

Towards Automatic Competence Assignment of Learning Objects

Ricardo Kawase¹, Patrick Siehndel¹, Bernardo Pereira Nunes^{2,1},
Marco Fisichella¹, and Wolfgang Nejdl¹

¹ L3S Research Center, Leibniz University Hannover, Germany
{kawase,siehndel,nunes,fisichella,nejdl}@L3S.de

² Department of Informatics - PUC-Rio - Rio de Janeiro, RJ - Brazil
bnunes@inf.puc-rio.br

Abstract. Competence-annotations assist learners to retrieve and better understand the level of skills required to comprehend learning objects. However, the process of annotating learning objects with competence levels is a very time consuming task; ideally, this task should be performed by experts on the subjects of the educational resources. Due to this, most educational resources available online do not enclose competence information. In this paper, we present a method to tackle the problem of automatically assigning an educational resource with competence topics. To solve this problem, we exploit information extracted from external repositories available on the Web, which lead us to a domain independent approach. Results show that automatically assigned competences are coherent and may be applied to automatically enhance learning objects metadata.

Keywords: Metadata Generation, Competences, e-Learning, Automatic Competence Classification.

1 Introduction

Understandability of resources by learners is one essential feature in the learning process. To measure it, a common practice is the use of competence metadata. A competence is the effective performance in a domain at different levels of proficiency. Educational institutions apply competences to understand whether a person has a particular level of ability or skill. Thus, an educational resource, enriched with competence information, allows learners to identify, on a fine-grained level, which resources to study with the aim to reach a specific competence target.

With the catch up of the Open Archives Initiative, plenty of learning materials are freely available. Through the utilization of the OAI-PMH protocol¹, a learning environment can list the contents of several external repositories. Although this open content strategy provides numerous benefits for the community, new challenges arise to deal with the overload of information. For example, every time a new repository is added to a library, thousands of new documents may come at once. This makes the experts' task of evaluating and assigning competences to the learning objects impossible.

¹ <http://www.openarchives.org/pmh>

Table 1. The competence classification of the OpenScout repository and the respective examples of most relevant keywords

Competences	Relevant Keywords
Business and Law	law,legal,antitrust,regulation,contract,formation,litigation. . .
Decision Sciences	decision,risk,forecasting,operation,modeling,optimization. . .
General Management	planining,plan,milestone,task,priority,management,evaluation. . .
Finance	finance,financial,banking,funds,capital,cash,flow,value,equity,debt. . .
Project Management	management,monitoring,report,planning,organizing,securing. . .
Accounting and Controlling	accounting,controlling,balance,budgets,bookkeeping,budgeting...
Economics	economics,economy,microeconomics,exchange,interest,rate,inflation. . .
Marketing and Sales	marketing,advertising,advertisement,branding,b2b,communication. . .
Organizational Behavior and Leadership	organizational,behavior,leadership,negotiation,team,culture. . .
Management Information Systems	management,information,system,IT,data,computer,computation...
Human Resource Management	resources,management,career,competence,employee,training,relation. . .
Entrepreneurship	entrepreneurship,entrepreneurs,start-up,opportunity,business. . .
Technology and Operations Management	technology,operation,ebusiness,egovernment,ecommerce,outsourcing. . .
Strategy and Corporate Social Responsibility	strategy,responsibility,society,sustainability,innovation,ethics,regulation. . .
Others	-

In this paper, we present our work towards an automatic competence assignment tool, taking into account the speed of educational resources development, exchange, and the problem of ensuring that these materials are easily found and understandable. Our goal is to provide a mechanism that facilitates learners in finding relevant learning materials and to enable them to better judge the required skills to understand the given material through the interpretation of competence levels.

2 Competences

Our work is contextualized within the OpenScout learning environment². The OpenScout portal is the outcome of an EU co-funded project³, which aims at providing skill-and-competence based search and retrieval Web services that enable users to easily access, use, and exchange open content for management education and training. Therefore, the project not only connects leading European Open Educational Resources (OER) repositories, but also integrates its search services into existing learning suites [2]. As the platform integrates different content repositories, many learning materials are daily added to the environment without the experts' annotations regarding competence levels. To tackle this problem we proposed a novel approach to automatically annotate the educational resources in OpenScout with competences.

Within the project, a management-related competence classification was developed (see Table 1), in order to support learners and teachers while searching for appropriated educational resources that meet a specific competence level. In a first major step, a

² <http://learn.openscout.net>

³ <http://openscout.net>

focus group was organized consisting of a sample of ten domain experts from Higher Education, Business Schools, and SMEs, including two professors, six researchers and two professionals with the aim to generate an initial competence classification from experience and academic literature.

