Large-Scale Information Extraction from Textual Definitions through Deep Syntactic and Semantic Analysis

Claudio Delli Bovi, Luca Telesca and **Roberto Navigli**
Department of Computer Science
Sapienza University of Rome
{dellibovi,navigli}@di.uniroma1.it
luca.telesca@gmail.com

Abstract

We present DEFIE, an approach to large-scale Information Extraction (IE) based on a syntactic-semantic analysis of textual definitions. Given a large corpus of definitions we leverage syntactic dependencies to reduce data sparsity, then disambiguate the arguments and content words of the relation strings, and finally exploit the resulting information to organize the acquired relations hierarchically. The output of DEFIE is a high-quality knowledge base consisting of several million automatically acquired semantic relations.[1]

1 Introduction

The problem of knowledge acquisition lies at the core of Natural Language Processing. Recent years have witnessed the massive exploitation of collaborative, semi-structured information as the ideal middle ground between high-quality, fully-structured resources and the larger amount of cheaper (but noisy) unstructured text (Hovy et al., 2013). Collaborative projects, like Freebase (Bollacker et al., 2008) and Wikidata (Vrandečić, 2012), have been being developed for many years and are continuously being improved. A great deal of research also focuses on enriching available semi-structured resources, most notably Wikipedia, thereby creating taxonomies (Ponzetto and Strube, 2011; Flati et al., 2014), ontologies (Mahdisoltani et al., 2015) and semantic networks (Navigli and Ponzetto, 2012; Nastase and Strube, 2013). These solutions, however, are inherently constrained to small and often prespecified sets of relations. A more radical approach is adopted in systems like TEXTRUNNER (Etzioni et al., 2008) and REVERB (Fader et al., 2011), which developed from the Open Information Extraction (OIE) paradigm (Etzioni et al., 2008) and focused on the unconstrained extraction of a large number of relations from massive unstructured corpora. Ultimately, all these endeavors were geared towards addressing the knowledge acquisition problem and tackling long-standing challenges in the field, such as Machine Reading (Mitchell, 2005).

While earlier OIE approaches relied mostly on dependencies at the level of surface text (Etzioni et al., 2008; Fader et al., 2011), more recent work has focused on deeper language understanding at the level of both syntax and semantics (Nakashole et al., 2012; Moro and Navigli, 2013) and tackled challenging linguistic phenomena like synonymy and polysemy. However, these issues have not yet been addressed in their entirety. Relation strings are still bound to surface text, lacking actual semantic content. Furthermore, most OIE systems do not have a clear and unified ontological structure and require additional processing steps, such as statistical inference mappings (Dutta et al., 2014), graph-based alignments of relational phrases (Grycner and Weikum, 2014), or knowledge base unification procedures (Delli Bovi et al., 2015), in order for their potential to be exploitable in real applications.

In DEFIE the key idea is to leverage the linguistic analysis of recent semantically-enhanced OIE techniques while moving from open text to smaller corpora of dense prescriptive knowledge. The aim is

[1] http://lcl.uniroma1.it/defie

Transactions of the Association for Computational Linguistics, vol. 3, pp. 529–543, 2015. Action Editor: Sebastian Riedel.
Submission batch: 5/2015; Revision batch: 8/2015; Published 10/2015.

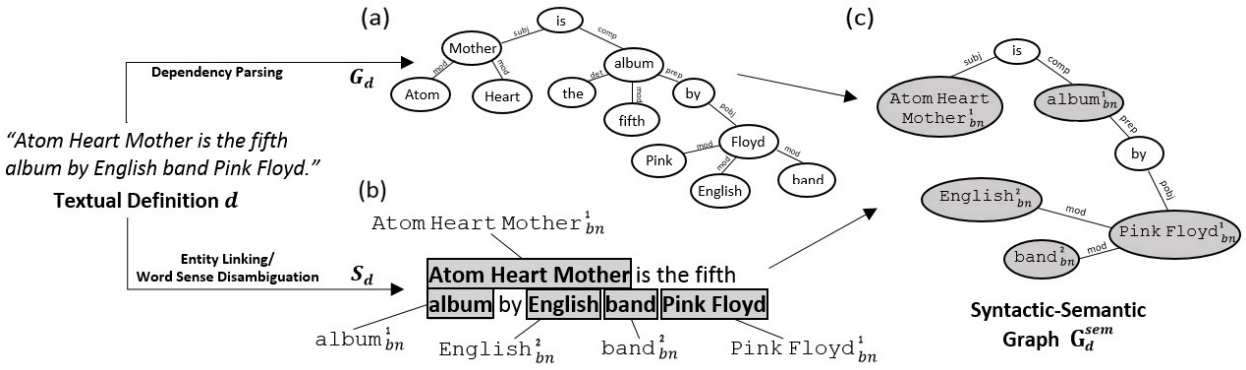

Figure 1: Syntactic-semantic graph construction from a textual definition

then to extract as much information as possible by unifying syntactic analysis and state-of-the-art disambiguation and entity linking. Using this strategy, from an input corpus of textual definitions (short and concise descriptions of a given concept or entity) we are able to harvest fully disambiguated relation instances on a large scale, and integrate them automatically into a high-quality taxonomy of semantic relations. As a result a large knowledge base is produced that shows competitive accuracy and coverage against state-of-the-art OIE systems based on much larger corpora. Our contributions can be summarized as follows:

- We propose an approach to IE that ties together syntactic dependencies and unified entity linking/word sense disambiguation, designed to discover semantic relations from a relatively small corpus of textual definitions;
- We create a large knowledge base of fully disambiguated relation instances, ranging over named entities and concepts from available resources like WordNet and Wikipedia;
- We exploit our semantified relation patterns to automatically build a rich, high-quality relation taxonomy, showing competitive results against state-of-the-art approaches.

Our approach comprises three stages. First, we extract from our input corpus an initial set of semantic relations (Section 2); each relation is then scored and augmented with semantic type signatures (Section 3); finally, the augmented relations are used to build a relation taxonomy (Section 4).

2 Relation Extraction

Here we describe the first stage of our approach, where a set of semantic relations is extracted from the input corpus. In the following, we refer to a *relation instance* as a triple $t = \langle a_i,\ r,\ a_j \rangle$ with a_i and a_j being the *arguments* and r the *relation pattern*. From each relation pattern r_k the associated *relation* R_k is identified by the set of all relation instances where $r = r_k$. In order to extract a large set of fully disambiguated relation instances we bring together syntactic and semantic analysis on a corpus of plain textual definitions. Each definition is first parsed and disambiguated (Figure 1a-b, Section 2.1); syntactic and semantic information is combined into a structured graph representation (Figure 1c, Section 2.2) and relation patterns are then extracted as shortest paths between concept pairs (Section 2.3).

The semantics of our relations draws on BabelNet (Navigli and Ponzetto, 2012), a wide-coverage multilingual semantic network obtained from the automatic integration of WordNet, Wikipedia and other resources. This choice is not mandatory; however, inasmuch as it is a superset of these resources, BabelNet brings together lexicographic and encyclopedic knowledge, enabling us to reach higher coverage while still being able to accommodate different disambiguation strategies. For each relation instance t extracted, both a_i, a_j and the content words appearing in r are linked to the BabelNet inventory. In the remainder of the paper we identify BabelNet concepts or entities using a subscript-superscript notation where, for instance, $\texttt{band}^i_{bn}$ refers to the i-th BabelNet sense for the English word *band*.

