# Open Knowledge Extraction Challenge 2017

René Speck[1], Michael Röder[1], Sergio Oramas[2], Luis Espinosa-Anke[2], and
Axel-Cyrille Ngonga Ngomo[1,3]

[1] AKSW Group, University of Leipzig, Germany
`{speck,roeder}@informatik.uni-leipzig.de`
[2] Universitat Pompeu Fabra, Spain
`{sergio.oramas,luis.espinosa}@upf.edu`
[3] University of Paderborn, Germany
`ngonga@upb.de`

**Abstract.** The Open Knowledge Extraction Challenge invites researchers
and practitioners from academia as well as industry to compete to the
aim of pushing further the state of the art of knowledge extraction from
text for the Semantic Web. The challenge has the ambition to provide a
reference framework for research in this field by redefining a number of
tasks typically from information and knowledge extraction by taking into
account Semantic Web requirements and has the goal to test the perfor-
mance of knowledge extraction systems. This year, the challenge goes in
the third round and consists of three tasks which include named entity
identification, typing and disambiguation by linking to a knowledge base
depending on the task. The challenge makes use of small gold standard
datasets that consist of manually curated documents and large silver
standard datasets that consist of automatically generated synthetic doc-
uments. The performance measure of a participating system is twofold
base on (1) Precision, Recall, F1-measure and on (2) Precision, Recall,
F1-measure with respect to the runtime of the system.

**Keywords:** Open Knowledge Extraction Challenge, Semantic Web

## 1 Introduction

The vision of the Semantic Web is an extension of the Document Web with the
goal to allow intelligent agents a better reuse, sharing and understanding of the
data in the Document Web. Agents are then able to automatically interpret the
content of the Document Web. Thus, implementing the vision of the Semantic
Web requires transforming unstructured and semi-structured data with knowl-
edge extraction approaches from the Document Web into structured machine
processable data for the current implementation of the Semantic Web, the Data
Web.

In summary, we expect to trigger attention from the knowledge extraction
community and foster their broader integration with the Semantic Web commu-
nity with the Open Knowledge Extraction (OKE) challenge.

The rest of this paper is structured as follows: We begin with defining the OKE tasks in Section 2. In Section 3, we explain the evaluation with its datasets and scenarios. In Section 4 we give a brief introduction of the participating systems. In Section 5, we compare the results achieved by our evaluation on the gold and silver standard datasets. Finally, we discuss the insights provided by the challenge and possible extensions in Section 6.

## 2    Open Knowledge Extraction Challenge Tasks

The OKE challenge consist of three tasks. The first two tasks comprise named entity identification and named entity linking to the DBpedia knowledge base. For measuring the system performance in different perspectives based on the size and noise of the data, each of this two tasks is subdivided into two scenarios. The size of the data in scenario A is small and the data generation process was curated. In contrast, the size of the data in scenario B is large and the data generation process was carried out automatically with the help of BENGAL[4,5] to produce synthetic data.

The third task comprises named entity recognition and linking to Linked Brainz[6], the music knowledge base that is based on MusicBrainz[7]. This knowledge base is provided by the challenge (see section 3.1) and dubbed MBL.

Both, the given input and the expected output are expressed with the help of the NIF[8][1] vocabularies and ontologies in an RDF serialisation. A participating system is not expected to process any preprocessing (e.g. pronoun resolution[2]) on the input data. In case a resource for an entity is missing in the knowledge base, a participating system is expected to generate a URI with the namespace of `http://aksw.org/notInWiki/` for this emerging entity.

For carrying out the evaluation, this year the OKE challenge is using the HOBBIT benchmarking platform and the benchmark implementation of the HOBBIT project[9] which rely on the GERBIL evaluation framework [12].

### 2.1    Task 1: Focused Named Entity Identification and Linking

The first task aims at the identification and linking of entities of a given, limited set of entity types. It is a two-step process with the identification of named entities (Recognition) and the linking of those entities to resources in DBpedia (D2KB). A competing system is expected to identify named entity mentions in a given document by its start and end index, further to generate a URI to link each identified entity to DBpedia if possible or generate a URI for an emerging entity.

