A FRAMEWORK FOR BUILDING A MULTILINGUAL INDUSTRIAL ONTOLOGY: METHODOLOGY AND A CASE STUDY FOR BUILDING SMARTPHONE ENGLISH-ARABIC ONTOLOGY

Amany K. Alnahdi

Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

As Web 3.0 is blooming, ontologies augment semantic Web with semi–structured knowledge. Industrial ontologies can help in improving online commercial communication and marketing. In addition, conceptualizing the enterprise knowledge can improve information retrieval for industrial applications. Having ontologies combine multiple languages can help in delivering the knowledge to a broad sector of Internet users. In addition, multi-lingual ontologies can also help in commercial transactions. This research paper provides a framework model for building industrial multilingual ontologies which include Corpus Determination, Filtering, Analysis, Ontology Building, and Ontology Evaluation. It also addresses factors to be considered when modeling multilingual ontologies. A case study for building a bilingual English-Arabic ontology for smart phones is presented. The ontology was illustrated using an ontology editor and visualization tool. The built ontology consists of 67 classes and 18 instances presented in both Arabic and English. In addition, applications for using the ontology are presented. Future research directions for the built industrial ontology are presented.

KEYWORDS

Ontology, Semantic Web, Computational Linguistics, Industrial Ontologies

1. INTRODUCTION

Ontologies are knowledge structures for conceptualization a domain of interest. Ontologies are used for computational purposes including education, information retrieval, and historical applications. Building ontologies depends on analysing textual information in a specific domain and extracting concepts and defining the relations between these concepts. Building ontologies can be a tedious task that would be expedited by adhering defined procedural steps.

Though ontologies are mostly built using one natural language, building multilingual ontologies is expected to support computational and communicational applications. Multilingual ontologies are used to structure concepts in a specific domain using different natural languages. Combining concepts in different languages can support semantic communication for both technology users and applications.

This paper proposes a framework model for building industrial multilingual ontologies. The model consists of several components that would help in facilitating building ontologies. These components emphasize on having multi languages. Issues to build and combine concepts of different languages such as translation are also discussed.

A case study to build a smartphone ontology is presented in this paper. This smartphone ontology is a bilingual English-Arabic ontology that combines knowledge of smartphone terminology. Building the ontology was basically based on recent commercial textual information. The use of contemporary information is expected to enable ontology uses of commercial market products. It also is expected to help product manufacturer and business owners to market their products.

This paper is organized as follows: Section 2 explore related work to this paper. The proposed framework is presented in Section 3. A case study of implementing the framework is explained in Section 4. The conclusion and indications for future research work are presented in Section 5.

2. RELATED WORK

Technological artifacts that are designed to support several languages are pointed out in the literature of computational linguistics to support several languages are pointed out in the literature of computational linguistics. The authors of [3] presented foundation principles of semantic Web. In addition, they presented an ontology building methodology consist of several steps. In [8], the authors proposed a Unified Process for Ontology building (UPON) as an incremental methodology for ontology building inspired from Unified Software Development Process. Their method is use-case driven and iterative. In [9], the authors presented UPON methodology with an example in the eBusiness domain.

In [5], the authors indicated the deficiency of the technology support of Arabic language, they also indicated some existing Arabic Semantic Web technologies. In addition, they pointed out some of multilingual support of Semantic Web technologies. The authors of [1] proposed a methodology for mapping ontologies from different languages. In [2], the authors proposed a bilingual ontology used to conceptualize knowledge in the Algorithm domain.

In [12], authors presented building blocks of ontologies built by several members. Authors of [10] introduced a Key phrase KP-Miner system in which the rules of the system are based on the nature of the Arabic or English documents. The authors of [7] suggested an ontology inference based semantic matching methodology for languages such as Hindi and Marathi. In [15], authors presented a model to measure the similarity between sociocultural factors in their Wikipedia communities with different languages Arabic, English, and Korean. Authors of [6] presented a methodology for building lexical Standard Arabic Language resources where the meaning of the words is based on domain ontology and the suggested Upper Merged Ontology.

There are few research works addressed industrial semantics of the industrial domain. Authors of [4] addressed issues regarding compiling multilingual thesauri for industry-specific translation (IST) specifically of Kazakh language. The controlled multilingual thesaurus proposed by authors was used to industry-specific information retrieval functionalities. Authors of [13] extended ontologies to be used for verbal communication between humans and robots. Authors of [14] addressed the shortcoming in knowledge exchange in the domain of smart manufacturing which is the protection and knowledge generation quality. They indicated the importance of applying ontologies for decision making, interoperability, and reasoning.

3. METHODOLOGY FOR BUILDING MULTILINGUAL INDUSTRIAL ONTOLOGIES

Building multilingual industrial ontologies are based on considering linguistic and technological factors. Industrial ontologies depend on the industrial domain as terms used to in the built ontologies are enterprise oriented. Enterprise terms can differ in several geographical locations where the language and dialects vary. Such factors determine the methodology that is used for collecting corpuses.

