Towards Constructing Linguistic Ontologies:

Mapping framework and preliminary experimental analysis

Mamoun Abu Helou

University Of Milano Biccoca
`mamoun.abuhelou@disco.unimib.it`

Abstract. Cross-language ontology matching methods leverage translation based approach in mapping ontological resources (e.g., concepts) lexicalized in different languages. Such methods are largely dependent on the structural information encoded in the matched ontologies. The paper presents an approach for large scale cross-language linguistic ontology matching. In particular, it shows how synsets belonging to two distinct language wordnets can be mapped, by means of machine translation and frequency analysis on word senses. A preliminary experimental analysis of the problem is presented, through which we conduct experiments on existing gold-standard datasets. The results are encouraging and show the feasibility of the approach and demonstrate that the sense-selection task is a crucial step towards high quality mappings. We conclude with our observations and draw our potential future directions.

Keywords: ontology matching, linguistic ontology, wordnet, translation

1 Introduction

The last decades witnessed remarkable efforts in the development of ontologies that aim at capturing relationships between words and concepts aiming to represent a commonsense knowledge in natural languages, and hence making the semantic connections between words in different languages explicit, such ontologies are often called *wordnets* or *linguistic ontologies* [11][13]. One such commonly used linguistic ontology is the English WordNet [4].

The success of WordNet motivated the construction of similarly structured lexicons for individual and multiple languages (multi-language lexicons). These include, among others, EuroWordNet [27], MultiWordNet [19]. Recently, wordnets for many languages have been constructed under the guidelines of Global Word-Net Association. However, the *manual* method of constructing these ontologies is expensive and time-consuming. *Automatic* construction of wordnets is another method for building and linking wordnets. De Melo and Weikum [5] used bilingual dictionaries to automatically provide equivalents in various languages for the English WordNet synsets. However, translation tools might remove the language barrier but not necessarily the socio-cultural one [2]. The main challenge is to find the appropriate *word sense* of the translated word [17].

To enable semantic interoperability across different languages, ontology based

cross-language matching was explored in the last years [23]. However, the cultural-linguistic barriers [8][2] still need to be overcome in terms of the mapping process and techniques, as well as to formally define the semantic of mappings that align concepts lexicalized across different natural languages[1]. In addition, languages do not cover exactly the same part of the lexicon and, even where they have common concepts, these concepts are lexicalized differently [11]. One of the problems in creating linguistic ontologies via cross-language matching approach is that one needs to map an unstructured or a weakly structured lexicon, to a structured lexicon [13]. This introduce an extremely difficult and challenging matching problem for many reasons: (i) the lack of structural information, which is often used in matching techniques [23] (ii) the large mapping space (e.g., WordNet(3.0) has 117659 concepts), (iii) the quality (uncertainty) and coverage of the translation resources. We need to assume that all sense distinctions given by the translations are correct and available in a translation resource, and (iv) the words ambiguity, that is, we need to select the most commonly and accepted meaning (sense) of a word [17]. The difficulties of this task rise due to the words polysemy (if a word can convey more than one meaning) and synonymy (if two or more words can convey the same meaning).

The research presented here aims to contribute to the construction of large linguistic ontologies [12]. The idea is to semi-automate this process by first matching source language concepts (synsets) to an established wordnet concepts in a target language, and then deriving the semantic relations among the source concepts using relations among concepts in the target wordnet. We argue that, the resultant relations can provide an initial set of relations that can be manually validated and corrected [1]. Before selecting and/or extending the more appropriate existing cross-language mapping techniques, we need to be able to compare alternative methods and to assess the quality of their output.

Contribution.

In this paper I introduce a semi-automatic mapping framework that tries to map synsets in different languages by combining translation tools and word sense disambiguation (WSD) into a hybrid task. I define a mapping algorithm for constructing linguistic ontologies, through mapping unstructured concepts to structured one, as a maximization problem [26] that retrieve *top k* mappings from a set of sorted candidate mappings.

