Poznań Contribution to TREC Clinical Trials 2021*

Jakub Dutkiewicz[1] and Czesław Jędrzejek[1]

Faculty of Computing, Poznań University of Technology, Plac Marii Skłodowskiej-Curie 5, 60-965 Poznań, Poland

Abstract. In this papaer we go over our approach to TREC Clinical Trials 2021. We discuss the architecture of our retrieval system. Specifically, we describe the used topic processing techniques. We recount the structure of a document and our methods of interpreting and extracting data from the document. This paper also covers the description of experiments we proposed for TREC Clinical Trials 2021. We conclude with a brief discussion of the obtained results.

1 Introduction

TREC Clinical Trials 2021 is a new TREC track, which fits into a line of clinical support related tracks. It is a natural continuation of the Clinical Decision Support (CDS) and Precision Medicine (PM) lines of tracks. In this edition, the information retrieval systems are asked to interpret complex, written in medical, natural language descriptions of patients. Contrarily to the previous editions, there is only a single entry for each patient. All the required medical information about the patient is supposed to be contained within this natural language entry. We use various techniques of regular expression based extraction and thesaurus based extraction, in order to obtain the most relevant data from the description. Processed topics are expressed as queries, which consist of a set of keywords. Information retrieval system is then tasked to search the most relevant Clinical Trials to the description of the patient. Clinical Trials are in a form of structured documents, which contain various exclusion and inclusion criteria for the trial, as well as the description of results obtained within the Clinical Trial. The documents are structured as JavaScript Object Notation files - a structure, which consists of a set of key-value pairs. Certain data, such as title, description, inclusion and exclusion criteria of the trial, list of keywords, used intervention or condition name are stored within designated parts of the document, which are denoted with a specific key. This is a very often used notation in Information Retrieval, however in classical approach, the number of key-value pairs is usually smaller. Here we have several dozen of pairs, while in the classical approach there are usually a few (such as title and body of the document), hence the assessment of importance of certain pairs for retrieval is of the upmost value in this task.

* Supported by PUT statutory funds. One of the authors (CJ) acknowledges the nVIDIA GPU Grant of Quadro P6000 card.

We have quite an experience in the medical search. We participated in TREC CDS 2016 [6], TREC PM 2017, TREC PM 2018 track [4], TREC PM 2019 [5] TREC PM 2020;and in bioCADDIE 2016 [3] . This track is unique in a sense it mostly derives the information from fields of various types. In this contribution we particularly focus on the source of the information, we try to choose the best set of document fields for the task.

2 Retrieval Architecture

We propose a dedicated Retrieval Architecture, which consists of three major layers. These layers compute the output for the following tasks:

1. Document Processing Layer - the main goal of this layer is the creation of the Elasticsearch index, which contains all the information about the Clinical Trials.
2. Topic Processing Layer - this layer is responsible of converting the input topics into a set of representative keywords.
3. Information Retrieval Layer - finally we use Elasticsearch to convert the sets of keywords into queries and perform the information retrieval task. This layer uses a classical BM25 weighting schema.

Layers are implemented separately, thus each of the layers can be replaced. We believe the modular implementation increases the robustness of the architecture. An overview of the application is illustrated in

2.1 Document Processing Layer

The document processing layer consists of three major components. First of all the documents are processed with use of a dedicated XML processor. The processor is implemented in C++ language, it is optimized towards the time of the execution. Its sole task is to convert the XML documents into JSON format. Once the documents are processed, a python script is used to filter out noisy data. We use a dedicated set of regular expressions in order to perform the preprocessing. Such prepared documents are ready to be used by a retrieval system based on Elasticsearch. The last step consists of two parts. First of all, we create an empty index. Then, we iterate over the list of created JSON files and index each file separately.

2.2 Topic Processing Layer

The task of topic processing layer is to extract vocabulary from the topics. The extracted vocabulary is used as a query in the Information Retrieval system. We use three separate subsystems for the term extraction.

