Combining Content Analytics and Activity Tracking to Identify User Interests and Enable Knowledge Discovery

Andrii Vozniuk, María Jesús Rodríguez-Triana, Adrian Holzer, and Denis Gillet

REACT Group, EPFL, Station 9, 1015 Lausanne, Switzerland.
Email: {andrii.vozniuk, maria.rodrigueztriana, adrian.holzer, denis.gillet}@epfl.ch

ABSTRACT

Finding relevant content is one of the core activities of users interacting with a content repository, be it knowledge workers using an organizational knowledge management system at a workplace or self-regulated learners collaborating in a learning environment. Due to the number of content items stored in such repositories potentially reaching millions or more, and quickly increasing, for the user it can be challenging to find relevant content by browsing or relying on the available search engine.

In this paper, we propose to address the problem by providing content and people recommendations based on user interests, enabling relevant knowledge discovery. To build a user interests profile automatically, we propose an approach combining content analytics and activity tracking. We have implemented the recommender system in Graasp, a knowledge management system employed in educational and humanitarian domains. The conducted preliminary evaluation demonstrated an ability of the approach to identify interests relevant to the user and to recommend relevant content.

Categories and Subject Descriptors

K.3.1 [**Computers and Education**]: Computer Uses in Education; H.5.2 [**Information interfaces and presentation**]: User interfaces

Keywords

Learning Analytics, Educational Data Mining, Interests Mining, Knowledge Discovery, Recommender System, Content Analytics, Text Mining, Activity Tracking, Information Retrieval

1. INTRODUCTION

Knowledge plays an essential role in value creation in the post-industrial economy. Knowledge is acquired and enriched in learning, which often takes place at a workplace or in an educational setting. While learning, people interact with content as a knowledge medium, located in various content repositories. In an educational setting, such content repositories are usually learning environments, where both students and teachers interact with the content found there. Teachers would regularly interact with the content, when preparing a course while students - when following a course or just collaborating with peers.

When working on a course in a learning environment, teachers enrich the system with relevant materials including text files, web links, videos, audio recordings coming from their device or the cloud. Other teachers can benefit from content already available in the platform when preparing their courses. Moreover, it may also be beneficial for the students to have access to the content that is relevant to their interests, but which the teacher did not directly include into her course [25]. In the case of learning environments with a vast number of content items, it may be hard for the user to find content items corresponding to her interests.

To address the mentioned issues, we propose to employ a recommender system that combines the content analytics, activity tracking, and information retrieval techniques to (1) build the user interests profile and afterward (2) to suggest content relevant to the user and users with similar interests enabling knowledge discovery. To perform the recommendation, first, for each item available in the content repository, we employ natural language processing techniques to identify a set of concepts related to the content in a similar way how humans would do it. Relying on high-level concepts instead of specific words present in the text when constructing user interests profile and afterward finding similar items, allows to identify the content that covers the same high-level concepts even if the specific words used in it are different. Next, we analyse the interactions of the users with the content items based on available user activity recordings and aggregate the concepts in the content that the user interacted with building in this way the user interest profile. Finally, we use information retrieval techniques to recommend to the user relevant content based on the similarity between the concepts in the content and concepts identified as user interests. In the same way, our approach allows finding relevant users based on the determined interests similarity. Our approach puts the user in control of her interests profile and allows to adjust the interests by removing concepts if necessary, as in the case when the user is not interested at the moment in some of the identified concepts.

To evaluate the usefulness and the performance of the algorithm, we have implemented the approach in Graasp, a knowledge management system used in educational [2] and humanitarian [24] settings. Afterward, we have evaluated the approach with teachers, identifying their interests and providing them with recommendations.

This paper describes the algorithm used, the implementation details of the approach in Graasp and the evaluation of the approach with users. The structure of this paper is as follows. First, Section 2 reviews some of the relevant approaches to content analytics, activity tracking, and knowledge discovery. Afterward, Section 3 explains our approach to constructing user interests profile and demonstrates how we make recommendations based on the interests. Section 4 illustrates an implementation of the proposal, while Section 5 talks about the evaluation methodology and the results. Finally, Section 6 presents the conclusions and highlights directions for the future work.

2. RELATED WORK

In this section, we review relevant work from the domains of content analytics, activity tracking, user interests mining and take a look at notable systems supporting knowledge discovery.

