Towards a Ranking Model for Semantic Layers over Digital Archives

Pavlos Fafalios, Vaibhav Kasturia, Wolfgang Nejdl
L3S Research Center, University of Hannover, Germany
{fafalios,kasturia,nejdl}@l3s.de

1. Motivation
- How to explore archives in a more advanced and exploratory way?
  - Find documents discussing about a specific category of entities (e.g., philanthropists), or about entities sharing some characteristics (e.g., born in Germany before 1960)?
- How to explore archives by integrating information from existing knowledge bases, like DBpedia?

2. Semantic Layer
- RDF repository describing metadata and annotation information for a collection of archived documents.
  - Allows running advanced, entity-centric SPARQL queries that combine metadata of the documents (e.g., publication date) and semantic information (e.g., mentioned entities)
- More as Fafalios et al., “Building and Querying Semantic Layers for Web Archives”, JCDL’17

Example for a news article:
- SELECT DISTINCT ?article WHERE {
  ?article dc:date ?date .
  FILTER(year(?date) = 1990)
  ?entity1 oae:hasMatchedURI "dbr:F_W_de_Klerk" .
  ?entity1 oae:hasMatchedURI "dbr:Fafalios" .
  ?entity1 oae:hasMatchedURI "dbr:Roger Федерали" .
  ?entity2 oae:hasMatchedURI "dbr:F_W_de_Klerk" .
  ?entity2 oae:hasMatchedURI "dbr:Fafalios" .
  ?entity2 oae:hasMatchedURI "dbr:Roger Федерали" .
}

Retrieve articles of 1990 discussing about Nelson Mandela and F. W. de Klerk

3. The problem
- The results returned by a SPARQL query:
  - can be numerous
  - all equally match the query
- How to rank them for identifying and promoting the most important ones?
  - What makes an archived document important for a given query?

4. Related Work
- Ranking of archived documents (for free-text queries)
  - Time-aware Retrieval and Ranking [Kanhabua and Anand, 2016]
  - Tempas [Holzmann and Anand, 2016], HistDiv [Singh et al., 2016]
  - Works by Kanhabua et al. (2016), Vo et al. (2016)
- Ranking in knowledge graphs
  - Learning to rank for RDF entity search [Dalil et al., 2012]
  - Swoogle [Ding et al., 2005], SemRank [Anyanwu et al., 2005]
  - NAGA [Kannewei et al., 2008], DING [Delbru et al., 2010], ReconRank [Hogan et al., 2006], Noc-order [Graves et al., 2008]
- Our approach: Ranking archived documents for structured queries in knowledge graphs
  - Availability of metadata and entity annotations
  - No access to full contents!