The steps of building multilingual enterprise ontology are illustrated in Figure1. As demonstrate in the figure the first step is to determine the corpus of the enterprise domain. The croupous of an enterprise domain can be built from several resources such as commercial booklets and pamphlets used to advertise products. Commercial online Websites are also helpful resource of building enterprise ontology. Online customer reviews of products are useful to be considered as they contain genuine jargon used by customers. Collection of these terms can be done by manual, semiautomatic, and fully automatic techniques. Manual collection for the ontology concepts can be done from selecting terms relating to the domain of interest from books, manual, or Website. Semi-automatic approach for collecting ontology concepts can be done by combining manual collection in addition using automated tools such as tokenizers. Fully automated approaches entirely depend on using automated tools with minimum human intervention.

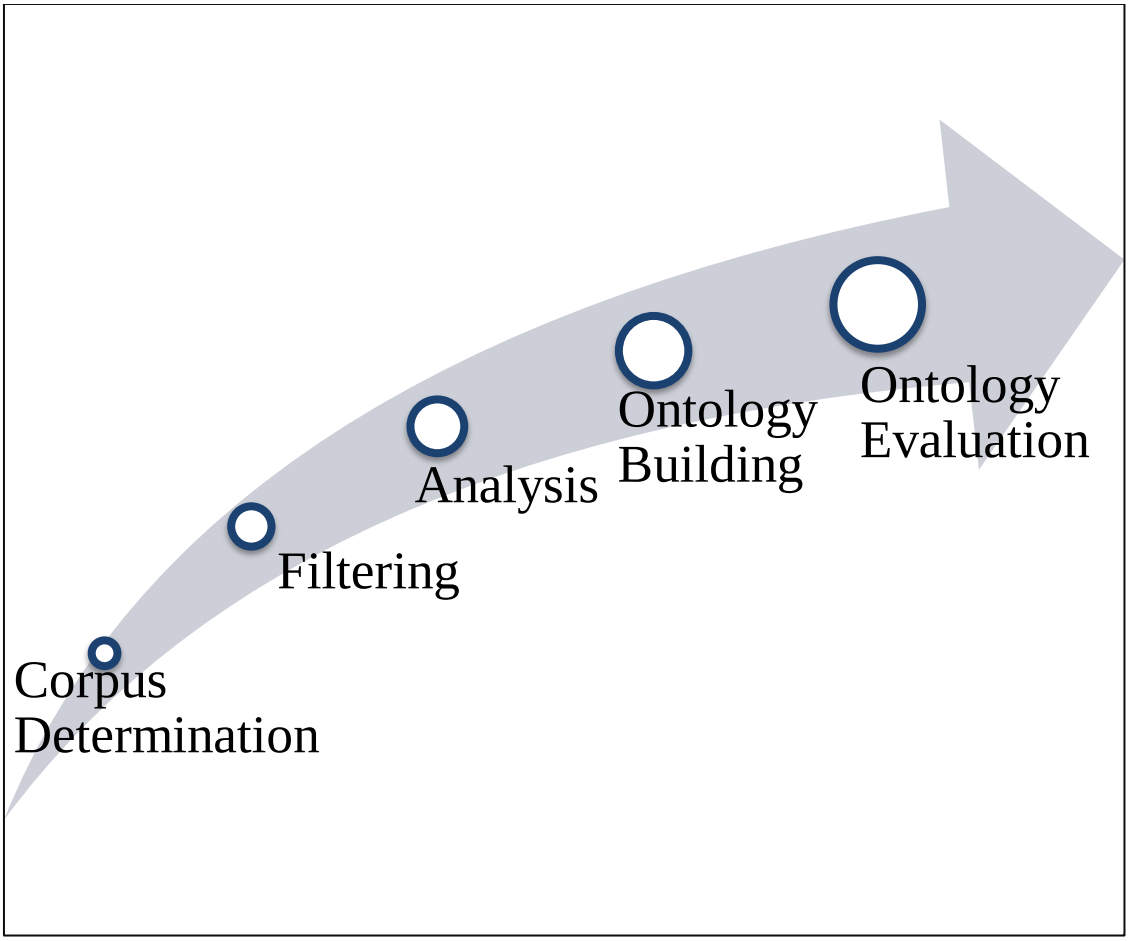

Figure 1. Multilingual Enterprise Ontology Building Model

A filtering phase is used to process the collected terms. The filtering would include removing of stop words and redundant terms in the corpus pool of enterprise terms. Stop words such as articles (e.g. The), propositions (e.g. or, and), and (e.g. while) can be eliminated using text processing tools that would remove words that belong to the list of stop words. Redundant terms that are collected from different text collection can be removed using text analysis tools. Handling the translation of the filtered terms can be attained by using some of the already available translations in enterprise booklet provided in multi languages. In addition, some of the translations can be obtained from people who are working in the enterprise area and exposed to use languages for which the ontology to be built. Meetings and online surveys can also be used to collect translation of terms to be used in the ontology. In addition, analysing online commercial videos is also beneficial for identifying appropriate translations used for specific terms.

