A Survey of the State of Art for Multilingual Ontologies Classified Based on Application

Abstract. Ontologies are semantic components that organize knowledge and enable context-aware communication among Web services. When modeling ontologies, different aspects are to be considered such as the natural language which represents ontology concepts. An ontology can be modeled using one language, two languages (bi-lingual), or several languages (multi-lingual). Multilingual ontologies have several applications which include biomedical, translation, and educational systems. This paper surveys and classifies the work in the domain of multilingual ontologies based on their applications. It summarizes and demonstrates the application of multilingual ontologies in each domain in terms of application, purpose of study, methodology, and multilingualism. The paper also indicates future research directions for modeling and using multilingual ontologies.

Keywords: Mono-lingual Ontologies, Multilingual Ontologies, computational Linguistics, Semantic Graphs

1. Introduction

Building and using semantic web artifacts are flourishing. Ontologies are basic structures for the semantic web for conceptualizing, organizing and sharing knowledge of a domain of interest [10]. They provide information concerning retrieval capabilities for knowledge-based intelligent systems. The conceptualization of ontologies can be modeled using one or more natural languages. When semantic data in ontologies are exhibited using one natural language, they are referred to as mono-lingual ontologies. While mono-lingual ontologies organize knowledge efficiently in a specific domain, multilingual ontologies conceptualize a specific domain using several languages. Semantic web facilitates the representation of web resources growth in various languages [19]. Multilingual ontologies augment ontologies with linguistic dimension(s). They expand the area of ontology applications by facilitating the development of multi-language user systems.

Multilingual ontologies can be developed by many approaches such as modeling, merging and matching techniques [32]. According to Ibrahim in [40], cross-
lingual ontology enrichment techniques is one of the most common methodologies that can be applied to build multilingual ontologies. However, the task of multilingual ontologies development and enrichment is time-consuming and prone to error particularly through human labor. Thus, many researchers studied and automated the process of building multilingual ontologies by developing algorithms and techniques. Ontologies, generally, and multilingual ontologies, specifically, are published and shared on the World Wide Web using the markup language OWL (Web Ontology Language) [37].

Ontologies are created and maintained using ontology editors such as the Protégé framework, which is a software that is widely used for building and managing ontologies. The research in [28] reflects on the importance of developing computational infrastructure for AI to the same degree as the Protégé project. Such tools are needed for the engineering and evaluation of the knowledge base of intelligent decision-support systems.

Hence, multilingual ontologies gained popularity in many disciplines including healthcare, education, and information technology. They also enable several digital applications such as translation and context communication among online Web services [15]. Therefore, this paper surveys the research carried out on multilingual ontologies. It classifies the research work based on applications of multilingual ontologies. In addition, it highlights the scientific methodologies used for building applications based on multilingual ontologies.

The organization of this paper is as follows: Section 2 presents the methodology used for the survey. Section 3 demonstrates the related work classified according to the application. Section 4 presents current multilingual ontologies repositories. Section 5 includes the conclusion and future research directions.

2. Methodology

In this review, research work published during the past few years concerning applications of multilingual ontologies is selected. The research work is analyzed then classified based on the application domain of the multilingual ontologies. This method of organizing and presenting the multilingual ontologies state-of-the-art work assists researchers and stakeholders in different domains to utilize ontologies and incorporate the built knowledge into their applications. Focusing on the application field when classifying the literature also introduces inquisitive researchers and scholars to new domains where multilingual ontologies have potential applications. Generally, most of the reviewed work fall into one of the domains shown in Figure 1.

3. Applications of multilingual ontologies

The following section reviews the related research work that employed multilingual ontologies in various domains. For each domain, the research work was presented in chronological order, as presented in Table 1. Each research paper is addressed in terms of purpose of study, application of multilingual ontology, methodology, multilingualism (bi-lingual, multilingual), and time. Note that the multilingual ontologies are not evaluated in terms of domain coverage, ontology validation, nor ontology consistency. It represents a collective of literature
reviews of work in the domain of multilingual ontologies.

