

# A Mathematical Ontology for a Pertinent Research of Didactic Exercises

Imane Lmati, Habib Benlahmar and Naceur Achtaich

**Abstract** This paper presents a semantic annotation of mathematics exercises based on the canonical form of variables or formulas used in the text of exercise. A MathML representation of mathematical formulas has been exploited for the extraction of canonical forms. An extension of the educational ontology Math-Bridge is used for semantic annotation. The ontology consists of all the themes and sub-themes of algebra hierarchically structured by generalization relationships. Each themes of ontology contain useful new attributes for annotation.

**Keywords** Exercise · Evaluation · Ontology · Mathematic

## 1 Introduction

The term “semantic annotation” refers to the activity of fixing the interpretation of a document by associating a formal and explicit semantics [1]. It leads immediately to a multitude of practices as: the comments of the reviewers, the indexing affixed by librarians...

Semantic annotation is one of the best known process in the field of search for knowledge. It was studied by several scientific works dedicated to the extraction of

---

I. Lmati (✉) · N. Achtaich  
Analysis, Modeling and Simulation Laboratory,  
University Hassan II, Casablanca, Morocco  
e-mail: lmati2010@gmail.com

N. Achtaich  
e-mail: nachtaich@gmail.com

H. Benlahmar  
Laboratory of Information Technology and Modeling,  
University Hassan II, Casablanca, Morocco  
e-mail: h.benlahmer@gmail.com

knowledge from logical content such as mathematics. For example, we find the Mias project [2] which operates semantic annotations in the conception and architecture of its system for the recovery of mathematical knowledge. The system adds to mathematics texts (including mathematical formulas) additional representations concerning semantic information (formulas developed as text, canonical text ...). The system is dedicated to research applications that use the library DML (Digital Libraries mathematics). It uses techniques of Natural Language Processing NLP and MathML (“Mathematical Markup Language”) representation.

Another case study was conducted by Kristianto [3]. The approach allows the annotation of scientific articles in XML format for research mathematical formulas represented by MathML. Although these formulas can be indexed and searched by their XML tree structures, they usually do not have enough information to semantic interpretation. The approach provides an annotation model to connect mathematical formulas to descriptions in natural language based text that surrounds it.

The project [4] also studied the semantic annotation; he introduced a new Framework for adding semantics in e-learning system. The proposed approach is based on RDFa [5] and MathML for collaborative annotation of the content of the e-learning and also on ontology to categorize the content of e-learning. The annotation of the Framework adds great value to meet the semantic queries (for example, SPARQL [6]) to retrieve the information requested or desired by a user.

The exercises were also treated in the field of indexing and annotation. The project [7] indexes the geometry exercises by the properties and theorems that serve for their resolution thereby facilitate their research. Indexing is performed using automatic theorem prover Argo, it generates rules (in relation to the themes of ontology of geometry theorems) from the properties that have been provided to it.

## 2 Problematic

Most systems have used semantic annotation for information search. They have annotated the mathematical content by text only (natural language).

The idea is to use logical expressions in the annotation process to facilitate research especially for pedagogical exercises. For the mathematical text, the annotation formalism can be difficult caused by the mix of textual expression and logical relationships. To overcome this problem, we use the ontology Math-Bridge [8] for the textual part and MathML representation for the logic part.

In the following paragraph, we introduce an extension of the educational ontology Math-Bridge [8] useful for annotation. Then we present the semantic annotation algorithm and we conclude with some perspectives.

### 3 Extension of Ontology Math-Bridge

Math-Bridge European project [8] is financed by the european program eContent Plus and project partner institutions. The purpose is to provide a broad base of customized courses in mathematics data, computerized in an online platform. The target group is students in first or second year of post-baccalaureate training, having mathematics in their courses.

During the preparations didactic project, all mathematical themes were organized hierarchically in the form of ontology of concepts relevant to the target group. See for example Fig. 1 for concepts in algebra.

The organization of mathematical concepts in such a tree structure is not obvious.

Some branches of mathematics such as the theory of categories do not appear, because in the beginning of the university, they are not taught in any european country, other mathematical concepts are relevant in one country but not in another.

In our study, we thought to extend the ontology by other attributes and concepts useful for annotation of mathematical exercises. For the topic of digital functions, each polynomial function has a degree and each degree has a canonical form and name.