---

[4] `http://github.com/aksw/bengal`

[5] `http://project-hobbit.eu/wp-content/uploads/2017/04/D2.2.1.pdf`

[6] `http://linkedbrainz.c4dmpresents.org/content/linkedbrainz-summary,`

[7] `http://musicbrainz.org`

[8] `http://persistence.uni-leipzig.org/nlp2rdf`

[9] `http://project-hobbit.eu`

The task is limited to a subset of resources in DBpedia, i.e., resources of the DBpedia ontology types: `Person`, `Place` and `Organisation`.

```
1  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3  @prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-
       core#> .
4
5  <http://example.com/example-task1#char=0,91>
6    a nif:RFC5147String , nif:String , nif:Context ;
7    nif:beginIndex "0"^^xsd:nonNegativeInteger ;
8    nif:endIndex "124"^^xsd:nonNegativeInteger ;
9    nif:isString "Leibniz was born in Leipzig in 1646 and attended the
       University of Leipzig from 1661-1666."@en .
```
Listing 1.1: Example request document in task 1.

```
1  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3  @prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .
4  @prefix dbr: <http://dbpedia.org/resource/> .
5  @prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-
       core#> .
6
7  <http://example.com/example-task1#char=0,7>
8    a nif:RFC5147String , nif:String ;
9    nif:anchorOf "Leibniz"@en ;
10   nif:beginIndex "0"^^xsd:nonNegativeInteger ;
11   nif:endIndex "7"^^xsd:nonNegativeInteger ;
12   nif:referenceContext <http://example.com/example-task1#char=0,91> ;
13   itsrdf:taIdentRef dbr:Gottfried_Wilhelm_Leibniz .
14
15 <http://example.com/example-task1#char=20,27>
16   a nif:RFC5147String , nif:String ;
17   nif:anchorOf "Leipzig"@en ;
18   nif:beginIndex "20"^^xsd:nonNegativeInteger ;
19   nif:endIndex "27"^^xsd:nonNegativeInteger ;
20   nif:referenceContext <http://example.com/example-task1#char=0,91> ;
21   itsrdf:taIdentRef dbr:Leipzig .
22
23 <http://example.com/example-task1#char=53,74>
24   a nif:RFC5147String , nif:String ;
25   nif:anchorOf "University of Leipzig"@en ;
26   nif:beginIndex "53"^^xsd:nonNegativeInteger ;
27   nif:endIndex "74"^^xsd:nonNegativeInteger ;
28   nif:referenceContext <http://example.com/example-task1#char=0,91> ;
29   itsrdf:taIdentRef dbr:Leipzig_University .
```
Listing 1.2: Example of the expected response document in task 1.

Listing 1.1 is an example request document of task 1 and Listing 1.2 is the expected response document for the given request document. Both documents are formalized with NIF.

## 2.2 Task 2: Broader Named Entity Identification and Linking

This task extends the former task towards the DBpedia ontology types. Beside the three types of the first task, a competing system might have to identify other types of entities and to link these entities as well. In the first column in Table 1, a complete list of types that are considered in this task is provided. The middle column contains example subtypes of the corresponding class if any such class is available and the last column contains example instances in DBpedia for the related class respectively subtypes.

Table 1: Types, subtypes examples and instance examples for task 2.

| Type | Subtypes | Instances |
|------|----------|-----------|
| Activity | Game, Sport | Baseball,Chess |
| Agent | Organisation, Person | Leipzig_University |
| Award | Decoration, NobelPrize | Humanitas_Prize |
| Disease | | Diabetes_mellitus |
| EthnicGroup | | Javanese_people |
| Event | Competition, PersonalEvent | Battle_of_Leipzig |
| Language | ProgrammingLanguage | English_language |
| MeanOfTransportation | Aircraft, Train | Airbus_A300 |
| PersonFunction | PoliticalFunction | PoliticalFunction |
| Place | Monument, WineRegion | Beaujolais, Leipzig |
| Species | Animal, Bacteria | Cat, Cucumibacter |
| Work | Artwork, Film | Actrius, Debian |

## 2.3 Task 3: Focused Musical Named Entity Recognition and Linking

Task 3 composes of two subtasks (1) focused musical NE identification and classification and (2) linking to the MBL knowledge base that is based on MusicBrainz. Thus the domain of this task is music. A competing system has to fulfil both tasks in order to participate.

