AI technique to generate course specification's item

By

H.H.Sadoma

Researcher, Computer Teacher Preparation Dep.
Faculty of Specific Education.
Mansoura University. Egypt
dev.haytham.hassan@gmail.com

E.A.Elayat

Computer lecturer, Computer
Teacher Preparation Dep.
Faculty of Specific Education.
Mansoura University. Egypt
ealayat@yahoo.com

A. F. Elgamal

Prof of Computer Science, Computer
Teacher Preparation Dep.
Faculty of Specific Education.
Mansoura University. Egypt
amany_elgamal@hotmail.com

AI technique to generate course specification's item

H.H.Sadoma*   **E.A.Elayat**[**]   **A. F. Elgamal** [***]

Abstract:

This paper presents a proposed system for course specification based on the combination of text mining techniques and semantic web services.

The course specification depends on firstly, text mining techniques by its methods such as (tokenization and stemming). Secondly by using semantic web services such as ontology web language (OWL).

OWL is defined as language for defining ontologies on the web, it describes a domain in terms of classes, properties, individuals and description of the characteristics of those objects.

Although OWL is not a programming language, the research relied on some utilities such as Java, SPARQL, and JENA library.

Finally, the proposed system produced a various educational objectives (skilled, cognitive and emotional) for the specific lesson.

Keywords:

Text mining, Web Ontology Language, Resource Description Framework, Jena Ontology API and SPARQL query language.

1- Introduction:

Educational goals defined as the basic goals that we would like students to achieve when passing educational learning experiences in different curriculums. It is understood that educational goals are either public or private. In terms of general goals, educational goals are that a student can reach in a long period of time and cannot be achieved in one

[*] Researcher, Computer Teacher Preparation Dep. Faculty of Specific Education. Mansoura University. Egypt

[**] Computer lecturer, Computer Teacher Preparation Dep. Faculty of Specific Education. Mansoura University. Egypt

[***] Prof of Computer Science, Computer Teacher Preparation Dep. Faculty of Specific Education. Mansoura University. Egypt

class or a week of study, such as the general objectives of teaching science at the primary level. These objectives can be communicated to the student at the end of the curriculum or curriculums of science for the primary stage.

Private objectives are the objectives of a teaching unit or subject, and these types of goals are sometimes called behavioral goals as the behaviors the school aims to impart to its students, or performance goals as that part of the learner's behavior that can be observed and evaluated.

Setting the goal helps with clarity, any successful actions must be directed towards specific and acceptable goals. Otherwise the work becomes a kind of trial and error based on randomness and impromptu. This leads to a waste of effort, time and money, which our energy cannot accept or tolerate. This randomness is really what we avoid in our education.

Artificial intelligence (AI) is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. [9]

The impact of AI as a powerful technology can be witnessed in diverse industry verticals. The education industry across the globe is no exception to this. AI in education is being used by different schools in the country. The use of AI in education has given a completely new perspective of looking at education to teachers, students, parents, and of course the educational institutions as well.

Natural Language Processing (NLP) is an effective technology for a group of computational systems for representing and analyzing naturally occurring texts, at one or more levels of linguistic analysis for the aim of simulating human-like language processing for a set of applications.

NLP techniques are utilized as a part of applications that extract data from texts and store it to database, for example, generating of learning objectives, retrieving relevant documents from a collection or translation, by generating text responses, or identifying spoken words then turn them into text.

Generating learning objectives depends on understanding the lesson content to be studied in order to obtain measurable educational outcomes using text-mining technology as a branch of NLP.

Text mining technique is defining as: the process of exploring and analyzing large amounts of unstructured text data aided by software that can identify concepts, patterns, topics, keywords and other attributes in the data. It's also known as text analytics. [10]

Text mining techniques consist of three stages;

1) Text preprocessing.

2) Text Transformation

3) Text Mining Methods

(OWL) is a semantic web language designed to represent rich and complex knowledge about things, groups of things, and relations between things, OWL documents, known as ontologies, can be published in the World Wide Web and may refer to or be referred from other OWL ontologies. OWL is part of the W3C's Semantic Web technology stack, which includes RDF, RDFS, SPARQL, etc. [11]

Resource Description Framework (RDF) is a standard model for data interchange on the web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.

RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications. [12].

Protégé is an open-source tool for editing and managing terminologies and ontologies. It is the most widely used domain-independent, freely available, platform-independent technology for developing and managing terminologies, ontologies, and knowledge bases in a broad range of application domains, there are several plug-ins for

Protégé, enabling it to import and export ontologies in many different formats, including XML, RDF, OWL, and the native Protégé format. [13]

Jena: is a Java framework for building semantic web applications. It provides extensive java libraries for helping developers develop code that handles RDF, RDFS, RDFa, OWL and SPARQL in line with published W3C recommendations. Jena includes a rule-based inference engine to perform reasoning based on OWL and RDFS ontologies, and a variety of storage strategies to store RDF triples in memory or on disk. [14]

SPARQL: can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports aggregation, sub queries, negation, and creating values by expressions, extensible value testing, and constraining queries by source RDF graph. The results of SPARQL queries can be result sets or RDF graphs. [15]

In this paper, an intelligent system for generating educational objectives is proposed and implemented on educational content for one of the curricula for basic education students in the Egyptian Ministry of Education. The proposed system can generate random of educational objectives (skilled, cognitive and emotional) for the lesson entered by the user of the system.

The rest of this paper is organized as follows: Section 2 presents the components of the proposed system; section 3 describes the future work, and finally section 4 presents the conclusion.

2- The proposed system:

The proposed intelligent educational objectives generation system contains three phases, namely Text Mining Phase, Connecting to OWL phase and Text Matching Phase, each of them consists of many sub-components, as shown in figure (1).
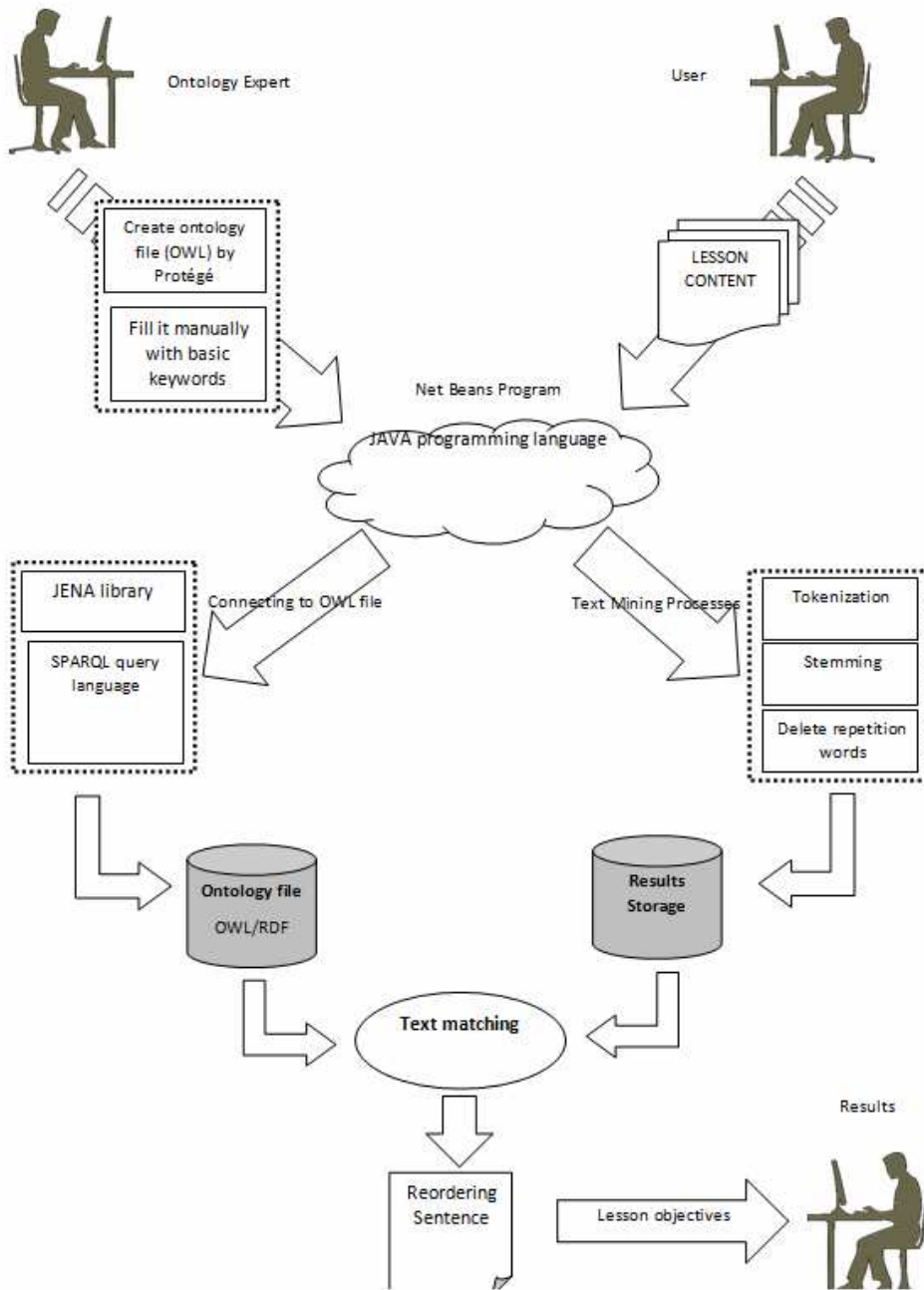
2.1: Text Mining Phase

