

# Ranking Archived Documents for Structured Queries on Semantic Layers

Pavlos Fafalios  
L3S Research Center,  
Leibniz University of Hannover  
Hannover, Germany  
fafalios@L3S.de

Vaibhav Kasturia  
L3S Research Center,  
Leibniz University of Hannover  
Hannover, Germany  
kasturia@L3S.de

Wolfgang Nejdl  
L3S Research Center,  
Leibniz University of Hannover  
Hannover, Germany  
nejdl@L3S.de

## ABSTRACT

Archived collections of documents (like newspaper and web archives) serve as important information sources in a variety of disciplines, including Digital Humanities, Historical Science, and Journalism. However, the absence of efficient and meaningful exploration methods still remains a major hurdle in the way of turning them into usable sources of information. A semantic layer is an RDF graph that describes metadata and semantic information about a collection of archived documents, which in turn can be queried through a semantic query language (SPARQL). This allows running advanced queries by combining metadata of the documents (like publication date) and content-based semantic information (like entities mentioned in the documents). However, the results returned by such structured queries can be numerous and moreover they all equally match the query. In this paper, we deal with this problem and formalize the task of “*ranking archived documents for structured queries on semantic layers*”. Then, we propose two ranking models for the problem at hand which jointly consider: i) the relativeness of documents to entities, ii) the timeliness of documents, and iii) the temporal relations among the entities. The experimental results on a new evaluation dataset show the effectiveness of the proposed models and allow us to understand their limitations.

## CCS CONCEPTS

• **Information systems** → *Probabilistic retrieval models*;

## KEYWORDS

Semantic Layers; Archived documents; Ranking; Probabilistic modeling; Stochastic modeling

## ACM Reference Format:

Pavlos Fafalios, Vaibhav Kasturia, and Wolfgang Nejdl. 2018. Ranking Archived Documents for Structured Queries on Semantic Layers. In *JCDL '18: The 18th ACM/IEEE Joint Conference on Digital Libraries, June 3–7, 2018, Fort Worth, TX, USA*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3197026.3197049>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

JCDL '18, June 3–7, 2018, Fort Worth, TX, USA  
© 2018 Association for Computing Machinery.  
ACM ISBN 978-1-4503-5178-2/18/06...\$15.00  
<https://doi.org/10.1145/3197026.3197049>

## 1 INTRODUCTION

Despite the increasing number of digital archives worldwide (like news and web archives), the absence of efficient and meaningful exploration methods still remains the major bottleneck in the way of turning them into usable information sources [6]. Semantic models try to solve this problem by offering the means to describe and publish metadata and semantic information about a collection of archived documents in the standard RDF format. A repository of such data, called Semantic Layer [11], allows running advanced queries which combine metadata of the documents (like publication date) and content-based semantic information (like entities mentioned in the documents). For example, we can access a Semantic Layer over a newspaper archive and find articles of a specific time period discussing about a specific category of entities (e.g., *philanthropists*) or about entities that share some characteristics (e.g., *lawyers born in Germany*), while we can also integrate information coming from other knowledge bases like DBpedia. Such advanced information needs can be directly expressed through structured (SPARQL) queries or through user-friendly interactive interfaces which transparently transform user interactions to SPARQL queries (e.g., Faceted Search-like browsing interfaces [39]).

However, the results returned by such queries can be numerous and moreover they all equally match the query: there is no relevance ranking like in the case of keyword-based information retrieval. Thus, there arises the need for an effective method to rank the returned results for discovering and showing to the users the most important ones. For instance, when requesting articles from a news archive published within a specific time period and mentioning one or more query entities, important documents may be those whose main topic is about an important event related to the query entities during the requested time period. Thus, an effective ranking method should consider the different factors that affect the importance of documents to the query, while at the same time relying only on the data available in the semantic layer (there is no access to the full contents of the documents).

Although there is a plethora of works on ranking archived documents for keyword-based temporal queries, the problem of ranking such documents for the case of structured queries on knowledge graphs has not yet been recognized and studied. In this paper, we address this gap by first introducing and formalizing this type of problem. Then, to cope with this problem, we propose two ranking models, a probabilistic one and a Random Walk-based one, which jointly consider the following aspects: i) the *relativeness* of a document to the query entities, ii) the *timeliness* of a document's publication date, iii) the temporal *relatedness* of the query entities to other entities mentioned in the documents. The idea is to promote

documents that mention the query entities many times, that have been published in important (for the query entities) time periods, and that mention many other entities co-occurring frequently with the query entities in important time periods. Such an approach is widely applicable since it exploits the least amount of metadata.

In a nutshell, in this paper we make the following contributions:

- We formulate and formalize the problem of ranking archived documents for structured queries over semantic layers.
- Due to lack of evaluation datasets for this problem, we have created a new ground truth dataset for a news archive which we make publicly available.
- We propose two ranking models for the problem at hand: a probabilistic one and a stochastic one (Random Walk with restart).
- We present the results of an experimental evaluation which illustrate the effectiveness of the proposed models. We also analyze problematic cases for understanding when and why the models fail to provide good rankings.

The rest of the paper is organized as follows: Section 2 presents the required background and related literature. Section 3 defines the problem and describes a new evaluation dataset. Section 4 introduces the probabilistic model. Section 5 introduces the stochastic model. Section 6 presents evaluation results. Section 7 concludes the paper and discusses interesting directions for future research.

## 2 BACKGROUND AND RELATED WORKS

### 2.1 Semantic Layers

A Semantic Layer is an RDF repository (RDF graph) of structured data about a collection of archived documents [11]. Structured data includes not only metadata information about a document (like publication date), but also *entity annotations*, i.e., disambiguated entities mentioned in each document extracted using an entity linking system [34]. Figure 1 shows an example of (a part of) a Semantic Layer describing metadata and annotation information for a news article. We notice that the document was published on 6 January 2012 and mentions the entity name “Giuliani” at character position 512, which probably (with confidence score 0.9/1.0) corresponds to the known American lawyer and former politician Rudy Giuliani.

A semantic layer allows running advanced, entity-centric queries that can also directly integrate information from other knowledge bases like DBpedia. Listing 1 shows an example of a SPARQL query that can be answered by a semantic layer over a collection of old news articles. The query requests articles of 1989 mentioning New York lawyers born in Brooklyn. By accessing DBpedia at query-execution time, the query retrieves the entities that satisfy the query as well as additional information, in particular the birth date of each lawyer. We see that, using the entity URIs and type/category information, we can find documents about entities even if we do not know the names of the entities, or without needing to specify a long list of all the entity names. Moreover, using the SPARQL operators COUNT and GROUP BY, and exploiting the expressive power of SPARQL, we can aggregate information related to documents and entities.

As shown in [11], a semantic layer can answer information needs that existing keyword-based systems (like Google news) are not

able to sufficiently satisfy. Such advanced (but also common) information needs can be directly expressed through SPARQL queries or by exploiting a user-friendly interface that transforms user interactions to SPARQL queries, like Sparklis [14] or SemFacet [2]. Since the results returned by such structured queries can be numerous and have no ranking, in this paper we study how we can rank them based on their importance to the query entities.

### 2.2 Related Works

We report works on the related problems of *time-aware document ranking* and *ranking in knowledge graphs*, and we discuss the commonalities and differences of these research areas to our problem.

