Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 159 (2019) 1614–1628

Procedia Computer Science

www.elsevier.com/locate/procedia

23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Knowledge Repository of Ontology Learning Tools from Text

Agnieszka Konys*

West-Pomeranian University of Technology in Szczecin, Faculty of Computer Science and Information Technology, Żołnierska 49, 71-210 Szczecin, Poland

Abstract

Ontologies are one of the fundamental elements of the Semantic Web, and they have gained a lot of popularity and recognition because they are viewed as the answer to the need for interoperable semantics in modern information systems. The intermingling of techniques in areas such as natural language processing, information retrieval, machine learning, data mining, and knowledge representation provide a lot of possibilities for development of ontology learning approaches. A rise in focus on the ability to cope with the scale of Web data required for ontology learning forces the potential growth of cross-language research, emphasizing the automatic or semi-automatic generation of the tools dedicated to text mining and information extraction. This paper presents the integration of ontology learning tools from text in the knowledge repository to incorporate the applied techniques and outputs of an ontology learning algorithm into the one complex multifunctional solution. The proposed knowledge repository covers various applicability of existing techniques of learning ontologies from text, and offers competency question-based reasoning mechanism for individuals to specify their profiles of ontology learning tools. The validation stage is also provided in the form of applied reasoning.

Keywords: Ontology learning tools; Learning techniques; Ontology learninng from text; Knowledge repository

1. Introduction

The Web is evolving from a huge information and communication space into a massive knowledge and service repository. Extracting useful information from Web was an erroneous process. To tackle with this problem, the

* Corresponding author. Tel.: +48-91-449-5662; fax: +48-91-449-5662.
E-mail address: akonys@zut.edu.pl

10.1016/j.procs.2019.09.332

concept of Semantic Web was introduced [1]. This concept has marked another stage in the ontology field, emphasizing the ontologies fundamental role to implement the idea of the Semantic Web [2,3]. Plethora of textual information derived from both of the read and write of the Web resources, merged with the increasing demand for ontologies to power the Semantic Web have made semi- and automatic ontology learning from text a very promising research area [3]. For this reason, ontologies are often regarded as the response for the need for interoperable semantics in modern information systems [4,5]. Ontologies offer a formal and structural way of representing the concepts and relations of a shared conceptualization [6]. Knowledge representation and reasoning enables the ontological elements to be formally specified and represented such that new knowledge can be deduced [3]. However, the lack of formally expressed semantics in Web, complemented with the increasing number of information, is one of the main problems of handling knowledge effectively [4,5].

Ontologies are treated as models representing reality [6]. Acquiring domain knowledge for building ontologies requires much time and many resources. To ensure complete reflection of reality, they require frequent updates, making them both costly and difficult to maintain. To overcome this problem, many attentions are put on development of Ontology Learning (OL). Predominantly, OL approaches are based on the Ontology Learning Layer Cake model [3,4]. In this context, they contribute many features, such as statistical based information retrieval, machine learning, and data and text mining, and using linguistics based techniques [2-5,7,8]. Advances in areas such as natural language processing, information retrieval, machine learning, data mining, and knowledge representation have been fundamental for the OL development [4].

Driven by the great potential of ontology learning in ontology development, a knowledge repository of ontology learning tools is proposed. A comparative analysis of selected tools exploiting numerous techniques of various fields (e.g. machine learning, text mining, knowledge representation and reasoning, information retrieval and natural language processing) is being proposed to bring some level of automation in the process of ontology acquisition from unstructured text. It was a basis for further knowledge repository of ontology learning tools elaboration.

This paper is structured as follows. Section 2 presents the main issues related to ontology learning from text. In section 3, selective problems of existing ontology learning techniques are introduced. Next, the current state of the art of ontology learning is shown. To aver this, ontology learning tools from text are presented in detail and a comprehensive analysis of them is highlighted in this field. Based on this, in section 4, the knowledge repository towards the intelligent handling of ontology learning tools from text is provided. It contains a road map of selected OL tools with its features collection. This topic is of particular interest considering that many of the current ontology learning tools employ readily available off-the-shelf tools or incorporate well-known techniques. The knowledge repository mostly focuses on ensuring relevant information on the automatic or semi-automatic generation of ontologies offered by given tools, concomitantly offering competency question-based reasoning mechanism for individuals to specify their profiles of ontology learning tools.

2. Ontology learning from text

The process of acquisition of domain knowledge for building ontologies involves much effort and time and many resources. From a formal point of view, ontologies are effectively formal and explicit specifications in the form of concepts and relations of shared conceptualizations [6]. Ontologies represent the intensional aspect of a domain for governing the way the corresponding knowledge bases (i.e., extensional aspect) [4,9]. Particularly, every knowledge base is engaged to a conceptualization twofold: implicitly or explicitly. This conceptualization exactly reflects ontologies [6]. Undoubtedly, ontology is one of the fundamental cornerstones of the semantic Web [1]. What is more, ontology provides a sound semantic ground of machine-readable description of digital content [4]. The process of adapting ontologies that can operate in the Web resources and cooperate with other systems needs to consider a huge amount of research effort towards developing methodologies and technologies for building and maintaining domain ontologies in a dynamic environment [2,5,8].

Ontology learning aims discovering ontological knowledge from various forms of data automatically or semi-automatically [5]. Generally, ontology learning can be defined as the set of methods and techniques used for building an ontology from scratch, enriching, or adapting an existing ontology in a semi-automatic fashion using several sources [9]. In the literature, learning is interpreted as the process of creating the ontology and populating it [5]. Ontology learning has benefited from the adoption of established techniques such as machine learning, data

mining, natural language processing, information retrieval, and knowledge representation [3,4]. Based on the classification proposed by Alexander Maedche and Steffen Staab [1], ontology learning approaches were distinguished by considering focus on the type of input used for learning. Thus, common classification contains ontology learning from text, from dictionary, from knowledge base, from semi-structured schemata and from relational schemata [3,8]. Each of them demands a number of research activities to reach a common domain conceptualization.

The group of ontology learning methods from texts consists of extracting ontologies by applying natural language analysis techniques to texts. This process derives high-level concepts and relations as well as the occasional axioms from information to form an ontology [4,7]. Fundamental in the development of ontology learning systems have techniques from established fields, such as information retrieval, data mining, and natural language processing [3,9]. Most frequently used techniques are pattern-based extraction, association rules, clustering, ontology pruning, and concept learning [3,4,10]. Generally, ontology learning from text requires identifying terms, concepts, relations, and optionally, axioms from textual information and using them to construct and maintain an ontology [4,10-12].

