Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A NEW APPROACH FOR GENERATING EXERCISES BASED ONTOLOGIES: A CASE STUDY OF COURSES IN MATHEMATICAL ANALYSIS

[1]IMANE LMATI, [2]HABIB BENLAHMAR, [3]NACEUR ACHTAICH
[1]Analysis, Modeling and Simulation Laboratory, University Hassan II, Morocco.
[2] Laboratory of Information Technology and Modeling, University Hassan II, Morocco.
[3]Analysis, Modeling and Simulation Laboratory, University Hassan II, Morocco.

ABSTRACT

We present in this paper a methodology to generate an exercise model from the mathematical content of analysis taught in high school and university. The approach allows to a non-computer designer (teacher, educationalist ...) to set the parameters of a pedagogical object PO (Theorem, Definition ...) and to dynamically generate instances of the model by exploiting the semantic relations of PO with other pedagogical objects of mathematical corpus.

Keywords: *Automatic Generation Of Exercise, Evaluation, Ontology.*

1. PROBELEMATIC

The process of exercises generating in higher education is completely manual, which implies a high cost in terms of time writing statements and also a variable quality of exercises especially in the mathematical sciences because it is difficult to treat both the natural language and logic expressions [1] , hence the interest to exploit ontologies.

The use of ontologies in educational settings is increasing, they are a suitable formalism for the automatic evaluation [2].

Ontologies provide domain's knowledge in the form of: Concepts, individuals, Relationships..., they also incorporate a reasoning mechanism to derive the facts of knowledge explicitly defined [3].

This work allows to present the approach to generate a model of exercise based on a pedagogical ontology to conceive from a corpus analysis courses and a ontology Math-Bridge project [4].

In the next section, we will introduce the approach for generating an exercise model, especially the conception and the settlement of the educational ontology and the presentation of the generation algorithm, then we present a case study applied to the theme "Sequences" followed by an assessment of the approach and we conclude with some open perspectives for this work.

2. APPROACH

The author approach allows to generate from a pedagogical object (Theorem, Definition ...) a model of mathematical exercise by exploiting the PO semantic relations with other pedagogical objects corpus of mathematics. As result, an extensive exercise which operates many knowledge instead of only one knowledge.

We can summarize the generation process in two steps:

- Automatic conception of domain ontology from mathematics teaching content.
- Automatic generation of an exercise model based on domain ontology.

2.1 Conception and populating the ontology

2.1.1 Educational content

According to [5], an organization of pedagogic objects types (Figure 1) can be designed with a pedagogical ontology regardless of the different fields of application. Indeed, the terms of the ontology for an object type are the same in different fields such as computer science, mathematics, management ...because these terms are used by the author to express an educational information.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

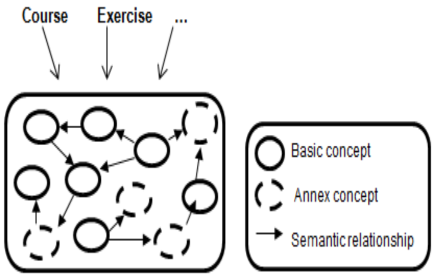


Figure. 1: Educational ontology

Depending on [6], the ontology may be described by:
- Simple nodes: correspond to the basic concepts and annex.
- Complex nodes: inputs in a sub-graph for (concept of complex base).
- Arcs: represent the semantic relations that bind the various nodes.

2.1.2 Ontology's core
To conceive the ontology's core we based on the project Math-Bridge [4].

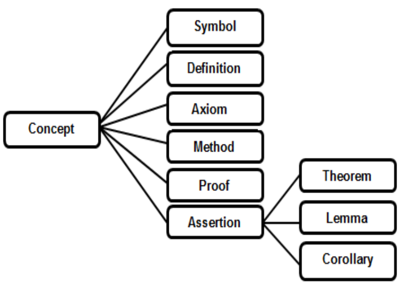


Figure 2: Ontology's core of Math-Bridge project

All pedagogical objects are linked together (by different types of possible relations) and with the concepts of the ontology. For example, the pedagogical object "Proof of intermediate value theorem" is connected to the object "Theorem intermediate values " and this object is connected to the concept of "continuous functions" in the ontology.

