

Extracting Knowledge from the Bible: A Comparison between the Old and the New Testament

Ramona Cristina Popa
Department of Engineering in Foreign Languages
University Politehnica of Bucharest
Bucharest, Romania
popa.ramona91@yahoo.com

Nicolae Goga
Faculty of Natural Sciences
University of Groningen
The Netherlands
n.goga@rug.nl

Maria Goga
Departamentul pentru Pregătirea Personalului Didactic
Universitatea Tehnica de Construcții
București
Bucharest, Romania
goga.maria@yahoo.com

Abstract—The objective of this work is to present a comparison between the Old Testament and the New Testament in terms of knowledge extraction and ontology learning. It is a known fact that these two books have many differences in terms of size, practice of worship, prophecy and there is also a difference in the time period when they were written. By applying the ontology learning and knowledge extraction methods we were interested to see if these differences are revealed and what are the similarities among them. 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 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. 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 Text2Onto as the main tool in order to obtain the most relevant concepts from the New Testament and then from The Old Testament. After that we analyze the most relevant concepts and the range of similarity for each domain identified in the New Testament and in The Old Testament. We can mention that there are no studies reported in the literature using ontology extraction for this religious domain. Those methods can be employed for automatic generation of content that can be further used in the educational process.

Keywords—Automatic Ontology Generation, corpus, data analyses, knowledge extraction

I. INTRODUCTION

In our days, ontologies are playing an important role in the field of computer science. Many computer science domains, including software engineering and knowledge extraction, are using ontologies in order to organize and share information in a semantic way. Since the first study that we have conducted [1] on the New Testament proved the importance of knowledge extraction methodologies from the Biblical texts, we have decided to continue our research by analyzing the text from the

Old Testament. Also, the research reported in this article is an example of how automatic generation of concepts through ontology generation can be used for dynamically constructing content that can be further used in computer science or other engineering processes.

The Old Testament has the same structure as the New Testament. The information is semi-structured, organized as a collection of scriptures that were written across multiple years. The Bible has many authors and it was written in different places around the world such as Israel, Babylon, etc.

While the Bible is one collection of books with one unique message, there are numerous differences between the Old Testament and New Testament. But there are also similarities. We wanted to analyze those similarities and differences.

There is a difference in time period. The Old Testament books were written from the time of Moses until about 400 BC. This approximately 1,000-year period covers a wide range of events from the creation of the universe to the return of the Jewish people to Jerusalem under Ezra and Nehemiah. The New Testament was written between approximately AD 50 and AD 150.

There is difference in focus. The Old Testament chronicles God's power and works, looking toward the coming Messiah. The New Testament reveals Jesus as the Messiah, looking back at His life and teachings as the basis for the Church and the spread of the Gospel.

There is a difference in prophecy. Most Old Testament prophecies look forward to the future and were unfulfilled at the time of the Old Testament's completion. The New Testament highlights the fulfillment of many of the Old Testament's prophecies, including more than 300 in the life of Jesus Christ.

There is a difference in the practice of worship. Much of the Old Testament focused on the tabernacle or temple as the central place of worship. Many details were included in the development of offerings, festivals, and related practices. In the New Testament, Jesus offered Himself as the focus of worship,

claiming to be the way, the truth, and the life, and that no one could come to the Father except through Him (John 14:6).

There is a difference in the covenant. The Old Testament emphasized the Mosaic Covenant with its focus on the Law of the Lord. The New Testament began a new covenant in which Jesus came as a fulfillment of the Law. No longer would the Jewish people (or Gentiles) need to fulfill the Law to find atonement; Jesus offered forgiveness and eternal life to all who believed in Him.

In many ways the Old Testament serves as the foundation for the New Testament. While the Old Testament comprises approximately three-fourths of the Bible, its contents point toward the Messiah who is revealed in the New Testament. Rather than an entirely new message, the New Testament reveals how Jesus fulfilled the predictions regarding the Messiah and provides many of His teachings and miracles, along with the writings of His earliest followers.

