Interconnecting and Enriching Higher Education Programs using Linked Data

Fouad Zablith
Olayan School of Business
American University of Beirut
PO Box 11-0236, Riad El Solh, 1107 2020
Beirut, Lebanon
fouad.zablith@aub.edu.lb

ABSTRACT

Online environments are increasingly used as platforms to support and enhance learning experiences. In higher education, students enroll in programs that are usually formed of a set of courses and modules. Such courses are designed to cover a set of concepts and achieve specific learning objectives that count towards the related degree. However we observe that connections among courses and the way they conceptually interlink are hard to exploit. This is normal as courses are traditionally described using text in the form of documents such as syllabi and course catalogs. We believe that linked data can be used to create a conceptual layer around higher education programs to interlink courses in a granular and reusable manner. We present in this paper our work on creating a semantic linked data layer to conceptually connect courses taught in a higher education program. We highlight the linked data model we created to be collaboratively extended by course instructors and students using a semantic Mediawiki platform. We also present two applications that we built on top of the data to (1) showcase how learning material can now float around courses through their interlinked concepts in eLearning environments (we use moodle as a proof of concept); and (2) to support the process of higher education program reviews.

Categories and Subject Descriptors

E.1 [**Data**]: Data Structures—*Graphs and networks*; J.1 [**Computer Applications**]: Administrative Data Processing—*Education*

General Terms

Design, Management

Keywords

Curriculum data management, education graph, learning analytics, learning environments, linked data, semantic web

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2741740.

1. INTRODUCTION

With the advancements of Internet and Web technologies, higher education is undergoing major changes in the way courses and programs are delivered. For example we are witnessing new teaching methodologies, the explosion of open education in the form of massive online courses (MOOCs), and new means of interaction between teachers and students. This move from an offline delivery mode, to blended and online modes is opening up new opportunities for educators to explore.

In higher education, most of the programs are designed around courses that follow a certain sequence. This sequence is usually guided by the knowledge covered in certain courses, which are entry points to further knowledge covered in subsequent courses. Traditionally, courses' syllabi and program catalogs are used as the main source of course description. The problem is that this description is usually at a high level, prohibiting course designers the ability to build a holistic representation of what a course covers, and how the knowledge in a course connects to the knowledge covered in other courses. This lack of knowledge representation creates boundaries around courses that are hard to break, even in online settings.

We observe that online environments are not being exploited to their full extent in higher education. We believe that the Web of data can be used to break the knowledge islands in courses delivered in higher education. We present in this paper our work on conceptually interlinking courses within a higher education program. We propose a data model that goes down to the concepts taught in courses, and we use the concepts to cross connect the courses and learning material. We use a semantic Mediawiki [6] platform as an entry point to collaboratively build this linked data graph around the higher education program. We implement this wiki environment at the Olayan School of Business of the American University of Beirut. We then exploit the linked data generated in two pilot applications: the first is used to demonstrate how the taught concepts in courses can be used to cross connect moodle pages and enrich them with relevant learning material; while the second includes a new unprecedented view of how courses interlink and share concepts, which proved to be useful in the program review exercise of the school.

We present in Section2 an overview of the related work. Then we discuss the linked data model in Section 3. We later present in Section 4 how we generated the linked data layer by focusing on the deployment of the semantic Mediawiki and the steps we followed in building the knowledge graph. In Section 5 we highlight the two applications we developed on top of the data, followed by our concluding notes and future work in Section 6.

2. RELATED WORK

Linked data is getting increasingly used in various contexts, higher education is no exception. In their survey, Dietze et al. [7] highlight the growing adoption of linked data by various universities. Various platforms have emerged where linked data is made available for direct consumption and reuse. This includes for example the OU's linked open data platform (`http://data.open.ac.uk`), the University of Muenster (`http://data.uni-muenster.de`), the University of Oxford (`http://data.ox.ac.uk`), the University of Southampton (`http://data.southampton.ac.uk`) among others.