2.1 Content Analytics and Activity Tracking

Content Analytics. Content analytics allows the machine to gain an understanding of the content, similarly to how a human would do it by, among others, extracting the main topics, concepts, and entities present in the content. Kovanovic et. al. did an extensive overview of content analytics as one of the often employed techniques in the domain of learning analytics [12]. For instance, in the line of our work, Bosnic et al. proposed to use automatic extraction of keywords from textual content as a foundation for content recommendations [3]. It is worth noting that existing papers focus mainly on analysis of textual content [12], while recent progress in the understanding of multimedia formats, such as object recognition in images or videos [13], or speech recognition allow broadening the scope of content analytics from purely textual information to the various multimedia formats.

Understanding the content alone is not sufficient for understanding the learning since, according to Moore, learner-content interaction is a defining characteristic of education [14]. Moore argues that such learner-content interaction is necessary to happen for the education to take place since "it is the process of intellectually interacting with content that results in changes in the learner's understanding, the learner's perspective, or the cognitive structures of the learner's mind" [14]. Recognising the importance of the interaction, below we consider approaches to capturing and persisting the interactions through activity tracking.

Activity Tracking. User activities tracked by a learning platform is a common data source in the field of learning analytics [18, 19]. Usually, a learning management system or a learning environment have a logging infrastructure in place that records how the user interacts with the platform [18]. The more modern educational platforms support a structured representation of user activities using well-defined formats including ActivityStreams used in [23], xAPI employed in [11] or IMS Caliper outlined in [19]. On a high-level, all these three formats record user-platform interactions in the form of the actor-verb-object triplet capturing who did what with what on the platform. However, on a more detailed level, each format captures additional aspects of the interaction. In the triplet, the verb indicates the type of interaction, for instance, the verb "accessed" would mean that the user viewed content, "downloaded" - downloaded the content and so on. Having a common set of verbs with a well-defined meaning is critical for being able to benefit from user interactions captured by several platforms [11].

Combining Both. While there is a considerable number of studies employing content analytics or relying on interaction analysis, the number of studies combining both is still somehow limited even taking into account that it is considered a promising direction [18, 12]. One noticeable recent proposal combining the both approaches is by Kim et. al. [10] where they use content analytics and recorded interaction data to understand better how students learn with video and eventually to improve their experience, for instance by explaining better the identified confusing topics. Following these recommendations, we consider the combination of both content analytics and activity tracking as a core part of our proposal.

2.2 Mining User Interests

The obtained user interests can be used for different purposes, including privacy awareness and recommendations. Harkous et. al. proposed in [9] to employ a content analysis of the files located on Google Drive of a user to understand the topics, concepts, and entities relevant to the user. They used the obtained information with the goal to improve the user awareness through a new permissions model called Far-reaching Insights. This model informs the user about the insights that third-party applications can derive about her based on the accessible Drive data given the requested permissions are granted. In our approach, we want to explore how identified interests can be used to provide the user with relevant content. In the following subsection, we review some of the systems enabling knowledge discovery with such recommendations.

2.3 Knowledge Discovery Systems

Klamma et al. have formulated a set of requirements for a collaborative adaptive learning platform [16]. One of the requirements is "Support for personalized learning resource delivery through an intelligent adaptive engine, being able to connect people to the right knowledge and deliver quality learning resources that are tailored to the learner's preferences and learning goals." [16]. Learning platforms often integrate such engine in a form of a recommender system. Drachsler et. al. have conducted an extensive review of 82 recommender systems used to support learning in [5]. Below, we take a look at several proposals, particularly relevant to our approach.

Zaldivar et. al. address in [25] the problem of discovering by the instructors relevant learning resources used by students when learning, that are not part of the materials provided by the instructor but still can be beneficial for the students. In their approach, the authors record the web pages that students visit and perform a lexical analysis of the page content. Afterward, they apply information retrieval techniques to identify the online content (webpages) that are the most similar to the content provided by the instructors as part of the course.

In [6] El Helou et. al. proposed a recommender system that considers user interactions with content items to construct a user-content associations graph. After the graph is built, the system applies a ranking algorithm to provide the user with personalized recommendations of relevant actors, activity spaces and knowledge assets taking into account the context.

Motivated by the presented approaches, in the next section, we propose to employ a recommender system that combines content analysis, activity tracking to identify user interests and information retrieval techniques to suggest relevant content and people.

3. INTERESTS-BASED RECOMMENDER

In this section, we explain how our approach works by first automatically identifying user interests and after using the interests to obtain relevant content and people.