This article is organized in the following four sections. The next section contains the state of art where the most relevant studies for our research were analyzed. Section three contains the theoretical foundation and the automatic extraction of ontologies in religious books. Since it was an important part of our research, the same section describes the methodology used to obtain the most Section four contains the results and the validation methods. The last section contains the conclusions that we draw based on this research.

II. STATE OF ART

Up to now we found some articles that are relevant for searching and extracting information from religious books [2,3,4]. However most of the studies are exploiting the knowledge of the Quran [5] and only a few articles are using the Bible in order to extract knowledge [6,7] – but not generation of ontologies as in our case.

In our previous paper [1] we have developed a meta-ontology using the innovational process of automatic generation of ontologies from unstructured text and we have presented the steps that were followed for this automatic process together with the results that were obtained while extracting the knowledge from the New testament. In the mentioned research we took into consideration only the most important terms with a biblical meaning. A sample of our final ontology was presented among with the relevance range of the concepts that were used to build the meta-ontology.

In [3] the authors are describing a statistical and numerical analyses on the Bible, based on the frequency of occurrence of some terms in the Bible and the location of those concepts within the text. The study is relevant for determining the fractal aspect of the Bible, and the research is focused on Luke and

Amos Books. According to the study many tests were made on each section of Luke and Amos Books. The experimental results showed the fractal structure of the Luke and Amos biblical books. Finding the fractal structure helps computer science engineers to show that the biblical books are organized even when they seem to be chaotic.

According to [4] the Bible is also used in order to determine its fractal properties and analyze the specific style characteristics of each author that has contributed in writing the Bible.

We have to mention that the papers [5] and [8] are relevant for the process on extracting knowledge and automatically generating ontologies using religious books even though they are using the book of Quran as text.

The main tasks in knowledge extraction and ontology learning are described in the article [9,10,11,12]. The concepts, relations among concepts and individuals are extracted from a corpus of unstructured text using Natural Language Processing and Machine Learning algorithms. Today, there are many software products that are capable of extracting knowledge, but most of them are developed for industrial use and only a few for research purposes one of the most developed one being Text2Onto [13] – that we are using it in this article.

Since a large amount of work is needed in order to develop domain ontologies manually, computer science engineers are researching different approaches to generate ontologies automatically.

Text2Onto framework is capable of parsing a large amount of unstructured text and it runs Machine Learning (ML) and Natural Language Processing (NLP) algorithms for knowledge extraction in order to generate the most important terms for a given domain. Extracting information from the Old Testament is our first step in the process of the automatic generation of ontologies for the biblical domain.

The system Text2Onto has three important components:

- A Natural Language Processing engine
- Algorithm Controller
- Probabilistic ontology model (POM)

The POM component stores our special entities. POM primitives are used in order to classify the ontology objects as follows:

- concepts that are represented by class
- concept inheritance is represented by subclass_of
- concept instantiation represented by instance_of
- relations represented by relation
- domain restrictions represented by domain
- equivalence.

Text2Onto uses algorithms for concept inheritance, mereological (part of) relations, general relations, concept instantiation and equivalence. The results obtained when running the tool Text2Onto using a corpus can be exported in different formats. The most popular formats are RDF, KAON and OWL.

Due to the fact that manual ontology creation is expensive, many learning tools were developed. We have chosen to use the framework text2onto and enhanced it with three algorithms:

- First, we added new rules for the ontology generation
- An automatic process for finding similar terms between two types of ontologies.
- We added an algorithm that analyzes the similarity between concepts by taking into account the synonyms of each concept.

III. PAPER RATIONALE

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).

A. Paper theoretical foundation and Ontologies:

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.

B. 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 (Petersen, 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.

C. 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; Vasilăţeanu 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 analyze 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 and Old 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.

D. Limitations

Before presenting the relevant results, we need to discuss some limitations of the method presented in this paper.

First, we 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.

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.

IV. RESULTS AND VALIDATION

Multiple steps were required for the process of automatically generate an ontology for the biblical domain using the New Testament and Old Testament as corpus.