2.1.3 Instance of concept
To populate the ontology with instances of concept, the project [5], directed by SMINE
allows to extract textual segments (objects) reflecting an educational content (definition, example, exercise ...) as "is defined as, is defined by ..." for the definition or "Exercise, Tutorial..." for the exercise. They are explicitly indicated by linguistic indicators identified in the texts (verbs, nouns, adjectives). The connexions between indicators and indices are defined by rules. A rule (IDR) is triggered at the time of identification of one of its indicators, then it tries to locate linguistic indexes in the left context (CL1, CL2) and / or right (CR1, CR2) of the indicator confirming or not the semantic value expressed by the indicator. Each type of pedagogical object is a set of rules. Examples of rules are shown in the following table.

Table 1: Examples of rules

Rule	CL1	CL2	Indicator	CR1	CR2	Type /Subtype PO
R1		Is / are	Defined	By...		Definition
R2			Is /are	The/a...		Definition
R3		The/a...	Characteristic(s)	Of...	Is/are	Characteristic
R4		for	Example(s)	It/the/a...		Example

2.1.4 Attribute of concept
In order to be recovered and used, an educational resource has to be described by a set of metadata according to [7]. This metadata can be classified into two broad categories. The first describes the educational characteristics of the resource (authors, title, language, media, time ...) that is modeled by a set of pairs "attribute - value". This part is similar to the metadata described in existing standards such as LOM [8]. The second describes the semantics of the resource relative to the domain model. This semantics is itself divided into three parts: the prerequisites (the entrance of the resource), the content and (its outputs) acquisition function.
In our study we are interested in the second category which deals the semantics of pedagogical objects structured in three parts: Variable (parameter used in the PO), Prerequisites (constraint) and output (result of PO).

2.1.4.1 Variable
The variables used in OP can be extracted using the ontology algebra (Figure 3) of Math-Bridge [4].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

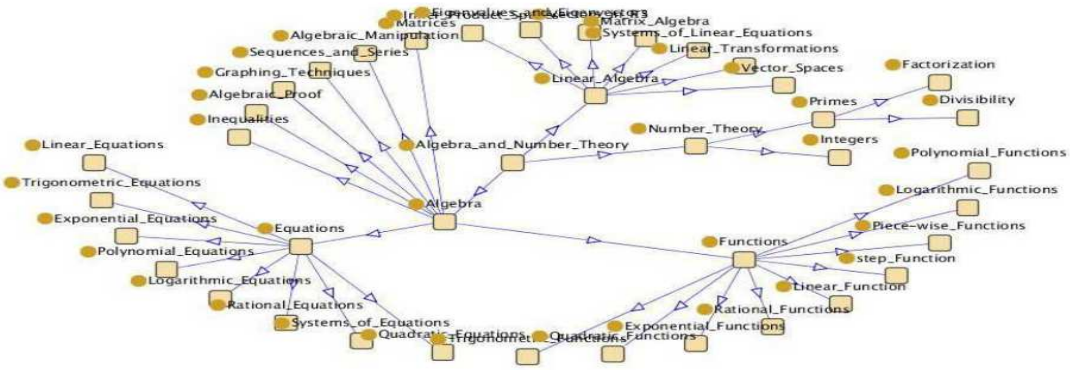


Figure. 3: Part of the ontology around algebra

In general, variables are preceded or followed by a concept or sub-concept of ontology algebra (Figure 4).

2.1.4.2 Prerequisites and output

Knowledge extraction (Prerequisite, output), related to variables entered in the form, is a combination of textual expression and mathematical functions [1].

Definition 8.2. *By f_n converges uniformly to f on E we mean that $\forall\varepsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall n > N$ and $\forall x \in E$*

$$|f_n(x) - f(x)| < \varepsilon.$$

Figure 4: Pedagogical object « uniform continuity»

In Figure (4), « *fn converges uniformly to f* » is a prerequisite for PO and « $|fn(x) - f(x)| < \varepsilon$ » its output. Some indicators and linguistic rules presented by [5] can help us to do the extraction of prerequisites and content of the PO:

Table 2: Some indicators and linguistic rules

Rule	CL1	CL2	Indicator	CR1	CR2	Attribute Type
R1			Suppose that	variable	Then	Prerequisites
R2		Variable	Is/are		If	Ouput
R3		Variable	Defined	by		Prerequisites
R4		Variable	Verify		Then	Prerequisites

According to Boukacem [9], Prerequisites may be listed by: Punctuation marks (",", "." ":" ",", "white", "paragraph" "!" "?"), Conjunctions ("or "," and "...)...

