



## UvA-DARE (Digital Academic Repository)

### Overview of the INEX 2013 Linked Data Track

Gurajada, S.; Kamps, J.; Mishra, A.; Schenkel, R.; Theobald, M.; Wang, Q.

*Published in:*  
CEUR Workshop Proceedings

[Link to publication](#)

*Citation for published version (APA):*

Gurajada, S., Kamps, J., Mishra, A., Schenkel, R., Theobald, M., & Wang, Q. (2013). Overview of the INEX 2013 Linked Data Track. *CEUR Workshop Proceedings*, 1179.

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<http://dare.uva.nl>)*

# Overview of the INEX 2013 Linked Data Track

Sairam Gurajada<sup>1</sup>, Jaap Kamps<sup>2</sup>, Arunav Mishra<sup>1</sup>, Ralf Schenkel<sup>3</sup>,  
Martin Theobald<sup>4</sup>, and Qiuyue Wang<sup>5</sup>

<sup>1</sup>Max Planck Institute for Informatics, Saarbrücken, Germany

<sup>2</sup>University of Amsterdam, Amsterdam, The Netherlands

<sup>3</sup>University of Passau, Passau, Germany

<sup>4</sup>University of Antwerp, Antwerp, Belgium

<sup>5</sup>Renmin University of China, Beijing, China

**Abstract.** This paper provides an overview of the INEX Linked Data Track, which went into its second iteration in 2013.

## 1 Introduction

As in the previous year [7], the goal of the INEX Linked Data Track<sup>1</sup> was to investigate retrieval techniques over a combination of textual and highly structured data, where rich textual contents from Wikipedia articles serve as the basis for retrieval and ranking techniques, while additional RDF properties carry key information about semantic relationships among entities that cannot be captured by keywords alone. As opposed to the previous year, the Linked Data Track employed a new form of a *reference collection*, which was purely based on openly available dumps of English Wikipedia articles (using a snapshot from June 1st, 2012, in MediaWiki XML format) plus two canonical subsets of the DBpedia 3.8 [3] and YAGO2 [4] collections (in RDF NT format). In addition to this reference collection, we provided *two supplementary collections*, one in an article-centric XML format and one in a pure text format, respectively, in order to allow for a large variety of retrieval techniques, based on either RDF, XML, or text to be incorporated into this retrieval setting. Moreover, links among the Wikipedia, DBpedia 3.8, and YAGO2 URI's were provided (again in RDF NT format) in order to allow for an easy integration of all of the above data sources. Participants were thus free to choose their preferred format of the collections in order to submit their runs. The goal in organizing this new track thus follows one of the key themes of INEX, namely to explore and investigate if and how structural information could be exploited to improve the effectiveness of ad-hoc retrieval. In particular, we were interested in how this combination of data could be used together with structured queries Jeopardy-style natural-language clues and questions. The Linked Data Track thus aims to close the gap between IR-style keyword search and Semantic-Web-style reasoning techniques, with the goal to bring together different communities and to foster research at the intersection of Information Retrieval, Databases, and the Semantic Web. For INEX 2013, we specifically explored the following two retrieval tasks:

<sup>1</sup> <https://inex.mmci.uni-saarland.de/tracks/lod/>

- The **Ad-hoc Retrieval Task** investigates informational queries to be answered mainly by the textual contents of the Wikipedia articles.
- The **Jeopardy Task** employs natural-language Jeopardy clues which are manually translated into a semi-structured query format based on SPARQL with additional keyword-based filter conditions.

## 2 Data Collections

### 2.1 Reference Collection

As for the *reference collection*, the Linked Data Track employed a combination of three data collections from Wikipedia, DBpedia 3.8 and YAGO2.

- Core of the reference collection is the dump of the English Wikipedia articles from June 1st, 2012, which is available from the following URL:  
<http://dumps.wikimedia.org/enwiki/20120601/enwiki-20120601-pages-articles.xml.bz2>.

The following subsets of the canonicalized datasets from DBpedia 3.8 were included in the reference collection:

- DBpedia Infobox Types  
[http://downloads.dbpedia.org/3.8/en/instance\\_types\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/instance_types_en.nt.bz2)
- DBpedia Infobox Properties  
[http://downloads.dbpedia.org/3.8/en/mappingbased\\_properties\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/mappingbased_properties_en.nt.bz2)
- DBpedia Titles  
[http://downloads.dbpedia.org/3.8/en/labels\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/labels_en.nt.bz2)
- DBpedia Geographic Coordinates  
[http://downloads.dbpedia.org/3.8/en/geo\\_coordinates\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/geo_coordinates_en.nt.bz2)
- DBpedia Homepages  
[http://downloads.dbpedia.org/3.8/en/homepages\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/homepages_en.nt.bz2)
- DBpedia Person Data  
[http://downloads.dbpedia.org/3.8/en/persondata\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/persondata_en.nt.bz2)
- DBpedia Inter-Language Links  
[http://downloads.dbpedia.org/3.8/en/interlanguage\\_links\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/interlanguage_links_en.nt.bz2)
- DBpedia Article Categories  
[http://downloads.dbpedia.org/3.8/en/article\\_categories\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/article_categories_en.nt.bz2)
- DBpedia Page IDs  
[http://downloads.dbpedia.org/3.8/en/page\\_ids\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/page_ids_en.nt.bz2)
- DBpedia External Links  
[http://downloads.dbpedia.org/3.8/en/external\\_links\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/external_links_en.nt.bz2)
- DBpedia to Wikipedia Page Links  
[http://downloads.dbpedia.org/3.8/en/page\\_links\\_en.nt.bz2](http://downloads.dbpedia.org/3.8/en/page_links_en.nt.bz2)
- DBpedia Links to YAGO2  
[http://downloads.dbpedia.org/3.8/links/yago\\_links.nt.bz2](http://downloads.dbpedia.org/3.8/links/yago_links.nt.bz2)
- DBpedia links to other Linked Data collections (not required for queries, only for interlinking with external collections)  
<https://inex.mmci.uni-saarland.de/tracks/lod/dbpedia-links.txt>

Additionally, the following subsets of the canonicalized datasets from YAGO2s were also included in the reference collection:

- YAGO2: domains, ranges and confidence values of relations  
<http://mpii.de/yago-naga/yago/download/yago/yagoSchema.ttl.7z>
- YAGO2: rdf:type class instances  
<http://mpii.de/yago-naga/yago/download/yago/yagoTypes.ttl.7z>
- YAGO2: rdf:subclassOf taxonomy facts  
<http://mpii.de/yago-naga/yago/download/yago/yagoTaxonomy.ttl.7z>
- YAGO2: facts between instances  
<http://mpii.de/yago-naga/yago/download/yago/yagoFacts.ttl.7z>
- YAGO2: facts with labels  
<http://mpii.de/yago-naga/yago/download/yago/yagoLabels.ttl.7z>
- YAGO2: facts with literals  
<http://mpii.de/yago-naga/yago/download/yago/yagoLiteralFacts.ttl.7z>
- YAGO2: links to DBpedia 3.8 instances  
<http://mpii.de/yago-naga/yago/download/yago/yagoDBpediaInstances.ttl.7z>

Since Wikipedia, DBpedia 3.8 and YAGO2 employ different URI's as identifiers for their target entities, valid results were restricted to a provided list of valid DBpedia URI's<sup>2</sup>, which contains one RDF triple of the form

```
<DBpedia-URI> lod:isValid "true"
```

for each valid result entity; other forms of this list (such as valid Wikipedia article ids) were available on request. If a run included an entity not in this list, the entity was considered as non-relevant.

As in the previous year, the Linked Data Track was explicitly intended to be an “open track” and thus invited participants to include more Linked Data sources (see, e.g., <http://linkeddata.org>) or other sources that go beyond “just” DBpedia and YAGO2. Any inclusion of further data sources was welcome, however, workshop submissions and follow-up research papers should explicitly mention these sources when describing their approaches.

## 2.2 Supplementary XML Collection

The new version of the XML-based Wikipedia-LOD (v2.0) collection (compare to [6, 2]) was again hosted at the Max Planck Institute for Informatics and has been made available for download in March 2013 from the Linked Data Track homepage. The collection consists of 4 compressed tar.gz files and contains an overall amount of 12.2 Million individual XML articles with more than 1 Billion XML elements. Each Wikipedia-LOD article consists of a mixture of XML tags, attributes, and CDATA sections, containing infobox attributes, free-text contents of the Wikipedia articles which describe the entity or category that the article captures, and a section with RDF properties exported from the DBpedia

<sup>2</sup> [http://inex-lod.mpi-inf.mpg.de/2013/List\\_of\\_Valid\\_DBpedia\\_URIs.ttl](http://inex-lod.mpi-inf.mpg.de/2013/List_of_Valid_DBpedia_URIs.ttl)

3.8 and YAGO2 subsets of the reference collection that are related to the article's entity. All sections contain links to other Wikipedia articles (including links to the corresponding DBpedia and YAGO2 resources), Wikipedia categories, and external Web pages. Figure 1 shows the structure of such a Wikipedia article in XML format about the entity `Albert_Einstein`. It depicts the five main components of the XML markup of these articles:

- i) the **metadata** section, which contains information about the author, title, and id of the article, as well as possible links to other Linked Data URI's,
- ii) the **infobox** properties taken from the original attributes and values from the Wiki markup of this article,
- iii) the **Wikipedia** section with additional Linked Data links to related entities in Wikipedia, DBpedia 3.8, YAGO2, and links to external web pages,
- iv) the **DBpedia properties** section, with RDF properties from DBpedia 3.8 about the entity that is described by this article, and
- v) a similar section with **YAGO2 properties** about the entity that is described by this article.