The first step was based on Text2Onto framework. By running Text2Onto using the Old Testament as corpus we obtained the initial list of concepts. The following sample contains a part of our generated terms:

New Testament		Old Testament	
Label	Relevance	Label	Relevance
heart	0.006215399246000	god	0.552156469
glory	0.006113507455082	earth	0.545135406
kingdom	0.005671976361104	man	0.542126379
place	0.005570084570186	day	0.537111334
flesh	0.005400264918656	year	0.531009328
woman	0.005230445267126	heaven	0.318054162
death	0.005230445267126	life	0.317354167
life	0.005230445267089	wife	0.317051153
spirit	0.005230445061077	bird	0.316862188

Fig. 1. Text2Onto generated concepts

For our research we took into account only the concepts with the highest grade of relevance. The relevance is a positive number with values between zero and one. Our final ontology contains only those terms having the relevance score higher than a given threshold. The threshold was calculated by taking into account the corpus size and the domain of our ontology.

Below we present a sample of the ontologies that were obtained. The first ontology contains the concepts that were generated for the New Testament and the second one contains the concepts that were obtained from the Old Testament.

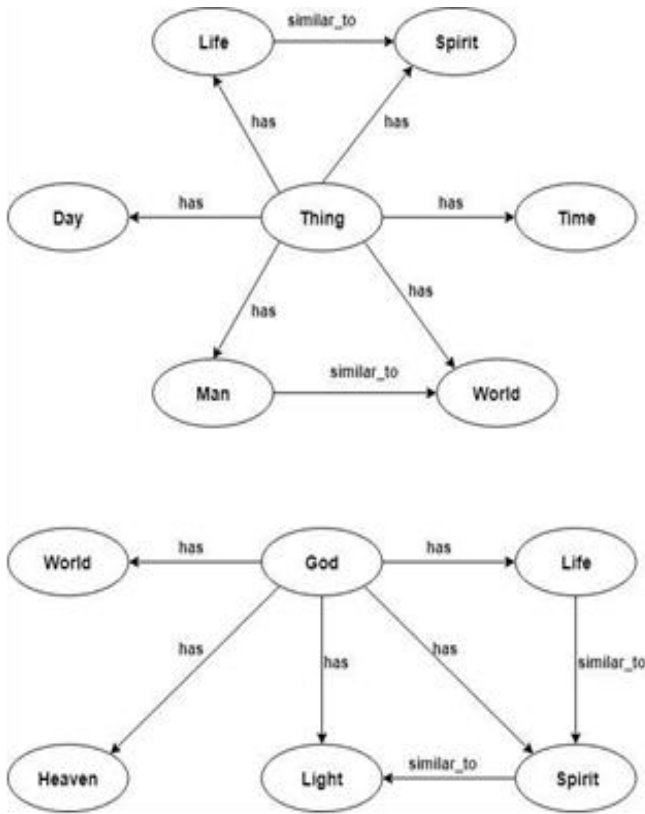


Fig. 2. Ontology sample

However, as discussed above, the method itself has limitations. Some of them were already presented.

From the terms that were generated, we can have a look to the theological meaning of the most important concepts for the Old Testament and New Testament.

Differences and similarities between Old and New Testament:

First it is worth to note that while in the Old Testament the most important term is the concept of God (Old Testament can be said to be centered on the revelation of God on earth – earth is the second term as importance generated by the Ontology), in the New Testament the most important term is Man which shows a center on the Man as God the Savior – through Jesus Christ who become the brother of the mankind – the second term as importance is brother for the New Testament – and mankind in general. However, although the first most important concepts of the New Testament shows a centering on the people from earth (the first most important are: man, brother, people), on the next positions from the Old Testament we find also earth and man which speaks about the activity of God that is for people on earth – in this way the two Testaments are similar, but the emphasis are different – New Testament on Man while the Old Testament on God. A next emphasis both on Old and New Testament is on heaven (which in the New Testament is on the fifth position while on the Old Testament is on the

seventh position). This is an important theological concept for the whole Bible and represents the final goal of the salvation plan. In the Old Testament the fourth and fifth position are occupied by: day and year. As noted in [13] this is because it seems that Text2Onto does not distinguish between the normal day and year (as normal meaning of time that is used in the narrative text in the Bible) and prophetic day and year that are also used in the Old Testament (Ezekiel 15, Daniel 9, etc.) which pinpoint most of the time for the prophecies regarding Messiah – day of his first coming (on the Earth as Jesus Christ – God the Savior) or second coming for the resurrection. We can conclude that the next emphasis on the Old Testament is on prophetic times. In the New Testament after the emphasis on heaven, interestingly comes the notion of law and sin (which can be defined as a transgression of the law of God), eighth and ninth positions that speaks that the New Testament place a good importance on the sin and Law (of God) that are important notions in the salvation of people.

