Parameterized Fielded Term Dependence Models for Ad-hoc Entity Retrieval from Knowledge Graph

Fedor Nikolaev
Department of Computer Science, Wayne State University, Detroit, Michigan 48202
fedor@wayne.edu
Kazan Federal University, 35 Kremlyovskaya Street, Kazan, Russia 420008
feanikolaev@kpfu.ru

Alexander Kotov*
Department of Computer Science, Wayne State University, Detroit, Michigan 48202
kotov@wayne.edu

Nikita Zhiltsov
Kazan Federal University, 35 Kremlyovskaya Street, Kazan, Russia 420008
nzhilcov@kpfu.ru

ABSTRACT

Accurate projection of terms in free-text queries onto structured entity representations is one of the fundamental problems in entity retrieval from knowledge graph. In this paper, we demonstrate that existing retrieval models for ad-hoc structured and unstructured document retrieval fall short of addressing this problem, due to their rigid assumptions. According to these assumptions, either all query concepts of the same type (unigrams and bigrams) are projected onto the fields of entity representations with identical weights or such projection is determined based only on one simple statistic, which makes it sensitive to data sparsity. To address this issue, we propose the Parametrized Fielded Sequential Dependence Model (PFSDM) and the Parametrized Fielded Full Dependence Model (PFFDM), two novel models for entity retrieval from knowledge graphs, which infer the user's intent behind each individual query concept by dynamically estimating its projection onto the fields of structured entity representations based on a small number of statistical and linguistic features. Experimental results obtained on several publicly available benchmarks indicate that PFSDM and PFFDM consistently outperform state-of-the-art retrieval models for the task of entity retrieval from knowledge graph.

CCS Concepts

•**Information systems** → **Structured text search;**

Keywords

Entity Retrieval; Structured Document Retrieval; Feature-based Models; Learning-to-rank Models; Knowledge Graph

*The first two authors provided equal contribution.

SIGIR '16, July 17 - 21, 2016, Pisa, Italy

ISBN 978-1-4503-4069-4/16/07...$15.00

DOI: http://dx.doi.org/10.1145/2911451.2911545

1. INTRODUCTION

Recent studies [16, 22] indicate that more than 75% of queries issued to Web search systems aim at finding information about entities, which could be material objects or concepts that exist in the real world or fiction (e.g. people, products, scientific papers, colors). Most common information needs underlying this type of queries include finding a certain entity (e.g. *"Einstein relativity theory"*), a particular attribute or property of an entity (e.g. *"Who founded Intel?"*) or a list of entities satisfying a certain criteria (e.g. *"Formula 1 drivers that won the Monaco Grand Prix"*). Such information needs can be efficiently addressed by presenting the target entity or a list of entities, either directly as search results or in addition to the ranked list of documents. This scenario gives rise to the problem of ad-hoc entity retrieval from knowledge graph, when the output of retrieval models is a list of entities given their (potentially verbose) textual description.

Recent successes in the development of Web information extraction methods have resulted in the emergence of a number of large-scale publicly available and proprietary knowledge repositories, such as DBpedia[1], Freebase[2], Google's Knowledge Graph and Microsoft's Satori. All these repositories adopt a simple knowledge representation model based on subject-predicate-object triples that can be conceptualized as a directed labeled multi-graph (commonly referred to as a knowledge graph), in which the nodes correspond to entities and the edges denote typed relations between entities. This model makes knowledge graphs a natural choice for addressing different types of entity-centric information needs. However, since the structure of knowledge graphs has been optimized for automated reasoning and answering structured graph pattern queries, finding entities in knowledge graphs that conceptually match unstructured free-text queries presents certain challenges to existing retrieval models.

First, since entities in knowledge graphs are designated only by an identifier (e.g. machine ID */m/0jcx*, in the case of Freebase, or URI *http://dbpedia.org/page/Albert_Einstein*,

[1] http://dbpedia.org
[2] http://freebase.com

in the case of DBpedia), there is no notion of document in this retrieval scenario. A simple workaround for this issue used in recent work is to aggregate the objects and predicates in all the triples, which include each distinct entity as a subject, into the fields of a structured document for that entity. Such aggregation can be done by predicate type [21], according to the importance weights manually assigned to predicates [6], based on a flat structure with fixed number of fields (e.g. title and content [19] or entity name, attributes, categories, similar and related entity names [33]) or a two-level hierarchy, in which the first level corresponds to predicate types, while the second level corresponds to individual predicates [18]. Accurately matching unstructured keyword queries with such structured entity representations, however, is another fundamental theoretical problem, which has received much less research attention.

In particular, existing retrieval models are based on a set of rigid assumptions, which limit their effectiveness for retrieval of structured entity documents. On one hand, retrieval models incorporating term dependencies, such as Sequential Dependence Model [17] (which allows to assign different importance weights to matching query concepts of different type), Weighted Sequential Dependence Model (WSDM) [4] (which estimates the importance of each matching query concept individually) and retrieval model based on copulas [10] disregard entity document structure by considering the matches of the same query concept in different fields of entity documents as equally important. On the other hand, although existing models for ad-hoc structured document retrieval, such as the Mixture of Language Models (MLM) [20] and BM25F [23], factor in document structure when determining the degree of relevance of queries to documents, they do not take into account term dependencies (i.e. are agnostic to bigram query concepts). Furthermore, these models calculate the retrieval score of a structured document as a sum of the matching scores of each query term in a linear combination of the language models (LMs) for each document field. As a result, the field weights in this linear combination, which effectively determine the projection of query terms onto document fields, are the same for all query terms. Although the recently proposed Fielded Sequential Dependence Model (FSDM) [33] partially addressed this limitation by allowing different importance weights to the matches of a query concept in different fields of structured entity documents, such parametrization still lacks flexibility, as those weights are the same for all query concepts of the same type (unigrams, ordered and unordered bigrams). This can create a problem, which is illustrated by an example query *"capitals in Europe which were host cities of summer Olympic games"*. First, contrary to the assumption of FSDM, query concepts of the same type in this query should be projected onto different fields of relevant entity documents (e.g. *"capitals"* should be mapped onto the *categories* field, while *"Europe"* should be mapped onto the *attributes* field). Incorrect projection of any query concept (e.g. mapping *"Europe"* and *"Olympic games"* to the *entity names* field) is likely to substantially degrade the accuracy of retrieval results for this query. Probabilistic Retrieval Model for Semistructured Data (PRMS) [14] is a unigram bag-of-words model for ad-hoc structured document retrieval that learns a simple statistical relationship between the intended mapping of terms in free-text queries and their frequency in different document fields. Robust estimation of this relationship, however, requires query terms to have a non-uniform distribution across document fields and is negatively affected by data sparsity, particularly in the case of entity representations with a large number of fields. As a result, PRMS has inferior performance on entity retrieval tasks not only relative to FSDM [33], but also to both MLM and BM25F [2, 33].