3.1. Agriculture

Authors of [10] address the importance of adopting multilingualism in ontologies in many domains including the domain of agriculture. They indicated that the Food and Agriculture Organization (FAO) must provide standard definitions for agricultural terminologies in official languages. In this regard, Thenmozhi and Aravindan in [16] have proposed a Cross-Lingual Information Retrieval System (CLIR) based on ontology in the agricultural domain. CLIR aims to help farmers who speak Tamil to retrieve information in both Tamil and English from the Internet. The system translates a query from Tamil into English both syntactically and semantically. In the translation process, a module utilizes a Word Sense Disambiguation (WSD) for ambiguity removal from Tamil query. Then it uses an agriculture ontology in English to reformulate the translated query and remove ambiguities before amending it into the search engine for information retrieval. The system performed several query experiments in the domain of agriculture and outperformed other techniques when comparing their precision.

3.2. Pedagogy

Multilingual ontologies play a role in pedagogy. Data and knowledge need to be represented in different native languages besides English to make it available to a broader audience for efficient use [43]. Researchers in [2] engineered an algorithm ontology using the Arabic and English languages. The authors followed the Noy and McGuinness approach to construct the AEOA for it to be utilized in educational purposes. In the medical domain of education, the researchers in [36] proposed an automatic ontology-based approach to reduce the time required for generating multilingual Comprehensive Integrative Puzzle (CIP) assessment questions. They also provided the structure and scheme of the CIP Generation Method as well as performance evaluation.

Researcher Florrence [33] proposed a new algorithm called Multilingual Ontology (MOnto). In the MOnto methodology, multilingual ontologies are merged and mapped to build multilingual ontology applications in the education field. Moreover, the researcher proposed a new plug-in called MLGrafViz to visually display ontologies in non-English natural languages such as Tamil and Zulu.

In [43], Ivanova proposed a bilingual ontology model for learning to allow content reuse in E-Learning systems. The Bulgarian-English ontological model ensures that domain-specific knowledge in the learning resources is strictly separated from pedagogical or general knowledge.

3.3. Linguistics

Bouma in [17] proposed a system for ontology alignment using extensive cross-lingual resources. The system was implemented by mapping a thesaurus of the Netherlands Institute for Sound and Vision (GTAA) to both dbpedia and the English WordNet. It uses intermediaries namely EuroWordNet and Dutch Wikipedia respectively, to map from Dutch to English Wordnet and to extract a database from English Wikipedia (dbpedia) [17]. However, this method has its limitations regarding the coverage and richness of the resources available in some languages.

3.4. Organizational structure of a university-administrative or information systems

Authors of [18] proposed an architecture for multilingual ontologies matching by translating the ontologies to a common global language, then applying ontology matching tools to the translated versions. They utilized two existing ontologies: An ontology in the German language for Freie Universität Berlin in Germany and the other expressed in the Arabic language for Fayoum University in Egypt. Both ontologies cover the organizational structure of a university and are represented using the Web Ontology Language (OWL 2). The paper discussed several approaches for ontology matching and compared them. To develop the multilingual ontologies, each ontology was manually translated into an English version then applied different approaches for ontology matching such as: Falcon, SAMBO, RiMOM, ... etc. They found several shortcomings in finding correct matchings in the translated versions of the ontologies due to the difference in the ontology models.

3.5. Information technology

In multilingual environments, mapping and matching for cross-lingual ontologies are essential. The research [14] addressed matching cross-lingual
and multilingual ontologies. It outlines different techniques and approaches to systematically evaluate the available data sets, approaches, and systems for multilingual matching. Additionally, Ivanova [42] surveyed the mapping approaches for available cross-lingual and multilingual ontologies. The following subsection illustrates the applications of multilingual ontologies that are related to ontology mapping and matching.