Since the creation of wordnets uses mappings among concepts expressed (lexicalized) in different languages, one of the research areas more relevant to the ontology creation problem is cross-language ontology matching (CLOM) . CLOM techniques [6] can play a crucial role in bootstrapping the creation of large linguistic ontologies and, for analogous reasons, in enriching existent ontologies. We also remark that the above considerations are general and can be reused for different languages. We demonstrate this by considering benchmark datasets for different pairs of languages. In particular, we discuss the results of investigating and assessing the mapping of an unstructured set of synsets in one language to an existent wordnet in different language. The proposed algorithm leverages on translation tools and tries to map synsets lexicalized in one language (e.g., Ara-

bic) to their correspondence synsets in other language (e.g., English), different translating settings were considered to investigate the appropriate translation methods for obtaining the correct translation. the algorithm ranks the translated synsets in order to select the most appropriate senses. This is followed by an experimental analysis and discussion. By this experiment we aim to respond on the following questions: (i) what is the best translation tool in term of coverage and correctness. (ii) what is the impact of the correct translation on the sense disambiguation task. (iii) what is the impact of providing (partial/semi)-structural knowledge among the source synsets. This paper will focus on the first two questions while leaving the third question for more investigation in the future work.

The rest of the paper is structured as follows. Section 2 overviews the related work. Section 3 illustrates the mapping algorithm. In section 4 the experiment, the evaluation settings, and a discussion about the obtained results are given. In section 5 we conclude and outline the future steps.

2 Background and State of the Art

The last decade witnessed a wide range of *ontology matching* methods which have been successfully developed and evaluated in OAEI [23]. The majority of the proposed techniques in these systems have mainly focused on mapping between ontological resources that are lexicalized in the same natural language (so-called, mono-language ontology matching, MOM). However, methods developed for MOM systems cannot directly access semantic information when ontologies are lexicalized in different natural languages [6]. There is a need for a method that automatically reconciles information when ontologies are lexicalized in different natural languages [8].

Manual mapping (by experts) was used to generate and review the mappings quality [14]. The mappings generated by such approaches are likely to be accurate and reliable. However, this can be a resource consuming process specially for maintaining large and complex ontologies. An *unsupervised* method was suggested based on (non-parallel) *bilingual corpora* [16].This approach, as it happens with most unsupervised learning methods, heavily relies on corpus statistics. De Melo and Weikum [5] constructed a binary classification learning problem to *automatically* determine the appropriate senses among the translated candidates. To create their classifier they used several scores that take into account structural properties as well as semantic relatedness and corpus frequency information where used. The authors claimed that, this technique is imperfect in terms of their quality and coverage of language-specific phenomena [5].

In general, to resolve the cross-lingual issue, a *translation based* approach [6] is considered in order to transform the CLOM problem into a MOM one. These systems deeply leverage the structural information derived from the mapped ontologies. Furthermore, Spohr et al. [25] approach, like all supervised learning methods, requires a significant number of labeled training samples and well-designed features to achieve good performance.

Another interesting work for resolving the cross-lingual issue exploits *Wikipedia*, a collaborative and multilingual resource of world and linguistic knowledge. Hassan and Mihalcea [9] developed a cross-lingual version of Explicit Semantic Analysis (ESA, [7]), *CL-ESA*. Similar works that used ESA in linking concepts across different languages are also presented in [3] and [15]. *WikiMatch*, a matching system presented in [10], searches Wikipedia pages (articles) titles for a given term (e.g., the ontology labels, and comments) and retrieve all language links describing the term, making use of the *inter-lingual* links between Wikipedia pages. However, such approaches are limited and highly dependent on the lexical coverages which are provided by Wikipedia inter-lingual links.

A notable approach for disambiguating and linking cross-lingual senses was presented in *BabelNet* [18]. Since Wikipedia inter-lingual links do not exist for all Wikipedia pages, Navigli and Ponzetto [18] proposed a *context-translation* approach. They automatically translate (using Google translator) a set of English sense-tagged sentences. After they applied the automatic translation, the most *frequent* translation is detected and included as a variant for the mapped senses in the given language. However, it is not clear if they employed any specific NLP techniques in this process, or if they aligned the translated words with words in the original (English) sentence (cf, the word aligner KNOWA [20]). Moreover, such frequency counts do not necessarily preserve the part of speech of the translated words. They only translated Wikipedia entries whose lemmas (page titles) do not refer to named entities [1]. They *contextless* translated lemmas in WordNet which are *monosemous* (i.e., words that have only one meaning), they simply included the translations returned by Google Translate. They [18] reported that the majority of the translated senses are monosemous.

