INEX 2009 Entity-Ranking Track Entity Ranking Retrieval Task and Result Submission Specification

Gianluca Demartini, Tereza Iofciu, Arjen P. de Vries

Version 1.0

Abstract

What's New in 2009? The INEX 2009 EntityRanking track sees a continuation of the Entity Ranking Task, and of the List completion (LC) task, with the new challenge of an annotated collection as working dataset.

Entity Ranking Track

The motivation for the entity ranking task is to return entities that satisfy a topic described in natural language text. Given preferred categories, relevant entities are assumed to loosely correspond to those Wikipedia pages that are labeled with these preferred categories (or perhaps sub-categories of these preferred categories). Retrieval methods need to handle the situation where the category assignments to Wikipedia pages are not always consistent, and also far from complete. For example, given a preferred category 'art museums and galleries', an article about a particular museum such as the 'Van Gogh Museum' (155508) may not be labeled by 'art museums and galleries' but labeled by a sub-category of the preferred category instead, such as category 'art museums and galleries in the Netherlands'. Therefore, when searching for "art museums in Amsterdam", correct answers may belong to other categories close to this category in the Wikipedia category graph, or may not have been categorized at all by the Wikipedia contributors. The category 'art museums and galleries' is only an indication of what is expected, not a strict constraint (like in the CAS title for the ad-hoc track).

Within the Entity Ranking Track we define the following two tasks:

- Entity ranking (without examples)
- Entity list completion (with examples). Entity list completion is a special case of entity ranking where a few examples of relevant entities are provided as relevance feedback information.

Entity ranking

Display

the results are presented as a ranked-list of entities to the user.

Users

view the result list top-down, one-by-one. Users do not want overlapping results in the result-list.

What we hope to learn from this task is: How can we identify entity results of the right type vs. documents containing pieces of information? How can we best make use of the additional information available, such as the category information and annotations?

Results to Return

The aim of the Entity Ranking Task is to return a ranked-list of distinct entity documents.

Entity List Completion

List completion is a sub-task of entity ranking which considers relevance feedback information. Instead of knowing the desired category (entity type), the topic specifies a number of correct entities (instances) together with the free-text context description. Results consist again of a list of entities (Wikipedia pages). The task makes a number of assumptions:

Display

the results are presented as a ranked-list of entities to the user.

Users

view the result list top-down, one-by-one. Users do not want overlapping results in the result-list.

What we hope to learn from this task is: How do results from the Entity List Completion Task differ from the ones from Entity Ranking Task? What features from the given examples are used in task?

Results to Return

The aim of the Entity List Completion Task is to return a ranked-list of distinct entity documents.

Result Submission

Fact sheet:

- For the two tasks, we allow up to 6 runs. That is, a participant can never submit more than 12 runs in total.
- **Mandatory runs:**
    - Entity Ranking: an automatic run which can only use the title and the category information - without using the topic description and example entities
    - List Completion: an automatic feedback run which must use the title and example entities - without using the topic description and category
- All participants are invited to submit runs with other combinations of query parts, for example, assuming the scenario where the user doesn't know the category information and she would only search with the title of the topic. We do not outlaw the use of the other topic fields, to allow participants to conduct their own experiments involving them, and would even encourage experiment in this area since such deviating runs may in fact improve the quality of the assessment pool.
- There are additional requirements on the submissions for the tasks:
    - Entity Ranking and List Completion: for the same topic, results should be unique (they should not overlap)

    Runs that violate these requirements in any way will not be taken into account.

The participants should indicate which parts of the topic have been used in their submitted runs for both the entity ranking and relation search tasks.

INEX 2009 Topics

There is only one set of topics to be used for all entity ranking tasks at INEX 2009. Such set is built taking genuine XER topics developed in 2007 and the final set of topics from 2008. Topics have been adapted in order to consider changes in document IDs and categories.

Example of topic:

<inex_topic topic_id="9999">
<title>Impressionist art in the Netherlands</title>
<description>
I want a list of art galleries and museums in the Netherlands that have impressionist art.
</description>
<narrative>Each answer should be the article about a specific art gallery or museum that contain impressionist or post-impressionist art works.
</narrative>
<categories>
<category>art museums and galleries</category>

```
</categories>
<entities>
<entity id="155508">Van Gogh Museum</entity>
<entity id="892971">Kröller-Müller Museum</entity>
</entities>
</inex_topic>
```

Runs

For each of the two tasks, we allow up to 6 submissions. The results of one run must be contained in one submission file (i.e. up to 12 files can be submitted in total). A submission may contain up to 500 retrieval results for each of the INEX topics included within that task.

Submission format

Run file should be submitted in the format that can be evaluated by TREC trec_eval program, and each article should be uniquely identified by WP+article's id number. The format is the following:

topicID 0   WP+articleID   rank   score   runID

Please name the run files differently, encoding your institute and the task, the fields used, and the order of priority for judging in the file name (1 is the highest priority). The filename convention bellow:

[priority:1|2|3..]_[institute:name]_[task:ER|LC]_[topic_fields:TDNEC](_runDescription), where runDescription is not mandatory.

For example, for L3S submitting a first priority run for entity ranking using only title and category the file name would be:

1_L3S_ER_TC_mandatoryRun

Please record within the run, in the last element, your run identifier (as specified above). In the following example, the run identifier is "1_L3S_ER_TC_mandatoryRun". The run identifier will be unique for each run.

Here's an artificial example for the entity ranking task:

100 0 WP112876 1 1.165 1_L3S_ER_TC_mandatoryRun

100 0 WP21934 2 1.016 1_L3S_ER_TC_mandatoryRun

100 0 WP4801 3 1.017 1_L3S_ER_TC_mandatoryRun

...

100 0 WP2218400 471 0.001 1_L3S_ER_TC_mandatoryRun

100 0 WP176879 472 0.001 1_L3S_ER_TC_mandatoryRun

100 0 WP1028765 473 0.001 1_L3S_ER_TC_mandatoryRun

Please submit up to 500 results per topic. The RSV value (5th value) is assumed to be higher for the docs to be retrieved first; if you do not use such a score in your system, please give a fictitious value such as 501-rank as score.

Result Submission Procedure

Prior to submission, please use the validation script: check_trec.pl.

The validation script works for trec_eval format only. It creates a log file with warnings and errors. The log file should be checked before submitting. If there are errors then we will not accept the run submission.

Usage: perl check_ check_trec.pl generic <file_path>

Upload files using the online submission tool, where each run should be compressed (gzip or zip) and uploaded individually.