2.1 Textual Definition Processing

The first step of the process is the automatic extraction of syntactic information (typed dependencies) and semantic information (word senses and named entity mentions) from each textual definition. Each definition undergoes the following steps:

Syntactic Analysis. Each textual definition d is parsed to obtain a dependency graph G_d (Figure 1a). Parsing is carried out using C&C (Clark and Curran, 2007), a log-linear parser based on Combinatory Categorial Grammar (CCG). Although our algorithm seamlessly works with any syntactic formalism, CCG rules are especially suited to longer definitions and linguistic phenomena like coordinating conjunctions (Steedman, 2000).

Semantic Analysis. Semantic analysis is based on Babelfy (Moro et al., 2014), a joint, state-of-the-art approach to entity linking and word sense disambiguation. Given a lexicalized semantic network as underlying structure, Babelfy uses a dense subgraph algorithm to identify high-coherence semantic interpretations of words and multi-word expressions across an input text. We apply Babelfy to each definition d, obtaining a sense mapping S_d from surface text (words and entity mentions) to word senses and named entities (Figure 1b).

As a matter of fact, any disambiguation or entity linking strategy can be used at this stage. However, a knowledge-based unified approach like Babelfy is best suited to our setting, where context is limited and exploiting definitional knowledge as much as possible is key to attaining high-coverage results (as we show in Section 6.4).

2.2 Syntactic-Semantic Graph Construction

The information extracted by parsing and disambiguating a given definition d is unified into a *syntactic-semantic graph* G_d^{sem} where concepts and entities identified in d are arranged in a graph structure encoding their syntactic dependencies (Figure 1c). We start from the dependency graph G_d, as provided by the syntactic analysis of d in Section 2.1. Semantic information from the sense mappings S_d can be incorporated directly in the vertices of G_d by attaching available matches between words and senses to the corresponding vertices. Dependency graphs, however, encode dependencies solely on a word basis, while our sense mappings may include multi-word expressions (e.g. $\texttt{Pink Floyd}^1_{bn}$). In order to extract consistent information, subsets of vertices referring to the same concept or entity are merged to a single *semantic node*, which replaces the subgraph covered in the original dependency structure. Consider the example in Figure 1: an entity like $\texttt{Pink Floyd}^1_{bn}$ covers two distinct and connected vertices in the dependency graph G_d, one for the noun *Floyd* and one for its modifier *Pink*. In the actual semantics of the sentence, as encoded in G_d^{sem} (Figure 1c), these two vertices are merged to a single node referring to the entity $\texttt{Pink Floyd}^1_{bn}$ (the English rock band), instead of being assigned individual word interpretations.

Our procedure for building G_d^{sem} takes as input a typed dependency graph G_d and a sense mapping S_d, both extracted from a given definition d. G_d^{sem} is first populated with the vertices of G_d referring to disambiguated content words, merging those vertices covered by the same sense $s \in S_d$ into a single node (like $\texttt{Pink Floyd}^1_{bn}$ and $\texttt{Atom Heart Mother}^1_{bn}$ in Figure 1c). Then, the remaining vertices and edges are added as in G_d, discarding non-disambiguated adjuncts and modifiers (like *the* and *fifth* in Figure 1).

2.3 Relation Pattern Identification

At this stage, all the information in a given definition d has been extracted and encoded in the corresponding graph G_d^{sem} (Section 2.2). We now consider those paths connecting entity pairs across the graph and extract the relation pattern r between two entities and/or concepts as the shortest path between the two corresponding vertices in G_d^{sem}. This enables us to exclude less relevant information (typically carried by adjuncts or modifiers) and reduce data sparsity in the overall extraction process.

Our algorithm works as follows: given a textual definition d, we consider every pair of identified concepts or entities and compute the corresponding shortest path in G_d^{sem} using the Floyd-Warshall algorithm (Floyd, 1962). The only constraint we enforce is that resulting paths must include at least one verb node. This condition filters out meaningless single-node patterns (e.g. two concepts connected

Algorithm 1 Relation Extraction

procedure EXTRACTRELATIONSFROM(D)

1: $\mathbf{R} := \emptyset$
2: **for each** d **in** D **do**
3: $G_d := dependencyParse(d)$
4: $S_d := disambiguate(d)$
5: $G_d^{sem} := buildSemanticGraph(G_d, S_d)$
6: **for each** $\langle s_i, s_j \rangle$ **in** S_d **do**
7: $\langle s_i, r_{ij}, s_j \rangle := shortestPath(s_i, s_j)$
8: $\mathbf{R} := \mathbf{R} \cup \{\langle s_i, r_{ij}, s_j \rangle\}$
9: $filterPatterns(\mathbf{R}, \rho)$

return R;

with a preposition) and, given the prescriptive nature of d, is unlikely to discard semantically relevant attributes compacted in noun phrases. As an example, consider the two sentences "*Mutter is the third album by German band Rammstein"* and "*Atom Heart Mother is the fifth album by English band Pink Floyd"*. In both cases, two valid shortest-path patterns are extracted. The first extracted shortest-path pattern is:

$$\text{X} \rightarrow \texttt{is} \rightarrow \texttt{album}^1_{bn} \rightarrow \texttt{by} \rightarrow \text{Y}$$

with a_i=$\texttt{Mutter}^3_{bn}$, a_j=$\texttt{Rammstein}^1_{bn}$ for the first sentence and a_i=$\texttt{Atom Heart Mother}^1_{bn}$, a_j=$\texttt{Pink Floyd}^1_{bn}$ for the second one. The second extracted shortest-path pattern is:

$$\text{X} \rightarrow \texttt{is} \rightarrow \text{Y}$$

with a_i=$\texttt{Mutter}^3_{bn}$, a_j=$\texttt{album}^1_{bn}$ for the first sentence and a_i=$\texttt{Atom Heart Mother}^1_{bn}$, a_j=$\texttt{album}^1_{bn}$ for the second one. In fact, our extraction process seamlessly discovers general knowledge (e.g. that $\texttt{Mutter}^3_{bn}$ and $\texttt{Atom Heart Mother}^1_{bn}$ are instances of the concept $\texttt{album}^1_{bn}$) and facts (e.g. that the entities $\texttt{Rammstein}^1_{bn}$ and $\texttt{Pink Floyd}^1_{bn}$ have an `isAlbumBy` relation with the two recordings).

A pseudo-code for the entire extraction algorithm is shown in Algorithm 1: given a set of textual definitions D, a set of relations is generated over extractions $\mathbf{R}$, with each relation $R \subset \mathbf{R}$ comprising relation instances extracted from D. Each $d \in D$ is first parsed and disambiguated to produce a syntactic-semantic graph G_d^{sem} (Sections 2.1-2.2); then all the concept pairs $\langle s_i, s_j \rangle$ are examined to detect relation instances as shortest paths. Finally, we filter out from the resulting set all relations for which the number of extracted instances is below a fixed threshold ρ.[2] The overall algorithm extracts over 20 million relation instances in our experimental setup (Section 5) with almost 256,000 distinct relations.