We will stop with our analysis here. Similar analysis can be done for the rest of the generated terms. As a conclusion, the tool generates core concepts that are in agreement with the broad meaning of the text from Old Testament and New Testament but also pinpoints about special similar or different emphasis for the Two Testament. However, currently the tool cannot distinguish yet the difference between normal meaning of some concepts and places in which they are used in prophetic context.

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.

V. 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 cannot 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

- [1] Ramona Cristina Popa, Andrei Vasileanu, Nicolae Goga, “Ontology Learning Applied in Education: A Case of the New Testament”, ISI, Edu World 2016- 7th International Conference, 2016
- [2] S. P. Ulrik., “Genesis 1:1-3 in Graphs: Extracting conceptual structures from biblical Hebrew.”, 2007, Retrieved from https://www.researchgate.net/publication/228734851_Genesis_1_1_3_in_Graphs_Extracting_Conceptual_Structures_from_Biblical_Hebrew
- [3] Roland MEYNET, Andrei Doncescu. “The fractal structure of Biblical Books. A mathematical model explaining and formalizing the ‘chaotic’ structures of books of the Bible through the concepts of Biblical and Semitic Rhetoric.”
- [4] Roland MEYNET, Andrei Doncescu, “Propriétés fractales des structures issues de la rhétorique biblique et sémitique : premiers exemples”
- [5] O. Ahmad, I. Hyder, R. Iqbal, Azmi Murad, M. A. Mustapha, A. Sharef, N. M. Mansoor. “A survey of searching and information extraction on a classical text using ontology-based semantics modeling: A case of Quran”, Life Science Journal, 2013, Retrieved from http://www.lifesciencesite.com/ljsj/life1004/181_19656life1004_1370_1377.pdf
- [6] Ramona Cristina Popa., Andrei Vasileanu., Nicolae Goga, “Ontology based multi-system for SME knowledge workers”, IEEE ISSE, Edinburg, 2016 Retrieved from https://www.researchgate.net/publication/310808-122_Ontology_based_multisystem_for_SME_knowledge_workers
- [7] A. Maedche, G. Neumann, and S. Staab. “Bootstrapping an ontology-based information extraction system,” Intelligent exploration of the web. Physica-Verlag HD, pp. 345-359, 2003.
- [8] Shamsuzzaman Sadi, A. B. M. Towfique, A. Mohamed, H. A. Abdillahi, Z. K. Sazid, R. M. Mohamed, S. Ghassan, “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). 2016 Retrieved from <https://arxiv.org/ftp/arxiv/papers/1604/1604.03318.pdf>
- [9] Mike Uschold, Michael Gruninger Ontologies – “Principles Methods and Applications”, 1996
- [10] P. Cimiano, J. Völker, “Text2Onto,” Natural language processing and information systems, Springer Berlin Heidelberg, pp. 227-238, 2005.
- [11] Ramona Popa, Nicolae Goga, Andrei Doncescu, “Extracting knowledge from the Old Testament: A semantic approach analyses used in education”, ISI, 11th annual International Conference of Education, Research and Innovation Seville (Spain). 12th - 14th of November, 2018
- [12] Ramona Popa, Andrei Vasileanu, Maria Goga, "Ontology Learning Applied In Education: A Case Of The New", IEEE, Edu World 2016-7th International Conference, Retrieved from: <https://www.futureacademy.org.uk/files/images/upload/127.%20EduWorldF%202017.pdf>
- [13] P. Cimiano, J. Völker, “Text2Onto - A Framework for Ontology Learning and Data-driven Change Discovery”, 2005