Text mining techniques consist of three stages:

2.1.1: Text preprocessing:

It is the process of cleaning and interpreting data into its implementing format. Being a core aspect of NLP, text preprocessing comprises the use of many techniques such as language identification, tokenization, part-of-speech tagging, Stemming and many more.

2.1.1.1: Tokenization

The tokenization process is implemented into each sentence of the entered lesson. For example " Scanner: An input device can convert contents of paper document into digital image " is converted to tokens as follows: [Scanner:] [An] [input] [device] [can] [convert] [contents] [of] [paper] [document] [into] [digital] [image].

2.1.1.2: Stemming

Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma. Stemming is important in natural language understanding (NLU) and natural language processing (NLP). [16]

Stems are created by removing the suffixes or prefixes used with a word, for example, deleting parentheses, breakpoints, quotes, etc…

2.1.2: Text transformation

Text transformation: conducts features generation. Feature generation reflects documents by words they contain and words occurrences where the order of word is not significant. It deploys bag-of-words or vector space models. [17]

Bag-of-words is one of the most fundamental algorithms in NLP. Its main work is information retrieval from data. It depends on cleaning of text by removing all the words that are less significance to our algorithm, such as duplicated words, determiner (a, an, the), and pronouns, Etc.

2.1.3: Text Mining Methods

Various text-mining techniques and methods such as classification, clustering, summarization, Term Based Method and more.

2.1.3.1: Term Based Method

Term is defined as the word that has a well-explained meaning in a document. Under term based method, the document is inspected on the basis of terms and takes the benefit of productive computational performance while capturing the theories for term weighting, Over the time, after the association of information retrieval and machine learning, the term based techniques have developed. It has some disadvantages as well, such as:-

Polysemy: means a word having multiple meanings, and Synonymy: means multiple words have the same meaning such as central processing unit = CPU, Random Access memory = RAM and Read Only memory = ROM

The results of this stage will be sets of processing tokens.

2.2: Ontology Phase

In this point, the proposed system using an expert to create the ontology file, which will be used as a words store containing all the keywords for the lesson, Protégé is the most suitable program.

2.2.1: Creating OWL file

The proposed system supports the ontology file with some keywords, which will be entered manually by the expert, first creating the main class then creating the subclasses.

2.2.2: Filling owl file with basic keywords

In this stage, keywords has been added to the lesson by creating individuals for each subclass, then a data property is created for each individual which is the property that will be called in the code later, the data type will be string, the file is in RDF/XML format.