3.1 Identifying User Interests

To identify user interests, our system needs, first, to understand the concepts covered in the content. Second, it requires recorded user activities to know how the user interacts with the content items. Having both the concepts and the activities, the system can construct the user interests profile. Below, we explain each component of the approach.

Content Analytics. Content can be available in multiple formats, and a data processing pipeline needs to be built to extract concepts from the content and after store them in an index for further use. A general representation of the key steps of the pipeline is shown on Figure 1 where different types of content may go through different processing steps to obtain the concepts.

On the first step, textual content is extracted from the stored items. In the second step, the content analysis is performed. For the content analysis, we considered using named entity recognition (NER), concept extraction, and topic modelling. Since NER picks entities only from the words present in the text, using such entities for recommendations would limit the discovery only to the content containing them directly. Differently, high-level concepts allow identifying relevant content even if the specific words used in it are different. When we applied topic modelling to real data, the identified topics having a high level of abstraction did not seem to capture well the content particularities. For these reasons, we use a set of concepts to describe the content. Finally, on the third step the extracted content and the concepts are tokenized and put into a searchable index so that they can be used on the recommendation step.

Activity Tracking. Our approach requires recording user-content interactions, namely the triplet user-verb-object. We consider different types of interactions as a manifestation of different interest strength. For instance, intuitively when a user downloads the content it manifests a stronger interest in the content compared to just viewing it online. Our approach does not assume a specific activity recording technique or data format used, but it requires the approach to capture the user identifier, the verb indicating the type of interaction and, the identifier of the resource the user has interacted with.

Computing User Interests. As the user interacts with the content, the system aggregates the concepts identified in the content, weighting them according to the type of interaction as demonstrated on Figure 2. The aggregated concepts constitute the user interests profile.

Let's look into more details how the system can compute the user interest profiles at any point in time. We denote by n the number of users on the platform, by p - the number of possible interaction types, by m - the number of content items on the platform and by k - the number of concepts identified in the content. Then at any point in time the user interests profiles can be computed in the following way:

$$UC_{n*k} = \sum_{v=1}^{p} w_v * UA^v_{n*m} * DC_{m*k}, \quad (1)$$

where UC_{n*k} is the matrix of user concepts of interest, hence UC_{ij} is the relevance of the concepts c_j for the user u_i; w_v is the weight assigned to specific interaction type v indicating how strongly specific action of the user expresses her interest in the content; UA^v_{n*m} is the matrix capturing user-content interactions of type v, UA^v_{ij} is the number of times the user u_i has done interaction of type v with the content d_j; DC_{m*k} contains the concepts represented in the content so DC_{f*r} is the relevance of concept c_r to the content item d_f.

While the formula presented above is suitable for computing the profile first time when the recommender is deployed, the profile does not need to be recomputed from scratch and can be updated incrementally. On every user-content interaction, we update in real-time the user concepts of interest based on the ones that were found in the content as follows:

$$UC^{after}_{1*k} = UC^{before}_{1*k} + w_v * UA^v_{n*m} * DC_{m*1}, \quad (2)$$

where UC^{before}_{1*k} is the vector of user concepts before the interaction and UC^{after}_{1*k} - after the interaction; UA^v_{n*m} is a matrix having 1 in position (i, j) if the user u_i had interaction of type v with the content item d_j, all other elements are 0; and DC_{m*1} contains relevance values for the item concepts.

Once the profile constructed, in the next section we explain how it can be used for recommendations.

3.2 Recommending Relevant Content and Users

Connecting right people with right knowledge is a possible way to improve knowledge sharing. We aim to improve knowledge discovery by facilitating connection creation between knowledge sources and users in need of knowledge. Knowledge sources can be individual content items or other users with similar interests possessing the knowledge. We propose an approach that can suggest 1) content relevant to users and 2) users with similar interest. Below, we present two main steps of our approach.

Step 1. Computing term weights with TF-IDF. On the first step, we compute the relevance of specific terms (including concepts) for the content items by using a known information retrieval technique, namely term frequency - inverse document frequency (TF-IDF) as explained in [17]. When computing the weight, TF-IDF considers the frequency of the term inside of a document and its frequency in the whole corpus. In this way, for each content item we obtain a vector that contains weights of individual words or concepts cw_{ci} present in the content:

$$cw_{ci} = tf_{ci} * idf_{ci}, \quad (3)$$

where tf_{ci} is the term frequency representing how often the term ci appears in the document and idf_{ci} is the inverse document frequency indicating how common is the term ci in all documents.