3 Relation Type Signatures and Scoring

We further characterize the semantics of our relations by computing semantic type signatures for each $R \subset \mathbf{R}$, i.e. by attaching a proper semantic class to both its *domain* and *range* (the sets of arguments occurring on the left and right of the pattern). As every element in the domain and range of R is disambiguated, we retrieve the corresponding senses and collect their direct hypernyms. Then we select the hypernym covering the largest subset of arguments as the representative semantic class for the domain (or range) of R. We extract hypernyms using BabelNet, where taxonomic information covers both general concepts (from the WordNet taxonomy (Fellbaum, 1998)) and named entities (from the Wikipedia Bitaxonomy (Flati et al., 2014)).

From the distribution of direct hypernyms over domain and range arguments of R we estimate the quality of R and associate a confidence value with its relation pattern r. Intuitively we want to assign higher confidence to relations where the corresponding distributions have low entropy. For instance, if both sets have a single hypernym covering all arguments, then R arguably captures a well-defined semantic relation and should be assigned high confidence. For each relation R, we compute:

$$H_R = -\sum_{i=1}^{n} p(h_i)\, log_2\, p(h_i) \qquad (1)$$

where $h_i (i = 1, ..., n)$ are all the distinct argument hypernyms over the domain and range of R and probabilities $p(h_i)$ are estimated from the proportion of arguments covered in such sets. The lower H_R, the better semantic types of R are defined. As a matter of fact, however, some valid but over-general relations (e.g. *X is a Y*, *X is used for Y*) have inherently high values of H_R. To obtain a balanced score,

[2] In all the experiments of Section 6 we set $\rho = 10$.

Pattern	Score	Entropy
X *directed by* Y	4 025.80	1.74
X *known for* Y	2 590.70	3.65
X *is* election district$^1_{bn}$ *of* Y	110.49	0.83
X *is* composer$^1_{bn}$ *from* Y	39.92	2.08
X *is* street$^1_{bn}$ *named after* Y	1.91	2.24
X *is* village$^2_{bn}$ *founded in 1912 in* Y	0.91	0.18

Table 1: Examples of relation scores

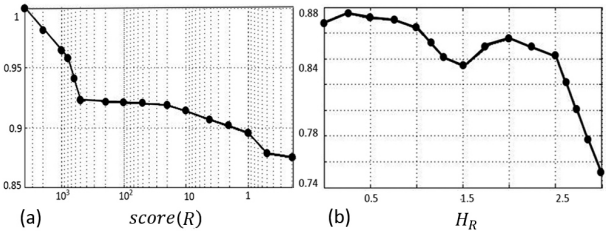

Figure 2: Precision against $score(R)$ (a) and H_R (b)

we therefore consider two additional factors, i.e. the number of extracted instances for R and the length of the associated pattern r, obtaining the following empirical measure:

$$score(R) = \frac{|S_R|}{(H_R + 1)\ length(r)} \qquad (2)$$

with S_R being the set of extracted relation instances for R. The +1 term accounts for cases where $H_R = 0$. As shown in the examples of Table 1, relations with rather general patterns (such as *X known for Y*) achieve higher scores compared to very specific ones (like *X is* village$^2_{bn}$ *founded in 1912 in Y*) despite higher entropy values. We validated our measure on the samples of Section 6.1, computing the overall precision for different score thresholds. The monotonic decrease of sample precision in Figure 2a shows that our measure captures the quality of extracted patterns better than H_R (Figure 2b).

4 Relation Taxonomization

In the last stage of our approach our set of extracted relations is arranged automatically in a relation taxonomy. The process is carried out by comparing relations pairwise, looking for hypernymy-hyponymy relationships between the corresponding relation patterns; we then build our taxonomy by connecting with an edge those relation pairs for which such a relationship is found. Both the relation

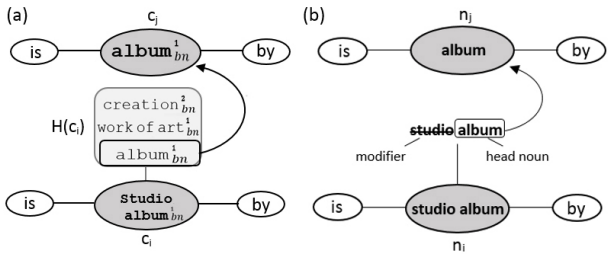

Figure 3: Hypernym (a) and substring (b) generalizations

taxonomization procedures described here examine noun nodes across each relation pattern r, and consider for taxonomization only those relations whose patterns are identical except for a single noun node.[3]

4.1 Hypernym Generalization

A direct way of identifying hypernym/hyponym noun nodes across relation patterns is to analyze the semantic information attached to them. Given two relation patterns r_i and r_j, differing only in respect of the noun nodes n_i and n_j, we first look at the associated concepts or entities, c_i and c_j, and retrieve the corresponding *hypernym sets*, $H(c_i)$ and $H(c_j)$. Hypernym sets are obtained by iteratively collecting the superclasses of c_i and c_j from the semantic network of BabelNet, up to a fixed height. For instance, given c_i = album$^1_{bn}$, $H(c_i)$ = {work of art$^1_{bn}$, creation$^2_{bn}$, artifact$^1_{bn}$}, and given c_j = Rammstein$^1_{bn}$, $H(c_j)$ = {band$^2_{bn}$, musical ensemble$^1_{bn}$, organization$^1_{bn}$}. Once we have $H(c_i)$ and $H(c_j)$, we just check whether $c_j \in H(c_i)$ or $c_i \in H(c_j)$ (Figure 3a). According to which is the case, we conclude that r_j is a generalization of r_i, or that r_i is a generalization of r_j.

4.2 Substring Generalization

The second procedure focuses on the noun (or compound) represented by the node. Given two relation patterns, r_i and r_j, we apply the following heuristic: from one of the two nouns, be it n_i, any adjunct or modifier is removed, retaining the sole head word $\hat{n}_i$. Then, $\hat{n}_i$ is compared with n_j and, if $\hat{n}_i = n_j$, we assume that the relation r_j is a generalization of r_i (Figure 3b).

[3]The simplifying assumption here is that two given relation patterns may be in a hypernymy-hyponymy relationship only when their plain syntactic structure is equivalent (e.g. *is* N_1 *by* and *is* N_2 *by*, with N_1 and N_2 being two distinct noun nodes).

