A Machine Learning Approach to Identify Dependencies Among Learning Objects

Carlo De Medio[1], Fabio Gasparetti[2], Carla Limongelli[2], Filippo Sciarrone[2] and Marco Temperini[1]

[1]*Sapienza University of Rome, Dept. of Computer, Control and Management Engineering, Via Ariosto, 25 - 00189 Rome, Italy*

[2]*Engineering Department, Artificial Intelligence Laboratory, Roma Tre University, Via della Vasca Navale, 79, 00146 Rome, Italy*

Keywords: E-learning, Data Mining, Wikipedia.

Abstract: Selecting and sequencing a set of Learning Objects (LOs) to build a course may turn out to be quite a challenging task. In this paper we focus on such an aspect, related to the verification and respect of the relationships of pedagogical dependence existing between two LOs added to a course (meaning that if a given LO has another one as "pre-requisite", then any sequencing of the LOs in the course will need to have the latter LO taken by the learners before of the former). In our approach the sequencing of LOs in the course can still be managed by the instructor, basing on her/his taste and preferences, yet s/he can also be helped by a set of suggestions, related to the pre-requisite relationships existing among the LOs selected for the course. Such suggestions (such relationships, in effect) can be computed automatically and provide the instructor with significant help and guidance. We show a light-weight formalization of the LO, and how it can be "represented" by a set of WikiPedia Pages ("topics"); then we show how such set of topics, together with a set of relevant hypotheses we previously defined, can help establish the dependence relationship existing between two LOs. In this endeavor we exploit the classification in categories available for the WikiPedia topics, and obtain interesting results for our framework, in terms of precision and recall of the dependence relationships.

1 INTRODUCTION

When an instructor is faced with the task of building or maintaining a web-based course, her/his work can be complex, in several respects. On the one hand, the availability of an ever increasing amount of educational material on the web offers a problem of sheer quantity of possible choices. An idea of how vast and, possibly, confusing the selection task for an instructor can be, is given by simply limiting the consideration to learning material available under the formalized shape of standardized Learning Objects (LOs), from Learning Object Repositories (LORs) such as Connexion[1], Ariadne[2], or Merlot[3]. In this kind of task, recommender and filtering tools might be of help (Revilla Muñoz et al., 2015; Limongelli et al., 2010; Limongelli et al., 2012; Limongelli et al., 2015; Limongelli et al., 2016).

On the other hand, one of the main responsibilities of the instructor, while assembling a course through LOs, is to ensure the fundamental pedagogical aspects of the course, such as the preservation of the existing relationships of dependence between two LOs in the course: in other words, to ensure that a LO_j in the course is depending on another object LO_i in the course (i.e. "LO_j has LO_i as a prerequisite"), then LO_i will precede LO_j in every admissible sequencing of the course's LOs.

How the instructor sequences the learning material strongly depends on the actual learning content of the LOs, as well as on a teacher's taste/preference. Nevertheless, having automated suggestions on how certain LOs should be necessarily sequenced, in order to preserve dependency relationships, can be of great help for the instructor, since it can ease a part of the selection and sequencing task, and allow the instructor to focus on less automatable aspects.

During course construction the instructor usually labels each selected LO with a set of pre-requisite concepts (knowledge and/or skills that will definitely

[1]Connexions is a Learning Object Repository, available at http://www.cnx.org (Accessed July 24 2015).

[2]Ariadne Foundation, available at http://ariadne-eu.org (Accessed July 24 2015).

[3]Merlot is a Learning Object Repository, available at http://www.merlot.org (Accessed July 24 2015)

Medio, C., Gasparetti, F., Limongelli, C., Sciarrone, F. and Temperini, M.
A Machine Learning Approach to Identify Dependencies Among Learning Objects.
In *Proceedings of the 8th International Conference on Computer Supported Education (CSEDU 2016) - Volume 1*, pages 345-352
ISBN: 978-989-758-179-3

need be covered prior to the learner engaging the LO). These pre-requisites statements provide a set of constraints that have to be verified during the course sequencing definition. Because of the intrinsic dependence of such constraints on the LOs' contents, neither their definition, nor their evaluation is a trivial task, the former possibly being quite time-consuming and error-prone.

In a previous work (Gasparetti et al., 2015b) we have proposed six hypotheses that help to decide, given two LOs, their prerequisite relationship. We proposed and validated the hypotheses (described in Section 3) and showed that, although these are statistically significant, we get a low precision, at most 0.4.