To overcome the limitations of existing retrieval models, we propose the Parametrized Fielded Sequential Dependence Model (PFSDM) and the Parametrized Fielded Full Dependence Model (PFFDM), two novel feature-based retrieval models, which *infer the user's intent behind each individual query concept (unigram or bigram) by dynamically estimating its probabilistic mapping onto the fields of structured entity representations* using a small number of statistical and linguistic features. We also provide a learning-to-rank algorithm to learn the weights of these features that maximize the target retrieval metric from the training data. The key contributions of this work are as follows:

1. We proposed PFSDM and PFFDM, two *novel feature-based models for structured document retrieval*, which account for sequential and full term dependencies as well as provide flexible parametrization allowing to dynamically project each query concept onto document fields. To the best of our knowledge, there are no previous studies of feature-based models for structured document retrieval, in general, and ad-hoc entity retrieval from knowledge graph, in particular;
2. We propose a set of statistical and linguistic features of query concepts that enable their accurate projection onto the fields of structured entity documents;
3. We experimentally demonstrate that, for the task of ad-hoc entity retrieval from knowledge graph, dynamic projection of query concepts onto entity representations is more effective than dynamic estimation of their importance. We also found out that retrieval models accounting for all dependencies between query terms provide more accurate retrieval results for this task than the models that account only for sequential dependencies.

The rest of this paper is organized as follows. Section 2 provides a brief overview of previous research along the directions relevant to the present work. Ranking functions of PFSDM and PFFDM, features to estimate the mapping of query concepts onto the fields of entity documents along with the method to learn the weights of those features are presented in Section 3. Experimental results are reported and analyzed in Section 4, while Section 5 concludes the paper and outlines the directions for future work.

2. RELATED WORK

In this section, we provide an overview of the recent work in ad-hoc entity retrieval from knowledge graph as well as term dependence and feature-based retrieval models, the three research directions most closely related to this work.

Ad-hoc Entity Retrieval from Knowledge Graph. Every information access task involving knowledge graphs requires finding entities matching a keyword query, either as an intermediate step or a final goal. Entity retrieval *methods* are typically designed to address one particular entity-centric information need, such as entity search [29, 32, 33], list search [7] or entity-based question answering [26, 31], and consist of two stages. Entities retrieved in the first stage of those methods using a standard retrieval model, such as BM25 [2, 29], BM25F [6, 11, 29] or MLM [18, 32], are re-ranked or expanded in the second stage with their immediate neighbors in the knowledge graph, which can be obtained using SPARQL queries [29] or through random walk [24].

In particular, Tonon et al. [29] proposed a hybrid method combining IR and structured graph search that starts by retrieving an initial set of entities using BM25 retrieval model and then extends it using SPARQL queries with the entities that are directly related to the ones in the initial search results. Zhiltsov and Agichtein [32] proposed a learning-to-rank method for re-ranking the results of MLM using query term and structural entity similarity features calculated in latent space. The SemSets method [7] proposed for entity list search utilizes the relevance of entities to automatically constructed categories (i.e. semantic sets) measured according to structural and textual similarity. This approach combines a basic TF-IDF retrieval model with spreading activation over the link structure of a knowledge graph and evaluation of membership in semantic sets. Sawant and Chakrabarti [25] proposed a learning-to-rank method for handling Web queries aimed at finding entities that belong to a particular category. Several approaches that translate free-text questions into structured SPARQL queries have been proposed for question answering over linked data [26, 31]. The proposed *retrieval models* can be leveraged in the first stage of the above *methods* to improve their overall performance. Models designed specifically for ad-hoc entity retrieval can also be leveraged to obtain a set of entities related to a keyword query (a process known as entity linking [12]), which is an important step for the methods utilizing knowledge bases for query expansion in ad-hoc document retrieval [9, 15, 30].

Several entity representation schemes, in which the objects from RDF triples involving an entity are grouped into the fields of a structured entity document based on the predicates in those triples, have also been proposed [6, 18, 19]. In [6], objects are grouped into three fields based on manually designated predicate type (important, neutral, and unimportant). A simple scheme, in which the entities are represented as documents with two fields (title and content) was proposed in [19]. Experimental comparison of a structured entity representation scheme based on four fields with an unstructured and more complicated hierarchical scheme indicated superior performance of a simple structured representation [18].

Term Dependence and Feature-based Retrieval Models. Markov Random Fields (MRF) based retrieval framework [17], proposed by Metzler and Croft, provided a theoretical foundation for incorporating term dependencies (in the form of query bigrams and unordered two-word phrases) into retrieval models. Sequential Dependence Model (SDM), which only considers two-word sequences of query terms, and Full Dependence Model (FDM), which considers all possible two-word combinations of query terms, are the two most popular variants of the MRF retrieval models for ad-hoc document retrieval. Subsequent work along this direction [1, 8] demonstrated strong positive effect of accounting for query term dependencies and proximity on both ad-hoc and Web document retrieval.

SDM was later extended into WSDM by Bendersky et al. [4]. WSDM estimates the relative importance of query concepts as a linear combination of statistical features based on the frequency of occurrence of these concepts in the collection and external resources. Superior retrieval performance of different variants of WSDM for ad-hoc document retrieval has been demonstrated through extensive experimental evaluation in [13]. The utility of linguistic analysis for accurate processing of verbose queries in ad-hoc document retrieval has been demonstrated in [3]. While feature-based retrieval models have been shown to be effective for weighting concepts in verbose queries [4, 5], this work examines their effectiveness for ad-hoc entity retrieval from knowledge graph.

3. APPROACH

In this section, we introduce the ranking functions of Parameterized Fielded Sequential Dependence (PFSDM) and the Parameterized Fielded Full Dependence (PFFDM) retrieval models, the features used by these models to determine the projection of query concepts onto the fields of entity documents along with an algorithm to learn the weights of those features that maximize the target retrieval metric.

3.1 PFSDM and PFFDM

The quality of retrieval results for entity-centric free-text queries depends on the correctness of inference of implicit query structure and the accuracy of matching the intent behind each query concept with different aspects of semantics of relevant entities encoded in their structured representations. However, the ambiguity of natural language can lead to many plausible interpretations of a keyword query, which combined with the requirement to accurately project those interpretations onto entity representations, makes entity retrieval from knowledge graph a challenging IR problem.

PFSDM is a parametric extension of FSDM [33], a recently proposed MRF-based entity retrieval model, which takes into account *both term dependencies and document structure*. FSDM uses the following function to score each entity profile E with respect to a given keyword query Q:

$$\begin{aligned} P_\Lambda(E|Q) \stackrel{rank}{=} & \lambda_T \sum_{q\in Q} \tilde{f}_T(q_i, E) + \\ & \lambda_O \sum_{q\in Q} \tilde{f}_O(q_i, q_{i+1}, E) + \\ & \lambda_U \sum_{q\in Q} \tilde{f}_U(q_i, q_{i+1}, E) \end{aligned} \quad (1)$$

where $\tilde{f}_T(q_i, E)$, $\tilde{f}_O(q_i, q_{i+1}, E)$, $\tilde{f}_U(q_i, q_{i+1}, E)$ are the potential functions and λ_T, λ_O, λ_U are the relative importance weights for unigram, ordered and unordered bigram query concepts, respectively. The potential function for unigrams is defined as:

$$\tilde{f}_T(q_i, E) = \log \sum_{j=1}^{F} w_j^T P(q_i|\theta_j^E) = \log \sum_{j=1}^{F} w_j^T \frac{tf_{q_i,E_j} + \mu_j \frac{cf_{q_i,j}}{|C_j|}}{|E_j| + \mu_j}$$