Listing 1.3 is an example input document and listing 1.4 the expected annotated document for the given input, both formalized with NIF.

**Task 3A: Focused Musical Named Entity Recognition** This subtask consists of the identification (Recognition) and classification (Typing) of named entities. The task is limited to a subset of resources in MBL, i.e., resources of the

MBL ontology types: `Artist`, `Album` and `Song`. A competing system is expected to identify elements in a given text by its start and end index, further to assign one of the three types to each element.

**Task 3B: Musical NE Linking** In this subtask a participating system has to link the recognised entities of the former subtask to the corresponding resources in MBL if existing or to generate a URI for the emerging entity.

```
1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3 @prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-
      core#> .
4
5 <http://example.com/example-task3#char=0,40>
6   a nif:RFC5147String , nif:String , nif:Context ;
7   nif:beginIndex "0"^^xsd:nonNegativeInteger ;
8   nif:endIndex "40"^^xsd:nonNegativeInteger ;
9   nif:isString "When Simon & Garfunkel split in 1970,..."@en .
```
<div align="center">Listing 1.3: Example request document in task 3</div>

```
1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
3 @prefix itsrdf: <http://www.w3.org/2005/11/its/rdf#> .
4 @prefix dbr: <http://dbpedia.org/resource/> .
5 @prefix mo: <http://purl.org/ontology/mo/> .
6 @prefix artist: <http://musicbrainz.org/artist> .
7 @prefix nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-
      core#> .
8
9 <http://example.com/example-task3#char=5,22>
10   a nif:RFC5147String , nif:String ;
11   nif:anchorOf "Simon & Garfunkel"@en ;
12   nif:beginIndex "5"^^xsd:nonNegativeInteger ;
13   nif:endIndex "22"^^xsd:nonNegativeInteger ;
14   nif:referenceContext <http://example.com/example-task3#char=0,40> ;
15   itsrdf:taIdentRef artist:5d02f264-e225-41ff-83f7-d9b1f0b1874a ;
16   itsrdf:taClassRef mo:MusicArtist .
```
<div align="center">Listing 1.4: Example of the expected response document in task 3.</div>

## 3 Evaluation

Overall, we follow two main evaluation approaches: subjective and objective. The subjective evaluation is based on paper reviews and the objective evaluation is based on computing relevance measures.

The knowledge bases DBpedia and MBL are used and the performance of a system is measured using Recall, Precision, F1-measure and $\beta$. Note that we

reuse the ability of the Gerbil project enabling the benchmarking of systems that link to another knowledge base than DBpedia as long as there exist sameAs links between the two knowledge bases[8].

## 3.1 Datasets

The documents in the datasets might contain emerging entities, i.e., entities that are not part of the KB. These entities have to be marked and a URI has to be generated for them.

The datasets for the challenge are available at the challenge website[10]. Table 2 shows all the datasets available on the site assigned to the tasks and scenarios.

Table 2: Datasets.

| Task | Scenario | File |
| --- | --- | --- |
| 1 | A | Task1/A/training.tar.gz <br> Task1/A/evaluation.tar.gz |
| | B | Task1/B/scenario-b-eval.zip |
| 2 | A | Task2/A/training.tar.gz <br> Task2/A/evaluation.tar.gz |
| | B | Task2/B/scenario-b-eval.zip |
| 3 | A | Task3/A/training.tar.gz <br> Task3/A/evaluation.tar.gz |

The music knowledge base MBL used in task 3 is provided by the challenge at the website in the file MusicBrainzRDF.tar.gz as well.