3. Selective problems of existing ontology learning techniques

Nowadays, diversity of formatted and multi-lingual data as well as lack of fully automatic ontology validation may cause some challenges for ontology learning techniques [1-4,12,13]. In this context, the diversity of Web data has also contributed to the rise of cross-language ontology learning. Some of the proposed approaches are suitable for lightweight ontologies [3,4], however, the role of formal ontology language becomes much more significant, and supposably heavyweight ontologies will take the center stage [7,9,15]. Oftentimes, a human intervention is required for better quality of learned ontologies [3,15-17]. Thus, the involvement of consensus and high-level abstraction requires human cognitive processing [50]. This makes the process of fully automating ontology learning impossible [4]. Another important issue refers to scalability of ontology learning techniques. Extracting knowledge from the growing amounts of data on the Web in different formats requires scalable and efficient approaches [3,5,18-21,47]. To address these challenges, various strategies are currently being developed (e.g. distributed computation for horizontally scaling ontology learning, incremental learning approaches for re-using existing knowledge, or sampling and modularization to improve the efficiency of ontology learning algorithms) [3,4,6,9,13,43,46]. More research efforts will be concentrated on creating new or adapting existing techniques to work with the noise, richness, diversity, and scale of Web data [7,20,40,48,51]. Furthermore, in regard to noise, there is currently little mention of data cleanliness during ontology learning [4]. Though, ontology population will also continue to play a crucial role in taming and structuring the large amount of unstructured data available [10,13,42,49].

Literature review pointed at some missing aspects related to ontology learning from text. Particularly, much attention has been conducted on discovering taxonomic relations [4,5,9], while non-taxonomic relations were given less attention [2,10,11]. To discover relations between concepts (especially fine-grained), much effort is needed. There is also a visible lack of common evaluation platforms for ontologies. Apart from that, increasing of the interest in harnessing the Web to address the knowledge acquisition bottleneck is clearly visible and also making ontology learning to be operational on a Web-scale [13,14]. Moreover, a rise in interest for constructing very large text corpora from the Web is noticeable [4]. This issue requires the design of new techniques for exploiting the structural richness of collaboratively maintained Web data [7,10]. Another important issue is ensuring formal correctness, completeness and consistency of an ontology as assessment criteria of quality standard [3,4]. Ontology learning methodologies and the adoption of ontology design patterns can help to further improve the results [22].

3.1. Literature review

This section presents an overview of the most relevant ontology learning tools. Based on the investigated literature [3-5,9-11,13], a variety of approaches have been applied to ontology learning. This paper presents more extended version of previously elaborated work [3]. In this paper, selected tools were analysed: Acquisition of SemantIc knowledge Using Machine learning methods (ASIUM) [17], Caméléon [18], CORPORUM-Ontobuilder [19], DOE: Differential Ontology Editor [20], KEA: Keyphrases Extraction Algorithm [21], LTG (Language Technology Group) Text Processing Workbench [23], Mo'K Workbench [24], OntoLearn Tool [25], Prométhée

[26], SOAT: a Semi-Automatic Domain Ontology Acquisition Tool [27], SubWordNet Engineering Process tool [28], SVETLAN [29], TFIDF-based term classification system [30], TERMINAE [31], Text-To-Onto [32], TextStorm and Clouds [33], Welkin [34], WOLFIE (WOrd Learning From Interpreted Examples) [35], SYNDIKATE [36], CRCTOL (concept-relation-concept tuple-based ontology learning) [37], OntoGain [38], and TermExtractor [39]. These tools have some similar features, however there are some differences.

ASIUM aims to learn semantic knowledge from texts and use the knowledge for the further expansion. It exploits linguistics and statistics-based techniques to perform its ontology learning tasks. ASIUM is a semi-automated ontology learning system that is part of an information extraction infrastructure called INTEX [17]. The reference to linguistic-based approach is also presented in TERMINAE tool, which helps in integration of linguistic tools and knowledge engineering tools. In TERMINAE, the linguistic tool allows defining terminological forms from the analysis of term occurrences in a corpus [31]. The analysis is based on the term in the corpus to define the meanings of the terms. The knowledge engineering tool involves an editor and a browser for the ontology. Another tool, Caméléon, assists the learning of conceptual relations to enrich conceptual models for the REX Knowledge management System [18].

Further, the ontology extraction using NLP is applied in CORPORUM-Ontobuilder, which extracts information from structured and unstructured documents using the tools named OntoWrappper [19]. Other way is shown by KEA, which automatically extracts key-phrases from the full text of documents [21]. Machine learning approach is used by Prométhée, where the tool offers the extraction and refinement of lexical-syntactic patterns relative to conceptual specific relations from technical corpora [26]. The patterns bases used in Prométhée enable enriching the extracted items in the learning. Clustering approach is applied in SVETLAN, a domain-independent tool enabling creating clusters from words appearing on texts [29].

Some tools offer support to detect relevant domain terms and to learn from relations that hold among them (e.g. TFIDF-based term classification system) [30]. Semi-automatic domain ontology acquisition from a domain corpus is offered by SOAT, which aims to extract relationships from parsed sentences based on applying phrase-rules to identify keywords with strong semantic links like hyperonym or synonym [27]. Semi-automatically constructing a semantic network is proposed by TextStorm and Clouds. The aim of this approach is to build and refine domain ontology by using logic and linguistics-based techniques to perform its ontology learning tasks [33]. To compare, SYNDIKATE provides a stand-alone automated ontology learning system, and uses only linguistics-based techniques to perform its ontology learning tasks [36]. Welkin offers automatically generating e-learning materials from unrestricted texts either [34]. Likewise OntoGain is designed for the unsupervised acquisition of ontologies from unstructured text [38]. Then, TermExtractor uses a sentence parser to parse texts and extract syntactic structures [39].

Mo'K Workbench is a configurable workbench that supports the semiautomatic construction of ontologies from a corpus using different conceptual clustering methods [24]. Next, LGT is a set of computational tools for uncovering internal structure in natural language texts written in English. The main idea behind the workbench is the independence of the text representation and text analysis [23]. SubWordNet Engineering Process tool provides the architecture to interactively acquire and maintain sublanguage. The applied architecture builds upon WordNet semantic structure and contains integrated capabilities for concept element discovery, concept identification, and concept maintenance [23]. Moreover, WOLFIE learns a semantic lexicon from a corpus of sentences paired with representations of their meaning, where the lexicon learned is composed of words paired with representations of their meaning, and allows for both synonymy and polysemy [35]. The construction of the ontologies from domain-specific documents is yielded by CRCTOL, which uses linguistics and statistics-based techniques to perform its ontology learning tasks [37].

Overall, some ontology learning approaches depend more strongly on external knowledge than others. Lots of functions need to be supported by ontology tools and technologies. For example, some functions can facilitate ontology development, especially ensuring knowledge elicitation by capturing domain knowledge from users directly, and also ontology retrieval, exploiting searching based on terms, relations and other forms [44,45]. Another important aspect encompasses ontology validation and evaluation using expert knowledge or user intervention [44]. Key issues refer to ensure collaborative support in the form of methodology, concurrency control and security mechanisms as well as further ontology development.

3.2. Comparative analysis

The selected tools vary significantly depending on goal and scope. For example, some tools concentrate in finding taxonomic relations (ASIUM), Text-To-Onto) [17,32] or finding non-taxonomic relations (Text-To-Onto) [32] as well as building a taxonomy focusing only on subclass relations (TextStorm and Clouds) [33]. There are tools referring to learning a semantic lexicon (e.g. WOLFIE) [35] or automatic extension of existing general-purpose ontologies with new terms identified in unrestricted text (Welkin) [34]. Furthermore, enriching a domain ontology with concepts and relations was included in OntoLearn tool [25]. Apart from that, some tools offer discovering internal structure of texts in natural language (e.g. KEA) [21].