where F is the number of fields in structured entity documents, θ_j^E is the language model of field j in structured document for entity E smoothed using the field-specific Dirichlet prior μ_j; $|E_j|$ is the length of field j in E and w_j^T are the field weights for unigrams with the following constraints: $\sum_{j=1}^{F} w_j^T = 1, w_j^T \geq 0$; tf_{q_i,E_j} is the frequency of query unigram q_i in field j of E; $cf_{q_i,j}$ is the frequency of q_i in the field j across structured documents for all entities in the collection; $|C_j|$ is the total number of terms in field j across all entity documents in the collection. The potential function for ordered bigrams is defined as:

$$\tilde{f}_O(q_{i,i+1}, E) = \log \sum_{j=1}^{F} w_j^O \frac{tf_{\#1(q_{i,i+1}),E_j} + \mu_j \frac{cf_{\#1(q_{i,i+1}),j}}{|C_j|}}{|E_j| + \mu_j}$$

while the potential function for unordered bigrams is defined as:

$$\tilde{f}_U(q_{i,i+1}, E) = \log \sum_{j=1}^{F} w_j^U \frac{tf_{\#uw_n(q_{i,i+1}),E_j} + \mu_j \frac{cf_{\#uw_n(q_{i,i+1}),j}}{|C_j|}}{|E_j| + \mu_j}$$

where $tf_{\#1(q_{i,i+1}),E_j}$ is the frequency of query bigram $q_i q_{i+1}$ in field j of structured document for entity E, $cf_{\#1(q_{i,i+1}),j}$ is the collection frequency of $q_i q_{i+1}$ in field j and $tf_{\#uw_n(q_{i,i+1}),E^j}$ is the number of times the terms q_i and q_{i+1} co-occur within a window of n words in field j of E, regardless of their order.

In the case of entity descriptions with F fields, FSDM has $3 * F + 3$ parameters: F field mapping weights plus λ_T, λ_O and λ_U. We believe that this parametrization lacks the necessary degrees of freedom, which can potentially limit the accuracy of this retrieval model. We propose to address this issue by dynamically estimating $w^T_{q_i,j}$, the relative contribution of each individual query unigram q_i, and $w^{O,U}_{q_{i,i+1},j}$, the relative contribution of each individual query bigram $q_{i,i+1}$, that are matched in field j of structured entity document for E towards the retrieval score of this entity, as a weighted linear combination of features:

$$w^T_{q_i,j} = \sum_k \alpha^U_{j,k} \phi_k(q_i, j)$$

$$w^{O,U}_{q_{i,i+1},j} = \sum_k \alpha^B_{j,k} \phi_k(q_{i,i+1}, j)$$

under the following set of constraints:

$$\sum_j w^T_{q_i,j} = 1, w^T_{q_i,j} \geq 0, \alpha^U_{j,k} \geq 0, 0 \leq \phi_k(q_i, j) \leq 1$$

$$\sum_j w^{O,U}_{q_{i,i+1},j} = 1, w^{O,U}_{q_{i,i+1},j} \geq 0, \alpha^B_{j,k} \geq 0, 0 \leq \phi_k(q_{i,i+1}, j) \leq 1$$

where $\phi_k(q_i, j)$ and $\phi_k(q_{i,i+1}, j)$ are the values of the k-th non-negative feature function for query unigram q_i and bigram $q_{i,i+1}$ in field j of entity document, respectively. $w^T_{q_i,j}$ and $w^{O,U}_{q_{i,i+1},j}$, which can also be considered as posteriors $p(E_j|q_i)$ and $p(E_j|q_{i,i+1})$, provide probabilistic projection of query unigram q_i and bigram $q_{i,i+1}$ onto the fields of structured entity representations (to reduce the number of parameters in the model, we set $w^O_{q_{i,i+1},j} = w^U_{q_{i,i+1},j}$). PFSDM determines this projection based on multiple features, unlike PRMS [14], which estimates it directly from the data based only on the total number of occurrences of a query term in a particular field across all documents in a collection. Feature-based estimation of this projection increases its robustness by overcoming the issues of sparsity and uniform distribution of occurrences of a query concept across the fields of entity documents.

PFFDM is different from PFSDM in that it accounts for all dependencies between the query terms, rather than only sequential ones.

3.2 Features

The features of different type that we propose to estimate the projection of a query concept κ onto field j of structured entity representations are presented in Table 1. In particular, PFSDM and PFFDM use two types of features: the ones whose value depends on a query concept and a field of entity representation and the ones that depend only on a query concept itself. The intuition behind also having the latter type of features is that the relation between them and the fields will be learned in the process of estimating their weights. For example, one can expect that the weight of a feature indicating whether a query concept is plural non-proper noun (NNS) will be higher in the *categories* field than in all other fields. For the features that depend on a field, one can expect that the value of the feature in that field will indicate the likelihood of a concept to be mapped to it. Nevertheless, we still learn the weights for these features, since: (1) ranges of values for particular features can be different in different fields and optimizing their weights is one of the ways to perform adequate scaling (2) contribution of the feature to the relevance of a field can be different for different fields.

Two of the features (FP, TS) depend on the collection statistics of a particular field. During optimization and retrieval, these two real-valued features (FP, TS) were rescaled to $[0, 1]$ range. The FP feature was rescaled logarithmically.

The other group of field mapping features (NNP, NNS, JJS, NPP, NNO) are binary and take particular values based on the output of Standford POS Tagger or Parser. NNP takes positive values for the query concepts that are proper nouns (e.g. entity names) and, thus, should be mapped to the *names*, *similar entity names* and *related entity names* fields. The NNS, NPP, and NNO features take positive values for the concepts that designate a broader class or type of the desired entities and, therefore, should be mapped to the *categories* field, while the JJS feature should project superlative adjectives to the *attributes* field. Constant feature (INT), which has the same value for all concepts, is known to be useful for mapping bigrams concepts.

3.3 Parameter estimation

In total, PFSDM and PFFDM have $F * U + F * B + 3$ parameters ($F * U + F * B$ feature weights as well as λ_T, λ_O and λ_U), where F is the number of fields, while U and B are the number of field mapping features for unigrams and bigrams, respectively. An efficient two-stage block optimization algorithm for learning the parameters of PFSDM and PFFDM with respect to the target retrieval metric is presented in Algorithm 1.

Algorithm 1 An algorithm for learning the feature weights in PFSDM and PFFDM.