## 3.2 Measures

Equation 1, 2, 3 and 4 formalize Precision $p_d$, Recall $r_d$, F1-measure $f_d$ and beta $\beta$ the performance measures we compute on the evaluation datasets for each document $d \in D$. They consist of the number of true positives $TP_d$, false positives $FP_d$ and false negatives $FN_d$.

We micro average the performances over the documents[11].

$$p_d = \frac{TP_d}{TP_d + FP_d} \tag{1}$$

$$r_d = \frac{TP_d}{TP_d + FN_d} \tag{2}$$

---

[10]http://hobbitdata.informatik.uni-leipzig.de/oke2017-challenge/

[11]The macro averages for the performance measures can be retrieved from the official Hobbit SPARQL endpoint at http://db.project-hobbit.eu/sparql.

$$f_d = 2 \cdot \frac{p_d \cdot r_d}{p_d + r_d} \tag{3}$$

Let $D$ be a set of documents for which $\beta$ should be calculated. Let $f_d$ be the F1-measure a benchmarked annotation system achieved for a given document $d \in D$ and let $t_d$ be the time (in seconds) the annotation system needed for the annotation of $d$. Then the $\beta$ value is the amount of F1-measure points a system achieves per second for a given amount of documents.

$$\beta = \frac{\sum_{d \in D} f_d}{\sum_{d \in D} t_d} \tag{4}$$

For matching the entity annotation positions of the benchmarked system and the correct entity markings of the datasets we used the *weak annotation matching* defined in [12]. Thus, an entity is counted as having the correct position, if its position overlaps with the correct position of the entity inside the dataset.

For example, our dataset considered "`Franziska Barbara Ley`". If a tool generated a URI for the emerging entity "`Barbara Ley`" and omitted "`Franziska`", it was assigned as a match.

### 3.3 Platform

The benchmark suite for named entity recognition and linking implemented within HOBBIT [4][12] reuses some of the concepts developed within the opensource project GERBIL. These concepts were migrated and adapted to the HOBBIT architecture. The Platform provides two different implementations of the benchmark described in the following subsections. It calculates values of Precision, Recall and F1-measure, measures the time a system needs to answer a request and counts the number of documents that cause errors in the benchmarked system.

**Scenario A: Quality-focused benchmarking** The first type of benchmarking provided by our suite focuses on the measurement of quality a system achieves on a given set of documents. We assume that each benchmark dataset consists of a set of documents. The documents are sent to the benchmarked system one at a time. The benchmarked system generates a response and sends it back before receiving the next document. That means that the benchmarked system can be configured to concentrate all its resources on a single request and does not need to scale to a large number of requests. In this benchmarking, Scenario A, we rely on manually created gold standards.

The goal in this scenario is to achieve a high F1-measure in a quality-focused benchmarking.

---

[12]`http://project-hobbit.eu/wp-content/uploads/2017/04/D2.2.1.pdf`

**Scenario B: Performance-focused benchmarking** The second approach to benchmarking implemented by our platform aims to put a high load on the benchmarked system and to evaluate its runtime and quality in terms of Precision, Recall and F1-measure. This approach hence focuses on the ability of a system to annotate documents in parallel with an increasing amount of load.

The benchmark creates a large amount of synthetic documents from the given KB using BENGAL[13]. These documents are sent to the system in parallel without waiting for responses for previous requests but with predefined delays between the single documents. During a first phase of the benchmark, the generated work load equals 1 document per second. After the 80 documents of this first phase have been sent, the next phase is started using half of the delay of the previous time. This is done for 6 phases. In the seventh and last phase all 80 documents of the phase are sent without a delay, this leads to workloads of $\{1, 2, 4, 8, 16, 32, 80\}$ documents per second during the different phases.

The performance of a system is measured by $\beta$ which is defined in equation 4. The scenarios goal is to achieve a high $\beta$ value in a performance-focused benchmarking.