2.1.5 Relationship

The semantic description of resources helps to define the interesting relationships to navigate or search for resources according to [7]:

- High substitution: a resource R1 substitutes highly a resource R2 when the prerequisites of R1 are equal to prerequisites of R2.
- Low substitution: a resource R1 substitutes low a resource R2 when the prerequisites of R1 are included in prerequisites of R2.
- Equivalent resources: a resource R1 is equivalent to a resource R2 when R1 can be substitute R2 and the content of R1 is equal to the content of R2.
- Low precedence: a resource R1 precedes low a resource R2, if the content of R1 is included in the prerequisites of R2.
- High precedence: a resource R1 precedes highly a resourceR2, if the contents of R1 equal the prerequisite R2.

2.2 Exercise model

Our approach allows to build an exercise in reverse: For example, to demonstrate the result of a mathematical exercise, it is deducted from the result of a concept C(i), or we use its predecessor C(i-1) (C(i) and C(i-1) related by a high precedence relationship) which gives as a result equal to the prerequisite C(i) (Figure 5). In this case we ask for the prerequisites C(i-1) to achieve the result of C(i). At every iteration we can

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

integrate concepts that substitute highly or lowly concept being processed. The process may thus repeat for C(i-1) and so on...In the methodology, we did not use the equivalence relation since it allows to extract just the same concepts. We can generalize the process as follows:

Let C(1), C(2) ... instances of concepts, C(i-1) is the predecessor of C(i), C(n) the starting concept, C(n)(1), C(n)(2)… all adjacent concepts of C(n), Rpft(C(i), C(i-1)) the high precedence relationship between C(i) and C(i-1), Rpfb(C(i), C(i-1)) the low precedence relationship, Rsft(C(i), C(i-1)) the high substitution relationship, Rsfb(C(i), C(i-1)) the low substitution relationship and Pr1_C(n), Pr2_C(n)... the prerequisites of C(n).

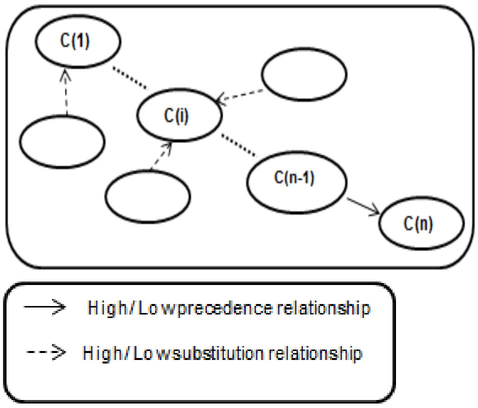


Figure 5 : Diagram of educational ontology

We start from the last exercise question based on the starting concept C(n) (Figure 5), and repeat the process until you get to the first question (depends on the number of concept used in the exercise).

```
Input (s)   : Choice of starting concept C(n) and these parameters
Output (s)  : Deduct the result of C(n)
Character   i =1
Start
 Request the output C(n)
  While there is an adjacent of C(n)
    If Rpft( C(n), C(n)(i) )
          If there is a high or low substitution relationship for C(n)(i)
            Ask the output of the first adjacent found
          End If
          Request the prerequisites of C(n)(i)
                  Pr1_ C(n)(i)
                  Pr2_ C(n)(i)
                  ...
                  Pr i_ C(n)(i)
                  ...
          C(n) ← C(n)(i) : replace C(n) by C(n)(i)
   If Rpfb( C(n), C(n)(i) )
          If C(n) is the starting concept
            Request the prerequisites C(n)
          End If
          If there is a High or low substitution relationship for C(n)(i)
            Ask the output of the first adjacent found
          End If
        Request the prerequisites C(n) using the prerequisites C(n)(i)
                  Pr1_C(n)
                  Pr2_C(n)
                  ...                  Pr1_C(n)(i)
                  Pr i_C(n)  ⟵        Pr2_C(n)(i)
                  ...                  ...
          C(n) ← C(n)(i) : replace C(n) by C(n)(i)
    Else
       i++ : Go to the next adjacent of C (n)
End while
End
```