Step 2. Scoring relevant items with cosine similarity. To obtain for the user u suggested content items or relevant users, we compute the relevance score for the item d using a cosine similarity between the two vectors representing the user and the content:

$$S(u, d) = \frac{V(u) \cdot V(d)}{|V(u)||V(d)|}, \quad (4)$$

where $V(u)$ and $V(d)$ are the vectors containing weights

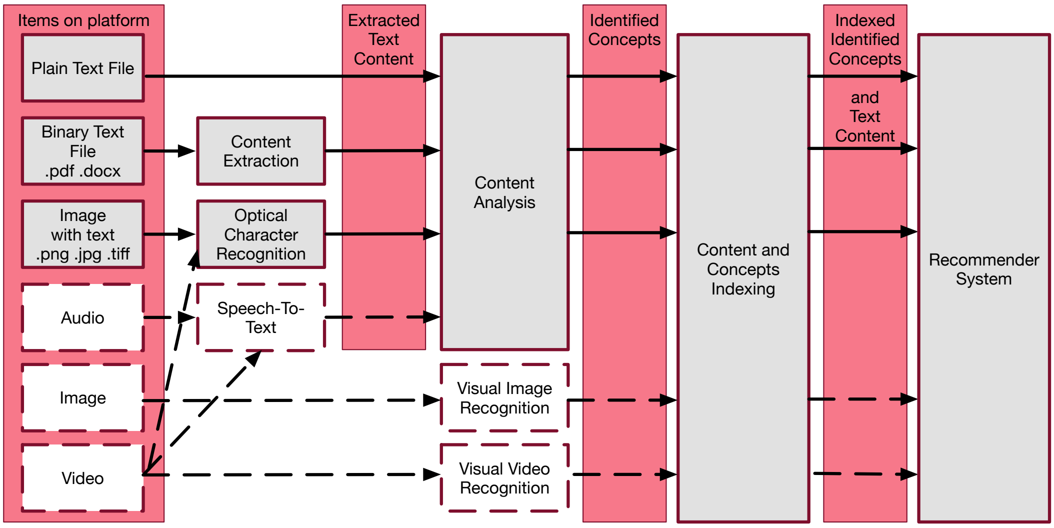

Figure 1: A possible pipeline architecture to extract concepts from diverse content types. Dotted lines mark the parts yet to be implemented in Graasp.

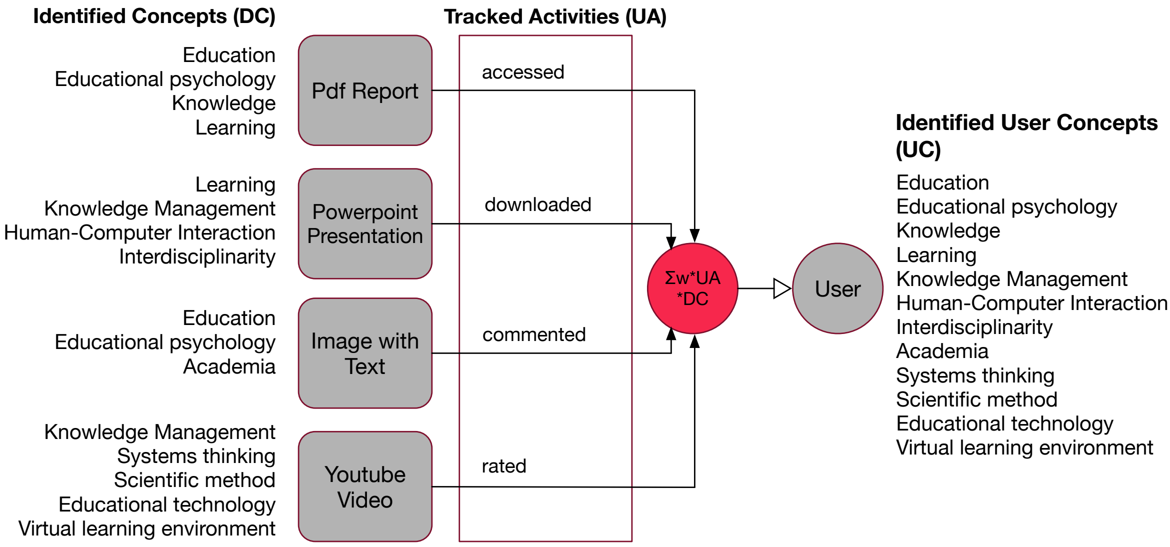

Figure 2: A schematic representation of the proposed approach. The system aggregates the concepts from the content as the user interacts with the content.

of the user terms and the document terms computed at Step 1; $V(u) \cdot V(d)$ is a scalar product of the two vectors; $|V(u)|$ and $|V(d)|$ are Euclidean norms of the vectors.