**Wikipedia to XML Conversion.** For converting the raw Wikipedia articles into our XML format, we modified and substantially extended the `wiki2xml` parser<sup>3</sup> as it is provided for the `MediaWiki`<sup>4</sup> format. The parser generates an XML file from the raw Wikipedia article (originally stored in Wiki markup) by transforming infobox information to a proper XML representation, matching the Wikipedia URI's to their corresponding DBpedia 3.8 and YAGO2 URI's, and finally by annotating each article with a list of RDF properties from the DBpedia 3.8 and YAGO2 sources.

**Collection Statistics.** The Wikipedia-LOD v2.0 collection currently contains 12.2 Million XML documents in 4 compressed `tar.gz` files, thus counting to the size of 90.7 GB in uncompressed form and 11.1 GB in compressed form, respectively. Table 1 provides more detailed numbers about different statistics of this supplementary collection.

### 2.3 Supplementary Text Collection

As a second supplementary text collection, all XML articles of the Wikipedia-LOD v2.0 collection were once more transformed into a plain text format by extracting all the CDATA sections of the content-related XML elements (including the infobox and RDF properties sections). In order to keep the original text structure of the Wikipedia articles intact as much as possible, our transformation tools marks links, infobox tags, and RDF properties by additional brackets. All full-text dumps of this second supplementary collection are available from the Linked

<sup>3</sup> <http://www.mediawiki.org/wiki/Extension:Wiki2xml>

<sup>4</sup> <http://www.mediawiki.org/>

```

<lodxml xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance' xml:lang='en'
  xmlns:xhtml='http://www.w3.org/1999/xhtml' encoding='UTF-8'>
  <article title='Albert_Einstein'>
    <wikipedia>
      <template type='Metadata'>
        <tag name='id'>736</tag>
        <tag name='title'>Albert_Einstein</tag>
      </template>
      <infobox type='person'>
        <tag name='name'>Albert Einstein</tag>
        <tag name='image'>Einstein 1921 portrait2.jpg</tag>
        .....
      </infobox>
      .....
      <paragraph>Albert Einstein was born in<space/>
      .....
      <link>
        <wikilink href='http://en.wikipedia.org/wiki/Ulm'>Ulm</wikilink>
        <dbpedia href='http://dbpedia.org/resource/Ulm' />
        <yago ref='Ulm' />
      </link>
      .....
    </wikipedia>
    <dbpediaproperties>
      <property name='http://dbpedia.org/property/birthPlace'>
        <object name='http://dbpedia.org/resource/German_Empire' />
      </property>
      .....
    </dbpediaproperties>
    <yagoproperties>
      <property name='isKnownFor'><object name='Einstein_field_equations' /></property>
      <property name='type'><object name='wikicategory_German_pacifists' /></property>
      <property name='type'><object name='wikicategory_German_philosophers' /></property>
      <property name='type'><object name='wikicategory_German_physicists' /></property>
    </yagoproperties>
  </article>
</lodxml>

```

Fig. 1. XML-ified Wikipedia articles with DBpedia 3.8 and YAGO2 properties.

Data Track homepage (<https://inex.mmci.uni-saarland.de/tracks/lod/>). A provided file<sup>5</sup> again maps each DBpedia entity to its corresponding text file. The resulting size of this text collection amounts to 11,945,084 files with non-empty text contents, with an overall size of 17 GB in uncompressed form and 5.5 GB in compressed form, respectively.

### 3 Retrieval Tasks and Topics

#### 3.1 Ad-hoc Task

The goal of the Ad-hoc Task is to return a ranked list of results in response to a search topic that is formulated as a keyword query. Results had to be represented by their Wikipedia page ID's, which in turn had to be linked to the set of valid

<sup>5</sup> <http://inex-lod.mpi-inf.mpg.de/2013/dbpedia-textfiles-map.ttl>

Property	Count
XML documents	12,216,083
XML elements	1,169,642,510
Internal Wikipedia links resolved	144,481,793
Wikipedia URI's resolved	215,621,680
DBpedia URI's resolved	144,497,401
YAGO2 URI's resolved	156,763,342

**Table 1.** Wikipedia-LOD v2.0 collection statistics.

DBpedia URI's (see above). A set of 144 Ad-hoc Task search topics for the INEX 2013 Linked Data track had been released in March 2013 and was made available for download from the Linked Data Track homepage. In addition, the set of QRels from the 2012 Ad-Hoc Task topics was provided for training.