## 4 Participants

The challenge attracted four research groups. Two systems were not passing the subjective evaluation. The two remaining groups participated with there system in the challenge, ADEL and FOX.

### 4.1 Adel

ADEL [7], base on previous works[5, 6], is an adaptive entity recognition and linking framework based on an hybrid approach that combines various extraction methods to improve the recognition level and an efficient knowledge base indexing process to increase the efficiency of the linking step. It deals with fine-grained entity types, either generic or domain specific. It also can flexibly disambiguate entities from different knowledge bases.

### 4.2 FOX

FOX [9] has been introduced in 2014 as an ensemble learning-based approach combining several diverse state of the art named entity recognition approaches and is based on the work in [3]. The FOX framework[14][10] outperforms the current state of the art entity recognizers. It relies on AGDISTIS [11] to perform named entity disambiguation. AGDISTIS is a pure entity linking approach (D2KB) based on string similarity measures, an expansion heuristic for labels to cope with co-referencing and the graph-based HITS algorithm. The authors

---

published datasets[15] along with their source code and an API[16]. AGDISTIS can only be used for the D2KB task. FOX together with AGDISTIS can be use on the A2KB and the RT2KB task. FOX serves as the baseline system in this OKE challenge.

## 5  Results

In this section we present the results the participating systems reach on the three OKE challenge tasks. Tables 3 and 4 comprise the results for task 1 and 2 on both scenarios A and B. Tables 5 and 6 comprise the results for task 3A and 3B. The tables show the overall measures for Precision, Recall and F1-measure in the first three rows. The last two rows in each table show the averaged time in seconds a system needs to perform a document and the errors a system triggers. Further Tables 3 to 5 show the interim results for step (i) in the next three rows and for step (ii) in the following three rows. For task 3.2 there are no interim results since there are no interim steps in this subtask.

### 5.1  Task 1

The measured values for scenario A in Table 3 show that ADEL outperforms FOX slightly with +1.09% F1-measure in step (i) Recognition. In step (ii) D2KB, FOX outperforms ADEL clearly with +16.82% F1-measure. Overall, FOX outperforms ADEL with +18.29% in Task1 1 in scenario A.

In scenario B, the results are similar to scenario A. In step (i) ADEL outperforms FOX slightly as well as FOX outperforms ADEL clearly in step (ii). Overall, FOX reaches the highest value in scenario B with 65.15% F1-measure while ADEL reaches 20.12% F1-measure. With 6 and 1 errors, the error rates of ADEL and FOX are low compared to the number of 560 documents they had to annotate in this scenario.

Figure 1 depicts on the left side the detailed results for task 1 in scenario B. Surprisingly, ADEL reaches a clearly higher $\beta$ value than FOX in the first phase for one document request per second. This is caused by the fast runtime of ADEL compensating its lower F1-score during that phase. In the following phases, the runtime of both systems increases—a clear sign that they are receiving requests to annotate document while they are still working on other documents. However, compared to FOX, the time that ADEL needs per document increases much more. Since the F1-score of both systems are similar over all phases but the time needed per document of FOX does not increase as much as it does for ADEL the $\beta$ value of FOX remains higher than the value for ADEL. The observation of the increasing

---

[15] http://github.com/AKSW/n3-collection

[16] http://github.com/AKSW/AGDISTIS

[17] http://w3id.org/hobbit/experiments#1497453653558

[18] http://w3id.org/hobbit/experiments#1497440615203

[19] http://w3id.org/hobbit/experiments#1497533785404

[20] http://w3id.org/hobbit/experiments#1497533898908

Table 3: Results on task 1.

