Supporting Organizations:

The 14th eLearning and Software for Education Conference - eLSE 2018 - organized by the Romanian Advanced Distributed Learning Association is hosted in Bucharest, April 19th - 20th, 2018.

The purpose of the annual international scientific conference on "eLearning and Software for Education" is to enable the academia, research and corporate entities to boost the potential of the technology enhanced learning environments, by providing a forum for exchange of ideas, research outcomes, business case and technical achievements.

The central theme of eLSE 2018 is **"eLearning challenges and new horizons"**.

The conference is indexed by:

ISSN 2360-2198
ISSN-L 2360-2198

All the papers published in this book were verified for plagiarism using the Plagiat-Sistem Antiplagiat prin Internet software

Sistemantiplagiat.ro

14th International Conference eLearning and Software for Education
eLearning challenges and new horizons

Ion ROCEANU
Editor

14th International Conference eLearning and Software for Education

eLearning challenges and new horizons

Volume IV

Bucharest, April 19-20, 2018

The 14th International Scientific Conference
eLearning and Software for Education
Bucharest, April 19-20, 2018
10.12753/2066-026X-18-261

Alignment of Content, Prerequisites and Educational Objectives: Towards Automated Mapping of Digital Learning Resources

Michele FIORAVERA, Marina MARCHISIO
Department of Mathematics, University of Turin, Via Carlo Alberto 10, Turin, Italy
michele.fioravera@unito.it, marina.marchisio@unito.it

Luigi DI CARO, Sergio RABELLINO
Department of Computer Science, University of Turin, Via Pessinetto 12, Turin, Italy
luigi.dicaro@unito.it, sergio.rabellino@unito.it

***Abstract**: The emergence of Technology Enhanced Learning environments has led to the continual growth of the availability of digital educational resources. In this paper, the potential of enabling their reuse into student-centric services – such as recommender systems or adaptive tutoring tools – is discussed through the proposal and comparison of procedures for automatically detecting the mutual relatedness among learning objects. Since the choice of the similarity measure is fundamental for clustering digital materials, this paper addresses the investigation on two distinct approaches: the content-based semantic similarity, compared to the closeness measure on natural language descriptions of metadata – namely prerequisites and educational objectives. The analysis is conducted on a collection of mathematical problems, equipped with metadata which facilitate their retrieval in Virtual Learning Environments, created by Secondary School teachers with the support of University experts. Natural Language Processing techniques are exploited for extracting relevant information from the metadata, while the developments in the emergent field of Mathematical Language Processing are proposed for the treatment of mathematical expressions included in the resources. The distinct similarity measures presented are examined considering the compared results, and their correlation is evaluated. This study is intended to be the first step towards the definition of a model for structuring shared materials available in disciplinary repositories of virtual communities. This model will be used for implementing a system for the delivery of learning objects trajectories on a digital map automatically generated. The system's efficacy will be tested through its integration to a Learning Management System hosting secondary school classrooms' courses. The research is part of a PhD in Pure and Applied Mathematics in apprenticeship, conducted in partnership with leading providers of software based on Computer Algebra System engine.*

***Keywords:** Semantic similarity; Natural Language Processing; mathematical problems.*

I. INTRODUCTION

This paper presents a preliminary study for a project aimed at the implementation of a system for adaptively providing digital materials in a Virtual Learning Community (VLC). The aim of the research is to propose a model to facilitate various types of learning and teaching. This involves structuring shared learning materials, that is automatically matching similarities according to learning goals and pre-requisites knowledge, to provide users with learning object trajectories on a digital map generated from the collection of resources. A trajectory is intended to be a path whose nodes are references to resources available in the VLC, and whose edges are created by matching commonalities between outcome goals and income requirements related to the two learning objects referenced by the nodes.

The clarification of objectives and prerequisites is a fundamental component of student-centred processes of formative and proactive assessment [1]. When these are not explicit – that is declared by the author of the resource, in the form of a metadata – they must be inferred by the content itself. In this paper, Natural Language Processing (NLP) techniques are proposed in order to infer the semantic similarity among learning materials, basing on both external descriptions and content of the resources, and compared clustering analyses are discussed.