3.5.1. Ontology mapping and matching

As knowledge linking and mapping is essential for sharing knowledge across languages, the researchers in [45, 46] proposed an approach based on concept annotation for increasing cross-lingual knowledge linking. According to the results of the experiment on the Chinese and English Wikipedia data, the quantity and quality of CL prediction can be effectively improved when using concept annotation.

Khiat introduced the Cross-Lingual Ontology Matching (CroLOM) [6], an automatic system. The system was based on handling multilingualism challenges based on NLP techniques, translator, and similarity computations. Aside from the main disadvantage of CroLOM which is the execution time, the system performance was promising on the track of the Multifarm of OAEI 2016 evaluation campaign run. The experiment included ADOM, an Arabic dataset describing the conference domain, which was integrated within the Multifarm dataset [7].

The researchers in [11,12] introduced a Semantic-Oriented Cross-Lingual Ontology Mapping (SOCOM) which acts as an upgrade to software data exchange, specifically with multilingual ontology-based systems. This approach aims to select the most appropriate translation of a given ontology label based on prescribed mapping context. Throughout the study, the effectiveness of the SOCOM structure is determined by mapping semantically equivalent English and Chinese ontologies in the research community domain. The SOCOM approach has generated higher quality matches in cross-lingual mapping than the generic approach.

Later in [10], the authors presented SOCOM++ which is an improvised approach that enables the user to configure the properties of selecting label translations in order to tune the intermediate resultant mapping. For evaluation, a set of six distinct SOCOM++ configurations were applied to two experiments for mapping which involved three languages: English, French, and Chinese. When cross-lingual mapping problems are translated to monolingual mapping problems via translation techniques in a CLOM system, the final matches are usually subjected to degradation [13]. Therefore, the authors of [13] have worked on proposing a new approach to enhance the matching qualities in a CLOM system through the pseudo feedback technique. The validation process includes ontologies from benchmark dataset of the OAEI 2009 in English and French in the field of the bibliography. The experimentation shows improvement of monolingual ontology matching by the consolidation of the pseudo feedback.

The researchers in [15] provide a novel machine learning ontology matching approach based on ranking SVM matching function. They distinguish between cross-lingual and multilingual ontology matching by providing precise definitions of them. To quantitatively measure the effectiveness of multilingual information based on different matching algorithms, multiple matching scenarios in different languages with distinctive learning settings were simulated.

In [22], Gracia addressed the importance of creating, presenting, and storing linkages of cross-lingual type in the semantic web. In like manner, the paper addresses different perspectives of the problems of cross-lingual ontology matching. Correspondingly, the major contribution of [23] is to provide an overview of the web of data challenges when it comes to cross-lingual linking. They did not only address three challenges of cross-lingual linking: discovery, representation, and reuse, but they also highlighted some research areas and techniques that will help to get closer to a completely multilingual Web of Data.

The authors of [24] presented a framework for ontology alignment composed of main features: background information use and imprecision handling in the alignment and the matching process. The suggested method produces a fuzzy knowledge base which enables the discovery of many-to-many cross-concept relationships. The benchmark used for this approach is the MultiFarm dataset which is composed of 7 ontologies and the conference domain is presented in 8 different languages.

LYAM++ is a cross-lingual system that matches ontologies based on a fully automatic approach [9]. Similarly, the approach makes use of background knowledge to control the shortcomings of existing cross-lingual matching systems. This system was experimentally evaluated on the basis of the Multifarm benchmark.

The authors in [21] proposed a system that computes the context ontological similarities between
a pair of terms called CIDER-CL. In CIDER-CL, class and property similarities of cross-lingual and monolingual ontologies are computed using neural networks. The alignment system was evaluated based on different ontologies while considering different languages. For monolingual ontology matching, CIDER-CL works well with the SoftTFIDF measure. On the other hand, CIDER-CL is not suitable for large ontologies, especially for cross-lingual matching.

