






SALMON: Sharing, Annotating and Linking Learning Materials Online

Farbod Aprin[✉], Sven Manske, and H. Ulrich Hoppe

COLLIDE Research Group, University of Duisburg-Essen,
Lotharstr. 63, 47045 Duisburg, Germany
farbod.aprin@stud.uni-due.de,
{manske, hoppe}@collide.info

Abstract. In consideration of the growing availability of mobile devices for students, web-based and shared annotations of learning materials are becoming more popular. Annotating learning material is a method to promote engagement, understanding, and independence for all learners in a shared environment. Open educational resources have the potential to add valuable information and close the gap between learning materials by automatically linking them. However, current popular web-based text annotation tools for learners, such as Hypothesis and Diigo, do not support learners in discovering new learning resources based on the context, metadata and the content of the annotated resource. In this article, we present SALMON, a collaborative web-based annotation system, which dynamically links and recommends learning resources based on annotations, content and metadata. It facilitates methods of semantic analysis in order to automatically extract relevant content from lecture materials in the form of PDF web documents. SALMON categorizes documents automatically in a way that finding similar resources becomes faster for the learners and they can discover communities for interesting topics.

Keywords: Open educational resources · CSCL · Recommender system · Annotation · PDF annotator · Semantic analysis · Web-annotation · Education · OER · TEL

1 Introduction

The digital annotation of learning materials is a common activity in the process of self-regulated learning and has a long tradition in the field of technology-enhanced education. Collaborative annotation systems like *Hypothesis* allow learners to share their annotations and create annotations of web resources together. Such annotations can encompass explanations on text, excerpts, or individual highlights. Incorporating digital technologies can be beneficial when it comes to the (automatic) enrichment of such annotations. Open Educational Resources (OER) have the potential to add valuable information and to connect learning materials by automatically linking them. With an increased level of digital technology in education, quick access to information and learning resources has become more relevant. To access a large number of digital texts, the learners need robust information systems and search engines to find appropriate

resources. Today the collection of OER found on YouTube, Vimeo or wikis has become an inevitable part of information seeking procedure for learners to enrich their knowledge in a specific area. Without the need to search actively for related OERs, recommendations based on an automatic tagging of the respective learning materials might be supportive in this context.

The Semantic Web (SW) can be considered in designing learning tools. SW is an extension of the current interactive web, which means: “Information is given well-defined meaning, better enabling computers and people to work in cooperation” [1]. Web 3.0 technologies will assist online learning instructors in the areas of course development, learner support, assessment, and record keeping [2]. The outcome of such a shared environment could be distributing knowledge among learners with different expertise level. The critical aspects behind achieving the SW vision are enabling metadata and the mapping of it onto learning resources.

In this article, SALMON has been presented, a collaborative web application to share, annotate and automatically link learning materials with each other and with external OERs. Tags, metadata and annotations are extracted automatically from learning materials. Links between resources are based on semantic similarity of the aspects of the materials. This linking is facilitated to recommend OERs to learners in order to expand their collections of learning materials.

2 Related Work

2.1 Generating Metadata in Collaboration Learning Systems

The information extraction process is observed as the method of identifying essential parts of the learning document and content of the annotation [3]. Knowledge about documents has traditionally been accomplished through the use of metadata that can involve the world around the document like a cloud. A learning object is an entity such as digital or non-digital object that may be used for learning, education or training [4]. Metadata is used to describe the learning object and make it possible to select and integrate relevant learning experiences from a collection of learning objectives. This will provide reusable learning objects permitting in the lessons to be generated and customised for specific groups [5]. Learning Object Metadata (LOM) is a standard to integrate educational metadata into learning objects. Such metadata makes learning objects shareable and accessible through indexing and a description of the specific context it occurs. Cardinaels, Michael, and Duval presented a method to automatically extract metadata and transform it into LOM [6, 7]. The automatically generated metadata can be approved by filtering and observation to retrieve more relevant learning objects. Open-Calais and CiceroLite are two tools that enable API to generate semantic tags from the extracted text. This metadata can be utilised for enriching the LOM [16].

Collaborative tagging is the process wherein learners work on the shared resource and assign keywords to it. If it were possible to generate tags to aid learners also enable them to have feedback on the shared tags, then we would have a common understanding of target learning material. The integration of collaborative e-learning systems and SW, which includes application and platform for social and collaborative exchange is increasing the quality of learning experience [8].