Figure. 6: Generation algorithm of model exercise

As shown in figure (6), Once the user selects a starting concept C(n) and these parameters, the system ask for the output of C(n) and tests the existence of semantic relations to each adjacent of C(n) : C(n)(1), C(n)(1)…C(n)(i)…

- If the high precedence relationship (Rpft(C(n), C(n)(1))) exist, the system ask the output of the first adjacent who highly or lowly substitutes C(n)(1) since the pre-requisites of the concept adjacent to C (n)(1) are equal to or included in the prerequisite C(n)(1). Then the prerequisites of C (n) (1) are required.
- - If the low precedence relationship (Rpfb(C(n), C(n)(1))) exists, we do the same treatment before, but in this case, we asked the prerequisites of C(n) instead of C(n)(1) and we replace Pri_C(n) by Pr1_C(n)(1) Pr2_C(n)(1)… since the output of C(n)(1) is included in the pre-

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

requisites of C(n), the process is repeated for C(n)(2) and so on ...

- Or go to the next adjacent of C(n) and repeats the process.

The generation algorithm is representing schematically in the figure below:

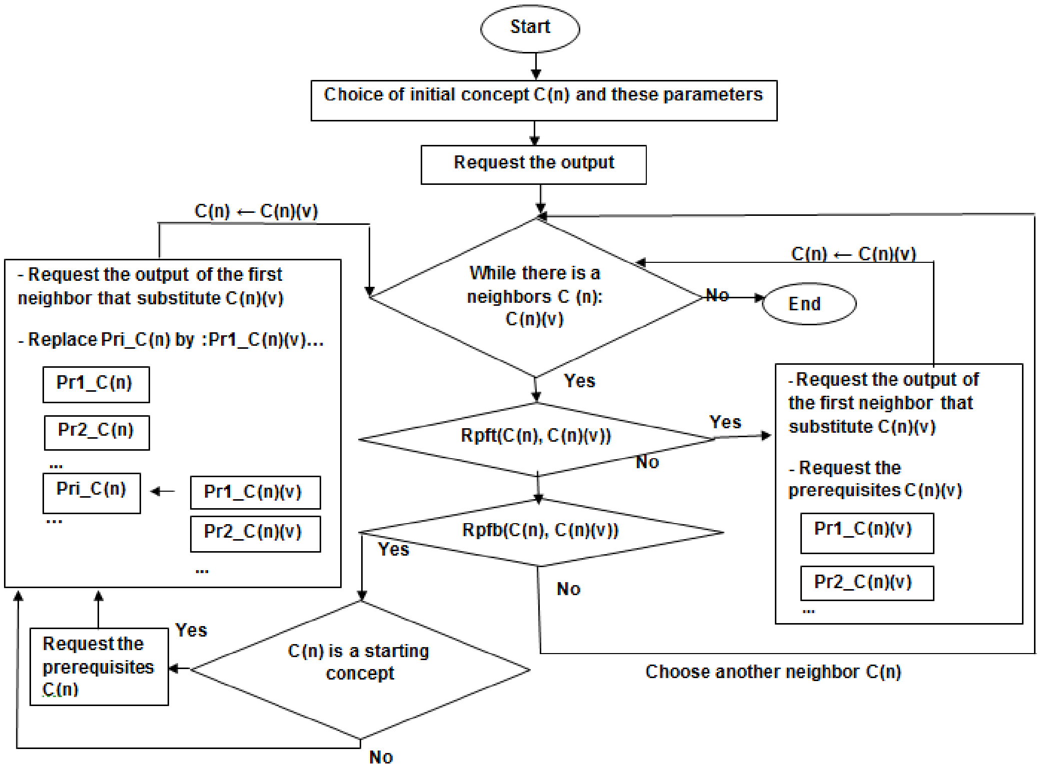


Figure 7: Generation process of model exercise

According to the process of figure (7), we deduce that: more the number of concepts increase, the Exercise model is larger.

