Mastery Profiling through Entity Linking

to Support Project Team Formation in Higher Education

Olieman, A.M.; Nack, F.

DOI
10.5220/0004764001820190

Publication date
2014

Document Version
Final published version

Published in
CSEDU 2014

Mastery Profiling Through Entity Linking
To Support Project Team Formation in Higher Education

Alexander M. Olieman and Frank Nack
Informatics Institute, University of Amsterdam, Science Park 904, Amsterdam, Netherlands
{olieman, nack}@uva.nl

Keywords: Collaborative Learning, Computer-Supported Group Formation, Information Extraction, Entity Linking, Mastery Profiling, Graph Exploration, DBpedia Spotlight, Gremlin.

Abstract: Computer-supported group formation enables educators to assign students to project teams. The focus in this paper is placed on gathering data about student attributes that are relevant in the context of specific course projects. We developed a method that automatically produces learner models from existing documents, by linking students to topics and estimating the levels of skill, knowledge, and interest that students have in these topics. The method is evaluated in an experiment with student participants, wherein its performance is measured on two levels. Our results demonstrate that it is possible to link students to topics with high precision, but suggest that estimating mastery levels is a more challenging task.

1 INTRODUCTION

Group-based learning has taken an important role in curricula across the educational spectrum. One aspect of group-based learning that has attracted considerable attention from researchers, is the formation of groups of learners. Group formation influences the interactions that group members have, and thereby affects the results of the learning experience (Kyprianidou et al., 2011). Poorly formed groups may suffer from, for example, an unproductive use of time or incompatible personality types. The way students are partitioned into groups also raises the question if they can be assessed fairly, e.g. due to an unbalanced spread of skills (Livingstone and Lynch 2000).

In higher education, common group forming methods are student self-selection and random assignment. These methods do not necessarily lead to good learning experiences, but are often the only practical alternatives for instructors who teach large numbers of new students each year. Instructors might lack the necessary information about the students to implement a more elaborate group formation process, or face the impracticality of manually solving a large combinatorial problem (Craig et al., 2010). This has motivated research towards the development of tools that can aid instructors in forming groups, which is known as Computer-Supported Group Formation (CSGF) (Ounnas et al. 2009).

Regardless of the algorithms that are used, the criteria that can be used in the group formation process are limited by the data that is used to describe the students. Hence much of the previous work makes use of student attributes for which standardized tests are available, such as team roles, personality types, and learning styles (Magnisalis et al., 2011). Important disadvantages of gathering student data in this way are the dependence on lengthy questionnaires, and the need to ask students new questions when course-specific characteristics are taken into account.

In computer-supported collaborative learning settings where the majority of learning occurs in a virtual environment, there are opportunities to gather relevant data about students continuously. In more traditional settings, it may instead be viable to use data from existing resources that describe students, specifically to model students' mastery of topics. Previous suggestions are to use text mining techniques on curricula vitae (CVs), academic transcripts, and personal websites (Ounnas et al., 2009).

The objective of this study is to develop and evaluate a method which allows existing data sources that describe students' mastery levels (e.g. of knowledge, skills, and interests) to be automatically combined into a learner profile for use in CSGF algorithms. The scope of group formation problems that need to be addressed by this method is restricted to the domain of team project-based higher education. To work towards this objective, the main research question that this paper addresses is formulated as follows:

Can existing data sources that describe students' mastery of topics be fused into a learner profile, to facilitate computer-supported group formation?

The remainder of this paper is structured as follows. In Section 2 we discuss related work. Section 3 serves to briefly describe a further exploration of the problem domain. In Section 4 we discuss gathering existing resources from student participants. Based on these resources we produce learner profiles, using a method that is described in Section 5. In Section 6 we discuss our evaluation approach and results, followed by the conclusion and future work.

2 RELATED WORK

2.1 Computer-supported Group Formation

CSGF is based on the idea that instructors can assign students to groups by making explicit educational criteria according to which groups should be formed (Craig et al., 2010). The essence of CSGF is: the synthesis of groups by applying criteria that optimize aspects of each group, by making use of data about the individual learners (Magnisalis et al., 2011).

A classification, originating from literature on team diversity, divides relevant attributes into *task-related* (e.g. knowledge, skills, experience) and *relations oriented* (e.g. gender, culture, attitude, social ties) (Jackson et al., 1995). Task-related attributes of individual students are relevant because they indicate which cognitive resources will be available in any possible grouping. Relations oriented attributes indicate how group members are expected to interact. A common approach to recording task-related attributes is to ask students for their grades in selected prerequisite courses (Lingard and Berry, 2002). Another approach is to measure skill levels for a few domain-specific skills by questionnaire (Winter, 2004).