In this paper, starting from the results of the previous work, we propose a different study based on a traditional machine learning approach (Mitchell, 1997). We expand the set of data on which the analysis is done, and identify some features derived from the above assumptions and feed the training system (Weka).

The protocol of application of the criteria/features, proceeds by passing the learning objects of a course through Wikipedia Miner (Milne and Witten, 2013), collect data and then feed Weka (Hall et al., 2009) in such a way to obtain a set of suggestions about the possible dependence relationships among LOs and wiki pages. The results show good precision and recall values, increasing with the number of LOs employed in the course. This led us to think that our protocol can provide the teacher with valuable suggestions about the possible sequencing of their learning material.

In the following section we present some related work. Then, in Section 3 we illustrate the six revised hypotheses on which we base the inference of dependence relationships; in this section we also provide a lightweight formal background to discuss the characteristics of a LO, based on its representation as set of Wikipedia Pages (*topics*). Here we also discuss the quantitative features computed for a LO during the relationships inference. Section 4 presents the results obtained through the use of Weka (Hall et al., 2009), that we trained with five different sets of LOs. Section 5 comments on the experimental results and proposes how to reinforce and generalize the present results.

2 RELATED WORK

Associating learning material, of diverse origin, in one's own course material is a delicate task, since the learning material is often not to be treated as a mere additive on the activities proposed to students, yet the new resources have to undergo some pedagogical adaptation.

Curriculum Sequencing is an interesting challenge in the educational research area: research in this field aims to automatically produce a personalized sequence of didactic materials or activities (Brusilovsky and Vassileva, 2003), or to allow a sequence of instructional material "on demand", by means of systems that deliver training to workers when and where they need it (Capuano et al., 2009). Our proposed approach focuses on the possibility that the learning objects through which the teacher makes her selection have already been labeled as to their possible dependency relationships. The teacher has full control of such dependency determinations: s/he can still accept or discard such suggestions. On the other hand, such hints can also simply add information that the teacher had not noticed, resulting even more fruitful. Such assistance is admittedly specific, even narrow, yet it can still be of good use and let the teacher save time to be dedicated to other pedagogical aspects.

Resources from the Web, and in particular from Wikipedia, have been investigated as a source for enrichment of the learning contents of a LMS or LOR (Parker and Chao, 2007; Cole, 2009; Stuurman et al., 2012; Gentili et al., 2001; Sciarrone, 2013). Other kinds of resources have also been considered for automated treatment and inclusion in the learning resources of a course, such as the podcasts (Cebeci and Tekdal, 2006), and the use of wikis also as means for contribution from students into the course's material (Allen and Tay, 2012; Sun and Qiu, 2014).

The task of including such external material in a course remains, however, a hard one, both on a technical and pedagogical level.

Wikipedia provides a wealth of information and possibly learning resource documents, so the idea of using such resources to enrich a course's learning material is compelling. In addition Wikipedia is instrumented on the web, so that the analysis and evaluation of semantic relationships between documents and concepts is supported, down to the comparison with words or text excerpts (Strube and Ponzetto, 2006; Gabrilovich and Markovitch, 2009; Milne and Witten, 2008; Biancalana et al., 2013). This can allow for an effective treatment of the wikipedia contents in advance of their uses to support learning in a formal environment.

In this paper we propose a method to exploit the informative resources of weak-semantic taxonomies, in particular Wikipedia, to allow sequencing LOs with external resources, and to make the LOs enrichment possible, by annotating them through the external resources.

To our knowledge our approach is novel, yet we have found correspondences in literature.

An approach to the identification of prerequisite relationships among "knowledge components" is in (Scheines et al., 2014), where causal discovery is used on components represented as latent (unmeasured) variables. To validate the approach, simulated data are used, representing a dataset of student-skills measures.

Voung *et al.* (Vuong et al., 2011) propose to analyze large-scale assessment to determine the dependency relationships between knowledge units. Given sufficient user data, the authors prove that prerequisites for each instructional unit can be identified. On the contrary, the methodology cannot be applied to new curriculums, that is, units to which student performances have not been extensively evaluated.