**2.2.1 Time-aware Document Ranking.** The impact of *time* on information retrieval has received a large share of attention in the last decade. The surveys in [8] and [23] provide a comprehensive categorization and overview of temporal IR approaches and related applications. As regards time-aware *ranking*, existing works are classified into two different types based on two main notions of relevance with respect to time [8, 23]: 1) recency-based ranking, and 2) time-dependent ranking. Since recency-based ranking methods promote documents that are recently created or updated (this preference of freshness is common in general web searching), below we discuss only time-dependent ranking methods which are useful when searching archived collections of documents.

Jin et al. [22] proposed a ranking algorithm to sort results by applying a linear interpolation of text similarity, temporal information, and page importance (based on PageRank), where temporal similarity is the ranking score of temporal relevance based on the set of intersection conditions between the temporal query and the temporal expressions found in the web page. Arikan et al. [3] and Berberich et al. [5] introduced approaches that integrate temporal expressions extracted from the documents into language modeling frameworks. Metzler et al. [27] proposed a time-dependent ranking model to adjust the document scores based on an analysis of web query logs and a set of document fields to estimate the time of both the query and the document. The work by Perkio et al. [9] automatically detects topical trends and their importance over time within a news corpus using a statistical topic model and a simple variant of TF-IDF. These trends are then used as the basis for temporally adaptive rankings. Dakka et al. [10] consider the publication time of documents to identify the important time intervals that are likely to be of interest to an implicit temporal query. Then, time is incorporated into language models to assign an estimated relevance value to each time period. Aji et al. [1] proposed a term weighting model that uses the revision history analysis of a document to redefine the importance of terms which is then incorporated into BM25 and statistical language models. Kanhabua and Nørvag [24] proposed a time-sensitive ranking model based on learning-to-rank techniques for explicit temporal queries. To learn the ranking model, temporal and entity-based features are applied.

Regarding more recent works, Singh et al. [36] introduced the notion of *Historical Query Intents* and modeled it as a search result diversification task which intends to present the most relevant results from a topic-temporal space. For retrieving and ranking historical documents like news articles, the authors propose a retrieval algorithm, called HistDiv, which jointly considers the dimensions

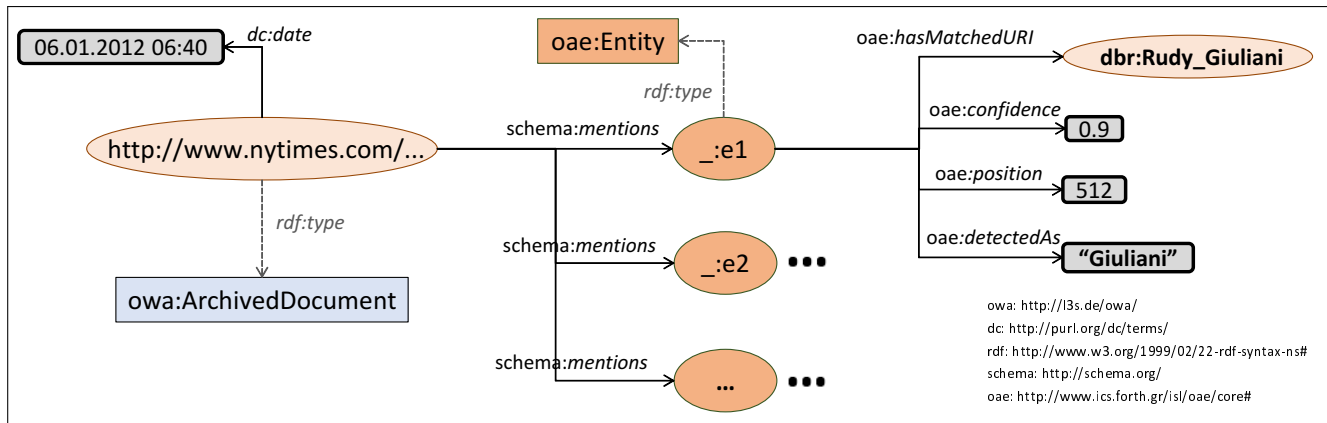


Figure 1: A part of a Semantic Layer describing metadata and entity annotations for a news article.

```

1 SELECT ?article ?title ?date ?nylawyer ?bdate WHERE {
2   SERVICE <http://dbpedia.org/sparql> {
3     ?nylawyer dc:subject dbc:New_York_lawyers ; dbo:birthPlace dbr:Brooklyn .
4     OPTIONAL { ?nylawyer dbo:birthDate ?bdate } }
5   ?article dc:date ?date FILTER(year(?date) = 1989) .
6   ?article schema:mentions ?entity .
7   ?entity oae:hasMatchedURI ?nylawyer .
8   ?article dc:title ?title
9 } ORDER BY ?nylawyer

```

Listing 1: Example SPARQL query over a Semantic Layer of a collection of old news articles. The query requests articles of 1989 discussing about New York lawyers born in Brooklyn.

of aspect and time. Expedition [35] is a time-aware search system for scholars which allows users to search articles in a news collection by entering free-text queries and choosing from four retrieval models: Temporal Relevance, Temporal Diversity, Topical Diversity, and Historical Diversity. Tempas [19] is a search system for web archives that exploits a social bookmarking service (Delicious) for temporally searching an archive by indexing tags and time. The new version of Tempas [20] exploits temporal link graphs and the corresponding anchor texts. The authors show how temporal anchor texts can be effective in answering queries beyond purely navigational intents, like finding the most central web pages of an entity in a given time period.

*Difference of our case.* Similar to the above works, our objective is to rank documents for a user information need. However, our case has the following three distinctive characteristics:

- i) The full contents of the documents are not available and there are no term-based indexes on top of them. We have access only on the RDF triples existing in the semantic layer, i.e., on metadata about the documents and on the entities mentioned in the documents. Moreover, these entities have been extracted using automated entity linking systems and thus are prone to disambiguation errors.
- ii) The information needs are expressed through structured SPARQL queries, not keywords. These SPARQL queries request documents of a specific time period mentioning one or more specific entities, while these query entities are specified through URIs which means that there is no ambiguity about them.

- iii) We already know the documents that match the query, however there are no relevance scores, i.e., all documents *equally* match the query. Our objective is to identify (and rank higher) the documents that discuss important information about the entities given in the query.

2.2.2 *Ranking in Knowledge Graphs.* There is a plethora of works on ranking entities, concepts and resources in knowledge graphs [7, 18, 30], as well as on ranking the results returned by SPARQL queries [15, 25, 29]. The majority of these works exploits the structure of the graph and applies some variation of a popular link analysis algorithm. The survey in [32] formalizes and contextualizes the problem of ranking in the Web of Data and provides an analysis and contrast of the similarities, differences and applicability of the different approaches.

An interesting related line of research tackles the problem of *ad-hoc object retrieval* [31, 37]. In this problem, the input is a keyword query and the output is one or more resources (entity URIs) that satisfy the corresponding information need. To tackle this kind of problem, Pound et al. [31] propose an adaptation of TF-IDF, while Tonon et al. [37] combine an inverted index with entity graph traversal. In the same context, the SemSearch challenge [16] focused on finding the entity identifier of a specific entity described by a user query, while the TREC entity track [4] studied two related search tasks: i) finding all entities related to a given entity, and ii) finding entities with common properties given some examples.