2.2 Learning Resource Recommendation

Open educational resources can be seen as an educational paradigm based on a simple but powerful idea that the knowledge of the world is a public good. OER provides a great possibility for learners to share, use, and reuse knowledge [9]. Integration of this idea to the daily learning procedure would add valuable free accessible information for learners to enhance their knowledge. A recommender system in a learning context is an information system examining what a learner is doing and tries to support actions [10]. Popular search engines are currently quite powerful regarding retrieval OER, but they still return the huge instructional list of information; in that, they can be easier by embedding recommender systems in the learning application.

Based on Almudena Ruiz-Iniesta research, integration of OER resources into educational tool helps students find resources faster. It also helps to retrieve the resources that matched with their interest and knowledge. According to evaluations, users are interested in using tools with OER repositories in different domains and mostly had positive feedback. Besides, the supplementary qualitative analysis evaluation showed promising improvements in learning performance and a considerable decrease in dropout rates of the students using OER in learning environment [11].

2.3 Related Collaborative Annotation Tools

A digital annotation learning tool provides an explanation, comment or feedback that is added to a text or diagram on the learning material. The semantic web envisions that technologies can make it possible to generate a kind of “intelligent” documents that were imagined almost twenty-five years ago [12]. There are specific tools that can generate semantic digital annotation automatically or manually to assist users.

One user-friendly interface tool is Mangrove system from the University of Washington that supports automated tagging of HTML documents. The first requirement is the facility of authoring. MANGROVE presents a graphical webpage annotation tool allowing users to immediately and accumulatively annotate the HTML content [13].

Hypothesis¹ is a web-based annotation tool where the users annotate a web resource (as URL) collaboratively in a textual representation [14]. It provides a second layer on the web to enable teachers and students to highlight and annotate on web-based links. Diigo² is a web-based annotation tool that lets students bookmark and tag web pages. It would also be able to highlight a part of a webpage like *Hypothesis*, and for more emphasis attaching the sticky notes to the text is considered.

The Open Annotation and Tagging System (OATS) is a multi-purpose annotation tool that can be integrated into any e-learning system. It will allow the students to tag and annotate HTML based on learning materials. Also, annotation in OATS consists of notes and tag categories on a highlighted section in a selected learning material [15]. Contrary to other systems, OATS provided retrieval functionalities. The user can search for pages, notes and tags among the index of the whole OATS system.

¹ Hypothesis website, hypothesis project (2013), <https://web.hypothes.is/>, retrieved:2019-6-25.

² Diigo Blog, <https://blog.diigo.com/2014/09/03/annotating-PDF-docs-with-diigo-a-tutorial/>.

All the mentioned tools focus on collaboration on the textual annotation tool. They are not integrating OER by recommendation. Besides, they are unable to provide an implementation for automatic metadata extraction and tagging.

3 SALMON Approach

The SALMON web application offers a collaborative textual annotation environment, which makes use of semantic methods for linking and recommending and learning open educational resources. In SALMON, collaboration is through the artefact. Learners can work together in a group around the learning material, for example, PDF slides of a lecture at university. The PDF format can be seen as the de facto standard for delivering digital lecture materials. Thus, the web-based environment consists of a PDF document viewer in the annotation component. On this material, learners perform different activities in a collaborative space, supports them in creating mutual knowledge. Such activities encompass tagging, immediate feedback on automatically generated semantic tags, highlighting and notes on the text in public or anonymous mode. The initial document of the learning material used that serves as a seed for the recommendation and discovery of new resources is called Pdf-core in this work. Besides, the learners can browse different topics of interest, vote on the recommended OER links to provide relevant feedback and watch multimedia formats in annotation environments. The application generates metadata based on the semantic tags extracted from the content and the from the interaction data obtained from the learners. Consequently, it categorises learning material into different topics by calculating the cosine similarity between semantic metadata.