Recently (Gasparetti et al., 2015c) proposed an early attempt to exploit Wikipedia as a source of learning materials. Analyzing the links present in the Wikipedia pages, they build courses based on the Grasha teaching styles and on a social didactic approach. A further early attempt to exploit Wikipedia for the sequencing task has been proposed in (Gasparetti et al., 2015a).

3 HYPOTESES AND FEATURES FOR COMPARING LOS

In the previous work (Gasparetti et al., 2015b) we have proposed an approach for extracting prerequisite relations from textual LOs by means of wikipedia miner (Milne and Witten, 2013) that is a tool to access to Wikipedia's structure and content.

The process goes from information extraction out of the involved LOs, through determination of the WikiPedia pages (topics henceforth) that are associated to each one of the LOs, up to the analysis of a set of features of such topics, allowing to establish the possible dependency relationships between the LOs. It can be summarized as follows:

- given the learning objects LO_i and LO_j, the Wikipedia Miner Toolkit is activated on them;
- for each LO the textual content is extracted and analyzed, so as to pair portions of it with annotations, relating to categories in the Wikipedia taxonomy (made by metatags inside the Wikipedia's pages and perfected by the community of Wikipedia, it is the division into categories of the information; the system generates a graph of the categories that can be questioned without fetch the whole page);
- for each LO, the set of annotations is used to relate the LO to a set of topics;
- then we apply certain criteria of evaluation to the two sets of topics associated to LO_i and LO_j;
- we infer the existence of dependency relationships on the basis of a set of hypotheses.

The dependency relation of prerequisites is expressed as $LO_i \rightarrow LO_j$ meaning that LO_i is a prerequisite for LO_j.

In the following we provide a revised set of hypotheses and explain the features of the topics associated with the LOs, which we use to test such Hypotheses.

H1
if the distance that connects the Wikipedia two categories to which LO_i and LO_j belong respectively (even through a common ancestor category) is less than a given threshold, the prerequisite relation exists. The threshold is set to 2, with values higher than 2 the execution of the program slows down dramatically and the performance is not improved. The hypothesis is inspired by the semantic similarity defined over a IS-A taxonomy proposed by Resnik (Resnik, 2011).

H2
more general topics need longer discussions to be described, compared to very specific ones.

H3
if a Wikipedia author has found the necessity to define a concept by means of other concepts, probably the former one is more specific than the latter.

H4 more general concepts create several connections to specific ones, especially if the author wants to give an overview of the domain leaving it up to the reader to deepen discussions on specialized pages.

H5 In this hypothesis we extract the nouns from the two articles and assume that nouns in a Wikipedia article correspond to concepts. Therefore, articles dealing with multiple concepts should be considered more general.

H6 The last hypothesis analyzes the length of LOs in terms of number of words included in the description (i.e., first paragraph) of the articles. If the number of words of the article associated with LO_i is much greater than the ones of LO_j, $LO_i \rightarrow LO_j$ exists. Again, we follow a similar rationale of the hypotheses H4 and H5, but the computational complexity is limited in this case.

In the proposed approach, the features the hypotheses are based upon are considered as an input of a machine learning algorithms. Eleven features representing relevant aspects of each LO's are defined. Before describing the features we introduce some definitions.

3.1 Formal Background

Following a trend used in several sources, a WikiPedia Page associated to a LO will be named "topic".

In a *topic* anchors are spread, to connect such resource to other topics. Such references will be named "links".

For our purposes a *link* is then a URI (*Uniform Resource Identifier*, i.e. a usual web address). Notice that another significant link is the topic address.

As a web resource, the topic comprises multimedia, yet in our approach we will consider only its textual contents; such textual contents are basically composed by all the topic's text and links to other topics. We differentiate between simple word and a nouns , they will be analyzed separately as suggested in $H5$.

Moreover, a Wikipedia page usually comprises a *first section*, containing a summary of relevant aspects of the page.