	DEFIE	NELL	PATTY	REVERB	WISENET	Freebase	DBpedia
Distinct relations	255 881	298	1 631 531	664 746	245 935	1 894	1 368
Distinct relations (disambiguated)	240 618	-	-	-	-	-	-
Average extractions per relation	81.68	7 013.03	9.68	22.16	9.24	127 727.99	24 451.48
Distinct relation instances	20 352 903	2 089 883	15 802 946	14 728 268	2 271 807	241 897 882	33 449 631
Distinct concepts/entities involved	2 398 982	1 996 021	1 087 907	3 327 425	1 636 307	66 988 232	10 338 501

Table 2: Comparative statistics on the relation extraction process

5 Experimental Setup

Input. The input corpus used for the relation extraction procedure is the full set of English textual definitions in BabelNet 2.5 (Navigli and Ponzetto, 2012).[4] In fact, any set of textual definitions can be provided as input to DEFIE, ranging from existing dictionaries (like WordNet or Wiktionary) to the set of first sentences of Wikipedia articles.[5] As it is a merger for various different resources of this kind, BabelNet provides a large heterogeneous set comprising definitions from WordNet, Wikipedia, Wiktionary, Wikidata and OmegaWiki. To the best of our knowledge, this set constitutes the largest available corpus of definitional knowledge. We therefore worked on a total of 4,357,327 textual definitions from the English synsets of BabelNet's knowledge base. We then used the same version of BabelNet as the underlying semantic network structure for disambiguating with Babelfy.[6]

Statistics. Comparative statistics are shown in Table 2. DEFIE extracts 20,352,903 relation instances, out of which 13,753,133 feature a fully disambiguated pattern, yielding an average of 3.15 disambiguated relation instances extracted from each definition. After the extraction process, our knowledge base comprises 255,881 distinct semantic relations, 94% of which also have disambiguated content words in their patterns. DEFIE extracts a considerably larger amount of relation instances compared to similar approaches, despite the much smaller amount of text used. For example, we managed to harvest over 5 million relation instances more than PATTY, using a much smaller corpus (single sentences as opposed to full Wikipedia articles) and generating a number of distinct relations that was six times less than PATTY's. As a result, we obtained an average number of extractions that was substantially higher than those of our OIE competitors. This suggests that DEFIE is able to exploit the nature of textual definitions effectively and generalize over relation patterns. Furthermore, our semantic analysis captured 2,398,982 distinct arguments (either concept or named entities), outperforming almost all open-text systems examined.

Evaluation. All the evaluations carried out in Section 6 were based on manual assessment by two human judges, with an inter-annotator agreement, as measured by Cohen's kappa coefficient, above 70% in all cases. In these evaluations we compared DEFIE with the following OIE approaches:

- NELL (Carlson et al., 2010) with knowledge base beliefs updated to November 2014;
- PATTY (Nakashole et al., 2012) with Freebase types and pattern synsets from the English Wikipedia dump of June 2011;
- REVERB (Fader et al., 2011), using the set of normalized relation instances from the ClueWeb09 dataset;
- WISENET (Moro and Navigli, 2012; Moro and Navigli, 2013) with relational phrases from the English Wikipedia dump of December 2012.

In addition, we also compared our knowledge base with up-to-date human-contributed resources, namely Freebase (Bollacker et al., 2008) and DBpedia (Lehmann et al., 2014), both from the dumps of April/May 2014.

[4] babelnet.org

[5] According to the Wikipedia guidelines, an article should begin with a short declarative sentence, defining what (or who) is the subject and why it is notable.

[6] babelfy.org

	Top 100	Top 250	Rand 100	Rand 250
DEFIE	0.93 ± 0.01	0.91 ± 0.02	0.79 ± 0.02	0.81 ± 0.08
PATTY	0.93 ± 0.05	N/A	0.80 ± 0.08	N/A

Table 3: Precision of relation patterns

	NELL	**PATTY**	**REVERB**	**WISENET**	**Freebase**	**DBpedia**
Top 100	.571	.238	.214	.155	.571	.461
Rand 100	.942	.711	.596	.635	.904	.880

Table 4: Novelty of the extracted information

6 Experiments

6.1 Quality of Relations

We first assessed the quality and the semantic consistency of our relations using manual evaluation. We ranked our relations according to their score (Section 3) and then created two samples (of size 100 and 250 respectively) of the top scoring relations. In order to evaluate the long tail of less confident relations, we created another two samples of the same size with randomly extracted relations. We presented these samples to our human judges, accompanying each relation with a set of 50 argument pairs and the corresponding textual definitions from BabelNet. For each item in the sample we asked whether it represented a meaningful relation and whether the extracted argument pairs were consistent with this relation and the corresponding definitions. If the answer was positive, the relation was considered as correct. Finally we estimated the overall precision of the sample as the proportion of correct items. Results are reported in Table 3 and compared to those obtained by our closest competitor, PATTY, in the setting of Section 5. In PATTY the confidence of a given pattern was estimated from its statistical strength (Nakashole et al., 2012). As shown in Table 3, DEFIE achieved a comparable level of accuracy in every sample. An error analysis identified most errors as related to the vagueness of some short and general patterns, e.g. *X take Y*, *X make Y*. Others were related to parsing (e.g. in labeling the head word of complex noun phrases) or disambiguation. In addition, we used the same samples to estimate the novelty of the extracted information in comparison to currently available resources. We examined each correct relation pattern and looked manually for an equivalent relation in the knowledge bases

Gold Standard	**DEFIE**	**WISENET**	**PATTY**
163	131	129	126
	REVERB	**Freebase**	**DBpedia**
	122	69	39

Table 5: Coverage of semantic relations

of both our OIE competitors and human-contributed resources. For instance, given the relation *X born in Y*, NELL and REVERB have the equivalent relations `personborninlocation` and `is born in`, while Freebase and DBpedia have `Place of birth` and `birthPlace` respectively. We then computed the proportion of 'new' relations among those previously labeled as correct by our human judges. Results are shown in Table 4 for both the top 100 sample and the random sample. The high proportion of relations not appearing in existing resources (especially across the random samples) suggests that DEFIE is capable of discovering information not obtainable from available knowledge bases, including very specific relations (*X is blizzard in Y*, *X is Mayan language spoken by Y*, *X is government-owned corporation in Y*), as well as general but unusual ones (*X used by writer of Y*).

6.2 Coverage of Relations

To assess the coverage of DEFIE we first tested our extracted relations on a public dataset described in (Nakashole et al., 2012) and consisting of 163 semantic relations manually annotated from five Wikipedia pages about musicians. Following the line of previous works (Nakashole et al., 2012; Moro and Navigli, 2013), for each annotation we sought a relation in our knowledge base carrying the same semantics. Results are reported in Table 5. Consistently with the results in Table 4, the proportion of novel information places DEFIE in line with its closest competitors, achieving a coverage of 80.3% with respect to the gold standard. Examples of relations not covered by our competitors are `hasFatherInLaw` and `hasDaughterInLaw`. Furthermore, relations holding between entities and general concepts (e.g. `criticizedFor`, `praisedFor`, `sentencedTo`), are captured only by DEFIE and REVERB (which, however, lacks any argument semantics).

We also assessed the coverage of resources based

	Freebase	DBpedia	NELL
Random 100	83%	81%	89%

Table 6: Coverage of manually curated resources

	PATTY	WISENET
Random 100	66%	69%

Table 7: Coverage of individual relation instances

	Hyp. Gen.	Substr. Gen.	PATTY (Top)	PATTY (Rand)
Precision	0.87 ± 0.03	0.90 ± 0.02	0.85 ± 0.07	0.62 ± 0.09
# Edges	44 412		20 339	
Density	1.89×10^{-6}		7.64×10^{-9}	

Table 8: Precision and coverage of the relation taxonomy

on human-defined semantic relations: we extracted three random samples of 100 relations from Freebase, DBpedia and NELL and looked for semantically equivalent relations in our knowledge base. As shown in Table 6, DEFIE reports a coverage between 81% and 89% depending on the resource, failing to cover mostly relations that refer to numerical properties (e.g. `numberOfMembers`).