Many proven techniques from established fields exist, and they may vary depending on the tasks to be accomplished. The investigated tools adapt distinguished learning techniques from conceptual clustering (ASIUM, Mo'K Workbench, SVETLAN, TERMINAE) [3,17,24,29,31] by learning from patterns (Caméléon, Prométhée, SOAT) [3,18,26,27], using linguistic and semantic techniques (CORPORUM-Ontobuilder, SYNDIKATE, CRCTOL) [19,36,37] and similarities measure (Welkin) [34]. Oftentimes, classical learning techniques such as statistical approach are widely adopted by KEA [21] and LGT [23], as well as machine learning (KEA, OntoLearn Tool) [21,25] and NLP (OntoLearn Tool, SubWordNet Engineering Process tool, TextStorm and Clouds, WOLFIE) [25,28,33,35]. Some tools exploit lexical processing (e.g. KEA) [21] and text mining (e.g. TFIDF-based term classification system, SYNDIKATE, CRCTOL, and TermExtractor) [30,36,37,39]. Besides, association rules (Text-To-Onto) [32], programming algorithms (TextStorm and Clouds) [33] and differential semantics (TextStorm and Clouds) [33] are also practised.

The analysis of the tools according to method followed for ontology learning shows that most frequently used is own method (e.g. Caméléon, CORPORUM-Ontobuilder, LTG, Mo'K Workbench, OntoLearn Tool, Prométhée, SOAT, TFIDF-based term classification system, TERMINAE, Text-To-Onto, TextStorm and Clouds, SYNDIKATE, CRCTOL, OntoGain, TermExtractor) [3,18,19,23-28,31-33,36-39] customarily provided and connected with the tool. In the case of the applied methods followed for ontology learning, such approaches as factorisation (ASIUM) [17], clustering (SVETLAN) [29], hyponymy patterns (WOLFIE) [35], syntactic analysis (SVETLAN) [29], Bachimont's method (DOE) [20] and unsupervised hybrid text-mining approach (TFIDF-based term classification system) [30] are used. However, not every tool is supported by a method (e.g. KEA) [21]. Sometimes the obtained results should be refined or supervised by expert participation (e.g. Caméléon, DOE, KEA, OntoLearn Tool, Prométhée, SOAT, SubWordNet Engineering Process tool, SVETLAN, TFIDF-based term classification system, TERMINAE, Text-To-Onto, Welkin, WOLFIE, SYNDIKATE, CRCTOL, OntoGain, and TermExtractor) [3,18,20,21,25-32,34-39] or user participation (e.g. ASIUM, Caméléon, DOE, KEA, OntoLearn Tool, Prométhée, SOAT, SubWordNet Engineering Process tool, SVETLAN, TFIDF-based term classification system, TERMINAE, Text-To-Onto, Welkin, WOLFIE, SYNDIKATE, CRCTOL, OntoGain, and TermExtractor) [3,17,18,20,21,25-32,34-39]. Nonetheless, in some cases, an expert or a user participation is not necessary (e.g. CORPORUM-Ontobuilder, LTG, Mo'K Workbench, TextStorm and Clouds) [19,23,24,33].

Moreover, the tools vary from types' sources used by method. Most frequently used is text (ASIUM, Caméléon, CORPORUM-Ontobuilder, DOE, KEA, Prométhée, SOAT, SubWordNet Engineering Process tool, SVETLAN, TFIDF-based term classification system, TERMINAE, WOLFIE, SYNDIKATE, CRCTOL, OntoGain, and TermExtractor, LTG, Mo'K Workbench, TextStorm and Clouds) [17-21,23,24,26-31,33,35-39], however pattern bases (OntoLearn Tool) [25] and machine readable dictionaries (Text-To-Onto) [32] and ontologies (Mo'K Workbench, Text-To-Onto, Welkin) [24,32,34] are also applied. Another criterion used in the assessment process was interoperability with other tools. Some tools is based on own tool (e.g. ASIUM, Caméléon, CORPORUM-Ontobuilder, LTG, Mo'K Workbench, TFIDF-based term classification system, TERMINAE, WOLFIE) [17-19,23-29,32-34,36-39], although some of them interoperates with WEKA machine (e.g. KEA) [21]. Most tools do not provide any information about this (e.g. DOE, OntoLearn Tool, Prométhée, SOAT, SubWordNet Engineering Process tool, SVETLAN, Text-To-Onto, TextStorm and Clouds, Welkin, SYNDIKATE, CRCTOL, OntoGain, TermExtractor) [20,25-29,32-34,36-39].

Evaluation is an important aspect of ontology learning, useful to assist users and experts in assessing the obtained results. Evaluation measures can be applied to help in this process. To compare, evaluation of ontology learning is assessed using common metrics such as Precision measure (ASIUM, SYNDIKATE, OntoGain) [17,36,38], F-

measure (OntoLearn Tool, Text-To-Onto, SYNDIKATE, CRCTOL) [25,32,36,37], Recall (SYNDIKATE) [36] and Accuracy measure (Text-To-Onto, TextStorm and Clouds, SYNDIKATE) [32,33,36]. The analysis of learning output results are distinguished in the form of terms (ASIUM, OntoLearn Tool, Text-To-Onto, TextStorm and Clouds, SYNDIKATE, CRCTOL, OntoGain) [17,25,32,33,36-38] concepts (ASIUM, OntoLearn Tool, Text-To-Onto, SYNDIKATE, CRCTOL, OntoGain) [17,25,32,36-38], taxonomic relations (ASIUM, OntoLearn Tool, Text-To-Onto, TextStorm and Clouds, SYNDIKATE, CRCTOL, OntoGain) [17,25,32,33,36-38], non-taxonomic relations (Text-To-Onto, TextStorm and Clouds, SYNDIKATE, CRCTOL, OntoGain) [32,33,36-38], and axioms (TextStorm and Clouds) [33]. Generally, there are five types of output in ontology learning, oftentimes called as ontology learning layer cake [3,5,9]. The final comparative analysis is presented in table included in Appendix A. [3,17-21,23-39].

4. A knowledge repository of ontology learning tools

Based on the previously elaborated analysis in section 3.1 and 3.2, including summarization of selected works done in domain of ontology learning tools from text, the highlights of them were shown in table included in Appendix A. Anyway, ontology learning is a complex and largely domain-oriented process [40,41] and it requires much more paying attention to constructing very large text corpora from the Web [42-44]. In-depth analysis allowed detailing the set of features determining the range of functionalities of selected OL tools. The main focus of the review was to introduce an approach in the form of knowledge repository for comparing ontology learning tools from text. For this purpose, a set of tools was identified and studied: ASIUM Acquisition of SemantIc knowledge Using Machine learning methods, Caméléon, CORPORUM-Ontobuilder, DOE: Differential Ontology Editor, KEA: Keyphrases Extraction Algorithm, LTG (Language Technology Group) Text Processing Workbench, Mo'K Workbench, OntoLearn Tool, Prométhée, SOAT: a Semi-Automatic Domain Ontology Acquisition Tool, SubWordNet Engineering Process tool, SVETLAN', TFIDF-based term classification system, TERMINAE, Text-To-Onto, TextStorm and Clouds, Welkin, WOLFIE (WOrd Learning From Interpreted Examples), SYNDIKATE CRCTOL (concept-relation-concept tuple-based ontology learning), OntoGain, and TermExtractor [3,17-21,23-39]. With the need of knowledge repository construction, the set of criteria was established as follows:

- Goal and scope of the tool – this criterion presents the main assumptions of analysed tools, and it includes the following 17 sub-criteria: finding taxonomic relations, tuning generic lexico-syntactic, extracting an initial ontology and refining, supporting of building an ontology, finding non-taxonomic relations , obtaining concept taxonomy from domain tagged text, extracting key-phrases that represent the content of a document, discovering internal structure of texts in natural language, enriching a domain ontology with concepts and relations, extraction and refinement of lexical-syntactic patterns relative to conceptual specific relations, acquisition of relationships using a predefined knowledge representation framework, building Sublanguage WordNets, building a hierarchy of concepts, learning concepts and relations between them, building an ontology, building a taxonomy focusing only on subclass relations, automatic extension of existing general-purpose ontologies with new terms identified in unrestricted text, and learning a semantic lexicon.
- Learning technique used by the tool – this criterion refers to applied learning technique and it is possible to fulfil by a given tool more than one of the 12 options: conceptual clustering, learning from patterns, linguistic and semantic techniques, differential semantic, statistical approach, machine learning, lexical processing, NLP, text mining, association rules, programing algorithm, and similarities measure.
- Method followed for ontology learning – this criterion considers the used method for ontology learning from the distinguished sub-set as follows: own method, factorisation, clustering, Bachimont's method, semantic interpretation, syntactic analysis, hyponymy patterns, and unsupervised hybrid text-mining approach. In case of the lack of data the option 'not propose' was used.
- User and expert intervention in the process – this criterion describes the fact of necessity of expert and user participation in the process. Thus, these two options were considered. In some cases, the option 'not necessary' was used.
- Types of sources used by the method – this criterion defines the types of sources that were employed by a given approach, especially exploiting text, pattern bases, machine readable dictionaries, and ontologies.
- Interoperability with other tools – this criterion describes the possibilities of interoperability with other

solutions, showing the potential options such as interoperability with WEKA machine or interoperability with own tool. If this option is not distinguished, the lack of data is used.

- Evaluation metrics – this criterion provides the selective options of used evaluation metrics by a given tool, including the most popular measures: Precision measure, F-measure, Recall, and Accuracy measure.
- Output – this criterion defines the expected output provided by a given tool in the form of terms, concepts, taxonomic relations, non-taxonomic relations, and axioms.

The final set of criteria includes 8 main criteria and 57 sub-criteria characterizing the selected OL tools from text in details. Due to the fact that ontology learning is a multidisciplinary task that corresponds with borrowing and using techniques from different domains and extracts important aspects in various forms, the knowledge repository was constructed to integrate and homogenize such data.

4.1. Formalization of knowledge repository

On base of the defined set of criteria and sub-criteria, the semi-formal description using set theory is proposed to present the mathematical background of constructed knowledge repository. The defined sets are input data to knowledge repository.

In the domain and range of a relation, if R is a relation from set Ol and F, then the set of all taxons (all of the first components of the ordered pairs) belonging to R is called the domain of R. Thus, Dom is defined as follows:

$$(R) = \{ol \in Ol: (ol, f) \in R \text{ for some } f \in F\} \quad (1)$$

The set of all second components of the ordered pairs (the set of all taxons) belonging to R is called the range of R. Thus, the range of R is defined as follows:

$$R = \{p \in P: (ol, p) \in R \text{ for some } ol \in T\} \quad (2)$$

If supplier ontology learning tools Ol and features F are two non-empty sets, then the Cartesian product T of Ol and F, denoted Ol x F, is the set of all ordered pairs (ol, f) such that ol $\in$ Ol and f $\in$ F:

$$Ol \times F = \{(ol, f): ol \in Ol, f \in F\} \quad (3)$$

Features F contain the finite set of taxons, defined as follows:

$$F = \{Em, Gs, It, Lt, Ml, Ot, Ts, Ue\} \quad (4)$$

where Em represents evaluation metrics, Gs refers to goal and scope of the tool, It shows interoperability with other tools, Lt defines learning technique used by the tool, Mt depicts method followed for ontology learning, Ot presents output, Ts shows types of sources used by the method, and Ue determines user or expert intervention in the process.

Ontology learning tools Ol contains the sub-set called OL tool with defined finite set of taxons as follows:

$$Ol = \{Ol1, Ol2, Ol3, Ol4, Ol5, Ol6, Ol7, Ol8, Ol9, Ol10, Ol11, Ol12, Ol13, Ol14, Ol15, Ol16, Ol17, Ol18, Ol19, Ol20, Ol21\} \quad (5)$$

where Ol1 presents ASIUM Acquisition of SemantIc knowledge Using Machine learning methods, Ol2 shows Caméléon, Ol3 depicts CORPORUM-Ontobuilder, Ol4 determines DOE: Differential Ontology Editor, Ol5 refers to KEA: Keyphrases Extraction Algorithm, Ol6 presents LTG (Language Technology Group) Text Processing Workbench, Ol7 refers to Mo'K Workbench, Ol8 means OntoLearn Tool, Ol9 shows Prométhée, Ol10 refers to SOAT: a Semi-Automatic Domain Ontology Acquisition Tool, Ol11 depicts SubWordNet Engineering Process tool, Ol12 refers to SVETLAN', Ol13 means TFIDF-based term classification system, Ol14 determines TERMINAE, Ol15 refers to Text-To-Onto, Ol16 means TextStorm and Clouds, Ol17 depicts Welkin, Ol18 refers to WOLFIE (WOrd Learning From Interpreted Examples), Ol19 presents SYNDIKATE CRCTOL (concept-relation-concept tuple-based ontology learning), Ol20 means OntoGain, and Ol21determines TermExtractor.

4.2. Intelligent and interoperable expression of semantic relations in the knowledge repository

The knowledge repository was implemented using the Protégé application. The applied technology standard is OWL (Ontology Web Language). By using the reasoning paradigm of OWL, complex contexts can be deduced in the proposed knowledge repository. The input data was gathered from previously elaborated comparison of OL tools from text, whereas the final result of this work was presented in table included in Appendix A. The proposed knowledge repository describes the terminology of the context model in the form of conceptual classes and relationships between these classes. These classes and relations construct the structure of the underlying knowledge repository model and are stored inside a terminological box (TBox) [8,12]. Instances of conceptual classes, their attributes and relationships are stored inside an assertional box (ABox) [8,12]. Ultimately, the class hierarchy contains 92 elements, axiom counts 417 elements, logical axiom counts 321 elements, and object property counts 2 elements. By utilizing knowledge repository, context information can be described semantically. Figure 1 presents the class hierarchy, whereas figure 2 depicts the analysed OL tools. A schematic view of the description of selected OL tool is given in figure 3.

