

Detection of Online Learning Activity Scopes

Syeda Sana e Zainab¹ and Mathieu D'Aquin¹

Data Science Institute, National University of Ireland Galway, Ireland
{firstname.lastname}@insight-centre.org

Abstract. During last ten years, online learning has been on the ascent as the advantages of access, accommodation, and quality learning are beginning to take shape. A key challenge is to identify learning activities and recognize how they participate in the learner's progress. In this paper, we look at the way this problems becomes even more challenging when considering the full set of online activities carried out by a learner, as compared to what is achieved on specific online platforms that are dedicated to learning. We in particular show how the integration of linked data-based information can help resolve the issue of representing activities for the purpose of identifying the key topics on which a learner is focusing, in a hierarchical clustering method. We apply this approach in the context of the AFEL project, and show how it performs in realistic use cases of an online learner's profile, comparing general browsing with the use of a dedicated learning platform.

Keywords: eLearning· Text mining· Hierarchical Clustering

1 Introduction

Online learning, often called “eLearning”, is a new class of learning that has been flourishing during the last ten years, following the growing adoption of eLearning in universities, schools, and companies. Online learning provides a convenient environment where instructors and students may access the course, study, and perform interactive learning activities with fewer time and space restrictions. However, the analysis of such online learning activities often requires systems and applications using learning analytics and data mining approaches. In many cases learning activities which are contributing to the learner's progress are not well detected and are outside of these systems and applications.

In the AFEL (Analytics for Everyday Learning) project¹, we aim to address this challenge by defining a process for detecting online learning activities enriched with specific “topic-based” learning trajectories and visualize them in an application that enables potential learners to reflect on their activities and ultimately improve the way they focus their learning.

The aim is to apply this to both online platforms dedicated to learning and to general online activities, showing how learning happens in everyday activities.

¹ <http://afel-project.eu>

Didactalia² is the online social learning platform we use as a testbed. We can directly connect to Didactalia through its API and retrieve metadata about its resources, allowing us to classify them in different topics, and to assess their value to the learning of the user.

However, when generalising this approach to multiple platform, we cannot guaranty that a meaningful description will be available for the resource considered. To take an extreme case, we are testing the detection and assessment of learning activities within the overall web browsing activities of a user. In this case, we need to face issues related to the differences in scale and velocity of the data we received. We also need to find ways to retrieve information that can help understanding the content of resources and assess them with respect to learning.

In this paper, we describe a method to identify key areas in which users of online resources are learning, by: 1- enriching the resources considered based on linked data, 2- clustering activities using this enriched information, and 3- identify the clusters that are most representative of the user’s learning through learning indicators also using the enriched information.

2 Related Work

Research demonstrates that assessment of the student learning is a key to success of any education system, and is “one of the main considerations separating capable schools and instructors from incapable ones”[4]. This also applies to online learning, e.g Helic et al. [10] contend that good online learning requires monitoring of a student’s progress with the material and testing of the gained information and aptitudes all the time. Various measurements enable educators to monitor the student input, feedback, and progress towards goals and is crucial in online education [16]. Few works have been done to analyze learning through a social media [3] but they are more focused towards the learning perspective of social network rather than detecting learning aspects in online activities. Some researches consider the combination of semantic web technologies with learning activities analysis [5,8], but they are more focused towards the integration of data.

Ferguson and Buckingham Shum [7] proposed the concept of learning activities analysis to specifically capture the social interactions underlying the social learning processes. Similarly, a model for mobile and ubiquitous learning environments has been proposed by Aliohani and Davis [1] for the learning analysis. A further action towards learning analysis methods which take into account the amount of data produced by the learner’s activities both in formal and informal settings is offered by the semantic web [2], that pointed out the practical educational benefits for teaching and learning. Some of the learning analytical work has been focused on health care systems. Such systems [17,19] typically employ sensors e.g. wearable sensors and visual sensors, collect a users activities data such as eating or exercise and apply machine learning algorithms to identify

² <https://didactalia.net>

activities progress. A significant work have been done in applications and tools development [9,20,11,13,12] for learning activities analysis. However these applications enable learners to track their progress on certain areas (games, music, document writing etc) or communities (Players, Moodles etc).

