing
- Investigation of ranking function
- Further formulae unifications

Questions?