II. STATE OF THE ART

Technology Enhanced Learning Environments (TELEs) enable the development of adaptive strategies to help overcome general individual differences among learners, such as incoming knowledge and skills, relevant abilities and disabilities, and demographic and sociocultural differences [2]. Integrating TELEs with semantic-capturing technologies can not only automatically connect learners to appropriate educational materials, but also foster reflections among instructors about the effectiveness of digital-based methodologies adopted in a VLC, based on aggregate analysis of students' results. No fixed learning pathway will be appropriate for all learners [3]: bearing in mind the final goal of designing an approach for the integration of learning materials into personalised learning programmes [4], this paper discusses the use of similarity measures for learning path sequencing [5].

Recent advances on Semantic Web models and technologies of mathematical documents led to the adoption of knowledge-based measures relying on newly available ontologies, which enabled the implementation of recommender systems [6] and adaptive learning systems [7]. Nevertheless, one of the obstacles in the propagation of ontological approaches is the laboriousness of the development of models for each knowledge domain separately. [8] provides a survey of the main semantic methods for the solution of fundamental tasks in mathematical knowledge management, but in literacy there is still not a single widely accepted formalism for computer mathematics, even though mathematics is full of formalisms. On the contrary, the need for a more computable mathematics with robust and standardized representation led to the proposal of increasing the interoperability between existing systems and languages [9].

Inspired by a document-based Mathematical Language Processing pipeline [10], an approach to extract information about mathematical questions from their surface text is presented. This is compared to an approach that considers only external knowledge, designed to help organize collections of materials and facilitate online learners in the design of their learning strategies and their search for the target concepts [11].

III. SEMANTIC SIMILARITY OF MATHEMATICAL PROBLEMS

A *semantic measure* is a mathematical tool used to estimate the strength of the semantic relationship between elements through a (numerical) description obtained from the comparison of information supporting their meaning [12]. Two strategies for the measurement of the semantic similarity among digital learning materials for automatic assessment in mathematics are presented. The first strategy considers what is 'inside' a learning material, while the second only uses information that define a learning material 'from the outside'. Cluster analysis techniques are presented to evaluate the different semantic measures.

The elements of the present analysis are products from web-based assessment platforms: response areas from digital resources for automatic assessment [13]. An example of material of automatic assessment is in Figure 1: it embeds three (related) mathematical problems.

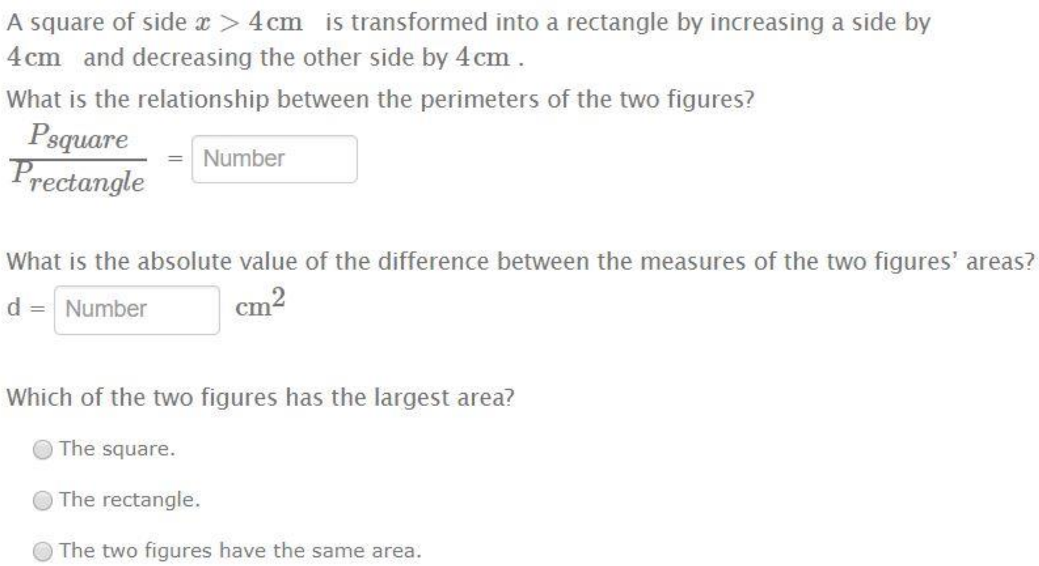

Figure 1. Example of three mathematical problems of a material for automatic assessment

For the sake of this research, each mathematical problem is considered as a single element, even though it is embedded in a question containing more than one response area. The empirical rule for identifying the 'content' (also referred to as the 'surface text') of an element – namely, a mathematical problem – is to consider the first not none content *before* the response area.