*Difference of our case.* In all these works, the result is a ranked list of *well-structured* RDF resources (like entities, properties, or triples). On the contrary, in our problem the units of retrieval represent unstructured (textual) documents returned by structured (SPARQL) queries. These works operate over knowledge graphs like DBpedia, where entities are described through properties and associations with other entities. A semantic layer is a special kind of a knowledge graph that represents metadata and annotation information about a collection of textual documents like news articles. Given a non-ambiguous SPARQL query and its result (the documents that match the query), our aim is to identify those documents that discuss important information about the query entities in the requested time-period by exploiting only the contents of the semantic layer.

### 3 PROBLEM DEFINITION AND EVALUATION DATASET

In this section, we formalize the problem of ranking documents returned by structured queries and describe a new ground truth dataset for the problem at hand. First we introduce the required notions and notations.

#### 3.1 Notions and Notations

*Entities.* In our problem, an *entity* is anything with a separate and meaningful existence that also has an identity expressed through a reference in a knowledge base (e.g., a Wikipedia/DBpedia URI). This does not only include persons, locations, organizations, etc., but also events (e.g., *US 2016 presidential election*) and more abstract concepts such as *democracy* or *abortion*. Let  $E$  be a finite set of entities, e.g., all Wikipedia entities, where each entity  $e \in E$  is associated with a unique URI in the reference knowledge base.

*Documents and extracted entities.* Let  $D$  be a set of documents (e.g., a set of news articles) published within a set of time periods  $T_D$  of fixed granularity  $\Delta$  (e.g., day). For a document  $d \in D$ , let  $t_d \in T_D$  be the time period of granularity  $\Delta$  in which  $d$  was published, while for a time period  $t \in T_D$ , let  $docs(t) \subseteq D$  be the set of all documents published within  $t$ , i.e.,  $docs(t) = \{d \in D \mid t_d = t\}$ . Let also  $ents(d) \subseteq E$  be all entities mentioned in  $d$  extracted using an entity linking system [34]. Inversely, for an entity  $e \in E$ , let  $docs(e) \subseteq D$  be all documents that mention  $e$ , i.e.,  $docs(e) = \{d \in D \mid e \in ents(d)\}$ .

#### 3.2 Problem Definition

Given a corpus of documents  $D$ , a set of entities  $E_D \subseteq E$  mentioned in documents of  $D$ , and a SPARQL query  $Q$  requesting documents from  $D$  published within a *time period*  $T_Q \subseteq T_D$  and related to one or more query entities  $E_Q \subseteq E_D$  with logical AND (mentioning all the query entities) or OR (mentioning at least one of the query entities) semantics, the problem is how to rank the documents  $D_Q \subseteq D$  that match  $Q$ .

Listing 2 shows an example SPARQL query requesting documents published in 1990 discussing about the entities *Nelson Mandela* and *Frederik Willem de Klerk* (logical AND semantics), while the query in Listing 3 requests articles of 1990 discussing about *state presidents of South Africa* (logical OR semantics). Our objective is to rank the results returned by such SPARQL queries.

#### 3.3 Evaluation dataset

Due to the lack of benchmark datasets for our problem, and to enable empirical evaluations, we have created a new ground truth dataset. We used the New York Times (NYT) annotated corpus [33] as the underlying document collection. The corpus contains over 1.8 million articles published by NYT between 1987 and 2007. We used Babelfy [28] for extracting DBpedia entities from each article, using a configuration proposed by the Babelfy developers<sup>1</sup>. We tested this configuration in the AIDA/CONLL-Test B ground truth dataset [17] and got the following evaluation scores: micro precision: 0.818, micro recall: 0.684, micro F1: 0.745 (the accuracy is

<sup>1</sup>The configuration is available at: <https://github.com/dice-group/gerbil/blob/master/src/main/java/org/aksw/gerbil/annotator/impl/babelfy/BabelfyAnnotator.java>

---

```

1 SELECT DISTINCT ?article WHERE {
2   ?article dc:date ?date FILTER(year(?date) = 1990) .
3   ?article oae:mentions ?entity1, ?entity2 .
4   ?entity1 oae:hasMatchedURI dbr:Nelson_Mandela .
5   ?entity2 oae:hasMatchedURI dbr:F._W._de_Klerk }

```

---

**Listing 2: SPARQL query for retrieving articles of 1990 discussing about *Nelson Mandela* and *Frederik Willem de Klerk* (logical AND semantics).**

---

```

1 SELECT DISTINCT ?article WHERE {
2   SERVICE <http://dbpedia.org/sparql> {
3     ?p dc:subject dbc:State_Presidents_of_South_Africa }
4   ?article dc:date ?date FILTER(year(?date) = 1990) .
5   ?article oae:mentions ?entity .
6   ?entity oae:hasMatchedURI ?p }

```

---

**Listing 3: SPARQL query for retrieving articles of 1990 discussing about *state presidents of South Africa* (logical OR semantics).**

almost the same with the one reported in the Babelfy paper for the same dataset [28]). Based on these annotations, we constructed a semantic layer following the process described in [11]. Then, we created 24 SPARQL queries, each one requesting articles published in a specific *time period* and mentioning one or more *entities*. The queries are grouped into 4 categories:

- **Single-entity queries (Q1-Q6):** 6 queries requesting articles related to 1 entity (e.g., articles of 1990 discussing about *Nelson Mandela*).
- **Multiple-entity AND queries (Q7-Q12):** 6 queries requesting articles related to 2 or more entities with logical AND semantics (e.g., articles of 1990 discussing about *Nelson Mandela* and *F.W. de Klerk*).
- **Multiple-entity OR queries (Q13-Q18):** 6 queries requesting articles related to 2 or more entities with logical OR semantics (e.g., articles of 1990 discussing about *Nelson Mandela* or *F.W. de Klerk*).
- **Category queries (Q19-Q24):** 6 queries requesting articles related to entities belonging to a DBpedia category (e.g., articles of 1990 discussing about *presidents of South Africa*<sup>2</sup>). This category is a special case of multiple-entity OR queries where the number of entities can be very large (hundreds or even thousands).

We manually evaluated all the results returned by these queries (773 results totally) using a graded relevance scale (from 0 to 3), following the criteria described below:

- **Score 0:** The document has almost nothing to do with the query entities.
- **Score 1:** The topic of the document is **not** about the query entities, however the query entities are related to the document context.
- **Score 2:** The topic of the document is **not** about the query entities, however the query entities are important for the document context.
- **Score 3:** The topic of the document is about the query entities and discusses something important about them.

Table 1 shows the number of results per relevance score for each of the queries. The semantic layer, the SPARQL queries, and

<sup>2</sup>Entities that have the value `<http://dbpedia.org/resource/Category:Presidents_of_South_Africa>` in their subject property (`<http://purl.org/dc/terms/subject>`).

**Table 1: Number of results per relevance score for each query of the evaluation dataset.**

Query	Total number of results	Score 0 results	Score 1 results	Score 2 results	Score 3 results
1	65	13	13	4	35
2	28	20	5	1	2
3	28	24	2	0	2
4	23	15	3	1	4
5	38	18	6	7	7
6	24	16	4	1	3
7	29	13	6	8	2
8	61	10	27	15	9
9	42	14	10	10	8
10	28	17	4	6	1
11	27	13	10	4	0
12	24	13	2	8	1
13	30	21	1	4	4
14	37	15	8	8	6
15	27	18	5	2	2
16	23	5	14	2	2
17	21	13	2	2	4
18	25	6	4	9	6
19	41	24	4	8	5
20	22	11	8	3	0
21	29	16	5	5	3
22	47	33	7	3	4
23	31	18	3	8	2
24	23	14	3	3	3

the relevance scores (together with explanations for the provided scores) are publicly available<sup>3</sup>.

