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
August-Bebel-Strasse 88, Potsdam 14482 Germany
`mariana.neves@hpi.de, milena.kraus@hpi.de`

Abstract

Despite the importance of the biomedical domain, there are few reliable applications to support researchers and physicians for retrieving particular facts that fit their needs. Users typically rely on search engines that only support keyword- and filter-based searches. We present Olelo, a question answering system for biomedicine. Olelo is built on top of an in-memory database, integrates domain resources, such as document collections and terminologies, and uses various natural language processing components. Olelo is fast, intuitive and easy to use. We evaluated the systems on two use cases: answering questions related to a particular gene and on the BioASQ benchmark.

Olelo is available at: `http://hpi.de/plattner/olelo`.

1 Introduction

Biomedical researchers and physicians regularly query the scientific literature for particular facts, e.g., a syndrome caused by mutations on a particular gene or treatments for a certain disease. For this purposes, users usually rely on the PubMed search engine[1], which indexes millions of publications available in the Medline database. Similar to classical information retrieval (IR) systems, input to PubMed is usually in the form of keywords, and alternatively MeSH concepts, and output is usually a list of documents.

For instance, when searching for diseases which could be caused by mutations on the CFTR gene, the user would simply write the gene name in PubMed's input field. For this example, he would be presented with a list of 9227 potentially relevant publications (as of February/2017).

There are plenty of other Web applications for searching and navigating through the scientific biomedical literature, as surveyed in (Lu, 2011). However, most of these systems rely on simple natural language processing (NLP) techniques, such as tokenization and named-entity recognition (NER). Their functionalities are restricted to ranking documents with the support of domain terminologies, enriching publications with concepts and clustering similar documents.

Question answering (QA) can support biomedical professionals by allowing input in the form of natural questions and by providing exact answers and customized short summaries in return (Athenikos and Han, 2010; Neves and Leser, 2015). We are aware of three of such systems for biomedicine (cf. Section 2), however, current solutions still fail to fulfill the needs of users: (i) In most of them, no question understanding is carried out on the questions. (ii) Those that do make use of more complex NLP techniques (e.g., HONQA (Cruchet et al., 2009)) cannot output answers in real time. (iii) The output is usually in the form of a list of documents, instead of short answers. (iv) They provide no innovative or NLP-based means to further explore the scientific literature.

We present Olelo, a QA system for the biomedical domain. It indexes biomedical abstracts and full texts, relies on a fast in-memory database (IMDB) for storage and document indexing and implements various NLP procedures, such as domain-specific NER, question type detection, answer type detection and answer extraction. We evaluated the methods behind Olelo in the scope of the BioASQ challenge (Tsatsaronis et al., 2015), the most comprehensive shared task on biomedical QA. We participated in the last three challenges and obtained top results for snippets retrieval and

---

[1]`http://www.ncbi.nlm.nih.gov/pubmed`

ideal answers (customized summaries) in the last two editions (Neves, 2014, 2015; Schulze et al., 2016).

Olelo provides solutions for the shortcomings listed above: (i) It detects both the question type and answer type. (ii) It includes various NLP components and outputs answers in real time (cf. Section 5). (iii) It always outputs a short answer, either exact answers or short summaries, while also allowing users to explore the corresponding documents. (iv) Users can navigate through the answers and their corresponding semantic types, check MeSH definition for terms, create document collections, generate customized summaries and query for similar documents, among other tasks. Finally, Olelo is an open-access system and no login is required. We tested it in multiple Web browsers, but we recommend Chrome for optimal results.

2 Related Work

MEDIE[2] was one of the first QA-inspired system for biomedicine (Miyao et al., 2006). It allows users to pose questions in the form of subject-object-verb (SOV) structures. For instance, the question "What does p53 activate?" needs to be split into its parts: "p53" (subject), "activate" (verb), and no object (i.e., the expected answer). MEDIE relies on domain ontologies, parsing and predicate-argument structures (PAS) to search Medline. However, SOV structures are not a user-friendly input, given that many of the biomedical users have no advanced knowledge on linguistics.

We are only aware of three other QA systems for biomedicine: AskHermes[3], EAGLi[4] and HONQA[5]. All of them support input in the form of questions but present result in a different ways.