3.1 Content-based similarity

Mathematical problems considered in this research contain words, phrases, and formulae. In natural language, words and phrases imply the semantics. Formulae components are regarded as 'words' in the mathematical language and entire formulae as 'sentences'. Formulae components are extracted from the internal data structure representation of a formula. The strength of the semantic relationship which links problems is inferred from observations regarding the distribution of words/components, based on the assumption that semantically related words tend to co-occur.

The mathematical problems were created using the Automatic Assessment System (AAS) Maple T.A. [14]. Subsuming the semantics from the formula structure helps to deal with the numerous different modalities for authoring problems enabled by an AAS: for instance, algorithmic variables allow to have potentially 'infinite' versions of the same question. Maple, the Advanced Computing Environment (ACE) on which Maple T.A. is based, was used for parsing Mathematical problems: a formula is expressed as series of embedded function calls (Figure 2), which can be represented as a Directed Acyclic Graph (DAG).

$$x^2$$

$$\_Inert\_POWER(\_Inert\_NAME(\texttt{"x"}),\ \_Inert\_INTPOS(2))$$

Figure 2. DAG representation of *x square two* in Maple

The semantic similarity between a pair of problems is obtained calculating the cosine similarity for each pair of vectors represented in the term frequency–inverse document frequency (tf-idf) vector space model. The tf-idf model expects a 'bag-of-words' training corpus during initialization. Then it takes a vector and returns another vector of the same dimensionality, except that those features which were rare in the training corpus will have their value increased.

Parsing is the first step for the transformation of problems into their bag-of-words representation, where also math formulae components are referred to as 'words'. The contents of a mathematical problem in the collection are parsed by following these consecutive steps: tokenization, stop words removal, stemming. Ignoring formatting, the input string is demarcated into its component tokens (words). A group of selected stop words is used to filter out common words. The remaining tokens are stemmed. Unique formulae components are extracted and identified by their names from Maple's syntax.

Once the iteration for parsing the collection of mathematical problems is completed, the following phases are executed to implement the bag-of-words model:

1. *Vector space model generation*. The descriptors are transformed into bag-of-words representation: sparse vectors whose values are the token id and its number of occurrences.
2. *Transformation model initialization*. The corpus of vectors is used to initialize the transformation model. The "training" consists in going through the supplied corpus once and computing document frequencies of all its features.
3. *Bag-of-words representation*. The transformation model is used to convert any vector from the bag-of-words representation to the representation based on the tf-idf statistic. The tf-idf model returns real-valued vectors normalized to (Euclidean) unit length.

After these phases, a similarity matrix is constructed by calculating the cosine similarity for each pair of vectors.

3.2 Metadata-based similarity

Automatic assessment provides feedback to both the learner and the instructor about the learning process. Metadata in the form of natural language description are proposed to activate a reflection on the structure of the materials used online: they shall be useful to the teacher both in the design phase and during the research and afterwards. The presence of descriptors associated to a learning material shall also help the development of an instructional strategy, the development and selection of instructional materials, the construction of tests and other instruments for assessing and then evaluating students' learning outcomes. Considering these principles, a model for the clarification of outcome goals and income requirements related to materials for automatic assessment is presented.

Each mathematical problem is equipped with three 'descriptors': *performance*, *requisites*, *objectives*. Resources descriptions are "operationalized", namely they are formulated in an empirically controllable way:

- *Performance* (also known as "instructional objectives", "behavioural objectives" or "learning objectives") is a specific statement of observable behaviours. A well-written performance should meet the following criteria: describe a learning outcome (what the student will be able to do, that *can be observed* directly), be student-oriented (describing the conditions under which the student will perform the task), be observable (indicating criteria for evaluating student's performance). Optionally, a degree of mastery needed can be explicated.
- *Requisites* (or "prerequisites") indicates the learning objectives that should be acquired before attempting to answer the response area. It states the instructor's belief about the necessary and sufficient conditions for providing the element to a student.
- *Objectives* (or "goals") specifies what learners are required to be able to do as a result of a learning activity. The statement should not simply describe a list of topics, that being too abstract, too narrow, nor being restricted to lower-level cognitive skills. 'Action verbs' are suggested to be used to compose this descriptor.