Among the diverse works that have been proposed on the analysis of learning activities, a few has been focused on scopes detection in online learning activities. Our work is focused towards the identification of online learning activities scopes and develop an application that helps the learner in their online learning progression.

3 Motivation: The AFEL Project

In several other areas than learning where self-directed activities are prominent (e.g. fitness), there has been a trend in recent years following the technological development of tools for self-tracking. Those tools quantify a specific user's activities with respect to a certain goal (e.g. being physically fit) to enable self-awareness and reflection, with the purpose of turning them into behavioral changes. While the actual benefits of self-tracking in those areas are still debatable, our understanding of how such approaches could benefit learning behaviors as they become more self-directed remains very limited.

AFEL is an European Horizon 2020 project which aim is to address both the theoretical and technological challenges arising from applying learning analytics in the context of online, social learning. The pillars of the project are the technologies to capture large scale, heterogeneous data about learner's online activities across multiple platforms (including social media) and the operationalization of theoretical cognitive models of learning to measure and assess those online learning activities. One of the key planned outcomes of the project is therefore a set of tools enabling self-tracking on online learning by a wide range of potential learners to enable them to reflect and ultimately improve the way they focus their learning.

Below is a specific scenario considering a learner not formally engaged in a specific study program, but who is, in a self-directed and explicit way, engaged in online learning. The objective is to describe in a simple way how the envisioned AFEL tools could be used for self-awareness and reflection, but also to explore what the expected benefits of enabling this for users/learners are:

Jane is 37 and works as an administrative assistant in a local medium-sized company. As hobbies, she enjoys sewing and cycling in the local forests. She is also interested in business management, and is considering either developing in her current job to a more senior level or making a career change. Jane spends a lot of time online at home and at her job. She has friends on Facebook with whom she shares and discusses local places to go cycling, and others with whom she discusses sewing techniques and possible projects, often through sharing YouTube videos. Jane also follows MOOCs and forums related to business management, on different topics. She often uses online resources such as Wikipedia and online

magazines. At school, she was not very interested in maths, which is needed if she wants to progress in her job. She is therefore registered on Didactalia³, connecting to resources and communities on maths, especially statistics.

Jane has decided to take her learning seriously: She has registered to use the AFEL dashboard through the Didactalia interface. She has also installed the AFEL browser extension to include her browsing history, as well as the Facebook app. She has not included in her dashboard her emails, as they are mostly related to her current job, or Twitter, since she rarely uses it.

Jane looks at the dashboard more or less once a day, as she is prompted by a notification from the AFEL smart phone application or from the Facebook app, to see how she has been doing the previous day in her online social learning. It might for example say “It looks like you progressed well with sewing yesterday! See how you are doing on other topics...” Jane, as she looks at the dashboard, realizes that she has been focusing a lot on her hobbies and procrastinated on the topics she enjoys less, especially statistics. Looking specifically at statistics, she realizes that she almost only works on it on Friday evenings, because she feels guilty of not having done much during the week. She also sees that she is not putting as much effort into her learning of statistics as other learners, and not making as much progress. She therefore makes a conscious decision to put more focus on it. She adds new goals on the dashboard of the form “Work on statistics during my lunch break every week day” or “Have achieved a 10% progress compared to now by the same time next week”. The dashboard will remind her of how she is doing against those goals as she goes about her usual online social learning activities. She also gets recommendations of things to do on Didactalia and Facebook based on the indicators shown on the dashboard and her stated goals.

While this is obviously a fictitious scenario, it highlights the key challenges faced by the project. The one specifically addressed by this paper is the identification of the topics on which the learning of the user is mostly focusing, so that the activities related to those topics can be assessed in the context of the relevant learning scope.

