Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 137 (2018) 127–138

Procedia Computer Science

www.elsevier.com/locate/procedia

SEMANTiCS 2018 – 14th International Conference on Semantic Systems

Cross-Lingual Ontology Enrichment Based on Multi-Agent Architecture

Mohamed Ali[a], Said Fathalla[a,b,*], Shimaa Ibrahim[b,c], Mohamed Kholief[d], Yasser Hassan[a]

[a] *Faculty of Science, University of Alexandria, Egypt*
[b] *Smart Data Analytics (SDA), University of Bonn, Germany*
[c] *Institute of Graduate Studies and Research, University of Alexandria, Egypt*
[d] *College of Computing and Information Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt*

Abstract

The proliferation of ontologies and multilingual data available on the Web has motivated many researchers to contribute to multilingual and cross-lingual ontology enrichment. Cross-lingual ontology enrichment greatly facilitates ontology learning from multilingual text/ontologies in order to support collaborative ontology engineering process. This article proposes a cross-lingual ontology enrichment (CLOE) approach based on a multi-agent architecture in order to enrich ontologies from a multilingual text or ontology. This has several advantages: 1) an ontology is used to enrich another one, written in a different natural language, and 2) several ontologies could be enriched at the same time using a single chunk of text (*Simultaneous Ontology Enrichment*). A prototype for the proposed approach has been implemented in order to enrich several ontologies using English, Arabic and German text. Evaluation results are promising and showing that CLOE performs well in comparison with four state-of-the-art approaches.

Keywords: cross-lingual ontology enrichment, multi-agent, knowledge management, ontology learning.

1. Introduction

The next decade is likely to witness a considerable rise in ontology learning and enrichment due to the massive utilization of ontologies in different fields of science. In particular, ontologies are widely used in various areas of computer science such as information retrieval, text classification [13], scholarly communication [17, 15, 16] and bioinformatics [14, 12]. Getting into that, more and more ontological resources will become usable on the Semantic Web in different languages. Therefore, it is necessary to develop innovative algorithms which are capable of digesting multilingual information, as well as enriching an ontology using a well-formed one in a different natural language. Most of the existing ontologies have been developed and maintained with a human intervention which is extremely laborious, costly process and time-consuming.

* Corresponding author. Tel.: +4-9163-8260501.
E-mail address: fathalla@cs.uni-bonn.de

10.1016/j.procs.2018.09.013

Generally, the problem of ontology learning still needs a lot of work [3], in particular, multilingual and cross-lingual ontology learning. According to Gracia et al. [19], multilingual and cross-lingual ontology enrichment is accomplished by using cross-lingual ontology matching. Multilingual and cross-lingual ontology matching process is used for mapping and establishing relationships between ontological resources in different natural languages [26]. Most studies have only tended to focus on learning a specific language. Therefore, we can conclude that there is a big need to create a standard framework for learning and matching ontologies from different languages. This raises a key question: what would be a suitable approach to simultaneously enrich ontologies using multilingual text or another ontology written in a different language? This article proposes a cross-lingual ontology enrichment (CLOE) approach based on a multi-agent architecture in order to enrich ontologies from a multilingual text or ontology. Multi-agent systems have been widely employed in many domains due to the distributed processing offered, efficiency, maintainability and scalability. We introduce a new terminology, *Simultaneous Ontology Enrichment* (SOE), which refers to enriching several ontologies simultaneously. We propose two novel algorithms, for simultaneous ontology enrichment and agents communication. The most prominent feature of the proposed approach is that agents could learn from each other, using a predefined communication scheme, to get the benefit of already-learned concepts found in the input ontologies. In addition, the enrichment of an ontology is performed using another ontology or text in different languages. A prototype for the proposed approach has been implemented in order to enrich several ontologies using English, Arabic and German text. We choose such three languages because they are much familiar to the authors. However, concepts and relationships extraction tasks become more challenging when it concerns the Arabic language. In Arabic, the complexity of spelling, morphology, and semantics makes pre-processing and extraction tasks quite difficult compared to Latin languages [8].

The remainder of this article is structured as follows: we present an overview of related work in Section 2. The proposed approach is described in Section 3. A case study is presented in Section 4. Prototype implementation and preliminary evaluation are presented in Section 5. Finally, we conclude with an outline of the future research in Section 6.