First one is based on the MeSH database [2]. In this subsystem we parse the topic text and for each entry in in the MeSH database we search for equivalent

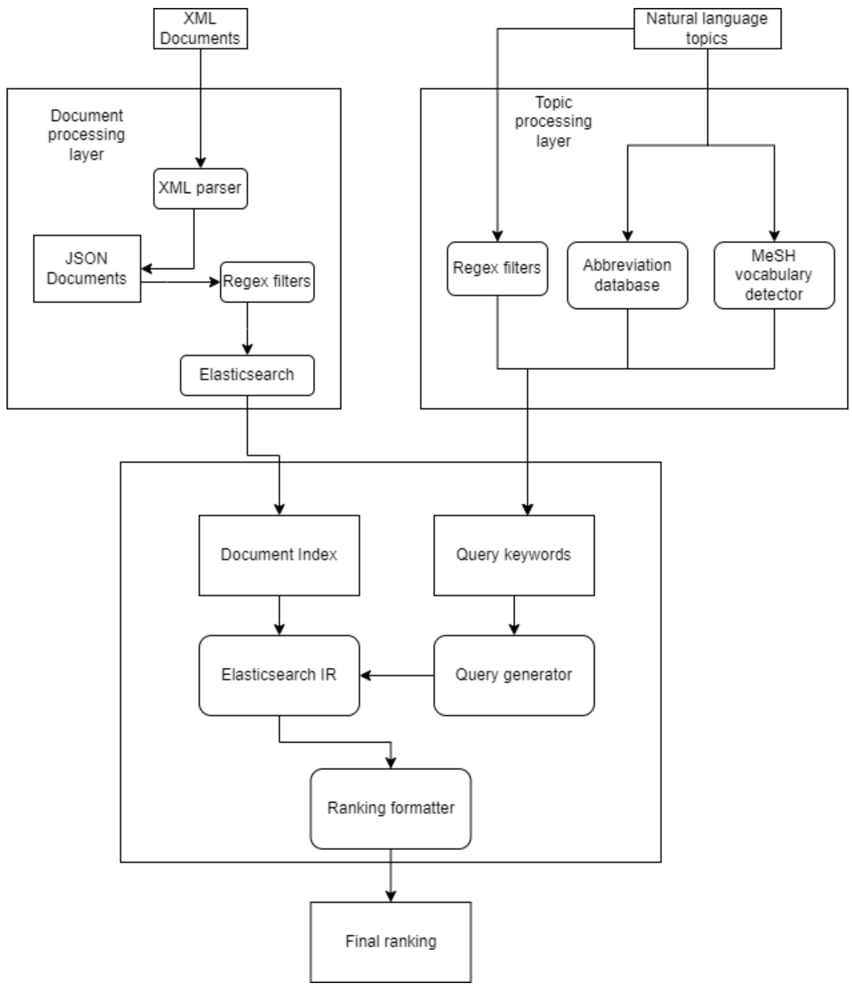

Fig. 1. Retrieval System Architecture

set of tokens within the text. We use the MeSH classifier entries denoted with '*NEWRECORD' identifier. For each entry, we extract the main name of the concept, denoted by 'NM' and its synonyms, denoted by 'SN'. Similarly, we use descriptor names denoted by 'MH' and 'ENTRY' identifiers. This approach is very similar to Query Expansion using MeSH vocabulary [1], although, contrarily to the Query Expansion, our goal is the extraction of the most informational pieces of text.

The second subsystem uses a dedicated set of medical abbreviations. We use search for the abbreviations in the topic text. Every matched abbreviation is included in the query. Finally, we use a set of regular expressions in order to find patient characteristics, such as his age or sex. Regular expressions are designed in a way, which is able to capture the value and type of each characteristic.

2.3 Information Retrieval Layer

In this layer we perform the main Information Retrieval task. We employ the BM25 framework, implemented in the Elasticsearch in order to generate retrieval lists. It should be noted that we use various document fields in the information retrieval. We use a set of fields, which consists of keywords and specific terms, such as "MeSH terms" or "keywords"; as well as fields, which contain natural language passages, such as "description" or "title". For each field, we generate a list. We exclude documents which do not meet exclusion criteria, such as sex or age. Then the lists are merged in order to create a final ranking.