## 4 PROBABILISTIC MODELING

The question is: "what makes an archived document important given a time period and one or more query entities"? We focus on an approach that makes use of the least amount of metadata information about the archived documents. This includes the *publication date* and the *extracted entities* of each document. Given this data, we have identified the following aspects that can affect the importance of a document to a query:

- the *relativeness* of a document with respect to the query entities (the document should talk about the query entities, ideally as its main topic).
- the *timeliness* of a document with respect to its publication date (the document should have been published in a time period which is important for the query entities).
- the *relatedness* of a document with respect to its reference to other entities (the document should discuss the relation of the query entities with other entities that are important for the query entities in important time periods).

The idea is to promote documents that: i) mention the query entities many times in their contents (because then, the topic of the document may be about these entities), ii) have been published in important (for the query entities) time periods, and iii) mention many other entities that co-occur frequently with the query entities in important time periods. For example, in case we want to rank articles of 1990 discussing about *Nelson Mandela*, we want to favor articles that i) discuss about *Nelson Mandela* as their main topic, ii)

have been published in important (for *Nelson Mandela*) time periods (e.g., February of 1990 since during that period he was released from prison), and iii) mention other entities that seem to be important for Nelson Mandela during important time periods (e.g., *Frederik Willem de Klerk* who was South Africa's State President in 1990).

### 4.1 Relativeness

We consider that if the query entities are mentioned multiple times within a document, the document should receive a high score since the document's topic may be about these entities. The term frequency (in our case entity frequency) is a classic numerical statistic that is intended to reflect how important a word (entity) is to a document [26].

We first define a *relativeness* score of a document  $d \in D_Q$  based on the *frequency* of the query entities in  $d$ . First, let  $count(e, d)$  be the number of occurrences of  $e$  in document  $d$ . For the case of AND semantics (" $\wedge$ "), the score is defined as:

$$score_{\wedge}^f(d, E_Q) = \frac{\sum_{e \in E_Q} count(e, d)}{\sum_{e' \in ents(d)} count(e', d)} \quad (1)$$

Notice that the score of a document will be 1.0 if it contains the query entities and no other entity. For the case of OR semantics (" $\vee$ "), we can also consider the number of query entities mentioned in the document (since a document does not probably contain all the query entities as in the case of AND semantics). In this case, the *relativeness* score can be defined as follows:

$$score_{\vee}^f(d, E_Q) = \frac{\sum_{e \in E_Q} count(e, d)}{\sum_{e' \in ents(d)} count(e', d)} \cdot \frac{|ents(d) \cap E_Q|}{|E_Q|} \quad (2)$$

where  $\frac{|ents(d) \cap E_Q|}{|E_Q|}$  is the percentage of query entities discussed in the document. The score of a document will be 1.0 if it contains all the query entities and no other entity. This formula favors documents mentioning many of the query entities multiple times.

Now, the probability of a retrieved document  $d \in D_Q$  given only the query entities can be defined as:

$$P(d|E_Q) = \frac{score^f(d, E_Q)}{\sum_{d' \in D_Q} score^f(d', E_Q)} \quad (3)$$

### 4.2 Timeliness

Previous works on searching document archives have shown that, considering the fraction of documents published in a time period and mentioning the query entities can improve the effectiveness of document retrieval (compared to an approach that assumes equal distribution) [36]. In a similar way, and for the case of AND semantics, we define the following importance score of a *time period*  $t \in T_Q$ :

$$score_{\wedge}^t(t) = \frac{|docs(t) \cap D_Q|}{|D_Q|} \quad (4)$$

This scoring formula favors time periods in which there is a large number of documents discussing about the query entities.

For the case of OR semantics, in a time period  $t$  there may be a large number of documents discussing only for one of the query entities, while in another time period  $t'$  there may be a smaller number of documents discussing though for many of the query

<sup>3</sup>[http://l3s.de/~fafalios/jcdl/evaluation\\_dataset.zip](http://l3s.de/~fafalios/jcdl/evaluation_dataset.zip)

entities. For also taking into account the number of query entities discussed in documents of a specific time period, we consider the following formula:

$$score_{\vee}^t(t) = \frac{|docs(t) \cap D_Q|}{|D_Q|} \cdot N(E_Q, t) \quad (5)$$

where,  $N(E_Q, t)$  is the average percentage of query entities discussed in articles of  $t$ , i.e.:

$$N(E_Q, t) = \frac{\sum_{d \in docs(t) \cap D_Q} \frac{|ents(d) \cap E_Q|}{|E_Q|}}{|docs(t) \cap D_Q|} \quad (6)$$

Now, the probability of a retrieved document  $d \in D_Q$  given only its publication date  $t_d$  can be defined as:

$$P(d|t_d) = \frac{score^t(t_d)}{\sum_{d' \in D_Q} score^t(t_{d'})} \quad (7)$$

### 4.3 Relatedness

Recent works have shown that the co-occurrence of entities in documents of a specific time period is a strong indicator of their relatedness during that period [38, 40]. We also consider that entities that are co-mentioned frequently with the query entities in important time periods are probably important for them. However, there may be also some general entities that co-occur with the query entities in almost all documents (independently of the time period). Thus, we should also avoid over-emphasizing documents mentioning such ‘‘common’’ entities.

For the case of AND semantics, we consider the following *relatedness* score for an entity  $e \in E_D \setminus E_Q$ :

$$\begin{aligned} score_{\wedge}^r(e) &= idf_{\wedge}(e) \cdot \sum_{t \in T_Q} (score_{\wedge}^t(t) \cdot \frac{|docs(t) \cap D_Q \cap docs(e)|}{|docs(t) \cap D_Q|}) \\ &= idf_{\wedge}(e) \cdot \sum_{t \in T_Q} \frac{|docs(t) \cap D_Q \cap docs(e)|}{|D_Q|} \end{aligned} \quad (8)$$

where  $idf_{\wedge}(e)$  is the inverse document frequency of entity  $e$  in the set of documents discussing about the query entities in the entire corpus, which can be defined as follows:

$$idf_{\wedge}(e) = 1 - \frac{|docs(e) \cap (\cap_{e' \in E_Q} docs(e'))|}{|\cup_{e' \in E_Q} docs(e')|} \quad (9)$$

The formula considers the percentage of documents in which the entity co-occurs with the query entities in important time periods.

For the case of OR semantics, the above formula does not consider the number of different query entities discussed in documents together with the entity  $e$ . To also handle this aspect, we consider the following *relatedness* score for the case of OR semantics:

$$\begin{aligned} score_{\vee}^r(e) &= idf_{\vee}(e) N(E_Q, e) \sum_{t \in T_Q} (score_{\vee}^t(t) \frac{|docs(t) \cap D_Q \cap docs(e)|}{|docs(t) \cap D_Q|}) \\ &= idf_{\vee}(e) N(E_Q, e) \sum_{t \in T_Q} (N(E_Q, t) \frac{|docs(t) \cap D_Q \cap docs(e)|}{|D_Q|}) \end{aligned} \quad (10)$$

where  $N(E_Q, e)$  is the average percentage of query entities discussed in articles together with entity  $e$ , i.e.:

$$N(E_Q, e) = \frac{\sum_{d \in docs(e) \cap D_Q} \frac{|ents(d) \cap E_Q|}{|E_Q|}}{|docs(e) \cap D_Q|} \quad (11)$$

Now the inverse document frequency  $idf_{\vee}(e)$  includes documents mentioning at least one of the query entities, i.e.:

$$idf_{\vee}(e) = 1 - \frac{|docs(e) \cap (\cup_{e' \in E_Q} docs(e'))|}{|\cup_{e' \in E_Q} docs(e')|} \quad (12)$$

This formula favors related entities that i) co-occur frequently with many of the query entities, ii) are discussed in documents published in important (for the query entities) time periods.