2. Related Work

A recent review of the literature on ontology learning found that most studies tended to focus on learning ontologies from monolingual data sources (e.g., French [18], Arabic [6], German [20], and English [29]). Cross-lingual ontology mapping approaches are used for the automatic construction and enrichment of multilingual or large lexical ontologies [21, 10]. Since more than a million datasets have been published online as linked open data (LOD) from 43 countries in 24 different natural languages [22], cross-lingual ontology mapping becomes a challenging area in the field of ontology learning. Yang and Callan [30] presented a technique to extract concepts from a corpus of public comments. Benabdallah et al. [6] proposed an approach for automatic ontology construction from Arabic corpus by using statistical method for extracting simple and complex terms. In addition, they link these terms by semantic relations using automatic learning of lexical and syntactic markers. Albukhitan and Helmy [1] proposed an ontology learning system from unstructured Arabic text based on NLP tools. Concepts are extracted based on several statistical and data mining based algorithms. Beseiso et al. [7] proposed a multilingual ontology learning algorithm from email. Albukhitan and Helmy [2] presented an interesting automatic ontology maintenance system using multi-agent based approach for multilingual ontologies for two domains: food and health. The maintenance tasks are distributed among a set of agents. Furthermore, agents are used for finding and mapping ontologies. Validating the extracted concepts is performed using the explicit semantic analysis. Rafi et al. [25] proposed a multi-agent based approach towards supporting ontology maintenance involving ontology enrichment and integration. One of the limitations of their approach is that it does not support the maintenance of the ontologies. A good description of the available literature in the field of multilingual ontologies is provided in [27]. A key problem with much of the literature on ontology enrichment is focusing on learning ontologies from monolingual data sources. We address this limitation in our proposed approach by supporting the process of enriching ontologies using multilingual text.

3. Cross-Lingual Ontology Enrichment (CLOE)

The proposed approach comprises three phases: pre-processing and concept extraction, candidate sentence selection, and multi-agent based enrichment. The input is the multilingual text and ontologies to be enriched, and the output is the enriched ontologies. At the end, experts validate the extracted concepts and relations before ingestion to the ontology. The main phases of the proposed architecture are (see Figure 1):

3.1. Pre-processing and Concept Extraction

Text pre-processing is the process of cleaning and preparing the text for further processing. The main aim of this phase is to process input text and make it ready for the next steps, by employing a variety of NLP techniques, *OpenNLP*[1] packages are used. Due to the multilingualism of the input text, it might contain text written in two or more natural languages, so it is important to detect and separate text from different languages.

Language identification: plays a key role in several NLP applications. Depending on the language of the text, appropriate techniques are used for stop words and unnecessary word removal, part-of-speech (POS) tagging, and text annotation. For instance, given an input text T=" مشغل الاقراص لا يعمل your DVD drive may not be found" , T will be divided into two sentences T1=" مشغل الاقراص لا يعمل " and T2= "your DVD drive may not be found" with labels $L_1 = Ar$ and $L_2 = En$ of the identified languages respectively. Various approaches have been made to master this task [24]. *Apache Tika*[2] package is used for language identification.

Sentence splitter: splits the text into a list of sentences. It is worth to mention that we keep each sentence as a block of text in the further preprocessing tasks.

Tokenization: divides each sentence to a set of tokens. Tokens separated by delimiters such as white-space characters.

Stop words and unnecessary words filter:, remove words with high frequencies of occurrence, but have no contribution to the subject of text such as pronouns, prepositions, conjunctions, and unnecessary words such as "very", "really", and non-alphanumeric contents.

POS tagger, analyzes the text to find out phrase structures and group words according to their syntactic and semantic property. There are nine default tags: Adjective (ADJ), Adverb (ADV), Conjunction (CONJ), Determiner (DET), Noun (N), Number (NO), Preposition (PREP), Pronoun (PRO), and Verb (V). Token verbs are identified to represent a relation. At this stage, the identified verbs in each sentence are counted as a two-argument relation with nouns in the same sentence.

3.2. Candidate Sentence Selection

In this phase, we filter pre-processed sentences in order to eliminate unnecessary/useless sentences. Iterate on sentences to count the number of nouns and noun phrases in a sentence. Then, candidate sentences selected according to the following rules:

- *Rule 1:* Single-concept sentences are ignored, for instance *"Cydia has been installed successfully"*. This sentence has only *"Cydia"* as a noun and does not has any useful information to be kept, therefore it will be ignored.
- *Rule 2:* If the sentence has two POS or more as nouns or noun phrases, then keep the sentence, otherwise discard it. For example *"Cydia has been installed successfully on the iOS device"*. This sentence has two nouns: *"Cydia"* and *"iOS device"*, therefore it will be kept. This is required to extract the relations between concepts.

[1] https://opennlp.apache.org/docs/1.8.2/apidocs/opennlp-tools/overview-summary.html
[2] https://tika.apache.org/1.17/api/

