Olelo: A Question Answering Application for Biomedicine

Mariana Neves, Hendrik Folkerts, Marcel Jankrift, Julian Niedermeier, Toni Stachewicz, Sören Tietböhl, Milena Kraus, Matthias Uflacker

Hasso Plattner Institute at University of Potsdam, Germany

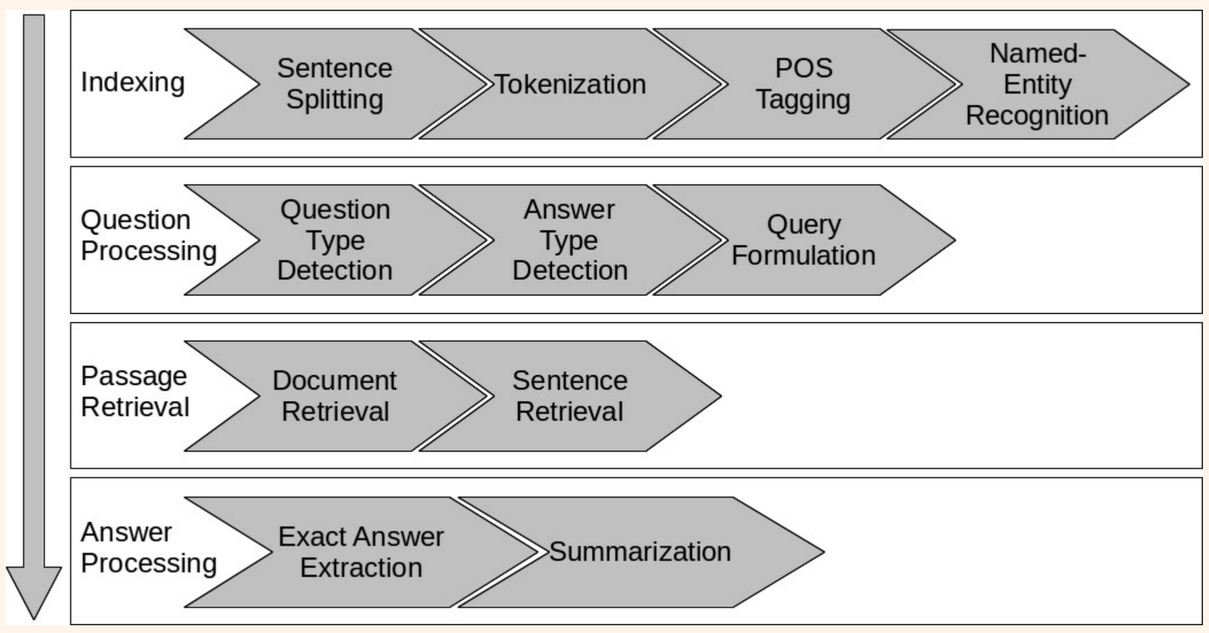
Olelo - Question Answering for Biomedicine

- Despite the importance of the biomedical domain, there are few reliable applications to support researchers and physicians to retrieve specific information or facts from the large corpus of literature.
- We present Olelo, a question answering system for biomedicine. It is built on top of an in-memory database (IMDB), integrates domain resources, such as document collections and terminologies, and uses various NLP components.
- We evaluated Olelo on two use cases: answering questions related to a particular gene and on the BioASQ benchmark.