BabelNet [18] mappings were evaluated against manually mapped random sample and gold-standard datasets. However, it is important to understand if the obtained mappings were achieved through the *context- or contextless-translation* approaches. More importantly, the monosemous and polysemy translated senses were not quantified in their evaluation, noticing that monosemous senses form a substantial large portion of the evaluated wordnets. It is important to measure both the contribution and the quality of the context-translation approach. For instance, the Italian wordnet contains about 66% monosemous senses, while in BabelNet they covered less than 53% and 74% of the Italian WordNet's [19] senses (words) and synsets, respectively. They determined the coverage as the percentage of gold-standard synsets that share a term (one synonym overlap) with the correspondence synset obtained from the mapping algorithm. This does not necessarily implies a high quality mappings. It worth noting that BabelNet context-translation approach covers about 25% of the Arabic WordNet's [22] words (which was used as benchmark).

---

[1] About 90% of Wikipedia pages are named entities which were included directly in BabelNet without translation [18]

Algorithm 1 Mapping Algorithm

1: **procedure**
2: **For each** $syn_i^{L_1} \in wn^{L_1}$
3:     $trans(syn_i^{L_1}, L_2) \longleftarrow \cup_{i=1}^{n} trans(w_i^{L_1}, L_2)$; $w_j^{L_1} \in syn_i^{L_1}$
4:     $candSenses(syn_i^{L_1}, L_2) \longleftarrow \cup_{i=1}^{n} sense(w_i^{L_2})$; $w_i^{L_2} \in trans(syn_i^{L_1}, L_2)$
5:     $rSenses \longleftarrow rank(\ candSenses(syn_i^{L_1}, L_2)\ )$
6:     $Map(syn_i^{L_1}) \longleftarrow select(rSenses, k)$; $k \in \{1, \dots, n\}$

---

3 Mapping Algorithm

Given a pair of wordnets wn^{L_1} and wn^{L_2} in different languages ($L_1 \neq L_2$), respectively called the source and target wordnet (we call it source wordnet although it has no relations among the synsets), the mapping algorithm tries to find for each synset $syn^{L_1} \in wn^{L_1}$ an equivalent synset from the target wordnet $syn^{L_2} \in wn^{L_2}$, such that it maximizes the probability of providing an appropriate correspondence to syn^{L_1}. In order to map the synsets, we make use of the mapping algorithm whose pseudocode is presented in *Algorithm* 1. The following steps are performed for each synset $syn^{L_1} \in wn^{L_1}$:

line 3: the translation function *trans* lookup all the possible translations in the target language for each word in the source synset, $w_i^{L_1} \in syn_i^{L_1}$.

line 4: the *sense* function ($sense(w^L) = \{syn_1^L, .., syn_n^L\}$) lookup all the candidate senses, *candSenses*, from the target wordnet for each translated word.

line 5: the *rank* function accepts the *candSenses* and returns a ranked senses (*rSenses*) ordered by the most appropriate senses, this is performed by counting the frequency of each sense in the *candSenses* set, the *rank* function also gives a higher priority (weight) to the synsets which obtained from translating different words in the source synset (i.e., synonym words give the same translation).

line 6: the *select* function selects the top k mappings (a set of mappings) from the (*rSenses*). For a single WSD mapping task $k = 1$. If a tie occurs, the senses will be selected based on the higher ratio between the number of translated words and the number of synonym words in the candidate senses (i.e., the ratio between the size of the mapped synsets), else it is randomly selected. As a result of executing the algorithm, a set of ranked mappings is returned.

4 Experiment

Wordnets. Three wordnets have been used in the experiment; the Arabic WordNet (2.0) [22], the Italian component of the MultiWordNet [19], and the English WordNet (3.0) [4]. The wordnets, respectively, have 11214, 34728 and 117659 synsets, such that each wordnet has 15964, 40178 and 155287 words, this forms 23481, 61558 and 206941 total word senses, respectively for each wordnet.

The Arabic WordNet and the Italian WordNet were used to benchmarked the mapping algorithm. Moreover, details on each wordnet are provided.