4 Overview

Figure 1 provides a general view of the data architecture used in the AFEL project. In summary, an external platform provides, through the POST API, a stream of activities, including an identifier of the user performing the activity u , the time at which the activity happen t , and a reference to the resource being used as part of the activity r . The POST API will then both index those activity descriptions, and call an internal service called the “Resource Indexer”, which role is to obtain additional information about the resources being used by the learner, and enrich them with information necessary for both detecting

³ <http://didactalia.net>

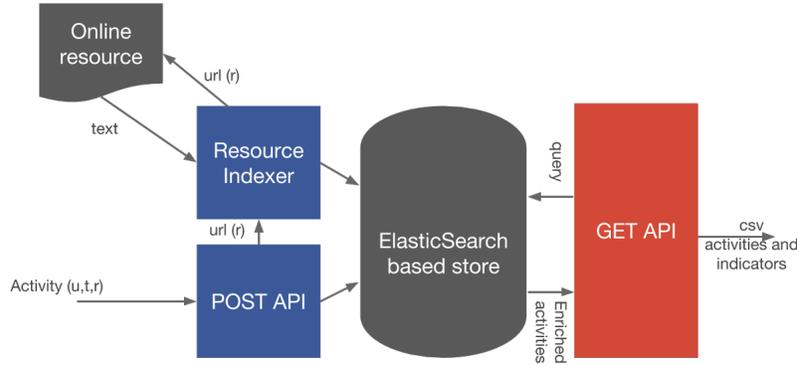


Fig. 1. Overview of the AFEL data architecture.

the learning scopes (the topics on which they are learning) and in assessing their learning progress within those scopes. Based on all those information being stored and indexed, the role of the GET API is to retrieve them and provide to the AFEL application all the relevant data for a given user, including the activities performed, in which learning scope they belong, and indicators of how much they contribute to the learning trajectory of the learner.

Although not the topic of this paper, a key challenge here is in identifying the indicators that can support assessing the progress of a learner in a certain learning scope. Based on theoretical work in educational psychology within the AFEL project (see for example [15]), the approach taken is to try to recognize to what extent encountering and processing a certain artifact (a resource) induced learning, based on representing “frictions” that are bringing new knowledge or new forms of knowledge to the learning. At the moment, we distinguish three forms of “frictions”, leading to three categories of indicators of learning (see [6]):

- New concepts and topics: The simplest way in which we can think about how an artifact could lead to learning is through its introduction of new knowledge unknown to the learner. This is consistent with the traditional “knowledge acquisition metaphor” of learning. In our scenario, this kind of friction happens for example when Jane watches a video about a sewing technique previously unknown to her. We call the indicator associated with this form of friction *coverage*.
- Increased complexity: While not necessarily introducing new concepts, an artifact might relate to known concepts in a more complex way, where complexity might relate to the granularity, specificity or interrelatedness with which those concepts are treated in the artifact. In a social system, the assumption of the co-evolution model is that the interaction between indi-

viduals might enable such increases in understanding of the concepts being considered through iteratively refining them. In our scenario, this kind of friction happens for example when Jane follows a statistics course which is more advanced than the ones she had encountered before. We call the indicator associated with this form of friction *complexity*.

- New views and opinions: Similarly, known concepts might be introduced “in a different light”, through varying points of views and opinions enabling a refinement of the understanding of the concepts treated. This is consistent with the co-evolution model in the sense that it can be seen either as a widening of the social system in which the learner is involved, or as the integration into different social systems. In our scenario, this kind of friction happens for example when Jane reads a critical review of a business management methodology she has been studying. We call the indicator associated with this form of friction *diversity*.

While in the implementation of the AFEL application, all three indicators are being used, we will here focus only on the first one, as the two others are the subject of specific work. To simplify, we will therefore consider the problem tackled as:

How to identify learning scopes in a stream of activities, where learning scopes represent a set of activities that cover a particular topic by the learner?