| Experiment Type | Micro measures | Scenario A | | Scenario B | |
|---|---|---|---|---|---|
| | | ADEL[17] | FOX[18] | ADEL[19] | FOX[20] |
| A2KB | Precision | 33.24 | 53.61 | 18.28 | 59.12 |
| | Recall | 30.18 | 46.72 | 22.36 | 72.51 |
| | F1-measure | 31.64 | **49.93** | 20.12 | **65.15** |
| Recognition | Precision | 91.62 | 92.47 | 74.39 | 73.27 |
| | Recall | 83.20 | 80.58 | 90.98 | 89.85 |
| | F1-measure | **87.21** | 86.12 | **81.85** | 80.72 |
| D2KB | Precision | 40.15 | 61.96 | 28.03 | 93.87 |
| | Recall | 27.82 | 41.47 | 19.26 | 66.99 |
| | F1-measure | 32.87 | **49.69** | 22.83 | **78.19** |
| | Time | 7.98 | 6.98 | 231.31 | 179.29 |
| | Errors | 0 | 0 | 6 | 1 |
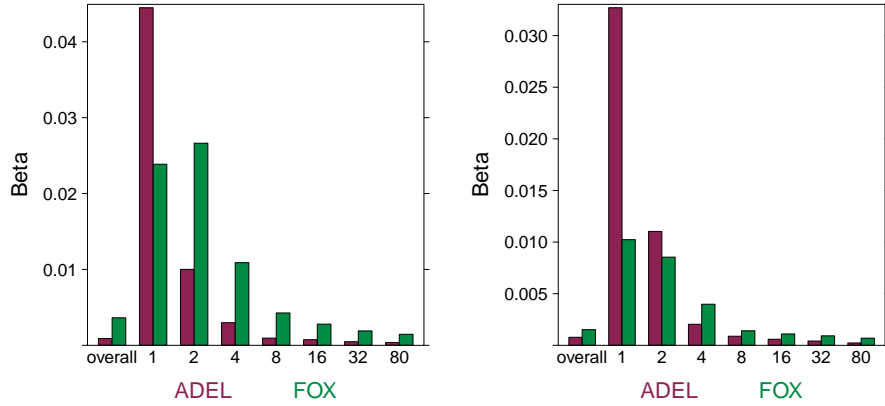


(a) $\beta$ on Task 1.

(b) $\beta$ on Task 2.

Fig. 1: $\beta$ values on several numbers of requests and overall.

of processing time can be also seen in the comparison of the overall values of scenario A and B. While in A, ADEL needs 14% more time per document on average in scenario A this increases to 29% in scenario B. Together with the higher F1-score, the lower runtime of FOX leads to an overall $\beta$ value which is four times higher than the value of ADEL.

## 5.2   Task 2

The measured values for scenario A in Table 4 show that ADEL outperforms FOX slightly with +4.83% F1-measure in step (i) Recognition. In step (ii) D2KB, FOX outperforms ADEL clearly with +14.02% F1-measure. Overall, FOX outperforms ADEL with +16.02% in Task 2 scenario A. In difference to Task 1, ADEL is nearly twice as fast as FOX in scenario A.

In scenario B, the results are similar to A. In step (i) ADEL outperforms FOX as well as FOX outperforms ADEL clearly in step (ii). Overall, FOX reaches the highest value in scenario B with 42.22% F1-measure while ADEL reaches 18.15% F1-measure.

Table 4: Results on task 2.

| Experiment Type | Micro measures | Scenario A | | Scenario B | |
| --- | --- | --- | --- | --- | --- |
| | | ADEL[21] | FOX[22] | ADEL[23] | FOX[24] |
| A2KB | Precision | 31.40 | 56.15 | 17.44 | 44.90 |
| | Recall | 28.14 | 38.53 | 18.93 | 39.83 |
| | F1-measure | 29.68 | **45.70** | 18.15 | **42.22** |
| Recognition | Precision | 87.68 | 95.90 | 72.31 | 74.64 |
| | Recall | 78.57 | 65.80 | 78.50 | 66.21 |
| | F1-measure | **82.88** | 78.05 | **75.27** | 70.17 |
| D2KB | Precision | 39.93 | 63.42 | 28.57 | 82.38 |
| | Recall | 25.76 | 35.28 | 17.47 | 36.92 |
| | F1-measure | 31.32 | **45.34** | 21.68 | **51.00** |
| | Time | 4.60 | 7.66 | 261.48 | 245.99 |
| | Errors | 0 | 1 | 57 | 0 |