**Submission Format.** Participants were allowed to submit up to 3 runs. Each run could contain a maximum of 1,000 results per topic, ordered in decreasing value of relevance. As in the previous year, a result is an article or an entity, identified by its Wikipedia page ID (so only entities from DBpedia or, equivalently, articles from Wikipedia were counted as valid results). The results of one run had to be contained in a single submission file, so up to three files could be submitted by each participant in total. Submissions were required to be in the familiar TREC format.

```
<qid> Q0 <file> <rank> <rsv> <run_id>
```

Where:

- The first column is the topic number.
- The second column is the query number within that topic. As of the early TREC days, this field is unused and should always be Q0.
- The third column is the ID of the result Wikipedia page.
- The fourth column is the rank of the result.
- The fifth column shows the score (integer or floating point) that generated the ranking.
- The sixth column is called the “run tag” and should be a unique identifier for the participating group *and* for the method used. Run tags must contain 12 or fewer letters and numbers, with no punctuation, to facilitate labeling graphs with the tags.

An example submission thus might have looked as follows:

```
2013001 Q0 12 1 0.9999 2013UniXRun1
2013001 Q0 997 2 0.9998 2013UniXRun1
2013001 Q0 9989 3 0.9997 2013UniXRun1
```

This run contains three results for the topic 2013001. The first result is the Wikipedia target entity that is associated with the page ID “12”. The second result is the page with ID “997”, and so on. Mappings between Wikipedia page ID’s and DBpedia URI’s were available from the DBpedia-to-Wikipedia-Page-Links file which is part of the reference collection. Results were restricted to target entities in the list of valid DBpedia URI’s (see above).

### 3.2 Jeopardy Task

As in 2012, the Jeopardy Task continued to investigate retrieval techniques over a set of natural-language Jeopardy clues, which were manually translated into SPARQL query patterns with additional keyword-based filter conditions. A set of 105 Jeopardy Task search topics, out of which 74 topics were taken over from 2012 and 31 topics were newly added to the 2013 setting. 72 single-entity topics (with one query variable) were also included into the set of 144 Ad-hoc topics. All topics were made available for download in March 2013 from the Linked Data Track homepage. In analogy to the Ad-hoc Task, the set of topics from 2012 was provided together with their QRels for training.

We illustrate the topic format with the example of topic 2012374 from the set of the 2013 topics. It is represented in XML format as follows:

```
<topic id="2012374" category="Politics">
  <jeopardy_clue>
    Which German politician is a successor of another politician
    who stepped down before his or her actual term was over,
    and what is the name of their political ancestor?
  </jeopardy_clue>
  <keyword_title>
    German politicians successor other stepped down before
    actual term name ancestor
  </keyword_title>
  <sparql_ft>
    SELECT ?s ?s1 WHERE {
      ?s rdf:type <http://dbpedia.org/class/yago/GermanPoliticians>.
      ?s1 <http://dbpedia.org/property/successor> ?s.
      FILTER FTContains (?s, "stepped down early").
    }
  </sparql_ft>
</topic>
```

The `<jeopardy_clue>` element contains the original Jeopardy clue as a natural-language sentence; the `<keyword_title>` element contains a set of keywords that has been manually extracted from this title and has been reused as part of the Ad-hoc Task; and the `<sparql_ft>` element contains the result of a manual conversion of the natural-language sentence into a corresponding SPARQL query. The `category` attribute of the `<topic>` element may be used as an additional



hint for disambiguating the query. In the above query, `?s` is a variable for an entity of type `http://dbpedia.org/class/yago/GermanPoliticians` (in the first triple pattern), and it should be in a `http://dbpedia.org/property/successor` relationship with another entity denoted by the variable `?s1`. The `FTContains` filter condition restricts `?s` to those entities that should be associated with the keywords “stepped down early” via its linked Wikipedia article.

Since this particular variant of SPARQL with full-text filter conditions cannot be run against a standard RDF collection (such as DBpedia 3.8 or YAGO2) alone, participants were again encouraged to develop individual solutions to index both the RDF and textual contents of the Wikipedia reference or supplementary collections in order to process these queries.

**Submission Format.** Similar to the Ad-hoc Task (see above), each participating group was allowed to submit up to 3 runs. Each run could contain a maximum of 1,000 results per topic, ordered by decreasing value of relevance (although we expect most topics to have just one or a combination of a few target entities). The results of one run must be contained in a single submission file, that is, up to 3 files could be submitted per group in total. For relevance assessments and evaluation of the results, the runs were again required to be in the familiar TREC format, however containing one row of target entities (denoted by their Wikipedia page ID’s, which are available in the reference collection through the `http://dbpedia.org/ontology/wikiPageID` properties) for each query result. Each row of target entities must reflect the order of query variables as specified by the `SELECT` clause of the Jeopardy topic. In case the `SELECT` clause contained more than one query variable, the row should consist of a comma- or semicolon-separated list of such target entity ID’s. Thus, an example submission may have looked as follows:

```
2012374 Q0 12;24 1 0.9999 2012UniXRun1
2012374 Q0 997;998 2 0.9998 2012UniXRun1
2012374 Q0 9989;12345 3 0.9997 2012UniXRun1
```