2.4 System Specification

We run the system on a machine with information retrieval and machine learning dedicated specification. For physical memory, we employ a disk array, which uses five disks directly connected to peripheral component interface with Non-Volatile Memory Express technology. We use this technology, to create a fast link between graphical processing unit, operating memory and physical memory. It is vital in information retrieval, as the textual corpora tend to get very large (e.g. the corpus used in Trec PM 2020 comprises of 220 GB of unprocessed data). We employ 126 GB of operational memory - it is convenient to store corpus indices directly in the operational memory for research purposes. Here, we use Elasticsearch tool for instant access to the documents within the corpus. Medium sized operational memory allows us to work on relatively large document and query matrices. We use nVIDIA P6000 graphics card for calculations - distributed, GPU based machine learning allows us to process 20000 samples per minute. We use classical driver-CUDA-CudNN-tensorflow-keras software stack. The intermediate layer, which multiplies Document matrices by Query matrices is still implemented in Python3 with numpy library and at this moment is our bottleneck. Currently we can process roughly 400 documents per minute in this layer. Considering the size of Clinical Trials, and other IR corpora, this part requires a redo. Our system is maintained on a Clear Linux machine with remote access.

3 Conducted Experiments

We propose five experiments. Each experiment uses a combination of query and document specifications. In various experiments we take only a subset of generated query terms. We propose several steps of processing the topic entry. We extract keywords with use of dedicated regular expression patterns. For example

```
[A-Za-z ]*([0-9]+)[ -]?(y/?o|(year|day|month)s?[ -](old)?)
([0-9a-zA-Z-]* ([0-9a-zA-Z-]* )*)?(girl|female|woman|[Ff] )
```

this pattern is used for recognition of age of female patients. We know, that the pattern works for this particular set of topics, and perhaps can make be a source of mistakes with different set of topics. We also can imagine a pattern, which would apply for a vast majority of documents in any set of topic documents. We would need a large set of patient descriptions in order to create such a pattern. We believe that regular expressions are a fit tool for extraction of basic patient properties, which are often expressed in the same manner.

We extract vocabulary, which intersects with MeSH vocabulary. The extracted vocabulary expands the query. In order to do that, we parse the topic descriptions and extract all MeSH terms from the text. For each concept in the MeSH vocabulary, we chech, wether it exists in the topic. If it does we append the proper name of this concept as well as all synonyms to the query.

We use a similar method to recognize the abbreviations used within the description. We use a dedicated set of abbreviations. The set in this case consists only of abbreviations, which appear within the topics. Usually each abbreviation has one proper name, but there are some, which consist of several forms. We add both the abbreviation as well as all its synonyms to the query. Here's a sample of the abbreviations database:

T-L	thoracolumbar
RLE	right lower extremity
MRI	Magnetic resonance imaging
CPT-11	Current Procedural Terminology
HTN	hypertension
SOB	shortness of breath
LV	liver
LE	lower extremity
EF7	ejection fraction
RV	Right Ventricle
BB	beta-blocker

Table 1. Sample abbreviations used in the retrieval system.

We propose three runs, which are designed to test the performance of various query processing strategies:

1. Trms - In this run we use only terms extracted with MeSH dictionary.

2. Abbr - In this run we use a set of abbreviations, keywords extracted with MeSH, keywords extracted with regular expressions. We also use all fields from the documents.

3. Add - In this run, on top of the Abbr setting, we use a set of additional terms for the retrieval.

Another important aspect of the retrieval is the processing of documents. Here the documents contain vast number of fields. We focus on the proper selection of the fields to be used in the retrieval process. We analyze the following fields:

1. brief summary - a field, which contains a natural language description of the clinical trial. It contains a brief information about the clinical trial. Interpretation of the information is hindered by the properties of natural language.
2. brief title and full title - there are several formulation of the title of specific clinical trial. Title is usually formed as a single, simple sentence, which makes sense in natural language.
3. brief description - a field, which is similar to brief summary. The information in this field is extended, but it is still written in natural language.
4. condition - a field, which contains information about the condition, which is being researched in a particular clinical trial. We treat this field as a keyword based field.
5. mesh terms - a set of keywords related to the clinical trial, each of the keywords appear in the mesh ontology.
6. criteria - a field, which contains both information about the exclusion and inclusion criteria of the patient. This field requires additional processing. It is usually formatted as a set of one, or two lists. The lists correspond to and are titled as Inclusion Criteria and Exclusion Criteria of the patients, which partake in a clinical trial. We did not use this particular field in the final submissions, however we believe incorporating this particular field into the retrieval process is crucial towards the improvement of the retrieval system.
7. intervention name - name of a technique, drug or behaviour tested during the clinical trial.
8. minimum age, maximum age and sex - those fields denote the demographics of patients partaking in the clinical trial. We use those two fields in order to exclude documents, which would not apply to the given description of a patient. Except for the last set of logical fields and criteria, we create a separate ranking for each field. A position of a document within the ranking with regard to a specific topic is denoted as $r_{criteria}(d, t)$ We calculate the final score of the document as

$$s(d, t) = \frac{1}{\sum_{c \in C} r_c(d, t)}$$

With regard to this aspect of document processing we propose two experiments, in order to judge usefulness of the certain sets of fields:

1. FT - In this run we take only the document parts, which are in the form of natural language texts.

2. MT - In this run we take only the document parts, which are in the form of keywords or tokens.

The results of specific runs are presented in Table. 2. Our runs generally performed better than median. Our best settings are Abbr and Add. Additional terms very slightly change the results. Using partial information generally worsens the results.

Run	NDCG@10	P@10	MRR
TREC median	0.304	0.161	0.294
TREC best[7] [8]	0.712	0.593	0.826
Abbr	0.399	0.271	0.480
Add	0.404	0.270	0.481
FT	0.372	0.248	0.487
MT	0.391	0.275	0.471
Trms	0.211	0.140	0.332

Table 2. Comparison of our runs to median TREC runs.

4 Conclusions

In this paper we briefly discuss the techniques of building a query. We test the strategy of building a query, which consists of keywords extracted or generated upon the natural language text. After comparing the results we obtained with the results of other participants. Even the best performing solutions have slightly weaker baselines and use a process of query reformulation, which gives similar results to ours [7]. We believe that our strategies of formulation of a query are valid. Additionally, it is recommended to use both the natural language information and semi structural information from the documents. Limiting the system to only one type of the information worsens the efficiency of the retrieval system.

References

1. Aronson, A.R., Rindflesch, T.C.: Query expansion using the UMLS metathesaurus. In: AMIA 1997, American Medical Informatics Association Annual Symposium, Nashville, TN, USA, October 25-29, 1997. AMIA (1997),

https://knowledge.amia.org/amia-55142-a1997a-1.585351/t-001-1.587519/f-001-1.587520/a-096-1.587728/a-097-1.587725
2. Bodenreider, O.: The unified medical language system (umls): integrating biomedical terminology. Nucleic Acids Res. **32**(Database-Issue), 267–270 (2004), http://dblp.uni-trier.de/db/journals/nar/nar32.htmlBodenreider04
3. Cieslewicz, A., Dutkiewicz, J., Jedrzejek, C.: Baseline and extensions approach to information retrieval of complex medical data: Poznan's approach to the biocaddie 2016. Database **2018**, bax103 (2018). https://doi.org/10.1093/database/bax103, https://doi.org/10.1093/database/bax103
4. Cieslewicz, A., Dutkiewicz, J., Jedrzejek, C.: POZNAN contribution to TREC PM 2018 (notebook paper). In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of the Twenty-Seventh Text REtrieval Conference, TREC 2018, Gaithersburg, Maryland, USA, November 14-16, 2018. NIST Special Publication, vol. 500-331. National Institute of Standards and Technology (NIST) (2018), https://trec.nist.gov/pubs/trec27/papers/Poznan-PM.pdf
5. Cieslewicz, A., Dutkiewicz, J., Jedrzejek, C.: Poznan contribution to TREC-PM 2019. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of the Twenty-Eighth Text REtrieval Conference, TREC 2019, Gaithersburg, Maryland, USA, November 13-15, 2019. NIST Special Publication, vol. 1250. National Institute of Standards and Technology (NIST) (2019), https://trec.nist.gov/pubs/trec28/papers/Poznan.PM.pdf
6. Dutkiewicz, J., Jedrzejek, C., Frackowiak, M., Werda, P.: PUT contribution to TREC CDS 2016. In: Voorhees, E.M., Ellis, A. (eds.) Proceedings of The Twenty-Fifth Text REtrieval Conference, TREC 2016, Gaithersburg, Maryland, USA, November 15-18, 2016. NIST Special Publication, vol. 500-321. National Institute of Standards and Technology (NIST) (2016), http://trec.nist.gov/pubs/trec25/papers/IAII_PUT-CL.pdf
7. Lin, J., Chen, H., Hu, C., Lin, S.C., Li, Y., Ma, X., Ronak, P., Yang, J.H., Wang, C.J., Yates, A., Zhang, X.: New nails for old hammers: Anserini and pyserini at trec 2021
8. Roberts, K., Demner-Fushman, D., Voorhees, E.M., Steven, B., William, R.H.: Overview of the trec 2021 clinical trials track