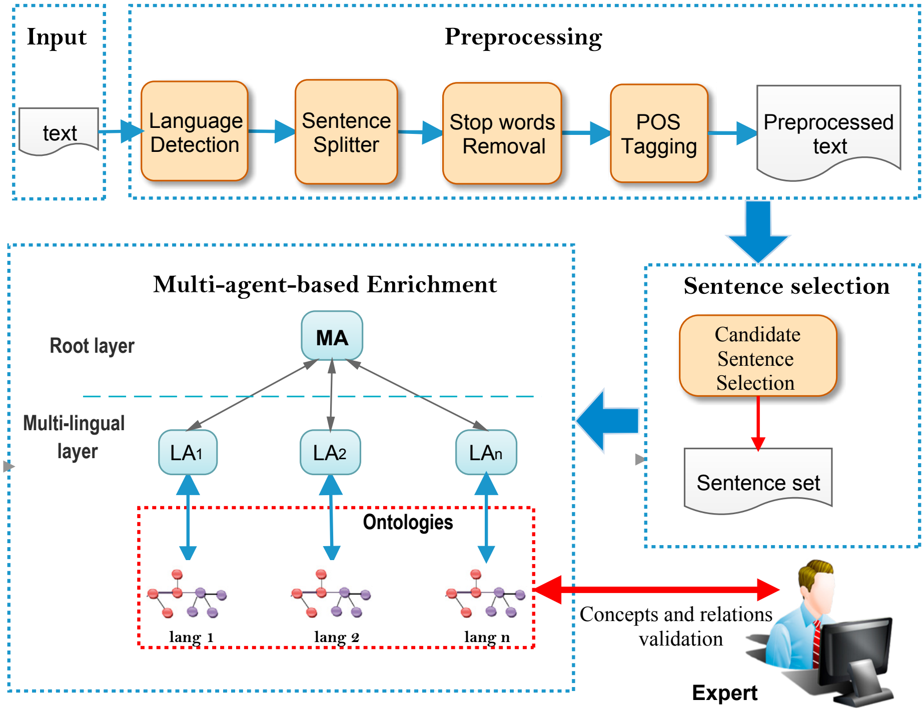

Figure 1: CLOE architecture

3.3. *Multi-agent based enrichment*

We proposed an innovative enrichment algorithm that allows a set of intelligent agents to interact with each other to learn about their ontologies. There are two types of agents: Master Agent (MA) and Language Agent (LA). Two-level hierarchical multi-agent model has been built where each node is modeled as an agent. The language agents LA_i are at the lowest hierarchy supervised by a single master agent MA as shown in Figure 1. In other words, language agents have the task of learning concepts, relations, and instances for a particular ontology language. For instance, the master agent MA is responsible for managing a set of n language agents $(LA_{i,i=1,2,\ldots n})$ The number of language agents is the number of languages of the ontologies.

3.3.1. *Agents Registration*

At birth, LAs register themselves at MA by sending the language of the associated ontology, then MA creates the look-up table LT and gives each one of them a unique id. The look-up table is a 2-dimensional table consists of tuples of $\prec id, Language \succ$ which is used by MA to decide which LA is responsible for learning the contents of each incoming text based on the language of the text. For each incoming text, MA takes: candidate sentences S, text language L, and look-up table LT as input from the previous phase and then selects the appropriate LA based on L (calling `FindLA`). A sketch of the main routine is reported in lgorithm 1 assuming the following assumptions:

- *Assumption 1:* candidate sentences are vectors of $\prec word, POS \succ$ pairs,
- *Assumption 2:* input text might have more than one language,
- *Assumption 3:* LAs are registred at MA,
- *Assumption 4:* each LA is responsible for learning concepts and relations between those concepts (taxonomic and non-taxonomic) for a specific language.

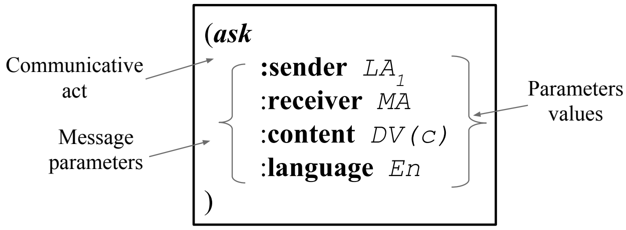

Figure 2: Components of the ACL message

Each LA maintains a comprehensive *translation table* (TT). Each translation table lists the concepts that the agent knows and maps them to the corresponding concepts in the associated ontology. English is the common language between all translation tables of all agents. For example, translation table of LA_{Gr} translates concepts from German to English and vice-versa, while the translation table of LA_{Ar} translates concepts from Arabic to English and vice-versa. Only translations that are credible will be recorded in the translation table. Google Cloud Translation API [3] for Java is used by MA to provide the translations for concepts and relations of the given ontologies to be enriched. Currently, Google Translate is considered as the best free online translator and can also be used offline [4]. One of the most prominent features in the proposed approach is that agents can exchange queries and messages to learn about incoming concepts and relations using ontologies, in different languages, of the other agents.

3.3.2. Agents Communication

Communications between agents are carried out using the Foundation for Intelligent Physical Agents (FIPA) Agent Communication Language (ACL) [23]. ACL is based on speech act theory: messages are communicative acts sent by the sender agent to the receiver agents in order to perform some actions. Language agents can learn using two ways: 1) users can teach them by supplying a pre-built ontology and its translation table, and 2) an agent can learn a new experience through its communications with its neighbors by *querying another agent for a certain concept*.