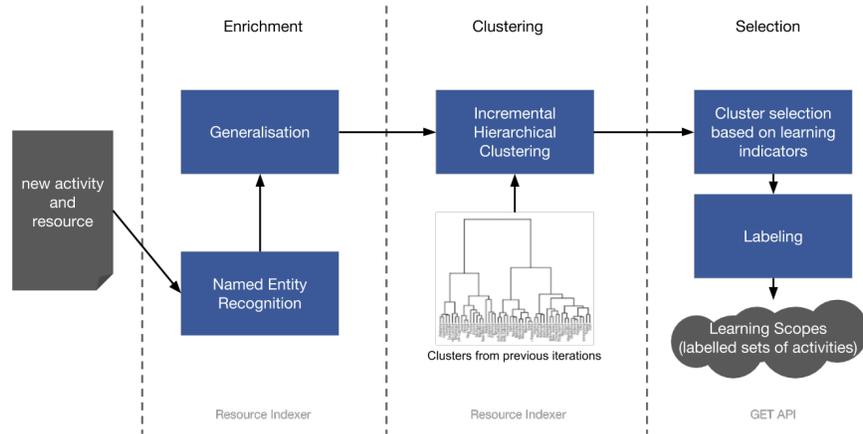


Fig. 2. Overview of the method used for identifying learning scopes.

While there are many ways in which this question could be answered, we here consider the key requirements that the method should equally work in cases

where resources are already described with rich metadata (e.g. tags) and in cases where we have no other information about the resources than their content. It also needs to take into account the dynamic aspect of the scenario (that new activities are constantly being added).

We therefore describe in the next sections the three main steps of the method represented in Figure 2.

5 Enrichment

The goal of the enrichment phase of the process described in Figure 2 is to extract from the resources used in learning activities information about their content, in the case such information is not already available. Indeed, in many cases if we focus on specific platforms, resources will be classified and associated with subjects and topics. This is the case for example of Didactalia which associates with each resource a set of tags, as provided by the users who contributed the resources. Those tags can then be used to represent a general overview of the content of the resource. However, when working with general online resources, we cannot rely on the availability of such metadata.

Using named entity recognition is a common approach to this problem. By extracting from the textual content of the resource key entities and concepts that are being mentioned, we can build a profile of the resources that can potentially be used to replace the tags available in existing platforms.

We use DBpedia Spotlight⁴ as an off-the-shelf named entity recognition tool. The advantage of DBpedia spotlight is that, practically, it is open source and can be deployed locally, removing the need to rely on an external service. This is especially important in our scenario as we need to process thousands of potentially large texts for each user. Since it is based on Wikipedia, in addition, the vocabulary of entities being covered is very wide and domain-independent, with millions of entities being recognizable on all sorts of topics.

The other advantage of DBpedia Spotlight is that the entity extracted are part of the DBpedia Linked Dataset.⁵ DBpedia is a linked data version of Wikipedia and, as such, also contains the taxonomy of categories of Wikipedia. This is especially useful here since our goal is to group activities based on their associated resources covering similar topics. To achieve this however, entities extracted are insufficient. Indeed, resources might be on the same topic (e.g. geography) without mentioning the same entities. Nevertheless, if they are indeed of the same topic, it is more likely that the categories of the entities mentioned, or their super-categories, will overlap.

The goal of the generalization process is therefore to augment the profile created by named entity recognition by adding the categories that relate to the entities found through named entity recognition to the profile of each resource. To make this efficient, it relies on an index of the “direct” categories of each entity and on an index of the DBpedia category hierarchy. Each entity is then

⁴ <https://github.com/dbpedia-spotlight/dbpedia-spotlight>

⁵ <http://dbpedia.org>

iteratively assigned categories at a higher level than its direct ones by climbing up this hierarchy until a given level. This process is of course carried out offline, with the online process only executing named entity recognition, and adding the pre-computed set of categories to the entity.

It is important to note here that the profile of a resource in this case is richer than the set of tags that might be available in online platforms. It is not only larger (the number of entities and categories found will generally be bigger than the number of tags manually used to describe a resource), it also contains more information, as multiple mentions of a given entity are counted, and multiple references to a category are also taken into consideration. In this way, if more than one entity are related to a given category, this category will have a stronger weight during the clustering phase.