Finally, we tested the coverage of DEFIE over individual relation instances. We selected a random sample of 100 triples from the two closest competitors exploiting textual corpora, i.e. PATTY and WISENET. For each selected triple $\langle a_i, r, a_j \rangle$, we sought an equivalent relation instance in our knowledge base, i.e. one comprising a_i and a_j and a relation pattern expressing the same semantic relation of r. Results in Table 7 show a coverage greater than 65% over both samples. Given the dramatic reduction of corpus size and the high precision of the items extracted, these figures demonstrate that definitional knowledge is extremely valuable for relation extraction approaches. This might suggest that, even in large-scale OIE-based resources, a substantial amount of knowledge is likely to come from a rather smaller subset of definitional sentences within the source corpus.

6.3 Quality of Relation Taxonomization

We evaluated our relation taxonomy by manually assessing the accuracy of our taxonomization heuristics. Then we compared our results against PATTY, the only system among our closest competitors that generates a taxonomy of relations. The setting for this evaluation was the same of that of Section 6.1. However, as we lacked a confidence measure in this case, we just extracted a random sample of 200 hypernym edges for each generalization procedure. We presented these samples to our human judges and, for each hypernym edge, we asked whether the corresponding pair of relations represented a correct generalization. We then estimated the overall precision as the proportion of edges regarded as correct.

Results are reported in Table 8, along with PATTY's results in the setting of Section 5; as PATTY's edges are ranked by confidence, we considered both its top confident 100 subsumptions and a random sample of the same size. As shown in Table 8, DEFIE outperforms PATTY in terms of precision, and generates more than twice the number of edges overall. HARPY (Grycner and Weikum, 2014) enriches PATTY's taxonomy with 616,792 hypernym edges, but its alignment algorithm, in the setting of Section 5, also includes transitive edges and still yields a sparser taxonomy compared to ours, with a graph density of 2.32×10^{-7}. Generalization errors in our taxonomy are mostly related to disambiguation errors or flaws in the Wikipedia Bitaxonomy (e.g. the concept $\texttt{Titular Church}^1_{bn}$ marked as hyponym of $\texttt{Cardinal}^1_{bn}$).

6.4 Quality of Entity Linking and Disambiguation

We evaluated the disambiguation stage of DEFIE (Section 2.1) by comparing Babelfy against other state-of-the-art entity linking systems. In order to compare different disambiguation outputs we selected a random sample of 60,000 glosses from the input corpus of textual definitions (Section 5) and ran the relation extraction algorithm (Sections 2.1-2.3) using a different competitor in the disambiguation step each time. We eventually used the mappings in BabelNet to express each output using a common dictionary and sense inventory.

The coverage obtained by each competitor was assessed by looking at the number of distinct relations extracted in the process, the total number of relation instances extracted, the number of distinct concepts or entities involved, and the average number of semantic nodes within the relation patterns. For each competitor, we also assessed the precision obtained by evaluating the quality and semantic consistency of the relation patterns, in the same manner as in

	# Relations	# Triples	# Entities	Average Sem. Nodes
Babelfy	96 434	233 517	79 998	2.37
TagME 2.0	88 638	226 905	89 318	1.67
WAT	24 083	56 503	38 147	0.39
DBpedia Spotlight	67 377	140 711	38 254	1.45
Wikipedia Miner	39 547	88 777	37 036	0.96

Table 9: Coverage for different disambiguation systems

	Relations	Relation instances
Babelfy	82.3%	76.6%
TagME 2.0	76.0%	62.0%
WAT	84.6%	72.6%
DBpedia Spotlight	70.5%	62.6%
Wikipedia Miner	71.7%	56.0%

Table 10: Precision for different disambiguation systems

	# Definitions	Proportion (%)
Wikipedia	3 899 087	89.50
Wikidata	364 484	8.35
WordNet	41 356	0.95
Wiktionary	39 383	0.90
OmegaWiki	13 017	0.30

Table 11: Composition of the input corpus by source

	# Relations	# Relation instances	Avg. Extractions
Wikipedia	251 954	19 455 992	77.58
Wikidata	5 414	1 033 732	191.01
WordNet	2 260	128 200	56.73
Wiktionary	2 863	143 990	50.52
OmegaWiki	1 168	45 818	39.45

Table 12: Impact of each source on the extraction step

Section 6.1, both at the level of semantic relations (on the top 150 relation patterns) and at the level of individual relation instances (on a randomly extracted sample of 150 triples). Results are shown in Tables 9 and 10 for Babelfy and the following systems:

- TagME 2.0[7] (Ferragina and Scaiella, 2012), which links text fragments to Wikipedia based on measures like sense commonness and keyphraseness (Mihalcea and Csomai, 2007);
- WAT (Piccinno and Ferragina, 2014), an entity annotator that improves over TagME and features a re-designed spotting, disambiguation and pruning pipeline;
- DBpedia Spotlight[8] (Mendes et al., 2011), which annotates text documents with DBpedia URIs using scores such as prominence, topical relevance and contextual ambiguity;
- Wikipedia Miner[9] (Milne and Witten, 2013), which combines parallelized processing of Wikipedia dumps, relatedness measures and annotation features.

As shown in Table 9, Babelfy outperforms all its competitors in terms of coverage and, due to its unified word sense disambiguation and entity linking approach, extracts semantically richer patterns with 2.37 semantic nodes on the average per sentence. This reflects on the quality of semantic relations, reported in Table 10, with an overall increase of precision both in terms of relations and in terms of individual instances; even though WAT shows slightly higher precision over relations, its considerably lower coverage yields semantically poor patterns (0.39 semantic nodes on the average) and impacts on the overall quality of relations, where some ambiguity is necessarily retained. As an example, the pattern *X is station in Y*, extracted from WAT's disambiguation output, covers both railway stations and radio broadcasts. Babelfy produces, instead, two distinct relation patterns for each sense, tagging *station* as $\texttt{railway station}^{1}_{bn}$ for the former and $\texttt{station}^{5}_{bn}$ for the latter.

6.5 Impact of Definition Sources

We carried out an empirical analysis over the input corpus in our experimental setup, studying the impact of each source of textual definitions in isolation. In fact, as explained in Section 5, BabelNet's textual definitions come from various resources: WordNet, Wikipedia, Wikidata, Wiktionary and OmegaWiki. Table 11 shows the composition of the input corpus with respect to each of these definition sources. The distribution is rather skewed, with the vast majority of definitions coming from Wikipedia (almost 90% of the input corpus).

