Automatic Generation of Learning Path

Claudia Perez-Martinez[1], Gabriel Lopez Morteo[1], Magally Martinez Reyes[2] and Alexander Gelbukh[3]
[1]*Instituto de Ingeniería, Universidad Autónoma de Baja California, México*
[2]*Universidad Autónoma del Estado de México, México*
[3]*Centro de Investigación en Computación, Instituto Politécnico Nacional, México*

1 RESEARCH PROBLEM

The diversity of forms to access to knowledge is one of the most important features of the current learning society (UNESCO, 2005). Consequently, the transmission of knowledge process turns into a relevant task. Instructional Design (ID) plays an important role by establishing methods for creating learning experiences which helps to develop and enhance student skills and student knowledge. One of the phases in ID is curriculum sequencing; its main objective is to select the most suitable individually planned sequence of knowledge and tasks. The sequence of knowledge units is named High Level Active Learning Path, or simply Learning Path (Brusilovsky, 1999).

A learning path is designed for one new unit knowledge to be learned. Generally, the knowledge units of a learning path are prior knowledge, which is necessary to understand the new knowledge. The learning path design turns more challenging in web-based adaptive educational systems because the student profile in web environments can be more diverse than the profile student in a classroom (Brusilovsky and Peylo, 2003).

Different learning path generation approaches has been developed, many of them are based on specific characteristics of each particular student, for example the results of a pre-test, the current emotional state of the student or the previous statistical count of use of educational resources.

However, to determine the learning path, is necessary to know what the ideal state of knowledge is. Before to recognise the current knowledge of one particular student, it is necessary to establish what a generic learning path is, independently of the particular student profile.

This means that for each new knowledge unit, a new generic learning path need to be built. After, this generic learning path could be personalized by applying some learning strategy. The problem is: given a particular knowledge unit to learn, how to automatically establish a generic learning path?

2 OUTLINE OF OBJECTIVE

In the area of instructional design is necessary to establish the knowledge units that will present in an instructional session, the instructional session helps to student to learn one particular new knowledge unit.

Usually the set of knowledge units are selected by the professor based on the student profile. The professor –or the knowledge expert in instructional design- knows which knowledge are necessary to learn a new concept, and he selects some of them to remember at student in an instructional session.

The objective of this research is to find a mechanism for automatically to establish a generic learning path for any particular knowledge unit. To get the objective is necessary to know how this problem has been resolved, which strategies has been implemented. Besides it is necessary to propose the methodology to get the objective and to prove the obtained results.

2.1 Prior Proposal

Based on previous documental revisions, in this research has proposed the use of Natural Language Processing (NLP) techniques to resolve the problem. Particularly the propose is to use those based on external knowledge sources techniques.

So, to generalize the generic learning path building process, we should have a very complete knowledge base to extract the necessary information for each particular request in all time.

A useful and well-known structure for knowledge representing is the ontology, it names and defines the types, properties, and interrelationships of the concepts in a domain of knowledge, such characteristics made it convenient to find the learning path.

Nevertheless, to build an ontology results in high cost; besides always it is limited to a domain knowledge. This problem has been confronted from the Natural Language Processing area, but they have

Perez-Martinez C., Lopez Morteo G., Martinez Reyes M. and Gelbukh A..
Automatic Generation of Learning Path.

founded one alternative using one great resource as ontology, Wikipedia. This research uses this source to design a method to automatically generate a learning path to one particular knowledge unit.

3 STATE OF THE ART

The adaptive multimedia instruction authoring producing suitable learning content that matches student learning styles. This is one of the challenging tasks in the emerging multimedia technologies for e-learning (Lau et al., 2014).

3.1 Learning Path Generation Process

The learning path generation process has been studied from diverse perspectives as follows. Based on the flow theory, one learning path is selected taking care of the state of mind of the student (Katuk and Hokyoung Ryu, 2010). In (Chih-Ming Chen, 2008) the authors constructed a personalized learning path based on simultaneously considering courseware difficulty level and learning concept continuity during learning processes, a genetic-based curriculum sequence scheme was developed. The algorithm constructs a learning path according to the incorrect response patterns of a pre-test. Other approach takes into account eventual competency dependencies among learning objects. The authors propose a learning design recommendation system based on graph theory, they using the concept of cliques, a loop generating sub graphs, until one such clique is generated whose prerequisites are a subset of the learner's competencies (G. Durand et al., 2013). One proposed methodology is inspired to the Knowledge Space Theory, and it proposes some heuristics to transform one original ontology in a weighted graph where the A* algorithm is used to find the path. The ontology is the result of the semantics of the relations among concepts (Pirrone et al., 2005).