6 Clustering

The objective of the enrichment method described above is to build a profile for resources that the learner has been using that captures the key topics, so that they could be grouped into general learning scopes. The next step is therefore to cluster activities based on those profiles. Here too, there are many ways to cluster based on the kind of resource profiles we have constructed in the previous step, which very much represent vectors of term frequencies. However, considering the nature of the scenario in which we are operating, we need to consider three key requirements:

- *The clustering cannot be static*: As the building the learning scopes is to be carried out from the stream of user activities, it obviously cannot be built in advance.
- *The clustering needs to be incremental*: Very related to the previous requirement, but adding that the clustering mechanism needs to be fast, the approach is required to be incremental, i.e. new activities and resources need to be added into an already constructed set of clusters, from previous activities and resources. It is also important that the set of clusters do not evolve dramatically due to the appearance of new activities and resources.
- *The number of clusters cannot be fixed in advance*: Many clustering mechanisms take as a given parameter the target number of clusters. However, this should in our case be automatically set, since we cannot decide in advance for every user how many topics they are learning about.

For those reasons, we adopted an incremental hierarchical clustering approach. Hierarchical clustering [14] is a method by which clusters are formed by progressively grouping items, creating a hierarchy of larger and larger clusters. One advantage of hierarchical clustering is that it creates many clusters of different sizes and levels, from which we can select afterwards based on their properties (see next section), rather than having to choose a number of clusters in advance.

However, the basic algorithm for hierarchical clustering assumes that the whole set of items to clusters are available. We therefore adapt it so that it can work incrementally, and fulfill our two first requirements. The basic algorithm we use is described below and broad terms. It represents the function to add a new item to an already existing cluster.

```
function add(item, clusterset)
  if clusterset is empty then
    c = new cluster([item])
    insert c in cluster set
  else
    find c the cluster in clusterset most similar to item
    nc1 = new cluster(concat(c, item))
    nc2 = new cluster(item)
    p = parent(c)
    remove c from children of p
    add nc1 as child of p
    add c as child of nc1
    add nc2 as child of nc1
```

By using this algorithm for every new items one after the other, a set of clusters is built organized as a hierarchy. The advantage is that the clusterset can be kept from one activity to the other, meaning that a new activity would only have to be added to it, and will not be require large amounts of computation. It is useful to note that

- The similarity measure used can vary. In our initial tests, we use an euclidean distance on the frequency vectors of terms (entities and categories) as obtained above. Further tests with other similarity measures (e.g. based on a cosine distance over TF.IDF vectors) are being conducted, giving better results.
- The creation of new a cluster computes a term frequency vector aggregating the vectors of all included items using their average. Therefore, the similarity is applied between those aggregated vectors and the original vector of the item to be added.

It is also useful to mention that this approach is not guarantied to obtain the same clusterset as one that would have applied hierarchical clustering in a non-incremental, static manner. Methods exist to achieve better approximations than the proposed algorithm (see e.g. [18]) but those tend to be computationally expensive. Also, as mentioned above, an advantage of this approach is that the clusterset only changes in a very localized manner, keeping the established structure mostly intact and is therefore less likely to lead to dramatically changing results for the user.

7 Cluster selection and labelling

The last step to be considered is the selection of the set of clusters of activities that are most representative of the learner's interests, and their labelling. Indeed, we have obtained from the previous step a clusterset which is hierarchically organised (it is, actually, a binary tree). The idea is to identify the ones that seem to be most interesting from the point of view of the learning indicators considered (see earlier description of the learning indicators).

