Ontology Coverage Tool and Document Browser for Learning Material Exploration

Christian Grévisse, Jeff Meder, Jean Botev and Steffen Rothkugel
Computer Science and Communications Research Unit
University of Luxembourg
6, Avenue de la Fonte L-4364 Esch-sur-Alzette, Luxembourg
Email: christian.grevisse@uni.lu, jeff.meder.001@student.uni.lu, jean.botev@uni.lu, steffen.rothkugel@uni.lu

***Abstract*—Document collections in e-learning can cause issues to both learners and teachers. On one hand, inquiry from the vast corpus of available resources is non-trivial without adequate formulation support and semantic information. Implicit links between documents are hardly understood without a proper visualization. On the other hand, it is difficult for teachers to keep track of the topics covered by a large collection. In this paper, we present an ontology coverage tool and document browser for learning material exploration. Both learners and teachers can benefit from a visualization of an ontology and the documents related to the comprised concepts, overcoming the limitations of traditional file explorers. Guiding users through a visual query process, learners can quickly pinpoint relevant learning material. The visualization, which has been implemented as a web application using the D3.js JavaScript library, can be integrated into different e-learning applications to further enhance the workflow of learners. Finally, teachers are provided an overview of topic coverage within the collection.**

***Index Terms*—Document Collection Visualization; Learning Material; Visual Query Support; Ontology Coverage; Document Browser.**

I. Introduction

The ever-growing amount of learning material can be an issue for both learners and teachers. Students can become overwhelmed by the vast corpus of available resources on a given set of topics. In addition, Meno's paradox of inquiry aggravates the problem: If a learner does not know what he is looking for, a successful search for relevant documents is unlikely. Furthermore, a set of learning resources, like any other document collection, can contain both explicit and implicit links [1]. Explicit links between documents can be provided in the form of cross-references or hyperlinks, whereas implicit links are realized by documents covering the same set of topics. File management on current operating systems is mostly based on hierarchical file systems. Traditional file browsers are unable to present to the user implicit links between documents [1]. Also, users need to remember file names and their location to be able to retrieve them. If the precise content is unknown, this is nearly impossible [2]. On the other hand, graph-based hypermedia systems like the World Wide Web (WWW) excel at handling explicit links, but require sophisticated search engines to properly discover implicit links between a user query and a document collection. However, formulating a successful query is hampered both by the mentioned paradox of inquiry and the ambiguity of natural language. Latter issue can be overcome by annotating learning material with semantic web technologies, such as ontologies. Even the use of concepts from lightweight ontologies can lower the risk of ambiguity [3].

From a teacher's point of view, learning material provided in a course ideally covers all the topics belonging to the learning outcomes. Depending on the type of subject and study domain, some aspects might be presented in the lecture, while other topics are covered in a practical exercise session or a project. In any case, the established list of concepts required to know at the end of a course should all have been covered in one way or another by the provided resources. However, depending on the length of the course period, the complexity of a subject or the granularity of a study domain, it may quickly become difficult to keep track of the already covered and still to be covered topics.

In this paper, we present an ontology coverage tool and document browser for learning material exploration. Both learners and teachers can benefit from a visualization of an ontology and the documents related to the comprised concepts. Implicit links between concepts or documents can be grasped easier in such a visualization than in a traditional file system. In addition, query formulation is fostered through a visual browsing process. Integrating this visualization into different e-learning contexts, such as programming environments, authoring tools or serious games, can lower the risk of the inquiry paradox. Learners are enabled to quickly pinpoint resources relevant to their ad-hoc study necessities from the vast corpus, which may reduce the information overload and improve their work flow. Teachers, on the other hand, are given an overview of which ontology concepts have already been covered by the provided learning material and which still need to be covered. Our approach has been implemented as a web application using the D3.js JavaScript library.

The remainder of this paper is organized as follows: in section II, we discuss related work from different research dimensions. We present our tool in section III. Various ways of integrating the tool into different e-learning contexts are discussed in section IV. We conclude in section V, providing some ideas for future research.