Objectives and *Performance* statements differ in 'general learning objective' and 'specific task' [15]. Differently from *Performance*, *Objectives* do not depend on the type of response area embedded in the question.

Referring to the learning material shown in Figure 1, an example of *Performance* for the first mathematical question would be: "Given the values of the side of a square and of the base and height of a rectangle – all expressed through monomials and binomials – the student will be able to write the relationship between the perimeters of the two polygons". The third question could have the following *Requisites*: "Knowing the concept of perimeter and how to calculate it for squares and rectangles". An example of *Objectives* for the second question would be: "Comparing polygons by the calculus of areas".

A similarity measure based on metadata in the form of natural language descriptions is proposed. Each descriptor gives rise to a distinct measure.

The semantic similarity between a pair of problems is obtained similarly by following the phases previously described for the Content-based similarity, except for the fact that the parsing step does not involve the treatment of formulae.

IV. RESULTS AND DISCUSSION

This research considers materials created for providing automated formative assessment in Mathematics and scientific disciplines [16]. 196 mathematical problems from 93 questions for automatic assessment were selected. Two experts separately inserted the triples of descriptors to each of the 196 mathematical problems. The questions were prepared by secondary school teachers in the VLC of the Problem Posing and Solving (PP&S), with the support of expert tutors for classrooms usage. PP&S is a community of teachers and a community of virtual classrooms for Italian Secondary Schools courses: teachers collaborate in the creation of learning materials, manage Moodle courses for their students and share learning materials through tools available as Moodle integrations or other resources of the common courses, under the supervision of tutors – experts in the use of digital tools.

The collection has the following characteristics.

- Mathematical problems concern the following 8 disciplinary areas: Contextualized problem about Algebra (4), Monomials (68), Polynomials (38), Special products (24), Contextualized problem about Probability (7), Statistics (36), Probability (13), Contextualized problem about Statistics (6).
- Each question expects the input response in a different format, according to the question type. For example, a Maple-graded area expects a mathematical expression as response, to be evaluated by a customizable grading code. Mathematical problems can be grouped by their response area type: Maple-graded (23), Matching (1), Mathematical formula (57), Multiple choice (32), Multiple selection (13), List (23), Essay (3), Numeric (44).

These features provide two 8-clusters labellings to be compared with the results from the clustering algorithm executed on the similarity matrices generated from the distinct similarity measures as input. Mini Batch k-Means is the clustering algorithm chosen [17]. It returns a list of labels: each mathematical problem is labelled with one out of k clusters, where k (set to 8) is the number of clusters to be generated.

To enhance the influence of semantically relevant concepts, a thesaurus of mathematical terms (both natural language words and formulae components names) and action verbs was manually constructed and used to filter tokens: the parsing step described in section 3.1 is affected by the following rules:

1. Words that appear in less than 2 input strings are filtered out.
2. Words that appear in more than the half of the input strings are filtered out.
3. Words are kept regardless of the previous rules, if they belong to the set of concepts contained in the thesaurus.
4. After the previous rules, only the first j most frequent words are kept.

The clustering algorithm was executed firstly without filtering, then with the value of j between 7 and 15 in steps of 2. In all the cases, the experiment was repeated 10 times.

The range for number j was chosen considering the average lengths of the vectors generated. The vectors generated from surface texts vary from 1 to 51: 10 is the approximated value of both its mean and standard deviation. On average, the approximate value of the length of the vectors generated from the descriptors is 6 for *Requirements* and *Objectives*, while it is 14 for the *Performance* of the first author and 10 for the performance of the second author.