We therefore start by ranking the clusters in the clusterset according to a score. While not taking into account the hierarchy, an important aspect of this score is that it takes into account the temporal aspects of activities. Indeed, if taking into account only the coverage indicator, as described above, the idea is to measure for each activity how many new concepts (entities and categories) it introduces into each of the clusters as a proportion of the concepts already present there from previous activities. Taking that into account, we can compute the average coverage of each cluster based on the included activities, effectively corresponding to a measure of how much, on average, an activity in this cluster increases the coverage of the topics of the cluster. It is worth noting that, while new activities (and therefore new clusters) affect the score of certain clusters and, obviously, their ranking based on this score, this can still be achieved incrementally, i.e. we can update the scores of clusters based on new activities, and update the ranking based on those changes without having to recompute it every time.

The results of the previous step is a ranked clusterset *rankedcl*, sorted in descending order of the score considered. The selection process from there can be reduced to selecting non-overlapping clusters that are highly ranked. To support the non-overlapping property, we exploit the hierarchical relationship that exist between clusters from the hierarchical clustering process, i.e. selected clusters should be taken from distinct branches of the tree as per the algorithm below:

```
function select(rankedcl) returns selected
  foreach c in ranked
    select = true
    foreach s in selected
      if c is a (direct/indirect) child/parent of s
        select = false
    if select
      insert c in selected
```

This approach has the two advantages that, because of the non-overlapping property of clusters being selected, we do not need to set a threshold for the cluster score to be selected, since, once a cluster is selected, all overlapping clusters are automatically removed from the candidate list. This means that the results are a limited subset of all the clusters that are complementary in terms of topic, and have the best coverage in score. This also means that clusters are selected to represent learning activities, and further filters can be applied later

based on other indicators to emphasise the clusters related mostly to learning activities.

The final results of this phase are a set of selected clusters grouping similar activities that have on average a good contribution to the coverage of the topic of the cluster, and characterised by a vector of term frequencies from the entities and categories in DBpedia. The last step is to label those clusters. For this, we take the simple approach to choose as label the entity or category in the vector with the greatest weight as label for a given cluster.

8 Application in two case studies

In order to test and validate the method described above, we apply it in two different scenarios with different sources of data. Those two scenarios actually rely on the same application, and on the same architecture depicted in Figure 1, in the context of the AFEL project. These scenarios however differ in the data sources they consider. In one case, the activity data is taken from an online social platform, Didactalia, where there is a limited amount of activities relating to well described resources, and clearly associated with learning. In the other case, the data is taken from the general browsing history of the user, generating much more and much more frequent activities, associated to resources that have no or very little metadata, and which are not necessarily directly related to learning.

Didactalia is an online, social learning platform. It includes more than 100,000 resources that are contributed and annotated by users, and shared within various communities. The whole platform relies on linked data technologies, and enable connecting resources with each other. We collect activity data from Didactalia through using a javascript snippet similar in style to the one used by web analytics tools such as Google Analytics or Piwik. The information about the resources are provided through an RDF-based API on the Didactalia platform.

The results of applying our application to Didactalia are shown in an example in Figure 3. The application first shows the detected learning scopes for the user in a world cloud depending on the indicators considered. Each learning scope can then be further explored, showing specific indicators, recommendations, etc. The detection of learning scopes here is therefore critical. While there are tens of thousands of users on Didactalia, and we therefore collect millions of data points over time, each user will only carry out between a few activities and in the order of one hundred. The tags associated with the resources can be used as the initial profile vector for each activity/resource, possibly in combination with the entities and categories from DBpedia. Considering the low number of activities, the complexity of the process is less critical, although it needs to be handled for a large number of users. Also, compared to the second scenario, each of the learning scopes is more or less guaranteed to be related to learning, since they only capture activities that are carried out on a learning platform.