To facilitate cross-lingual ontology mapping, researchers in [27] studied the effectiveness of automatic translations obtained from Google Translate and BabelNet to evaluate the effectiveness of automatic translation based on the coverage and correctness of the translation. Also, one of the most important contributions of this research is the efficiency investigation for automatic translations of specific synsets types.

As resources are developed into various languages, these resources and ontologies must be aligned appropriately through multilingual ontology alignment techniques. Therefore, M. Kachrudi in [35] introduces Cross-Lingual Ontology alignment (CLONA) that functions with respect to information retrieval techniques and external resources. This experimental evaluation uses the Multifarm dataset that is translated in nine different languages. Alongside best practice methods in the Ontology Alignment Evaluation Initiative OAEI’2015, CLONA ranked second.

In the area of cross-lingual mapping for ontologies, authors of [26] proposed a semi-automatic Interactive Cross-lingual Mapping (ICLM) application to improve automatic alignment. The approach aims to integrate data from multilingual resources by incorporating multi-users feedback. The application automatically matches the source concepts with the potential target concepts. It considers the mapping task's difficulties and requires a mapping task for validation by several users. For source concepts that require user feedback, the application assigns tasks of mapping selection to the users according to estimated efforts to validate the mapping and asks them to select the best match from a group of resultant candidates. The approach was implemented on mapping Arabic source concepts from Arabic Wordnet (ArWN) to the English target concepts and performed well compared to the crowdsourcing approach.

The researchers of [20] proposed the Interactive Compact Differential Evolution (ICDE) algorithm used for cross-lingual ontologies matching. They distinguish the various cross-lingual classes by introducing the similarity metric of cross-lingual ontology. They combine both automatic and user matches to obtain alignments timely with high quality standards.

The authors of [41] proposed a method that incorporates the concept of semantic web to recommend news items written in various languages. The proposed method surpassed the keyword-based method in relating English news based on the Bengali news.

3.5.2. Ontology augmentation and enrichment

In [39], the authors proposed a new approach for ontology enrichment to construct multilingual ontologies based on Cross-Lingual Matching (OECM) technique. It is a fully automated enrichment approach for monolingual ontologies based on various natural languages. The best terms that semantically match the target ontology are selected. In this paper, the English Scholarly Communication Ontologies were enriched using ontologies in Arabic and German. The results were compared with five advanced approaches and scored higher using recall and precision.

The use of cross-lingual ontologies to create and enhance multi-lingual ontologies has been proposed in numerous studies. Authors in [29, 30] put forward a novel multi-agent-based Cross-Lingual Ontology Enrichment (CLOE) approach. The key contribution of this approach is enriching ontologies simultaneously based on multilingual sources. The input of the proposed approach is texts presented in multilingual forms and ontologies that require enrichment, while the ontologies post-enrichment is the output. The approach is composed of three phases: preprocessing, candidate sentence, and selection phase. At the preprocessing step, the input texts are cleaned and preprocessed for further processing. In the candidate sentence selection phase, the preprocessed texts are filtered and candidate sentences are selected based on specific rules. The last phase involves a proposed innovative algorithm that allows the input ontologies to interact and learn from each other. To validate the approach, experts evaluate extracted terms and their relations before adding them to the ontologies to be enriched. The approach was implemented on text presented in several languages including Arabic, English, and German. It performed well compared to four other approaches [29].

3.6. Tourism

In the tourism industry, information including accommodations and attractions is frequently
available across many websites and in various formats; however, only in a few languages. End users and managers can both benefit from using semantic data sources available in different languages. The authors of [8] listed many ontologies that were proposed in the literature on tourism. Hontolo [31] is one of the accommodation-related ontologies which is beneficial in many resources in the accommodation industry. The terms of the ontology are available in English, Portuguese, Spanish and French. The ontology was built based on existing ontologies, translated manually, then expanded using online reviews.