When building an enterprise multilingual ontology, stakeholders who are involved in building the ontology are suggested to consider the following issues.

- What if there is more than one synonym or translation for the same concept? How this issue can be handled in the enterprise ontology? By using the translated term that is common in the context of the enterprise domain. In addition, several translations can be embedded in the ontology.
- How building multilingual enterprise ontologies can differ in different domains? The source of concepts to be modeled is different. Industrial ontology terminologies can be found in advertisement and booklets of devices.
- What are available information resources for collecting the terminologies? Contemporary multilingual industrial catalogues are main resource for concepts in the domain. Online stores also have categories for common terms for industrial products. Manufacturer manuals are also significant resource for industrial terminologies.
- Are there any software automation tools that can help in filtering and analyzing the multimedia information including textual, audio, and video? While natural language processing tools can be of help in building ontologies, Simi-automation techniques are more accurate and suitable for modelling ontologies. Experts revisions can filter the precise terms and accurate classifications.
- Do these software automation tools work in only one language, or all the languages supported by the multilingual ontologies? To the best of the author knowledge there is still gap in building multi-lingual ontologies using automated tools as most translation techniques have limitation in semantic aspects.
- What ontology editor(s) that support languages used in building the ontology? As ontologies are textual files, text editors for experts can be used. In addition, several ontology editors such as protégé [16] allows ontology modeling and visualization.
- What ontology evaluation methodologies can be used for ontology verification? Statistical methodologies that provide reflection about quantitative aspects of the ontology such as number of axioms and classes in the ontology are considered reflective metrics. In addition, the comprehensive and accuracy of the built ontology can be evaluated by experts in the domain.

4. CASE STUDY FOR BUILDING BILINGUAL SMARTPHONE ONTOLOGIES

Smartphones are pervasive technology devices that are used for telephone communication, Internet browsing, and application uses to name few. With the several functionalities that a smartphone would provide, terms associated in the domain of smartphone are ample. Organizing concepts in the smartphone domain in an ontological form would provide a semantic Web infrastructure component to people who are interested in the smartphone domain. The use of this

ontology can help consumers to search about smartphone products. Reviews also can be easily found and read about smart phone devices. Moreover, Website marketing designer can use this classification in their Websites. As using common terms can help in accessing these terms by users which is expected to improve the marketing of the products.

Uses of the built smartphone ontology include the following:

- Catalogue annotation
- Online search about smartphone
- Search filtering
- Supporting Website with search features
- Marketing smartphone device on online commercial websites

The ontology was built based on the proposed theoretical model in this paper. The corpus of the smartphone ontology includes collected commercial pamphlets for around one year. local commercial Website that support both English and Arabic languages. The collected terms were filtered using tokenizers and word counter tools. Then, the terms were analysed by categorizing the relationships between terms. The translations of the terms were coordinated.

The ontology was built using Protégé [16] an ontology editor that supports building ontologies in several format including owl and rdf. Figure 2 illustrates the implemented smart phone ontology using protégé tool.