4. IMPLEMENTATION

To validate the feasibility of the approach and further evaluate it, we have implemented it in Graasp, a social media platform employed for knowledge management. Graasp supports uploading and storage of content from user devices or the cloud. Graasp provided extraction of text content from multiple file formats, and the activity logging infrastructure was already in place. Still, we needed to extend the platform to enable content analytics with concepts extraction, construction of the interests profile, and items recommendations with Elasticsearch[1]. Below, we explain the architecture of the implemented solution.

4.1 Concept Extraction and Activity Tracking

Concept Extraction. The concepts extraction is done as soon as content is uploaded to Graasp. To extract concepts, we have implemented a processing pipeline presented on Figure 1. On the first step, the type of the content is identified, and Graasp tries to extract textual information when possible. For plain text files, it just reads the text content of the file. For binary text files including pdfs and

[1]Elasticsearch Open Source Engine `https://github.com/elastic/elasticsearch`

Microsoft Office formats, we use the textract library[2]. For images, Graasp tries to perform Optical Character Recognition and read the text presented on the image using the tesseract[3] library. In the future, we foresee extracting text from Audio and Video files relying on Speech-To-Text technologies (shown with dotted lines on Figure 1) and obtaining concepts for images and videos with the help of visual recognition tools [13], for instance using clarifai[4]. Once the text is available, we analyse its content, identifying the concepts present there. For this purpose, we concatenate the item name, the item description and the extracted content and, at the moment of writing, employ AlchemyAPI[5] Concept Tagging to get the concepts. It is worth noting that our approach does not assume a specific concept identification technology, and AlchemyAPI was picked for the reasons of minimal administration and scalability. After the system identifies the concepts, it indexes them in Elasticsearch together with the text content extracted before.

Activity Tracking. Graasp uses ActivityStreams format for capturing user activities on the platform. Some of the actions that the platforms records include access, download, rating, commenting, inviting members and, searching.

4.2 Interests and Recommendations

Constructing Interests Profile. Graasp continuously updates interests profile of the users as they interact with the content. Users interests are displayed next to their profile information as demonstrated on Figure 3. The user can adjust her profile by removing individual concepts by pressing the X button and in this way influence in real-time the content and users suggested by the recommender.

Computing Recommendations. In Graasp, we rely on Elasticsearch for computing recommendations whenever the user wants to see them. Elasticsearch is built on the Lucene[6] text search engine that internally employs vector space model, TF-IDF, and cosine similarity when finding relevant items[7], similarly as in our proposed approach described in Section 3.2. We assemble into a single search query all of the concepts from the user interests profile and, whenever present, the terms from the user description as on Figure 3 (1). We run this query against the name, description, content and, concepts fields of the items, assigning different boost weights for matches happening in different fields. The obtained results are presented to the user next to her profile as illustrated on Figure 3.

5. EVALUATION

To understand opinions regarding the approach and its performance when put into practice, we have conducted a preliminary evaluation of the approach implementation in Graasp with pre-service teachers. This section explains in more details the methodology used and the main outcomes.

5.1 Methodology

We have conducted a survey-based preliminary evaluation of the developed approach. Surveys are one of the common ways of evaluating recommender systems allowing to collect opinions regarding the system from multiple users in a reasonable timeframe [7, 21]. Our goal was to validate if the approach, in general, is useful, if its implementation in Graasp is usable, as well as if the system can identify relevant interests and recommend relevant items.

Participants. We have conducted the survey with six participants of a workshop on inquiry-based learning for pre-service teachers in secondary education. During the workshop the participants registered in Graasp and carried out on the platform a set of activities during 2 hours. At the end of the session, we asked them to fill in the survey.

Survey Structure. Our survey had three parts[8]. The first part asked about general disposition towards the interests identification and the interests-based recommender. The second part was the System Usability Scale (SUS) [4] evaluating the usability of the implemented system. We have selected SUS because of its understood interpretation and robustness [1]. In the third part, we evaluated the quality of the identified interests and recommendations. Two types of questions formed the survey. The first type was questions to indicate the level of agreement with specific statements, where we followed the 5-point Likert scale ranging from 1 - Strongly Disagree to 5 - Strongly Agree to obtain quantitative results. The second type was open questions where we asked the responders to provide us with qualitative feedback regarding the approach and its implementation.