Clusterings are compared with the 'ground truth' labellings as follows (*dGT* stands for the labelling by disciplinary area, *rGT* stands for the labelling by response area type). Figure 3 and Figure 4 show the mean values of the v_measure scores (in what follows, standard deviation values are about two orders of magnitude smaller than the means) between a ground truth labelling and each of the clusterings: the clustering by surface text (*ST*), and the clusterings by each combination of descriptor (*P* stands for *Performance*, *O* for *Objectives*, *R* for *Requisites*) and author (*1* stands for the first author, *2* for the second). The homogeneity metric v_measure expresses how successfully *homogeneity* and *completeness* criteria have been satisfied among two clusterings. A clustering satisfies *homogeneity* with respect of the classes of a second clustering if all of its clusters contain only data points which are members of a single class. It satisfies *completeness* if all the data points that are members of a given class are elements of the same cluster [18].

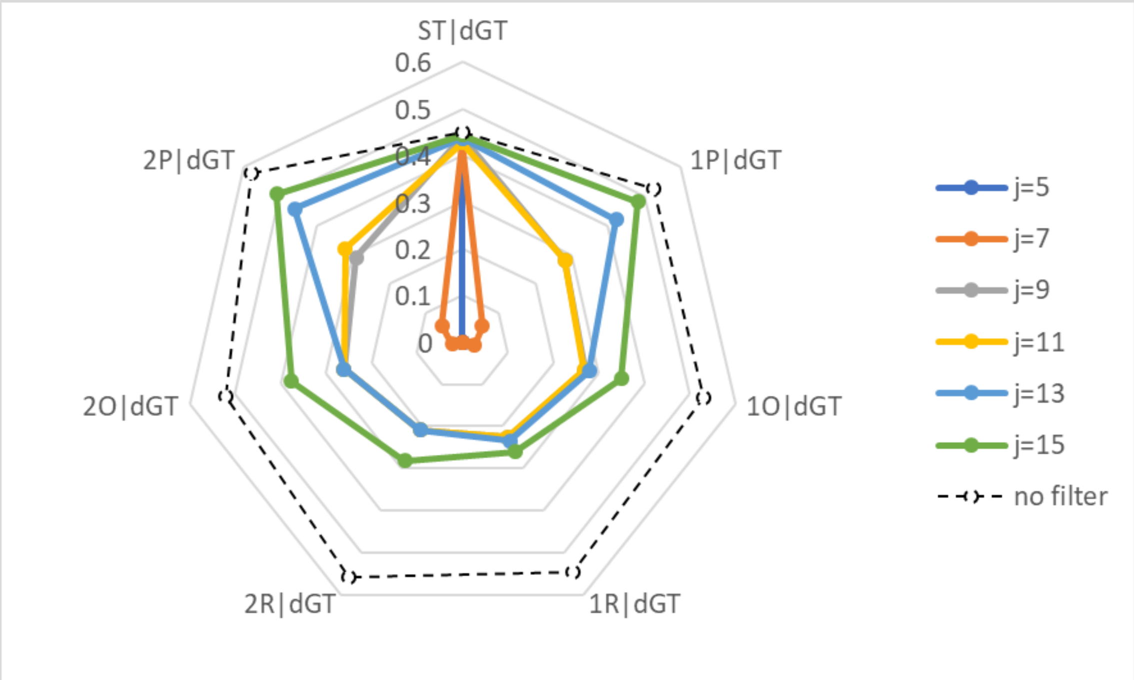

Figure 3. Mean values of the *v_measure* between the ground truth labelling by disciplinary area and different clusterings, in case of *k* = 8 and *j* between 7 and 15 in steps of 2

