Edu World 2016
7th International Conference

ONTOLOGY LEARNING APPLIED IN EDUCATION: A CASE OF THE NEW TESTAMENT

Ramona Popa (a), Andrei Vasilateanu(b), Maria Goga (c)*

* Corresponding author

(a) Department of Engineering in Foreign Languages, Universitatea Politehnica din Bucuresti, Spaiul Independentei 313, Bucuresti, Romania, popa.ramona91@yahoo.com
(b) Department of Engineering in Foreign Languages, Universitatea Politehnica din Bucuresti, Spaiul Independentei 313, Bucuresti, Romania, andrei.vasilateanu@upb.ro
(c) Departamentul pentru Pregatirea Pesonalului Didactic, Universitatea Tehnica de Constructii Bucuresti, Bd. Lacul Tei nr. 122-124, Bucuresti, Romania, goga.maria@yahoo.com

Abstract

The objective of this work is to show how ontology learning can be applied for the semantic analysis of text, in order to extract concepts, relations, which can be further used for automated summaries or critical comparison. Such activities are very important in education as they can allow dynamic creation of content or analyses that can be further used in the educational process. Since ontology-learning methods require large corpus of unstructured data, we have chosen the Bible as source for the text. Another reason for the choice was that the Bible is one of the most studied texts by scholars and we can test our conclusions, such as importance of some terms, with existing knowledge. In this way, the new developed methods are validated and they can be used successfully in other educational domains. The Bible is the religious text of Christians and Jews. The Bible contains a collection of scriptures that was written by many authors, at different time and locations. Computationally, the Bible contains semi-structured information due to its organized structure of scriptures and numbered chapters. We have used Text2nto as the main tool in order to obtain the most relevant concepts from the New Testament. After that we analyze extracting the most relevant concepts and the range of similarity for each domain identified in the New Testament. Those methods can be employed for automatic generation of content that can be further used in the educational process.

© 2017 Published by Future Academy www.FutureAcademy.org.uk

Keywords: Automatic Ontology Generation; Theological Education; ICT in Education.

1. Introduction

The objective of this work is to show how ontology learning can be applied for the semantic analysis of text, in order to extract concepts, relations, which can be further used for automated summaries or critical comparison. Such activities are very important in education as they can allow dynamic creation of content or analyses that can be further used in the educational process.

In the field of computer science, the ontologies have gained significant importance in last years, due to their applicability for different fields including information extraction. Information extraction in computer science deals with the analyses of unstructured sources of text in order to extract relevant information. The process of information extraction requires statistical analysis, natural language processing techniques and machine learning methods.

Since ontology-learning methods require large corpus of unstructured data, we have chosen as source for the text the Bible. Another reason for the choice was that the Bible is one of the most studied texts by scholars and we can test our conclusions, such as importance of some terms, with existing knowledge. In this way, the new developed methods are validated and they can be used successfully in other educational domains.

The Bible is the religious text of Christians and Jews. The Bible contains a collection of scriptures that was written by many authors, at different time and locations. This book is divided in two parts: The Old Testament and The New Testament. Computationally, the Bible contains semi-structured information due to its organized structure of scriptures and numbered chapters. We have used Text2nto (Cimiano & Volker, 2015) as the main tool in order to obtain the most relevant concepts from the New Testament. After that we analyze extracting the most relevant concepts and the range of similarity for each domain identified in the New Testament. Those methods can be employed for automatic generation of content that can be further used in the educational process - i.e. online learning (Istrate, 2010).

2. Paper Theoretical Foundation and Related Literature

Ontologies are defined as the theory of objects and their ties according to the articles written by (Berners-Lee et al., 2001). In computer science, an ontology is the formal specification of a domain and is composed of properties, types, and interrelationships of entities that should exist for a specific domain according to the articles (Maedche et al., 2003) and (Gruber, 1993).

They provide many principles in order to distinguish different types of objects and their dependencies. Thus, we can distinguish three different types of ontologies: *descriptive,* formal and formalized ontologies.

Sharing information is an important issue for researchers and ontologies represent to them a formal way to store and organize the information. The concepts defined in ontologies can be stored and shared in many formats, the most common ones being OWL/RDF.