Here, there are 3 results for topic “2012374”; and we can see this topic requests two entities per result, since it has two variables in the `SELECT` clause. The first result is the entity pair (denoted by their Wikipedia page ID’s) with the ID’s “12” and “24”, the second result is the entity pair with the ID’s “997” and “998”, and the third result is the entity pair with the ID’s “9989” and “12345”. For the evaluation, symmetric results, where the order of the returned entities did not matter, were considered as duplicates and were automatically removed at the lower rank of the run at which the duplicate occurred. Mappings between DBpedia URI’s and Wikipedia page ID’s were available from the DBpedia-to-Wikipedia-Page-Links file which was part of the reference collection. And, again, results were restricted to target entities contained in the list of valid DBpedia URI’s (see above).

## 4 Run Submissions & Evaluation

All run submissions were to be uploaded via the INEX website via the URL: <https://inex.mmci.uni-saarland.de/>. The due date for the submission of all Linked Data Track runs was May 15, 2012. In total, 5 Ad-hoc search runs were submitted by 2 participants, i.e., Oslo and Akershus University College of Applied Sciences (OAUC), Renmin University of China (RUC), and 3 Jeopardy runs were submitted by the Max-Planck Institute for Informatics (MPI).

### 4.1 Assessments

For the Ad-hoc Task, assessments for the 72 single-entity Jeopardy topics were done on Amazon Mechanical Turk by pooling the top-100 ranks from the 8 submitted runs in a round-robin fashion. Assessments for the remaining 72 Ad-hoc Task topics from INEX 2009 and 2010 were taken over from the previous years. (Notice that the latter provide only an approximation of the actual relevance judgments for these topics, since the collection has meanwhile changed.) Table 2 provides detailed statistics about the assessments of the 144 Ad-hoc Task topics.

Topic Set	Number of topics	Number of relevant results per topic					Total
		Min	Max	Median	Mean	Std. Deviation	
2009/2010	72	24	95	63	63	16	4542
Jeopardy	72	3	72	26	27	12	1929
all	144	3	95	42	45	23	6471

**Table 2.** Statistics of the assessment results for the 72 Ad-hoc Task topics from INEX 2009/2010 and the 72 Jeopardy Task topics of INEX 2013.

For the Jeopardy Task, assessments for 77 single- and multi-entity topics were additionally done on Crowdfunder by pooling the top-10 results from the 3 Jeopardy submissions for the single-entity topics and by pooling the top-20 for the multi-entity topics, respectively, again in a round-robin fashion. These assessments were done based on an entity-centric rather than a document-centric evaluation mode, i.e., there was usually just a single target entity (or a short list of target entities) to marked as relevant for a given SPARQL-FT topic. Overall, 144 Ad-hoc topics and 77 Jeopardy topics were finally assessed this way.

### 4.2 Metrics

The TREC-eval tool was adapted to calculate the following well-known metrics (see [1, 5]) used in ad-hoc and entity ranking settings: Precision, Recall, Average-

Precision (AP), Mean-Average-Precision (MAP), Mean-Reciprocal-Rank (MRR), and Normalized-Discounted-Cumulated-Gain (NDCG).

For the Ad-hoc Task, we employed the usual binary relevance assessments obtained from a majority vote over the judgments obtained from AMT for each result. For the Jeopardy Task, which yielded different QRels than the Ad-hoc Task, we additionally had to distinguish between four types of search topics in order to obtain similar binary relevance assessments. These four types divide the set of 105 Jeopardy topics as follows:

- **46 single-entity, single-target topics:** these are typical Jeopardy clues which have just one relevant target entity as result.
- **27 single-entity, multiple-target topics:** these are entity-centric topics which may have an entire list of relevant target entities as result.
- **17 multiple-entity, single-target topics:** these are enhanced Jeopardy clues which have just one combination of relevant target entities as result.
- **15 multiple-entity, multiple-target topics:** these are enhanced entity-centric topics which may have an entire list of combinations of relevant target entities as result.