II. Related Work

In this section, we will review related work on ontology visualization, document (collection) visualization, ontology-based retrieval and visual query formulation support.

A. Ontology Visualization

Visualizing ontologies as a graph of interlinked concepts is common practice. The well-known ontology editor *Protégé*[1] provides the *OntoGraf* plugin, which helps the creator of an ontology to get an overview of existing classes, properties and entities. In general, graph visualizations of ontologies are often rendered in a force-directed or hierarchical layout [4]. Several visualizations for ontologies have been proposed recently, including node-link diagrams, treemaps, or UML class diagrams. Lohmann et al. propose *Visual Notation for OWL Ontologies (VOWL)*, a visual language for ontology representation [5]. The specification defines graphical primitives, color schemes and visual elements. VOWL has been implemented as a Protégé plugin and a standalone Web tool. Psyllidis stated that most ontology visualizations are desktop-based and require a certain set of skills from the user [6]. The author proposes *OSMoSys*, a Web interface enabling a graph-based visualization of RDF data and ontology browsing. An experimental prototype was deployed for smart cities applications to exemplify its potential.

B. Document (Collection) Visualization

The objective of a proper visualization of a document collection should be to give a holistic view of the collection [1]. Such visualizations have been proposed in different domains. Padia et al. present different strategies to visualize digital collections at Archive-It, a subscription service from the Internet Archive [7]. Xu et al. suggest an interactive visual analytics application to help archivists analyze large-scale digital collections [8]. Bergström and Atkinson state that web-based digital libraries fail to show the context in which a scientific paper is situated [9]. The authors propose *PaperCube*, a set of visualizations to augment the exploration capabilities of digital libraries. Windhager et al. perform a review of visualization approaches for digital cultural heritage collections [10]. Another broad overview of current research on document (collection) visualization is provided by Gan et al. [11]. There exist also commercial document management systems and personal information managers such as *DEVONthink*[2] or *Keep It*[3].

C. Ontology-based Retrieval & Visual Query Formulation Support

Kim et al. propose *Ontalk*, an ontology-based document management and retrieval tool [2]. The tool generates metadata for a resource in a semi-automatic way and provides an ontology-based search engine. Zhuhadar et al. describe a visual ontology-based information retrieval system for the *HyperManyMedia* platform [12]. Clicking on a node in the graph representation of their ontology is considered as a query and updates the search results. Alhenshiri et al. state that the search relevance may be improved by integrating the user in the process of query reformulation [13]. The authors present an interactive Visual Search Engine (VSE) which visualizes the query reformulation and results presentation. Soylu et al. review different strategies of ontology-based end-user visual query formulation [14]. Munir and Anjum emphasize that existing database visualization tools do not help in semantic data retrieval [15]. Instead, ontology-based interactive query formulation systems can retrieve data from databases in a visual and semantically enhanced way, easing the human-computer communication.

[1]https://protege.stanford.edu/

[2]https://www.devontechnologies.com/products/devonthink/

[3]https://reinventedsoftware.com/keepit/

III. Ontology Coverage Tool & Document Browser

As indicated in the previous section, our work is related to the disciplines of ontology visualization, document (collection) visualization, ontology-based retrieval and visual query formulation support. However, previous work did not approach the challenges particular to the field of e-learning. Foremost, students shall be enabled to quickly pinpoint documents relevant to their studying process, without being hampered by Meno's paradox of inquiry. Teachers, on the other side, shall be given an overview of which topics have or have not been covered in the provided collection of learning resources.

In this section, we will present our tool which allows to browse through both the ontology and the learning material collection while giving an idea which topics are well or less covered. We assume that the comprised documents have been annotated and indexed with well-defined semantic concepts. In the upcoming example, these resources come from the domain of computer programming and are annotated with concepts from the ALMA ontology, a lightweight modular ontology for programming education [16]. These annotations could either have been done in a manual way, or through a semi-automatic entity linking process [17].