Ontologies are used today in order to obtain the most relevant terms for a given domain. We can search for the documents containing their terms over the web. The documents containing the most relevant terms, are the most relevant for a search. Also, documents with a larger number of terms from a given ontology, are more relevant than documents with a small number of terms. Therefore, given an

ontology automatically generated from a bunch of text, one can get the most relevant documents from a web search, documents that contains terms from ontologies (concepts, instances) and use them for education purposes.

2.1 Automatic Extraction of Ontologies in Religious Books

Currently we found only a few articles about searching and information extraction from religious books using ontologies.

According to the article written by (Shamsuzzaman Sadi et al., 2016), people are often exploiting the knowledge of the Quran. Searching mechanisms and extracting information algorithms requires an ontological modeling of the data. In the results section of that research, verses and concepts of interest were retrieved from the Quran using SPARQL queries. Searching for particular concepts inside the book of Quran and retrieving verses that contains those concepts is an important issue for the followers of Islam that are eager in gaining the Quranic knowledge. This article presents a research problem that is not too different for the Bible's New Testament, and other religious texts.

In the article written by (Ahmad et al., 2013), ontologies are also used to represent and encapsulate Quran's knowledge. In the same article a comparison based on a generic framework is done. The comparison is focused on the role of ontologies and how these ontologies are compared to other ontology applications in other areas.

When it comes to the Bible we found far less relevant research articles. A research for automatically transforming Old Testament texts in Hebrew into English-based conceptual graphs was made in an article by (Ulrik, 2007). The method utilizes ontology of the text plus syntactic analysis and transformational rules. The results presented are interesting but not very relevant for our research.

In conclusion, this research is relevant because no other studies were made in order to search and extract information from the Bible's New Testament. The framework used in this research in order to exploit the knowledge from the New Testament helps us to extract not only the most relevant concepts but also the ranges of similarities for the concepts extracted.

3. Methodology

Domain ontologies are very hard to develop manually and a better approach is to generate them automatically. The platform Text2Onto is an important part of our ontology generator (Popa et al., 2016; Vasilățeanu et al., 2015). In order to obtain the most relevant terms for our ontology we made some improvements to this platform. First an algorithm was added to analyse similarity in terms of concepts that are synonyms to another. Second, new rules for ontology generation were added to Text2Onto. The last improvement was automatizing the process of finding similar concepts between the two types of ontologies.

The list containing the most important concepts was obtained by drawing the information from the New Testament using Text2Onto. In the obtained list of concepts each term is associated with a score, called relevance. This relevance score is between zero and one. We have chosen for our research, all the concepts with a score that is larger than a given threshold. The threshold was chosen in a dynamic way in function of the domain and the corpus size.

Below a sample of the concepts generated by the Tet2Onto algorithm from New Testament.

Tabel 1. Concepts generated by the Tet2Onto algorithm.

Concepts	
Label	**Relevance**
heart	0.006215399246000
glory	0.006113507455082
kingdom	0.005671976361104
place	0.005570084570186
flesh	0.005400264918656
woman	0.005230445267126
death	0.005230445267126

4. Results

Before presenting the relevant results, we need to discuss some limitations of the method presented in this paper. First one should note that the best scenario was to use the original text in Hebrew/Greek for such an analysis. However, because the tool that we use has only the grammar of the English language, we were restricted to the New Testament text in English.

There are several translations of the text in English. Due to the online availability, we used the one from David Robert Palmer's translation (Palmer, 2015) that has a text that is also compatible with the tool grammar that we use.

Part of the ontology that we obtained is presented in the following figure – some of most important concepts.
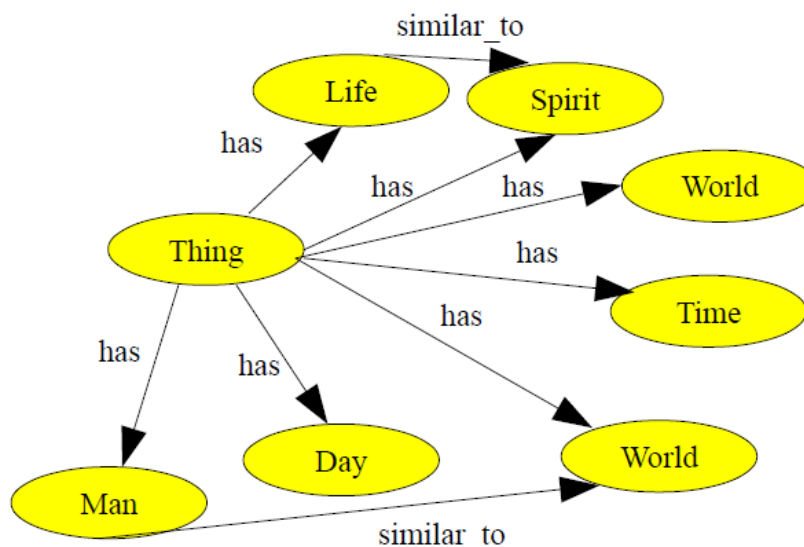
Fig. 2. Most important concepts.