1: $Q \leftarrow$ Train queries
2: $e_U = \{1, 0, 0\}, e_B = \{0, 1, 1\}$
3: **for** $s \in \{U, B\}$ **do**
4: $\quad \boldsymbol{\lambda} = \boldsymbol{e}_s$
5: $\quad \hat{\boldsymbol{\alpha}}_s \leftarrow CA(Q, \boldsymbol{\lambda})$
6: **end for**
7: $\hat{\boldsymbol{\lambda}} \leftarrow CA(Q, \hat{\boldsymbol{\alpha}}_U, \hat{\boldsymbol{\alpha}}_B)$

In the first stage (lines 3-6), the algorithm optimizes field mapping feature weights α separately for unigrams and bigrams. During optimization of the feature weights for un-

Table 1: **Features to estimate the contribution of query concept κ matched in field j towards the retrieval score of E. Column CT designates the type of query concept that a feature is used for (UG stands for unigrams, BG stands for bigrams).**

Source	Feature	Description	CT
Collection statistics	$FP(\kappa, j)$	Posterior probability $P(E_j\|w)$ obtained through Bayesian inversion of $P(w\|E_j)$, as defined in [14].	UG BG
	$TS(\kappa, j)$	Retrieval score of the top document according to SDM [17], when concept κ is used as a query and only the jth fields of entity representations are used as documents.	BG
Stanford POS Tagger[3]	$NNP(\kappa)$	Is concept κ a proper noun (singular or plural)?	UG
	$NNS(\kappa)$	Is concept κ a plural non-proper noun? We consider a bigram as plural when at least one of its terms is plural.	UG BG
	$JJS(\kappa)$	Is concept κ a superlative adjective?	UG
Stanford Parser[4]	$NPP(\kappa)$	Is concept κ part of a noun phrase?	BG
	$NNO(\kappa)$	Is concept κ the only singular non-proper noun in a noun phrase?	UG
	INT	Intercept feature, which has value 1 for all concepts.	UG BG

igrams, the feature weights for bigrams are not considered and vice versa. This is achieved by setting the corresponding λ weights to 0. After the algorithm is finished with optimizing the α weights, it proceeds to optimize the weights of MRF potential functions for different query concept types (λ_T, λ_O and λ_U in Eq. 1).

4. EXPERIMENTS

4.1 Experimental setup

Experimental results reported in this work were obtained on a publicly available benchmark developed by Balog and Neumayer [2], which uses DBpedia as the knowledge graph. For fair comparison, we used the same five field entity representation scheme and the same query sets as in [33] (SemSearch ES consisting primarily of named entity queries, ListSearch consisting primarily of entity list search queries, QALD-2 consisting of entity-focused natural language questions, and INEX-LD containing a mix of entity-centric queries of different type). We pre-processed both entity documents and queries by applying the Krovetz stemmer and removing the stopwords in the INQUERY stopword list.

4.2 Feature analysis

First, to evaluate the effectiveness of the proposed features, we performed an exploratory analysis of distributions of their values (for the features whose values depend on document fields) or frequencies of their occurrences (for the features whose values are independent of document fields) in different fields of entity representations. Specifically, we manually annotated each concept in all queries according to the user's intent with respect to a particular aspect of target entities as an *attribute* concept, *entity* concept, *relation* concept, or *type* concept. Our intuition is that the *attribute* query concepts (e.g. when a user is searching for an entity attribute rather than an entity itself) should be frequently occurring or have relatively higher feature values in the *attributes* field of entity representations. Query terms or phrases marked as *entity* concepts (e.g. when a user is searching for a particular named entity) should be primarily occurring in the *names* and *similar entity names* fields, while the *relation* concepts (when a query is about a relation between the two named entities) should be primarily occurring in the *similar entity names* and *related entity names* fields of entity representations. Finally, query concepts marked as *type* (when a query is about several entities with the same type) should be frequently occurring or have relatively greater feature values in the *categories* field.

Figure 1 visualizes the distributions of values of the Field Probability (FP) and Top Score (TS) features for the query concepts of different types in different fields of structured entity representations. As can be observed in Figure 1 (left), the median values of the FP feature for the query concepts of type *entity* in all three *entity* fields (*entity names*, *similar entity names* and *related entity names*) are significantly greater than the median values of the same feature for the query concepts of the same type in both the *attributes* and *categories* fields. The median values of the FP feature for the query concepts of type *attribute* in the *attributes* and *categories* fields are significantly greater than the median values of the same feature for the query concepts of the same type in all three *entity* fields. Furthermore, for the *type* and *relation* query concepts, the median values of the same feature in the *categories* and *related entity names* fields, respectively, are significantly greater than the median values of this feature in all other fields. It can also be observed in Figure 1 (right), that the median values of the TS feature for the query concepts of type *entity* in the *similar entity names* and *related entity names* fields are significantly greater than the median values of the same feature for the query concepts of the same type in all other fields. Furthermore, the median values of the TS feature for the *type* and *attribute* query concepts in the *attributes* field are greater than the median values of the same feature in all other fields.

To formally validate these observations, we conducted statistical significance tests. In particular, the Kruskal-Wallis

[3] http://nlp.stanford.edu/software/tagger.html
[4] http://nlp.stanford.edu/software/lex-parser.html

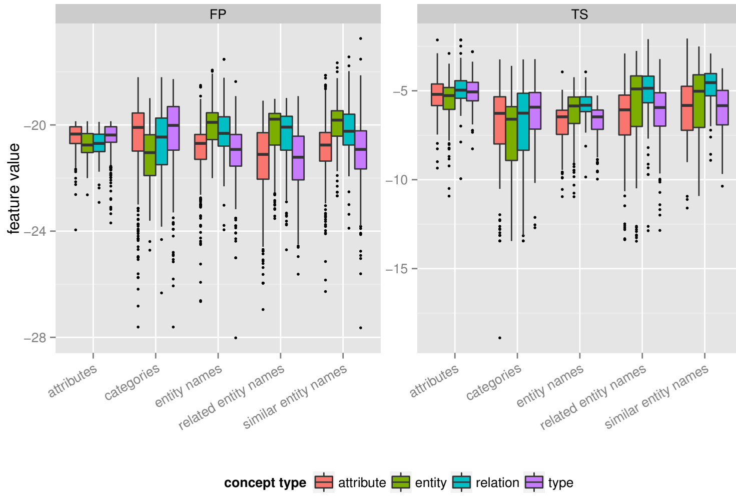


Figure 1: Boxplots for the distributions of values of real-valued features for the query concepts of different types in different fields of entity documents.

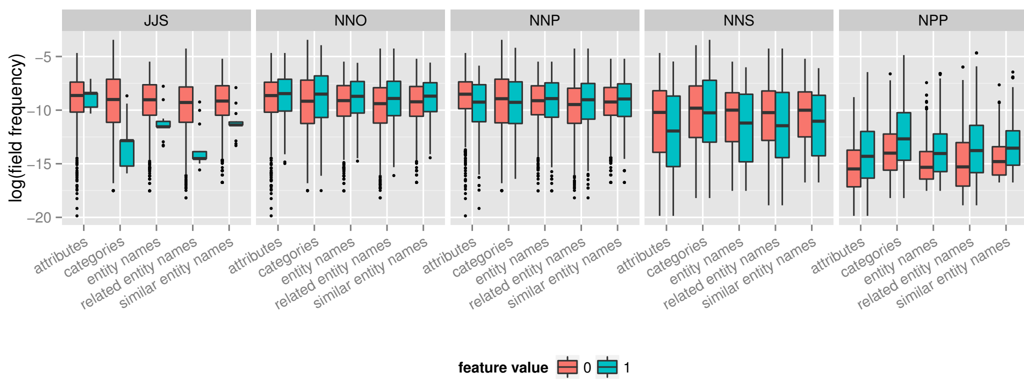


Figure 2: Boxplots for the distributions of normalized frequencies of query concepts with negative and positive feature values in different fields of entity documents.

test [27] indicated that the null hypothesis of the FP and TS feature values having the same median in different fields of entity representations for all concept types should be rejected ($p < 0.05$). Following the Kruskal-Wallis test, we performed a multiple comparison test (kruskalmc), which besides confirming the above empirical observations, indicated other statistically significant differences in feature values. In particular, the values of the TS feature for the query concepts of type *relation* in the *related entity names* field are different from its values in the *categories* and all three *entity* fields. This test also indicated that for all concept types the values of the FP feature in all three *entity* fields (*entity names*, *similar entity names* and *related entity names*) are significantly different from the values of the same feature in both *attributes* and *categories* fields, which in turn are significantly different from each other for the query concepts of type *relation*.

