

# 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

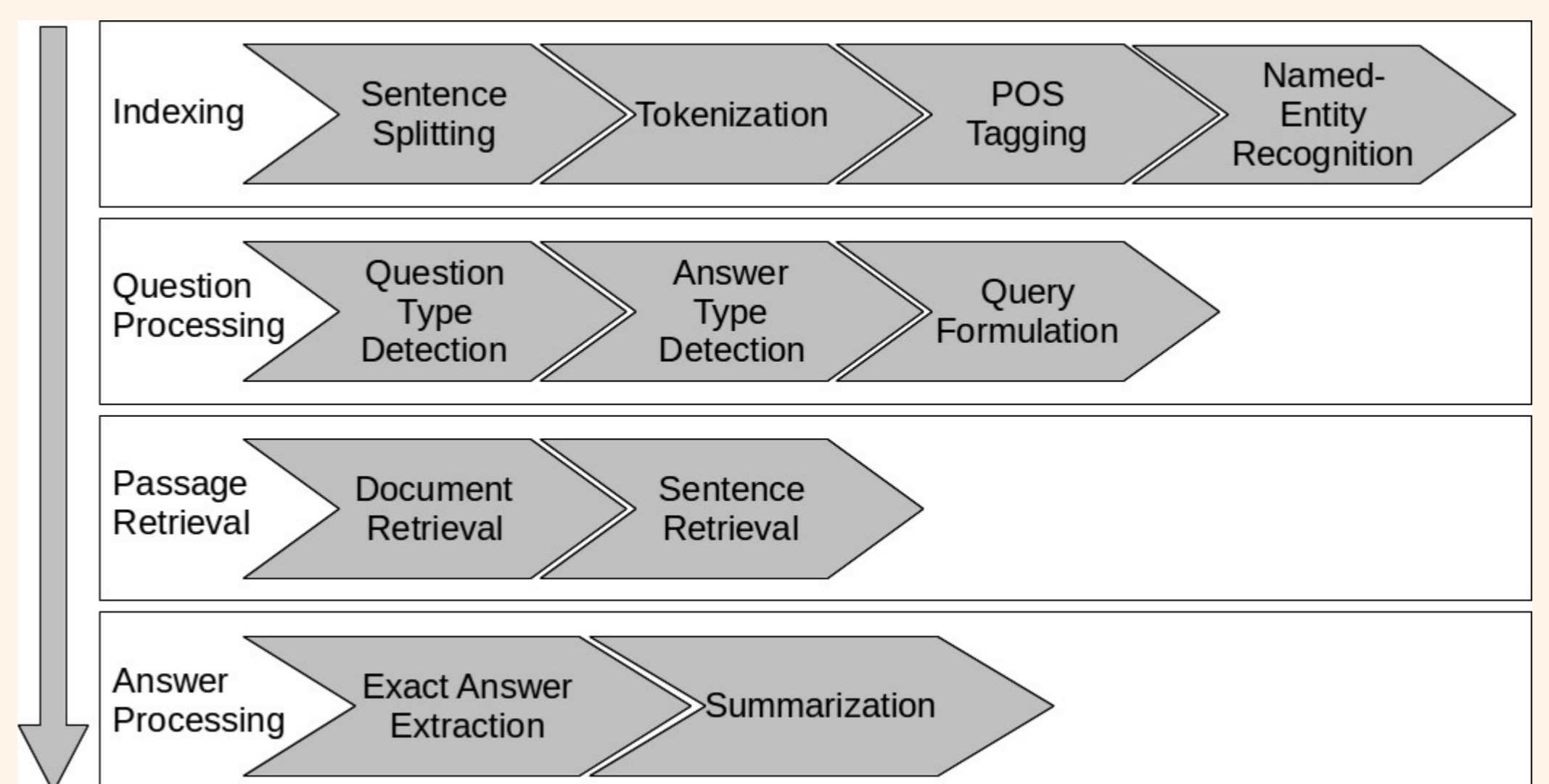
## Olelo - Question Answering for Biomedicine

- 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)

Olelo interface showing a search query: "List chromosomes that have been linked to Arnold Chiari syndrome in the literature". The results are displayed in a table with columns for "List of Chromosomes" and "Isochromosomes". The "List of Chromosomes" column includes Y CHROMOSOME, NUCLEOSOMES, ISOCHROMOSOMES, SEX CHROMOSOMES, X CHROMOSOME, TELOMERE, KARYOTYPE, CHROMATIN, CHROMATIDS, CENTROMERE, and CHROMOSOMES, HUMAN, PAIR 4. The "Isochromosomes" column includes a detailed description: "Metacentric chromosomes produced during MEIOSIS or MITOSIS when the CENTROMERE splits transversely instead of longitudinally. The chromosomes produced by this abnormal division are one chromosome having the two long arms of the original chromosome, but no...".

Olelo interface showing a search query: "what are the diseases related to mutations on the CFTR gene?". The results are displayed in a table with columns for "List of Diseases" and "Cystic Fibrosis". The "List of Diseases" column includes CYSTIC FIBROSIS, ASPERGILLOSIS, ALLERGIC BRONCHOPULMONARY, PANCREATIC NEOPLASMS, ASTHMA, ARTHRITIS, RHEUMATOID, SINUSITIS, LIVER CIRRHOSIS, BILIARY, PANCREATITIS, ALCOHOLIC, AZOOSPERMIA, LUNG DISEASES, and LUNG INJURY. The "Cystic Fibrosis" column includes a detailed description: "An autosomal recessive genetic disease of the EXOCRINE GLANDS. It is caused by mutations in the gene encoding the Cystic fibrosis TRANSMEMBRANE CONDUCTANCE REGULATOR expressed in several organs including the LUNG, the PANCREAS, the BILIARY SYSTEM, and the...".

### 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.