The tool has been implemented in form of a web application which makes use of *D3.js*[4], a JavaScript library for data-driven documents, in order to visualize the data retrieved from the ALMA ontology via API calls. The tool transposes an ontology scheme into a force-directed graph where every concept and its corresponding label is represented by a circle respectively a text SVG element. The user can interact with the visualization by clicking on the different concept circles to navigate through the ontology scheme so that he or she can retrieve the desired resources which are represented by rectangular SVG elements and related to one or more concepts. Moreover, visual indicators like distinct coloring of the circles or icons attached to concepts are meant to ease this navigation process.

[4]https://d3js.org

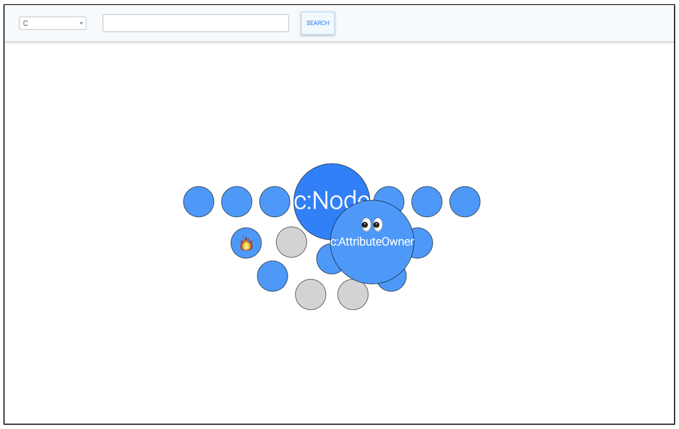

Fig. 1. Starting point showing the top-level concept and its children

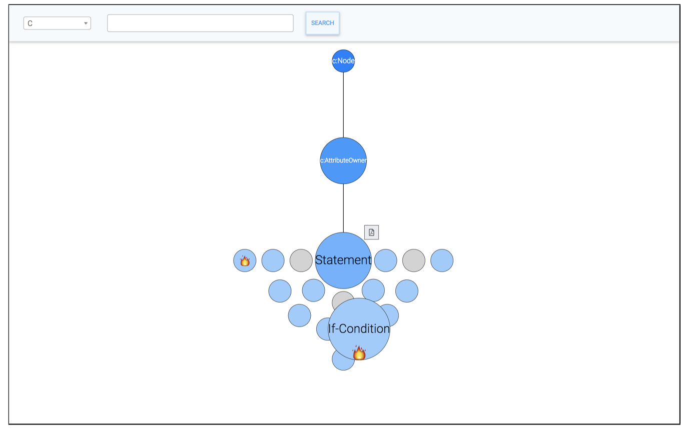

Fig. 2. Selected concept with its parent, children and documents

A. Ontology Browsing

The starting point is the root concept of an ontology (or an ontology scheme in the case of an SKOS[5]-based ontology such as the ALMA ontology). As shown in figure 1, the topmost concept `Node` is at the center. The ontology (or ontology scheme) can be selected in the top-left dropdown menu. The small circles around this concept represent its direct children. Note that the background color of the parent concept is darker than the one of its children. This is a color-based visual cue used throughout the tool. When hovering over a child concept, its name is shown. The eyes icon is a visual cue to indicate that this concept itself also has children, including learning resources related to at least one of the concepts.

Once a concept has been selected, a situation similar to the one in figure 2 is reached. Again, the deeper the level of a concept, the lighter its background color. Parent concepts are also displayed as a smaller screen object, similar to the top of a skyscraper appearing thinner from the ground perspective. Clicking on an upper concept allows to put this concept back into focus. In this case, the `Statement` concept is selected. The grey circles represent child concepts which do not have further children and have not been annotated on a resource. In opposition to this is the fire icon, which indicates that a concept has been heavily used in annotations throughout the document collection. Both visual cues help teachers in determining the coverage of ontology concepts in the learning material corpus. We have chosen this type of cue for showing that a concept is more or less used in annotations, and not a size-based one, as the size semantic is already used when showing the hierarchy. Also students can benefit from this feature, as a more frequently annotated concept might be an indicator for its importance. Finally, the PDF icon on the `Statement` concept indicates that there is a PDF document annotated with this concept. When hovering over the icon, a preview image of the resource is shown; clicking on the icon leads to a view as shown in figure 5.