In addition to the competence classification, within the OpenScout project we created a list of keywords that are mostly relevant to each competence (see Table 1). These descriptions are essential for our automatic competence assignment tool, further explained in Section 3.

In order to build the competence descriptions, eight researchers from the ESCP Europe Business School⁴ with different research focuses and knowledge about certain domains were asked to provide a list of terms that best fit their domains (competences). Participants had completed different diploma studies in Germany, the US, UK, Australia, or China and had an average of two years of work experience at the university; three of them had also been employed full-time in several industries before. All experts have emphasized that they provided a subjective assessment creating the keyword list related to each competence. Thus, due to their long years of experience and ongoing education in their respective field of knowledge, these experts fulfilled the necessary criteria for providing the most relevant keywords.

3 Automatically Assigning Competences

In order to solve the problem of automatically assigning competence annotations to learning objects, we developed an unsupervised method that can be applied to any repository of documents where the competences involved are known in advance. The method is a tag-based competence assigner. To better understand the proposed method, in the next subsection we briefly introduced the methodology involved to extract tags from learning objects, followed by the actual competence assigning method.

3.1 α -TaggingLDA

Our proposed competence annotation method is an extension of the α -TaggingLDA. This method is a state-of-the-art LDA based approach for automatic tagging introduced by Diaz-Aviles et al. [1]. α -TaggingLDA is designed to overcome new item cold-start problems by exploiting content of resources, without relying on collaborative interactions. The details involving the technical aspects of the automatic tagger are out of the scope of this paper and we refer to [1] for more details. The important abstraction to be considered is that, for a given LO, α -TaggingLDA outputs a ranked list of most representative tags.

3.2 Tag-Based Competences

On top of the automatic tagging method presented in Section 3.1, we added a new layer to identify which is the most probable competence a document includes. The

⁴ <http://www.escpeurope.eu>

classification layer uses two different inputs; (i) a ranked list of keywords that describes the resource to be classified (tags) and (ii) a list of competences that a document can belong to with a list of keywords describing each competence (see Table 1).

With these two inputs, the classification method assigns scores for each match found between the document's list of keywords and the competences' keywords. Since the document's tags are already properly ranked, we apply a linear decay on the matching-score. It means that the competences' keywords that matches the first document's keywords have a greater score. In the other hand, the higher a document's keywords is positioned in the ranking, the lower is the final score. After the matching process, we compute the sum of the scores for each competence and the document is assigned with the top scoring competence. The pseudo-code (Algorithm 1) depicts the matching method. It is important to remark that all keywords involved are first submitted to a stemming process.

Algorithm 1: Pseudocode for keyword-term matching method.

```

1 begin
2   for each document do
3     Get top N  $\alpha$ -TaggingLDA keywords;
4     for each keywords do
5       KeywordIndex++; for each competence do
6         Get competence's terms;
7         for each competence's terms do
8           if keyword == terms then
9             competence-score += 1/KeywordIndex;
10  return scoring competences;

```

3.3 Evaluation

To evaluate our method we used the OpenScout dataset containing 21,768 learning objects. We pruned these data to consider only objects that are in English, with the description with a minimum length of 500 characters, which resulted in a set of 1,388 documents. Thus, on these documents we applied the competence assignment method. Since the dataset is relatively new and very few items have been assigned with competences, we propose an automatic method to evaluate the outcomes of the automatic competence assignments.

Our evaluation method considers the similarity among the learning objects and a set of assumptions/cases that we believe can validate whether the automatic competence assigner produces optimum results or not. To measure the similarity among the documents, in our study, we used MoreLikeThis, a standard function provided by the Lucene search engine library⁵. MoreLikeThis calculates similarity of two documents by computing the number of overlapping words and giving them different weights based on

⁵ http://lucene.apache.org/core/old-versioned_docs/versions/3.4.0/api/all/org/apache/lucene/search/similar/MoreLikeThis.html

TF-IDF [3]. MoreLikeThis runs over the fields we specified as relevant for the comparison - in our case the description of the learning object - and generates a term vector for each analyzed item (excluding stop-words).

To measure the similarity between documents, the method only considered words that are longer than 2 characters and that appear at least 2 times in the source document. Also, words that occur in less than 2 different documents are not taken into account for the calculation. For calculating the relevant documents, the method used the 15 most representative words, based on their TF-IDF values, and generated a query with these words. The ranking of the resulting documents is based on Lucene's scoring function which is based on the Boolean model of Information Retrieval and the Vector Space Model of Information Retrieval [4].