- The *Arabic WordNet (ArWN)* consists of 11214 Arabic synsets which were constructed from 15964 vocalized Arabic words (13867 unvocalized word; i.e.,

without diacritics). The ArWN authors [22] mapped the Arabic synsets to their equivalent synsets in the English WordNet (2.0). As a result 10413 synsets have been manually mapped to the correspondent English WordNet (EnWN) synsets. Out of the 10413 Arabic-English mapped synsets, 54 mappings do not have Arabic lexicon (Arabic concepts without lexicalization, such concepts were considered as *lexical gaps* [27]), and 10 synsets have no English correspondence synsets in EnWN (3.0) due to the part of speech mismatching. Overall, the resulted mappings are **10349** Arabic-English equivalent synsets mappings.
- The *Italian WordNet (ItWN)*[2] authors [19] have manually mapped the Italian synsets to their equivalent synsets in the EnWN (1.6). Later under the Open Multilingual Wordnet initiative the ItWN was mapped to EnWN (3.0). As a result we have 34728 Italian-English equivalent synsets mappings including 997 (*lexical gaps*), mappings do not have the correspondence Italian lexicon. Overall, the resulted mappings were **33731** Italian-English equivalent synsets mappings.

Evaluation. The goal is to disambiguate the senses and to find the appropriate mappings between synsets lexicalized in different languages. The disambiguation task was evaluated using evaluation measures borrowed from the information retrieval field [17]. The *coverage* is defined as the percentage of mappings in the test set for which the mapping method provides a sense mapping. The *precision* of the mapping method is computed as the percentage of the correct mappings given by the method, this reflects how good are the mappings which obtained by the assessed mapping method. The *recall* is defined as the ratio between the number of the correct mappings provided by the method being assessed, and the total number of mappings to be provided (number of mappings in the test set; the benchmark dataset). The $F-measure$ is also used, the F-measure is defined as the weighted harmonic mean of precision and recall.
In addition, a *Baseline* (lower bound) setting was used; the *first sense heuristic* was used to measure the lower bound of the experiment. Note that in general most of the WSD algorithms hardly can beat these bounds [17]. The upper bound (having the correct translations, *Oracle* translation) was also computed in order to specify the highest expected performance of the proposed approach.

Translation Methods. In the experiment we used different resource for translation. (i) *Machine translation tools*: Google translator was used to obtain the English translation for all the Arabic and Italian words in the benchmark dataset. (ii) *Machine readable dictionaries*: Sina dictionary was used for Arabic to English translations, Sina dictionary is a result of the ongoing Arabic Ontology project [12], the dictionary was constructed by integrating several specialized and general domain dictionaries. An Italian to English translation dictionary should be considered in the future work. (iii) *Oracle translation*: An oracle is a hypothetic system which is always supposed to know the correct answer (i.e., the correct translation). We used the translations provided in the benchmark wordnets as an oracle (correct translation). An oracle translation was used to demonstrate

---

[2] http://multiwordnet.fbk.eu/

Mapping	No.	Experiment	Mappings Provided	Correct Mappings provided	Coverage	Precision		Recall		F-measure	
						k=1	K=100	k=1	K=100	k=1	K=100
ArWN-to-EnWN	1	Google Dictionariy	9425	7889	90.59	27.53	83.70	24.94	75.83	26.18	79.57
	2	Sina & Google Dict.	9523	8340	91.53	28.45	87.58	26.04	80.16	27.19	83.71
	3	Oracle Translation	10349	10349	100	46.41	99.90	46.15	99.34	46.28	99.62
	4	Extended Oracle Trans.	10349	10349	100	58.22	99.87	57.93	99.38	58.07	99.62
	5	All Dictionaries	10349	10349	100	45.34	99.90	45.12	99.42	45.23	99.66
		Baseline (First Sense Heuristic)									
	6	Google Dictionariy				25.8	55.4	23.4	42.0	24.6	47.8
	7	Sina & Google Dict.				26.4	48.3	24.2	44.2	25.2	46.2
	8	Oracle Translation				45.8	60.8	46.2	60.7	46.0	61.0
	9	Extended Oracle Trans.				48.9	59.6	48.6	59.3	48.7	59.4
	10	All Dictionaries				43.0	60.4	42.7	60.0	42.9	60.2
ItWN-to-EnWN	11	Google Dictionariy	23707	20572	92.66	43.96	86.77	50.68	80.4	48.25	83.5
	12	Extended Oracle Trans.	33731	33731	100	61.16	100	61.16	100	61.16	100
		Baseline (First Sense Heuristic)									
	13	Google Dictionariy				42.0	66.0	48.2	61.5	45.1	63.8
	14	Extended Oracle Trans.				59.0	73.0	59.0	73.0	59.0	73.0

Fig. 1. Overall experiment results.

the upper bounds. Moreover, (iv) an *extended-oracle* translation was obtained for the Arabic-English translations by considering all the synonyms of the translated word, not only the translations provided in the ArWN. For the Italian words one can only obtain the extend-oracle translation from the ItWN dataset. Finally,(v) *all dictionaries*: the translations above were combined to investigate the different translation resources accuracy.