[5]https://www.w3.org/2004/02/skos/

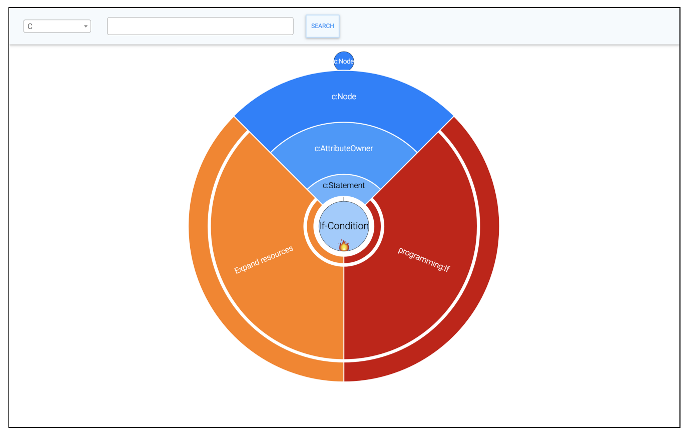

Fig. 3. Navigation wheel with shortcuts

Through a long press (i.e., clicking and holding) on a concept, a navigation wheel is shown, as seen in figure 3. Although it is possible to traverse the hierarchy up by clicking on parent concepts, the number of parent concepts shown is limited. This is particularly necessary for very deep hierarchies, as the chain of parent concepts would fill up the entire screen while top-level concepts would be unrecognizably small. Therefore, the navigation wheel, with the currently selected concept at the center, proposes shortcuts to higher-level concepts. Apart from these, it is also possible to explore concepts related to the currently selected one. In the example, the concept `If Condition` from the `C` scheme of the ALMA ontology is related to the `If` concept from the `Programming` scheme. A view similar to the one shown in figure 2 about this concept is shown when clicking on the red part of the wheel. Finally, it is also possible to see the resources annotated with the currently selected concept when clicking on the orange part of the wheel. A view similar to figure 4 is then shown. As long as a shortcut is only hovered over and not clicked, the selection will not change. However, once a click occurs, the selection changes and the view adapts.