AskHermes (Cao et al., 2011) outputs lists of snippets and clusters of terms, but the result page is often far too long. Their methods involve regular expressions for question understanding, question target classification, concept recognition and passage ranking based on the BM25 model. The document collection includes Medline articles and Wikipedia documents.

EAGLi (Gobeill et al., 2015) provides answers

based on concepts from the Gene Ontology (GO). Even when no answers are found for a question, EAGLi always outputs a list of relevant publications. It indexes Medline documents locally in the Terrier IR platform and uses Okapi BM25 to rank documents.

HONQA (Cruchet et al., 2009) considers documents from certified websites from the Health On the Net (HON) and supports French and Italian, besides the English language. The answer type detection is based on the UMLS database and the architecture of the systems seems to follow the typical QA workflow. However, no further details are described in their publication.

3 System Architecture

The architecture of Olelo follows the usual components of a QA system (Athenikos and Han, 2010), i.e., document indexing, question processing, passage retrieval and answer processing (cf. Figure 1). In this section we present a short overview of the many tasks inside each of these components. We previously published our methods for multi-document summarization (Schulze and Neves, 2016), which we applied not only for biomedical QA but also for gene-specific summaries. Finally, our participations on the BioASQ challenges also provide insights on previous and current methods behind our system (Neves, 2014, 2015; Schulze et al., 2016).

Document Indexing. We index the document collection and the questions into an IMDB (Plattner, 2013), namely, the SAP HANA database. This database stores data in the main memory and includes other desirable features for on-line QA systems, such as multi-core processing, parallelization, lightweight compression and partitioning. Our document collection currently consists of abstracts from Medline[6] and full text publications from PubMed Central Open Access subset[7]. The document collection is regularly updated to account for new publications.

When indexed in the database, documents and questions are processed using built-in text analysis procedures from the IMDB, namely, sentence splitting, tokenization, stemming, part-of-speech (POS) tagging and NER (cf. Table 1). The latter is
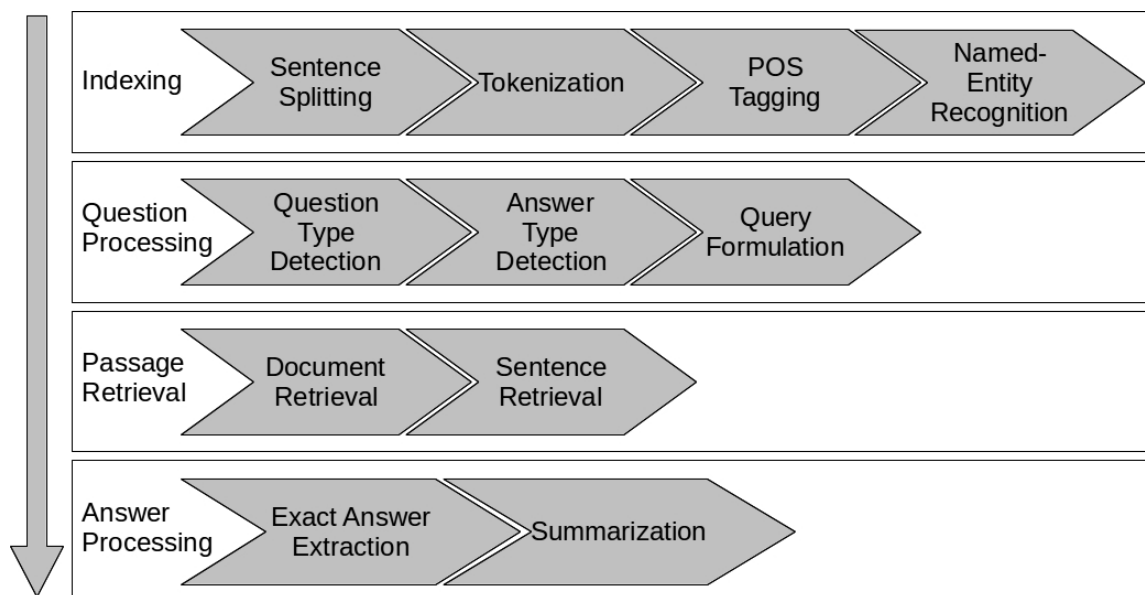

Figure 1: Natural language processing components of Olelo question answering system.