Results. The results are reported in Figure 1, the "Experiment" column specifies the translation method. The evaluation measures were reported, the reported measures evaluate (as %) if the equivalent mappings are among the *top k* mappings, $k \in [1, 100]$. The lower bounds (baseline) experiments were also reported. In Figure 2, four variants that exploit the structural information of the target wordnet were considered to select the equivalent mappings: (1)*isEquivalent* (isCorrect): the correct equivalent mappings appear among the *top k* candidate synsets. (2)*isHypernym*: the candidate synset is a hyponym of the correct mapping. (3)*hasHypernym* (or isHyponym): the hypernym of the candidate synset is an equivalent mapping. (4)*isSister*: the candidate synset is a sister node of a equivalent mapping (has the same hypernym synset). Figure 2, also shows the upper and lower bounds, and the precision of the obtained mappings using the Google translator for the ArWN synsets (Figure 1, experiment No.1).

Discussion. In experiment No.2, (*Sina & Google translation*) although it has a similar coverage to (experiment No.1) about 100 mappings more were covered-due to the correct translations by using Sina dictionary. From experiment No.3 (*Oracle translation*) we can notice that the missed mappings are only the lexical gaps synsets. From experiment No.4 (*Extended oracle translation*) the impor-
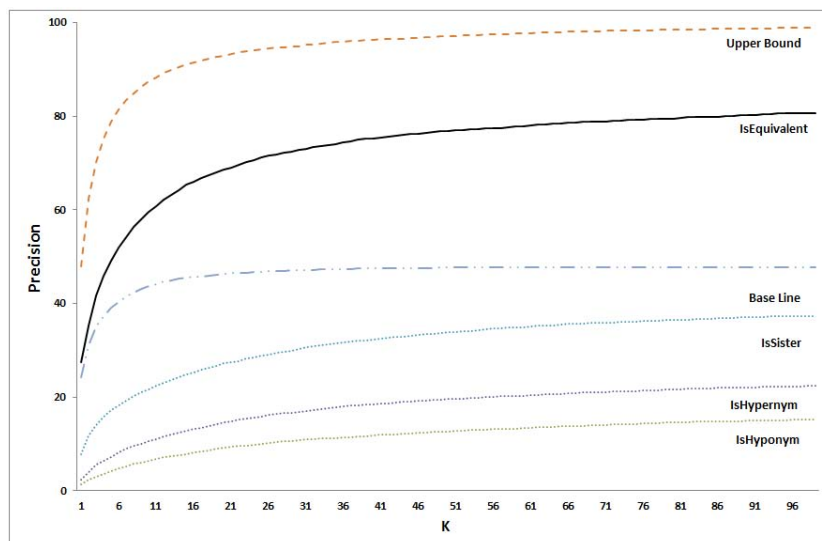
Fig. 2. Result of Using Google translation for mapping ArWN-to-EnWN.

tant observation is that as more accurate translation we have, the better we can rank and select the candidate senses, that is, the different between using the oracle translation and adding the synonyms words is obtaining the equivalent mappings with lower value of k.

Experiment No.5 (*All dictionaries*) highlights the fact that even we have the correct translations; the existence of inaccurate translations (by combining Google translations and the domain specific translations in Sina dictionary) introduced some noises and rise the ambiguity. This ranked the correct sense in lower positions, and increases the needed value of k to find the correct mappings. This means it is important to filter and discriminate the incorrect translations at first and then perform the ranking and selection steps. In the *Baseline* experiments all the performed experiments outperformed the lower bound settings.