5.2 Results

In this section we focus on the main outcomes of the evaluation. Complete survey results are available online[9].

Approach. The users valued positively the idea of using their interests to guide the recommendations and to find other users with similar interests (mean Likert score $\mu = 3.17$ and $\mu = 3.33$ respectively). Besides, the fact of being aware of the inferred interests and the possibility of editing interests were well appreciated ($\mu = 3.17$ and $\mu = 3.33$ respectively). Although the quantitative analysis does not illustrate a high adoption by the users, during the workshop, the participants were keen on understanding how the interests were extracted and highlighted the novelty of the approach. Further details with the survey results may be found following the URL mentioned above.

Usability. In general, the participants were eager to use the recommender with a certain frequency ($\mu = 3.33$) and did not report major issues regarding complexity, inconsistency or difficulty of usage. Just one person considered that she would need technical support or previous background to use the recommender. According to the discussion with this person after the workshop, these answers were partially conditioned by the cognitive load due to the short time available to get used to the platform itself and to integrate all the ideas presented in the workshop. The quantitative results of the SUS questionnaire are also available on-line.

Accuracy. Despite the limited amount of traces collected due to the short time of the user interaction, the results point out that both the interests extracted and the recommendations, in general, were relevant ($\mu = 3.17$) and diverse ($\mu = 3.50$). It is noteworthy that when we asked the users to check how many relevant interests and recommendations

[2]textract `https://github.com/dbashford/textract`
[3]tesseract Library `https://github.com/tesseract-ocr`
[4]Clarifai `http://clarif.ai/`
[5]AlchemyAPI `http://www.alchemyapi.com`
[6]Apache Lucene `https://lucene.apache.org`
[7]Relevance Scoring `https://www.elastic.co/guide/en/elasticsearch/guide/current/scoring-theory.html`
[8]Recommender Evaluation Survey `https://goo.gl/Wes6uP`
[9]Evaluation results `https://goo.gl/Wes6uP`

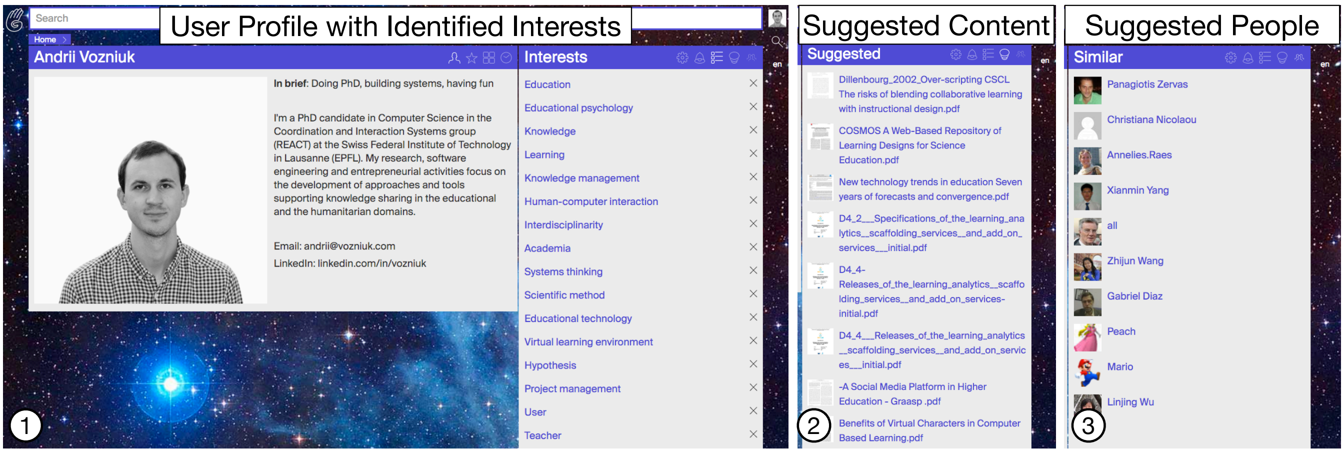

Figure 3: (1) User interests in Graasp as identified by our approach. Suggested content (2) and suggested people (3) based on the user profile information and identified interests.

appeared in the top 10, we discovered two groups. While most of the users reported more that six relevant items, two users got less than two relevant items. We have looked into this case and identified the reason covered below.