A Appendix A. Detailed Results

In this appendix we present detailed version of the achieved results. Results marked with green colour are better than TREC median, red colour denotes results worse than a median, yellow indicates the same value as TREC median. Results equal to the best achieved results are denoted with red border.

	ndcg@10							prec@10							MRR						
topic	Best	Median	Abbr	Add	FT	MT	Trms	Best	Median	Abbr	Add	FT	MT	Trms	Best	Median	Abbr	Add	FT	MT	Trms
1	0,7386	0,4351	0,6748	0,6748	0,6652	0,7384	0,3431	0,5	0,1	0,4	0,4	0,4	0,5	0,1	1	0,2	0,5	0,5	1	0,5	0,25
2	0,9633	0,6256	0,9195	0,9195	0,7813	0,9195	0,0392	0,9	0,4	0,8	0,8	0,7	0,8	0	1	0,5	1	1	0,3333	1	0,0714
3	0,7744	0,0851	0,6673	0,6673	0,6539	0	0,0851	0,6	0	0,5	0,5	0,4	0	0,1	1	0,0303	1	1	1	0,0149	0,2
4	0,7656	0,1526	0,2122	0,2122	0,4618	0,244	0,0734	0,7	0,1	0,2	0,2	0,3	0,2	0,1	1	0,2	0,5	0,5	1	0,5	0,1429
5	0,9286	0,4265	0,828	0,828	0,6792	0,7249	0,5167	0,8	0,3	0,8	0,8	0,6	0,8	0,5	1	0,5	1	1	1	0,5	1
6	0,9574	0,3286	0,6349	0,6349	0,5411	0,865	0	0,9	0,2	0,6	0,6	0,4	0,7	0	1	0,3333	0,5	0,5	0,5	1	0,0323
7	0,7453	0,2915	0,5536	0,5536	0,6328	0,5536	0,3359	0,8	0,3	0,4	0,4	0,5	0,4	0,3	1	0,3333	1	1	1	1	0,1667
8	0,9364	0,1918	0,251	0,251	0,4609	0	0	0,9	0,1	0,2	0,2	0,3	0	0	1	0,1111	0,33	0,3333	0,5	0,0093	0,0196
9	0,5474	0,3645	0,2639	0,2639	0,3418	0,2639	0,1795	0,2	0	0	0	0,1	0	0	1	0,0227	0,08	0,0769	0,125	0,0769	0,0526
10	0,7405	0	0,6081	0,6081	0,4592	0,6081	0,6698	0,5	0	0,2	0,2	0,2	0,2	0,4	1	0,002	1	1	0,125	1	1
11	0,808	0,359	0,6026	0,6026	0,6448	0,5744	0,2265	0,7	0,1	0,2	0,2	0,2	0,2	0	1	0,125	0,5	0,5	1	0,2	0,0667
12	0,9682	0,6146	0,8408	0,8408	0,7059	0,8166	0,5863	0,9	0,5	0,7	0,7	0,6	0,6	0,3	1	1	1	1	1	1	1
13	0,7671	0,1844	0,0367	0,0367	0,1939	0	0,2716	0,6	0,1	0	0	0,1	0	0,3	1	0,1667	0,07	0,0714	0,5	0,0032	0,2