Querying Another Agent for a Certain Concept: The language agent sends out a query to the master agent when it needs to obtain some information about an incoming concept. Agents communication actions are defined in Algorithm 2. For instance, in Figure 2, LA_1 is querying agent MA about the English concept c_i. If the queried agent responds positively with its own semantics, description vector(s) (DV) of matched concept(s), the querying agent takes two actions: 1) enrich the associated ontology by the DV of that concept and 2) update its translation table. When the queried agent processes the query, it first matches the concept to its translation table. If a credible translation is found, then the queried agent simply sends back a description vector about the associated concept translated ($trans_{En}(DV(C))$). Figure 3 shows the complete communication actions taken by two language agents, LA_1 and LA_2, and the master agent MA when LA_1 wants to get some information about an incoming concept C.

There are two possible dialogues of communication between agents. First, language agent asks the master agent about a certain concept. Then the master agent replies with the DV of this concept. Second, the master agent asks other language agents, except the asking one, about the DV of the concept. If a positive reply is received, then the master agent asks the language agent to *stream-out* the available n DV of this concept. With the completion of these actions, we have succeeded in achieving the goal, i.e. simultaneous ontology enrichment.

[3] https://developers.google.com/api-client-library/java/apis/translate/v2

4. A Case Study

In this case study, we will use an example scenario of a multilingual text, containing English, German and Arabic, from the Information Technology domain. This case study aims to show the whole process starting from submitting the text to the system till getting the candidate concepts and relations. In this scenario, the user submits the following text: "*Netbeans fits the pieces together. The license of Netbeans software is available online. Die Firma produziert Netbeans software ist Oracle.* تطور البرمجيات بواسطة المبرمج ". Then, the system processes the text and produces a set of candidate concepts and relations, taxonomic and non-taxonomic. Finally, experts edit and validate them before adding them to the ontology, to preserve the quality. Table 1 demonstrates the output of each phase starting from the pre-processing phase to producing the set of candidate concepts and relations. Figure 4 shows a small fragment from the enriched ontologies before and after submitting a multilingual text chunk containing English, German and Arabic text. As shown in Figure 4, the German concept (Firma) detected in the input text and had a match in the Arabic ontology with (شركه), then LA_{ar} returns the DV which describes it to the master agent which in turn translates it to German as (Firma) and sends it to LA_{de} to use it to enrich the German ontology. Similarly, the concept (license), which found in the Arabic ontology as (رخصه), is used to enrich the German ontology after translating it to German (Lizenz).

5. Implementation and Preliminary Evaluation

For performing the evaluation, we have implemented all phases of the proposed architecture in Java. In particular, java agent development framework (JADE) is used for the development of the intelligent agent [5]. In our experiments, we consider only text from three languages: English, German and Arabic. All experiments are carried out on Ubuntu 16.04 LTS operating system using Java (JDK 1.8) with an Intel Core i7-4600U CPU @ 2.10GHz x 4 CPU and 10 GB of memory. Within this framework, the expert will make the final decision to select relevant concepts and relations.

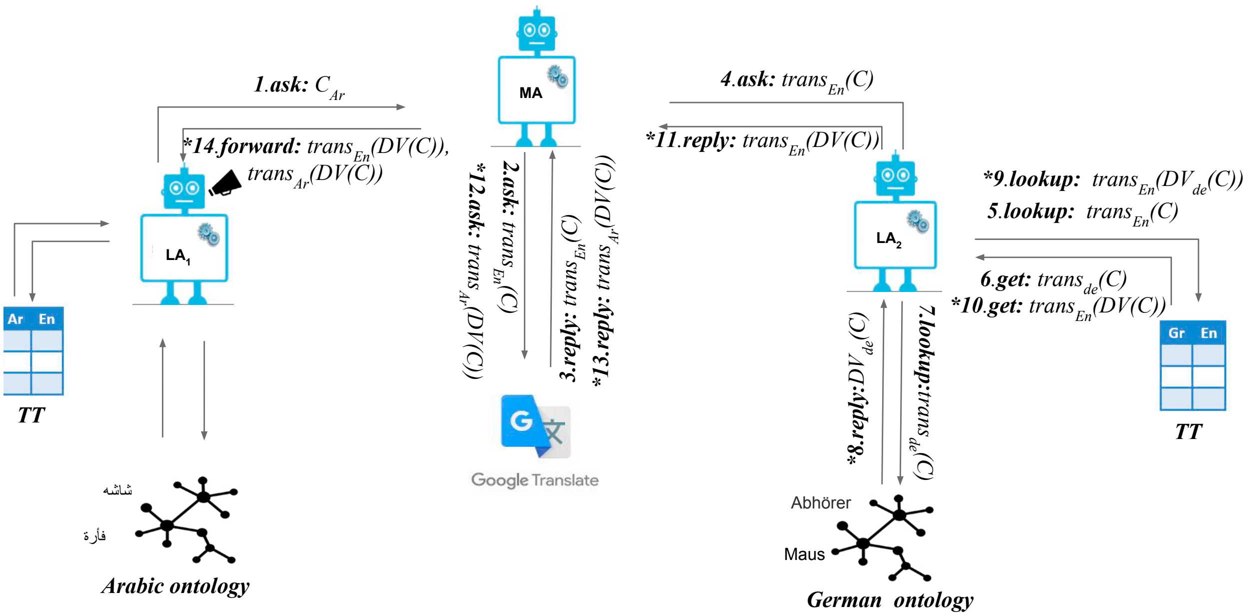

Figure 3: Agent communication actions when LA_1 wants to learn about a new incoming Arabic concept C. The asterisk "*" means that this step might repeat several times if there are more than one DV for a concept.