Totals	Abstracts	Full text
Documents	8,335,584	1,116,645
MeSH all	335,174,549	48,455,036
MeSH distinct	54,313	56,587
UMLS all	1,773,195,457	6,209,018,977
UMLS distinct	387,110	392,932

Table 1: Statistics on documents, sentences and named entities (as of February/2017).

based on customized dictionaries for the biomedical domain, which we compiled based on two domain resources: the Medical Subject Headings (MeSH)[8] and the Unified Medical Language System (UMLS)[9].

Question Processing. Olelo currently supports three types of questions: (i) factoid; (ii) definition; and (iii) summary. A factoid question requires one or more short answers in return, such as a list of disease names, definition questions query for a particular definition of a concept, while summary questions expect a short summary about a topic. Components in this step include the detection of the question type via simple regular expressions, followed by the detection of the answer type, in the case of factoid questions. This step also comprises the detection of the headword via regular expression and the identification of its semantic types with the support of the previously detected named entities. The semantic types correspond to

the ones defined by UMLS semantic types (Bodenreider, 2004). Finally, a query is built based on surface forms of tokens, as well as previously detected MeSH and UMLS terms.

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. We start by considering all keywords in the query and we drop some of them later if no document match is found.

Answer Processing. An answer is produced depending on the question type. In case of a definition question, the system simply shows the corresponding MeSH term along with its definition, as originally included in the MeSH terminology. In the case of factoid questions, Olelo returns MeSH terms which belong to the corresponding semantic type that was previously detected. Lastly, the system builds a customized summary for summary questions, based on the retrieved documents and on the query.

4 Use Cases

In this section we show two use cases of obtaining precise answers for particular questions. The examples include a question related to a specific gene and two questions from the BioASQ benchmark. We also present a preliminary comparison of our systems to three others on-line biomedical QA applications.

The "Tutorial" page in Olelo contains more details on the various functionalities of the system. Some few parameters can be set on the "Setting" page, such as the minimal year of publication, the size of the summary (in terms of number of sentence, default value is 5) and the number of documents considered when generating a summary (default value is 20).

Gene-related question. This use case focuses on the gene CFTR, which was one of the chosen #GeneOfTheWeek in a campaign promoted in Twitter by the Ensembl database of genes. Mutations on genes are common causes of diseases, therefore, a user could post the following question to Olelo: "What are the diseases related to mutations on the CFTR gene?". Olelo returns a list of potential answers to the question (cf. Figure 2), and indeed, "cystic fibrosis" is associated to the referred gene[10]. By clicking on "cystic fibrosis", its definition in MeSH is shown, and Olelo informs that 349 relevant document were found (blue button on the bottom). By clicking on this button, a document is shown and this is indeed relevant, as we can confirm by reading the first sentence of its abstract. At this point, the user has many ways to navigate further on the topic, for instance: (a) flick through the rest of the documents; (b) create a summary for this document collection; (c) click on a term (in blue) to learn more about it; (d) visualize full details on the publication (small icon besides its title); (e) navigate through the semantic types listed for cystic fibrosis; or (f) click on another disease name, i.e., "asthma".

BioASQ benchmark questions. Currently, BioASQ (Tsatsaronis et al., 2015) is the most comprehensive benchmark for QA systems in biomedicine. We selected one summary and one factoid question to illustrate the results returned by Olelo for different question types. For the question "What is the Barr body?" (identifier 55152c0a46478f2f2c000004), the system returns a short summary whose first sentence indeed contains the answer to the question: "The Barr body is the inactive X chromosome in a female somatic cell." (PubMed article 21416650). On the other hand, for the factoid question "List chromosomes that have been linked to Arnold Chiari syndrome in the literature.", Olelo presents

a list of chromosome names. Indeed, the following are the official answers in the BioASQ benchmark: "1", "3", "5", "6", "8", "9", "12", "13", "15", "16", "18", "22", "X", "Y". For this particular example, Olelo outputs an even more comprehensive answer than BioASQ, as the MeSH terms include the word "chomosome".