3.1 Scenario in SALMON

As a scenario, a student called Sara participating in application desires to annotate collaboratively with other students in the learning document environment online. The teacher added the document link with virtual reality topic to the application. After registration via email, Sara and others can access the Pdf-core environment by scanning a QR-code or by entering directly via URL. They can highlight text or capture screenshot of the document and annotate it. Sara can post under her username or as anonymous. This note will be shown in the sidebar. Adding the new Pdf-core is possible, but the content of the new material must be unique; otherwise, Sara will get a warning that the target resource is added before. Furthermore, the application automatically redirects her to the existing collaboration Pdf-core environment. Once she adds new unique Pdf-core, the system searches in local Pdf-core collections to find a match according to similarity function and assigns the new link to an associated collection. If it does not match with any current stored collection, then the system will generate a new collection with the title of the highest rank tag. Recommending relevant OER and Pdf-cores from other remote or local collections in SALMON could assist Sara and her friends to find their answers more quickly and enable them to find other communities for other related Pdf-cores in the same collection. Sara can pin a visual

card to save it in her own private collection. These personalisation features prevent repeating the information-seeking procedure. The annotation will be stored and usable for the next semester students (Figs. 1 and 2).

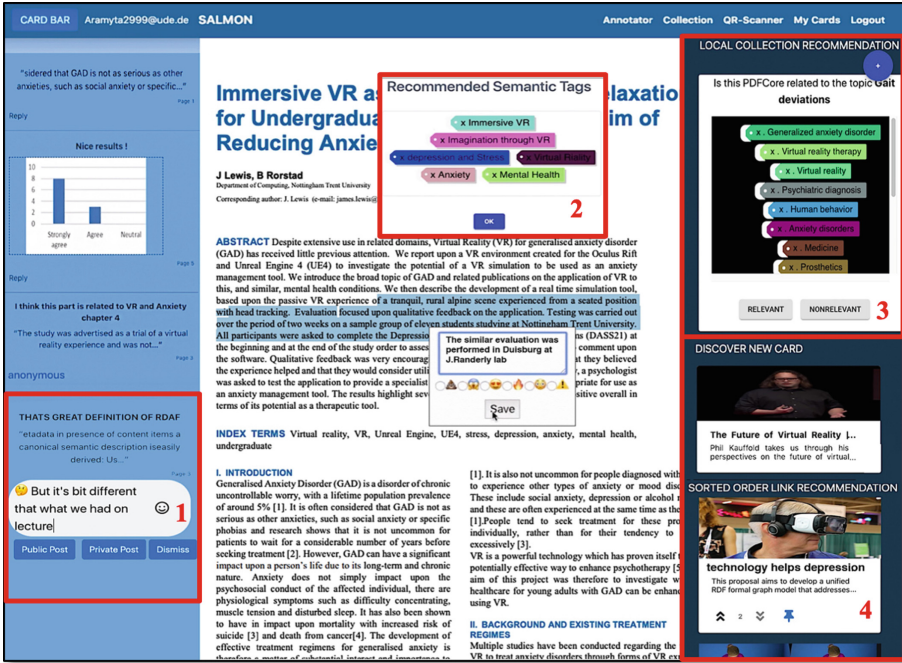


Fig. 1. Annotating PDFs: learners can add a new PDF file and annotate it. 1. Comment part: learners can ask questions or reply to it. 2. Automatically generated tags based on the semantic text analysis, learners can add or remove tags. 3. Recommended similar collection in SALMON. 4. Recommended Open Educational Resources.

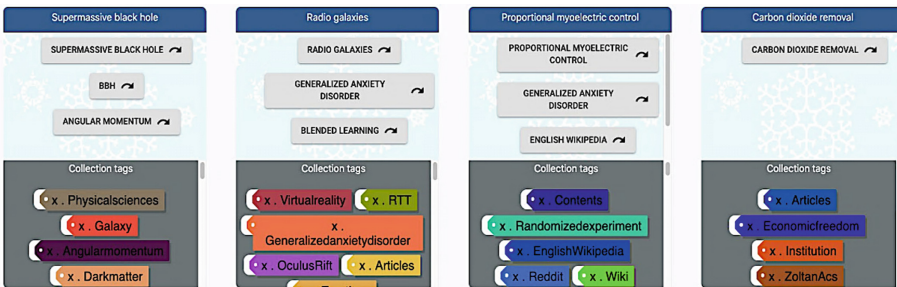


Fig. 2. A collection is a set of similar Pdf-cores (cosine similarity) and contains the union of their tags. Users can find a specific collection by querying the system and navigating through topics.

3.2 SALMON Life Cycle

SALMON is designed to act dynamically by modifying the metadata when the data is changed by each API service provider or by user interaction. The application receives the input data from learning material content, user activities, and feedback in different nonlinear steps. Afterwards, it filters and stores input data as metadata for the added Pdf-core. This data is needed for generating procedure of semantic tags. With the usage of extracted data, the system utilises external knowledge extraction services engines via the application user interface (API). We used Open-Calais³ to generate semantic tags from extracted filtered text and then added that to enrich metadata [16]. These keywords are exploited to index the OERs in the domain ontology of Pdf-cores and are exploited for retrieval and classification purposes.