Most of the criteria according to which groups should be formed can be classified as *homogeneous*, *heterogeneous*, or *apportioned* (Craig et al., 2010). Both homogeneous and heterogeneous criteria are concerned with the distance between students within a group for a specific attribute, while apportioned criteria serve to distribute a specific attribute as evenly as possible across the groups.

Hoogendoorn (2013) has recently conducted three field experiments which provide evidence for the effect of heterogeneity on the performance of student teams. His results suggest that gender diverse teams perform significantly better than male-dominated teams, and no worse than female-dominated terms. The effect of ethnic diversity on team performance is found to be positive for teams where at least half of the members have different backgrounds. Diversity in cognitive ability of team members only shows a positive effect on performance when the degree of heterogeneity is moderate. Heterogeneity of cognitive resources is suggested as an underlying mechanism for the effect of diversity in ethnicity and cognitive ability.

Other researchers argue for certain criteria without empirical support (e.g. based on expert opinion). Most arguments are made for heterogeneous criteria (i.e. complementary fit) on specific attributes, including skills, knowledge, abilities (Wells, 2002; Werbel and Johnson 2001; Wilkinson and Fung 2002), and learning styles (Magnisalis et al., 2011). Student interests and values should however be grouped homogeneously (Werbel and Johnson, 2001). Grades in prerequisite courses are most often apportioned (Craig et al., 2010; Ounnas, 2010).

2.2 Entity Linking

Entity linking (EL) is the information extraction task of automatically *"matching a textual entity mention [...] to a [knowledge base] entry, such as a Wikipedia page that is a canonical entry for that entity."* (Rao et al., 2013, p. 96). Three key challenges have been identified for EL to deal with: name variation, entity ambiguity, and absence (Dai et al., 2012; Rao et al., 2013). Name variation entails that an entity can be referred to by multiple different terms. Entity ambiguity refers to the issue that a single name string can match with several distinct entities. The issue with absence is that when no knowledge base (KB) entry exists for the entity that is mentioned in the text, no entity should be returned, rather than the highest-ranking KB entry.

There are, however, two relevant limitations present in the existing work on EL. Most research focuses explicitly on linking *named* entities (i.e. entities referred to by proper names), specifically on persons, locations, and organizations (Mendes, Daiber, et al., 2011; Rao et al., 2013). Additionally, many current approaches are evaluated only on English-language texts, with a focus on the news domain

(Mendes et al. 2011; Rao et al., 2013).

DBpedia Spotlight is an open-source system that can annotate any given input text with DBpedia resources (i.e. KB entries), which are based on semantic extraction from Wikipedia articles (Mendes, Jakob, et al., 2011). Several parameters provide the means to filter annotations according to task-specific requirements. By default, DBpedia Spotlight is not specialized towards specific entity types, but it may be configured to annotate only instances of specific types, either by selection of classes, or by arbitrary SPARQL[1] queries (Mendes et al., 2011).

When linking targets are known to have a specific type, e.g. genes (in the biomedical domain), annotating only those entities is quite straightforward (Dai et al., 2012). If a domain-specific vocabulary already contains links to DBpedia or Wikipedia then one can consider all DBpedia Spotlight candidate entities, and then check whether the top-ranked candidate has a corresponding entity in the local vocabulary (Mendes et al., 2011).

Multiple EL researchers have found it helpful to include a measure of *semantic-relatedness* between entities in the disambiguation process (e.g. Han et al., 2011). The intuition behind collective entity disambiguation is that the links between, e.g., Wikipedia articles reflect how closely the corresponding entities are related, and that texts are more likely to mention several related entities than entirely unrelated entities.

Besides using metrics of semantic-relatedness and disambiguation purposes, it might be feasible to use them to find additional topics in which a student has some mastery. For instance, when a student's CV mentions that she is skilled in technical drawing, we can infer that she has some skill in drawing in general.

3 VIEWS ON FORMING TEAMS

In the group formation literature arguments are made for the relevance of skills, knowledge, abilities, interest, and grades. All arguments are, however, made from the educator's perspective, and information about the student's perspective is lacking. We have therefore surveyed a group of university students and asked them about the considerations they have had while forming project teams in the past.