A proposal for a personalized e-learning system is based on Item Response Theory -which considers both course material difficulty and learner ability to provide individual learning paths for learners-. In the proposal a single difficulty parameter is used to model the course materials, and the maximum likelihood estimation is applied to estimate learner ability based on explicit learner feedback. Besides, a collaborative voting approach is used for adjusting course material difficulty (Chen et al., 2005).

Other proposed approach develop a genetic algorithm and case-based reasoning to construct an optimal learning path for each learner. (Huang et al., 2007).

All this approach needs one source of knowledge where to obtain the information to apply a learning strategy. So, they are limited by the domain of their sources of knowledge.

3.2 Assumptions

As result of a documental research, some assumptions have been useful to this work. To begin to describe the learning path building we have stated some assumptions as follows.

(1) The curriculum sequencing can be resumed as the knowledge unit selection to build the learning path from a complete universe of possibilities

(2) A learning path, for a specific objective knowledge (new knowledge unit), can be seen as an organized set of knowledge units, they correspond to prior knowledge for one new knowledge unit, named objective knowledge (Fig. 2.2). The last element in the learning path will be precisely the new knowledge unit. After, each knowledge unit is associated to one specific activity.

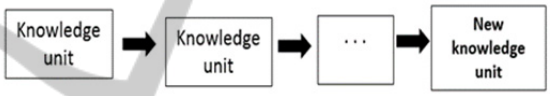

Figure 3.1: Learning path

(4) The learning path generation process has been explored under the NLP approach, particularly by statistical methods.

(5) It is known that, in the NLP area, the based on additional knowledge sources methods provides better results than the based on statistical approaches. Nevertheless, the size and domain of the additional knowledge resources is usually limited, because the construction of this kind of resources is costly.

(6) Wikipedia is now treated as a linguistic resource, it is used in PLN tasks, the performance of some of them results even better than those using other resources as Wordnet (Medelyan et al., 2009).

(7) In Wikipedia content, unlike the categories structure shapes one hierarchical structure, the articles structure shapes one cyclic graph, this can be seen resembling the human brain. We associate one event or object to some ideas or concepts. Depending the situation (context), but this same ideas can be evoked from another context. The figure 3.1 shows a snapshot at Wikipedia article "*Derivative*" and its anchors. "*Derivative*" has nodes which point to different articles and at the same

time, this articles point to other o the same articles. Derivative article points to "*Function*" article and "*Function of a real variable*" article too points to "*Function*".

Then, is possible that one teacher in his classroom, to teach "*Derivative*" concept, first address the "*Function*" concept, and after address the "*Function of a real variable"* concept and finally Derivative. Perhaps only selects *"Function of a real variable"* before "*Derivative*". Which will be the correct selection? Which others concepts must be select to build the learning path for "*Derivative*"? The selection depending on the learning strategy only? Perhaps before can answer this questions is necessary to know the structure of concepts.

The teacher undoubtedly knows this structure, but in an automatic system is necessary to provide this information. Once the system has the information, how the system select the appropriate concepts to build the learning path?

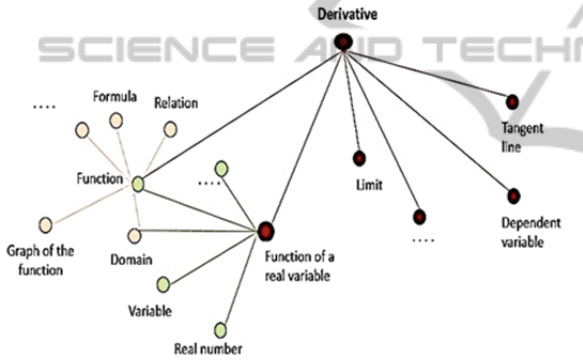

Figure 3.2: Knowledge units semantically related

In front of the special interest in provide the adequate learning path to each different student, one learning path construction method, first would know which are every the necessary knowledge units to understand a new concept, or a new knowledge units. After, some of the different strategies will decide which knowledge units to select, dealing with the student profile.

A learning path can visualize as one acyclic directed graph with a topological sort, where each node represent a concept (knowledge unit) which would be learned by the student before to try to learn the new concept. The next section describes how to carry out the learning path generation.