Sensitivity to Inaccurate Concepts. In the case when an item that the user interacted with many times has concepts identified not accurately, these concepts appear on the top of user interests. We plan to mitigate this problem in the future by introducing a heuristic for not considering concepts with low relevance and by limiting the influence of a single item on the overall concept relevancy for the user. Our goal is to make sure that the identified concepts come from many items rather than from many visits to a single item with potentially misidentified concepts. Our expectation is that it will allow reducing the impact of faults in concepts identification on the user resulting user profile.

Privacy Implications. Right now, only the user can see her interests, but we consider putting in place a mechanism that will allow to make validated interests visible to other users of the platform and to make it possible to find the user based on her interests as it was proposed in [15]. However, based on the evaluation, while some of the participants were eager to make their interests visible, others were reluctant. Thus, it will be necessary to allow users configure the visibility of their interests to preserve their privacy, following the recommendations provided in codes of practice for learning analytics [20].

6. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a new approach to building user interests profile based on 1) content analytics providing the system with the concepts present in the content and 2) activity tracking allowing the system to know how the user interacted with the content. We have used the extracted interest concepts to recommend relevant content and people. Further on, we have implemented the proposed approach in Graasp, a knowledge management system. Graasp was used in a workshop to support teachers when building inquiry learning spaces for their students. Thus, we have evaluated the approach with the teachers, and the evaluation has demonstrated that the proposed approach can identify relevant user interests and recommend relevant content based on the identified interests. At the same time, the evaluation has unveiled sensitivity of the approach to inaccurately identified concepts that we plan to overcome in the future. While we draw our experience and motivation from the educational context, our contributions have a broad impact and can be applied for content repositories, where it is possible to obtain content analytics and track activities performed by the users (e.g., Google Drive and Dropbox).

Looking Outside. In this study, we analyzed the content and recorded the activities limited to the scope of the content repository. However, in the current technological landscape, the interactions are getting more distributed often spanning across multiple platforms. Studies suggest that combining data obtained from several platforms could allow creating a more accurate user interests profile [8]. In the future, we plan to extend the architecture of Graasp to aggregate the content and the interactions outside of the system.

Incorporating Relevance Scores. At the moment, when computing the similarity score for the recommended items we consider the fact of the concept presence but do not take into account the available concept relevance scores. Incorporating the relevance scores available for user interest concepts and content concepts when computing the user-content relevance score may lead to more relevant recommendations since it will promote the results with similar highly relevant concepts.

Recommender Adaptability. One potential downside of our approach could be related to its limited ability to react timely to change in the user interests reflected in her interactions. This happens since the concepts the user accumulated at some point through her interaction history maintain the same score indefinitely. One of the possible solutions to this problem is to introduce the forgetting function as suggested in [22], so that as time goes the concepts that are not encountered anymore get their relevance score reduced.

Substantial Evaluation. This paper presented a preliminary evaluation of the recommender to provide early feedback. We are planning to conduct an evaluation with more users that used Graasp for longer periods of time so that more activity traces are available. We also expect these users to have established expectations regarding their interests when interacting with the platform.

7. ACKNOWLEDGMENTS

This research is partially funded by the European Union in the context of the Go-Lab project (Grant Agreement no. 317601) under the Information and Communication Technologies (ICT) theme of the 7th Framework Programme for R&D (FP7). This document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

8. REFERENCES

[1] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *Int. J. Hum. Comput. Interact.*, 24(6):574–594, 2008.

[2] E. Bogdanov, F. Limpens, N. Li, S. El Helou, C. Salzmann, and D. Gillet. A social media platform in higher education. In *Proceedings of the Global Engineering Education Conference*, EDUCON, pages 1–8, Apr. 2012.

[3] I. Bosnić, K. Verbert, and E. Duval. Automatic keywords extraction - a basis for content recommendation. In *Proceedings of the 4th International Workshop on Search and Exchange of e-learning Materials*, pages 51–60, Sept. 2010.

[4] J. Brooke. SUS: A quick and dirty usability scale. In P. W. Jordan, B. Weerdmeester, A. Thomas, and I. L. Mclelland, editors, *Usability evaluation in industry*, pages 189–194. Taylor and Francis, 1996.

[5] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis. Panorama of recommender systems to support learning. In *Recommender Systems Handbook*, pages 421–451. Springer US, 2015.