As in other group formation studies, we recruited participants from a subpopulation of students who study the same subject (Lingard and Berry, 2002; Winter, 2004). All participants were enrolled in the MSc program *Design for Interaction* at the Delft University of Technology (DUT), and were recruited through a mailing list. The same sample of students were respondents to the questionnaire as well as participants in the experiment that is described in subsequent sections.

The questionnaire was taken by 11 students. We focused our questions on seven attributes: *competence, education, experience, general ability, interest, knowledge, and skill*.

The results of the questionnaire broadly correspond to what was found in the CSGF literature. On this basis, we decide to include skill and knowledge in our learner profiles. We also include interest in our profiles because it is used in team formation criteria by many instructors (Werbel and Johnson 2001; Kyprianidou et al., 2011). Competence, we argue, is not a suitable choice because it depends on specific skills and knowledge.

4 GATHERING EXISTING RESOURCES

After finishing the initial questionnaire, students were asked to participate in creating a learner profile based on existing documents about them. To participate, they needed to provide access to a project portfolio, an academic transcript, and temporary access to their LinkedIn[2] profile, and/or provide the URL to a personal website. The terms that indicate relevant attributes need to be recognized in these resources, and should be linked to a shared vocabulary in which the learner profiles can be expressed. The quality of the resulting profiles is evaluated by comparing them with a ground truth that is given by the participants.

The decision to gather academic transcripts, CVs, and personal websites was motivated by suggestions found in literature (Ounnas et al., 2009). Although documents in a project portfolio do not describe students in the same sense as the other document types do, they can give a more detailed view of the specific topics a student has engaged with during previous projects (at least for a human reader).

Participants' academic transcripts and LinkedIn profiles were saved by a sign-up application. The course descriptions that correspond to the course identifiers in the academic transcripts were retrieved

[1] SPARQL 1.1 - <http://www.w3.org/TR/sparql11-overview/>

[2] LinkedIn - <http://www.linkedin.com/>

through the DUT API[3]. The project portfolios consist of deliverables, such as reports, presentation slides, and project blogs. The corresponding files were provided by Shareworks Solutions BV. For the two participants who provided a website, we saved the pages manually.

In total, 10 LinkedIn profiles (all English), 190 course descriptions (66 Dutch, 124 English), 54 portfolio documents (2 Dutch, 52 English), and two websites (both English) were gathered. Participants were associated with between 15 and 57 documents; on average with 37 documents. Most course descriptions and some portfolio documents were associated with more than one participant.

5 PRODUCING LEARNER PROFILES

Two existing implementations of DBpedia Spotlight are used to produce *annotations* that we define as "links between a phrase in a document and a topic, which is represented by a DBpedia URI". Subsequently, we estimate skill, knowledge, and interest levels by taking into account surrounding terms of each annotation, document origins, and annotation frequencies. The learner profiles that are produced by this method consist of *statements*, where a statement is "the relation between a student and a single topic, which is quantified by three mastery levels. Finally, the set of statements in each profile is expanded by inference over probabilistic and semantic relations between topics.

5.1 Vocabulary Selection and Modification

CSGF differs from the current applications of entity linking: when links are generated for the readers of an article, it is assumed that the readers are familiar with the majority of abstract concepts that are mentioned. For CSGF, abstract concepts are mostly relevant, and people and places less so. To annotate documents only with topics that are relevant for CSGF, we test the approach taken in Mendes, Daiber, et al. (2011) and in Wetz et al. (2012), for which we require an application-relevant subset of DBpedia entities.

We use the LinkedIn "Skills & Expertise" vocabulary[4] V_{LI} as a basis. Reasons to choose this vocabulary are that it is already partially linked to Wikipedia, and that it is used daily by thousands of people to describe their professional abilities. From V_{LI}, we define our vocabulary $V = V_{\mathrm{LI}} \cap DBpedia$. There are 26 292 topics t in V; nearly 70% of $t \in V_{\mathrm{LI}}$, but only 0.7% of all DBpedia resources. The links between V_{LI} and Wikipedia contain inaccuracies. We have manually corrected 40 of such links, but we estimate that at least 10% of $t \in V_{\mathrm{LI}}$ are needlessly missing a link, or are linked to an incorrect Wikipedia article.