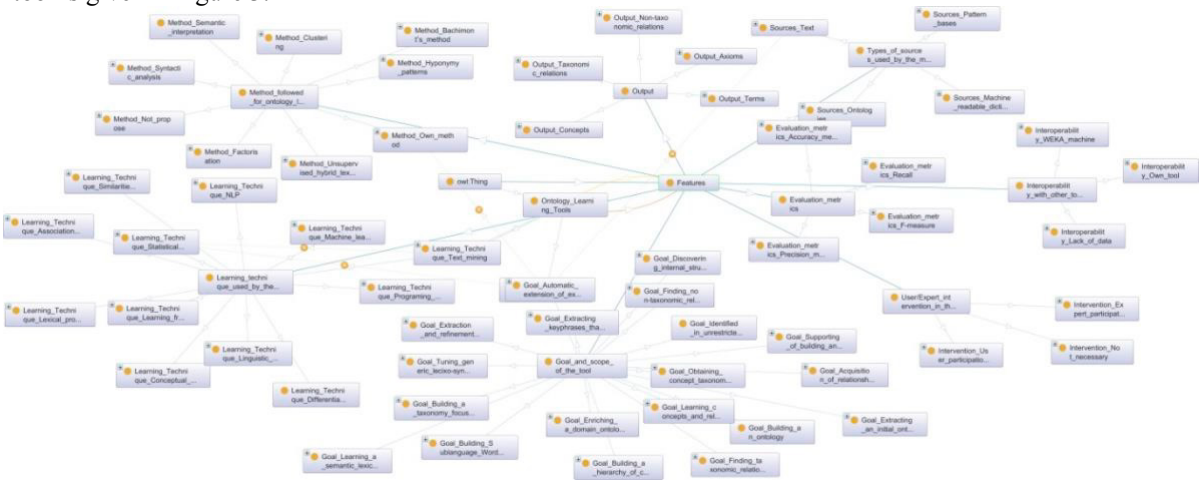

Fig. 1. The class hierarchy of elaborated knowledge repository.

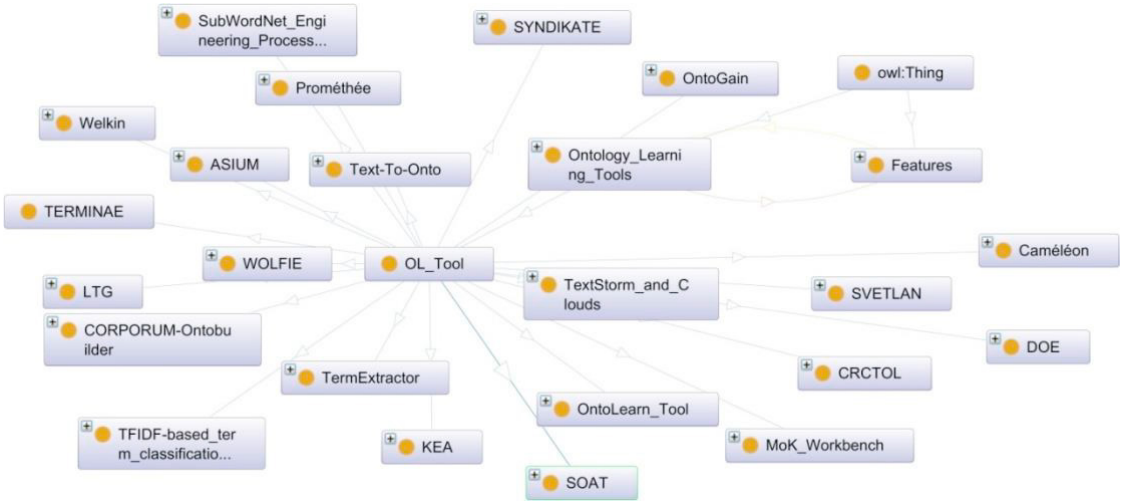

Fig. 2. The set of selected OL tools.

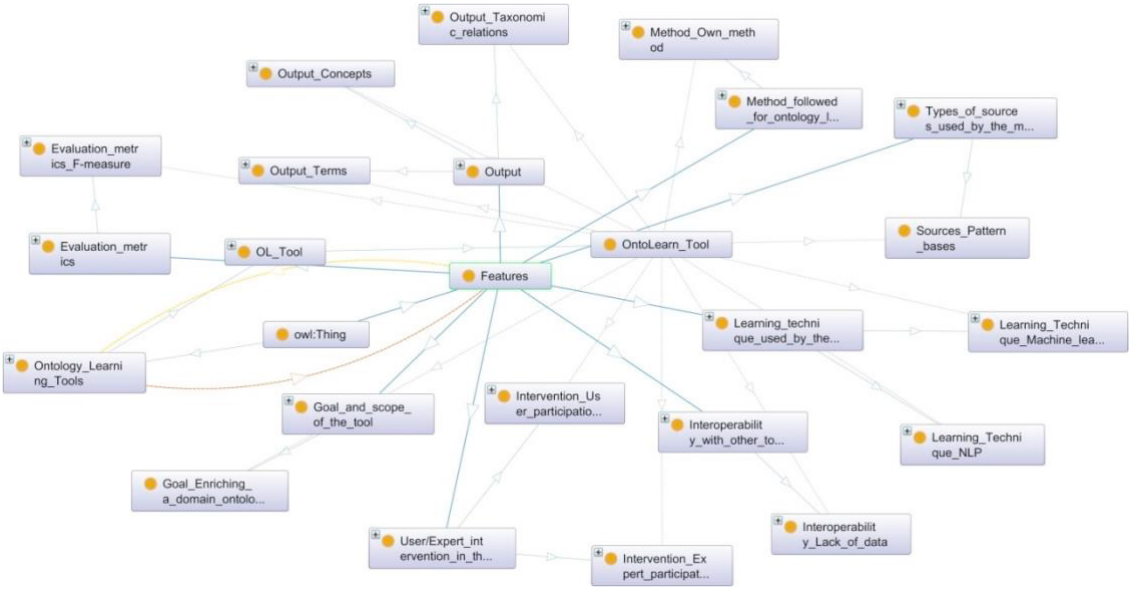

Fig. 3. OntoLearn Tool distinguished features

4.3. Proposed competency question-based reasoning mechanism

The validation process of presented knowledge repository is performed using predefined competency question-based reasoning mechanism. Thus, an interoperable knowledge concept can be refined in order to address a validation stage and consistency checking of elaborated knowledge repository. With the reasoning paradigm of OWL ontologies, the proposed knowledge repository can deduce complex contexts from already aggregated information. Using description logic-based reasoning, the example of competency question demonstrates the operation of implemented knowledge repository. In the fact that gathered knowledge in the repository and reasoning is explicitly integrated, the competency questions can be formulated in a natural way using concepts on appropriate level of abstraction, including additional constraints.

In this way, the exemplary competence question contains the followed constraints to be fulfilled, exploiting as a learning technique statistical approach, applying text as a type of sources used by the method, using own method followed for ontology learning, and demanding expert participation and intervention in the process. All constraints need to be considered by the reasoning mechanism. The competency question was implemented manually to Description Logic query mechanism. By utilizing reasoner HermiT, the knowledge repository can deduce complex contexts from already aggregated context information by using the reasoning paradigm in OWL. The final set contains 2 OL tools that meet these requirements (OntoGain and SubWordNet Engineering Process tool). By applying OWLViz tool implemented in Protégé, the set of results is as figure 4 follows.