Now, the probability of a document  $d \in D_Q$  given only other entities mentioned in the retrieved documents ( $E_{D_Q}$ ) can be defined as:

$$P(d|E_{D_Q}) = \frac{\sum_{e \in ents(d) \setminus E_Q} score^r(e)}{\sum_{d' \in D_Q} \sum_{e' \in ents(d') \setminus E_Q} score^r(e')} \quad (13)$$

### 4.4 Joining the Models

We can now combine the different models in a single probability score:

$$P(d|E_Q, t_d, E_{D_Q}) = \frac{P(d|E_Q)P(d|T_Q)P(d|E_{D_Q})}{\sum_{d' \in D_Q} P(d'|E_Q)P(d'|T_Q)P(d'|E_{D_Q})} \quad (14)$$

where the denominator can be ignored as it does not influence the ranking.

## 5 STOCHASTIC MODELING

Here we model the problem as a *random walker* on the graph (*Markov chain*) defined by the query-entities  $E_Q$ , the returned documents  $D_Q$ , and the entities mentioned in the documents  $E_{D_Q}$ . Then, we propose a biased (personalized-like) PageRank algorithm for analyzing the graph and scoring its nodes. Similar modeling and scoring methods have been applied for the problems of *results re-ranking* [13] and *enrichment* [12] in information retrieval.

### 5.1 The Transition Graph

The walker starts from a query-entity and can move either to a document mentioning the entity or to a related entity that co-occurs with the query-entity in at least one document. From a document, the walker can move to an entity mentioned in it, while from a no query-entity, the walker can only move to a document mentioning that entity.

In the case of logical AND semantics, the query-entities are connected with all documents, while in the case of OR semantics, each query-entity is connected with at least one document. Figure 2 shows an example of a transition graph for the case of OR semantics. In this example, the query-entities are two (the black nodes), while we notice that three of the documents mention both query-entities ( $d_2$ ,  $d_3$  and  $d_4$ ).

### 5.2 The Transition Probabilities

*From query-entities to documents or related entities.* When the walker lies at a query-entity  $e \in E_Q$ , he can either move to a document  $d$  mentioning the entity or to a related entity  $e'$ . When

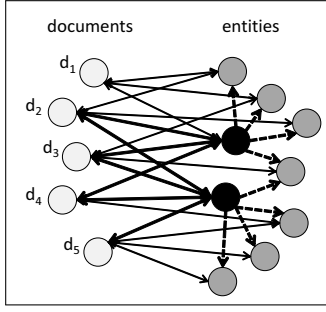


Figure 2: An example of the considered transition graph for the case of logical OR semantics (with the query-entities being the black nodes).

moving to a document, we consider its *relativeness* and *timeliness* scores as introduced in the previous section. Specifically, the weight of the edge from a query-entity  $e$  to a document  $d$  is defined as:

$$weight(e \rightarrow d) = \frac{score^f(d, E_Q) \cdot score^t(t_d)}{\sum_{d' \in docs(e) \cap D_Q} (score^f(d', E_Q) \cdot score^t(t_{d'}))} \quad (15)$$

For moving from a query-entity  $e$  to a related entity  $e'$ , we consider the *relatedness* score of  $e'$ :

$$weight(e \rightarrow e') = \frac{score^r(e')}{\sum_{e'' \in out(e)} (score^r(e''))} \quad (16)$$

where  $out(e)$  is the set of nodes connected to  $e$  through outgoing edges starting from  $e$ .

However, the weights of the outgoing edges of a single node represent transition probabilities and must sum to 1. Thus, the weight of the edge from a query-entity  $e$  to a connected node  $n$  (that can be either a document or a related entity) is defined as:

$$weight(e \rightarrow n) = \begin{cases} p_1 \cdot weight(e \rightarrow d) & n \in D_Q \\ (1 - p_1) \cdot weight(e \rightarrow e') & n \in E_{D_Q} \end{cases} \quad (17)$$

where  $p_1 \in [0, 1]$  is the probability that the walker selects to move to a document node.

*From documents to entities.* For moving from a document  $d$  to an entity  $e$  mentioned in  $d$ , it is more likely that the walker moves to an entity that is mentioned many times within its contents. Thus, we simply consider the normalized frequency of  $e$  in  $d$ . Specifically:

$$weight(d \rightarrow e) = \frac{count(e, d)}{\sum_{e' \in ents(d)} (count(e', d))} \quad (18)$$

*From no query-entities to documents.* Likewise, when moving from a no query-entity  $e$  to a document  $d$ , it is more likely that the walker moves to a document that mentions  $e$  many times. Thus, we again consider the normalized frequency of  $e$  in  $d$ . Specifically:

$$weight(e \rightarrow d) = \frac{count(e, d)}{\sum_{d' \in docs(e) \cap D_Q} (count(e, d'))} \quad (19)$$

### 5.3 The Stochastic Analysis (Random Walk with Restart)

For analyzing the transition graph, we follow a PageRank-like algorithm. The walker starts from the query-entities and can either follow an edge or perform a “restart”, i.e., jump to a query-entity and start again the traversal. Formally, the score of a graph node  $n$  is defined as:

$$r(n) = d \cdot Jump(n) + (1 - d) \cdot \sum_{n' \in in(n)} (weight(n' \rightarrow n) \cdot r(n')) \quad (20)$$

where  $d \in [0, 1]$  is the probability that the walker performs a restart,  $Jump(n)$  is the probability the walker to restart by jumping to node  $n$ ,  $in(n)$  is the set of nodes connected to  $n$  through incoming edges, and  $weight(n \rightarrow n')$  (as defined in Formulas 15-19) is the probability that the walker visits  $n$  when being at node  $n'$  connected to  $n$  (there should be an edge from  $n'$  to  $n$ ).

As regards the *restart*, we allow equiprobable jumps only to query entities, i.e.:

$$Jump(n) = \begin{cases} 1/|E_Q| & n \in E_Q \\ 0 & n \notin E_Q \end{cases} \quad (21)$$

Regarding the initial scores of the graph nodes, we assign the same score to the query entities ( $1/|E_Q|$ ), while the score of all other entities is zero. Finally, the algorithm must iteratively run to convergence.

## 6 EVALUATION

We evaluated the performance of the proposed models using the ground truth dataset described in Section 3.3 and two evaluation measures: *Normalized Discounted Cumulative Gain (NDCG)* at positions 5, 10, full list, and *Precision (P)* at 5 and 10 (precision for full list is the same in all cases). Precision is the fraction of relevant results among the top retrieved results, while NDCG uses a graded relevance scale and measures the usefulness, or gain, of a document based on its position in the result list [21]. For precision, we consider a document as relevant if its relevance score is either 2 or 3 and irrelevant if it is either 0 or 1. As regards timeliness, we considered *day* granularity. We also tested the case of random rankings of the results. In this case, we computed 10 different random lists for each query and considered the average NDCG and precision scores.

### 6.1 Effectiveness of Probabilistic Models

We investigated the performance of each model described in Section 4 as well as their combinations.