On the other hand, the application of our method to the user's whole browsing history (shown in an example in Figure 4) has to deal with a lot of activities for each users (up to thousands per day), with very little overlap and

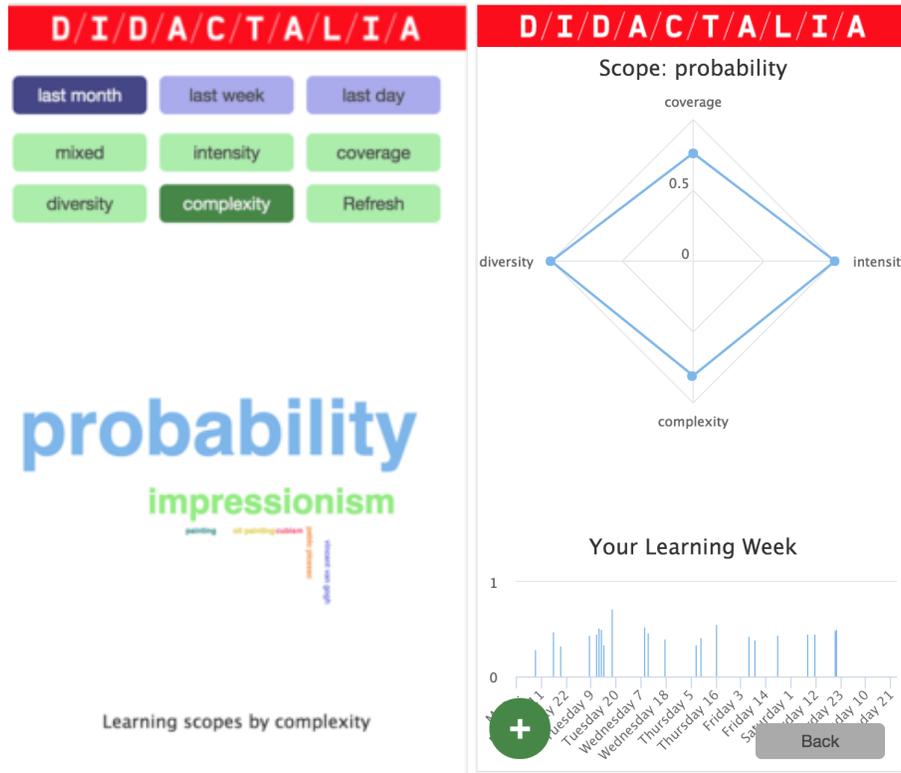


Fig. 3. Application for Didactalia, using learning scope detection (right) and showing learning indicators for a given learning scope (right).

not much description. The data in this case is collected through a specially developed browser extension⁶ (compatible with browsers supporting web-ext). The application functions very much in the same way, except that in this case the learning scopes detected are much broader and have to be based on the entities and categories from Dbpedia. In this case as well, the scale is much more of an issue. The built hierarchical cluster set can have tens of thousands of nodes for one user, and it would be unfeasible to rebuild it every time. It is therefore important, as we have chosen to do, to update it incrementally every time a new activity appears.

While the accuracy of the learning scope detection is subjective, the applications described above demonstrate that the proposed method, combining incremental hierarchical clustering with activity profiles based on named entity recognition and abstraction through DBpedia categories can indeed be applied and adapts well to different contexts.

⁶ <https://github.com/afel-project/browsing-history-webext>

Alternatives and versions of the techniques and tools used are available for other languages, but a language detection mechanism will be needed in order to be able to direct the processing of a resource to the right version depending on the main language of the textual component of that resource.

This also raises a clear other limitation, i.e. that our method is dedicated to online resources which content is mainly textual. This restricts its application especially in the eLearning process where more and more multimedia resources are being used. While this can be seen as a problem, methods are available to retrieve textual description and content from many types of media that can be used to alleviate this issue.

Finally, as already mentioned, it appears obvious that the set of topics one is mostly interested in learning about is a very subjective matter. We therefore cannot expect for our method to ever be entirely accurate. Enabling interaction with the constructed learning scopes and clusters seems therefore to be a promising solution here. Indeed, we plan to integrate feedback from the user into our approach, allowing them to mark certain clusters as “persistent” (i.e. they should not disappear even if other clusters appear more important), irrelevant (i.e. their activities should be either moved to other clusters or considered as not interesting from the point of view of learning), merged (i.e. two learning scopes might need to be combined) or as containing a different set of activities (i.e. the user should be able to re-assign activities manually). The incremental hierarchical clustering method described in this paper would therefore have to be updated to be able to integrate those feedback.