3.7. Health care

To facilitate international communication in the medical field, the Working Group on Health Informatics for Inter-regional Cooperation engineered and ontology for medical informatics [3]. Medical Informatics and Digital Health Multilingual Ontology (MIMO) is a multilingual ontology that presents digital health and health informatics concepts in more than 30 languages. It is hosted on the (HeTOP) ontology portal, has over 1,000 concepts and is continuously updated. MIMO is used as a collaboration tool to support self-learning and enable communication between researchers and medical workers.

In addition, the researchers in [38] have verified the convenience of MIMO by employing the multilingual ontology in indexing digital teaching resources. MIMO concepts are used to index 345 resources that are integrated into a semantic research engine to enable accessibility to digital health teaching resources. The enhanced quality and efficiency in the approach of these resources encourages members’ collaboration and communication in the industry and academic fields.

To aid health care providers and facilitate communication for screening patients in several languages in South Africa, the authors of [25] utilized an existing semantic machine architecture based on knowledge representation supporting awareness for Covid-19. The application that was intended for multilingual speech-to-speech translation has been extended into a knowledge base for Covid-19 screening. The research included ontology development to represent and integrate the semantic data of patients.

4. Repositories supporting multilingual ontologies

There are some articles that are concerned about multilingual repositories such as [19] which describes a methodology that automates ontology population and lexical acquisition semi-automatically. Another article is [1] which proposed a multi-lingual ontology server (MOS) that helps in searching and selecting appropriate web services even if it is available in the central repository UDDI in a different language. UDDI is an emerging web service framework called Universal Description Discovery and Integration.

The following lists the current repositories for multilingual ontologies. These repositories are addressed with regards to the number of concepts and representation languages.

- MCR - Multilingual Central Repository [44], which provides the Inter-Lingual-Index (ILI). ILI connects words to equivalent translations in other languages. Mappings among WordNet edits are automatically generated.
- MCR 3.0 | Adimen [4, 5], which is the recent version of the Multilingual Central Repository (MCR). MCR 3.0 is a new wordnet version for the five languages: Basque, Catalan, Galician, English, Spanish, and the ontological knowledge from WordNet Domains, Top Ontology and SUMO. The new version is created by (A Gonzalez-Agirre). This new version uses WordNet 3.0 as ILI.
- Arabic-Multilingual Database [34] presents formal conceptual representation in Arabic based on state-of-art science of ontological analysis. The ontology currently consists of approximately 1300 concepts that are well-investigated and 11,000 partially validated concepts. A mapping process was conducted for the proposed ontology with Princeton WordNet, Wikidata, and other resources. A lexicographic search engine is provided to make the ontology accessible and searchable. The database also contains 150 Arabic-multilingual lexicons enriched using the ontology.

5. Conclusion

This research paper surveys the literature in the domain of multilingual ontologies. An analysis of the state-of-art was carried out. The work was classified into categories based on the applications of the multilingual ontologies. Future research directions in
the domain include building repositories for multilingual ontologies. In addition, multilingual ontologies can be employed in robotic communications to enable multilingual human-robot interaction. IOT applications can use multilingual ontologies for context-aware web service communication. Moreover, tourism information centers can utilize multilingual ontologies for information shared in several languages. Additionally, multilingual ontologies can offer promising information structures for e-commerce applications. In the research area, multilingual bibliographic ontologies are expected to facilitate researchers' tasks. Pharmaceutical industries can also utilize multilingual ontologies modeled for the domain.

**Acknowledgement**

This research work was funded by Institutional Fund Projects under grant no (IFPRC-028-612-2020). Therefore, authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, Jeddah, Saudi Arabia.”