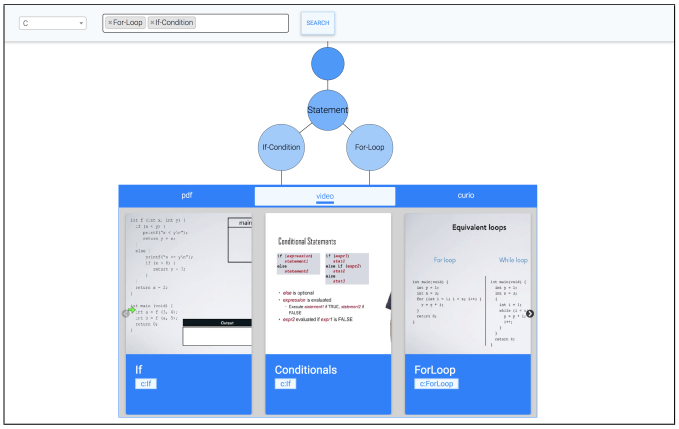

Fig. 4. Resources for a set of selected concepts

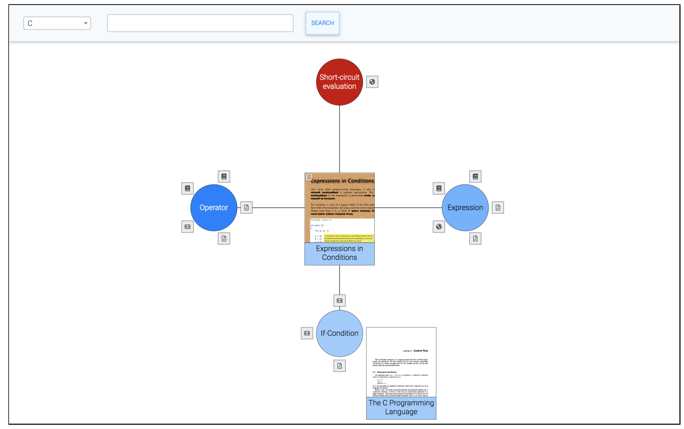

Fig. 5. Selected resource with surrounding annotated concepts

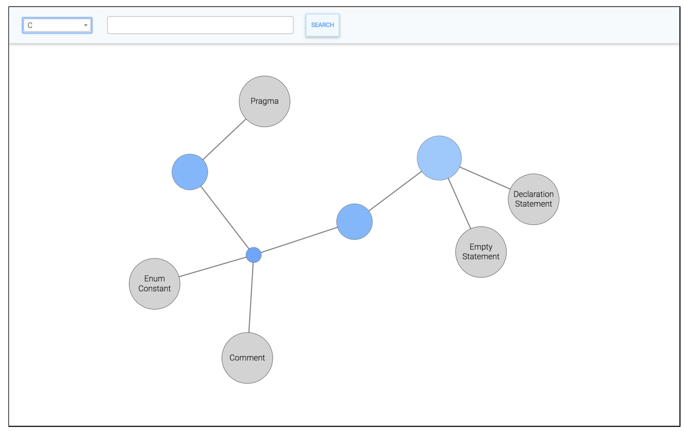

Fig. 6. Coverage view for teachers

B. Resource Browsing

Several concepts can also be selected at the same time by querying these concepts in the search field. The related resources are then presented as in figure 4. In order to avoid the inquiry paradox and support the user in the query, an integration of our visualization tool into different e-learning contexts is recommended, as discussed in section IV. As both concepts share the same parent, there is only a single ancestor tree. Learning material annotated with either one of the concepts are categorized with respect to their media type, e.g., PDF files or videos. For each resource, a preview image is shown, together with its name and the set of annotated concepts. The list can be horizontally scrolled.

Finally, once a resource has been selected, as in figure 5, it is centered on the screen, surrounded by the annotated concepts. Concepts from another ontology (or module) than the one currently explored are represented with a different background color: here, the red circle represents the `Short circuit evaluation` concept from the `Programming` scheme of the ALMA ontology. This allows to see how a document can be related to topics from different origins. For instance, if two different ontologies or ontology schemes are used in two subjects, this type of cross-curricular link of documents across subjects may support learners, particularly if they follow a global learning style, in understanding inter-disciplinary connections and dependencies. On another note, if related concepts have been annotated in other resources, these are again represented by respective icons. For example, this can be seen around the blue circle `If Condition`. Hovering over such an icon again shows a preview image of the corresponding document. This way, the implicit link between learning resources on the same topic can be easily grasped, both within the same subject and across subjects, due to the formerly mentioned cross-curricular relations within the learning material collection.

C. Coverage

To avoid that the teacher would have to browse through the whole ontology in order to discover grey circles representing non-covered concepts, a different visualization, particular for teachers, can be used as shown in figure 6. Here, all the non-covered concepts are in focus and shown at the same time. This view is limited to discovering individual non-covered concepts, and is not recommended to be used when no material at all has been indexed for any concept in the ontology. Note that the hierarchy of these concepts is still shown, but with less opacity, to be able to locate their position within the ontology. Clicking on some of the blue circles leads back to a situation as shown in figure 2.

As we have seen, our tool combines the visualization of an ontology with the exploration of related documents. Either a concept or a resource can be focused on, with the related resources respectively the annotated concepts around.

IV. Integration Opportunities

As discussed in section III, concepts respectively resources can be browsed through a visual navigation process. However, it is still possible to query a concept directly by searching for it, as shown in figure 4. In order to avoid both a lengthy visual navigation on one hand and the inquiry paradox on the other hand, our Web-based tool can be integrated into different e-learning contexts.