Figure 2 visualizes the distributions of normalized frequencies of query concepts with negative and positive values of the features that do not depend on a field in different fields of structured entity representations. Examining the properties of these distributions for concepts with positive and negative feature values in the same field as well as across different fields can give us an intuition about whether the concepts having positive values for a particular feature are more likely to occur in certain fields of entity representations than in the others. This in turn can give us an insight about whether certain linguistic properties of query concepts are indicative of the user's intent with respect to the projection of those concepts onto specific aspects of relevant entities. As follows from Figure 2, the query concepts that are superlative adjectives (JJS is true) much more frequently occur in the *attributes* field and are very likely to designate the attributes of relevant entities; plural non-proper unigrams and bigrams (NNS is true), bigrams (NPP is true) or singular non-proper nouns (NNO is true) that are part of a noun phrase are more likely to represent the categories of relevant entities, while singular or plural proper nouns (NNP is true) more frequently occur in three *entity* fields, than in any other field of entity documents and, thus, typically

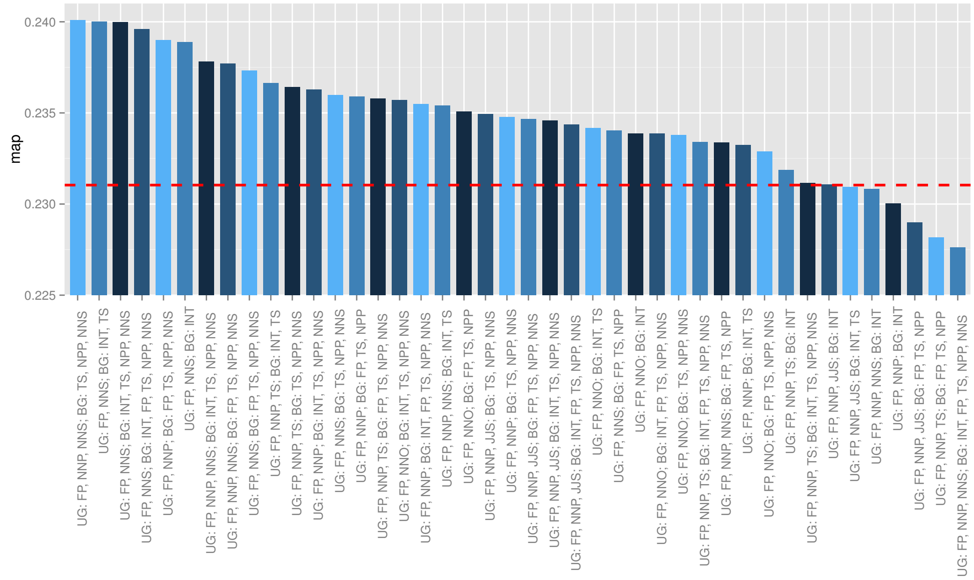


Figure 3: MAP of PFSDM on DBpedia knowledge graph and the benchmark in [2] depending on different combinations of field mapping features for unigram and bigram query concepts. Dashed red line represents the performance of FSDM [33].

designate the target entities directly. These observations indicate that entity-centric keyword queries have an implicit structure, with each element in that structure designating a particular aspect in multi-fielded representation of relevant entities.

4.3 Feature effectiveness

To determine the combination of features, which results in the best performance of PFSDM and PFFDM, we conducted a feature selection study. First, we determined the combinations of only unigram and only bigram features (using simplified versions of PFSDM that consider only unigrams or bigrams, respectively), which result in the best retrieval performance in terms of MAP (target retrieval metric). In particular, we identified 6 most effective unigram and 7 most effective bigram feature sets. Then we evaluated the performance of PFSDM using each possible combination of unigram and bigram feature sets to determine the best performing combined feature set. Retrieval effectiveness of different feature combinations for PFSDM is illustrated in Figure 3. As follows from Figure 3, most feature combinations result in higher MAP than FSDM. PFSDM achieves the highest MAP of 0.240 in conjunction with FP, NNS, NNP features for unigram query concepts and TS, NNS, NPP features for bigram ones. The weights of these features that result in the highest MAP of PFSDM are presented in Table 3.

From Table 3, it follows that the learned feature weights are similar to the distribution of frequencies of manually marked up query concept types across the fields of structured entity representation.

4.4 Comparison with baselines

Retrieval accuracy of PFSDM and PFFDM using the best feature combination is compared with that of state-of-the-art retrieval models for ad-hoc document (SDM [17] and WSDM [4] with *cf* and *df* features) and structured document (BM25F [23], PRMS [14], MLM [20] and FSDM [33]) retrieval in Table 2. Parameters of both the proposed retrieval models and the baselines have been optimized using 5-fold cross validation (except PRMS, which does not require optimization).

Several important conclusions can be made based on the results in Table 2. First, WSDM shows minor improvement and, on some query sets, is even worse than SDM, which indicates that feature-based query concept importance weighting is less effective for entity retrieval than for document retrieval. On the other hand, dynamic feature-based estimation of relative importance of matching query concepts in different fields of entity documents provides significant improvement of retrieval accuracy on verbose queries, such as the ones in QALD-2 query set. Second, the relatively small difference in performance between PFSDM and FSDM on SemSearch ES and ListSearch query sets can be explained by the fact that all concepts in those query sets map to only a few fields. In particular, most concepts in SemSearch ES queries map onto the *entity* field, while most concepts in ListSearch queries map onto the *categories* and *attributes* fields. In such cases, estimating field mapping degenerates to estimating relative importance of matching concepts in a particular field of entity representation, which nullifies the advantages of PFSDM. We can also observe that taking into account all dependencies between query terms can partially mitigate this problem, as evidenced by superior performance of PFFDM on both ListSearch and INEX-LD query sets. Third, although it can be seen that PFSDM achieves improvement over FSDM in terms of MAP, MRR and NDCG@5 at the expense of decreased P@10, taking into account all dependencies between the query terms allows PFFDM to achieve consistent improvement over both FSDM and FFDM in terms of all retrieval metrics.

Table 4 compares the retrieval accuracy of PFSDM and PFFDM with the same baselines on the knowledge graph

Table 2: Performance of retrieval models on DBpedia knowledge graph and the benchmark in [2]. Relative improvement over PRMS and FSDM is shown in parenthesis, while "*" and "†" indicate statistically significant improvement over the same baselines, according to the Fisher's randomization test ($\alpha = 0.05$) [28].