V_{LI} only links topics with English Wikipedia articles, and as such V would only include English identifiers for topics. Since our profiling method also needs to deal with documents in other languages, we incorporate alternative topic identifiers into V by using Wikipedia's *interlanguage links*[5]. For each $t \in V$ the Dutch identifier nl (if available) is retrieved from the nl.dbpedia.org SPARQL endpoint through the query:

```
SELECT ?nl WHERE {?nl owl:sameAs
<http://dbpedia.org/resource/t> .}
```

There are topics that are mentioned frequently in all types of the gathered resources, but that are not relevant for learner profiles. We exclude 56 of such topics in total from V.

5.2 Information Extraction Pipeline

The information extraction process that we employ lends itself to being described as a data transformation pipeline.

For each gathered document $d \in Documents$, a DBpedia Spotlight implementation, given a configuration, annotates the content per section. The resources that are returned are filtered with our vocabulary.

In each section we count *qualifying terms*, which indicate specific types of mastery, and linearly combine the normalized counts with $(Skill, Knowledge, Interest)$ weights that depend on the document origin $o(d)$. The resulting (s, k, i) score is assigned to each annotation within the section, after which the scores are summed per topic for the entire document. The summed scores are stored in the document (d_{OUT}) as indications of mastery, where one $indication = (t, (s, k, i))$.

Hereafter, we select for each profile $p \in Profiles$ the d_{OUT} that are associated with p by a document link $dl = (p, d)$. The indications in the associated documents are summed per (t, o), resulting in $1..|o|$ mastery levels scores, where one

[3] Delft University of Technology API - <http://apidoc.tudelft.nl/>

[4] LinkedIn Skills & Expertise - <http://www.linkedin.com/skills/>

[5] See: <http://en.wikipedia.org/wiki/Help:Interlanguage_links>

$mls = (s, k, i)$, for each unique topic that is associated with p. The indications that originate from course descriptions are, before summation, weighted by the grade that the student received for the corresponding course.

The maximum s, k, and i scores of the summed indications differ significantly between origins. For each origin, the scores are linearly transformed to the range $[0, 100]$, because we wish to compare the accuracy of statements from different origins in our evaluation. The resulting normalized indications represent the relative mastery levels per topic for a single student, i.e., they encode beliefs that the student has more mastery in one topic than in another.

For CSGF the aim is to compare the mastery levels per topic between students. A final data transformation is thus needed. Each normalized $mls_{t,o}$ is transformed to its percentile rank (PR) $mlpr_{t,o}$in the frequency distribution of the mls with the same (t, o) from all profiles. Finally, statements are saved as $statement_{p,o} = (t, mlpr)$.

5.3 Linking Documents to Topics

The first step in our approach to mastery level profiling is to ask for each student: in which topics does this student have any skill, knowledge, and/or interest? We use entity linking to answer this question based on the gathered documents, in lieu of more tailored information extraction techniques. This enables us to test the hypothesis that:

"From all entities that are mentioned in the documents associated with a student, a vocabulary can be used to select the entities that are topics in which this student has some mastery".

5.3.1 Annotation Method

Two DBpedia Spotlight implementations (Mendes, Jakob, et al. 2011; Daiber et al. 2013) are used and configured to produce annotations in our experiment. The original Information Retrieval-based implementation, with the default configuration, spots all phrases in the input text that also occur in a dataset of possible surface forms for all DBpedia. It selects candidate entities for each spotted phrase, and ranks them according to the prior probability that the observed phrase refers to the selected candidate. The candidates are then re-ranked by querying a Vector Space Model (VSM), in which entities are represented by the paragraphs that mention them in Wikipedia, with the context of the observed. Top-ranking candidates are the most likely disambiguations.

The newer statistical model uses a generative probabilistic model for disambiguation. This model is used to calculate a disambiguation score for entity e, given the spotted phrase s and its context c, by combining $P(e)$, $P(s|e)$, and P. The original phrase spotting method is used in parallel with a Natural Language Processing (NLP) method that is not limited to surface forms that occur in DBpedia. Any overlap in spotted phrases is resolved, after which the phrases that fall below a score threshold α are dropped from the annotation process.

In both implementations the *topical pertinence* of a candidate for the observed context is indicated by the disambiguation score. The relative difference in this score between the first and second ranking candidate indicates *contextual ambiguity*, i.e. how uncertain it is that the top-ranked candidate is the entity that is mentioned in this context. The *confidence* parameter, which is provided at runtime, applies a threshold of $(1 - confidence)$ to candidates' contextual ambiguity scores. A second runtime parameter, *support*, specifies the minimum number of Wikipedia inlinks a candidate resource must have to be further considered.