10 Acknowledgment

This work has received funding from the European Union’s Horizon 2020 research and innovation programme as part of the AFEL (Analytics for Everyday Learning) project under grant agreement No 687916, and supported by the Insight Centre for Data Analytics, funded by Science Foundation Ireland.

References

1. Naif R Aljohani and Hugh C Davis. Learning analytics in mobile and ubiquitous learning environments. 2012.
2. Patrick Carmichael and Katy Jordan. Semantic web technologies for education—time for a turn to practice? *Technology, Pedagogy and Education*, 21(2):153–169, 2012.
3. Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan. Mining social media data for understanding student’s learning experiences. *IEEE Transactions on Learning Technologies*, 7(3):246–259, 2014.
4. Kathleen J Cotton. Monitoring student learning in the classroom. school improvement research series close-up# 4. 1988.
5. Mathieu d’Aquin and Nicolas Jay. Interpreting data mining results with linked data for learning analytics: motivation, case study and directions. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, pages 155–164. ACM, 2013.

6. Mathieu d'Aquin, Alessandro Adamou, Stefan Dietze, Besnik Fetahu, Ujwal Gadiraju, Ilire Hasani-Mavriqi, Peter Holtz, Joachim Kimmerle, Dominik Kowald, Elisabeth Lex, et al. Afel: Towards measuring online activities contributions to self-directed learning. In *Proceedings of EC-TEL 2017 workshop ARTEL*, 2017.
7. Rebecca Ferguson and Simon Buckingham Shum. Social learning analytics: five approaches. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 23–33. ACM, 2012.
8. Giovanni Fulantelli, Davide Taibi, and Marco Arrigo. A semantic approach to mobile learning analytics. In *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, pages 287–292. ACM, 2013.
9. Danita Hartley and Antonija Mitrovic. Supporting learning by opening the student model. In *International Conference on Intelligent Tutoring Systems*, pages 453–462. Springer, 2002.
10. Denis Helic, Hermann Maurer, and Nick Scherbakov. Web based training: What do we expect from the system. In *Proceedings of ICCE*, pages 1689–1694, 2000.
11. Riccardo Mazza and Vania Dimitrova. Coursevis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, 65(2):125–139, 2007.
12. Riccardo Mazza and Christian Milani. Gismo: a graphical interactive student monitoring tool for course management systems. In *International Conference on Technology Enhanced Learning, Milan*, pages 1–8, 2004.
13. Colin McCormack and David Jones. *Building a web-based education system*. John Wiley & Sons, Inc., 1997.
14. Fionn Murtagh. A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4):354–359, 1983.
15. A. Oeberst, J. Kimmerle, and U. Cress. What is knowledge? who creates it? who possesses it? the need for novel answers to old questions. *Mass collaboration and education*, 2016.
16. Lawrence C Ragan. Good teaching is good teaching. an emerging set of guiding principles and practices for the design and development of distance education. *Cause/Effect*, 22(1):20–24, 1999.
17. Felipe Barbosa Araújo Ramos, Anne Lorayne, Antonio Alexandre Moura Costa, Reudismam Rolim de Sousa, Hyggo O Almeida, and Angelo Perkusich. Combining smartphone and smartwatch sensor data in activity recognition approaches: an experimental evaluation. In *SEKE*, pages 267–272, 2016.
18. Arnaud Ribert, Abdel Ennaji, and Yves Lecourtier. An incremental hierarchical clustering. In *Proceedings of the Vision Interface Conference*, pages 586–591, 1999.
19. Christian Seeger, Alejandro Buchmann, and Kristof Van Laerhoven. myhealthassistant: a phone-based body sensor network that captures the wearer's exercises throughout the day. In *Proceedings of the 6th International Conference on Body Area Networks*, pages 1–7. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.
20. Diego Zapata-Rivera and Jim E Greer. Exploring various guidance mechanisms to support interaction with inspectable learner models. In *International Conference on Intelligent Tutoring Systems*, pages 442–452. Springer, 2002.