3 CASE STUDY

For the case study, we extract the pedagogical objects from a set of a mathematical analysis course [11] for the theme «the sequences». Let T1, T2 ... all theorems, D1, D2 ... Definitions, C1, C2 ... Corollaries, E1, E2 ... Exercises…

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3. Extract of pedagogical Objects

OP	Prerequisites	Output
D1	Un convergesto l in ∞	$\lim_{-} Un = l$
D2	Un diverge to ∞ in ∞	$\lim_{-} Un = \infty$
D3	Un is increasing or decreasing	Un is monotonic
D4	∃ M ∈ R, $Un \le$ M	Un is an above sequence
D5	Un is increasing , Vn is decreasing, $Un \le$ Vn and $\lim_{-} Vn - Un = 0$	Un and Vn are adjacent sequences
D6	$if\ Un+1 \ge Un$ if $Un + 1 \le Un$	Un is increasing Un is decreasing
T1	$Un\ is\ a$ convergent sequence	It's limit is unique
T2	Un and Vn are adjacent sequences	Un and Vn converge to the same limit l, $Un \le$ l $\le Vn$
T3	If Un is increasing and above If Un is decreasing and below	Un converge
T4	If Un is increasing and not above	Un diverges
E1	Un converges	$\lim_{-} Un+1 - Un = 0$
E2	$Un\ is\ a$ convergent sequence	Un is abounded sequence

To build the ontology from corpora, we used the Protégé editor version 4.3.0 [10]. The editor works with a full support for language OWL Web Ontology Language 2, it supports the creation and editing of one or more ontologies into a single workspace via a fully customizable user interface. Visualization tools enable interactive navigation of ontology relationships, and more.

Let T_Prerequisite, T_Output and T_Variable: Prerequisite, Output and Variable of Theorem.
E_Prerequisite, E_Output and E_Variable: Prerequisite, Output and Variable of Exercise.
D_Prerequisite, D_Output and D_Variable: Prerequisite, Output and Variable of Definition.

Journal of Theoretical and Applied Information Technology
31st December 2014. Vol.70 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

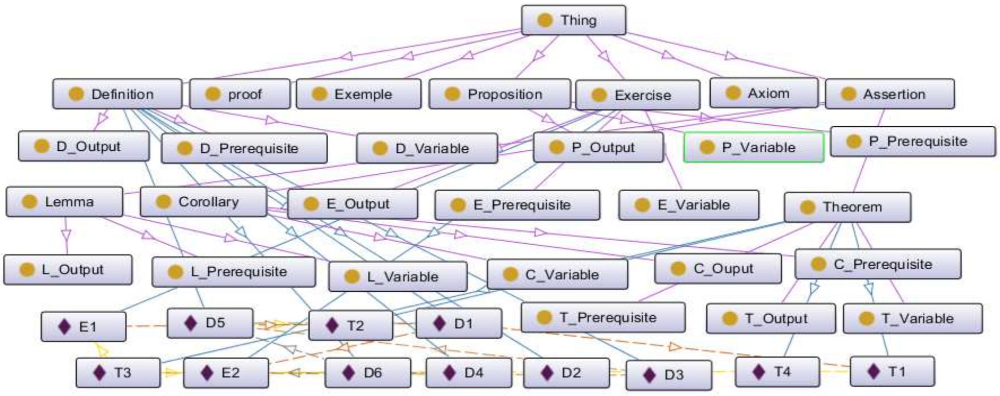
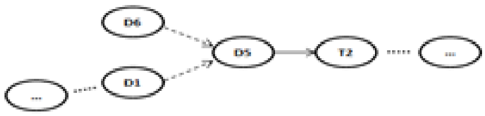


Figure 8: Extract ontology "Sequence" using Protégé

The user selects for example the starting concept « Theorem adjacent suites »: T2.
The values entered by the user satisfy the prerequisites of T2.

: Low precedence relationship.
: High precedence relationship.

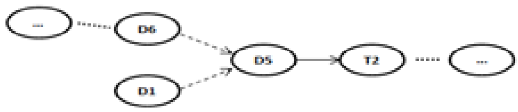
Version 1:



Let (Un), (Vn) two sequences defined by:
U0 = ?, Un =?, Vn = ?, V0 = ?

1) Prove that Un is increasing, Vn is decreasing. …(*)
2) Prove that Vn - Un are a convergent sequences
3) Confirm that $Un \leq Vn$
4) Deduce that $\lim_{\infty} Un = \lim_{\infty} Vn = l$ and $Un \leq l \leq Vn$.

Version 2:



Let (Un), (Vn) two sequences defined by:
U0 = ?, Un = ?, V0 = ? and Vn= ?
…(*)

1) Prove that $U2n+1 - U2n \geq 0$ and $V2n+1 - V2n \leq 0$.
2) Calculate the limit $\lim_{\infty} Vn - Un = 0$.
3) Prove that $Un \leq Vn$.
4) Deduce that $\lim_{\infty} Un = \lim_{\infty} Vn = l$ and $Un \leq l \leq Vn$.