Here are some definitions for understanding the formal background:

T the set of all the topics t associated to a LO

L the set of all the links to other topics in t: $t.L$

W the set of all the words used in t: $t.W$

N the number of the links to other topics in t: $t.N$

addr the topic address (another link): $t.addr$

FS the first section of the topic in its usual web-publication: $t.FS$. It comprises sets of links, nouns and words, that we consider subsets of, resp., the links, words and nouns of the topic:

- set of links in the first section: $t.FS.L \subseteq t.L$
- set of words in the first section: $t.FS.W \subseteq t.W$
- set of nouns in the first section: $t.FS.N \subseteq t.N$

3.2 Features of a LO

Given two learning objects LO_i and LO_j, the features can be formalized as follows:

1. $avgLen(LO_i)$: The average length of the text of the Wikipedia topics associated to LO_i defined in terms of words obtained by a text tokenization process.

$$avgLen(LO_i) = \frac{\sum_{t \in T_{LO_i}} |t.W|}{|T_{LO_i}|}$$

2. $avgLen(LO_j)$: Similar to $avgLen(LO_i)$ but evaluated on LO_j.

$$avgLen(LO_j) = \frac{\sum_{t \in T_{LO_j}} |t.W|}{|T_{LO_j}|}$$

3. $fsl(LO_i)$: number of links in the first section of the Wikipedia topics associated with LO_i

$$fsl(LO_i) = t.FS.L$$

4. $fsl(LO_j)$: number of links in the first section of the Wikipedia topics associated with LO_j

$$fsl(LO_j) = t.FS.L$$

5. $avgNL(LO_i)$: The average number of links in the topics associated to LO_i

$$avgNL(LO_i) = \frac{\sum_{\forall t \in T} t.L}{|T|}$$

6. $avgNL(LO_j)$: Similar to $avgNL(LO_i)$ but evaluated on LO_j.
7. $nouns(LO_i)$: Number of distinct nouns in LO_i extracted by the part-of-speech tagger.
8. $nouns(LO_j)$: Similar to $nouns(LO_i)$ but on LO_j.
9. $nounsIntersect(LO_i, LO_j)$: Given the two sets of nouns N_i and N_j extracted from LO_i and LO_j

$$nounsIntersect(LO_i, LO_j) = |N_i \cup N_j|$$

10. $avgFsLen(LO_i)$: The average length of the text of the Wikipedia topics associated to LO_i defined in terms of words obtained by the tokenization process limited to the first section of the topics

$$avgFsLen(LO_i) = \frac{\sum_{\forall t \in T} t.FS.W}{|T|}$$

11. $avgFsLen(LO_j)$: Similar to $avgFsLen(LO_i)$ but evaluated on LO_j.

All the features are represented by real or integer numbers.

4 EXPERIMENTAL RESULTS

The system, programmed to operate on the features described above, was tested through WEKA (Waikato Environment for Knowledge Analysis), a free software developed by the University of Waikata in new Zealand which allows to apply Machine Learning algorithms on large data-sets. In our case the training set consists of 5 courses structured as shown in Table 1.

The courses were chosen in order to have samples differing both in terms of number of LOs and in terms of content and materials. Some of the LOs content presents a very formal language (e.g.: Programming Language (Java) and Advanced Computer Science) and other LOs show historical references and

Table 1: Description of test courses.

Course Name	N. of LOs	N. of expected dependences
Advanced Computer Science	85	2256
Programming Language (Java)	18	41
Economy	5	10
Futurism	4	5
Basic Mathematics	4	5

descriptive content (e.g.: Futurism). The expected dependencies are the relationships between prerequisite and successor concepts represented by LOs. The LOs are represented by text files containing the entire text of the lessons; the system is implemented so as to accept `html` pages, automatically retrieved (or not) by the network. They are parsed and return the textual file automatically. The entire data-set was classified by the algorithm J48, an open source implementation of the algorithm C4.5, which in the first tests was the most encouraging one. Other algorithms are applicable to data-sets but require larger data sets to see significant results. The measures taken from the performance analysis of the implemented system are the following:

- π Precision in [0..1]: high precision means that an algorithm returned more relevant results than irrelevant results;
- ρ Recall in [0..1]: high recall means that an algorithm returned most of the relevant results;
- $F1$ F-measure is a combination of precision and recall, namely the harmonic mean;
- *ROC* Is the area under the ROC curve normalized between [0..1] and it is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

In addition, given a pair of LOs: LO_1 and LO_2 we distinguish two classes results for the classifier:

- Class 1 (Cl1) Set of all pairs: LO_i, LO_j for which there is the prerequisite relation $LO_i \rightarrow LO_j$;
- Class 2 (Cl2) Set of all pairs: LO_i, LO_j for which there isn't any prerequisite relation.