We ran the relation extraction algorithm (Sections 2.1-2.3) on each subset of the input corpus. As in previous experiments, we report the number of relation instances extracted, the number of distinct re-

[7] tagme.di.unipi.it
[8] spotlight.dbpedia.org
[9] wikipediadataminer.cms.waikato.ac.nz

	# Wikipages	# Sentences	# Extractions	Precision
All	14 072	225 867	39 684	61.8%
Top 100	10 334	161 769	13 687	59.0%

Table 13: Extraction results over non-definitional text

	# Relation instances	# Relations	# Edges
PATTY (definitions)	3 212 065	41 593	4 785
PATTY (Wikipedia)	15 802 946	1 631 531	20 339
Our system	20 807 732	255 881	44 412

Table 14: Performance of PATTY on definitional data

lations, and the average number of extractions for each relation. Results, as shown in Table 12, are consistent with the composition of the input corpus in Table 11: by relying solely on Wikipedia's first sentences, the extraction algorithm discovered 98% of all the distinct relations identified across the whole input corpus, and 93% of the total number of extracted instances. Wikidata provides more than 1 million extractions (5% of the total) but definitions are rather short and most of them (44.2%) generate only is-a relation instances. The remaining sources (WordNet, Wiktionary, OmegaWiki) account for less than 2% of the extractions.

6.6 Impact of the Approach vs. Impact of the Data

DEFIE's relation extraction algorithm is explicitly designed to target textual definitions. Hence, the result it achieves is due to the mutual contribution of two key features: an OIE approach and the use of definitional data. In order to decouple these two factors and study their respective impacts, we carried out two experiments: first we applied DEFIE to a sample of non-definitional text; then we applied our closest competitor, PATTY, on the same definition corpus described in Section 5.

Extraction from non-definitional text. We selected a random sample of Wikipedia pages from the English Wikipedia dump of October 2012. We processed each sentence as in Sections 2.1-2.2 and extracted instances of those relations produced by DEFIE in the original definitional setting (Section 5); we then automatically filtered out those instances where the arguments' hypernyms did not agree with the semantic types of the relation. We evaluated manually the quality of extractions on a sample of

Source	Label	Target
$\texttt{enzyme}^{1}_{bn}$	*catalyzes* $\texttt{reaction}^{1}_{bn}$ *of*	$\texttt{chemical}^{1}_{bn}$
$\texttt{album}^{1}_{bn}$	*recorded by*	$\texttt{rock group}^{1}_{bn}$
$\texttt{officier}^{1}_{bn}$	*commanded* $\texttt{brigade}^{1}_{bn}$ *of*	$\texttt{army unit}^{1}_{bn}$
$\texttt{bridge}^{1}_{bn}$	*crosses over*	$\texttt{river}^{1}_{bn}$
$\texttt{academic journal}^{1}_{bn}$	*covers* $\texttt{research}^{1}_{bn}$ *in*	$\texttt{science}^{1}_{bn}$
$\texttt{organization}^{1}_{bn}$	*has* $\texttt{headquarters}^{3}_{bn}$ in	$\texttt{city}^{1}_{bn}$

Table 15: Examples of augmented semantic edges

100 items (as in Section 6.1) for both the full set of extracted instances and for the subset of extractions from the top 100 scoring relations. Results are reported in Table 13: in both cases, precision figures show that extraction quality drops consistently in comparison to Section 6.1, suggesting that our extraction approach by itself is less accurate when moving to more complex sentences (with, e.g., subordinate clauses or coreferences).

PATTY on textual definitions. Since no open-source implementation of PATTY is available, we implemented a version of the algorithm which uses BABELFY for named entity disambiguation. We then ran it on our corpus of BabelNet definitions and compared the results against those originally obtained by PATTY (on the entire Wikipedia corpus) and those obtained by DEFIE. Figures are reported in Table 14 in terms of number of extracted relation instances, distinct relations and hypernym edges in the relation taxonomy. Results show that the dramatic reduction of corpus size affects the support sets of PATTY's relations, worsening both coverage and generalization capability.

6.7 Preliminary Study: Resource Enrichment

To further investigate the potential of our approach, we explored the application of DEFIE to the enrichment of existing resources. We focused on BabelNet as a case study. In BabelNet's semantic network, nodes representing concepts and entities are only connected via lexicograhic relationships from WordNet (hypernymy, meronymy, etc.) or unlabeled edges derived from Wikipedia hyperlinks. Our extraction algorithm has the potential to provide useful information to both augment unlabeled edges with labels and explicit semantic content, and create additional connections based on semantic relations. Examples are shown in Table 15.

	# Concept pairs	# Unlabeled	# Labeled
Type signatures	1 403	299	90
Relation instances	8 493 588	3 401 677	551 331

Table 16: Concept pairs and associated edges in BabelNet

We carried out a preliminary analysis over all disambiguated relations with at least 10 extracted instances. For each relation pattern r, we first examined the concept pairs associated with its type signatures and looked in BabelNet for an unlabeled edge connecting the pair. Then we examined the whole set of extracted relation instances in R and looked in BabelNet for an unlabeled edge connecting the arguments a_i and a_j. Results in Table 16 show that only 27.7% of the concept pairs representing relation type signatures are connected in BabelNet, and most of these connections are unlabeled. By the same token, more than 4 million distinct argument pairs (53.5%) do not share any edge in the semantic network and, among those that do, less than 14% have a labeled relationship. These proportions suggest that our relations provide a potential enrichment of the underlying knowledge base in terms of both connectivity and labeling of existing edges. In BabelNet, our case study, cross-resource mappings might also propagate this information across other knowledge bases and rephrase semantic relations in terms of, e.g., automatically generated Wikipedia hyperlinks.

7 Related Work

From the earliest days, OIE systems had to cope with the dimension and heterogeneity of huge unstructured sources of text. The first systems employed statistical techniques and relied heavily on information redundancy. Then, as soon as semi-structured resources came into play (Hovy et al., 2013), researchers started developing learning systems based on self-supervision (Wu and Weld, 2007) and distant supervision (Mintz et al., 2009; Krause et al., 2012). Crucial issues in distant supervision, like noisy training data, have been addressed in various ways: probabilistic graphical models (Riedel et al., 2010; Hoffmann et al., 2011), sophisticated multi-instance learning algorithms (Surdeanu et al., 2012), matrix factorization techniques (Riedel et al., 2013), labeled data infusion (Pershina et al., 2014) or crowd-based human computing (Kondreddi et al., 2014). A different strategy consists of moving from open text extraction to more constrained settings. For instance, the KNOWLEDGE VAULT (Dong et al., 2014) combines Web-scale extraction with prior knowledge from existing knowledge bases; BIPERPEDIA (Gupta et al., 2014) relies on schema-level attributes from the query stream in order to create an ontology of class-attribute pairs; RENOUN (Yahya et al., 2014) in turn exploits BIPERPEDIA to extract facts expressed as noun phrases. DEFIE focuses, instead, on smaller and denser corpora of prescriptive knowledge. Although early works, such as MindNet (Richardson et al., 1998), had already highlighted the potential of textual definitions for extracting reliable semantic information, no OIE approach to the best of our knowledge has exploited definitional data to extract and disambiguate a large knowledge base of semantic relations. The direction of most papers (especially in the recent OIE literature) seems rather the opposite, namely, to target Web-scale corpora. In contrast, we manage to extract a large amount of high-quality information by combining an OIE unsupervised approach with definitional data.