*6.1.1 Overall Results.* Table 2 shows the average NDCG and precision scores for all queries. We notice that joining all three models provides the best results. The improvement of the top-5 results compared to *relativeness* is about 17% in NDCG and 18% in precision, which is statistically significant (paired t-test,  $p \leq 0.05$ ). Note that *relativeness* can be considered a baseline model for our problem since it considers entity frequency which is a classic numerical statistic [26]. The results also show that *relatedness* performs very well, outperforming the joined models for P@10. This means that considering other entities that co-occur frequently



with the query entities in important time periods has a positive effect on the ranking.

**6.1.2 Detailed results per query type.** Tables 3-6 show the results per query type.

Regarding *single entity queries* (Table 3), we observe that combining relativeness with relatedness has almost the same performance with the case of joining all models. This means that timeliness seems to have a minor negative effect on the rankings. For this query type, query 4 has the best performance on the joined model with  $NDCG@5 = 1.0$  and  $NDCG@10 = 0.97$ . On the contrary, query 5 does not perform well with  $NDCG@5 = 0.29$  and  $NDCG@10 = 0.37$  (this query returns several "headlines" articles summarizing what is inside today's paper, which may confuse our modeling).

As regards *multiple-entity AND queries* (Table 4), we observe that joining relativeness with timeliness seems to perform better with small difference compared to the case of joining of all models, which means that here timeliness has a positive effect on the ranking. We also notice that all models perform worst compared to the single-entity case. For this query type, the  $NDCG@10$  score of all queries ranges from 0.35 (for query 8) to 0.67 (for query 12).

The case of *multiple-entity OR queries* (Table 5) is very similar to the case of *single entity queries* where timeliness seems to have a minor negative impact. Here almost all queries have very good performance with  $NDCG@10$  score  $> 0.55$ .

As regards *category queries* (Table 6), it is very interesting that the performance of relativeness is very low, while relatedness performs very well and timeliness has a positive impact especially for the top-5 lists. By thoroughly checking the results of queries that perform very bad in relativeness (like queries 19 and 22), we noticed that the reason for this bad performance is the presence of a large number of disambiguation errors in the returned documents. A characteristic of this type of queries is the big number of query entities which in turn increases the probability of disambiguation errors. Note that relativeness favors documents containing many instances of the query entities. However, if an irrelevant document contains many instances of entity names that have been wrongly linked to query entities, this document will receive a high relativeness score (since it contains many "false positive" query entities). For example, query 19 in our dataset has a very bad performance in relativeness ( $NDCG@5 = 0.02$ ,  $NDCG@10 = 0.10$ ). This query requests articles mentioning NASA civilian astronauts, however many instances of the entity name *Armstrong* have been incorrectly linked to the American astronaut Neil Armstrong (like the referee Hank Armstrong or the company Armstrong World Industries). Likewise, in query 22 several instances of the entity names *Teresa* and *Theresa* are incorrectly mapped to Mother Teresa. Nevertheless, we see that *relatedness* somehow fixes this problem. This model considers the co-occurrence of entities in important time periods and favors documents that contain many instances of other important entities. The probability that many false positive cases co-occur with the same related entities in many documents is very low, and thus considering entity associations has a positive effect on the ranking. In query 19, for example, the incorrect instances of the query entity Neil Armstrong co-occur with entities that do not exist in the documents in which Neil Armstrong has been correctly

identified, and thus the relatedness score (Equation 10) of these entities is low.

## 6.2 Effectiveness of Stochastic Model

We run experiments for different values of  $d$  probability (probability to perform a restart, cf. Equation 20) and different values of  $p_1$  probability (probability to move to a document node when being at a query-entity node, cf. Equation 17). For the restart probability, we tested the following 5 cases:  $d = 0.0, 0.2, 0.4, 0.6, 0.8$ . Note that testing  $d = 1.0$  does not make sense, since the walker will never reach document nodes. Regarding the  $p_1$  probability, we tested 6 cases ( $p_1 = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$ ). Empirically,  $d = 0.2$  provides the best results independently of the  $p_1$  value. Thus, all the results reported below correspond to  $d = 0.2$ . Moreover, in all experiments, we set the number of iterations to 30.

Tables 7-10 show the results per query type and for different values of  $p_1$  probability. For the first three types of queries, we observe that  $p_1 = 1.0$  provides the best results and that, as the probability gets lower, the results get worse. Using  $p_1 = 1.0$ , the walker can move only to document nodes. This means that for small values of  $p_1$  probability, the algorithm overemphasizes the association between the entities resulting in worst performance. Thus, the results show that reaching documents through related entities affects negatively the rankings for these types of queries.

Compared to the results of the probabilistic modeling, we see that: i) the performance of the two models is almost the same for single-entity queries, while the stochastic model outperforms the best probabilistic model for multiple-entity AND queries, and ii) the best probabilistic model outperforms the stochastic model for multiple-entity OR queries. Regarding (ii), this failure of the stochastic model is probably due to the fact that the graph in the case of OR queries is not as well-connected as in the case of AND queries (for OR queries, the graph may contain disconnected components, e.g., in cases where there are no documents mentioning all query entities).

However, for the category queries (Table 10), we see that  $p_1 = 1.0$  provides the worst results, while  $p_1 = 0.4$  performs better. This means that, for this type of queries, considering the associations of the query entities with other entities mentioned in the retrieved documents affects positively the results. However, at the same time, we see that we should not give too much emphasis to this (by giving very low value to  $p_1$ ). Notice that this is in correspondence to the evaluation results of the probabilistic models for the same type of queries. Moreover, we notice that the stochastic model for  $p_1 = 0.4$  performs better than the best probabilistic model, providing better rankings and more relevant results in the top positions.

## 6.3 Synopsis of Evaluation Results

Based on the evaluation results, we can conclude that:

- *Category queries* is a special case where the number of query entities can be very large and this makes the results more susceptible to disambiguation errors of the used entity linking system. Thereby, for this type of queries one should select a model which considers the associations between the query-entities and other entities mentioned in the returned



**Table 2: Average NDCG and Precision of the probabilistic models for all queries.**

Measure	Random ranking	Relativeness [A]	Timeliness [B]	Relatedness [C]	[A][B]	[A][C]	[B][C]	[A][B][C]
NDCG@5	0.26	0.48	0.27	0.41	0.53	0.52	0.45	<b>0.56</b>
NDCG@10	0.33	0.52	0.36	0.50	0.55	0.56	0.51	<b>0.58</b>
NDCG@all	0.68	0.79	0.69	0.76	0.80	0.81	0.76	<b>0.82</b>
P@5	0.27	0.44	0.28	0.48	0.48	0.50	0.45	<b>0.52</b>
P@10	0.29	0.38	0.30	<b>0.45</b>	0.37	0.42	0.41	0.41

**Table 3: Average NDCG and Precision of the probabilistic models for single-entity queries (Q1-Q6).**

Measure	[A]	[B]	[C]	[A][B]	[A][C]	[B][C]	[A][B][C]
NDCG@5	0.66	0.30	0.40	0.68	0.69	0.45	<b>0.70</b>
NDCG@10	0.69	0.38	0.51	0.70	<b>0.71</b>	0.51	0.70
NDCG@all	<b>0.88</b>	0.67	0.75	0.87	<b>0.88</b>	0.72	0.87
P@5	0.57	0.23	0.50	0.57	<b>0.60</b>	0.40	<b>0.60</b>
P@10	0.40	0.27	<b>0.45</b>	0.40	0.40	0.38	0.40