SALMON also obtains metadata from a particular part of the Pdf-core content and the keywords that have been set previously to the PDF file by the author. This metadata can be enriched through lifecycle. The current version of SALMON determines similarity by utilising the *DKPro* Similarity package for estimating the similarity between two sets of metadata text [17]. “The Semantic Web offers learners the possibility of having a wealth of related content delivered to their desktop without explicitly identifying or requesting it” [18].

Morris indicates that semantic web-based learning software system agents can provide instructors new information relevant to expertise area and professional interests [2]. A Semantic search will return multimedia rather than just a list of URLs. A smart agent can return relevant blogs and multimedia about the topic to the user [19].

Brindley stated, “Quality learning environments include opportunities for students to engage in interactive and collaborative activities with their peers” [20]. As illustrated, visual cards can be rated by the learners and pinned; pinned cards will be kept in layout and stored in learner personal collection. Semantic web technologies are used to provide contextualised feedback to instructors about learner personalisation [15].

4 SALMON Architecture

SALMON architecture follows the variant of the service-oriented architecture (SOA). It introduces a modular architecture that has the benefit of decomposing application into different smaller services besides improving modularity. These microservices (MS) are compiling and running independently with various application servers. “In short, the MS architectural style is an approach to developing a single application as a suite of small services, often an HTTP resource API⁴.” One benefit of utilising microservice architectures is that if one of the services is disrupted even for a short period, other services will continue their functioning independently.

Each microservice can be tested and run independently. SALMON has three main modules that include two services: SALMON-API microservice (MS), SALMON

³ Semantic knowledge extractor, Thomson Reuters (2018), <http://www.opencalais.com/opencalais-api/>, retrieved:2019-6-25.

⁴ Microservice architecture, Martin Fowler (2012), <https://martinfowler.com/>, retrieved: 2019-6-25.

backend MS and SALMON frontend module (Fig. 3). The user interacts with a ReactJs application and can add new data or modify the system status via environment interface. For instance, as the user adds new material and desire to annotate it, the state in the SALMON-Front will be changed. Besides, SALMON-Front talks to SALMON-Backend MS via REST-API to post the PDF link to the Backend MS⁵.

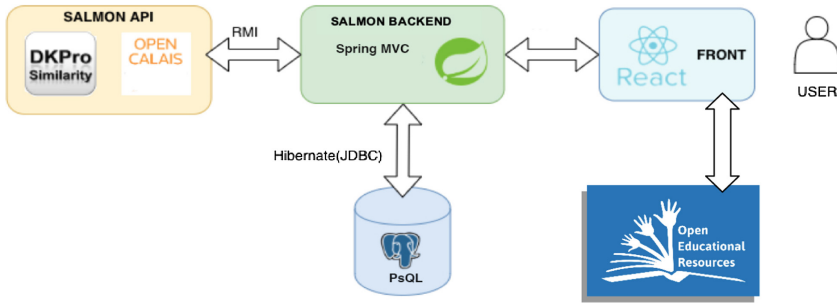


Fig. 3. SALMON architecture consists of three main modules: SALMON API microservices, SALMON backend microservices and SALMON frontend module.

5 Discussion and Conclusion

In this work, we have presented SALMON, a collaborative web-based text annotation tool with recommender systems. The system extracts the metadata for added learning material, which can add the valuable potential for integrating the OER and similar local resources, as a recommendation for learners.

In the SALMON environment, we have considered the nonlinear procedure to enrich the learning material metadata which can be updated, according to learner interaction and feedback and updating resources. This feature can increase the level of relevancy in the recommended resources for the learners. To find similarity for the local document recourses, cosine-similarity which is a useful method, was used to approximate the similarity between metadata for materials. By finding the similarity of metadata, we demonstrate categorizing learning materials into different associated collections. This feature will facilitate the procedure of discovering similar documents for learners. Also, classification helps to find other peers in the interested topic easier.

For future work, the application will be evaluated empirically in the context of a university course. In order to support internationalisation and multiple languages, future versions of SALMON will be connected to different text analysers like FRED⁶ to increase the preciseness of the extracted keyword and for using it for extracting more different languages, e.g., Russian and German. We will update the interaction features like an instant recommendation for highlighted text in the annotation tool.

