Assessing the performance of Olelo, a real-time biomedical question answering application

Mariana Neves, Fabian Eckert, Hendrik Folkerts, Matthias Uflacker

Hasso Plattner Institute at University of Potsdam, Germany

Olelo in BioASQ task 5b

- We recently developed **Olelo**, a **question answering (QA) system for biomedicine**.
- This year, we participated with the **current state of the application** and with an extension based on **semantic role labeling** (SRL) that we are currently investigating.
- We compared our system to other on-line biomedical QA systems (AskHermes, EAGLi, HONQA) in terms of the **response time and the quality of the answers**.

Systems & Methods

- Olelo includes various NLP components: question processing, document and passage retrieval, answer processing and multi-document summarization.
- For the SRL approach, we used BioKIT tool and propose a rule-based approach to investigate if SRL could improve Olelo. We created a set of rules to make use of SRL for answering yes/no, factoid, list and summary questions. Basically, the rules aim to rank answer snippets by their semantic conformity with the question. Semantic conformity was measured by comparing the PASs found in the question and a particular answer candidate.

Results for BioASQ task 5b

Batch	System	Doc. retr.	Pass. retr.
1	Olelo	0.0465	0.0441
2	Olelo	0.0318	0.0246
3	Olelo	0.0658	0.0386
4	Olelo	0.0449	0.0347
5	Olelo	0.0381	0.0386
top results		[0.0874,0.1157]	[0.0467,0.0898]

Results for mean average precision (MAP) for Olelo in BioASQ task 5b phase A, i.e., for document retrieval and passage retrieval. Range of top results in all batches are presented in the last row.

Batch	System	Rouge-2	Rouge-SU4
1	Olelo	0.2222	0.2710
	Olelo-GS	0.2958	0.3243
	SRL	0.0467	0.0510
	SRL2	0.0833	0.0870
2	Olelo	0.2751	0.2976
	Olelo-GS	0.2048	0.2500
	SRL2	0.0425	0.0418
3	Olelo	0.3426	0.3604
	Olelo-GS	0.2891	0.3262
	SRL2	0.0411	0.0416
4	Olelo	0.2261	0.2696
	Olelo-GS	0.3460	0.3516
	SRL2	0.0796	0.0740
5	Olelo	0.3418	0.3536
	Olelo-GS	0.2117	0.2626
	SRL2	0.0406	0.0413
top results		[0.5153,0.6891]	[0.5182,0.6789]

Results for ideal answers (summaries) in terms of Rouge metrics for Olelo and the SRL approach. Range of top results in all batches are presented in the last row.

Batch	System	Yes/No	Factoid	List
1	Olelo	-	0.0400	0.0240
	Olelo-GS	-	0.0400	0.0477
	SRL	0.8824	-	0.0038
	SRL2	0.8824	-	0.1183
2	Olelo	-	0.0430	0.0281
	Olelo-GS	-	0.0323	0.0287
	SRL2	0.9630	0.0129	0.1123
3	Olelo	-	0.0192	0.0408
	Olelo-GS	-	0.0192	0.0549
	SRL2	0.8065	0.0128	0.1715
4	Olelo	-	0.0253	0.0513
	Olelo-GS	-	0.0513	0.0513
	SRL2	0.5517	0.0379	0.0943
5	Olelo	-	-	0.0202
	Olelo-GS	-	-	0.0379
	SRL2	0.4615	0.0286	0.2870
top results		[0.8387,0.9630]	[0.3606,0.5713]	[0.3358,0.5001]

Results for Olelo and the SRL approach in the BioASQ task 5b phases B (exact answers). Results for yes/no questions are in terms of accuracy, MRR for factoid questions and f-measure for list questions. Range of top results in all batches are presented in the last row.

Systems	Output	Answers	Time
AskHermes	7/10	1/10	10.1 [2.09,19.74]
EAGLi	10/10	2/10	58.6 [21.41,107.72]
Olelo	10/10	4/10	8.84 [3.35,28.12]

Results in terms of number of correct answers and response time for the on-line QA applications.

Related publications:

- Kraus M, Niedermeier J, Jankrift M, Tietböhl S, Stachewicz T, Folkerts H, Uflacker M and Neves M. **Olelo: a web application for intuitive exploration of biomedical literature**, Nucleic Acids Research Web service issue, 2017.
- 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

http://hpi.de/plattner/olelo