Table 1: Case study: the output of each phase starting from pre-processing to produce the set of candidate concepts and relations.

Phase		Output
Pre-processing	Lang. identification	En, De, Ar
	Language separation	**En**: Netbeans fits the pieces together. The license of Netbeans software is available online, **DE**: Oracle-Firma produziert Netbeans-Software, **AR**: تطور البرمجيات بواسطة المبرمج
	Sentence splitter and Tokenization	1) Netbeans fits the pieces together. 2) The license of Netbeans software is available online. 3) Oracle-Firma produziert Netbeans-Software. 4) تطور البرمجيات بواسطة المبرمج
	Stop words filter	1) Netbeans fits 2) license Netbeans software is available online. 3) Oracle Firma produziert Netbeans Software. 4) تطور البرمجيات بواسطة المبرمج
	POS tagging	1) Netbeans(N) fits(V) 2) license(N) Netbeans(N) software(N) is(V) available (ADJ) online(N). 3) Oracle(N) Firma(N) produziert(V) Netbeans(N) Software(N). 4) تطور (V) البرمجيات (N) بواسطة (PPREP) المبرمج (N)
Candidate sentence selection		sentence 1 has been removed because it has just one POS as noun.
Ontology enrichment		1. LA_{de} ask MA for $DV_{de}(Firma)$ 2. MA translates $Firma$ to English: $Company$ 3. MA asks LA_{ar} for $Company$ 4. LA_{ar} looks up for $Company$ in its TT and get the translation, in case of found, to Arabic: شركة 5. LA_{ar} lookup(شركة) in its Arabic ontology and get its DV, in RDF triples as: <http://...../برمجيات> <http://..منإنتاج> <http://...../شركة> 6. LA_{ar} gets the English translation of these DVs using its TT and sends it back to MA as: <http://...../software> <http://...../producedBy> <http://...../company> 7. MA translates the DV of $Company$, using an online translator, to German: <http://...../software> <http:/...../produziertVon> <http://...../Firma> 8. MA sends $DV_{de}(Firma)$ and $DV_{En}(Company)$ to LA_{de} 9. LA_{de} updates its TT and then enriches its German ontology with the incoming DVs.

Algorithm 1 Ontology Enrichment

```
procedure OntologyEnrichment(S[], L, LT[][])
    MA forwards input data to LA;
    LA = FindLA(LT[][], L);
    LA.Enrich(S[]);
end procedure
procedure LA.Enrich(S[])
    for each sentence s_i in S[] do
        for each concept c_i in s_i do
            found = SearchFor(c_i, LA.TT)
            if (found = True)
                Skip this concept because it is already exist;
            else
                DV[] = AskMA(c_i)
                if (!DV.isEmpty)
                    Enrich its ontology with the returned DV;
                    update LA.TT;
                else
                    Use any of the state-of-the-arts tool for learning;
                end if
            end if
        end for
    end for
end procedure
```

Generally, evaluation of ontology learning systems is the assessment of the resulting ontologies which are used to guide and refine the learning process. According to [11], there are three ways to evaluate a learned ontology. The resulting ontology can be evaluated: 1) in an executable application; 2) by experts or even by 3) comparing it with a predefined reference ontology (gold standard). We followed two strategies for our evaluation: satisfaction questionnaire evaluation and Gold standard-based evaluation.

Satisfaction Questionnaire Evaluation: We first succinctly introduce the evaluation setup and then discuss the result. A total of 24 ontology engineers were recruited for this questionnaire. At the beginning of the evaluation, all the participants have completely understood the approach by giving them a presentation about our approach and a case study describes it. Then, we asked them to fill in a satisfaction questionnaire[4] with 12 questions to assess the distinct phases of the proposed approach. Figure 5 shows the results of the satisfaction questionnaire analysis.

By analyzing the responses to the questionnaire, we have observed the following observations:

- Among all participants, 70.8% strongly agree that our approach would help ontology engineers.
- About 87.5% of the participants believe that it is *very hard* to manually develop ontologies from a multilingual text, while only 4.2% responds with *Neutral*.
- 62.5% of the participants pointed out that they believe that this approach is considered as a step towards a language-independent approach for ontology enrichment.
- 79% of the participants were satisfied with the *completeness* of the candidate relations and concepts, involving 33% of them are *strongly satisfied*.
- Finally, the overall satisfaction is : 60.9% of the participants were *strongly satisfied*, 21.7% of them were *satisfied* and 17.4% were *Neutral*.