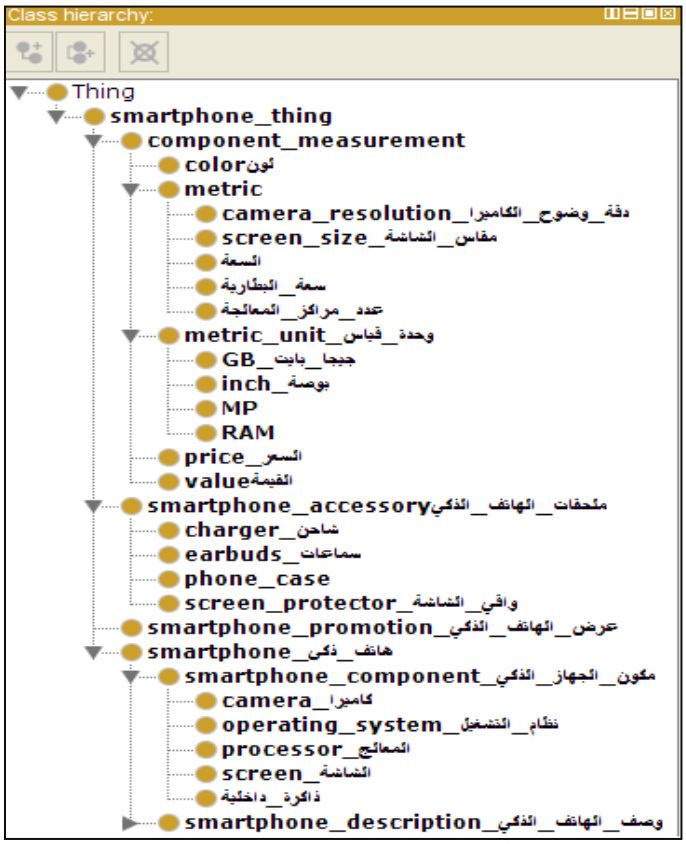

Figure 2. Snapshot of the English-Arabic Smartphone Ontology

To evaluate the ontology number of classes and individuals in the ontology are statistically determined by Protégé metrics plugin. As can be seen from Table 1, number of classes in the built smartphone ontology is 67 and number of instances is 18.

Table 1. Statistics of the built smartphone ontology

Ontology Metric	Count
Number of classes	67
Number of individuals	18

5. CONCLUSION

Enterprise ontologies are expected to support technological communication in the E-Commerce domain. This research provides a theoretical framework for building multilingual ontologies in industrial domain which include the phases Corpus Determination, Filtering, Analysis, Ontology Building, and Ontology Evaluation. The process for building the domain ontology is explained, as well a case study for implementing a bi-lingual ontology for smartphones was modelled using Arabic and English languages. In addition, this research points out the main issues to be considered when building multilingual ontologies.

Implementing combined text processing and semantic analysing tools is expected to help in expediting ontology creation. In addition, building multilingual ontologies in the industrial domain is a promising field for Webservices in the industrial domain. As interoperability is a main factor for successful Webservice communications and orchestration, domain specific ontology can be used for semantic data communication. In addition, multi-lingual ontologies are expected to improve contextual awareness of industrial global transactions.

ACKNOWLEDGEMENTS

“This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.”

REFERENCES

[1] Abusalah, M. et al.: Arabic-English Automatic Ontology Mapping Based on Machine Readable Dictionary. In: r2009 Mexican International Conference on Computer Science. pp. 367–372 IEEE (2009).
[2] Alnahdi, A. et al.: Building Bilingual Algorithm English-Arabic Ontology For Cognitive Applications. In: 2017 IEEE International Conference on Cognitive Computing (ICCC). pp. 124–127 IEEE (2017).
[3] Antoniou, G., Van Harmelen, F.: A semantic web primer. MIT press (2004).
[4] Bayekeyeva, A. et al.: Controlled Multilingual Thesauri for Kazakh Industry-Specific Terms. Social Inclusion. 9, 1, 35–44 (2021).
[5] Beseiso, M. et al.: A Survey of Arabic language Support in Semantic web. International Journal of Computer Applications. 9, 1, 35–40 (2010).
[6] Black, W. et al.: Introducing the Arabic wordnet project. In: Proceedings of the third international WordNet conference. pp. 295–300 (2006).
[7] Chaware, S., Rao, S.: Ontology supported inference system for Hindi and Marathi. In: 2012 IEEE International Conference on Technology Enhanced Education (ICTEE). pp. 1–6 IEEE (2012).
[8] De Nicola, A. et al.: A proposal for a unified process for ontology building: UPON. In: International Conference on Database and Expert Systems Applications. pp. 655–664 Springer (2005).

[9] De Nicola, A. et al.: A software engineering approach to ontology building. Information systems. 34, 2, 258–275 (2009).
[10] El-Beltagy, S.R., Rafea, A.: KP-Miner: A keyphrase extraction system for English and Arabic documents. Information systems. 34, 1, 132–144 (2009).
[11] Lohmann, S. et al.: WebVOWL: Web-based visualization of ontologies. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 154–158 Springer (2014).
[12] Metzinger, T., Gallese, V.: The emergence of a shared action ontology: Building blocks for a theory. Consciousness and cognition. 12, 4, 549–571 (2003).
[13] Qbadou, M. et al.: Multilingual verbal interaction between humans and robots-Modeling and Implementation. In: 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET). pp. 1–4 IEEE (2020).
[14] Shilov, N. et al.: Ontologies in Smart Manufacturing: Approaches and Research Framework. In: 2020 26th Conference of Open Innovations Association (FRUCT). pp. 408–414 IEEE (2020).
[15] Stvilia, B. et al.: Issues of cross-contextual information quality evaluation—The case of Arabic, English, and Korean Wikipedias. Library & information science research. 31, 4, 232–239 (2009).
[16] "protégé." [Online]. Available: https://protege.stanford.edu/. [Accessed: 20-Jul-2021].

AUTHOR

Dr. Amany Alnahdi is an assistant professor at King Abdulaziz University. She has PhD degree in Computer Science from Kent State University and a Master of Science degree from California State University-Fresno. Her research interests include Semantic Web and Web service computing.