Having the correct translation is an important step, still selecting the correct senses is a crucial task; One can notice from the results that although we have high (about 100%) coverage, still we need to consider a high value of k to obtain these correct mappings. In fact, the goodness and the better performance of the mapping algorithm is to provide the correct mappings while minimizing the value of k, and this depends on the *rank* function. Moreover, considering the *structural information* in the target wordnet (EnWN) improves the results. For instance, in experiment No.1 (Google dictionary) considering the neighbor nodes (e.g.,*isSister*) improves the results around 10% for small k values, $k \in [5, 20]$. For that, we believe that exploiting the structural information (e.g., graph based and the similarity based approaches [21]) is expected to help in ranking the correct senses in higher positions. Considering the candidate synsets *hypernyms*(*hasHypernym* plot, Figure2) is also expected to improve the results.

Error Analysis. The main reasons for missing some mappings can be divided into the following errors:

(a) Incorrect translation, number of mappings were missed due to the incorrect translation, this can be divided into 4 errors categories (i) compound words, (ii) named entity, (iii) translation (word) do not exist in WordNet, and (iv) bad translation, by bad translation we mean that Google translator only re-spell the source word in English alphabet. For instance, regarding the ArWN-EnWN mappings, out of the 988 missing mappings (about 10% of the mappings) 760 are compound (one) word synsets, and 33 mappings are named entity synsets. Moreover, 799 mappings do not have an equivalent EnWN lemma, and 189 mappings were considered as bad translations.

(b) Monosemous words, the ArWN has 4197 one word synsets, and it has 4370 monosemous synsets (all words in the synsets are monosemous). 2111 synsets are compound (one) word synsets. The majority of the compound one words (about 94%) which used in mapping the Arabic and the English synsets are monosemous synsets. Due to this fact, about 10% of the ArWN-EnWN mappings have only one candidate sense, and about 24% have up to 10 candidate senses. Regarding the ItWN-EnWN mappings, about 35% of the mappings have only one candidate sense, this high values is due to the fact that most of the ItWN synsets (about 66%) are monosemous synsets. About 74% synsets have up to 10 candidate senses. By studying the distribution of the rank values for the Google translation for both the ArWN and ItWN, about 85% of the candidate senses were equally ranked with one (the sense appear one time in the *candSenses*), thus the increment of k retrieves more correct equivalent mappings.

(c) Polysemous words, the ArWN has 3641 polysemous words, such that about 30% and 16% of the polysemous synsets are respectively, two and three words synsets. The ItWN has 10366 polysemous words, such that about 64% and 20% of the polysemous synsets are respectively, two and three words synsets. This highlights the fact that most of the benchmark polysemous sysnets are small size (number of words), this make it more difficult to distinguish the correct sense. For instance, in experiment No.4, when the synonym translations were considered (we increased the size of the translated synsets) the raking functions performs better than experiment No.3.

5   Conclusions & Future Works

This paper investigated how synsets from different languages can be mapped, especially the impact of translation tools and the selection of candidate synsets for mappings. A cross language mapping algorithm was presented, the algorithm tries to maximizes the probability that mappings with higher rank are considered correct mappings by users, based on the frequency of translated synsets and majority voting approaches. The experiments have demonstrated several outcomes that can summarized as follow:

Mamoun Abu Helou

1. the approach was successfully tested over two different pairs of languages, which demonstrates its adoptability across different languages.
2. using *structural information* encoded in the target wordnet improves the sense selection task.
3. combining the translations of Machine Translation (MT) tools with a bilingual dictionary translation improves the results (Figure 1, experiment No.2).
4. the proposed approach outperforms the baseline settings. The upper bounds indicate that there is a space for more improvements in terms of obtaining the correct translations and to better rank the candidate senses.

Moreover, features obtained from the MT tools (Google translation) such as the translation score and the synset translations need to be explored in order to filter the correct translations and to better rank the candidate senses. In addition, NLP techniques (e.g., stemming, headword extraction, ..etc) are expected to improve the MT coverage, and obtain more candidate senses (instead of using pure translation-lookup and word-sense exact matching).
Currently I am planning to consider other languages datasets, and to investigate the construction of partial structural source synsets and its impact on the mapping algorithm inspired by the work presented in [21]. Another interesting direction is to crowdsource the construction process by providing the workers with the *top k* mappings [24], at the same time this should simulates the majority of speakers agreement [1].