We carried out tests divided into courses; for each test we show the measures described above and the confusion matrix. This matrix is a specific table layout that allows to visualize the performance of an algorithm: each column of the matrix represents the instances classified by the system for a given class, while each row represents the instances in the expected class. The classifier works on the two classes described above, so the corresponding confusion matrix will be 2 X 2; on the primary diagonal we will find the number of instances correctly classified: True Positives (a prerequisite relation does exist and it is found) and True Negatives (a prerequisite relation does not exist and it is not found). On the secondary diagonal the classifier errors are reported: False Positives (a prerequisite relation does not exist while it is found) and False Negatives (a prerequisite relation does exist and it is not found). The general confusion matrix is shown in table 2.

Table 2: Confusion Matrix.

1	2	class
True Positives	False Positives	1
False Negatives	True Negatives	2

With the course of Advanced Computer Science, composed of 85 LO , The System has responded with good values of precision and recall as is shown in Table 3.

Table 3: Results for the course Advanced Computer Science.

instances n.	Correct	Wrong
1037	877	160

	Cl 1	Cl 2
Precision	0,839	0,853
Recall	0,858	0,833
F1	0,848	0,843
ROC	0,894	0,894

1	2	class
448	74	**1**
86	429	**2**

The results for the course Programming Language (Java) reported in Tab. 4 are lower than the course of Advanced Computer Science because the second course has more LOs than the first one, even if both belong to a branch of science that is well categorized in Wikipedia.

Table 4: Results for the course Programming Language (Java).

instances n.	Correct	Wrong
289	256	33

	Cl 1	Cl 2
Precision	0,633	0,915
Recall	0,463	0,956
F1	0,535	0,935
ROC	0,743	0,743

1	2	class
19	22	**1**
11	237	**2**

The Futurism course's results are illustrated in Table 5. The low number of training instances does not allow the system to reach acceptable levels of precision and recall on Class1.

Table 5: Results for the course Futurism.

instances n.	Correct	Wrong
25	19	6

	Cl 1	Cl 2
Precision	0,333	0,818
Recall	0,2	0,9
F1	0,25	0,857
ROC	0,53	0,53

1	2	class
1	4	**1**
2	18	**2**

The Economy course was created by taking the LOs text directly from the Wikipedia pages in order to see how the system works in an ideal situation. The performance reported in the Table 6 shows good results despite the few LOs in the course.

Table 6: Results for the course Economy.

instances n.	Correct	Wrong
25	21	4

	Cl 1	Cl 2
Precision	0,75	0,923
Recall	0,9	0,8
F1	0,818	0,857
ROC	0,87	0,87

1	2	class
9	1	**1**
3	18	**2**

In the Basic Mathematics course, results are reported in Table 7. All the zero values are explained by the fact that the system is unable to find a sufficient number of relationships between pairs of LOs. This is due to the mathematical formalism expressed with the use of formulas in the text: Wikipedia Miner service fails to find Wikipedia pages associated to the mathematical formulas and doesn't associate topics to concepts.

To see how the system works in a real situation, with LO taken from courses of different types, a dataset consisting of all instances of Advanced Computer Science, Programming Language (Java) and Futurism was created. We left out the mathematics course owing to the low results obtained, as shown above. The tests are shown with and without the ideal course of Economy so we can see the peak present at low numbers of LO's components in the course.

Table 7: Results for the course Basic Mathematics.

instances n.	Correct	Wrong
5	4	1

	Cl 1	Cl 2
Precision	0	0,8
Recall	0	1
F1	0	0,889
ROC	0,5	0,5

1	2	class
0	1	**1**
0	4	**2**

The test performed on the data set composed of 1-2-3 courses highlights how the training stabilizes the system levels of precision and recall around the 0.848 using a big training set.

Table 8: Results for the course course 1-2-3.

instances n.	Correct	Wrong
5498	4660	838

	Cl 1	Cl 2
Precision	0,835	0,859
Recall	0,845	0,85
F1	0,84	0,855
ROC	0,895	0,895

1	2	class
2199	403	**1**
435	2461	**2**

Here is the trend of results sorted by number of instances in the dataset of the course.

