

1 User logs description

Log queries Given a certain topic s , each search log includes user queries q_i :

$$s \rightarrow [q_1, \dots, q_n] \quad (1)$$

These queries were used to obtain responses by sending them to a Solr engine. Each query q includes keywords for *abstract* and *title* document parts. An example of user query q could be as follows:

(abstract:"autoimmune" OR title:"autoimmune") AND (abstract:"symptoms"
OR title:"symptoms")

Solr responses Each Solr response r for query q (r_q) has an XML structure, where **result** xml-node includes all documents that were found. Listing 1 illustrates response which includes 87 **doc** xml-nodes. An example of document information presented at Listing 2. Each document provides d^i – an unique MEDLINE identifier **medlineId** (line 7, Listing 2). Values of these identifiers were used to compose a list of **relevant** documents according to the certain response r_q :

$$r_q \rightarrow [d_1^i, \dots, d_n^i] \quad (2)$$

Listing 1: "Solr response format"

```
1 <response>
2   <result name="response" numFound="87" start="0">
3     <doc> ... <doc>
4     <doc> ... <doc>
5     ...
6   </result>
7 </response>
```

Listing 2: "Document format"

```
1 <doc>
2   <str name="abstract">
3     An unusual clinical presentation of ...
4   </str>
5   <str name="author">Amontree JS; Stuart TD; Bredfeldt JE.</str>
6   ...
7   <int name="medlineId">89327874</int>
8   ...
9   <str name="pubtype">JOURNAL ARTICLE.</str>
10  <str name="source">J Clin Gastroenterol 8911; 11(3):303-7</str>
11 </doc>
```

Evaluation We use an OHSUMED collection ¹ that provides an expert opinion of document relevance (and non relevance) according to the topics. Every topic s has a list of documents ² d^i , where each document has a relevance label l . This label might be as follows: **p** (possibly relevant), **d** (definitely relevant), and **n** (non relevant). Both **p** and **d** label types we combine as **r** – relevant. The latter means that we distinguish *relevant* and *non relevant* classes. Equation 3 summarises above:

$$s \rightarrow \left[\langle d_1^i, l_1 \rangle, \dots, \langle d_n^i, l_n \rangle \right], \text{ where } l_i = \mathbf{r|n} \quad (3)$$

We use *trec_eval* ³ tool for search performance evaluation. Next we describe an information that were provided both for documents of OHSUMED collection and Solr results to perform evaluation. Given a topic s and certain query, *trec_eval* requires to provide the following information per each document:

- **qid** – identifier of the query (topic s);
- **doc_no** – document identifier (d^i);
- **rel** – relevance, (label l , which is 1 for relevant and 0 otherwise); each document of the Solr response we decide a relevant.

Additionally, each document of Solr results should have a *sim* parameter – similarity, which is assumed to be higher for documents that were retrieved at first. For i -th document, we define sim_i as follows:

$$sim_i = docs_count - i \quad (4)$$

Where *docs_count* is an amount of documents that were found by Solr.

¹<https://dmice.ohsu.edu/hersh/ohsumed>

²<https://dmice.ohsu.edu/hersh/ohsumed/judged>

³https://github.com/eXascaleInfolab/pytrec_eval