The process to discriminate the unnecessary knowledge units to build the learning path can resume it as select one subset of knowledge units from a complete universe of possibilities. We propose to build one complete set of knowledge units surrounding the objective knowledge unit.

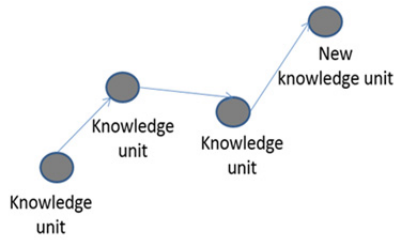

Figure 3.3: Graphic representation of learning path.

4 METHODOLOGY

The idea to build a learning path in automatic form is possible to build a learning path by use of NLP techniques and using as additional knowledge resource to Wikipedia, the steps are described as follows:

1. Take the textual content of one objective knowledge, and enrich it with explanatory links toward Wikipedia. Each link then will be a knowledge unit.
2. Calculate the semantic relatedness between the objective knowledge and each knowledge unit.
3. Take the knowledge units more closely related with the objective knowledge. The created list will be the learning path.

The use of Wikipedia as the knowledge source permit to have a broad space of concepts, whose semantic relatedness can be numerically measured. So, is possible to get a learning path for any concept which is stored as an article in the database of Wikipedia.

In case of the source of knowledge source is not Wikipedia it is possible to convert a document, for example a learning object, in a linked document like a Wikipedia article, as is shown in the Appendix A

5 EXPECTED OUTCOME

The expected outcome is, based on a research and software development process, to obtain a useful tool to get a learning path for a specific knowledge unit. This learning path will be useful to automatic instructional design purposes. This tool must be useful in educational virtual environments to carry out different learning strategies.

The tool consists of an algorithm whose input is only a text with the definition or description of one knowledge unit (objective knowledge). The output is a learning path, says, a group of knowledge units, which are closely related to the objective

knowledge.

One of the main challenge is to get the necessary knowledge resource for the algorithm.

5.1 Validation Process

As it has been described, the proposal method in this paper generates a learning path based on use Wikipedia as linguistic resource.

To test the results one survey has been developed. The survey was based on the result of a prior questionnaire applied to a group of professional in engineering. The evaluated group selected a learning path to the "Derivative" concept, the opinion was seemed but not identical. How to measure the closeness among the results and the automatically generated learning path and the interviewee people?

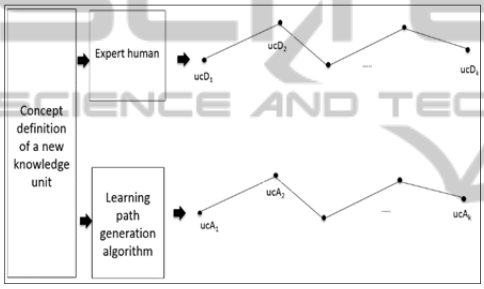

Figure 5.1: Graphic representation of learning path.

In each case the resulting product is an acyclic graph, whose nodes are the concepts or knowledge units which can be measured by some numeric values.

The validation method selected is clustering. When a cluster rather than a classifier is learned, the output takes the form of a diagram that shows how the instances fall into clusters. Clustering techniques apply when there is no class to be predicted but the instances are to be divided into natural groups (Witten et al., 2011).

We will use an algorithm that works in numeric domains, using the nearest neighbor method of instance-based learning. The method will be used to measure the closeness among the learning path automatically generated and the learning path established by one group of expert humans.

6 STAGE OF THE RESEARCH

The algorithm, to build a learning path based on use of Wikipedia as external knowledge resource, has been developed. Some of the main contributions are the follows:

a) One method to enrichment learning objects has proposed (see APENDIX).
b) One method to generate learning path has been developed. The method is based on NLP, and it use as knowledge source to Wikipedia. The method visualize Wikipedia as an Ontology

The main contributions in this approach are two: one proposal to carry out WSD based on the use of metadata as either an additional or alternative context, and the method to discriminate the relevant phrases based on the degree of semantic relatedness with the LO main subject.

The validation was developed for the "Derivative" knowledge unit, a survey was applied to a group of mathematics teachers.

REFERENCES

The content of this document is part of submitted to evaluation articles.

The content of this document is part of submitted to evaluation articles.

UNESCO. 2005. Hacia las sociedades del conocimiento. Informe Mundial de la UNESCO. Ediciones UNESCO.