[6] S. El Helou, C. Salzmann, and D. Gillet. The 3A personalized, contextual and relation-based recommender system. *J. Univers. Comput. Sci.*, 16(16):2179–2195, 2010.

[7] M. Erdt, A. Fernandez, and C. Rensing. Evaluating recommender systems for technology enhanced learning: A quantitative survey. *IEEE Trans. Learn. Technol.*, 8(4):326–344, Oct. 2015.

[8] I. Guy, U. Avraham, D. Carmel, S. Ur, M. Jacovi, and I. Ronen. Mining expertise and interests from social media. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW, pages 515–526, New York, NY, USA, 2013. ACM.

[9] H. Harkous, R. Rahman, B. Karlas, and K. Aberer. The Curious Case of the PDF Converter that Likes Mozart: Dissecting and Mitigating the Privacy Risk of Personal Cloud Apps. In *Proceedings of the 16th Privacy Enhancing Technologies Symposium*, PETS, Darmstadt, Germany, 2016.

[10] J. Kim, S.-W. Li, C. J. Cai, K. Z. Gajos, and R. C. Miller. Leveraging video interaction data and content analysis to improve video learning. In *Proceedings of the CHI 2014 Learning Innovation at Scale workshop*, 2014.

[11] K. Kitto, S. Cross, Z. Waters, and M. Lupton. Learning analytics beyond the LMS: the connected learning analytics toolkit. In *Proceedings of the 5th International Conference on Learning Analytics And Knowledge*, LAK, pages 11–15, New York, NY, USA, March 2015. ACM.

[12] V. Kovanović, S. Joksimović, D. Gašević, M. Hatala, and G. Siemens. Content analytics: the definition, scope, and an overview of published research. *Handbook of Learning Analyitcs*, 2015.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[14] M. G. Moore. Editorial: Three types of interaction. *The American Journal of Distance Education*, 3(2):1–6, 1989.

[15] W. Pohs, G. Pinder, C. Dougherty, and M. White. The lotus knowledge discovery system: Tools and experiences. *IBM Syst. J.*, 40(4):956–966, 2001.

[16] Ralf Klamma, Mohamed Amine Chatti, Erik Duval, Hans Hummel, Ebba Thora Hvannberg, Milos Kravcik, Effie Law, Ambjörn Naeve, and Peter Scott. Social software for life-long learning. *Journal of Educational Technology & Society*, 10(3):72–83, 2007.

[17] J. Ramos. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, pages 1–4. cs.rutgers.edu, 2003.

[18] M. J. Rodríguez-Triana, L. P. Prieto, A. Vozniuk, M. Shirvani Boroujeni, B. A. Schwendimann, A. C. Holzer, and D. Gillet. Monitoring, Awareness and Reflection in Blended Technology Enhanced Learning: a Systematic Review. *International Journal of Technology Enhanced Learning*, In press.

[19] J. L. Santos, K. Verbert, J. Klerkx, E. Duval, S. Charleer, and S. Ternier. Tracking Data in Open Learning Environments. *J. Univers. Comput. Sci.*, 21(7):976–996, 2015.

[20] N. Sclater. Code of practice for learning analytics. a literature review of the ethical and legal issues. Technical report, JISC, 2014.

[21] G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender Systems Handbook*, pages 257–297. Springer US, 2011.

[22] H. Thüs, M. A. Chatti, R. Brandt, and U. Schroeder. Evolution of interests in the learning context data model. In *Design for Teaching and Learning in a Networked World*, pages 479–484. Springer, 2015.

[23] A. Vozniuk, S. Govaerts, and D. Gillet. Towards portable learning analytics dashboards. In *Proceedings of the 13th International Conference on Advanced Learning Technologies*, ICALT, pages 412–416, Beijing, China, July 2013. IEEE.

[24] A. Vozniuk, A. Holzer, S. Govaerts, J. Mazuze, and D. Gillet. Graspeo: a social media platform for knowledge management in NGOs. In *Proceedings of the 7th International Conference on Information and Communication Technologies and Development*, page 63, Singapore, Singapore, 15 May 2015. ACM.

[25] V. A. R. Zaldivar, R. M. Crespo García, B. Daniel, P. Abelardo, and others. Automatic discovery of complementary learning resources. In *Proceedings of the 6th European Conference of Technology Enhanced Learning*, EC-TEL, pages 327–340, Palermo, Italy, 2011.