Figure 1 depicts on the right side the detailed results for task 2 for scenario B. Similar to task 1, ADEL reaches a clearly higher $\beta$ value than FOX in the first two phases. This is again caused by the lower runtime of ADEL that compensates its lower F1-score. In all other phases FOX reaches a higher $\beta$ value because as in

---

[21]http://w3id.org/hobbit/experiments#1497453720774

[22]http://w3id.org/hobbit/experiments#1497440635319

[23]http://w3id.org/hobbit/experiments#1497533810319

[24]http://w3id.org/hobbit/experiments#1497533871062

Task 1 the runtime of Adel increases much more than the runtime of Fox when it receives many requests in a short amount of time. Overall, Fox nearly reaches a $\beta$ value twice as high as the value achieved by Adel. It is also worth noting that this is the only experiment, in which the error rate of one of the systems is increased. For 57 of the 560 documents, Adel responded with an error code. Nearly all of these errors—9, 26 and 21—occurred during the last three phases. Since the documents are chosen randomly and Adel reported nearly no errors in the phases before, it is possible that they are related to the high load that Adel receives during these phases.

### 5.3 Task 3

Task 3 is composed of two subtask, 3A and 3B. In the following, we first summarize the results on the first subtask and then on the second.

Table 5: Results on Task 3A.

| Experiment Type | Micro measures | Adel[25] | Fox[26] |
| --- | --- | --- | --- |
| RT2KB | Precision | 26.99 | 0 |
| | Recall | 27.24 | 0 |
| | F1-measure | **27.12** | 0 |
| Recognition | Precision | 35.03 | 63.02 |
| | Recall | 74.57 | 49.21 |
| | F1-measure | 47.66 | **55.27** |
| Typing | Precision | 64.33 | 0 |
| | Recall | 64.91 | 0 |
| | F1-measure | **64.62** | 0 |
| | Time | 37.19 | 7.82 |
| | Errors | 16 | 0 |

**Task 3A** The measured values for task 3A are depicted in Table 5. Fox reaches a higher F1-measure than Adel, 55.27% to 47.66% in step (i). In step (ii) Adel reaches a higher F1-measure, since Fox is not supporting this subtask due to the lack of the support of the music entity types.

Overall, Adel reaches the highest value with 27.12% F1-measure on this task.

---

[25]http://w3id.org/hobbit/experiments#1497451343913

[26]http://w3id.org/hobbit/experiments#1497449000101

[27]http://w3id.org/hobbit/experiments#1497453361862

[28]http://w3id.org/hobbit/experiments#1497453423494

Table 6: Results on Task 3.2.

| Experiment Type | Micro measures | ADEL[27] | FOX[28] |
|---|---|---|---|
| D2KB | Precision | 6.82 | 10.10 |
| | Recall | 5.10 | 4.97 |
| | F1-measure | 5.83 | **6.66** |
| | Time | 36.96 | 9.15 |
| | Errors | 16 | 0 |

**Task 3.2** The measured values for task 3.2 are depicted in Table 6. Both systems, ADEL and FOX, reach low performance on this task. ADEL achieves 5.83% and FOX a slightly higher value with 6.66%.

It is noteworthy that FOX processed the documents faster with 9.15s/doc in this subtask than ADEL with 39.96s/doc. Additionally, FOX encountered no errors in comparison to ADEL for which 16 errors have been reported.

## 5.4 Overall

The winner of Task 1 and 2 in both Scenarios A and B is FOX. For task 3A the winner is ADEL, since FOX is not supporting all subtasks. For task 3B the winner is FOX again. Since the advantage ADEL has in Task 3A is larger than the difference between FOX and ADEL in Task 3B, ADEL is the overall winner of Task 3.