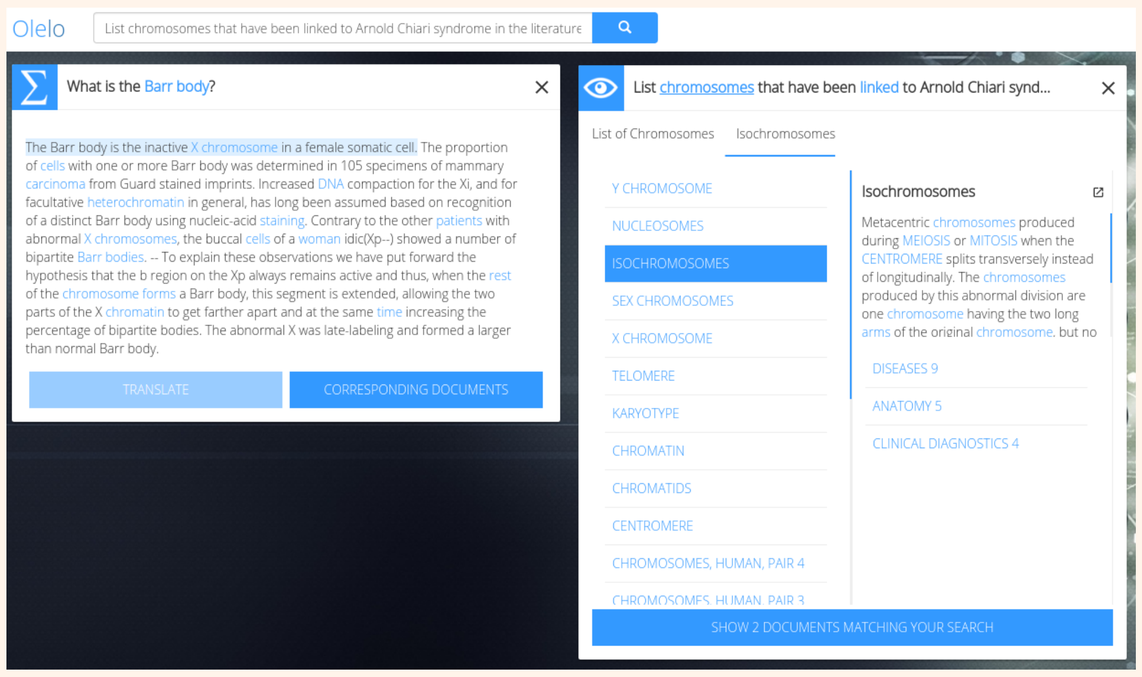
http://hpi.de/plattner/olelo

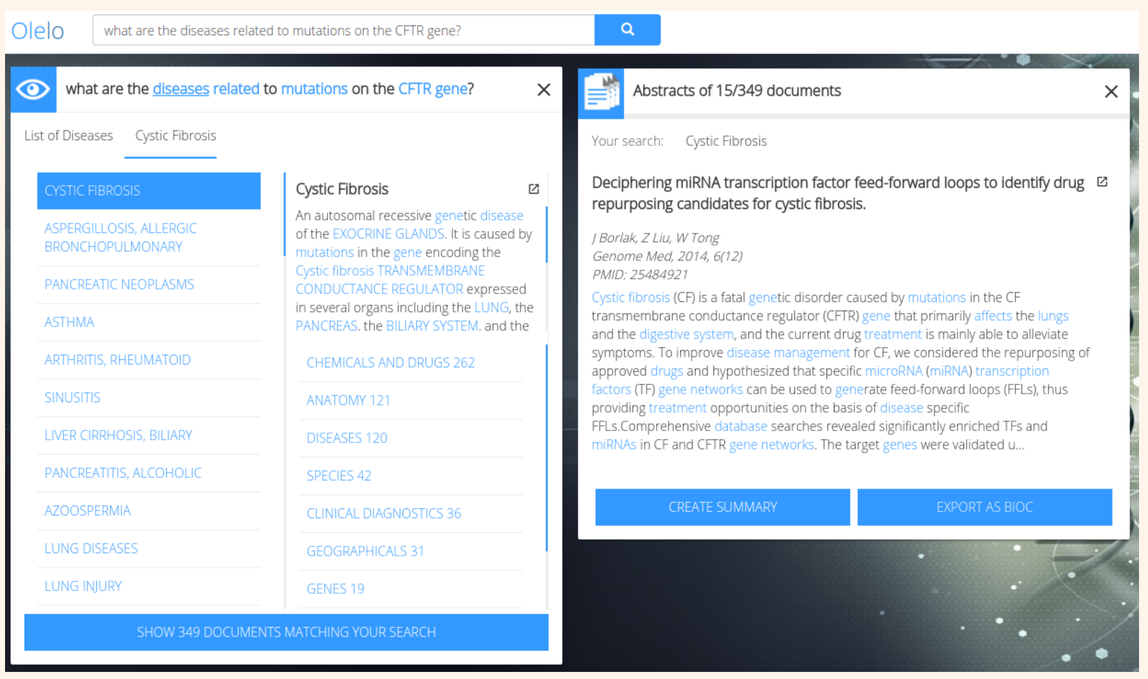
Architecture & Methods

- **Indexing**: We index Medline, PMC OA, MeSH and UMLS into an IMDB (SAP HANA), which includes sentence splitting, tokenization, stemming, part-of-speech (POS) tagging and NER.
- **Question processing**: Olelo currently supports three types of questions: (i) factoid, (ii) definition, and (iii) summary. It detects the question type via simple regular expressions, followed by the detection of the answer type, in the case of factoid questions.
- **Passage retrieval**: The system ranks documents and passages based on built-in features of the IMDB. It matches keywords from the query to the documents in an approximate way, including linguistic variations.
- **Answer processing**: The system simply shows the corresponding MeSH term for definition questions; returns MeSH terms which belong to the corresponding semantic type for factoid questions; and builds a customized summary for summary questions.



Use cases: Genomics (left) and BioASQ (right)





Related publications:

- Neves M, Eckert F, Folkerts H and Uflacker M. **Assessing the performance of Olelo, a real-time biomedical question answering application**, Biomedical Natural Language Processing (BioNLP) Workshop at ACL'17, Vancouver, Canada.
- Schulze F and Neves M. **Entity-Supported Summarization of Biomedical Abstracts**, Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining, Coling 2016, Osaka, Japan.
- Neves M and Kraus M. **BioMedLAT Corpus: Annotation of the Lexical Answer Type for Biomedical Questions**, Proceedings of the Open Knowledge Base and Question Answering Workshop at Coling 2016, pp. 49-58, Osaka, Japan.

HPI Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Contact

e-mail: mariana.neves@hpi.de