Preliminary evaluation. We recently compared Olelo to the three other biomedical QA systems (cf. Section 2) by manually posing 10 randomly selected factoid questions from BioASQ. We manually recorded the response time of each system and the experiments were carried out outside of the network of our institute. HONQA did not provide results for any of the questions because an error occurred in the system. Olelo found correct answers for four questions (in the returned summaries), EAGLi for two of them (in the titles of the returned documents) and AskHermes for one of them (among the many returned sentences). Regarding the response time, Olelo was the fastest one (average of 8.8 seconds), followed by AskHermes (average of 10.1 seconds) and EAGLi (average of 58.6 seconds).

5 Conclusions and Future Work

We presented our Olelo QA system for the biomedical domain. Olelo relies on built-in NLP procedures of an in-memory database and SQL procedures for the various QA components, such as multi-document summarization and detection of answer type. We have shown examples of the output provided by Olelo when obtaining information for a particular gene and for checking the answers for two questions from the BioASQ benchmark.

Nevertheless, the methods behind Olelo still present room for improvement: (a) The system does not always detect factoid questions correctly given the simple rules it uses for question type detection. In these cases, Olelo generates a short summary from the corresponding relevant documents. (b) Answers are limited to existing MeSH terms, which also support our system for further navigation (cf. Figures 2 and 3). Indeed, our experiments show that we cannot provide answers for many of the questions which expect a gene or protein name, both weakly supported in MeSH, but very frequent in BioASQ (Neves and Kraus, 2016). (c) Our document and passage retrieval components currently rely on approximate match-

---

[10]http://www.ensembl.org/Homo_sapiens/
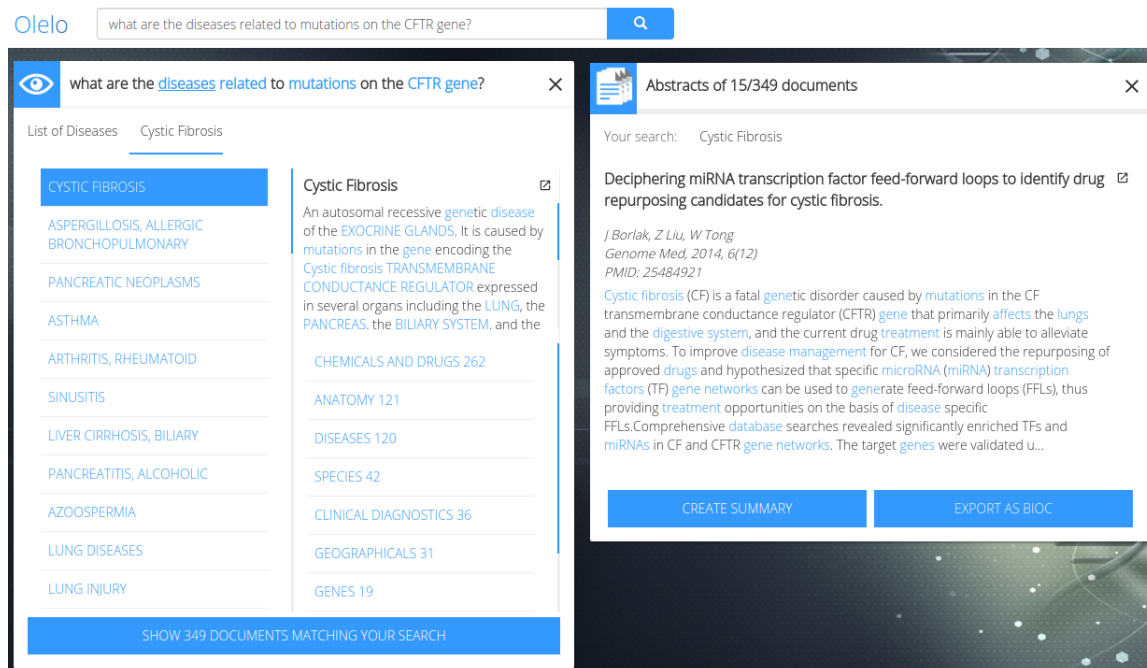Gene/Summary?g=ENSG00000001626