2.2.3: Connecting to owl file

The proposed system supports the Java programming language in addition to some tools such as the jena library and the query language Sparkle to connect to the ontology file.

2.2.3.1: Apache Jena Library

To connect to the owl file a Java Class is created, then adding some codes at the beginning of the class to call the Jena file packages.

2.2.3.2: SPARQL query language

The query statement consists of two parts: the SELECT clause identifies the variables to appear in the query results, and the WHERE clause provides the basic graph pattern to match against the data graph, The results of SPARQL queries will be result sets.

2.3: Text Matching Phase

In this stage, the text is matched between the results of the first stage (text mining stage) and the second stage (connecting to the OWL file), then the sentence is rearranged and finally the objectives of the input lesson are extracted.

3: Future work

The proposed system will be applied to a sample of students on the third year in the preparatory stage for computer subject at the Egyptian Ministry of Education, and the results will be studied under comparing them with the human factor to obtain the effectiveness of the proposed system.

4: Conclusion

In this paper, an intelligent system for educational objectives generation was produced, which contained three main stages, "Text Matching Phase" using NLP tools for extracting sets of processing tokens, the second stage "Connecting to OWL " to create the ontology file and fill it with the important keywords then bind it with the graphical user interface and the final stage " Text Matching Phase " to make text matching between the previous two stages and rearrange the statement to generate the educational objectives for the entered lesson.

The proposed system is not applied to computer lessons only, but can be applied to any course, if an ontology expert is present.

6: REFERENCES

1. Samaa Elnagar, Victoria Yoon, Manoj A.Thomas, "An Automatic Ontology Generation Framework with An Organizational Perspective ", Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020. available at https://doi.org/10.48550/arXiv.2201.05910, viewed on 08-02-2021.

2. Rafael Ferreira-Mello, Máverick André, Anderson Pinheiro, Evandro Costa and Cristobal Romero, "Text mining in education ",2019.

3. Matthew Horridge & Mark Musen, " Snap-SPARQL: A Java Framework for Working with SPARQL and OWL ", OWLED (2015): Ontology Engineering pp 154–165

4. Bob DuCharme, "Learning SPARQL, 2nd Edition ", O'Reilly Media, Inc., ISBN: 9781449371432, (2013),

5. AGOSTINO POGGI, "A Java and OWL based Approach for System Interoperability", (2011)

6. Zainab Salih Ageed, Rowaida Khalil Ibrahim and Mohammed A. M. Sadeeq, "Unified Ontology Implementation of Cloud Computing for Distributed Systems", Current Journal of Applied Science and Technology; Article no.CJAST.62378, ISSN: 2457-1024, 2020

7. Naresh Kumar, Minakshi Kumar and Manjeet Singh, "Automated ontology generation from a plain text using statistical and NLP techniques " International Journal of System Assurance Engineering and Management volume 7, pages 282–293 (2016).

8. PankajDeep Kaur, Pallavi Sharma and Nikhil Vohra "An_Ontology_Based_Text_Analytics_on_Social_Media" International Journal of Database Theory and Application, Vol.8, No.5 (2015), pp.233-240.

9. Available at: https://builtin.com/artificial-intelligence, viewed on 08-05-2020.

10. Available at: https://www.techtarget.com/searchbusinessanalytics/definition/text-mining , viewed on 17-12-2019.

11. Available at: https://www.w3.org/OWL/, viewed on 14-02-2021.

12. Available at: https://www.w3.org/RDF/, viewed on 15-05-2021.

13. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2039856/
, viewed on 24-11-2020

14. Available at: https://jena.apache.org/about_jena/about.html
, viewed on 03-03-2019

15. Available at: https://www.w3.org/TR/sparql11-query/
, viewed on 25-11-2021.

16. Available at:
https://www.techtarget.com/searchenterpriseai/definition/stemming, viewed on
14-10-2020

17. Available at: https://www.analyticssteps.com/blogs/what-text-mining-process-
methods-and-applications, viewed on 29-11-2021

18. Available at:
https://www.youtube.com/watch?v=AnKQJxp2ayI&ab_channel=AliSaidi,
viewed on 05-05-2022