For the multiple-entity topics, a combination of entities was considered to be relevant at a particular rank, only if all the entities of this combination formed a correct answer to the topic. That is, relevance judgments for Jeopardy topics were still based on binary assessments. Moreover, duplicate results (including duplicates due to symmetric answers for multi-entity topics) were removed from the lower ranks of the run files at which they occurred. For completeness, we next list the detailed definition of the above metrics.

$$Precision(P) = \frac{\text{Number of relevant results returned}}{\text{Total number of results returned}} \quad (1)$$

$$Precision\text{-at-}k(P@k) = \frac{\text{Number of relevant results at rank } k}{k} \quad (2)$$

*Precision(P)* is defined as the ability of a system to present all relevant items. It is a simple statistical set-based measure calculated as shown by Equation 1. *Precision-at-K (P@K)* is the portion of the relevant documents in the first *K* ranks and is calculated as shown by Equation 2.

$$Recall(R) = \frac{\text{Number of relevant results returned}}{\text{Total number of relevant results}} \quad (3)$$

*Recall(R)* is also a set-based measure that can be perceived as the probability of a system to return correct entities. It can be computed as shown in Equation 3. A standard technique to compute *Interpolated-Precision (iP)* at a given recall level is to use the maximum precision for any actual recall level greater than or equal to the recall level in question. This is modeled by Equation 4.

$$Interpolated\text{-Precision-at-}k = \max_{k' > k}(P(k')), \quad (4)$$

where  $k$  and  $k'$  are recall levels.

To measure the average performance of a system over a set of queries, each with different number of relevant entities, we compute the Interpolated-Precision at a set of 11 standard recall levels (specifically, 1%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%). *Average-Interpolated-Precision (AiP)* is a single-valued measure that reflects the performance of an engine over all the relevant results. We thus report the *Mean-Average-Interpolated-Precision (MAiP)* that reflects the performance of a system over all the results. This is simply the mean of the AiP for each topic as shown by Equation 5.

$$\text{Mean-Average-Interpolated-Precision}(MAiP) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} iP(RL_j) \quad (5)$$

where  $|Q|$  is the total number of topics,  
 $m_j$  is the total number of relevant results for topic  $q_j$ ,  
 $RL_j$  is the ranked list of results returned for topic  $q_j$ .

The *Reciprocal-Rank (1/r)* of a query can be defined as the rank  $r$  at which a system returns the first relevant entity. In our case, we report the average of the reciprocal rank scores over all topics, known as *Mean-Reciprocal-Rank (MRR)*.

Finally, we present the *Normalized-Discounted-Cumulative-Gain (NDCG)* at top 5, 10 and 15 results to evaluate systems in an ah-hoc and entity-oriented retrieval setting. Discounted-Cumulative-Gain (DCG) uses a graded relevance scale to measure the gain of a system based on the positions of the relevant entities in the result set. This measure gives a lower gain to relevant entities returned in the lower ranks to that of the higher ranks. This makes a sensible measure to use for our task as we reward engines that retrieve relevant results at the top ranks. NDCG reports a single-valued score by normalizing the DCG, thus accounting for differently sized output lists.  $NDCG(Q, K)$ , i.e., NDCG at  $K$  for a set of queries  $Q$ , is computed as shown in Equation 6.

$$NDCG(Q, k) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} Z_{kj} \sum_{m=1}^k \frac{2^{R(j,m)} - 1}{\log_2(1 + m)} \quad (6)$$

where  $|Q|$  is the total number of topics,  
 $R(j, e)$  is the binary relevance score obtained for an individual result of topic  $j$ ,  
 $Z_{jk}$  is the normalization factor,  
 $k$  is the rank at which NDCG is calculated.

## 5 Results

### 5.1 Ad-hoc Task

As mentioned above, 144 Ad-hoc Task topics were collected from two different sources: 72 of them are old topics from INEX 2009 and 2010, and 72 of them are single-entity Jeopardy topics. In this section, we will first present the evaluation results over the whole set of Ad-hoc Task topics for all the submitted runs, and then analyze the effectiveness of the runs for each of the two sets of topics. Table 3 presents the evaluation results for all the 8 submitted runs. Even though 3 runs were submitted to the Jeopardy Task, we evaluated them altogether since there are 72 Ad-hoc Task topics are the same as for the Jeopardy Task. The results show that the 3 Jeopardy runs have higher the Mean-Reciprocal-Rank (MRR), which means most of time they returned the first relevant results earlier than other runs. But in terms of MAiP and other metrics, the run from RUC performed best. Table 4 shows the evaluation results for the 5 Ad-hoc Task runs over the 72 old topics from the INEX 2009 and 2010 Ad-Hoc Tasks. Table 5 shows the results over the 72 single-entity Jeopardy topics for all the submitted runs, now evaluated by MRR. We can observe that the 3 runs submitted to the Jeopardy Task have much higher MRR. That means that most of time they returned the first relevant results earlier than the other 5 runs.

### 5.2 Jeopardy Task

Table 6 depicts the detailed retrieval results for the 3 runs submitted to the Jeopardy Task by the MPI group, which was the only group that participated in this task. These evaluations are based on the distinct set of QREs, which were specifically created for the Jeopardy Task (see above).