	SemSearch ES			
	MAP	**P@10**	**MRR**	**NDCG@5**
SDM	0.254 (+10.5%)	0.202 (+14.4%)	0.520 (−5.3%)	0.306 (−3.5%)
WSDM	0.246 (+7.2%)	0.201 (+13.5%)	0.507 (−7.7%)	0.298 (−5.9%)
BM25F-tc [2]	0.334 (+45.3%)	0.263 (+48.7%)	0.705 (+28.4%)	0.453 (+42.9%)
PRMS	0.230	0.177	0.549	0.317
MLM	0.320 (+39.3%)	0.250 (+41.3%)	0.680 (+23.9%)	0.423 (+33.3%)
FSDM	0.386 (+68.1%)	0.286 (+61.7%)	0.737 (+34.3%)	0.476 (+50.3%)
PFSDM	0.394^{*} (+71.4%/+1.9%)	0.286^{*} (+61.7%/0.0%)	0.757^{*} (+38.0%/+2.7%)	$0.494^{*}_{\dagger}$ (+55.9%/+3.7%)
FFDM	0.389^{*} (+69.3%/+0.7%)	0.286^{*} (+61.7%/0.0%)	0.734^{*} (+33.8%/−0.4%)	0.479^{*} (+51.3%/+0.6%)
PFFDM	0.380^{*} (+65.3%/−1.7%)	0.286^{*} (+61.7%/0.0%)	0.739^{*} (+34.6%/+0.2%)	0.477^{*} (+50.6%/+0.2%)
	ListSearch			
	MAP	**P@10**	**MRR**	**NDCG@5**
SDM	0.197 (+78.3%)	0.252 (+63.8%)	0.463 (+30.5%)	0.282 (+60.1%)
WSDM	0.194 (+75.4%)	0.257 (+66.7%)	0.457 (+28.8%)	0.280 (+58.7%)
BM25F-tc [2]	0.159 (+43.9%)	0.221 (+43.5%)	0.390 (+9.8%)	0.217 (+23.2%)
PRMS	0.111	0.154	0.355	0.176
MLM	0.190 (+71.7%)	0.252 (+63.8%)	0.439 (+23.5%)	0.280 (+58.5%)
FSDM	0.203 (+83.9%)	0.256 (+66.1%)	0.447 (+25.8%)	0.274 (+55.2%)
PFSDM	0.201^{*} (+81.8%/−1.1%)	0.253^{*} (+64.4%/−1.0%)	0.443^{*} (+24.8%/−0.8%)	0.278^{*} (+57.5%/+1.5%)
FFDM	$0.226^{*}_{\dagger}$ (+104.4%/+11.2%)	$0.282^{*}_{\dagger}$ (+83.1%/+10.2%)	$0.499^{*}_{\dagger}$ (+40.6%/+11.7%)	$0.313^{*}_{\dagger}$ (+77.2%/+14.2%)
PFFDM	$0.228^{*}_{\dagger}$ (+106.4%/+12.3%)	$0.286^{*}_{\dagger}$ (+85.9%/+11.9%)	0.487^{*} (+37.2%/+9.1%)	$0.302^{*}_{\dagger}$ (+71.3%/+10.4%)
	INEX-LD			
	MAP	**P@10**	**MRR**	**NDCG@5**
SDM	0.117 (+83.5%)	0.258 (+77.9%)	0.567 (+38.7%)	0.341 (+57.4%)
WSDM	0.118 (+85.3%)	0.257 (+77.2%)	0.549 (+34.4%)	0.341 (+57.4%)
BM25F-tc [2]	0.117 (+83.0%)	0.249 (+71.7%)	0.559 (+36.7%)	0.341 (+57.4%)
PRMS	0.064	0.145	0.409	0.216
MLM	0.102 (+60.2%)	0.238 (+64.1%)	0.530 (+29.7%)	0.306 (+41.3%)
FSDM	0.111 (+74.4%)	0.263 (+81.4%)	0.546 (+33.7%)	0.322 (+48.7%)
PFSDM	0.116^{*} (+81.7%/+4.2%)	0.259^{*} (+78.6%/−1.5%)	0.579^{*} (+41.5%/+5.9%)	0.341^{*} (+57.6%/+6.0%)
FFDM	$0.122^{*}_{\dagger}$ (+91.3%/+9.7%)	0.273^{*} (+88.3%/+3.8%)	0.560^{*} (+37.0%/+2.5%)	$0.345^{*}_{\dagger}$ (+59.5%/+7.3%)
PFFDM	$0.121^{*}_{\dagger}$ (+89.9%/+8.9%)	0.274^{*} (+89.0%/+4.2%)	0.556^{*} (+36.0%/+1.8%)	0.343^{*} (+58.7%/+6.7%)
	QALD-2			
	MAP	**P@10**	**MRR**	**NDCG@5**
SDM	0.184 (+52.9%)	0.106 (+35.5%)	0.287 (+52.0%)	0.215 (+46.5%)
WSDM	0.183 (+52.8%)	0.112 (+42.7%)	0.288 (+52.6%)	0.214 (+45.7%)
BM25F-tc [2]	0.107 (−11.1%)	0.062 (−20.9%)	0.158 (−16.0%)	0.117 (−20.6%)
PRMS	0.120	0.079	0.188	0.147
MLM	0.152 (+26.3%)	0.103 (+30.9%)	0.215 (+14.0%)	0.170 (+15.7%)
FSDM	0.195 (+62.7%)	0.136 (+73.6%)	0.283 (+50.0%)	0.229 (+55.7%)
PFSDM	$0.218^{*}_{\dagger}$ (+81.9%/+11.7%)	0.140^{*} (+78.2%/+2.6%)	0.308^{*} (+63.2%/+8.8%)	$0.253^{*}_{\dagger}$ (+72.5%/+10.8%)
FFDM	0.200^{*} (+66.5%/+2.3%)	0.139^{*} (+76.4%/+1.6%)	0.292^{*} (+54.9%/+3.3%)	0.237^{*} (+61.6%/+3.8%)
PFFDM	$0.219^{*}_{\dagger}$ (+82.1%/+11.9%)	0.147^{*} (+87.3%/+7.9%)	0.310^{*} (+64.2%/+9.5%)	$0.267^{*}_{\dagger}$ (+81.5%/+16.6%)
	All queries			
	MAP	**P@10**	**MRR**	**NDCG@5**
SDM	0.192 (+41.5%)	0.198 (+45.0%)	0.449 (+21.3%)	0.281 (+31.5%)
WSDM	0.189 (+39.6%)	0.200 (+46.5%)	0.441 (+19.1%)	0.278 (+30.2%)
BM25F-tc [2]	0.182 (+34.3%)	0.192 (+40.8%)	0.442 (+19.5%)	0.277 (+29.5%)
PRMS	0.136	0.136	0.370	0.214
MLM	0.196 (+44.3%)	0.206 (+50.6%)	0.458 (+23.7%)	0.292 (+36.4%)
FSDM	0.231 (+70.4%)	0.231 (+69.2%)	0.498 (+34.5%)	0.325 (+52.0%)
PFSDM	$0.240^{*}_{\dagger}$ (+77.1%/+3.9%)	0.231^{*} (+68.9%/−0.2%)	$0.516^{*}_{\dagger}$ (+39.5%/+3.7%)	$0.342^{*}_{\dagger}$ (+59.9%/+5.2%)
FFDM	$0.241^{*}_{\dagger}$ (+77.5%/+4.2%)	$0.240^{*}_{\dagger}$ (+75.7%/+3.8%)	$0.515^{*}_{\dagger}$ (+39.2%/+3.4%)	$0.342^{*}_{\dagger}$ (+60.1%/+5.3%)
PFFDM	$0.244^{*}_{\dagger}$ (+79.9%/+5.6%)	$0.244^{*}_{\dagger}$ (+78.4%/+5.4%)	$0.518^{*}_{\dagger}$ (+39.9%/+4.0%)	$0.347^{*}_{\dagger}$ (+62.5%/+6.9%)

from the 2009 Billion Triple Challenge (BTC-2009). This knowledge graph consists of 1.14 billion RDF triples and contains entities from other knowledge bases besides DBpedia. For this experiment, we used the queries from the Entity Search (ES) track of 2010[5] and 2011[6] Yahoo! Sem-