Let $c(LO_i)$ be a function returning the competence for a specific learning object LO_i and let $s(LO_i, LO_j)$ be a function measuring the similarity between two resources LO_i and LO_j . Then, given the set of learning objects, the similarity scores $s(LO_i, LO_j)$ and the competence assignments $c(LO_i)$, we evaluate the results through four given cases:

- **Case 1:** If two LOs have the same competence and are similar to some extent, it is reasonable to assume that the competence assigner is coherent. If $c(LO_1) = c(LO_2)$ and $s(LO_1, LO_2) \geq 0.7$
- **Case 2:** If two LOs have been assigned with the same competence but are not similar, it is not completely implausible and means that the competence is broad. If $c(LO_1) = c(LO_2)$ and $s(LO_1, LO_2) < 0.7$
- **Case 3:** If two LOs have been assigned with different competences and are very similar, it suggests that the automatic competence assigner committed a fault. Thus, the lower the assignments that fall in this case the better the results. If $c(LO_1) \neq c(LO_2)$ and $s(LO_1, LO_2) \geq 0.7$
- **Case 4:** Finally, for the cases where two LOs have been assigned with different competences and the LOs are not similar, correctness can not be derived but a high value also demonstrates the coherence of the method. If $c(LO_1) \neq c(LO_2)$ and $s(LO_1, LO_2) < 0.7$

3.4 Evaluation Results

In this section, we present the results of the proposed automatic competence assigner method. In Table 2, we plot the results discerning the number of occurrences (in percentage) that fall in each case. Additionally, we alternate the number of competences considered in the evaluation. First, we used only the top scoring competence for a given LO and, in a second round of the evaluation, we considered the top two scoring competences.

The results show that very few items fell in the cases 1 and 3, meaning that most of the items did not meet the minimum threshold value of similarity, thus, showing that most of the classified documents are dissimilar. The low similarities are also caused by the short textual descriptions available. Regarding the documents that are similar (≥ 0.7), only around 1% of the items fell into case 3; given our assumptions in Section 3.3, we consider this 1% as a false assignment.

Table 2. Results of the automatic competence assignments according to the cases defined in Section 3.3, considering the top one and top two competences with similarity threshold at 0.7

Rule	Tags(1)	Tags(2)
1) Same Competence Sim. ≥ 0.7	0.24	0.24
2) Same Competence Sim. < 0.7	9.60	10.63
3) Dif. Competence Sim. ≥ 0.7	1.02	0.95
4) Dif. Competence Sim. < 0.7	89.12	88.16

4 Conclusion

In this work, we proposed a methodology to automatically assign competences to learning objects. Our proposed method is based on an automatic tagging tool that does not require a training set or any previous users' interaction over the resources. We also proposed an automatic methodology to evaluate the given competences through a set of cases that considers objects' textual similarities. Although the cases cannot guarantee the correctness of an assignment, the third case indeed exposes misassigned items. The results obtained showed very few occurrences where different competences were assigned to very similar items. We interpret that as evidence of the coherence and effectiveness of the proposed method that may be applied to effectively enhance competence metadata for learning objects.

As future work, we plan to improve the competence classifier by including the whole content of documents and representing it through a weighted term vector. Additionally we plan to automatically quantify necessary competence levels according to the European Qualification Framework (EQF).

Acknowledgement. This research has been co-funded by the European Commission within the eContentplus targeted project OpenScout, grant ECP 2008 EDU 428016 (cf. <http://www.openscout.net>) and by CAPES (Process n° 9404-11-2).

References

1. Diaz-Aviles, E., Georgescu, M., Stewart, A., Nejdl, W.: Lda for on-the-fly auto tagging. In: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys 2010, pp. 309–312. ACM, New York (2010)
2. Niemann, K., Schwertel, U., Kalz, M., Mikroyannidis, A., Fisichella, M., Friedrich, M., Dicerto, M., Ha, K.-H., Holtkamp, P., Kawase, R., Parodi, E., Pawlowski, J., Pirkkalainen, H., Pitsilis, V., Vidalis, A., Wolpers, M., Zimmermann, V.: Skill-Based Scouting of Open Management Content. In: Wolpers, M., Kirschner, P.A., Scheffel, M., Lindstaedt, S., Dimitrova, V. (eds.) EC-TEL 2010. LNCS, vol. 6383, pp. 632–637. Springer, Heidelberg (2010)
3. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1983)
4. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM 18(11), 613–620 (1975)