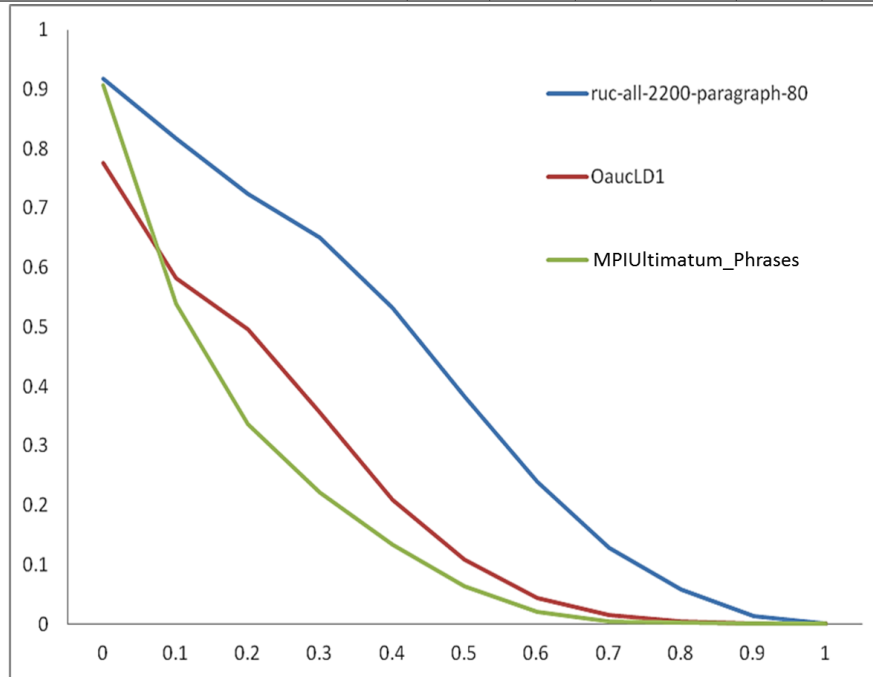
## 6 Conclusions

The Linked Data Track, which was continued track in INEX 2012, was organized towards our goal to close the gap between IR-style keyword search and Semantic-Web-style reasoning techniques. The track thus continues one of the earliest guiding themes of INEX, namely to investigate whether structure may help to improve the results of ah-hoc keyword search. As a core of this effort, we introduced a new and much larger supplementary XML collection, coined *Wikipedia-LOD v2.0*, with XML-ified Wikipedia articles which were additionally annotated with RDF properties from both DBpedia 3.8 and YAGO2. However, due to the very low number of participating groups, in particular for the Jeopardy, detailed comparisons of the underlying ranking and evaluation techniques can only be drawn very cautiously.

## 7 Acknowledgements

This work is partially supported by National 863 High-tech project, No: 2012AA011001.

Run	MAiP	MRR	P@5	P@10	P@20	P@30
ruc-all-2200	0.3733	0.8772	0.7028	0.6424	0.5979	0.5544
ruc-all-2200-rerank	0.2408	0.7957	0.5958	0.5229	0.4549	0.4134
<b>ruc-all-2200-paragraph-80</b>	<b>0.388</b>	0.8861	<b>0.725</b>	<b>0.6674</b>	<b>0.6174</b>	<b>0.5646</b>
ruc-all-2200-paragraph-80-rerank	0.2577	0.7922	0.5986	0.5403	0.4903	0.4426
OaucLD1	0.2112	0.7449	0.5458	0.509	0.4618	0.4127
MPISupremacy	0.1489	0.8684	0.4444	0.3204	0.2407	0.171
MPIUltimatum_Phrases	0.1677	<b>0.8888</b>	0.4689	0.3459	0.2615	0.1891
MPIUltimatum_NoPhrase	0.1629	0.8786	0.4689	0.3443	0.2607	0.1896



**Table 3.** Evaluation results for all submitted runs over all the 144 Ad-hoc Task topics.

## References

1. Amer-Yahia, Sihem and Lalmas, Mounia. XML search: languages, INEX and scoring. *SIGMOD Record*, 35, 2006.
2. P. Bellot, T. Chappell, A. Doucet, S. Geva, S. Gurajada, J. Kamps, G. Kazai, M. Koolen, M. Landoni, M. Marx, A. Mishra, V. Moriceau, J. Mothe, M. Pre-minger, G. Ramírez, M. Sanderson, E. SanJuan, F. Scholer, A. Schuh, X. Tannier, M. Theobald, M. Trappett, A. Trotman, and Q. Wang. Report on INEX 2012. *SIGIR Forum*, 46(2):50–59, 2012.