Fig. 4. Results of the first competency question presented using OWLViz tool

Additional competency question has been posed to find the relevant applied OL tools from text that meets the following constraints: user intervention in the process, required interoperability with other tools using own tool, and expected output in the form of terms. Optionally, the output may include taxonomic relations and goal and scope of

the tool should consider finding taxonomic relations. Similarly to the first competence question, the definition was formulated and implemented using Description Logic query mechanism. After reasoning process the final set of results was provided as presented on figure 5. The set of results contains 6 OL tools.

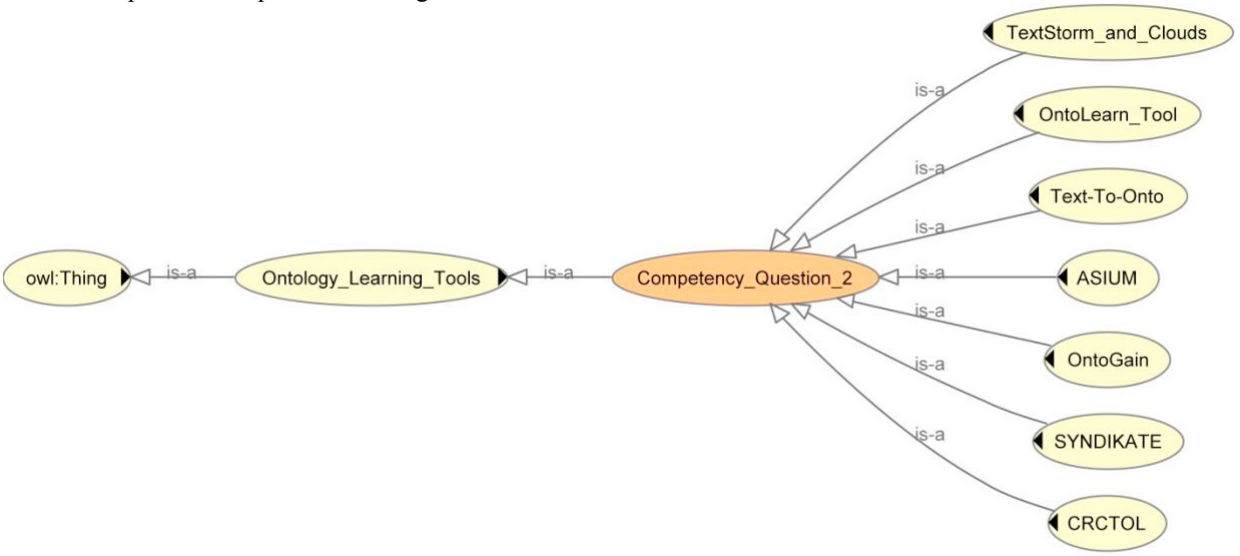

Fig. 5. Results of the second competency question presented using OWLViz tool.

5. Conclusions

Rapid development of advances in areas such as natural language processing, information retrieval, machine learning, data mining, and knowledge representation has been promising field for the ability to cope with the scale of Web data required for ontology learning. Ontology learning refers to the automatic discovery and creation of ontological knowledge using various techniques. However, ontology learning approaches differ among others by type of input, including ontology learning from text, from dictionary, from knowledge base, from semi-structured schemata and from relational schemata. This paper summarizes the first group containing tools dedicated to ontology learning from text, where the main aim consists of extracting ontologies by applying particular learning techniques to texts. Integrated information about the tools is encapsulated in the proposed knowledge repository. The knowledge representation of analyzed tools looks through the results of an ontology learning algorithm including a goal and scope of the tool and types of sources used by the method. Moreover, it evaluates the tools in the context of offered quality of an ontology (ontology evaluation task), as well as considering the process of extracting output (taxonomic and non-taxonomic relations, terms, axioms, and concepts discovery). Apart from that, the analysis included used ontology learning techniques, offered support in the form of method followed for ontology learning and optional interoperability with other tools. The key issue of user and expert intervention in the process was also incorporated to the knowledge repository.

Ultimately, the knowledge repository of ontology learning tools from text embodied 22 tools and their 65 features. They mostly help constructing the automatic or semi-automatic generation of ontologies by means of applied learning algorithm focusing on schematic structures or the data level. The validation of the proposed knowledge repository was checked by competency question-based reasoning mechanism. This work may help in limitation a lot of burden on knowledge engineers and domain experts and playing a crucial role in taming and structuring the large amount of unstructured data available by using dedicated tool. Lastly, individuals can specify their profiles of ontology learning tools with presented knowledge repository.

Due to the fact that it is difficult to find the best tool among all as the performance of ontology learning techniques is highly dependent on efficient preprocessing of data in target domain, the proposed knowledge repository facilitates knowledge transfer and it combines the speed of computers with the accuracy of human factor.

Undoubtedly more research involve the use of Web data for addressing the bottleneck of manual knowledge crafting, and with this need, the knowledge repository can be used to group related resources and provide an enhanced structure for the underlying data considering ontology learning tools from text.

Like any work, there are also shortcomings in the proposed approach. Probably the knowledge repository does not contain all the tools available. Further, the dynamic development of technology requires fast adapting to the design of new techniques for ontology learning. Thus, it imposes some duties for systematic update of the proposed solution. Public availability of the proposed knowledge repository may support these processes.

Appendix A. A comparative analysis of selected OL tools.

Features	Name	ASIUM	Caméléon	CORPORUM-Ontobuilder	DOE	KEA	LTG	Mo'K Workbench	OntoLearn Tool	Prométhée	SOAT	SubWordNet Engineering Process tool
Goal and scope of the tool	finding taxonomic relations	✓										
	tuning generic lecixo-syntactic		✓									
	extracting an initial ontology and refining			✓								
	supporting of building an ontology				✓							
	finding non-taxonomic relations											
	obtaining concept taxonomy from domain tagged text							✓				
	extracting keyphrases that represent the content of a document					✓						
	discovering internal structure of texts in natural language						✓					
	enriching a domain ontology with concepts and relations								✓			
	extraction and refinement of lexical-syntactic patterns relative to conceptual specific relations									✓		
	acquisition of relationships using a predefined knowledge representation framework										✓	
	building Sublanguage WordNets											✓
	building a hierarchy of concepts											
	learning concepts and relations between them											
	building an ontology											
	building a taxonomy focusing only on subclass relations											
	automatic extension of existing general-purpose ontologies with new terms identified in unrestricted text											
	learning a semantic lexicon											
Learning technique used by the tool	conceptual clustering	✓						✓				
	learning from patterns		✓							✓	✓	
	linguistic and semantic techniques			✓								
	differential semantic				✓							
	statistical approach					✓	✓					✓
	machine learning					✓			✓			
	lexical processing					✓						
	NLP								✓			✓
	text mining											
	association rules											
	Programing algorithm											
	Similarities measure											
Method followed for ontology learning	own method		✓	✓			✓	✓	✓	✓	✓	✓
	Factorisation	✓										
	Clustering	✓										
	Bachimont's method				✓							
	not propose					✓						
	semantic interpretation											
	syntactic analysis											
	hyponymy patterns											
	unsupervised hybrid text-mining approach											
User/Expert intervention in the process	not necessary			✓			✓	✓				
	User participation	✓	✓		✓	✓			✓	✓	✓	✓
	Expert participation		✓		✓	✓			✓	✓	✓	✓
Types of sources used by the method	text	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	pattern bases								✓			
	machine readable dictioraries											
	ontologies							✓				
Interoperability with other tools	lack of data				✓				✓	✓	✓	✓
	WEKA machine					✓						
	own tool	✓	✓	✓			✓	✓				
Evaluation metrics	Precision measure	✓										
	F-measure								✓			
	Recall											
	Accuracy measure											
Output	Terms	✓							✓			
	Concepts	✓							✓			
	Taxonomic relations	✓							✓			
	Non-taxonomic relations											
	Axioms											