Analyzing the ontology, the following conclusions can be formulated. First there are concepts which have theological relevance such as "*man*", "*brother*", "*spirit*" (of God), "*disciple*", "*heaven*", "*faith*", etc. On the other hand, there are also general concepts that are not theological relevant such as "*time*", "*day*" (in the text with the sense of *day* in general), and "*word*" (common, not *Word of God*). In this respect the concepts of the ontology that are automatically generated should be also analyzed by humans, in a post-generation process, to extract the ones that are theologically relevant.

Looking to the concepts of the ontology, we see that in the first 25 most important ones, the followings are found that are theological relevant:

- "*man*" - similar also with "*world*" (according to the process of automatic analysis)

- "*brother*"

- "*people*"

- "*spirit*" (of God) – similar with "*life*" (according to the process of automatic analysis)

- "*heaven*"

- "*faith*"

- "*world*" – similar with "*man*", "*earth*" (according to the process of automatic analysis)

- "*law*" (of God) – similar with "*life*" (according to the process of automatic analysis)

- "*sin*"

- "*angel*"

- "*glory*"

- "*kingdom*"

Looking to the generated concepts from the New Testament (NT) we find that:

a) *on the human part*: 1) they speak about this world ("*man*", "*people*", "*earth*"); 2) about the existing "*sin*" and "*law*" (of God); 3) about the "*faith*" and "*kingdom*". From a theological perspective, this can be correlated with the message of NT that this *world* is under the *sin*, and through *faith*, one can reach the *kingdom* of heaven – concepts present also in *The Orthodox Nicene Creed* (381), *Athanasian Creed*

(293-373), etc. It is also interesting to observe the existence of the concept of "*brother*" in the first three generated concepts, which shows that the early Christians see each others as brothers and also that among them this concept was of importance.

b) *on heaven side*: 1) they speak about "*spirit*" (of God), "*kingdom*", "*angels*", "*heaven*" which from a theological perspective this can be correlated with the coming *kingdom* of *heaven* and the work of *angels* and *spirit* of God. One can note that some important concepts of NT are missing such as "*God*". "*Jesus Christ*". The reason for this is the fact that in the automatic generation of the concepts proper noun are eliminated (as well as conjunctions, articles, etc.). This fact – elimination of proper nouns - which is suitable in other domains is not suitable for theology. This is a difference between the theological domain and other domains that should be taken into account when applying this method.

5. Discussions

From the analysis performed (see the discussion in the section of results) the followings can be deducted:

- Such methods are good to generate ontology in a given domain. Looking to the generated concepts one can get a grasp of some important terms and concepts that play a role in a given domain. With respect to the automatic generation of content for learning, those concepts can be: 1) keywords that can be used for content searching and 2) used for weighting of different texts that contains them for selecting the most relevant ones as content for e-learning.

- However, the method itself has limitations. We discussed some of them. First we note that the best was to use original Greek/Hebrew text in place of English, but we were restricted by the existence of English grammar used by the tool. Second, the tool generated only part of the existing important concepts. By the elimination of proper nouns, tool eliminates also important concepts such as "God", "Jesus Christ" that are central to the New Testament theology. Finally, such analysis cannot reveal things such as Trinity for example.

As a conclusion, such methods are important and useful, but, as this case study reveals, some human analysis in a post-generation phase might be needed for improvement of the ontology.

6. Conclusions

The objective of this work is to show how ontology learning can be applied for the semantic analysis of text, in order to extract concepts, relations, which can be further used for automated summaries or critical comparison. Such activities are very important in education as they can allow dynamic creation of content or analyses that can be further used in the educational process.

Since ontology-learning methods require large corpus of unstructured data, we have chosen as source for the text the Bible. Another reason for the choice was that the Bible is one of the most studied texts by scholars and we can test our conclusions, such as importance of some terms, with existing knowledge. In this way, the new developed methods are validated and they can be used successfully in other educational domains.