[4] https://goo.gl/DnnPnv

Algorithm 2 Agents Communications

```
procedure AskMA(c_i)
    C_en = GoogleAPI.trans_en(c_i); //translate concept c_i to english
    for each LA connected to MA do
        DV[] = AskLA(C_en);
        if (!DV.isEmpty)
            break;
        end if
    end for
    return DV[]
end procedure
procedure AskLA(C_en)
    found = SearchInTT(C_en, LA.TT);
    if (found = True)
        C_LA.lang = Trans_LA.lang(C_en, LA.TT);
        DV_lang = SearchInOntology(C_LA.lang);
        DV_en = TranslateFromTT(C_LA.lang, LA.TT);
        return DV_en;
    else
        return null
    end if
end procedure
procedure FindLA(LT[][], L)
    for each integer i in LA do
        if (LT[i][1] = L)
            return i;
    end for
    return null
end procedure
```

Gold standard-based evaluation: The objective of the gold standard-based evaluation is to determine whether the concepts/relations used in the learning process are correct after comparing them with the gold standard results. We have used a corpus of texts from the 20 Newsgroups dataset[5]. It is a common benchmark used for many learning tasks [9]. The 20 Newsgroups dataset contains approximately 20,000 newsgroup posts, partitioned across 20 different newsgroups which are organized into more extensive categories involving computers, religion, science, and politics. We consider only the computers category (2936 documents). We have divided this category into two subsets. The first one is used for manually, by ontology experts, creating a reference standard ontology. Then, we have fed the system with this subset and saved the output. The second subset is used for evaluating CLOE.

We randomly selected 100 documents from the second subset. Fifty documents out of the 100 were translated using Google API to Arabic and German to be able to submit multilingual text to the system. Inspired by the information retrieval community, the quality of the learned ontology, without expert validation, can also be measured by Precision, Recall and F-measure metrics. The lexical precision (LP) and lexical recall (LR) reflect how good the extracted terms covering the target domain. Lexical precision (LP) is the fraction of retrieved concepts and relations that are relevant, while lexical recall (LR) is the fraction of relevant items retrieved by the system [28]. F-measure (F) is the harmonic mean of precision and recall.

[5] http://qwone.com/~jason/20Newsgroups/

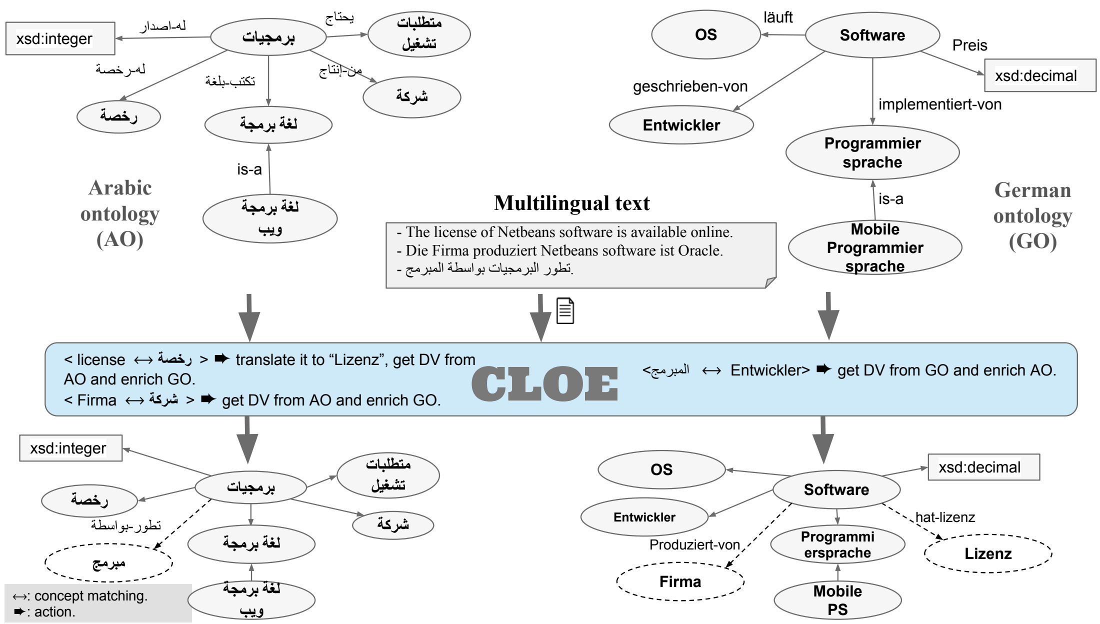

Figure 4: Small fragment from the enriched ontologies before and after submitting a multilingual text chunk containing English, German and Arabic text