We define a third runtime parameter which chooses between *single* and *multiple candidate filtering*. In single candidate filtering, we take the set of top-ranking entities E_s that Spotlight produces for a section and select as topics the entities that occur in our vocabulary ($T_s = E_s \cap V$). In multiple candidate filtering we instead initialize $T_s = \{\}$, take the set of ranked candidate vectors $R_s = \{\bar{r}_1, \bar{r}_2, \dots, \}$, and from each vector we add the top-ranking topic to T_s (denoted $T_s \cup \{t_1\}$, where $t_1 = e\colon e_{\min(j)} \in \{e_j\colon e_j \text{ in } \bar{r} \text{ and } e \in V\}$).

5.3.2 Exploration of the Parameter Space

To assess the suitability of various configurations for producing learner models (before the ground truth is given by the participants), we have manually annotated a small test collection of documents, and have measured the performance of our annotation method on this collection.

We use the measures *precision*, *recall*, and *F-score* (F_β) to evaluate the performance of the annotation method. The definitions of these measures are adapted from the prevailing definitions in entity linking (Han et al. 2011) to better suit our annotation task. Let $learner$ be the person who is profiled in p, with the set of associated documents $\{d_{\text{OUT}}\colon dl(p, d)\}$. $Generated_p$ is the set of all topics with which $\{d_{\text{OUT}}\colon dl(p, d)\}$ has been annotated. $Truth_l$ is the set of all topics in which $learner$

claims to have some mastery.

$$precision = \frac{|Generated_p \cap Truth_l|}{|Generated_p|} \quad (1)$$

$$recall = \frac{|Generated_p \cap Truth_l|}{|Truth_l|} \quad (2)$$

$$F_\beta = (1 + \beta^2) \times \frac{precision \times recall}{(\beta^2 \times precision) + recall} \quad (3)$$

In this work $\beta = 0.5$, which reflects our assumption that recall is only half as important as precision for this annotation task.

For the test documents we have had to use $Generated_d$ and have created $Truth_d$. In $Truth_d$ we have only included the topics for which d implied some mastery on the part of any student associated with d.

To find which configurations perform best for our annotation task, we have performed a parameter sweep on 3 Dutch and 8 English test documents. Because we were not able to assign a value to the threshold α, we have instead manipulated a spot score weight which, when enlarged, increases the probability that spotted terms are annotated.

With both languages, a clear tradeoff between precision and recall can be observed. Multiple candidate filtering, as expected, results in higher recall, but lower precision, than single candidate filtering. Increasing the confidence value causes higher precision, but lower recall. We found the effect of the support parameter to be negligible.

Based on these results, we continue our main experiment with the IR-based implementation and $confidence = 0.3$ for English documents, and the statistical implementation with a spot score weight of 0.4 and $confidence = 0.2$ for Dutch documents. For both languages we use single candidate filtering and $support = 0$.

5.4 Estimating Mastery Levels

To estimate which type(s) of mastery a student has in a topic, we have selected the descriptions of the 25 most attended courses, and recorded all terms that imply skill or knowledge (i.e. qualifying terms).

We represent the qualifying terms as sets, and count for every section that contains annotations how many terms indicate skill and knowledge. Stemming is used to also count lexical variations of the terms. Predefined weights per document origin are linearly combined with the fractions of skill and knowledge term counts, to produce a (s, k, i) score per annotation in a section. In the defined weights, we assume that portfolio documents and websites indicate each type of mastery, but that course descriptions do not indicate any interest.

In our definition of mastery levels we need to take into account that the ground truth that we will use in our evaluation is provided by the participants. The scale and unit in which mastery levels are expressed need to be understood by students to allow them to accurately correct their profiles (Bull and Kay 2007). In our model a mastery level means that a student has more knowledge, skill, or interest in a topic than a percentage of his or her peers. *"Paul (Knowledge, 75) Archery"*, for example, would indicate that Paul has more knowledge about archery than 75% of his peers.

To estimate mastery levels, we use the intuition that a student will have more mastery in topics that are mentioned more often in the associated documents. For the topics that originate from course descriptions we also incorporate the grade that a student received and the extent of the course. Each indication of mastery that originates from a course description is multiplied by a weight $g_{p,d}$, which is calculated as:

$$g_{p,d} = \frac{0.5 \times (2 + ECTS\_credits_d)}{11 - grade_{p,d}}. \quad (4)$$

All indications in the associated documents for a single student are subsequently summed per (t, o). Hereafter, the indications are normalized per origin, but across profiles, to the range $[0, 100]$, so that for each (t, o) there exist maximum s, k, and i scores with the value 100. This enables us to generate indications for a fifth origin ALL, in which we have compensated for differences in annotation frequency and weighting between topics and origins. The indications for *ALL* are generated by summing the existing indications per profile.