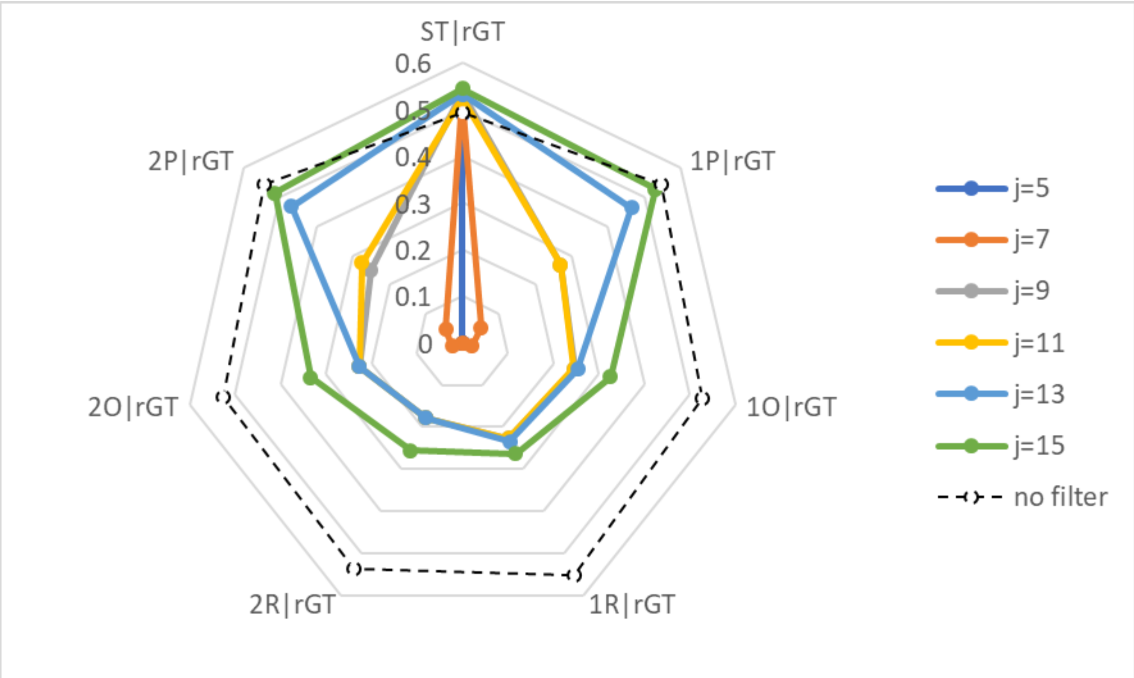

Figure 4. Mean values of the *v_measure* between the ground truth labelling by response area type and different clusterings, in case of *k* = 8 and *j* between 7 and 15 in steps of 2

The results highlight that clusterings generated by the descriptors highly reflect both the disciplinary area and the response area groupings (daGT and raGT) of the collection of mathematical problems, since the *v_measure* mean value – in case of no filtering – is higher than 0.5.

The influence of the filtering affects the descriptors accordingly with their respective average lengths of the vectors generated. Few relevant concepts can be found in the short statements which define the metadata-based similarity. On the contrary, content-based similarity appears not to be influenced by filtering. This is in line with the fact that only 8.7% of mathematical problems of the collection are contextualized in “real life situation”: the majority of them are explicitly expressed questions, possibly with formulae which enrich the semantic expressivity.

The simmetry of the graphs suggests a certain degree of inter-annotation agreement between the two authors.

Little correlation appears between the clusterings from metadata-based similarity and the clustering from content-based similarity, since the *v_measure* mean value is approximately lower than 0.3 (Figure 6). This suggests that these two approaches express different concepts of mathematical problems.