We have identified four state-of-the-art approaches from the literature as the comparison approaches. The average of precision, recall, and F-measure are used to evaluate CLOE in comparison to the identified approaches. Table 2 shows the comparison results with the four identified approaches. CLOE outperforms all other systems in recall and F-measure. In average, CLOE achieves a high recall of 99% and a high F-measure of 94% among all the other four systems. CLOE achieved F-measure from 86% to 100% for the extraction of concepts and relations. In 2014, Albukhitan and Helmy [2] achieve a high precision by using explicit semantic analysis to validate the extracted concepts while CLOE outperforms it by 22% in terms of recall and 9% in terms of F-measure. In their paper in 2016, Albukhitan and Helmy [1] do not achieve

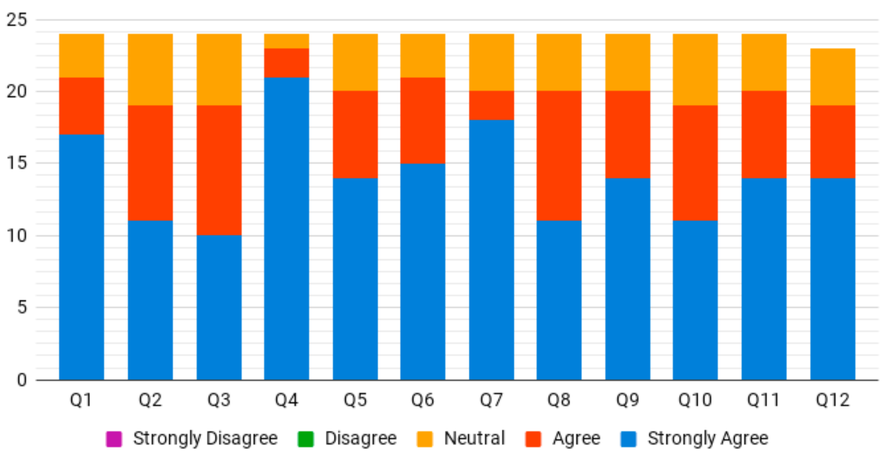

Figure 5: Results of the satisfaction questionnaire analysis

Table 2: Comparison with the state-of-the-art systems.

Metrics	CLOE	Benabdallah et al. [6]	Albukhitan and Helmy 2014 [2]	Albukhitan and Helmy 2016 [1]	Beseiso et al. [7]
LP	0.90	0.90	0.95	0.76	0.89
LR	0.99	0.71	0.77	0.45	0.73
F	0.94	0.79	0.85	0.56	0.80

satisfactory results by using statistical measures for concept recognition based on the frequency of the word in the corpus. Overall, the evaluation results are promising and show a considerable precision, recall, and F-measure. This finding confirms the usefulness of multi-agents in simultaneous ontology enrichment.

6. Conclusion

We present a novel multi-agent-based approach (CLOE) in order to enrich several ontologies (Simultaneous Ontology Enrichment) from multilingual data sources. Each agent is responsible for enriching an ontology for a particular language. Two-level hierarchical multi-agent model has been built where each node is modeled as an agent. The most prominent feature of the proposed approach is that agents could learn from each other, using a predefined communication scheme, to get the benefit of already-learned concepts found in the input ontologies. The main contribution of this work is the usage of a text segment to enrich several ontologies from different languages than the language of the input text. The results are satisfying compared to four state-of-the-art approaches. CLOE outperforms all other systems in terms of lexical precision and lexical recall, and consequently F-measure. In conclusion, this study has gone some way towards cross-lingual ontology enrichment.

Interesting lines of the future research are adapting the proposed approach to be domain independent approach and translate concepts based on language entries in DBpedia instead of Google Translation. Furthermore, techniques to generate multilingual ontologies from monolingual ones will also to be considered.

Acknowledgement

Said Fathalla and Shimaa Ibrahim would like to thank and acknowledge the Ministry of Higher Education (MoHE) of Egypt for providing scholarships to conduct this study.

References

[1] Saeed Albukhitan and Tarek Helmy. 'Arabic Ontology Learning from Un-structured Text'. In: *Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on*. IEEE. 2016, pp. 492–496.

[2] Saeed Albukhitan and Tarek Helmy. 'Multi-agent Based System for Multilingual Ontologies Maintenance'. In: *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on*. Vol. 1. IEEE. 2014, pp. 419–423.

[3] Mohamed Ali et al. 'The problem learning Non-Taxonomic Relationships of Ontologies from unstructured data sources'. In: *23rd International Conference on Automation and Computing (ICAC)*. 2017, pp. 1–6.

[4] Gabor Bella, Fausto Giunchiglia and Fiona McNeill. 'Language and domain aware lightweight ontology matching'. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 43 (2017), pp. 1–17.

[5] Fabio Bellifemine et al. 'JADE—a java agent development framework'. In: *Multi-Agent Programming*. 2005, pp. 125–147.

[6] Ali Benabdallah, Mohammed AlaEddine Abderrahim and Mohammed El-Amine Abderrahim. 'Extraction of terms and semantic relationships from Arabic texts for automatic construction of an ontology'. In: *International Journal of Speech Technology* (2017), pp. 1–8.