14	0,7444	0,4186	0,5847	0,5847	0,4167	0,525	0,2558	0,5	0,1	0,3	0,3	0,2	0,2	0	1	0,2	0,2	0,2	0,1667	0,25	0,0625
15	0,8938	0,4618	0,0851	0,0851	0,117	0,7447	0,0663	0,8	0,3	0,1	0,1	0,1	0,6	0,1	1	0,5	0,2	0,2	0,2	1	0,1111
16	0,6138	0,1582	0,1737	0,1737	0,2832	0,2048	0	0,6	0,1	0,2	0,2	0,3	0,2	0	1	0,125	0,33	0,3333	0,3333	0,3333	0,0323
17	0,8959	0,4627	0,0734	0,0734	0,1389	0,3218	0	0,8	0,2	0,1	0,1	0,1	0,2	0	1	0,5	0,14	0,1429	0,5	0,5	0,006
18	0,6106	0,3658	0,4574	0,4574	0,5	0,3423	0,2571	0,4	0,1	0	0	0	0,1	0	1	0,1	0,06	0,0556	0,0208	0,1667	0,0172
19	0,839	0,43	0,4774	0,4774	0,4724	0,5121	0,2201	0,8	0,4	0,2	0,2	0,3	0,2	0,1	1	0,5	1	1	0,25	1	1
20	0,8283	0,11	0,3837	0,3837	0,211	0,2201	0,2985	0,7	0,1	0,3	0,3	0,1	0,1	0,2	1	0,1667	0,25	0,25	0,1111	1	1
21	1	0	0		0	0	0	1	0	0	0	0	0	0	1	0,0345	0,01	0,0077	0,0058	0,0084	0,0312
22	0,9669	0,3211	0,0739	0,0739	0	0,0739	0	0,9	0,2	0	0	0	0	0	1	0,1429	0,03	0,0323	0,0385	0,0208	0,0455
23	0,8832	0,4207	0,09	0,09	0,1024	0,09	0	0,8	0,1	0	0	0	0	0	1	0,2	0,02	0,0179	0,0192	0,0179	0,0088
24	0,9669	0,4986	0,8461	0,8461	0,5998	0,8048	0,4888	0,9	0,4	0,7	0,7	0,5	0,7	0,4	1	0,5	1	1	1	1	1
25	0,8669	0,1514	0,0636	0,0636	0,142	0	0	0,8	0,1	0,1	0,1	0,2	0	0	1	0,2	0,1	0,1	0,1667	0,0069	0,0015
26	0,9306	0,3129	0,3786	0,3786	0,2935	0,3933	0	0,9	0,2	0,3	0,3	0,2	0,3	0	1	0,3333	1	1	1	1	0,0111
27	0,7263	0,2569	0,1791	0,1791	0,1939	0,0474	0	0,5	0,1	0,1	0,1	0,1	0	0	1	0,1	0,14	0,1429	0,5	0,0066	0,0455
28	0,9682	0,3537	0	0	0	0,471	0,11	0,9	0,2	0	0	0	0,5	0,1	1	0,2	0,03	0,0312	0,0141	0,5	0,3333
29	0,8093	0,374	0,5045	0,5045	0,4256	0,5045	0,1024	0,6	0,1	0,1	0,1	0	0,1	0	1	0,1	0,17	0,1667	0,0625	0,1667	0,0357
30	0,9134	0,4264	0,3866	0,3866	0,5665	0	0,0367	0,8	0,2	0,1	0,1	0,2	0	0	1	0,25	0,17	0,1667	0,125	0,05	0,0055