References

1. M. Abu Helou, M. Palmonari, M. Jarrar, Ch. Fellbaum. Towards Building Linguistic Ontology via Cross-Language Matching. GWN 2014
2. P. Cimiano, E.Montiel-Ponsoda, P. Buitelaar, M. Espinoza, A. Gmez-Prez. A note on ontology localization. Applied Ontology. 2010.
3. P. Cimiano, A. Schultz, S. Sizov, P. Sorg, S. Staab, Explicit vs. latent concept models for cross-language information retrieval, in: Proc. of the 21st IJCAI,2009.
4. Ch. Fellbaum. Wordnet: An electronic lexical database. MIT Press,1998.
5. G. De Melo and G. Weikum. Constructing and utilizing wordnets using statistical methods. Language Resources and Evaluation,2012.
6. B. Fu, R. Brennan, and D. O'Sullivan. A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes. J. Web Sem., 2012.
7. E. Gabrilovich, S. Markovitch, Computing semantic relatedness using Wikipedia-based explicit semantic analysis, in: Proceedings of the 20th IJCAI, India, 2007.
8. J. Gracia, E. Montiel-Ponsoda, P. Cimiano, A. Gmez-Prez, P. Buitelaar, J. McCrae, Challenges for the multilingual web of data, JWS, 2012.
9. S. Hassan and R. Mihalcea, Cross-lingual Relatedness using Encyclopedic Knowledge, to appear in Proc. EMNLP 2009.
10. S. Hertling and H. Paulheim. WikiMatch - Using Wikipedia for Ontology Matching. In Proceedings OM, 2012.

11. G. Hirst, Ontology and the Lexicon, in Handbook on Ontologies and Information Systems. eds. S. Staab and R. Studer. Heidelberg: Springer, 2004.
12. M. Jarrar, Building A Formal Arabic Ontology (invited paper), Proc. of the Experts Meeting On Arabic Ontologies And Semantic Networks. Alecso, Arab League. Tunis, July 26-28, 2011.
13. M. Jarrar, Lexical Semantics and Multilingualism. Lecture Notes, Sina Institute, Birzeit University, 2013.
14. A. Liang, M. Sini, Mapping AGROVOC & the Chinese Agricultural Thesaurus: Definitions, Tools Procedures. New Review of Hypermedia & Multimedia, 2006.
15. F. Narducci, M. Palmonari, and G. Semeraro. Cross-language semantic matching for discovering links to e-gov services in the LOD cloud . Know@LOD,ESWC, 2013.
16. G. Ngai, M. Carpuat, P. Fung, Identifying Concepts Across Languages: A First Step towards A Corpus-based Approach to Automatic Ontology Alignment. In: Proceedings of the 19th COLING, 2002.
17. R. Navigli. Word Sense Disambiguation: a Survey. ACM Computing Surveys, 2009.
18. R. Navigli and S. Ponzetto. BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. AI. 2012.
19. E. Pianta, L. Bentivogli, C. Girardi, MultiWordNet: Developing an aligned multilingual database, in: Proceedings of the 1st International GWC. 2002.
20. E. Pianta, L. Bentivogli, Knowledge Intensive Word Alignment with KNOWA. In Proceedings of COLING 2004.
21. M. T. Pilehvar and R. Navigli. A Robust Approach to Aligning Heterogeneous Lexical Resources. Proc. of the ACL 2014
22. H. Rodrguez, D. Farwell, J. Farreres, M. Bertran, M. Alkhalifa, M. Antonia Mart, W. Black, S. Elkateb, J. Kirk, A. Pease, P. Vossen, Ch. Fellbaum. Arabic WordNet: Current State and Future Extensions in: Proc. of the GWC 2008.
23. P. Shvaiko and J. Euzenat. Ontology matching: State of the art and future challenges. IEEE Trans. Knowl. Data Eng 2013.
24. C. Sarasua, E. Simperl, N.F. Noy: CrowdMAP: Crowdsourcing Ontology Alignment with Microtasks. In: Cudre-Mauroux, P., et al. (eds.) ISWC 2012, Part I. LNCS, Springer, Heidelberg (2012)
25. D. Spohr, L. Hollink, and Ph. Cimiano. A machine learning approach to multilingual and cross-lingual ontology matching. ISWC 2011.
26. Venetis, Petros and Garcia-Molina, Hector and Huang, Kerui and Polyzotis, Neoklis. Max Algorithms in Crowdsourcing Environments. Proc. WWW 2012.
27. P. Vossen. EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index. IJL 2004.