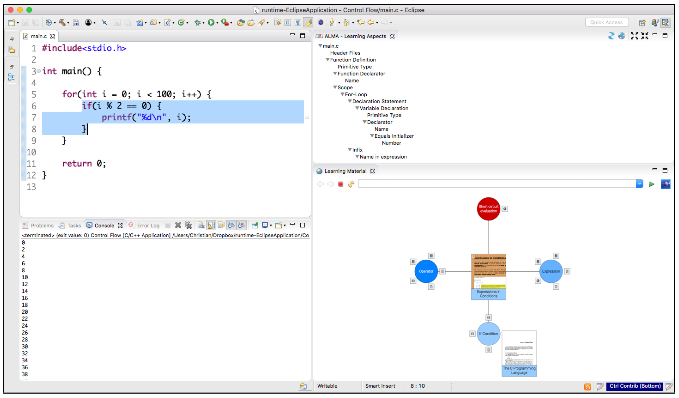

Fig. 7. Integration of our tool in Eclipse

In [16], learning resources are suggested to computer science students within the Eclipse IDE. When selecting a piece of code a learner has understanding issues with, the respective elements from the abstract syntax tree are retrieved. Those elements are mapped to concepts in the ALMA ontology, and related learning material annotated with these concepts is proposed. This way, students can consult a set of resources directly from within the IDE. Using the visualization of our tool, as shown in figure 7, related concepts as well as implicitly related resources can be discovered. Through this process, students do not need to know the name of a concept a priori, but can rely on the mapping to the ontology and the visualization of concepts and documents.

Integrating related learning material is also the focus of the work presented in [17], where authors of Office documents are enabled to have the main topics in the text recognized and linked to corresponding ontology concepts through entity linking tools. Users can then retrieve resources annotated with the recognized concepts, including knowledge base articles for biochemistry-related ontologies. Combining this Office plugin with our visualization, users could easily grasp the context of a document and the related topics.

Finally, mobile serious games such as provided within the Yactul platform [18] can benefit from the integration of learning material. When a learner has issues with a quiz activity, providing related resources can help to clarify her doubts. As those activities on Yactul are also annotated with ontology concepts, and in combination with our tool, the previously mentioned advantages can also be brought to the serious game.

Using these different integration strategies, the query for relevant documents becomes entirely visual. Learners are not required to have any prior knowledge on a concept to retrieve those resources, can grasp implicit links between documents in the collection and understand cross-curricular relations between concepts. Throughout this purely visual exploration process, neither file names nor locations have to be remembered.

V. Conclusions & Future Work

In this paper, we have presented an ontology coverage tool and document browser for learning material exploration. Users can browse through a vast corpus of semantically enhanced learning material and are enabled to pinpoint documents relevant to their study process through a visual query formulation support. Implicit and cross-curricular relations between concepts and resources can be easily grasped. A variety of visual cues including icons, coloring schemes and interactions also help teachers to evaluate the coverage of concepts within the document collection. Furthermore, the integration of this Web-based visualization into different e-learning applications overcomes the limitations of traditional file explorers, improves the workflow of learners and avoids Meno's paradox of inquiry.

There are several directions for future work. First, the visualization could not only benefit from learning resources explicitly indexed on a repository, but it could also suggest related knowledge base articles, which is particularly interesting when using life science ontologies. Furthermore, our tool currently relies only on already indexed resources, but a next step could allow teachers to drag and drop documents onto the interface. A semi-automatic annotation process based on concept recognition and entity linking tools would then index this new material, which would then appear in the visualization, as well. Finally, popular query patterns could be logged on a per-user basis, so that teachers can understand their students' search behavior, whereas students could be given a personalized list of implicit bookmarks.

References

[1] K. A. Olsen, R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams, "Visualization of a Document Collection with Implicit and Explicit Links: the ViBe System," *Scandinavian Journal of Information Systems*, vol. 5, pp. 79–95, 1993.