Fig. 2. Topics 1 - 30.

	ndcg@10							prec@10							MRR						
topic	Best	Median	Abbr	Add	FT	MT	Trms	Best	Median	Abbr	Add	FT	MT	Trms	Best	Median	Abbr	Add	FT	MT	Trms
31	0,6348	0,2181	0,2114	0,2114	0,2294	0,311	0,1012	0,4	0,1	0	0	0,1	0,2	0	1	0,1667	0,08	0,0769	0,1429	0,1667	0,0118
32	0,7472	0,0948	0,4819	0,4819	0,7001	0,0331	0,1389	0,7	0,1	0,4	0,4	**0,7**	0	0,1	1	0,1	**1**	**1**	**1**	0,0345	0,5
33	1	0,3223	0,1168	0,1168	0,1168	0,1168	0,0734	1	0,3	0,1	0,1	0	0,1	0,1	1	0,5	0,13	0,125	0,0244	0,125	0,1429
34	0,8078	0,1389	0,3699	0,3699	0,3996	0,2548	0,1389	0,7	0,1	0,2	0,2	0,2	0,1	0,1	1	0,125	**1**	**1**	**1**	**1**	0,5
35	1	0,3757	0,839	0,839	0,7549	0,839	0,615	1	0,1	0,8	0,8	0,6	0,8	0,6	1	0,3333	**1**	**1**	**1**	**1**	**1**
36	1	0,3301	0,2283	0,2283	0,3485	0,1584	0,0784	1	0,2	0,1	0,1	0,2	0,2	0,1	1	0,3333	0,2	0,2	**1**	0,25	0,1667
37	1	0,4703	0,3566	0,3566	0,3517	0,8568	0,1389	1	0,3	0,3	0,3	0,2	0,8	0,1	1	0,5	0,25	0,25	0,5	0,5	0,5
38	0,9653	0,3476	0,7218	0,7218	0,5697	0,9024	0,581	0,9	0,3	0,5	0,5	0,3	0,8	0,5	1	0,5	0,5	0,5	0,3333	**1**	**1**
39	0,8375	0,0694	0,5536	0,5536	0,5012	0,5536	0,7701	0,7	0	0,4	0,4	0,5	0,4	**0,7**	1	0,0303	0,5	0,5	0,3333	0,5	0,5
40	0,8938	0,1993	0	0	0	0	0	0,8	0,1	0	0	0	0	0	1	0,3333	0,01	0,009	0,0167	0,009	0
41	1	0,4746	0,4746	0,4746	0,2807	0,4746	0,359	1	0,2	0,3	0,3	0,2	0,3	0,2	1	0,3333	0,5	0,5	0,3333	0,5	**1**
42	0,6529	0,0367	0,055	0,055	0,1062	0,055	0	0,6	0	0	0	0,1	0	0	1	0,0029	0,02	0,0156	0,1	0,0189	0
43	0,825	0,4926	0,3295	0,3295	0,1467	0,2259	0,2532	0,7	0,3	0,3	0,3	0	0,1	0,1	1	0,5	0,5	0,5	0,0323	0,1	**1**
44	0,7063	0,3685	0,5789	0,5789	0,4985	0,5789	0,4182	0,1	0	0	0	0	0	0	1	0,0074	0,06	0,0625	0,0357	0,0625	0,04
45	0,9653	0,4314	0	0	0	0,4314	0	0,9	0,3	0	0	0	0,4	0	1	0,5	0,02	0,0244	0,0278	0,5	0,0435
46	1	0,3786	0,5619	0,5619	0,5479	0,5619	0,2985	1	0,3	0,5	0,5	0,5	0,5	0,2	1	0,5	**1**	**1**	**1**	**1**	**1**
47	1	0,2618	0,2112	0,2112	0,1498	0,2112	0,2399	1	0,1	0,1	0,1	0,1	0,1	0,3	1	0,25	0,5	0,5	0,25	0,5	0,3333
48	1	0,5107	0,8338	0,8338	0,7722	0,9315	0,9574	1	0,4	0,7	0,7	0,7	0,8	0,9	1	1	**1**	**1**	**1**	**1**	**1**
49	0,8693	0,4303	0	0	0	0,5345	0	0,8	0,2	0	0	0	0,4	0	1	0,5	0,03	0,0345	0,037	0,3333	0,0345
50	0,7893	0,2489	0,0948	0,0948	0,0636	0,11	0	0,5	0,1	0,1	0,1	0,1	0,1	0	1	0,5	0,25	0,25	0,1	0,3333	0,0455
51	0,6526	0,1751	0,2489	0,2489	0,2675	0,2532	0,0367	0,5	0,1	0,1	0,1	0,1	0,1	0	1	0,2	0,5	0,5	**1**	**1**	0,04
52	0,8912	0,1952	**0,8912**	**0,8912**	0,6726	0,7156	0,0948	0,8	0,1	**0,8**	**0,8**	0,5	0,6	0,1	1	0,3333	**1**	**1**	**1**	**1**	0,25
53	0,4235	0	**0,4235**	**0,4235**	0,1467	0,0663	0,0851	0,3	0	**0,3**	**0,3**	0,1	0,1	0,1	1	0,0111	**1**	**1**	0,3333	0,1111	0,2
54	0,8117	0,4437	0,4921	0,4921	0,3568	0,7083	0,3765	0,7	0,3	0,1	0,1	0,2	0,5	0	1	0,3333	0,13	0,125	0,25	0,5	0,0833
55	1	0,0474	0,0347	0,0347	0	0,0347	0	1	0	0	0	0	0	0	1	0,027	0,01	0,0053	0,0097	0,0053	0
56	0,8189	0,5017	0,3777	0,3777	0,3839	0,4239	0,1514	0,8	0,3	0,2	0,2	0,2	0,3	0,2	1	1	**1**	**1**	**1**	**1**	0,2
57	1	0,3241	0,3697	0,3697	0,1787	0,3697	0,2685	1	0,2	0,4	0,4	0,2	0,4	0,3	1	0,3333	0,2	0,2	0,1667	0,2	0,3333
58	0,9682	0,5496	0,8115	0,8115	0,8192	0,842	0,4969	0,9	0,3	0,6	0,6	0,6	0,7	0,4	1	0,5	**1**	**1**	**1**	**1**	0,3333
59	1	0,6061	0,8568	0,8568	0,7202	0,8568	0,6137	1	0,3	0,8	0,8	0,6	0,8	0,5	1	0,5	**1**	**1**	**1**	**1**	**1**
60	0,9669	0,478	0,766	0,766	0,7338	0,36	0,3997	0,9	0,3	0,6	0,6	0,6	0,1	0,2	1	1	**1**	**1**	**1**	0,5	**1**