Finally, indications are transformed into statements with the desired semantics by using the frequency distributions of s, k, and i scores, again per (t, o), over all profiles. Each score is transformed into a mastery level by calculating its percentile rank in the corresponding frequency distribution. But because mastery levels are defined relative to a student's peers, we would need frequency distributions that include all peers.

To compensate for the limited number of participants, we apply a form of additive smoothing in the calculation of percentile ranks. Into each frequency distribution $m_{t,o}$ we insert the values 0.0, $1.5 \times max(m_{t,o})$, and $(|m_{t,o}| - 1)$ evenly spaced values in between. The percentile ranks of individual indications are computed from these modified frequency distributions.

5.5 Expanding Profiles by Inference

Because our vocabulary is not a KB, in the sense that it contains no information about relationships between topics, we use DBpedia as a source for semantic relations between topics. We aim to predict for each student, on the basis of the topics that are linked from their profile, in which other topics they are likely to have some mastery.

We use Gremlin[6], which implements efficient graph traversal as described in Rodriguez & Neubauer (Rodriguez & Neubauer 2011), to traverse the semantic network of DBpedia. For each topic (node) in a profile, the `dcterms:subject` links are followed to the categories the topic is a member of, from where `skos:broader` and `skos:narrower` are followed to neighboring categories, up to two levels outward. At each category node that is visited during the traversal, the contained topics are also visited, and the frequencies of these visits are counted as a side effect. When the traversal is finished, we take the frequency table, and store it as a measurement of the relatedness between the starting node and the visited topics.

Then, to infer which newly found topics should be included in the profile, we summate the frequency tables of all topics in the profile. From the resulting table, we ignore any topics that are already in the profile, and take the top-10 related topics that are in our vocabulary, and the top-10 topics that are not in our vocabulary.

LinkedIn uses a proprietary algorithm, which likely incorporates aspects of collaborative filtering, to display 20 "related skills" on each of the pages that we used as the basis for our vocabulary. Such lists of related skills are added into a frequency table for each profile, and are further treated identically to the inferences from DBpedia.

6 EVALUATION

The ground truth against which we measure the performance of the method is provided by 8 participants. We have provided them with an interface that allowed them to review and correct their own profile. First they were presented with 184–353 topics that were extracted from all types of documents ($o = ALL$). Participants were asked to remove all topics in which they had no mastery by clicking the corresponding buttons.

The second step for the participants was to correct the estimated mastery levels. Here, statements were presented as boxes (again including the topic name and description) with three sliders for the skill, knowledge, and interest level. Due to the large amount of extracted statements, we randomly omitted 50% of the statements that were based only on extraction from course descriptions or portfolio documents.

The third step was similar to the first, except with 20 inferred topics from DBpedia, and 20 inferred topics from LinkedIn. In the fourth and final step, the participants were asked to add any topics in which they had mastery that were missing from their profile.

It is worth noting that people are prone to over- and underestimating themselves (Dunning et al., 2003). This is, however, not a weakness of our experiment in particular. In CSGF it is still quite common to base a profile of task-related attributes solely on the information that is provided by the learner in question.

The measures *precision*, *recall*, and $F_{0.5}$*-score*, which have been defined in Section 5.3.2, are used to evaluate the performance of our annotation process. We do not average our measures over the profiles, but rather take the counts of $Generated_p$, $Truth_l$, and their intersection per profile, sum the counts, and then compute precision, recall, and $F_{0.5}$ over all profiles. To assess how successful we were at estimating mastery levels, we test for correlation between the estimated and actual levels. We use Pearson's correlation coefficient (r) as a measure, and we report on statistical significance at the levels of 0.10 (˙) and 0.001 (*). Because we included a limited number of inferred topics in the participants' profiles, we cannot use exactly the same measures as with the extracted topic links. Instead, we use *Precision at 10*, which denotes the fraction of inferred topics in which the participants claim to have mastery.

Our results (see Table 1) indicate that the combination of extracted information from all document origins leads to the most accurate profiles. The profiles included a large amount of topics in which the participants actually had some mastery. Course descriptions are the only type of document that could be used to produce profiles of similar quality by itself. Documents from other origins lead to topic links with a comparable precision, but in a quantity that is likely not sufficient for application in CSGF.