**Table 4: Average NDCG and Precision of the probabilistic models for multiple-entity AND queries (Q7-Q12).**

Measure	[A]	[B]	[C]	[A][B]	[A][C]	[B][C]	[A][B][C]
NDCG@5	0.34	0.28	0.31	<b>0.42</b>	0.35	0.38	0.40
NDCG@10	0.43	0.33	0.40	0.46	0.45	0.46	<b>0.47</b>
NDCG@all	0.76	0.71	0.75	<b>0.78</b>	0.76	0.77	0.77
P@5	0.43	0.33	0.30	<b>0.50</b>	0.47	0.43	<b>0.50</b>
P@10	0.50	0.32	0.42	0.48	<b>0.53</b>	0.47	0.50

**Table 5: Average NDCG and Precision of the probabilistic models for multiple-entity OR queries (Q13-Q18).**

Measure	[A]	[B]	[C]	[A][B]	[A][C]	[B][C]	[A][B][C]
NDCG@5	0.68	0.24	0.44	0.70	<b>0.72</b>	0.46	0.71
NDCG@10	0.69	0.36	0.55	0.69	0.72	0.54	<b>0.73</b>
NDCG@all	0.87	0.69	0.79	0.87	<b>0.88</b>	0.79	<b>0.88</b>
P@5	0.60	0.27	0.57	0.60	<b>0.63</b>	0.43	<b>0.63</b>
P@10	0.42	0.30	<b>0.48</b>	0.40	0.45	0.42	0.45

**Table 6: Average NDCG and Precision of the probabilistic models for category queries (Q19-Q22).**

Measure	[A]	[B]	[C]	[A][B]	[A][C]	[B][C]	[A][B][C]
NDCG@5	0.22	0.27	0.48	0.34	0.34	<b>0.52</b>	0.41
NDCG@10	0.28	0.37	<b>0.53</b>	0.36	0.38	<b>0.53</b>	0.43
NDCG@all	0.66	0.68	<b>0.78</b>	0.70	0.71	0.77	0.73
P@5	0.17	0.27	<b>0.53</b>	0.23	0.30	<b>0.53</b>	0.33
P@10	0.20	0.30	<b>0.45</b>	0.20	0.28	0.37	0.28

**Table 7: Average NDCG and Precision of the stochastic model for single entity queries (Q1-Q6).**

Measure	$p_1=0.0$	$p_1=0.2$	$p_1=0.4$	$p_1=0.6$	$p_1=0.8$	$p_1=1.0$
NDCG@5	0.09	0.27	0.48	0.63	0.65	<b>0.67</b>
NDCG@10	0.18	0.35	0.54	0.69	0.72	<b>0.74</b>
NDCG@all	0.60	0.68	0.77	0.85	<b>0.87</b>	<b>0.87</b>
P@5	0.07	0.23	0.43	0.47	0.53	<b>0.60</b>
P@10	0.12	0.23	0.33	0.37	0.42	<b>0.43</b>

**Table 8: Average NDCG and Precision of the stochastic model for multiple-entity AND queries (Q7-Q12).**

Measure	$p_1=0.0$	$p_1=0.2$	$p_1=0.4$	$p_1=0.6$	$p_1=0.8$	$p_1=1.0$
NDCG@5	0.29	0.29	0.32	0.31	0.37	<b>0.48</b>
NDCG@10	0.36	0.36	0.40	0.45	0.47	<b>0.52</b>
NDCG@all	0.72	0.73	0.74	0.75	0.78	<b>0.80</b>
P@5	0.27	0.27	0.27	0.30	0.37	<b>0.57</b>
P@10	0.30	0.30	0.38	0.45	0.47	<b>0.50</b>

**Table 9: Average NDCG and Precision of the stochastic model for multiple-entity OR queries (Q13-Q18).**

Measure	$p_1=0.0$	$p_1=0.2$	$p_1=0.4$	$p_1=0.6$	$p_1=0.8$	$p_1=1.0$
NDCG@5	0.20	0.30	0.32	0.38	0.45	<b>0.59</b>
NDCG@10	0.27	0.38	0.42	0.51	0.52	<b>0.65</b>
NDCG@all	0.66	0.71	0.72	0.75	0.76	<b>0.84</b>
P@5	0.27	0.33	0.33	0.37	0.47	<b>0.53</b>
P@10	0.30	0.33	0.35	0.42	0.42	<b>0.45</b>

**Table 10: Average NDCG and Precision of the stochastic model for category queries (Q19-Q24).**

Measure	$p_1=0.0$	$p_1=0.2$	$p_1=0.4$	$p_1=0.6$	$p_1=0.8$	$p_1=1.0$
NDCG@5	0.43	0.47	<b>0.53</b>	0.50	0.46	0.23
NDCG@10	0.48	0.54	<b>0.55</b>	0.52	0.54	0.36
NDCG@all	0.74	0.77	<b>0.79</b>	0.77	0.77	0.68
P@5	0.57	0.57	<b>0.60</b>	<b>0.60</b>	<b>0.60</b>	0.23
P@10	0.42	<b>0.43</b>	0.40	0.40	<b>0.43</b>	0.28

documents, since this limits the negative impact of this problem. Finally, a stochastic model with  $p_1 = 0.4$  outperforms the best probabilistic model (relatedness).

- For the other types of queries, a model which considers both *relativeness* and *relatedness* should be selected, while *timeliness* does not seem to significantly affect the rankings. For single-entity and multiple-entity AND queries, one may select the stochastic model which seems to perform better than the probabilistic models, however for multiple-entity OR queries where the graph is not so well connected, one may opt for a probabilistic model.

## 7 CONCLUSIONS

We have formalized the problem of ranking archived documents for structured (SPARQL) queries on semantic layers, i.e., on RDF graphs describing metadata and annotation information about the documents. To cope with this problem, we have proposed two ranking models, a probabilistic one and a stochastic one, which jointly

consider the following aspects: i) the *relativeness* of the documents to the query entities, ii) the *timeliness* of the documents, and iii) the temporal *relatedness* of other entities mentioned in the documents to the query entities. To evaluate our approach, and due to lack of evaluation datasets for the problem at hand, we carefully created a new ground truth dataset which we make publicly available for fostering further research on similar problems. The evaluation results showed that the proposed models can identify important - for the query entities - documents, achieving high NDCG and precision scores and outperforming a classic, frequency-based baseline model. The results also showed that *relatedness*, i.e., considering the temporal association of the query entities with other entities mentioned in the results, has a high positive impact on the ranking and can limit the negative effect caused by disambiguation errors of entity linking.

Regarding future work, we plan to study the applicability of similar models on other types of archives, like web archives where much more noisy (and probably spam) data exists and the archived documents (web pages) contain multiple similar or identical versions. Another interesting direction is the study of diversity-aware ranking methods that can reflect the diversity of the retrieved documents. We also plan to build user-friendly interfaces on top of semantic layers which will make use of the proposed ranking models and will allow end-users to easily explore the archives.

## ACKNOWLEDGMENTS

The work was partially funded by the European Commission for the ERC Advanced Grant ALEXANDRIA (No. 339233).

