Olelo: A Question Answering Application for Biomedicine

<u>Mariana Neves</u>, 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	C List chromosomes that have been linked to Arnold Chiari syndrome in the literature				Olelo what are the diseases related to mutations on the CFTR gene?				
N Wh	at is the Barr body?	List <u>chromosomes</u> that have been linked to Arnold Chiari synd X			what are the <u>diseases</u> related to mutations on the CFTR gene?			Abstracts of 15/349 documents X	
of cells with carcinoma fi facultative h of a distinct abnormal X bipartite Bar hypothesis t of the chron parts of the percentage	body is the inactive X chromosome in a female somatic cell. The proportion ith one or more Barr body was determined in 105 specimens of mammary a from Guard stained imprints. Increased DNA compaction for the Xi, and for the heterochromatin in general, has long been assumed based on recognition net Barr body using nucleic-acid staining. Contrary to the other patients with al X chromosomes, the buccal cells of a woman idic(Xp) showed a number of Barr bodies To explain these observations we have put forward the sis that the b region on the Xp always remains active and thus, when the rest romosome forms a Barr body, this segment is extended, allowing the two the X chromatin to get farther apart and at the same time increasing the ge of bipartite bodies. The abnormal X was late-labeling and formed a larger mal Barr body.	List of Chromosomes Isochromoso	omes	List of Disea	ases Cystic Fibrosis		You	ir search: Cystic Fibrosis	
		Y CHROMOSOME	lsochromosomes 🛛	CYSTIC F	FIBROSIS	Cystic Fibrosis 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		ciphering miRNA transcription factor feed-forward loops to identify drug ourposing candidates for cystic fibrosis.	
		NUCLEOSOMES	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		ILLOSIS, ALLERGIC HOPULMONARY		j Bo	<i>J Borlak, Z Liu, W Tong</i> <i>Genome Med, 2014, 6(12)</i> <i>PMID: 25484921</i> Cystic fibrosis (CF) is a fatal genetic disorder caused by mutations in the CF transmembrane conductance regulator (CFTR) gene that primarily affects the lungs	
		SEX CHROMOSOMES			ATIC NEOPLASMS		Cyst		
		X CHROMOSOME					and	the digestive system, and the current drug treatment is mainly able to alleviate approximation of the system of the system and the current drug treatment is mainly able to alleviate approximation of the system of	



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.



IT Systems Engineering | Universität Potsdam



e-mail: mariana.neves@hpi.de