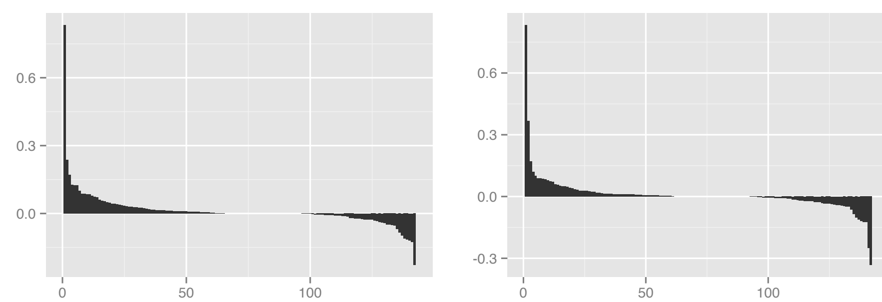
[5]http://km.aifb.kit.edu/ws/semsearch10/
[6]http://km.aifb.kit.edu/ws/semsearch11/

Table 3: Optimized weights of the best performing features for PFSDM (averaged over all folds).

concept type	feature	attributes	categories	names	related entity names	similar entity names
Unigram	FP	0.147	0.109	0.026	0.020	0.041
	NNS	0.110	0.141	0.019	0.023	0.014
	NNP	0.116	0.092	0.025	0.060	0.057
Bigram	TS	0.065	0.153	0.029	0.043	0.087
	NNS	0.039	0.183	0.028	0.046	0.057
	NPP	0.091	0.073	0.000	0.075	0.042

Table 4: Comparison of retrieval models on SemSearch ES queries and BTC-2009 knowledge graph.

	MAP	P@10	MRR	NDCG@5
SDM	0.102 (+4.4%)	0.210 (+6.0%)	0.518 (−4.9%)	0.248 (−8.0%)
WSDM	0.100 (+2.6%)	0.214 (+8.2%)	0.495 (−9.1%)	0.230 (−14.7%)
PRMS	0.098	0.198	0.545	0.269
MLM	0.121 (+23.6%)	0.243 (+22.8%)	0.588 (+8.0%)	0.312 (+16.0%)
FSDM	0.171 (+75.3%)	0.323 (+63.3%)	0.631 (+15.8%)	0.358 (+32.9%)
PFSDM	$0.182^{*}_{\dagger}$ (+87.0%/+6.7%)	0.335^{*} (+69.4%/+3.7%)	$0.657^{*}_{\dagger}$ (+20.7%/+4.2%)	0.371^{*} (+37.8%/+3.7%)
FFDM	$0.180^{*}_{\dagger}$ (+84.8%/+5.4%)	$0.330^{*}_{\dagger}$ (+66.9%/+2.2%)	0.647^{*} (+18.8%/+2.6%)	$0.373^{*}_{\dagger}$ (+38.6%/+4.3%)
PFFDM	0.187^{*} (+91.8%/+9.4%)	$0.342^{*}_{\dagger}$ (+72.6%/+5.7%)	0.650^{*} (+19.4%/+3.1%)	0.377^{*} (+40.2%/+5.5%)



(a) **(b)**

Figure 4: Topic-level differences in average precision on BTC-2009: a) between PFSDM and FSDM; b) between PFFDM and FFDM.

Search Challenge and publicly available relevance judgments for those queries[7]. We used the same 5-field entity representation scheme for this knowledge graph, as we did for the DBpedia one. As can be seen in Table 4, PFSDM and PFFDM demonstrate significant and consistent improvement relative to PRMS, as well as FSDM and FFDM, respectively.

From Figures 4a and 4b, it also follows that parameterizing the field importance weights in PFSDM and PFFDM results in more improved topics and greater magnitude of improvements than using static weights in FSDM and FFDM.

4.5 Success/Failure Analysis

Next, we provide a brief qualitative analysis of sample queries illustrating the strengths and weaknesses of PFSDM and PFFDM. The ability to map query concepts of the same type onto different fields of entity documents allows PFSDM to promote the correct entities for verbose and question queries. For example, for the query *"Who produces Orangina?"*, PFSDM returns the correct result *A.G. Barr* at the top, unlike FSDM, which ranks it as the 18th result. For this query, PFSDM correctly assigns higher weights to the matches of the query term *produce* in the *attributes* (0.49) and *categories* (0.41) fields, of the query term *Orangina* in the *related entity names* field (0.55) and of the bigram *produce Orangina* in the *categories* (0.45) field, unlike FSDM, which uses the same field weighting scheme (0.40 for *attributes*, 0.20 for *categories*, 0.30 for *related entity names* and 0.10 for *similar entity names* fields) for all query unigrams. The correct field mapping weights for these query concepts are determined by the FP and NNP features. The same effect was observed for the query *"Who is the governor of Texas?"*. PFSDM promoted *Rick Perry*, the only correct answer for this query, from the second to the first position by boosting the matches of query concepts *governor* (captured by the FP and NNP features) and *governor Texas* (captured by the TS feature) in the *categories* field.

Another type of queries with the highest relative MAP

[7] https://github.com/nzhiltsov/YSC-relevance-data

gain of PFSDM over FSDM are list search queries, such as *"Give me a list of all American inventions"* (from 0.032 to 0.232), *"Tom Hanks movies where he plays a leading role"* (from 0.073 to 0.181) and *"Give me all companies in Munich"* (from 0.114 to 0.252). For the first query, PFSDM promotes the correct entities *Aberdeen Chronograph*, *Lisp programming language* by boosting their matching scores in the *categories* field, while FSDM ranks *The Heroic Age of American Invention*, a science book for children, as the highest entity, by not taking into account the absence of an important term *invention* in its *categories* field.

We also observed that the common causes of PFSDM failures are assignment of uniform field weights to query concepts and a lack of concept statistics. For example, for the query *"Give me all people that were born in Vienna and died in Berlin"*, PFSDM underestimates the importance of relatively rare concepts *Vienna* and *Berlin*, but overestimates the importance of very popular concepts *born* and *die*. The issue can be addressed by using a minimum support matching strategy or by introducing additional features.