<table>
<thead>
<tr>
<th>Application Domain</th>
<th>Ref.</th>
<th>Year</th>
<th>Available ontologies</th>
<th>Purpose</th>
<th>Method</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[24]</td>
<td>2012</td>
<td>Multilingual fuzzy knowledge base</td>
<td>Ontology Matching based on fuzziness and cross-linguality framework</td>
<td>Background knowledge use(Yago)</td>
<td>5 languages</td>
</tr>
<tr>
<td></td>
<td>[9]</td>
<td>2015</td>
<td>7-languages conference ontology</td>
<td>Fully automated matching system based on cross-lingual ontology</td>
<td>LYAM ++ - Yet-Another-Matcher- Light use of background knowledge (BabelNet)</td>
<td>7 languages</td>
</tr>
<tr>
<td></td>
<td>[35]</td>
<td>2016</td>
<td>9-languages conference ontology</td>
<td>Alignment for cross-lingual ontology</td>
<td>CLONA method - Ontology alignment based on cross linguality</td>
<td>9 languages</td>
</tr>
<tr>
<td>Year</td>
<td>Documents</td>
<td>Language/Field</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>----------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>Bengali-English document ontologies</td>
<td>Recommendation System for news based on crosslingual features</td>
<td>Translation to cross-lingual ontology mapping</td>
<td>Bengali-English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Scholarly communication ontologies</td>
<td>Multilingual ontology building using ontology enrichment</td>
<td>OECM - ontology enrichment using cross-lingual matching</td>
<td>English-German-Arabic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>IT ontology</td>
<td>Simultaneous Ontology enrichment</td>
<td>CLOE - Cross-Lingual ontology enrichment</td>
<td>English-German-Arabic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>9-languages conference ontology</td>
<td>Cross-lingual ontology alignment</td>
<td>ICDE - interactive compact differential evolution algorithm</td>
<td>45 different language pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>Dutch-English ontologies</td>
<td>Thesaurus cross-lingual alignment</td>
<td>Ontology alignment using intermediate ontologies</td>
<td>Dutch-English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td>Patient information for COVID-19 knowledge representation</td>
<td>Extension of an existing speech-to-speech translation application for COVID-19 screening</td>
<td>Ontology development to represent and integrate semantic data of patients</td>
<td>English-multiple South-African languages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>MIMO</td>
<td>Development of multilingual ontology for medical informatics</td>
<td>Gradually developed in five steps from flat dictionary combining knowledge bases, generated knowledge and domain experts' opinions</td>
<td>More than 30 languages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>MIMO - Health resources concepts for education</td>
<td>Digital indexing for health teaching resources</td>
<td>MIMO - Health and medical digital informatics ontology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Tamil–English agriculture ontology</td>
<td>Ontology-based retrieval system Tamil–English cross-lingual</td>
<td>Translation of an automatically learnt English agriculture ontology utilizing a word sense disambiguation module</td>
<td>Tamil–English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>Arabic-English algorithms ontology (AEAO)</td>
<td>Cognitive applications</td>
<td>The Noy and McGuinness approach</td>
<td>Arabic-English</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td>OntoCIP ontology</td>
<td>Automatic CIP generation</td>
<td>Translation of OntoCIP ontology with an online translation service</td>
<td>English-French-Italian…etc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Hontology</td>
<td>Accommodation applications for information extraction</td>
<td>Manual translation of existing ontologies then expanded using online reviews</td>
<td>English, Portuguese, Spanish and French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>[33] 2020</td>
<td>Bulgarian-English bilingual ontologies</td>
<td>A reuse model based on ontology for e-learning resources</td>
<td>Annotation for bilingual ontology and mapping ontologies</td>
<td>English-Bulgarian</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>--------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>[43] 2019</td>
<td>Education ontology</td>
<td>Building multilingual ontologies for education</td>
<td>MOnto method-knowledge resources in different languages are merged and mapped</td>
<td>Tamil–English Zulu–English</td>
<td></td>
</tr>
<tr>
<td>Administration</td>
<td>[18] 2013</td>
<td>German-English-Arabic ontology</td>
<td>Organizational structure of a university</td>
<td>Manual translation of existing ontologies into a common language then using automatic matching tools</td>
<td>German-English-Arabic</td>
<td></td>
</tr>
</tbody>
</table>

References


M. Kachroud, G. Diallo, and S. B. Yahia, Initiating cross-lingual ontology alignment with information retrieval techniques, 2016.