The results on Task 1 and 2 suggest, that the Recognition component in ADEL achieved a higher F-measure than the respective component in FOX, but its linking component showed a worse performance than than the respective component in FOX. Thus, it would be interesting to investigate the performance of the composition of the Recognition component of ADEL together with the linking component in FOX in this tasks.

The results on task 3 in the music domain suggest that the Recognition component of FOX achieved a better F-measure than ADEL. While FOX is not supporting the music entity types in its current version. Thus, it would be interesting to investigate the performance of an extended version that supports this types compared to ADEL in this task.

## 6 Conclusion

The Open Knowledge Extraction challenge attracted four research groups coming from Knowledge Extraction and Semantic Web communities. Indeed, the challenge proposal was aimed at attracting research groups from these two communities in order to further investigate exiting overlaps between Knowledge Extraction and the Semantic Web.

Although the participation in terms of the number of competing systems remained quite limited, we believe that the challenge is a breakthrough in the

hybridisation of Semantic Web technologies with Knowledge Extraction methods. As a matter of fact, the evaluation framework is available online and can be reused by the community and for next editions of the challenge.

## 7 Acknowledgement

## References

1. S. Hellmann, J. Lehmann, S. Auer, and M. Brümmer. Integrating nlp using linked data. In *12th International Semantic Web Conference, 21-25 October 2013, Sydney, Australia*, 2013.
2. J. Hobbs. Pronoun resolution. *Lingua*, 44:339–352, 1978.
3. A.-C. Ngonga Ngomo, N. Heino, K. Lyko, R. Speck, and M. Kaltenböck. Scms - semantifying content management systems. In *ISWC 2011*, 2011.
4. A.-C. Ngonga Ngomo and M. Röder. HOBBIT: Holistic benchmarking for big linked data. In *ESWC, EU networking session*, 2016.
5. J. Plu, G. Rizzo, and R. Troncy. *A Hybrid Approach for Entity Recognition and Linking*, pages 28–39. Springer International Publishing, Cham, 2015.
6. J. Plu, G. Rizzo, and R. Troncy. *Enhancing Entity Linking by Combining NER Models*, pages 17–32. Springer International Publishing, Cham, 2016.
7. J. Plu, R. Troncy, and G. Rizzo. ADEL@OKE 2017: A generic method for indexing knowlege bases for entity linking. In *ESWC 2017, 14th European Semantic Web Conference, Open Extraction Challenge, 28th May-1st June 2017, Portoroz, Slovenia*, Portoroz, SLOVENIA, 05 2017.
8. M. Röder, R. Usbeck, and A.-C. Ngonga Ngomo. Techreport for gerbil 1.2.2 - v1. Technical report, Leipzig University, 2016.
9. R. Speck and A.-C. Ngonga Ngomo. Ensemble learning for named entity recognition. In *The Semantic Web – ISWC 2014*, volume 8796 of *Lecture Notes in Computer Science*, pages 519–534. Springer International Publishing, 2014.
10. R. Speck and A.-C. Ngonga Ngomo. Named entity recognition using fox. In *International Semantic Web Conference 2014 (ISWC2014), Demos & Posters*, 2014.
11. R. Usbeck, A.-C. Ngonga Ngomo, M. Röder, D. Gerber, S. Coelho, S. Auer, and A. Both. AGDISTIS - Graph-Based Disambiguation of Named Entities Using Linked Data. In P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz, and C. Goble, editors, *The Semantic Web – ISWC 2014*, volume 8796 of *Lecture Notes in Computer Science*, pages 457–471. Springer International Publishing, 2014.
12. R. Usbeck, M. Röder, A.-C. Ngonga Ngomo, C. Baron, A. Both, M. Brümmer, D. Ceccarelli, M. Cornolti, D. Cherix, B. Eickmann, P. Ferragina, C. Lemke, A. Moro, R. Navigli, F. Piccinno, G. Rizzo, H. Sack, R. Speck, R. Troncy, J. Waitelonis, and L. Wesemann. GERBIL – general entity annotation benchmark framework. In *24th WWW conference*, 2015.