5. CONCLUSION

In this paper, we proposed two novel models for ad-hoc entity retrieval from knowledge graph, which account for term dependencies and perform feature-based projection of query concepts onto the fields of entity documents. By demonstrating the possibility of inferring implicit structure of keyword queries using linguistic attributes and simple field statistics of query concepts, the proposed models constitute an important step in the evolution of models for structured document retrieval. We hypothesize that the proposed models can be effective in other structured information retrieval scenarios, such as product and social graph search, and leave verification of this hypothesis to future work.

Acknowledgments

This work was partially supported by the subsidy from the government of the Russian Federation to support the program of competitive growth of Kazan Federal University and by the Russian Foundation for Basic Research (grants # 15-07-08522, 15-47-02472).

6. REFERENCES

[1] J. Bai, Y. Chang, H. Cui, Z. Zheng, G. Sun, and X. Li. Investigation of Partial Query Proximity in Web Search. In *Proceedings of the 17th WWW*, pages 1183–1184, 2008.

[2] K. Balog and R. Neumayer. A Test Collection for Entity Search in DBpedia. In *Proceedings of the 36th ACM SIGIR*, pages 737–740, 2013.

[3] M. Bendersky and W. B. Croft. Discovering Key Concepts in Verbose Queries. In *Proceedings of the 31st ACM SIGIR*, pages 491–498, 2008.

[4] M. Bendersky, D. Metzler, and W. B. Croft. Learning Concept Importance Using a Weighted Dependence Model. In *Proceedings of the 3rd ACM WSDM*, pages 31–40, 2010.

[5] M. Bendersky, D. Metzler, and W. B. Croft. Parameterized Concept Weighting in Verbose Queries. In *Proceedings of the 34th ACM SIGIR*, pages 605–614, 2011.

[6] R. Blanco, P. Mika, and S. Vigna. Effective and Efficient Entity Search in RDF Data. In *Proceedings of the 10th ISWC*, pages 83–97, 2011.

[7] M. Ciglan, K. Nørvåg, and L. Hluchý. The SemSets Model for Ad-hoc Semantic List Search. In *Proceedings of the 21st WWW*, pages 131–140, 2012.

[8] R. Cummins and C. O'Riordan. Learning in a Pairwise Term-term Proximity Framework for Information Retrieval. In *Proceedings of the 32nd ACM SIGIR*, pages 251–258, 2009.

[9] J. Dalton, L. Dietz, and J. Allan. Entity Query Feature Expansion Using Knowledge Base Links. In *Proceedings of the 37th ACM SIGIR*, pages 365–374, 2014.

[10] C. Eickhoff, A. P. de Vries, and T. Hoffman. Modelling Term Dependencies with Copulas. In *Proceedings of the 38th ACM SIGIR*, pages 783–786, 2015.

[11] B. Fetahu, U. Gadiraju, and S. Dietze. Improving Entity Retrieval on Structured Data. In *Proceedings of the 14th ISWC*, pages 474–491, 2015.

[12] F. Hasibi, K. Balog, and S. E. Bratsberg. Entity Linking in Queries: Tasks and Evaluation. In *Proceedings of the 1st ACM ICTIR*, pages 171–180, 2015.

[13] S. Huston and W. B. Croft. A Comparison of Retrieval Models using Term Dependencies. In *Proceedings of the 23rd ACM CIKM*, pages 111–120, 2014.

[14] J. Y. Kim, X. Xue, and W. B. Croft. A Probabilistic Retrieval Model for Semistructured Data. In *Proceedings of the 31st ECIR*, pages 228–239, 2009.

[15] A. Kotov and C. Zhai. Tapping into Knowledge Base for Concept Feedback: Leveraging ConceptNet to Improve Search Results for Difficult Queries. In *Proceedings of the 5th WSDM*, pages 403–412, 2012.

[16] T. Lin, P. Pantel, M. Gamon, A. Kannan, and A. Fuxman. Active Objects Actions for Entity Centric Search. In *Proceedings of the 21st WWW*, pages 589–598, 2012.

[17] D. Metzler and W. B. Croft. A Markov Random Field Model for Term Dependencies. In *Proceedings of the 28th ACM SIGIR*, pages 472–479, 2005.

[18] R. Neumayer, K. Balog, and K. Nørvåg. On the Modeling of Entities for Ad-hoc Entity Search in the Web of Data. In *Proceedings of the 34th ECIR*, pages 133–145, 2012.

[19] R. Neumayer, K. Balog, and K. Nørvåg. When Simple is (more than) Good Enough: Effective Semantic Search with (almost) no Semantics. In *Proceedings of the 34th ECIR*, pages 540–543, 2012.

[20] P. Ogilvie and J. Callan. Combining Document Representations for Known-item Search. In *Proceedings of the 26th ACM SIGIR*, pages 143–150, 2003.

[21] J. R. Pérez-Aguera, J. Arroyo, J. Greenberg, J. P. Iglesias, and V. Fresno. Using BM25F for Semantic Search. In *Proceedings of the 3rd SemSearch Workshop*, 2010.

[22] J. Pound, P. Mika, and H. Zaragoza. Ad-hoc Object Retrieval in the Web of Data. In *Proceedings of the 19th WWW*, pages 771–780, 2010.

[23] S. Robertson, H. Zaragoza, and M. Taylor. Simple BM25 Extension to Multiple Weighted Fields. In *Proceedings of the 13th ACM CIKM*, pages 42–49, 2004.

[24] C. Rocha, D. Schwabe, and M. P. de Argão. A Hybrid Approach for Searching in the Semantic Web. In *Proceedings of the 13th WWW*, pages 374–383, 2004.

[25] U. Sawant and S. Chakrabarti. Learning Joint Query Interpretation and Response Ranking. In *Proceedings of the 22nd WWW*, pages 1099–1110, 2013.

[26] S. Shakarpour, A.-C. N. Ngomo, and S. Auer. Question Answering on Interlinked Data. In *Proceedings of the 22nd WWW*, pages 1145–1156, 2013.

[27] S. C. Siegel and J. Castellan. Nonparametric Statistics for the Behavioral Sciences. *McGraw-Hill*, 1998.

[28] M. D. Smucker, J. Allan, and B. Carterette. A Comparison of Statistical Significance Tests for Information Retrieval Evaluation. In *Proceedings of the 16th ACM CIKM*, pages 623–632, 2007.

[29] A. Tonon, G. Demartini, and P. Cudré-Mauroux. Combining Inverted Indices and Structured Search for Ad-hoc Object Retrieval. In *Proceedings of the 35th ACM SIGIR*, pages 125–134, 2012.

[30] C. Xiong and J. Callan. Query Expansion with Freebase. In *Proceedings of the 1st ACM ICTIR*, pages 111–120, 2015.

[31] M. Yahya, K. Berberich, S. Elbassuoni, and G. Weikum. Robust Question Answering over the Web of Linked Data. In *Proceedings of the 22nd ACM CIKM*, pages 1107–1116, 2013.

[32] N. Zhiltsov and E. Agichtein. Improving Entity Search over Linked Data by Modeling Latent Semantics. In *Proceedings of the 22nd ACM CIKM*, pages 1253–1256, 2013.

[33] N. Zhiltsov, A. Kotov, and F. Nikolaev. Fielded Sequential Dependence Model for Ad-Hoc Entity Retrieval in the Web of Data. In *Proceedings of the 38th ACM SIGIR*, pages 253–262, 2015.