Fig. 3. Topics 31 - 60.

	ndcg@10							prec@10							MRR						
topic	Best	Median	Abbr	Add	FT	MT	Trms	Best	Median	Abbr	Add	FT	MT	Trms	Best	Median	Abbr	Add	FT	MT	Trms
61	1	0,2577	0,2154	0,2154	0,2298	0,1266	0	1	0,2	0,3	0,3	0,3	0,1	0	1	0,25	0,17	0,1667	0,2	0,25	0,007
62	0,7126	0,4314	0,5049	0,5049	0,6852	0,5759	0,3028	0,5	0,1	0,1	0,1	0,4	0,2	0,1	1	0,2	0,14	0,1429	1	0,1667	0,1
63	0,6995	0,0851	**0,6995**	**0,6995**	0,4249	0,6489	0,2201	0,7	0,1	0,6	0,6	0,5	0,5	0,1	1	0,1429	**1**	**1**	0,25	**1**	**1**
64	0,785	0,2602	0,2058	0,2058	0,5679	0,0734	0	0,7	0,1	0,2	0,2	0,4	0,1	0	1	0,1667	0,11	0,1111	0,5	0,1429	0,0046
65	0,6476	0,0773	0,0663	0,0663	0	0,0663	0	0,6	0	0,1	0,1	0	0,1	0	1	0,0909	0,11	0,1111	0,04	0,1111	0,0189
66	0,7283	0,0367	0,11	0,11	0,2457	0,1635	0	0,5	0	0,1	0,1	0,3	0,2	0	1	0,0667	0,33	0,3333	0,3333	0,2	0,0455
67	0,8485	0,4212	0,2354	0,2354	0,2065	0,1188	0,2201	0,7	0,2	0,2	0,2	0,1	0	0,1	1	0,3333	0,25	0,25	0,2	0,0909	**1**
68	0,7854	0,5096	0,6622	0,6622	0,5096	0,7676	0,669	0,6	0,2	0,3	0,3	0,3	0,4	0,4	1	0,3333	**1**	**1**	**1**	**1**	0,3333
69	0,6197	0	0	0	0	0	0	0,5	0	0	0	0	0	0	1	0,0074	0	0	0	0	0
70	0,9653	0,2201	0,6105	0,6105	0,6286	0	0	0,9	0,1	0,4	0,4	0,4	0	0	1	0,2	**1**	**1**	**1**	0,0169	0,0161
71	0,9364	0,1884	0,1657	0,1657	0,2201	0,2134	0,0784	0,9	0,1	0,1	0,1	0,1	0,2	0,1	1	0,25	0,2	0,2	**1**	0,25	0,1667
72	0,8988	0,2736	**0,8988**	**0,8988**	0,3933	0,88	0,6931	0,8	0	**0,8**	**0,8**	0,3	0,7	0,7	1	0,0909	**1**	**1**	0,3333	**1**	0,5
73	0,8393	0,2181	0,6758	0,6758	0,6215	0,6758	0,5675	0,6	0,1	0,3	0,3	0,3	0,3	0,1	1	0,1667	**1**	**1**	**1**	**1**	**1**
74	0,9633	0,2369	0	0	0	0	0	0,9	0,2	0	0	0	0	0	1	0,5	0,02	0,0152	0,0312	0,0152	0
75	0,9302	0,4792	0,7118	0,7118	0,71	0,7118	0,2489	0,8	0,2	0,5	0,5	0,5	0,5	0,1	1	0,3333	**1**	**1**	**1**	**1**	0,5
	0,849143	0,303988	0,398725	0,404114	0,372596	0,391432	0,211307	0,748	0,161333	0,270667	0,270667	0,248	0,274667	0,14	1	0,294165	0,48	0,4805	0,486816	0,470999	0,331995

Fig. 4. Topics 61 - 75.