A deeper linguistic analysis constitutes the focus of many OIE approaches. Syntactic dependencies are used to construct general relation patterns (Nakashole et al., 2012), or to improve the quality of surface pattern realizations (Moro and Navigli, 2013). Phenomena like synonymy and polysemy have been addressed with kernel-based similarity measures and soft clustering techniques (Min et al., 2012; Moro and Navigli, 2013), or exploiting the semantic types of relation arguments (Nakashole et al., 2012; Moro and Navigli, 2012). An appropriate modeling of semantic types (e.g. selectional preferences) constitutes a line of research by itself, rooted in earlier works like (Resnik, 1996) and focused on either class-based (Clark and Weir, 2002), or similarity-based (Erk, 2007), approaches. However, these methods are used to model the semantics of verbs rather than arbitrary patterns. More recently some strategies based on topic modeling have been proposed, either to infer latent relation semantic types from OIE relations (Ritter et al., 2010), or to directly learn an ontological structure from a starting set of relation instances (Movshovitz-Attias and Cohen, 2015). However, the knowledge generated is often hard to interpret and integrate with existing

knowledge bases without human intervention (Ritter et al., 2010). In this respect, the semantic predicates proposed by Flati and Navigli (2013) seem to be more promising.

A novelty in our approach is that issues like polysemy and synonymy are explicitly addressed with a unified entity linking and disambiguation algorithm. By incorporating explicit semantic content in our relation patterns, not only do we make relations less ambiguous, but we also abstract away from specific lexicalizations of the content words and merge together many patterns conveying the same semantics. Rather than using plain dependencies we also inject explicit semantic content into the dependency graph to generate a unified syntactic-semantic representation. Previous works (Moro et al., 2013) used similar semantic graph representations to produce filtering rules for relation extraction, but they required a starting set of relation patterns and did not exploit syntactic information. A joint approach of syntactic-semantic analysis of text was used in works such as (Lao et al., 2012), but they addressed a substantially different task (inference for knowledge base completion) and assumed a radically different setting, with a predefined starting set of semantic relations from a given knowledge base. As we enforce an OIE approach, we do not have such requirements and directly process the input text via parsing and disambiguation. This enables DEFIE to generate relations already integrated with resources like WordNet and Wikipedia, without additional alignment steps (Grycner and Weikum, 2014), or semantic type propagations (Lin et al., 2012). As shown in Section 6.3, explicit semantic content within relation patterns underpins a rich and high-quality relation taxonomy, whereas generalization in (Nakashole et al., 2012) is limited to support set inclusion and leads to sparser and less accurate results.

8 Conclusion and Future Work

We presented DEFIE, an approach to OIE that, thanks to a novel unified syntactic-semantic analysis of text, harvests instances of semantic relations from a corpus of textual definitions. DEFIE extracts knowledge on a large scale, reducing data sparsity and disambiguating both arguments and relation patterns at the same time. Unlike previous semantically-enhanced approaches, mostly relying on the semantics of argument types, DEFIE is able to semantify relation phrases as well, by providing explicit links to the underlying knowledge base. We leveraged an input corpus of 4.3 million definitions and extracted over 20 million relation instances, with more than 250,000 distinct relations and almost 2.4 million concepts and entities involved. From these relations we automatically constructed a high-quality relation taxonomy by exploiting the explicit semantic content of the relation patterns. In the resulting knowledge base concepts and entities are linked to existing resources, such as WordNet and Wikipedia, via the BabelNet semantic network. We evaluated DEFIE in terms of precision, coverage, novelty of information in comparison to existing resources and quality of disambiguation, and we compared our relation taxonomy against state-of-the-art systems obtaining highly competitive results.

A key feature of our approach is its deep syntactic-semantic analysis targeted to textual definitions. In contrast to our competitors, where syntactic constraints are necessary in order to keep precision high when dealing with noisy data, DEFIE shows comparable (or greater) performances by exploiting a dense, noise-free definitional setting. DEFIE generates a large knowledge base, in line with collaboratively-built resources and state-of-the-art OIE systems, but uses a much smaller amount of input data: our corpus of definitions comprises less than 83 million tokens overall, while other OIE systems exploit massive corpora like Wikipedia (typically more than 1.5 billion tokens), ClueWeb (more than 33 billion tokens), or the Web itself. Furthermore, our semantic analysis based on Babelfy enables the discovery of semantic connections between both general concepts and named entities, with the potential to enrich existing structured and semi-structured resources, as we showed in a preliminary study on BabelNet (cf. Section 6.7).

As the next step, we plan to apply DEFIE to open text and integrate it with definition extraction and automatic gloss finding algorithms (Navigli and Velardi, 2010; Dalvi et al., 2015). Also, by further exploiting the underlying knowledge base, inference and learning techniques (Lao et al., 2012; Wang et al., 2015) can be applied to complement our model, generating new triples or correcting wrong ones. Fi-

nally, another future perspective is to leverage the increasingly large variety of multilingual resources, like BabelNet, and move towards the modeling of language-independent relations.

Acknowledgments

The authors gratefully acknowledge the support of the ERC Starting Grant MultiJEDI No. 259234.


This research was also partially supported by Google through a Faculty Research Award granted in July 2012.

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge. In *Proceedings of SIGMOD*, pages 1247–1250.

Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Toward an Architecture for Never-Ending Language Learning. In *Proceedings of AAAI*, pages 1306–1313.

Stephen Clark and James R. Curran. 2007. Wide-coverage Efficient Statistical Parsing with CCG and Log-Linear Models. *Computational Linguistics*, 33(4):493–552.

Stephen Clark and David Weir. 2002. Class-Based Probability Estimation Using a Semantic Hierarchy. *Computational Linguistics*, 28(2):187–206.

Bhavana Dalvi, Einat Minkov, Partha P. Talukdar, and William W. Cohen. 2015. Automatic Gloss Finding for a Knowledge Base using Ontological Constraints. In *Proceedings of WSDM*, pages 369–378.

Claudio Delli Bovi, Luis Espinosa Anke, and Roberto Navigli. 2015. Knowledge Base Unification via Sense Embeddings and Disambiguation. In *Proceedings of EMNLP*, pages 726–736.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge Vault: a Web-Scale Approach to Probabilistic Knowledge Fusion. In *Proceedings of KDD*, pages 601–610.

Arnab Dutta, Christian Meilicke, and Simone Paolo Ponzetto. 2014. A Probabilistic Approach for Integrating Heterogeneous Knowledge Sources. In *Proceedings of ESWC*, pages 286–301.

Katrin Erk. 2007. A Simple, Similarity-based Model for Selectional Preferences. In *Proceedings of ACL*, page 216–223.

Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld. 2008. Open Information Extraction from the Web. *Commun. ACM*, 51(12):68–74.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying Relations for Open Information Extraction. In *Proceedings of EMNLP*, pages 1535–1545.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.

Paolo Ferragina and Ugo Scaiella. 2012. Fast and Accurate Annotation of Short Texts with Wikipedia Pages. *IEEE Software*, 29(1):70–75.

Tiziano Flati and Roberto Navigli. 2013. SPred: Large-scale Harvesting of Semantic Predicates. In *Proceedings of ACL*, pages 1222–1232.