Figure 2: List of answers (disease names) potentially caused by the CFTR gene (on the left) and an overview of one of the relevant publications which contains the answer (on the right).
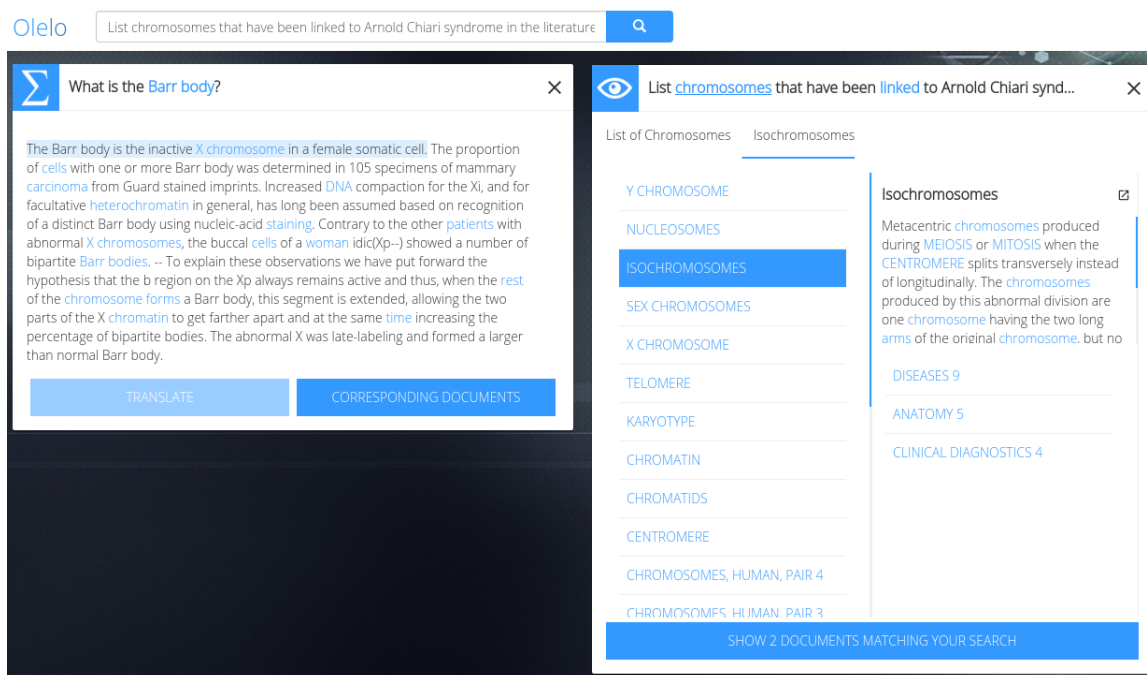

Figure 3: Short paragraph for a summary question (on the left) and list of answers (chromosome names) for a factoid question (on the right), both from the BioASQ dataset.

ing of tokens and named entities but do not consider state-of-the-art IR methods, such as TF-IDF. (d) The sentences that belong to a summary could have been better arranged. The fluency of the summaries is not optimal and we do not deal with co-references, such as pronouns (e.g., "we") which frequently occur in the original sentences. However, when compared to other biomedical QA systems, Olelo performs faster and provides focused answers for most of the questions, instead of a long list of documents. Finally, it provides means to further explore the biomedical literature.

Olelo is under permanent development and improvements are already being implemented on multiple levels: (a) integration of more advanced NLP components, such as chunking and semantic role labeling; (b) support for yes/no questions and improvement of the extraction of exact answers based on deep learning; (c) integration of additional biomedical documents, e.g., clinical trials, as well as documents in other languages.

Finally, in its current state, adaptation of our methods to a new domain would not require major changes. Minor changes are necessary on the question processing step, which relies on specific ontologies, as well as creating new dictionaries for the NER component. In summary, adaptation of the system would mainly consist on the integration of new document collections and specific terminologies.

References

Sofia J. Athenikos and Hyoil Han. 2010. Biomedical question answering: A survey. *Computer Methods and Programs in Biomedicine* 99(1):1 – 24. https://doi.org/10.1016/j.cmpb.2009.10.003.