⁵ SALMON GitHub repository, 02-06-2019, <https://github.com/SALMON2PROJECT>.

⁶ Keyword extraction Framework, Machine Reading for the Semantic Web, STlab 2015 <http://wit.istc.cnr.it/stlab-tools/fred/>.

References

1. Berners-lee, T., Hendler, J., Lassila, O.: The Semantic web a new form of web content that is meaningful to computers will unleash a revolution of new possibilities (2001)
2. Morris, R.D.: Web 3.0: implications for online learning. *TechTrends* **55**, 42–46 (2011). <https://doi.org/10.1007/s11528-011-0469-9>
3. Bittencourt, I.I.M., Costa, E., Isotani, S., Mizoguchi, R.: Towards a reference model to semantic web-based educational systems (2008)
4. Barker, P., Campbell, L.M.: Metadata for learning materials: an overview. *Technol. Instr. Cogn. Learn.* **7**, 225–243 (2010)
5. Baker, T.: Library Hi Tech Libraries, languages of description, and linked data: a Dublin Core perspective Article information. <https://doi.org/10.1108/07378831211213256>
6. Barker, P.: What is IEEE learning object metadata/IMS learning resource metadata? *Cetis Stand. Briefings Ser.* **1**, 4 (2005)
7. Learning Technology Standards Committee, IEEE Computer Society: IEEE Standard for Learning Technology — Extensible Markup Language (XML) Schema Definition. Institute of Electrical and Electronics Engineers (2005)
8. Torniai, C., Jovanović, J., Gašević, D., Bateman, S., Hatala, M.: E-learning meets the Social Semantic Web. In: Proceedings of 8th IEEE International Conference on Advanced Learning Technologies, ICAALT 2008, pp. 389–393 (2008). <https://doi.org/10.1109/ICALT.2008.20>
9. Deimann, M., Bastiaens, T.: Special session OER: integrating OER and instructional design – towards a more holistic view keywords: open educational resources, instructional design abstract: 1 introduction 2 the claim of open educational resources (OER). *Learning* **1**, 1–10 (2007)
10. Schafer, J.B.: The application of data-mining to recommender systems. In: *Encyclopedia of Data Warehousing and Mining*, Second Edition, pp. 45–50 (2011). <https://doi.org/10.4018/978-1-60566-010-3.ch008>
11. Ruiz-Iniesta, A., Jiménez-Díaz, G., Gomez-Albarran, M.: A semantically enriched context-aware OER recommendation strategy and its application to a computer science OER repository. *IEEE Trans. Educ.* **57**, 255–260 (2014)
12. Delphi Group: The document process (1994). <https://bit.ly/2MgNJYK>
13. McDowell, L., et al.: Mangrove: enticing ordinary people onto the semantic web via instant gratification, pp. 754–770 (2010). https://doi.org/10.1007/978-3-540-39718-2_48
14. Te Whaley, D.: Annotation is now a web standard. *hypothesis Blog* (2017). <https://web.hypothes.is/blog/annotation-is-now-a-web-standard/>. Accessed 12 June 2017
15. Brooks, C., Bateman, S., Greer, J., Mccalla, G.: Lessons Learned Using Social and Semantic Web Technologies for E-Learning (2009). <https://doi.org/10.3233/978-1-60750-062-9-260>
16. Gangemi, A.: A comparison of knowledge extraction tools for the semantic web. In: Cimiano, P., Corcho, O., Presutti, V., Hollink, L., Rudolph, S. (eds.) *ESWC 2013*. LNCS, vol. 7882, pp. 351–366. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38288-8_24
17. Bär, D., Zesch, T., Gurevych, I.: DKPro Similarity: An Open Source Framework for Text Similarity. (2013). www.aclweb.org/anthology/P13-4021
18. Daly, C.: The semantic web and e-Learning. *eLearn.* (2009). <https://doi.org/10.1145/1595384.1555528>
19. Ohler, J.: The semantic web in education: what happens when the read-write web gets smart enough to help us organize and evaluate the information it provides?, pp. 7–9 (2008)
20. Brindley, J., Blaschke, L.M., Walti, C.: Creating effective collaborative learning groups in an online environment. *Int. Rev. Res. Open Distrib. Learn.* **10**, 1–18 (2016). <https://doi.org/10.19173/irrodl.v10i3.675>