[6] Gremlin - < http://gremlin.tinkerpop.com/ >

Table 1: Performance of the Mastery Profiling Method.

Extracted						
Origin	**Topic Links**			**Levels Corr.**		
	Pr.	Re.	$F_{0.5}$	Sk.	Kn.	In.
Coursed.	0.859	0.711	0.825	.210*	.212*	—
LinkedIn	0.935	0.077	0.290	.060	.108	.047
Portfolio	0.811	0.184	0.482	.220*	.058	.073
Website	1.000	0.003	0.013	.853-	.696	.700
ALL	0.845	0.860	0.848	.268*	.241*	.299*
Inferred						
In-Vocabulary		**DBpedia Pr. at 10**			**LinkedIn Pr. at 10**	
Yes		0.825			0.925	
No		0.750			0.875	

Inference can be used to expand learner profiles with high precision. The additional topics that are found by taking into account probabilistic relations between topics from LinkedIn are more precise than those that are found from a traversal over the semantic network of DBpedia. It is more accurate, in both cases, to filter the inferred topics with our vocabulary. A large majority of the inferred topics that are not in our vocabulary is, however, also correct.

Our method was not able to estimate mastery levels with the accuracy that is necessary for CSGF. The estimated mastery levels show a weak but significant correlation with the levels that the participants reported. The differences between origins suggest that course grades and qualifying terms are both indicators for mastery levels, but that the number of sections that is annotated with a given topic is a worse indicator than we expected.

7 DISCUSSION

Our annotation process has produced results that are very promising for use in CSGF. It does not rely on optimizations specific for the field the students are training in. Instead, it relies upon the configurability of DBpedia Spotlight and the broad coverage of professional topics that is used on LinkedIn. Advantages of keeping the method and implementation field-agnostic are the reproducibility of the experiments with students of other fields, and a greater potential to collaborate in the development of the necessary software.

We found that the method makes mistakes that may, however, be overcome with domain-specific optimizations. Abbreviations of field-specific concepts which are commonly used with a different meaning are not disambiguated correctly. We also found that the coverage of the vocabulary was too broad. For example, "Schizophrenia" is in most disciplines never a main topic.

Our results in estimating mastery levels are less promising. It is possible that we have used suitable indicators and that the used data transformations are not right for this task. A post-hoc analysis of our results can clarify this matter to some extent. It will be interesting to see if a method that is based on machine learning, but uses the same features as we have, will fare better in future research.

To make further advances in mastery profiling, we may have to turn to techniques that are outside the scope of the current method. Portfolio documents that were the product of teamwork inherently describe the actions of multiple team members. We would want to distinguish between individuals, and discern "who did what". For course descriptions it holds that not all text indicates what the students will do or learn. Administrative remarks say something about the course or about the teacher, but give no relevant information about the students who have completed the course. Such mistakes ask for more focus on textual relations, as is done in Open Information Extraction (Etzioni et al. 2011).

8 CONCLUSIONS

In this paper, we have presented a method that produces learner profiles on the basis of existing documents that are associated with students. It is able to link students to a large amount of topics, in which they have skill, knowledge, and/or interest, with high precision. We have not yet succeeded in estimating the mastery levels that students have in these topics. Our method can be used as a baseline in future experiments that aim to produce learner profiles from existing documents. We aim to publish our current implementation under an open source license to facilitate this.

Our work is also a demonstration of a novel application of entity linking. We have shown that DBpedia Spotlight can be configured to accurately annotate course descriptions, portfolio documents, and websites. A customized vocabulary was used to filter annotations that are relevant to the mastery profiling task, and to combine the annotations from Dutch and English language documents into a single learner profile. The sets of topics that were extracted from students' associated documents have been successfully expanded by inference over semantic and probabilistic relationships between topics.

We hope that the work that has been described in this paper can serve as a starting point for the inclusion of detailed task-related attributes in learner profiles and, more generally, that it will assist in the adoption of CSGF in higher education.

ACKNOWLEDGEMENTS

Many thanks go to Wouter Weerkamp for valuable guidance and feedback; to Flip van Haaren and Raymond Jelierse at Shareworks for feedback and technical assistance; to the participants for responding to the questionnaire and for reviewing their own learner models; and to Joachim Daiber for clarifying the configurability of DBpedia Spotlight.