Tiziano Flati, Daniele Vannella, Tommaso Pasini, and Roberto Navigli. 2014. Two Is Bigger (and Better) Than One: the Wikipedia Bitaxonomy Project. In *Proceedings of ACL*, pages 945–955.

Robert W. Floyd. 1962. Algorithm 97: Shortest Path. *Communications of the ACM*, 5(6):345–345.

Adam Grycner and Gerhard Weikum. 2014. HARPY: Hypernyms and Alignment of Relational Paraphrases. In *Proceedings of COLING*, pages 2195–2204.

Rahul Gupta, Alon Halevy, Xuezhi Wang, Steven Euijong Whang, and Fei Wu. 2014. Biperpedia: An Ontology for Search Applications. In *Proceedings of VLDB*, pages 505–516.

Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S. Weld. 2011. Knowledge-based Weak Supervision for Information Extraction of Overlapping Relations. In *Proceedings of NAACL HLT*, pages 541–540.

Eduard Hovy, Roberto Navigli, and Simone Paolo Ponzetto. 2013. Collaboratively built semi-structured content and Artificial Intelligence: The story so far. *Artificial Intelligence*, 194:2–27.

Sarath Kumar Kondreddi, Peter Triantafillou, and Gerhard Weikum. 2014. Combining Information Extraction and Human Computing for Crowdsourced Knowledge Acquisition. In *Proceedings of ICDE*, pages 988–999.

Sebastian Krause, Hong Li, Hans Uszkoreit, and Feiyu Xu. 2012. Large-Scale Learning of Relation-Extraction Rules with Distant Supervision from the Web. In *Proceedings of ISWC*.

Ni Lao, Amarnag Subramanya, Fernando Pereira, and William W. Cohen. 2012. Reading the Web with Learned Syntactic-Semantic Inference Rules. In *Proceedings of EMNLP-CoNLL*, pages 1017–1026.

Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef,

Sören Auer, and Christian Bizer. 2014. DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web Journal*, pages 1–29.

Thomas Lin, Mausam, and Oren Etzioni. 2012. No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities. In *Proceedings of EMNLP-CoNLL*, pages 893–903.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian M. Suchanek. 2015. YAGO3: A Knowledge Base from Multilingual Wikipedias. In *CIDR*.

Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBPedia Spotlight: Shedding Light on the Web of Documents. In *Proceedings of I-Semantics*, pages 1–8.

Rada Mihalcea and Andras Csomai. 2007. Wikify!: Linking Documents to Encyclopedic Knowledge. In *Proceedings of CIKM*, pages 233–242.

David Milne and Ian H. Witten. 2013. An Open-Source Toolkit for Mining Wikipedia. *Artificial Intelligence*, 194:222–239.

Bonan Min, Shuming Shi, Ralph Grishman, and Chin-Yew Lin. 2012. Ensemble Semantics for Large-scale Unsupervised Relation Extraction. In *Proceedings of EMNLP-CoNLL*, pages 1027–1037.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant Supervision for Relation Extraction Without Labeled Data. In *Proceedings of ACL-IJCNLP*, pages 1003–1011.

Tom M. Mitchell. 2005. Reading the Web: A Breakthrough Goal for AI. *AI Magazine*.

Andrea Moro and Roberto Navigli. 2012. WiSeNet: Building a Wikipedia-based Semantic Network with Ontologized Relations. In *Proceedings of CIKM*, pages 1672–1676.

Andrea Moro and Roberto Navigli. 2013. Integrating Syntactic and Semantic Analysis into the Open Information Extraction Paradigm. In *Proceedings of IJCAI*, pages 2148–2154.

Andrea Moro, Hong Li, Sebastian Krause, Feiyu Xu, Roberto Navigli, and Hans Uszkoreit. 2013. Semantic Rule Filtering for Web-Scale Relation Extraction. In *Proceedings of ISWC*, pages 347–362.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *TACL*, 2:231–244.

Dana Movshovitz-Attias and William W. Cohen. 2015. KB-LDA: Jointly Learning a Knowledge Base of Hierarchy, Relations, and Facts. In *Proceedings of ACL*.

Ndapandula Nakashole, Gerhard Weikum, and Fabian M. Suchanek. 2012. PATTY: A Taxonomy of Relational Patterns with Semantic Types. In *Proceedings of EMNLP-CoNLL*, pages 1135–1145.

Vivi Nastase and Michael Strube. 2013. Transforming Wikipedia into a Large Scale Multilingual Concept Network. *Artificial Intelligence*, 194:62–85.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. *Artificial Intelligence*, 193:217–250.

Roberto Navigli and Paola Velardi. 2010. Learning Word-class Lattices for Definition and Hypernym Extraction. In *Proceedings of ACL*, pages 1318–1327.

Maria Pershina, Bonan Min, Wei Xu, and Ralph Grishman. 2014. Infusion of Labeled Data into Distant Supervision for Relation Extraction. In *Proceedings of ACL*, pages 732–738.

Francesco Piccinno and Paolo Ferragina. 2014. From TagME to WAT: a New Entity Annotator. In *Proceedings of ERD*, pages 55–62.

Simone Paolo Ponzetto and Michael Strube. 2011. Taxonomy Induction Based on a Collaboratively Built Knowledge Repository. *Artificial Intelligence*, 175(9-10):1737–1756.

Philip Resnik. 1996. Selectional Constraints: An Information-Theoretic Model and its Computational Realization. *Cognition*, 61(1-2):127–159.

Stephen D. Richardson, William B. Dolan, and Lucy Vanderwende. 1998. MindNet: Acquiring and Structuring Semantic Information from Text. In *Proceedings of ACL*, pages 1098–1102.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling Relations and Their Mentions without Labeled Text. In *Proceedings of ECML-PKDD*, pages 148–163.

Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M. Marlin. 2013. Relation Extraction with Matrix Factorization and Universal Schemas. In *Proceedings of NAACL HLT*, pages 74–84.

Alan Ritter, Mausam, and Oren Etzioni. 2010. A Latent Dirichlet Allocation Method for Selectional Preferences. In *Proceedings of ACL*, pages 424–434.

Mark Steedman. 2000. *The Syntactic Process*. MIT Press, Cambridge, MA, USA.

Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. 2012. Multi-instance Multi-label Learning for Relation Extraction. In *Proceedings of EMNLP-CoNLL*, pages 455–465.

Denny Vrandečić. 2012. Wikidata: A New Platform for Collaborative Data Collection. In *Proceedings of WWW*, pages 1063–1064.

William Yang Wang, Kathryn Mazaitis, Ni Lao, Tom M. Mitchell, and William W. Cohen. 2015. Efficient Inference and Learning in a Large Knowledge Base - Reasoning with Extracted Information using a Locally Groundable First-Order Probabilistic Logic. *Machine Learning*, 100(1):101–126.

Fei Wu and Daniel S. Weld. 2007. Autonomously Semantifying Wikipedia. In *Proceedings of CIKM*, pages 41–50.

Mohamed Yahya, Steven Euijong Whang, Rahul Gupta, and Alon Halevy. 2014. ReNoun: Fact Extraction for Nominal Attributes. In *Proceedings of EMNLP*, pages 325–335.