Features	Name	SVETLAN	TFIDF-based term classification system	TERMINAE	Text-To-Onto	TextStorm and Clouds	Welkin	WOLFIE	SYNDIKATE	CRCTOL	OntoGain	TermExtractor
Goal and scope of the tool	finding taxonomic relations				✓							
	tuning generic lecixo-syntactic											
	extracting an initial ontology and refining											
	supporting of building an ontology											
	finding non-taxonomic relations				✓							
	obtaining concept taxonomy from domain tagged text											
	extracting keyphrases that represent the content of a document											
	discovering internal structure of texts in natural language											
	enriching a domain ontology with concepts and relations											
	extraction and refinement of lexical-syntactic patterns relative to conceptual specific relations											
	acquisition of relationships using a predefined knowledge representation framework											
	building Sublanguage WordNets											
	building a hierarchy of concepts	✓										
	learning concepts and relations between them		✓									
	building an ontology			✓								
	building a taxonomy focusing only on subclass relations					✓						
	automatic extension of existing general-purpose ontologies with new terms identified in unrestricted text						✓					
	learning a semantic lexicon							✓				
Learning technique used by the tool	conceptual clustering	✓		✓								
	learning from patterns											
	linguistic and semantic techniques								✓	✓	✓	
	differential semantic											
	statistical approach		✓		✓			✓			✓	
	machine learning											
	lexical processing											
	NLP					✓		✓				
	text mining		✓						✓	✓		✓
	association rules				✓							
	Programing algorithm					✓						
	Similarities measure						✓					
Method followed for ontology learning	own method			✓	✓	✓			✓	✓	✓	✓
	Factorisation											
	Clustering	✓										
	Bachimont's method											
	not propose											
	semantic interpretation											
	syntactic analysis	✓										
	hyponymy patterns							✓				
	unsupervised hybrid text-mining approach		✓									
User/Expert intervention in the process	not necessary					✓						
	User participation	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
	Expert participation	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
Types of sources used by the method	text	✓	✓	✓		✓		✓	✓	✓	✓	✓
	pattern bases											
	machine readable dictioraries				✓							
	ontologies				✓		✓					
Interoperability with other tools	lack of data	✓			✓	✓	✓		✓	✓	✓	✓
	WEKA machine											
	own tool		✓	✓				✓				
Evaluation metrics	Precision measure								✓		✓	
	F-measure				✓				✓	✓		
	Recall								✓			
	Accuracy measure				✓	✓			✓			
Output	Terms				✓	✓			✓	✓	✓	
	Concepts				✓				✓	✓	✓	
	Taxonomic relations				✓	✓			✓	✓	✓	
	Non-taxonomic relations				✓	✓			✓	✓	✓	
	Axioms					✓						

References

[1] Maedche, Alexander, and Staab, Steffen. (2001) "Ontology Learning for the Semantic Web." *IEEE Intell. Syst.* **16 (2)**: 72 –79.

[2] Uschold, Mike, and Gruninger, Michael. (1996) "Ontologies: principles, methods, and applications." *Knowledge Engineering Review* **11**: 93–155.

[3] Gomez Perez, Asunción. (2003) "Deliverable 1.5: A survey of ontology learning methods and techniques". *IST Project IST-2000-29243 OntoWeb*.

[4] Wong, Wilson, Wei, Liu, and Bennamoun, Mohammed. (2012) "Ontology learning from text: A look back and into the future." *ACM Comput. Surv.* **44 (4)**: 1-36.

[5] Zhou, Lina. (2007) "Ontology learning: state of the art and open issues." *Information Technology and Management* **8(3):** 241-252.

[6] Gruber, Thomas. (1993) "A translation approach to portable ontologies." *Knowledge Acquisition* **5:** 199–220.

[7] Hazman, Maryam, El-Beltagy, Samhaa R., Rafea, Ahmed. (2011) "A Survey of Ontology Learning Approaches." *International Journal of Computer Applications* **22 (8)**: 36–43.

[8] Jankowski, Jarosław, Hamari Juho, and Wątróbski, Jarosław. (2019) "A gradual approach for maximising user conversion without compromising experience with high visual intensity website elements." *Internet Research* **29 (1)**: 194-217.

[9] Shamsfard, Mehrnoush, Andbarforoush, Ahmad. (2003). "The state of the art in ontology learning: A framework for comparison." *Knowl. Eng. Rev.* **18 (4)**: 293–316.

[10] Konys, Agnieszka, Wątróbski, Jarosław, and Różewski Przemysław. (2013). „Approach to Practical Ontology Design for Supporting COTS Component Selection Processes", ACIIDS 2013 - A. Selamat et al. (Eds.): ACIIDS 2013, Part II, *LNAI 7803*, *Springer, Heidelberg* 245-255.

[11] Sanchez, David, and Moreno, Antonio. (2008) "Learning non-taxonomic relationships from Web documents for domain ontology construction." *Data Knowl. Eng.* **64 (3)**: 600–623.

[12] Konys, Agnieszka. (2018) "An Ontology-Based Knowledge Modelling for a Sustainability Assessment Domain." *Sustainability* **10 (300).**

[13] Cimiano, Philipp, and Staab, Steffen. (2005) "Learning concept hierarchies from text with a guided agglomerative clustering algorithm." In *Proceedings of the Workshop on Learning and Extending Lexical Ontologies with Machine Learning Methods*.

[14] Ding, Ying, and Schubert, Foo. (2002) "Ontology research and development. Part 1-a review of ontology generation." *Journal of information science* **28 (2)**: 123-136.

[15] Wątróbski, Jarosław, Jankowski, Jarosław, Ziemba, Paweł, Karczmarczyk, Artur, and Zioło, Magdalena. (2019). Generalised framework for multi-criteria method selection. *Omega* **86:** 107-124.

[16] Konys Agnieszka. (2018) "Knowledge systematization for ontology learning methods." *Procedia Computer Science* **126:** 2194-2207

[17] Faure, David, and Poibeau, Thierry. (2000) "First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX." In: S. Staab, A. Maedche, C. Nedellec, P. Wiemer-Hastings (eds.), *Proceedings of the Workshop on Adapting lexical and corpus resources to sublanguages and applications*.

[18] Aussenac-Gilles, Nathalie, and Seguela, Patrick. (2000) "Les relations sémantiques: du linguistique au formel. Spécial sur la linguistique de corpus." *Cahiers de grammaire* **25**: 175-198.

[19] Engels, Robert. (2001) "CORPORUM-OntoExtract. Ontology Extraction Tool." *Deliverable 6 Ontoknowledge*.

[20] Bachimont, Bruno, (2000). "Engagement sémantique et engagement ontologique: conception et realisation d'ontologies en ingénierie des connaissances." *In Ingénierie des Conniassances : Evolutions récentes et nouveaux défis, Eyrolles* 305-323.