[7] Majdi Beseiso et al. 'Multilingual Ontology Learning Algorithm for Emails'. In: *Soft Computing Applications and Intelligent Systems*. Springer, 2013, pp. 229–244.

[8] Ibrahim Bounhas, Wiem Lahbib and Bilel Elayeb. 'Arabic domain terminology extraction: A literature review'. In: *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer. 2014, pp. 792–799.

[9] Ming-Wei Chang et al. 'Importance of Semantic Representation: Dataless Classification'. In: *AAAI*. Vol. 2. 2008, pp. 830–835.

[10] Gerard De Melo and Gerhard Weikum. 'Towards a universal wordnet by learning from combined evidence'. In: *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM. 2009, pp. 513–522.

[11] Klaas Dellschaft and Steffen Staab. 'On how to perform a gold standard based evaluation of ontology learning'. In: *International Semantic Web Conference*. Vol. 4273. Springer. 2006, pp. 228–241.

[12] Said Fathalla. 'Detecting Human Diseases Relatedness: A Spreading Activation Approach Over Ontologies'. In: *International Journal on Semantic Web and Information Systems (IJSWIS)* 14.3 (2018), pp. 120–133.

[13] Said M Fathalla, Yasser F Hassan and Maged El-Sayed. 'A hybrid method for user query reformation and classification'. In: *Computer Theory and Applications (ICCTA), 2012 22nd International Conference on*. IEEE. 2012, pp. 132–138.

[14] Said Fathalla and Yaman Kannot. 'A Bidirectional-Based Spreading Activation Method for Human Diseases Relatedness Detection Using Disease Ontology'. In: *Conference on Computational Collective Intelligence Technologies and Applications*. Springer. 2017, pp. 14–23.

[15] Said Fathalla et al. 'Analysing Scholarly Communication Metadata of Computer Science Events'. In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2017, pp. 342–354.

[16] Said Fathalla et al. 'SemSur: A Core Ontology for the Semantic Representation of Research Findings'. In: *Proceedings of the 14th International Conference on Semantic Systems*. ACM. 2018. In press.

[17] Said Fathalla et al. 'Towards a Knowledge Graph Representing Research Findings by Semantifying Survey Articles'. In: *International Conference on Theory and Practice of Digital Libraries*. Springer. 2017, pp. 315–327.

[18] David Faure and Thierry Poibeau. 'First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX'. In: *Proceedings of the ECAI workshop on Ontology Learning*. 2000.

[19] Jorge Gracia et al. 'Challenges for the multilingual web of data'. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 11 (2012), pp. 63–71.

[20] Udo Hahn and Martin Romacker. 'The SYNDIKATE text Knowledge base generator'. In: *Proceedings of the first international conference on Human language technology research*. 2001, pp. 1–6.

[21] Mamoun Abu Helou and Matteo Palmonari. 'Cross-lingual lexical matching with word translation and local similarity optimization'. In: *Proceedings of the 11th International Conference on Semantic Systems*. ACM. 2015, pp. 97–104.

[22] *International Open Government Data Search*. https://logd.tw.rpi.edu/iogds_analytics_2. [Online; accessed: 2017-12].

[23] Yannis Labrou, Tim Finin and Yun Peng. 'Agent communication languages: The current landscape'. In: *IEEE Intelligent Systems and Their Applications* 14.2 (1999), pp. 45–52.

[24] Marco Lui, Jey Han Lau and Timothy Baldwin. 'Automatic detection and language identification of multilingual documents'. In: *Transactions of the Association for Computational Linguistics* 2 (2014), pp. 27–40.

[25] Muhammad Rafi, Hilal Qureshi and Hasina Khatoon. 'Ontology Maintenance via Multi-agents'. In: *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on*. IEEE. 2009, pp. 955–959.

[26] Dennis Spohr, Laura Hollink and Philipp Cimiano. 'A machine learning approach to multilingual and cross-lingual ontology matching'. In: *The Semantic Web–ISWC 2011* (2011), pp. 665–680.

[27] Cássia Trojahn et al. 'State-of-the-art in multilingual and cross-lingual ontology matching'. In: *Towards the Multilingual Semantic Web*. Springer, 2014, pp. 119–135.

[28] Wilson Wong, Wei Liu and Mohammed Bennamoun. 'Ontology learning from text: A look back and into the future'. In: *ACM Computing Surveys (CSUR)* 44.4 (2012), p. 20.

[29] Takahira Yamaguchi. 'Acquiring Conceptual Relationships from Domain-Specific Texts.' In: *Workshop on Ontology Learning*. Vol. 38. 2001, pp. 69–113.

[30] Hui Yang and Jamie Callan. 'Ontology generation for large email collections'. In: *Proceedings of the 2008 international conference on Digital government research*. Digital Government Society of North America. 2008, pp. 254–261.