The Bible is the religious text of Christians and Jews. The Bible contains a collection of scriptures that was written by many authors, at different time and locations. Computationally, the Bible contains semi-structured information due to its organized structure of scriptures and numbered chapters. We have used Text2nto as the main tool in order to obtain the most relevant concepts from the New Testament. After that we analyze extracting the most relevant concepts and the range of similarity for each domain identified in the New Testament.

From the analysis performed, it results that such methods are good to generate ontology in a given domain. Looking to the generated concepts one can get a grasp of some important terms and concepts that play a role in a given domain. From automatic generation of content for learning, they can be keywords that can be used for content searching and also automatic weighting of different text that contains them.

However, the method itself has limitations. We discussed some of them. First we note that the best was to use original Greek/Hebrew text in place of English, but we were restricted by the existence of grammar used by the tool. Second, the tool generated only part of existing important concepts. By the elimination of proper nouns, tool eliminates also important concepts such as "God", "Jesus Christ" that are central to the New Testament theology. Finally, such analysis can not reveal things such are Trinity for example.

As a conclusion, such methods are important and useful, but, as this case study reveals, some human analysis in a post-generation phase might be needed for improvement of the ontology.

References

Ahmad, O., Hyder, I., Iqbal, R., Azmi Murad, M. A., Mustapha, A., Sharef, N. M., Mansoor, M. (2013). A survey of searching and information extraction on a classical text using ontology based semantics modeling: A case of Quran, *Life Science Journal*. Retrieved from http://www.lifesciencesite.com/lsj/life1004/181_19656life1004_1370_1377.pdf

Athanasian Creed (293-373) Retrieved from https://www.urcna.org/urcna/Creeds/Athanasian%20Creed.pdf

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The semantic web. Scientific american, vol. 284, no. 5, pp. 28-37.

Cimiano, P., V¨olker, J. (2015). Text2Onto A framework for ontology learning and data-driven change discovery. Retrieved from http://www.dfki.de/~horacek/text2onto.pdf

Gruber, T. R. (1993). A translation approach to portable ontology specifications, in knowledge acquisition, vol. 5 no. 2, pp: 199-220. Retrieved from http://tomgruber.org/writing/ontolingua-kaj-1993.pdf

Istrate, O. (2010). Efecte și rezultate ale utilizării TIC în educație, Retrieved from https://www.academia.edu/3356979/Efecte_si_rezultate_ale_utiliz%C4%83rii_TIC_%C3%AEn_educatie

Maedche, A., Neumann, G. and Staab, S. (2002). Bootstrapping an ontology-based information extraction system, Intelligent exploration of the web. Physica-Verlag HD, pp. 345-359. Retrieved from ftp://ftp.dfki.de/papers/local/bootstrapping.pdf

Palmer, D. R. translation. (2015). *Holly Bible - New Testament.*

Popa, R. C., Vasilateanu, A., Goga, N. (2016). Ontology based multi-system for SME knowledge workers, *IEEE ISSE, Edinburg*. Retrieved from https://www.researchgate.net/publication/310808122_Ontology_based_multi-system_for_SME_knowledge_workers

Shamsuzzaman Sadi, A. B. M., Towfique, A., Mohamed, A., Abdillahi, H. A., Sazid, Z. K., Mohamed, M. R., Ghassan, S. (2016). Applying ontological modeling on quranic nature domain published at

the International Conference on Information and Communication Systems, *7th International Conference on Information and Communication Systems (ICICS)*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1604/1604.03318.pdf

The Orthodox The Nicene Creed (381).

Ulrik, S. P. (2007). Genesis 1:1-3 in Graphs: Extracting conceptual structures from biblical hebrew. Retrieved from https://www.researchgate.net/publication/228734851_Genesis_1_1-3_in_Graphs_Extracting_Conceptual_Structures_from_Biblical_Hebrew

Vasilățeanu, A., Goga, N., Tanase, E. A., Marin, I. (2015). Enterprise domain ontology learning from web-based corpus, *6th ICCCNT 2015 Conference Proceedings, USA, pg. 112-117*. Retrieved from https://www.researchgate.net/publication/281649559_Enterprise_Domain_Ontology_Learning_From_Web-Based_Corpus