

Relevance Ranking

19 January 2017

This online document is available [here](#).

Scope of Task Force:

Path to Plenary 9 (Barcelona; 5-7 April):

Who is target for recommendations/outputs?

- Could tie into persona used in best practices for search, i.e. producer, repository, searcher. For repository (primary target) how to implement, what to implement, what tools are available in open code repositories. For producer - relates to choice of metadata standard, etc. For searcher - how to formulate queries to get appropriate results ranking.

- Search ranking:
 - Which weighting algorithm works better for short and structured records: vector space, BM25, language model? (Compare the three algorithm with same test collections)
 - Which weighting algorithm works better for which search task (e.g, known item search versus exploratory search - get a list of summarised search tasks from the Use Case task force)?
 - Is it helpful by adding more metadata - Core metadata for DOI (core) versus Core + extra?
 - Which metadata standards to recommend
 - How to deal with federated search across repositories using different standards
 - Are some fields (e.g., title, description) more important than others? Weight fields differently?
 - How to treat linked information: lump them to its data record (metadata) or propagate weight like pagerank?
 - Exploit personal or community search history (clicks and/or downloaded data)
 - Exploit personal profile (from registration, or MyData)
 - Do we consider “deep indexing” i.e. content, not just metadata
 - Applying machine learning to ranking, e.g., Solr Learning To Rank (LTR) ML plugin...
 - Exploit ontologies/semantics, especially important for cross-domain searches, for query expansion
 - Examples: MeSH and GENE in medicine
 - Query expansion: What evidences should be used for query expansion to achieve better ranking?

- Evaluation
 - Which measurement should be applied (precision, recall, F-measure, normalized discounted cumulative gain (NDCG), etc)? Given a search task, which measure should be used?

- What is baseline? Techniques for establishing baseline?
- Where to get test search topics? From members, query logs from service providers, researchers?
- Who is going to judge relevance of a record to a search topic? Do we need domain experts?
- How do we best categorize the relevancy of results? For example, in the bioCADDIE 2016 Dataset Retrieval Challenge they used the categories “relevant”, “partially relevant” and “not relevant”.
- Should we build test collections (like TERC or INEX style) or have people do A/B test? How large must test collections be?
- Search platform
 - Which search platform to be used
 - Solr
 - Lucene
 - Indri (Lemur)
 - Elasticsearch
 - Proprietary ...?
 - Which is used by most of service providers?
 - Which type of collection to take on: Generic collection versus disciplinary collectio

TODOs / Questions / Things to consider?

- Polish document, especially scope; create PDF with link to online Google Doc; Siri will propagate doc (via RDA list, which is on Wiki), but will check with Anita first
- Can one or more participants provide a test bed?
 - Peter will check with Elsevier and MingFang will check with ANDS
- Next steps:
 - Next telecon: 2 Feb; same time