[21] Jones, Steve, and Gordon W., Paynter. (2002) "Automatic extraction of document keyphrases for use in digital libraries: evaluation and applications." Jo*urnal of the American Society for Information Science and Technology* **53(8):** 653-677.

[22] Karczmarczyk, Artur, Jankowski, Jarosław, and Wątróbski, Jarosław. (2018). „Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks." *PloS one* **13 (12):** e0209372.

[23] Mikheev, Andrei, and Steven, Finch. (1997) In Proceedings of the fifth conference on Applied natural language processing. Association for Computational Linguistics, 372-379.

[24] Gilles, Bisson, Nédellec, Claire, and Canamero, Dolores. (2000) "Designing Clustering Methods for Ontology Building-The Mo'K Workbench." *ECAI workshop on ontology learning* **31**.

[25] Missikoff, Michele, Navigli, Roberto, and Velardi, Paola (2002). "Integrated approach to web ontology learning and engineering." *Computer* **35 (11)**: 60-63.

[26] Morin, Emmanuel. (1999) "Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique." *TAL. Traitement automatique des langues* **40 (1):** 143-166.

[27] Wu, Shih-Hung, and Wen-Lian Hsu. (2002) "SOAT: a semi-automatic domain ontology acquisition tool from Chinese corpus." *COLING 2002: The 17th International Conference on Computational Linguistics: Project Notes.*

[28] Gupta, Kalyan Moy, et al. (2002). "An architecture for engineering sublanguage WordNets." *In Proceedings of the First International Conference On Global WordNet* 207-215.

[29] Chaelandar, Gaë, and Grau, Brigitte. (2000) "SVETLAN'- A System to Classify Words in Context". In S. Staab, A. Maedche, C. Nedellec, P. Wiemer-Hastings (eds.) Proceedings of the Workshop on Ontology Learning, 14th European Conference on Artificial Intelligence ECAI'00.

[30] Xu, Feiyu, et al. (2002). "A Domain Adaptive Approach to Automatic Acquisition of Domain Relevant Terms and their Relations with Bootstrapping." In Proceedings of LREC 2002, The third international conference on language resources and evaluation.

[31] Biébow, Brigitte, and Szulman, Sylvie. (1999) TERMINAE: a linguistic-based tool for the building of a domain ontology. In EKAW'99 Proceedings of the 11th European Workshop on Knowledge Acquisition, Modelling and management. *LCNS Springer-Verlag*, 49-66.

[32] Maedche, Alexander, and Volz, Raphael. (2001) "The Text-To-Onto Ontology Extraction and Maintenance Environment." *In ICDM-Workshop on Integrating Data Mining and Knowledge Management*, USA.

[33] Pereira, Francisco, Câmara, A. Oliveira, and Amílcar, Cardoso (2000). "Extracting Concept Maps with Clouds." *Argentine Symposium of Artificial Intelligence* (ASAI 2000).

[34] Alfonseca, Enrique, et al. (2002) "Automatically Generating Hypermedia Documents depending on User Goals." *Workshop on Document Compression and Synthesis in Adaptive Hypermedia Systems*, AH-2002.

[35] Zelle, John Marvin. (1995) "Using Inductive Logic Programming to automate the construction of natural language parsers." *PhD Dissertation*, University of Texas at Austin.

[36] Hahn, Udo, Romacker, Martin, and Schulz, Stefan. (2000), "MedSynDiKATe--design considerations for an ontology-based medical text understanding system", *In Proceedings of the AMIA Symposium*, American Medical Informatics Association.

[37] Jiang, Xing, and Ah-Hwee, Tan. (2009) "CRCTOL: A semantic-based domain ontology learning system." *Journal of the American Societ for Information Science and Technology*, **61 (1)**: 150-168.

[38] Drymonas, Euthymios, Kalliopi Zervanou, and Euripides G.M., Petrakis. (2010) "Unsupervised ontology acquisition from plain texts: The OntoGain system." *Natural Language Processing and Information Systems*, 277-287.

[39] Sclano, Francesco, and Paola, Velardi. (2007) "TermExtractor: a Web Application to Learn the Shared Terminology of Emergent Web Communities." *Enterprise Interoperability II*. Springer, London, 287-290.

[40] Wątróbski, Jarosław, Ziemba, Ewa, Karczmarczyk, Artur, & Jankowski, Jarosław. (2018). „An index to measure the sustainable information society: the Polish households case." *Sustainability*, **10(9)**: 3223.

[41] Konys, Agnieszka. (2018). "Towards Knowledge Handling in Ontology-Based Information Extraction Systems." *Procedia Computer Science,* **126:** 2208-2218.

[42] Sałabun, Wojciech. (2015). "The Characteristic Objects Method: A New Distance-based Approach to Multicriteria Decision-making Problems." *Journal of Multi-Criteria Decision Analysis*, **22(1-2)**: 37-50.

[43] Piwowarski, Mateusz, et al. (2018) „Application of the Vector Measure Construction Method and Technique for Order Preference by Similarity Ideal Solution for the Analysis of the Dynamics of Changes in the Poverty Levels in the European Union Countries." *Sustainability*, **10:** 2858.

[44] Konys, Agnieszka. (2017) "Ontology-Based Approaches to Big Data Analytics." In: Kobayashi S., Piegat A., Pejaś J., El Fray I., Kacprzyk J. (eds) Hard and Soft Computing for Artificial Intelligence, Multimedia and Security. *Advances in Intelligent Systems and Computing*, *Springer, Cham,* **534:** 355-365.

[45] Wątróbski, Jarosław. (2016). " Outline of multicriteria decision-making in green logistics." *Transportation Research Procedia*, **16**: 537-552.

[46] Piwowarski, Mateusz. et al. (2018) "TOPSIS and VIKOR methods in study of sustainable development in the EU countries, *Procedia Computer Science,* **126**: 1683-1692.

[47] Sałabun, Wojciech. (2014). "Reduction in the number of comparisons required to create matrix of expert judgment in the comet method." Management and Production Engineering Review, 5(3): 62-69.

[48] Sałabun, Wojciech, and Piegat, Andrzej. (2017). "Comparative analysis of MCDM methods for the assessment of mortality in patients with acute coronary syndrome." Artificial Intelligence Review, 48(4): 557-571.

[49] Wątróbski, Jarosław, Jankowski, Jarosław, and Piotrowski, Zbigniew. (2014) „The selection of multicriteria method based on unstructured decision problem description." In: Hwang D., Jung J.J., Nguyen NT. (eds) Computational Collective Intelligence. Technologies and Applications. *Lecture Notes in Computer Science*, *Springer, Cham*, **8733**: 454-465.

[50] Nermend, Kesra, and Piwowarski, Mateusz. (2018) "Cognitive Neuroscience Techniques in Supporting Decision Making and the Analysis of Social Campaign." In: Proceedings book International Conference on Accounting, Business, Economics and Politics (ICABEP-2018), Erbil, Iraq, 1-12.

[51] Jankowski, Jarosław, Kolomvatsos, Kostas, Kazienko, Przemysław, and Watróbski, Jarosław. (2016) „Fuzzy Modeling of User Behaviors and Virtual Goods Purchases in Social Networking Platforms." *J. UCS*, **22(3)**: 416-437.