Olivier Bodenreider. 2004. The unified medical language system (umls): integrating biomedical terminology. *Nucleic Acids Res* 32(Database issue):D267–D270. https://doi.org/10.1093/nar/gkh061.

Yonggang Cao, Feifan Liu, Pippa Simpson, Lamont D. Antieau, Andrew S. Bennett, James J. Cimino, John W. Ely, and Hong Yu. 2011. Askhermes: An online question answering system for complex clinical questions. *Journal of Biomedical Informatics* 44(2):277–288. https://www.ncbi.nlm.nih.gov/pubmed/21256977.

Sarah Cruchet, Arnaud Gaudinat, Thomas Rindflesch, and Celia Boyer. 2009. What about trust in the question answering world? In *Proceedings of the AMIA Annual Symposium*. San Francisco, USA, pages 1–5.

Julien Gobeill, Arnaud Gaudinat, Emilie Pasche, Dina Vishnyakova, Pascale Gaudet, Amos Bairoch, and Patrick Ruch. 2015. Deep question answering for protein annotation. *Database* 2015:bav081. https://doi.org/10.1093/database/bav081.

Zhiyong Lu. 2011. Pubmed and beyond: a survey of web tools for searching biomedical literature. *Database* 2011. https://www.ncbi.nlm.nih.gov/pubmed/21245076.

Yusuke Miyao, Tomoko Ohta, Katsuya Masuda, Yoshimasa Tsuruoka, Kazuhiro Yoshida, Takashi Ninomiya, and Jun'ichi Tsujii. 2006. Semantic retrieval for the accurate identification of relational concepts in massive textbases. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL-44, pages 1017–1024. https://doi.org/10.3115/1220175.1220303.

Mariana Neves. 2014. Hpi in-memory-based database system in task 2b of bioasq. In *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18*. pages 1337–1347. http://ceur-ws.org/Vol-1180/CLEF2014wn-QA-Neves2014.pdf.

Mariana Neves. 2015. Hpi question answering system in the bioasq 2015 challenge. In *Working Notes for CLEF 2015 Conference, Toulouse, France, September 8-11*. http://ceur-ws.org/Vol-1391/59-CR.pdf.

Mariana Neves and Milena Kraus. 2016. Biomedlat corpus: Annotation of the lexical answer type for biomedical questions. In *Proceedings of the Open Knowledge Base and Question Answering Workshop (OKBQA 2016)*. The COLING 2016 Organizing Committee, Osaka, Japan, pages 49–58. http://aclweb.org/anthology/W16-4407.

Mariana Neves and Ulf Leser. 2015. Question answering for biology. *Methods* 74:36 – 46. Text mining of biomedical literature. https://doi.org/10.1016/j.ymeth.2014.10.023.

Hasso Plattner. 2013. *A Course in In-Memory Data Management: The Inner Mechanics of In-Memory Databases*. Springer, 1st edition.

Frederik Schulze and Mariana Neves. 2016. Entity-supported summarization of biomedical abstracts. In *Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016)*. The COLING 2016 Organizing Committee, Osaka, Japan, pages 40–49. http://aclweb.org/anthology/W16-5105.

Frederik Schulze, Ricarda Schüler, Tim Draeger, Daniel Dummer, Alexander Ernst, Pedro Flemming, Cindy Perscheid, and Mariana Neves. 2016. Hpi question answering system in bioasq 2016. In *Proceedings of the Fourth BioASQ workshop at the Conference of the Association for Computational Linguistics*. pages 38–44. http://aclweb.org/anthology/W/W16/W16-3105.pdf.

George Tsatsaronis, Georgios Balikas, Prodromos Malakasiotis, Ioannis Partalas, Matthias Zschunke, Michael R Alvers, Dirk Weissenborn, Anastasia Krithara, Sergios Petridis, Dimitris Polychronopoulos, et al. 2015. An overview of the bioasq large-scale biomedical semantic indexing and question answering competition. *BMC bioinformatics* 16(1):138. https://www.ncbi.nlm.nih.gov/pubmed/25925131.