UNIVERSIDADE DA CORUÑA

Exploiting Pooling Methods for Building Datasets for Novel Tasks

David Otero

Information Retrieval Lab, Computer Science Department, University of A Coruña, Spain

Overview

- **Test collections**, within the Crandfield paradigm, are the most widely used tools to evaluate new IR systems.
- They consist of a set of **documents**, **topics** and **relevance judgments** indicating which documents are relevant to those topics.
- The **quality** of the datasets is important, as they are used to compare the effectiveness of different retrieval models.
- **Objective:** build a platform to easily and cheaply build high-quality test collections to use in new appearing tasks.

Background

- Building test collections is **time and resource consuming** because it requires the work of human assessors.
- **Pooling strategies** can alleviate this cost, as they allow building larger test collections without the need of evaluate every document in the collection.
- In pooled collections only a **subset of the entire document set** –the pool– is assessed for relevance for every information need.
- It has been demonstrated that these strategies make the most of the assessor's work time.

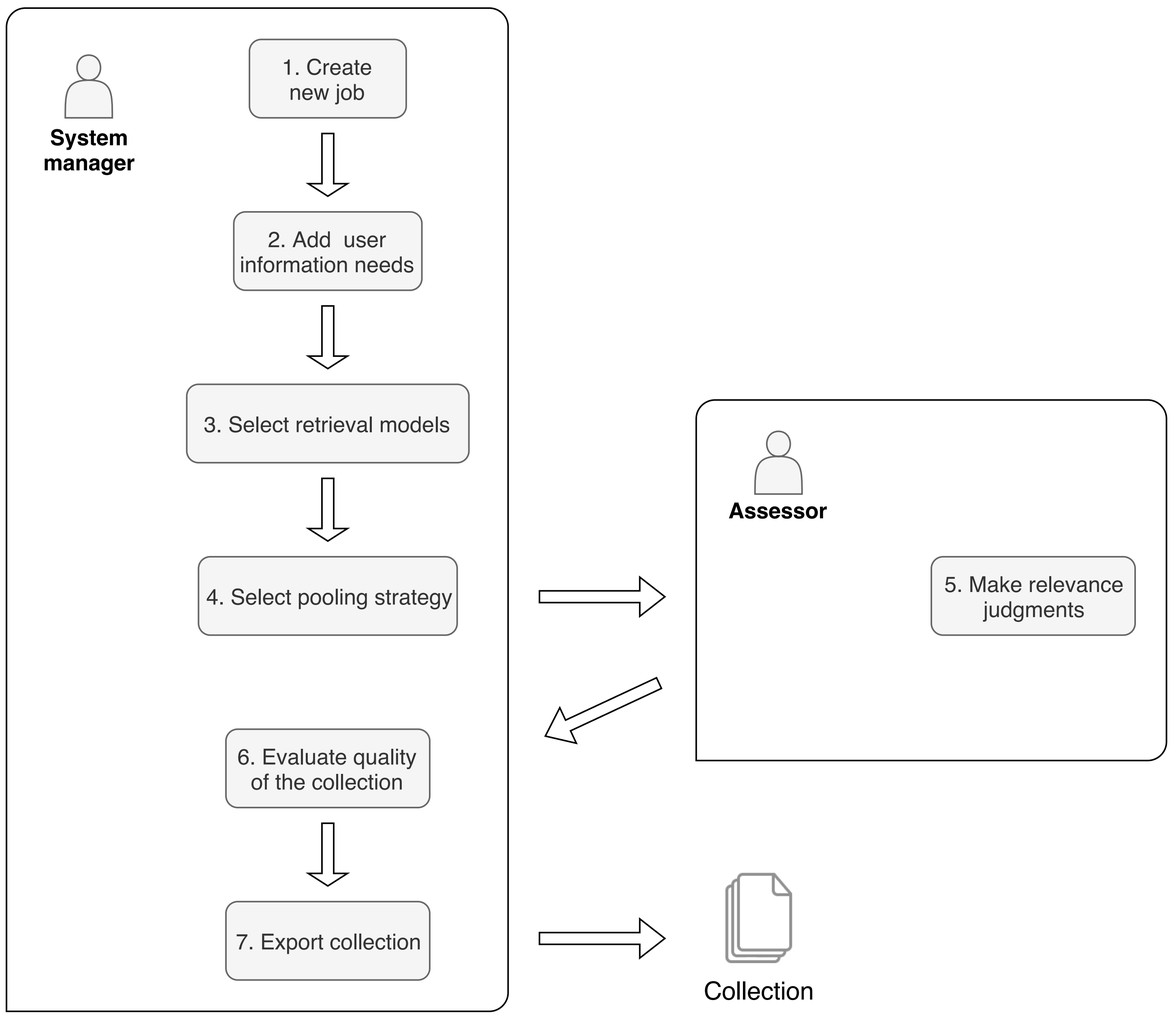
Proposal

- Platform with two roles: system manager and assessor.
- Instead of participant runs, use different **query variants and retrieval strategies**.
- Implement **state-of-the-art pooling methods** to make the most of the assessor's time.
- **Flexibility** in terms of adding new pooling strategies and new retrieval methods.
- Focus on the **evaluation** of the obtained collections.

Motivation

- Address the problem of building test collections at an affordable cost.
- Obtain relevance judgments without the need to use participant systems.
- Having judgments prior to competition.

Workflow



Pooling methods

Classic unbiased full pool assessment strategy:

- **DocID**. Documents in the pool are sorted by their identifier.

Algorithms that impose an order of evaluation intending to reduce the assessment effort:

- **MTF**. The systems are prioritized and the top document of the top prioritized system is assessed.
- **Multi-armed bandits**. A multi-armed bandit approach where each machine is represented by a run and playing a machine means examining the document supplied by it.
- **Hedge**. An on-line learning algorithm that was adapted for metasearch and pooling.

Conclusions and Future Work

- Obtaining human judgments is time and resource consuming.
- We designed a **flexible** platform in terms of adding new retrieval methods and pooling strategies.
- We want to investigate the quality of datasets built using the results of participant systems versus the collections built following this approach.

FDIA 2019, 9th PhD Symposium on Future Directions in Information Access, July 17th, 2019, Milan, Italy.