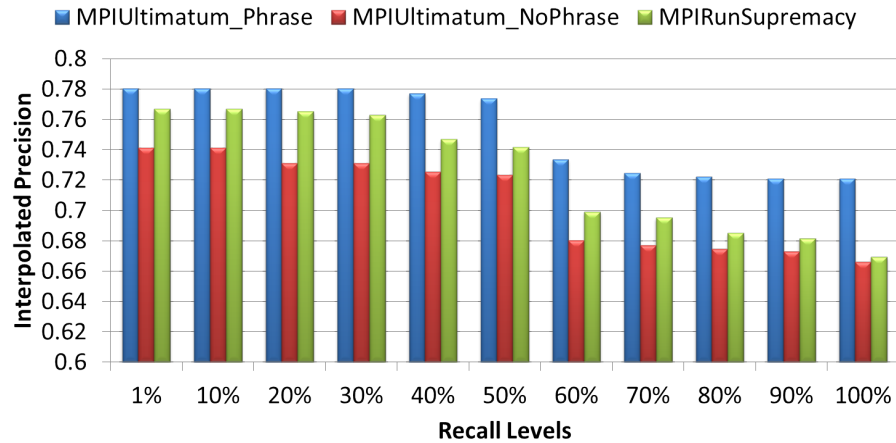
Run	MAiP	MRR	P@5	P@10	P@20	P@30
ruc-all-2200	0.454	0.9491	0.8389	0.8153	0.7833	0.7648
ruc-all-2200-rerank	0.3281	0.9144	0.7889	0.7208	0.65	0.613
ruc-all-2200-paragraph-80	<b>0.4631</b>	<b>0.9468</b>	<b>0.8611</b>	<b>0.8292</b>	<b>0.7958</b>	<b>0.7708</b>
ruc-all-2200-paragraph-80-rerank	0.3384	0.933	0.7778	0.7319	0.6813	0.6352
OaucLD1	0.2922	0.897	0.7444	0.7097	0.666	0.6231

**Table 4.** Evaluation results for all Ad-hoc Task runs over the 72 INEX 2009 and 2010 topics.

Run	MAiP	MRR	P@5	P@10	P@20	P@30
ruc-all-2200	0.2926	0.8054	0.5667	0.4694	0.4125	0.344
ruc-all-2200-rerank	0.1534	0.677	0.4028	0.325	0.2597	0.2139
ruc-all-2200-paragraph-80	<b>0.3128</b>	0.8255	<b>0.5889</b>	<b>0.5056</b>	<b>0.4389</b>	<b>0.3583</b>
ruc-all-2200-paragraph-80-rerank	0.1769	0.6513	0.4194	0.3486	0.2993	0.25
OaucLD1	0.1302	0.5929	0.3472	0.3083	0.2576	0.2023
MPISupremacy	0.1489	0.8684	0.4444	0.3204	0.2407	0.171
MPIUltimatum_Phrases	0.1677	<b>0.8888</b>	0.4689	0.3459	0.2615	0.1891
MPIUltimatum_NoPhrase	0.1629	0.8786	0.4689	0.3443	0.2607	0.1896

**Table 5.** Evaluation results for all Ad-hoc Task runs over the 72 INEX 2013 single-entity Jeopardy topics.

	MPIUltimatum_Phrase	MPIUltimatum_NoPhrase	MPISupremacy
MAiP	0.7491	0.701	0.719
MRR	0.7671	0.7358	0.7539
NDCG@5	0.7723	0.7307	0.7393
NDCG@10	0.7864	0.7347	0.7598
NDCG@15	0.7968	0.7484	0.7728
AiP@1%	0.7804	0.7411	0.7669
AiP@10%	0.7804	0.7411	0.7669
AiP@20%	0.7804	0.731	0.7653
AiP@30%	0.7804	0.731	0.763
AiP@40%	0.7772	0.7255	0.7468
AiP@50%	0.7737	0.7232	0.7417
AiP@60%	0.7337	0.6803	0.6991
AiP@70%	0.7245	0.6771	0.6952
AiP@80%	0.7223	0.6747	0.685
AiP@90%	0.7208	0.673	0.6817
AiP@100%	0.7208	0.6662	0.6694



**Table 6.** Evaluation results for the three Jeopardy Task runs over the set of 105 INEX 2013 topics.

3. C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia - A crystallization point for the Web of Data. *J. Web Sem.*, 7(3):154–165, 2009.
4. J. Hoffart, F. M. Suchanek, K. Berberich, and G. Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artif. Intell.*, 194:28–61, 2013.
5. J. Kamps, J. Pehcevski, G. Kazai, M. Lalmas, and S. Robertson. Focused Access to XML Documents. chapter on INEX 2007 Evaluation Measures, pages 24–33. Springer-Verlag, Berlin, Heidelberg, 2008.
6. A. Mishra, S. Gurajada, and M. Theobald. Design and evaluation of an IR-benchmark for SPARQL queries with fulltext conditions. In *ESAIR*, pages 9–10, 2012.



7. Qiuyue Wang and Jaap Kamps and Georgina Ramirez Camps and Maarten Marx and Anne Schuth and Martin Theobald and Sairam Gurajada and Arunav Mishra. Overview of the INEX 2012 Linked Data Track. In *CLEF (Online Working Notes/Labs/Workshop)*, 2012.