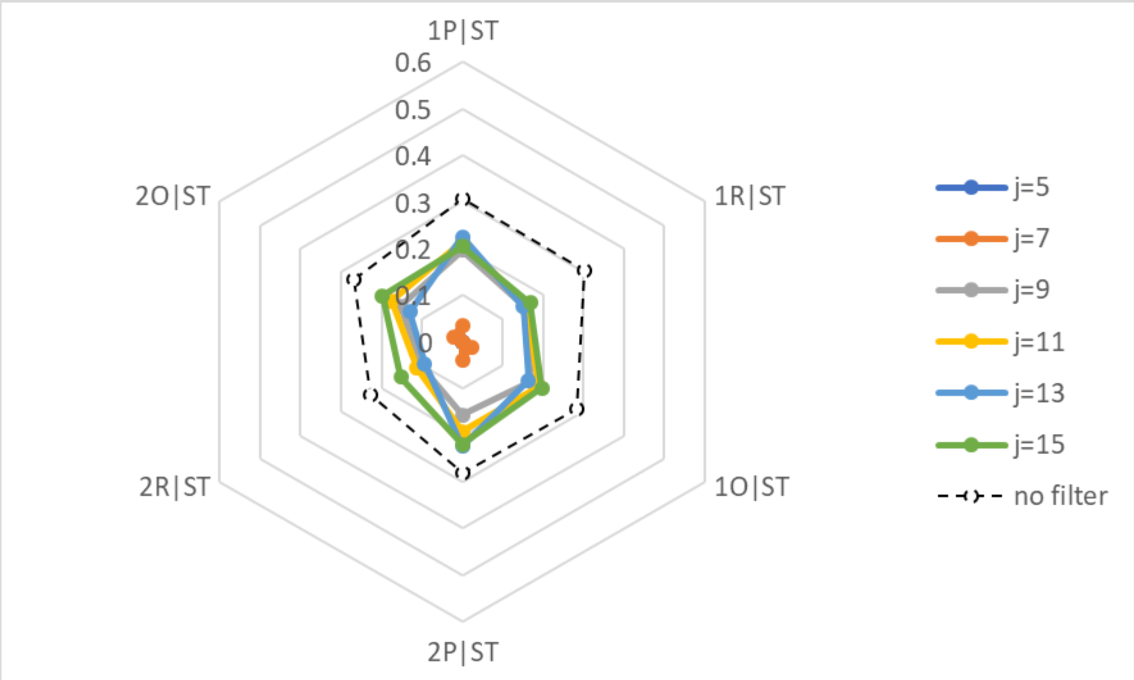

Figure 5. Mean values of the *v_measure* between clusterings from metadata-based similarity and the clustering from content-based similarity, in case of k = 8 and j between 7 and 15 in steps of 2

To guarantee the quality of the clusterings obtained, *1P*, *1R*, *1O*, *2P*, *2R*, *2O* and *ST* are compared to randomly generated clusterings (*Rdm*). The *v_measure* mean values are significantly close to zero (Figure 6), showing that the semantic similarity clusterings can be considered reliable.

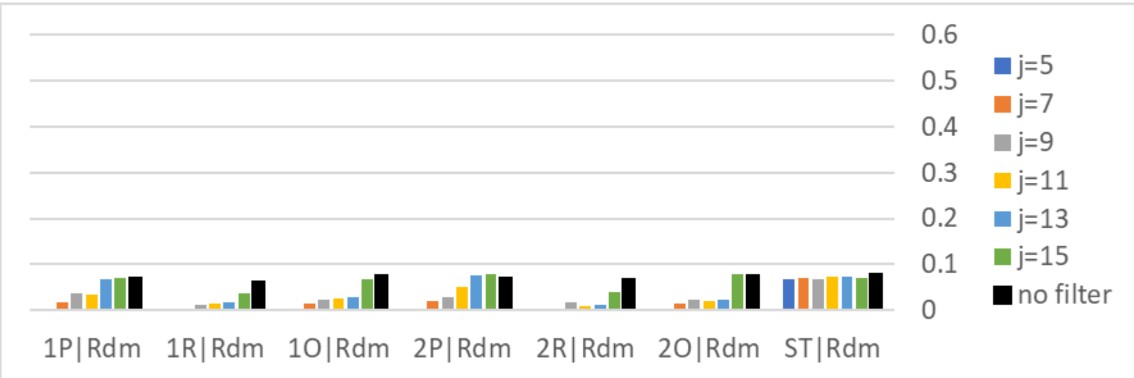

Figure 6. Mean values of the *v_measure* between the random labelling and each of the clusterings, in case of k = 8 and j between 7 and 15 in steps of 2

V. CONCLUSIONS

Student-centric services for the reuse of shared resources – such as recommender systems or adaptive tutoring tools – could be based on both similarity measures presented in this paper. Results showed how little they are related, together with their respective strong alignment with grouping established in the design phase. This led to the proposal of a semantic measure based on a mix between corpus-based and knowledge based, which will be the subject of further research.

This research will grow by activating projects at national and European scale, involving instructors and students to further investigate mathematical problems authoring and sharing.

The research is part of a PhD in Pure and Applied Mathematics in apprenticeship, conducted in partnership with leading providers of software based on ACE engine.

This work has been conducted by using Gensim, the software to realize unsupervised semantic modelling from plain text licensed under the OSI-approved GNU LGPLv2.1 license [19], nltk, the platform for building Python programs to work with human language data [20], and scipy [21].