[2] H. L. Kim, H. G. Kim, and K.-M. Park, "Ontalk: Ontology-based personal document management system," in *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters*, ser. WWW Alt. '04. ACM, 2004, pp. 420–421.

[3] P. Andrews, I. Zaihrayeu, and J. Pane, "A Classification of Semantic Annotation Systems," *Semantic Web*, vol. 3, no. 3, pp. 223–248, 2012.

[4] S. Lohmann, S. Negru, F. Haag, and T. Ertl, "Vowl 2: User-oriented visualization of ontologies," in *Knowledge Engineering and Knowledge Management*, K. Janowicz, S. Schlobach, P. Lambrix, and E. Hyvönen, Eds. Cham: Springer International Publishing, 2014, pp. 266–281.

[5] ——, "Visualizing ontologies with VOWL," *Semantic Web*, vol. 7, pp. 399–419, 2016.

[6] A. Psyllidis, "OSMoSys: A Web Interface for Graph-Based RDF Data Visualization and Ontology Browsing," in *Engineering the Web in the Big Data Era*, P. Cimiano, F. Frasincar, G.-J. Houben, and D. Schwabe, Eds. Cham: Springer International Publishing, 2015, pp. 679–682.

[7] K. Padia, Y. AlNoamany, and M. C. Weigle, "Visualizing Digital Collections at Archive-it," in *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, ser. JCDL '12. ACM, 2012, pp. 15–18.

[8] W. Xu, M. Esteva, S. D. Jain, and V. Jain, "Analysis of large digital collections with interactive visualization," in *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, October 2011, pp. 241–250.

[9] P. Bergström and D. C. Atkinson, "Augmenting the exploration of digital libraries with web-based visualizations," in *2009 Fourth International Conference on Digital Information Management*, November 2009, pp. 1–7.

[10] F. Windhager, P. Federico, E. Mayr, G. Schreder, and M. Smuc, "A review of information visualization approaches and interfaces to digital cultural heritage collections," in *FMT2016: Proceedings of the 9th Forum Media Technology*, 2016, pp. 74–81.

[11] G. Qihong, Z. Min, L. Mingzhao, L. Ting, C. Yu, and Z. Baoyao, "Document visualization: an overview of current research," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 6, no. 1, pp. 19–36, 2014.

[12] L. Zhuhadar, O. Nasraoui, and R. Wyatt, "Visual ontology-based information retrieval system," in *2009 13th International Conference Information Visualisation*, July 2009, pp. 419–426.

[13] A. Alhenshiri, S. Brooks, C. Watters, and M. Shepherd, "Augmenting the visual presentation of web search results," in *2010 Fifth International Conference on Digital Information Management (ICDIM)*, July 2010, pp. 101–107.

[14] A. Soylu, M. Giese, E. Jimenez-Ruiz, E. Kharlamov, D. Zheleznyakov, and I. Horrocks, "Ontology-based end-user visual query formulation: Why, what, who, how, and which?" *Universal Access in the Information Society*, vol. 16, no. 2, pp. 435–467, June 2017.

[15] K. Munir and M. S. Anjum, "The use of ontologies for effective knowledge modelling and information retrieval," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 116 – 126, 2018.

[16] C. Grévisse, J. Botev, and S. Rothkugel, "An Extensible and Lightweight Modular Ontology for Programming Education," in *Advances in Computing*, A. Solano and H. Ordoñez, Eds. Cham: Springer International Publishing, 2017, pp. 358–371.

[17] C. Grévisse, R. Manrique, O. Mariño, and S. Rothkugel, "Knowledge Graph-Based Teacher Support for Learning Material Authoring," in *Advances in Computing*. Springer International Publishing, in press.

[18] C. Grévisse, J. Botev, and S. Rothkugel, "Yactul: An Extensible Game-Based Student Response Framework for Active Learning," in *Ponencias del XVIII Encuentro Internacional Virtual Educa*, 2017.