Brusilovsky, P. (1999). Adaptive and Intelligent Technologies for Web-based Education. Special Issue on Intelligent Systems and Teleteaching, Künstliche.

Brusilovsky, P. & Peylo, C. 2003. Adaptive and Intelligent Web-based Educational Systems. *Int. J. Artif. Intell. Ed.* 13, 2-4. 159-172.

Christopher D., Manning and Hinrich Schütze . (1999). Foundations of Statistical Natural Language processing MIT Press, Cambridge, MA, USA.

Indurkhya, N. & Damerau, F. (2010). *Handbook of Natural Language Processing* (2nd ed.). Chapman & Hall/CRC.

Medelyan, O., Milne, D., Legg, C., and Witten, I. H. (2009). Mining Meaning from Wikipedia. Extraído el 19 de enero de http://arxiv.org/abs/0809.4530.

Katuk, N.; Hokyoung Ryu. (2010). Finding an optimal learning path in dynamic curriculum sequencing with flow experience. *Computer Applications and Industrial Electronics (ICCAIE), 2010 International Conference on*, vol., no., pp.227,232, 5-8.

Witten, I., Frank, E., Hall, M. (2011). Data Mining: Practical Machine Learning Tools and Techniques. 3nd Edition, Morgan Kaufmann, San Francisco.

Theodoridis, S., Koutroumbas, K. (2006). *Pattern Recognition, Third Edition*. Academic Press, Inc., Orlando, FL, USA.

Lau, R., Yen, N., Li, F., and Wah, B. (2014). Recent development in multimedia e-learning technologies. World Wide Web 17, 2. 189-198.

Chih-Ming Chen. (2008). Intelligent web-based learning system with personalized learning path guidance, Computers & Education, Volume 51, Issue 2, September 2008, Pages 787-814, ISSN 0360-1315.

G. Durand et. al. (2013). Graph theory based model for learning path recommendation, Inform. Sci. (2013).

Pirrone, R. & Pilato, G. & Rizzo, R. & Russo, G. (2005). Learning Path Generation by Domain Ontology Transformation. Advances in Artificial Intelligence. Lecture Notes in Computer Science. Springer Berlin Heidelberg. V 3673.

Chen, C., Lee, H, and Chen,Y. (2005). Personalized e-learning system using Item Response Theory. Comput. Educ. 44, 3 (April 2005), 237-255.

Huang, M., Huang, H, and Chen, M. (2007). Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach. *Expert Syst. Appl.* 33. 3. 551-564.

APPENDIX

Wikification process*

The wikification process was inspired by the *wikipedians,* the people who edited the Wikipedia articles. They select the relevant words or phrases in an article and link them towards other Wikipedia articles which titles correspond to the phrase.

It is possible that there would be more than one article that matches, and the appropriate article needs to be selected according to the context. In this case, there is a disambiguation page with a list of possibilities. As it is shown in Fig. a.1, the phrase "jaguar" corresponds to more than one sense.

Figure a.1: Human WSD in wikification process.

Since, there is one disambiguation page, "Jaguar (disambiguation)", which contains several senses to the word "jaguar". The *wikipedians* easily select the correct sense.

This easy human process turns out to be very difficult to be done automatically. The text wikification "*task of automatically extracting the most important* words *and phrases in the document, and identifying for each such keyword the appropriate link to a Wikipedia article*". The process involves two apparently easy tasks: The selection of the relevant phrases and the WSD. The wikification process proposed in this paper follows one sequence of tasks, which begin from the extraction of useful information from LO (metadata and textual content), until the LO delivering with explanatory links towards Wikipedia articles (see Fig a.1).

The current wikification learning object methodology proposes the use of the metadata as either an additional or alternative context. The machine learning approach was proved with different classifier algorithms, but the best results were obtained with c4.5 algorithm, evaluated by cross validation method.

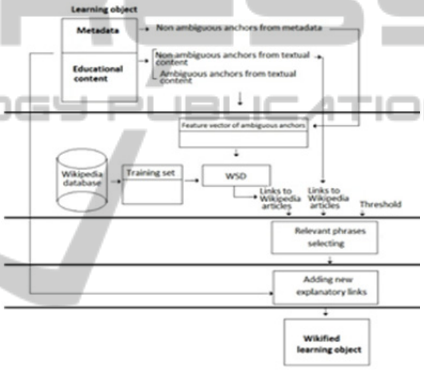

Figure a.2: Wikification process.