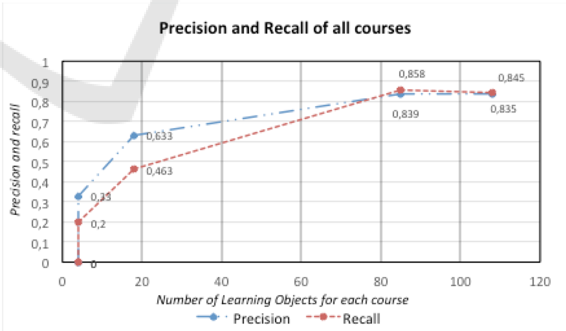

Figure 1: On the x-axis the courses are ordered by LO numerousness, on the y-axis the value of the precision and recall.

As expected, the results show an improvement in system performance with the growth of the size of the data set used as a training set. Plus, an analysis of decision trees generated by the algorithm J48 for each iteration shows that initially the features with a higher content of information are the two measures associated with the amount of nouns in the wikipedia page associated to the LO. Just below these measures, in the decision tree, two other measures are close to the root: the average length of the initial parts of wikipedia pages related to the LO, and the number of the internal links. Including in the graph the results of the Economy course we have a peak in the graph at a low level of instances as shown in Fig. 2.

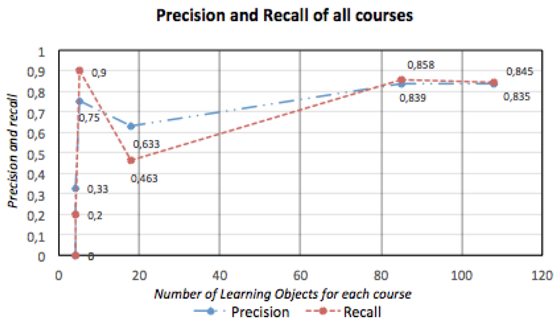

Figure 2: On the x-axis the courses are ordered by LO numerousness, on the y-axis the value of the precision and recall.

5 CONCLUSIONS

The composition of web based instructional courses, especially if personalization and adaptivity are supported (Sterbini and Temperini, 2009; De Marsico et al., 2013) can result in a burdensome task for the teacher, encompassing both the selection of suitable learning objects, and the control on their sequencing. In this paper we have presented an approach for supporting the teacher in the management of the relationships of dependencies between learning objects: such relationships can be suggested/discovered automatically, so as to allow the teacher to adopt or change them. An effective automated determination of such relationships can also be very useful in contexts of personalized e-learning, where the learner is proposed learning objects that are automatically sequenced.

Experimental results presented in this article have confirmed the suitability of an approach based on the data, namely a machine learning approach that provides precious indications that strengthen our working hypothesis. Obviously, since this approach is *data driven*, the provided information may be domain-dependent.

In order to produce results as *independent domain* as possible, in future we will consider resorting to different machine learning approaches (neural networks, Bayesian networks, etc.) and to substantiate the validity of our work hypotheses, on a theoretical level too.

REFERENCES

Allen, M. and Tay, E. (2012). Wikis as individual student learning tools: The limitations of technology. *Int. J. Inf. Commun. Technol. Educ.*, 8(2):61–71.

Biancalana, C., Gasparetti, F., Micarelli, A., and Sansonetti, G. (2013). Social semantic query expansion. *ACM Trans. Intell. Syst. Technol.*, 4(4):60:1–60:43.

Brusilovsky, P. and Vassileva, J. (2003). Course sequencing techniques for large-scale web-based education. *International Journal of Continuing Engineering Education and Life-long Learning*, 13:75–94.

Capuano, N., Gaeta, M., Orciuoli, F., and Ritrovato, P. (2009). On-demand construction of personalized learning experiences using semantic web and web 2.0 techniques. In *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on*, pages 484–488.

Cebeci, Z. and Tekdal, M. (2006). Using podcasts as audio learning objects. *Interdisciplinary Journal of E-Learning and Learning Objects*, 2(1):47–57.

Cole, M. (2009). Using wiki technology to support student engagement: Lessons from the trenches. *Computers & Education*, 52(1):141 – 146.

De Marsico, M., Sterbini, A., and Temperini, M. (2013). A framework to support social-collaborative personalized e-learning. In *Kurosu, M. (Ed.) Human-Computer Interaction (Part II), 15th Int. Conf. on Human-Computer Interaction, LNCS 8005*, pages 351–360. Springer.

Gabrilovich, E. and Markovitch, S. (2009). Wikipedia-based semantic interpretation for natural language processing. *J. Artif. Int. Res.*, 34(1):443–498.