REFERENCES

Bull, S. & Kay, J., 2007. Student Models that Invite the Learner In: The SMILI:-) Open Learner Modelling Framework. *International Journal of Artificial Intelligence in Education*, 17(2), pp.89–120.

Craig, M., Horton, D. & Pitt, F., 2010. Forming reasonably optimal groups: (FROG). In *Proceedings of the 16th ACM international conference on Supporting group work*. pp. 141–150.

Dai, H. et al., 2012. From Entity Recognition to Entity Linking: A Survey of Advanced Entity Linking Techniques. In *The 26th Annual Conference of the Japanese Society for Artificial Intelligence*. pp. 1–10.

Daiber, J. et al., 2013. Improving Efficiency and Accuracy in Multilingual Entity Extraction. In *Proceedings of the 9th International Conference on Semantic Systems*. Austria, Graz.

Dunning, D. et al., 2003. Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), pp.83–87.

Etzioni, O. et al., 2011. Open information extraction: The second generation. In *Proceedings of the 22nd international joint conference on Artificial Intelligence*. pp. 3–10.

Han, X., Sun, L. & Zhao, J., 2011. Collective entity linking in web text: a graph-based method. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. pp. 765–774.

Hoogendoorn, S.M., 2013. *Diversity and team performance: A series of field experiments.* PhD Thesis. University of Amsterdam. Available at: http://dare.uva.nl/record/440433.

Jackson, S., May, K. & Whitney, K., 1995. Understanding the dynamics of diversity in decision-making teams. In R. A. Guzzo, E. Salas, & Associates, eds. *Team Effectiveness and Decision Making in Organizations*. San Francisco, pp. 204–261.

Kyprianidou, M. et al., 2011. Group formation based on learning styles: can it improve students' teamwork? *Educational Technology Research and Development*, 60(1), pp.83–110.

Lingard, R. & Berry, E., 2002. Teaching teamwork skills in software engineering based on an understanding of factors affecting group performance. In *Frontiers in Education Conference*. Boston, MA: IEEE.

Livingstone, D. & Lynch, K., 2000. Group Project Work and Student-centred Active Learning: Two different experiences. *Studies in Higher Education*, 25(3), pp.325–345.

Magnisalis, I., Demetriadis, S. & Karakostas, A., 2011. Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field. *IEEE Transactions on Learning Technologies*, 4(1), pp.5–20.

Mendes, P., Jakob, M., et al., 2011. DBpedia Spotlight: Shedding Light on the Web of Documents. In *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics)*. Austria, Graz.

Mendes, P., Daiber, J., et al., 2011. Evaluating DBpedia Spotlight for the TAC-KBP Entity Linking Task. In *Proceedings of the TAC-KBP 2011 Workshop*. Gaithersburg, USA.

Ounnas, A., 2010. *Enhancing the Automation of Forming Groups for Education with Semantics*. PhD Thesis. University of Southampton. Available at: http://eprints.soton.ac.uk/171641/.

Ounnas, A., Davis, H. & Millard, D., 2009. A framework for semantic group formation in education. *Educational Technology & Society*, 12(4), pp.43–55.

Rao, D., McNamee, P. & Dredze, M., 2013. Entity linking: Finding extracted entities in a knowledge base. In T. Poibo et al., eds. *Multi-Source, Multilingual Information Extraction and Summarization*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 93–115.

Rodriguez, M. A. & Neubauer, P., 2011. The Graph Traversal Pattern. In S. Sakr & E. Pardede, eds. *Graph Data Management: Techniques and Applications*. IGI Global.

Wells, C. E., 2002. Teaching Teamwork in Information Systems. In E. Cohen, ed. *Challenges of Information Technology Education in the 21 st Century*. pp. 1–24.

Werbel, J.D. & Johnson, D. J., 2001. The Use of Person – Group Fit for Employment Selection : A Missing Link in Person – Environment Fit. *Human Resource Management*, 40(3), pp.227–240.

Wetz, P. et al., 2012. Matching Linked Open Data Entities to Local Thesaurus Concepts. In *Proceedings of the I-SEMANTICS 2012 Posters & Demonstrations Track*. pp. 6–11.

Wilkinson, I. A. G. & Fung, I. Y. Y., 2002. Small-group composition and peer effects. *International Journal of Educational Research*, 37(5), pp.425–447.

Winter, M., 2004. *Developing a Group Model for Student Software Engineering Teams*. MSc Thesis. University of Saskatchewan. Available at: http://hdl.handle.net/10388/etd-07052004-140018.