## REFERENCES

- [1] Ablimit Aji, Yu Wang, Eugene Agichtein, and Evgeniy Gabrilovich. 2010. Using the past to score the present: Extending term weighting models through revision history analysis. In *19th CIKM*. ACM.
- [2] Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, Dmitriy Zheleznyakov, and Ernesto Jimenez-Ruiz. 2014. SemFacet: semantic faceted search over yago. In *International Conference on World Wide Web*. ACM.
- [3] Irem Arikian, Srikanta Bedathur, and Klaus Berberich. 2009. Time Will Tell: Leveraging Temporal Expressions in IR. In *2nd ACM International Conference on Web Search and Data Mining*. ACM.
- [4] Krisztian Balog, Pavel Serdyukov, and Arjen P De Vries. 2010. Overview of the trec 2010 entity track. In *In TREC 2010*.
- [5] Klaus Berberich, Srikanta Bedathur, Omar Alonso, and Gerhard Weikum. 2010. A Language Modeling Approach for Temporal Information Needs. In *European Conference on Information Retrieval*. Springer, 13–25.
- [6] Karen Calhoun. 2014. *Exploring digital libraries: foundations, practice, prospects*. Facet Publishing.
- [7] Stéphane Campinas, Renaud Delbru, and Giovanni Tummarello. 2012. Effective retrieval model for entity with multi-valued attributes: Bm25mf and beyond. In *International Conference on Knowledge Engineering and Knowledge Management*. Springer, 200–215.
- [8] Ricardo Campos, Gaël Dias, Alípio M Jorge, and Adam Jatowt. 2015. Survey of temporal information retrieval and related applications. *ACM Computing Surveys (CSUR)* 47, 2 (2015), 15.
- [9] Mario Cataldi, Luigi Di Caro, and Claudio Schifanella. 2010. Emerging topic detection on twitter based on temporal and social terms evaluation. In *Proceedings of the tenth international workshop on multimedia data mining*. ACM, 4.
- [10] Wisam Dakka, Luis Gravano, and Panagiotis Ipeirotis. 2012. Answering general time-sensitive queries. *IEEE Transactions on Knowledge and Data Engineering* 24, 2 (2012), 220–235.
- [11] P. Fafalios, H. Holzmann, V. Kasturia, and W. Nejdl. 2017. Building and Querying Semantic Layers for Web Archives. In *ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '17)*. Toronto, Ontario, Canada.
- [12] P. Fafalios and Y. Tzitzikas. 2014. Post-Analysis of Keyword-based Search Results using Entity Mining, Linked Data and Link Analysis at Query Time. In *2014 IEEE Eighth International Conference on Semantic Computing (ICSC 2014)*. IEEE, Newport Beach, California, USA.
- [13] Pavlos Fafalios and Yannis Tzitzikas. 2017. Stochastic reranking of biomedical search results based on extracted entities. *Journal of the Association for Information Science and Technology* 68, 11 (2017), 2572–2586.
- [14] Sébastien Ferré. 2014. Sparklis: a sparql endpoint explorer for expressive question answering. In *ISWC Posters & Demonstrations Track*.
- [15] Azam Feyznia, Mohsen Kahani, and Fattane Zarrinkalam. 2014. COLINA: A Method for Ranking SPARQL Query Results through Content and Link Analysis. In *International Semantic Web Conference (Posters & Demos)*. 273–276.
- [16] Harry Halpin, Daniel M Herzig, Peter Mika, Roi Blanco, Jeffrey Pound, Henry S Thompson, and Thanh Tran Duc. 2010. Evaluating ad-hoc object retrieval. In *International Workshop on Evaluation of Semantic Technologies, IWEST*. Citeseer.
- [17] Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust disambiguation of named entities in text. In *Conference on Empirical Methods in Natural Language Processing*.
- [18] Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres, and Stefan Decker. 2011. Searching and browsing linked data with SWSE: the semantic Web search engine. *Web Semantics: Science, Services and Agents on the World Wide Web* 9, 4 (2011), 365–401.
- [19] Helge Holzmann and Avishek Anand. 2016. Tempas: Temporal Archive Search Based on Tags. In *International Conference on World Wide Web*.
- [20] Helge Holzmann, Wolfgang Nejdl, and Avishek Anand. 2017. Exploring Web Archives Through Temporal Anchor Texts. In *Proceedings of the 2017 ACM on Web Science Conference*. ACM, 289–298.
- [21] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.
- [22] Peiquan Jin, Jianlong Lian, Xujian Zhao, and Shouhong Wan. 2008. Tise: A temporal search engine for web contents. In *Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on*, Vol. 3. IEEE.
- [23] Nattiya Kanhabua, Roi Blanco, Kjetil Nørvåg, et al. 2015. Temporal information retrieval. *Foundations and Trends® in Information Retrieval* 9, 2 (2015), 91–208.
- [24] Nattiya Kanhabua and Kjetil Nørvåg. 2012. Learning to rank search results for time-sensitive queries. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2463–2466.
- [25] Sara Latifi and Mohammadali Nematbakhsh. 2014. Query-independent learning to rank RDF entity results of SPARQL queries. In *Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on*. IEEE.
- [26] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. 2014. *Mining of massive datasets*. Cambridge University Press.
- [27] Donald Metzler, Rosie Jones, Fuchun Peng, and Ruiqiang Zhang. 2009. Improving search relevance for implicitly temporal queries. In *32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 700–701.
- [28] Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics* 2 (2014).
- [29] Kunal Mulay and P Sreenivasa Kumar. 2011. SPRING: Ranking the results of SPARQL queries on Linked Data. In *Proceedings of the 17th International Conference on Management of Data*. Computer Society of India, 12.
- [30] Axel-Cyrille Ngonga Ngomo, Michael Hoffmann, Ricardo Usbeck, and Kunal Jha. 2017. Holistic and Scalable Ranking of RDF Data. In *2017 IEEE International Conference on Big Data*.
- [31] Jeffrey Pound, Peter Mika, and Hugo Zaragoza. 2010. Ad-hoc object retrieval in the web of data. In *19th international conference on World wide web*. ACM.
- [32] Antonio J Roa-Valverde and Miguel-Angel Sicilia. 2014. A survey of approaches for ranking on the web of data. *Information Retrieval* 17, 4 (2014), 295–325.
- [33] Evan Sandhaus. 2008. The new york times annotated corpus. *Linguistic Data Consortium, Philadelphia* 6, 12 (2008).
- [34] Wei Shen, Jianyong Wang, and Jiawei Han. 2015. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering* 27, 2 (2015), 443–460.
- [35] Jaspreet Singh, Wolfgang Nejdl, and Avishek Anand. 2016. Expedition: A Time-Aware Exploratory Search System Designed for Scholars. In *SIGIR conference on Research and Development in Information Retrieval*.
- [36] Jaspreet Singh, Wolfgang Nejdl, and Avishek Anand. 2016. History by diversity: Helping historians search news archives. In *ACM Conference on Human Information Interaction and Retrieval*.
- [37] Alberto Tonon, Gianluca Demartini, and Philippe Cudré-Mauroux. 2012. Combining inverted indices and structured search for ad-hoc object retrieval. In *ACM SIGIR*. ACM, 125–134.
- [38] Nam Khanh Tran, Tuan Tran, and Claudia Niderée. 2017. Beyond Time: Dynamic Context-Aware Entity Recommendation. In *European Semantic Web Conference*.
- [39] Yannis Tzitzikas, Nikos Manolis, and Panagiotis Papadakos. 2016. Faceted exploration of RDF/S datasets: a survey. *Journal of Intelligent Information Systems* (2016), 1–36.
- [40] Lei Zhang, Achim Rettinger, and Ji Zhang. 2016. A Probabilistic Model for Time-Aware Entity Recommendation. In *International Semantic Web Conference*.