Gasparetti, F., Limongelli, C., and Sciarrone, F. (2015a). A content-based approach for supporting teachers in discovering dependency relationships between instructional units in distance learning environments. In *Human-Computer Interaction. Theories, Methods, and Tools - 17th International Conference, HCI Los Angeles, CA, U.S.A., August 2-7, 2015*. To appear.

Gasparetti, F., Limongelli, C., and Sciarrone, F. (2015b). Exploiting wikipedia for discovering prerequisite relationships among learning objects. In *Proc. IEEE Int. Conf. on Information Technology Based Higher Education and Training, ITHET2015*.

Gasparetti, F., Limongelli, C., and Sciarrone, F. (2015c). Wiki course builder: a system for retrieving and sequencing didactic materials from wikipedia. In *Proc. IEEE Int. Conf. on Information Technology Based Higher Education and Training, ITHET2015*.

Gentili, G., Marinilli, M., Micarelli, A., and Sciarrone, F. (2001). Text categorization in an intelligent agent for filtering information on the web. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(3):527–549.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Limongelli, C., Lombardi, M., Marani, A., Sciarrone, F., and Temperini, M. (2016). A recommendation module to help teachers build courses through the moodle learning management system. *New Review of Hypermedia and Multimedia*, 22(1-2).

Limongelli, C., Miola, A., Sciarrone, F., and Temperini, M. (2012). Supporting teachers to retrieve and select learning objects for personalized courses in the moodle ls environment. In *Proc. 14th IEEE Int. Conf. on Advanced Learning Technologies, Workshop SPEL*, pages 518–520. IEEE Computer Society.

Limongelli, C., Sciarrone, F., Starace, P., and Temperini, M. (2010). An ontology-driven olap system to help teachers in the analysis of web learning object repositories. *Information Systems Management*, 27(3):198–206.

Limongelli, C., Sciarrone, F., and Temperini, M. (2015). A social network-based teacher model to support course construction. *Computers in Human Behavior*, 51:1077–1085.

Milne, D. and Witten, I. H. (2008). An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In *Proceeding of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy*, pages 25–30. AAAI Press.

Milne, D. and Witten, I. H. (2013). An open-source toolkit for mining wikipedia. *Artif. Intell.*, 194:222–239.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition.

Parker, K. and Chao, J. (2007). Wiki as a teaching tool. *Interdisciplinary Journal of E-Learning and Learning Objects*, 3(1):57–72.

Resnik, P. (2011). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *CoRR*, abs/1105.5444.

Revilla Muñoz, O., Alpiste Penalba, F., and Fernández Sánchez, J. (2015). The skills, competences, and attitude toward information and communications technology recommender system: an online support program for teachers with personalized recommendations. *New Review of Hypermedia and Multimedia*. Published online - Article in Press.

Scheines, R., Silver, E., and Goldin, I. (2014). Discovering prerequisite relationships among knowledge components. In Stamper, J., Pardos, Z., Mavrikis, M., and McLaren, B., editors, *Proceedings of the 7th International Conference on Educational Data Mining*, pages 355–356. European Language Resources Association (ELRA).

Sciarrone, F. (2013). An extension of the q diversity metric for information processing in multiple classifier systems: A field evaluation. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(6).

Sterbini, A. and Temperini, M. (2009). Collaborative projects and self evaluation within a social reputation-based exercise-sharing system. In *Workshop SPEL, for the IEEE/WIC/ACM Int. Joint Conferences on Web Intelligence and Intelligent Agent Technologies, WI-IAT'09, Vol. 3*, pages 243–246.

Strube, M. and Ponzetto, S. P. (2006). Wikirelate! computing semantic relatedness using wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, pages 1419–1424. AAAI Press.

Stuurman, S., van Eekelen, M., and Heeren, B. (2012). A new method for sustainable development of open educational resources. In *Proceedings of Second Computer Science Education Research Conference*, CSERC '12, pages 57–66, New York, NY, USA. ACM.

Sun, Z. and Qiu, X. (2014). Evaluating the use of wikis for efl: A case study of an undergraduate english writing course in china. *Int. J. Inf. Technol. Manage.*, 13(1):3–14.

Vuong, A., Nixon, T., and Towle, B. (2011). A method for finding prerequisites within a curriculum. In Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero, C., and J. Stamper, J., editors, *The 4th International Conference on Educational Data Mining (EDM 2011)*, pages 211–216.