Reference Text and Citations

[1] Trinchero, R., 2014. Valutare l'apprendimento nell'e-learning, Edizioni Centro Studi Erickson.

[2] Shute V., Zapata-Rivera D., 2012. Adaptive Educational Systems. Adaptive Technologies for Training and Education. 7-27. 10.1017/CBO9781139049580.004

[3] Chen, C.M. Peng, C. J. Shiue, J. Y., 2009. Ontology-based concept map for planning personalized learning path. 2008 IEEE Conference on Cybernetics and Intelligent Systems, Chengdu, 2008, 1337-1342.

[4] Tugsgerel, B., Anane, R., Theodoropoulos, G., 2010. B. (2010). An Integrated Approach to Learning Object Sequencing. 10th IEEE International Conference on Advanced Learning Technologies, Sousse, 2010, pp. 105-109.

[5] Shmelev, V., Karpova, M., Dukhanov A., 2015. An Approach of Learning Path Sequencing Based on Revised Bloom's Taxonomy and Domain Ontologies with the Use of Genetic Algorithms. Procedia Computer Science. 66. 711-719

[6] Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A., Solovyev V.D., Zhiltsov N.G., 2014. Mathematical Knowledge Representation: Semantic Models and Formalisms. Lobachevskii Journal of Mathematics. 35.

[7] Tam V., Lam E. Y., Fung S.T., 2014. A new framework of concept clustering and learning path optimization to develop the next-generation e-learning systems. *In Journal of Computers in Education*, Vol. 1, N. 4, 335.

[8] Elizarov A. M., Zhizhchenko A.B., Zhil'tsov N.G., Kirillovich A.V., Lipachev E.K., 2016. Mathematical knowledge ontologies and recommender systems for collections of documents in physics and mathematics. *In Doklady Mathematics*. 93. 231-233.

[9] Ion P., Trott M., Weisstein E., Wiedijk F., 2016. White Paper of the Semantic Representation of Mathematical Knowledge Workshop

[10] Schubotz M., Grigorev A., Leich M., Cohl H., Howard S., Meuschke N., Gipp B., Youssef, A., Abdou S., Markl V. 2016. Semantification of Identifiers in Mathematics for Better Math Information Retrieval. ACM SIGIR, Jul 18-20, 2016 Pisa.

[11] Huang, C., Li, X., Lei, W., Pong, T., Wang, F., Yin, M., 2015. Constructing Learning Maps for Lecture Videos by Exploring Wikipedia Knowledge. PCM.

[12] Harispe S., Ranwez S., Janaqi S., Montmain J., 2015. Semantic Similarity from Natural Language and Ontology Analysis.

[13] Petri I., 2011. Automated Assessment of Programming Assignments: Visual Feedback, Assignment Mobility, and Assessment of Students' Testing Skill.

[14] Barana A., Marchisio M., Rabellino S., 2015. Automated Assessment in Mathematics. COMPSAC Symposium on Computer Education and Learning Technologies (CELT), Taichung.

[15] Gronlund N., 2004. Writing Instructional Objectives for Teaching and Assessment. 7th ed. Upper Saddle River, New Jersey. Merrill Prentice Hall.

[16] Barana, A., Marchisio, M., 2016. Ten good reasons to adopt an automated formative assessment model for learning and teaching Mathematics and scientific disciplines. Procedia-Social and Behavioral Sciences, Vol. 228, 608-613.

[17] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E., 2011. Scikit-learn: Machine Learning in Python. JMLR 12. 2825-2830.

[18] Rosenberg A. and Hirschberg J., 2007. "V-Measure: A conditional entropy-based external cluster evaluation measure", Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 410-420.

[19] Řehůřek, R., Sojka, P., 2010. Software Framework for Topic Modelling with Large Corpora. 45-50. 10.13140/2.1.2393.1847.

[20] Bird S., Edward L., Ewan K., 2009. Natural Language Processing with Python. O'Reilly Media Inc.

[21] Jones E., Oliphant E., Peterson P., et al. SciPy: Open Source Scientific Tools for Python, 2001-, http://www.scipy.org/ [Online; accessed 2018-01-29]