Let:

$$F(x) = 3x^2 + 2x + 1. \tag{1}$$

**Fig. 1** Ontology Math-Bridge in the editor Protege [9]

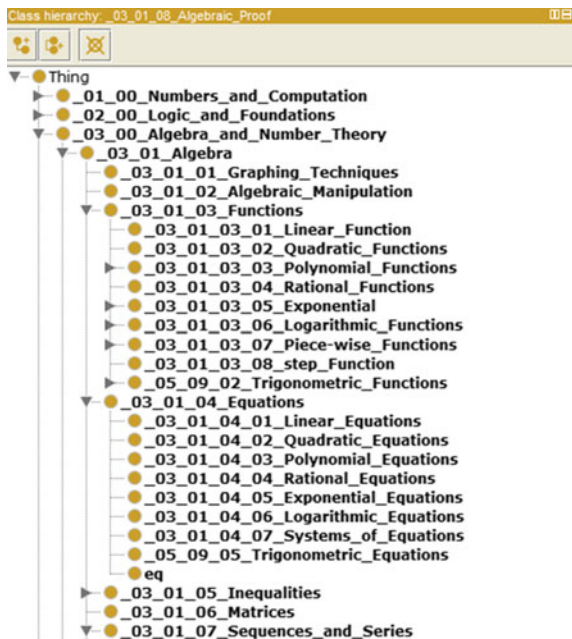




Fig. 2 Extract from the ontology of polynomial functions



Fig. 3 Extract from the ontology of rational functions

This function has the canonical form:  $Ax^2 + Bx + C$ . So we say that  $F(x)$  is a polynomial function of degree 2.

For the trigonometric functions, we can neglect the degree and keep only the canonical form and name. For example the canonical form of a cosine function is:

$$F(x) = \text{Cos}(x). \tag{2}$$

According to the previous examples, we can create other concepts such as Degree, Canonical\_form, and Name (Fig. 2).

The rational functions also have a canonical form  $(P(x))/(Q(x))$  with  $P(x)$  and  $Q(x)$  are polynomials. Since the numerators and denominators are polynomials, we can link them to the concept “polynomial” of ontology (Fig. 3).

Each theme of Math-Bridge ontology contains specific attributes that can be useful for semantic annotation.

## 4 Representation and Annotation Mathematical Formulas

The interpretation of mathematical texts and annotations is a complex process implementing the treatment of different types of information: data acquisition, segmentation data, the structural description of an expression, symbol recognition...

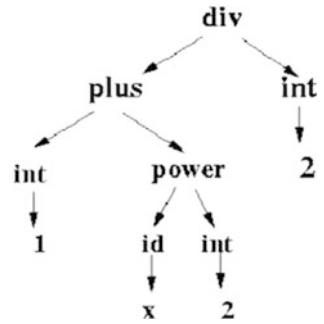
To minimize the work of the treatment, we just interpret the logic part of mathematical text based on abstract syntax tree (According to formalism). It’s very close to the MathML representation (Fig. 4).

Let:

$$F(x) = (1 + x^2)/2. \tag{3}$$

The abstract syntax tree of the function is:

**Fig. 4** Abstract syntax tree of the function (3)



Each node is represented by a mathematical function (div, int, plus...) can be a starting point to bring closer the sub-trees with canonical forms (Fig. 5).

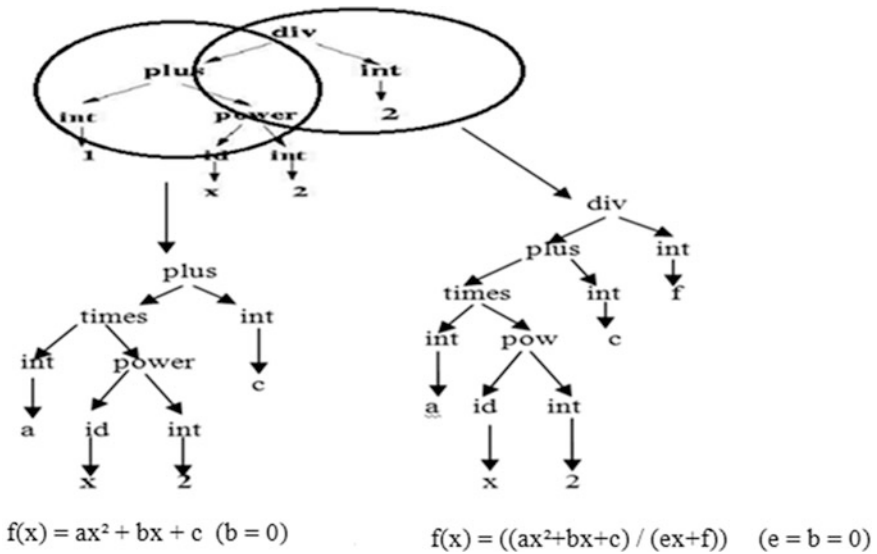
From the abstract syntax tree, we can generate two canonical forms,

- The first is a polynomial function of degree 2:

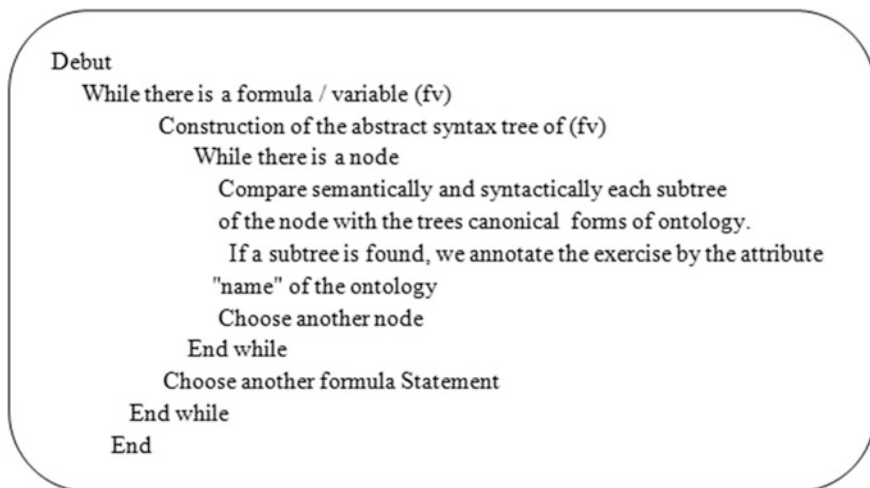
$$(x^2 + 1 \rightarrow ax^2 + bx + c) \quad \text{and} \quad (b = 0).$$

- The second represents a rational function:

$$((x^2 + 1)/2 \rightarrow (ax^2 + bx + c)/(ex + f)) \quad \text{and} \quad (e = b = 0).$$



**Fig. 5** Generation of canonical forms from the function (3)



**Fig. 6** Algorithm for semantic annotation

A semantic and syntactic comparison helps to bring closer the sub-trees with canonical forms, which allows annotating the exercises in several theme of ontology.

The following algorithm examines the canonical form of each formula found in a given exercise. Let (fv) a variable or formula found in the exercise.

As shown in Fig. 6, once the variable or formula (fv) is found, we build its abstract syntax tree:

- If there is a node in the abstract syntax tree, we compare semantically and syntactically each sub-tree of the node with the trees canonical forms of ontology.
- If a sub-tree is found, we annotate the exercise by the name of function (attribute "name" of the ontology) having a canonical forms similar to the one found.
- Or go to the next noeud and repeats the process.

We repeat the process with other formula or variable (fv) used in the exercise.

Semantic annotation requires the extraction of logical expression from mathematical text. The latter is crucial for the annotation, it was studied in several scientific works such as [10, 11]. They are based on labeling, segmentation, classification...

Furthermore, to reduce the complexity of logical expressions, we can develop or use existing patterns to simplify mathematical expressions, which facilitate the semantic and syntactic comparison between the tree MathML and the canonical forms.

## 5 Conclusion and Perspective

In this report, we presented a new method for annotating mathematical exercises based on ontology Math-Bridge. It represents a new support tool for students to target evaluation exercises. The approach uses the canonical form of variables and formulas in the annotation process and not the text as studied in other research. Since we work in mathematical content (logical expression, natural language), the method requires several essential steps to get to the stage of annotation. We hope in future works:

- Conceive patterns for the simplification of complex mathematical formulas.
- Develop a module that allows the semantic and syntactic comparison of mathematical formulas with canonical forms.
- Enrich the ontology by other extensions to facilitate extraction exercises.
- Realize a first prototype of our approach.

## References

1. Ma, Y., Audibert, L., Nazarenko, A.: Ontologies étendues pour l'annotation sémantique. In 20es Journées Francophones d'Ingénierie des Connaissances (2009, May)
2. Sojka, P.: Exploiting semantic annotations in math information retrieval. In: Proceedings of the Fifth Workshop on Exploiting Semantic Annotations In Information Retrieval. ACM (2012, November)
3. Kristianto, G.Y., Topić, G., Nghiem, M.Q., Aizawa, A.: Annotating scientific papers for mathematical formula search. In: Proceedings of the Fifth Workshop On Exploiting Semantic Annotations in Information Retrieval. ACM (2012, November)
4. Doush, I.A., Alkhateeb, F., Maghayreh, E.A., Alsmadi, I., Samarah, S.: Annotations, Collaborative Tagging, and Searching Mathematics in E-Learning. arXiv preprint [arXiv:1211.1780](https://arxiv.org/abs/1211.1780) (2012)
5. RDFa Primer: Bridging the Human and Data Webs, <http://www.w3.org/TR/xhtml-rdfa-primer/> (2010)
6. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>. (Recommendation): W3C (2008)
7. Hibou, M., Labat, J.M., Spagnol, J.P.: Génération de feuilles d'exercices de géométrie à l'aide d'énoncés indexés automatiquement. In: Environnements Informatiques pour l'Apprentissage Humain 2003. ATIEF;INRP (2003, April)
8. Durand-guerrier, V., Mercat, C., Zsidó, J.: Math-Bridge. Gazette des Mathématiciens, Issue 131 (2012)
9. Noy, N.F., Sintek, M., Decker, S., Crubezy, M., Fergerson, R.W., Musen, M.A.: Creating semantic web contents with protege-2000. IEEE Intell. Syst. **2**, 60–71 (2001)
10. Kacem, A., Belaïd, A., Ahmed, M.B.: Embedded formulas extraction. In: Proceedings of 15th International Conference IEEE Pattern Recognition, 2000 (2000)
11. Kacem, A., Belaïd, A., Ahmed, M.B.: Extraction de formules à partir de documents mathématiques. In: Reconnaissance des Formes et Intelligence artificielle-RFIA'2000 (2000)