The relation (Low Precedence) between D6 and D5 is the same between D5 and D1 (Table 3) which gave two possibilities to generate an exercise. In version (1),
To generate exercise, we begin with the last question (Question 4) which demonstrates the output of T2. And since the prerequisite of T2 is equal to the output of D5, you can create other questions based on the prerequisites of the D5 (Un is increasing, Vn is decreasing, $Un \leq Vn$ and $\lim_{\infty} Vn - Un = 0$).
To demonstrate the prerequisite « $\lim_{\infty} Vn - Un = 0$ » of D5, we used the prerequisite of D1 (Question 2). In this case, we can use adjacent concepts of D1 to generate more questions (*) by exploiting the semantic relationships.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In version (2), To demonstrate the prerequisite « Un is increasing, Vn is decreasing » of D5, we used the prerequisite of D6 (Question 1).

4 EVALUATION

The absence of a corpus exercises specific to OP leads us to a first manual evaluation of exercises generated. We applied our approach to a set of four instances of different pedagogical objects from a corpus analysis courses consisting of 8 course materials, 13 Tutorials: intermediate value, adjacent sequences, Monotonic sequences, Continuity of numerical functions.

To evaluate our approach, we use: the recall and precision.

$$\text{Precision} = \frac{\text{Number of exercises correctly generated}}{\text{umber of exercises generated}}$$

$$\text{Recall} = \frac{\text{Number of exercises correctly generated}}{\text{Number of exercises generated by experts}}$$

$$\text{F-Measure} = \frac{2.(\text{précision.rappel})}{\text{précision} + \text{rappel}}$$

Table 4: The results of the generation

Type OP	Instance of OP	Precision (%)	Recall (%)	F-Measure (%)
Theorem	intermediate value	83,33	71.42	76,91
Theorem	adjacent sequences	80	66,66	72,72
Corollary	Monotonic sequences	77,77	70	73,68
Definition	Continuity of functions	92,3	75	82,75

We observe that the Precision exceeds 70% for the four instances, which explain that the number of exercises generated by the method converges to the number of exercises generated by experts.

5 CONCLUSION AND PERSPECTIVE

We presented a method for the automatic generation of exercise model based on an educational ontology. We used the semantic relations between pedagogical objects of mathematical corpus to extract a model that exploits several knowledge instead of one. Our approach is applicable to most pedagogical objects of mathematical analysis unless some exception caused by problem of ambiguity of mathematical language. We hope, in future works to:

- Introduce intermediate steps in the model to facilitate the response to the statement in the case of complex parameter.
- Solve the problem of ambiguity using a formal language for alleviating side effects of the natural language.
- Develop a first prototype of our approach.

REFRENCES:

[1] C. Laborde, Natural language and symbolic writing: two codes in interaction in mathematical education, IMAG, Grenoble, 1982.

[2] A. Papasalouros , K. Kanaris and , K. Kotis, Automatic Generation Of Multiple Choice Questions From Domain Ontologies, e-Learning, 2008.

[3] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, 2003.

[4] V. DURAND-GUERRIER, C. MERCAT, and j. ZSIDÓ, Math-Bridge, Issue 131, 2012.

[5] B. Smine, R. Desclés, J. Faiz, Automatic extraction of pertinent pedagogic informations from textual documents. Automatic Natural Language Processing.

[6] N. Meksoud, An approach for reverse engineering and engineering to create, The sharing and reuse of educational content, Third Meeting Young Researchers in Human-Computer Interaction, 2006.

[7] A. Bouzeghoub, B. Defude, J.F. Duitama, C. Lecocq, A model for semantic description of educational resources based on a domain ontology. Science and Technologies of Information and Communication for Education and Training, Volume 12, 2005.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] http://ltsc.ieee.org/wg12/
[9] B. Boukacem, Segmentation mathematical articles for a navigation and improved exploitation of mathematical documents. Thèse de doctorat, INSA de Lyon, 2005.
[10] http://protege.stanford.edu/
[11] http://mathksar.weebly.com/uploads/1/4/4/0/14403348/livre_mp1.pdf