Methodologies and frameworks have been proposed to transform existing data sources into linked data [2, 8, 9]. In this context, available organizational data was transformed, using pre-programmed transformation patterns, into linked data. Following the trend and success of social and knowledge graphs functionalities provided by Facebook[1] and Google[2], there are discussions around the value of having an education graph[3] [4]. Heath et al. [4] proposed in their work to create an education graph by processing courses information and learning material from various universities in the UK. In their approach, they mainly rely on bibliographical data of material repositories to identify links to course resources [4].

In our context, the aim is to interlink courses within the same institution at the level of concepts covered in course topics. To our knowledge, there is no existing vocabulary that covers this granular information about courses information. To achieve our objective, we had to go at lower (i.e. more specific) conceptual levels by enabling users to have a direct impact on reshaping how courses interconnect among them and with learning resources. This controlled environment is necessary for aligning the conceptual coverage of courses delivered by more than one instructor. We also aimed to have direct input from students to connect to and expand the graph around the education program. For example when students find an interesting material online (e.g. video or article) relevant to a specific course, we wanted to enable them to connect it back to the course by extending the graph and creating the appropriate links to the course.

3. LINKED DATA GRAPH

We present in this section the data model we adopted and followed to capture courses' related information. We used the information included in the courses' syllabi template as a guide to what information to represent. A typical course syllabus follows a predefined template, which includes the course details such as course number, description, prerequisites, textbook, topical coverage and others. Furthermore, each course has well defined learning objectives. We visualize our proposed data model in Figure 1.

We focus in this part on the description of specific aspects that guided the development of the elements that are not covered in the syllabi content. Based on the need for conceptually connecting courses (i.e. beyond the analysis of topics in common), we captured in the graph concepts that are taught at the level of every course topic. This design offers many advantages. First we are getting a more granular view of what is covered in each course. Second, such concepts can be used as anchors between learning material and courses. This will enable course designers to know where exactly each piece of material fits in the course, and this enables learning material to float around not only course topics, but also around the program as a whole. In other words an interesting article that can be useful in one course, can also be used in other courses if the concepts are shared between the courses.

Following the linked data principles [3], we aimed to reuse the available ontologies that are relevant to our context. For example the CourseWare ontology[4] was reused to represent information related to course types, student interaction types, number of credits, assessment methods, and others. The AIISO[5] was adopted to capture the courses' unique codes. We have also used the Dublin Core Terms[6] to represent generic properties such as descriptions. In cases where no vocabularies were found, we created a local vocabulary used in our context.

4. LINKED DATA GENERATION

We followed a collaborative process to build the graph around the higher education program. This involved professors, students and program coordinators. As courses are delivered by different professors, we required to have a system that supports collaboration, through which each professor or involved person can see, expand and modify content as seen appropriate. Another requirement was to enable students to also enrich and add to the graph in a quick and easy way online material that they find interesting. In addition to the above, another needed feature was the ability to control linked data vocabularies. In this section, we discuss how we generated linked data around our higher education program using the semantic Mediawiki, and a semantic bookmarklet for linking online resources to the courses.

4.1 Semantic Mediawiki Deployment

We have put to the test a semantic Mediawiki available at: `http://linked.aub.edu.lb/collab`. For controlling the data generated and vocabularies used, we created several forms that are automatically loaded when new content is to be created. In the case of creating new courses for example, the course form[7] is automatically loaded[8]. The advantage of using such forms is that the user is guided around what

[1] `http://newsroom.fb.com/News/562/Introducing-Graph-Search-Beta`
[2] `http://www.google.com/insidesearch/features/search/knowledge.html`
[3] `http://hackeducation.com/2011/11/10/is-there-an-education-graph/`
[4] http://courseware.rkbexplorer.com/
[5] http://vocab.org/aiiso/schema
[6] http://purl.org/dc/terms/
[7] `http://linked.aub.edu.lb/collab/index.php/Special:FormEdit/Course/New_Course`
[8] For a filled form example, check the Foundations of Information Systems course at the following link: `http://linked.aub.edu.lb/collab/index.php?title=INFO200_-_Foundations_of_Information_Systems&action=formedit`

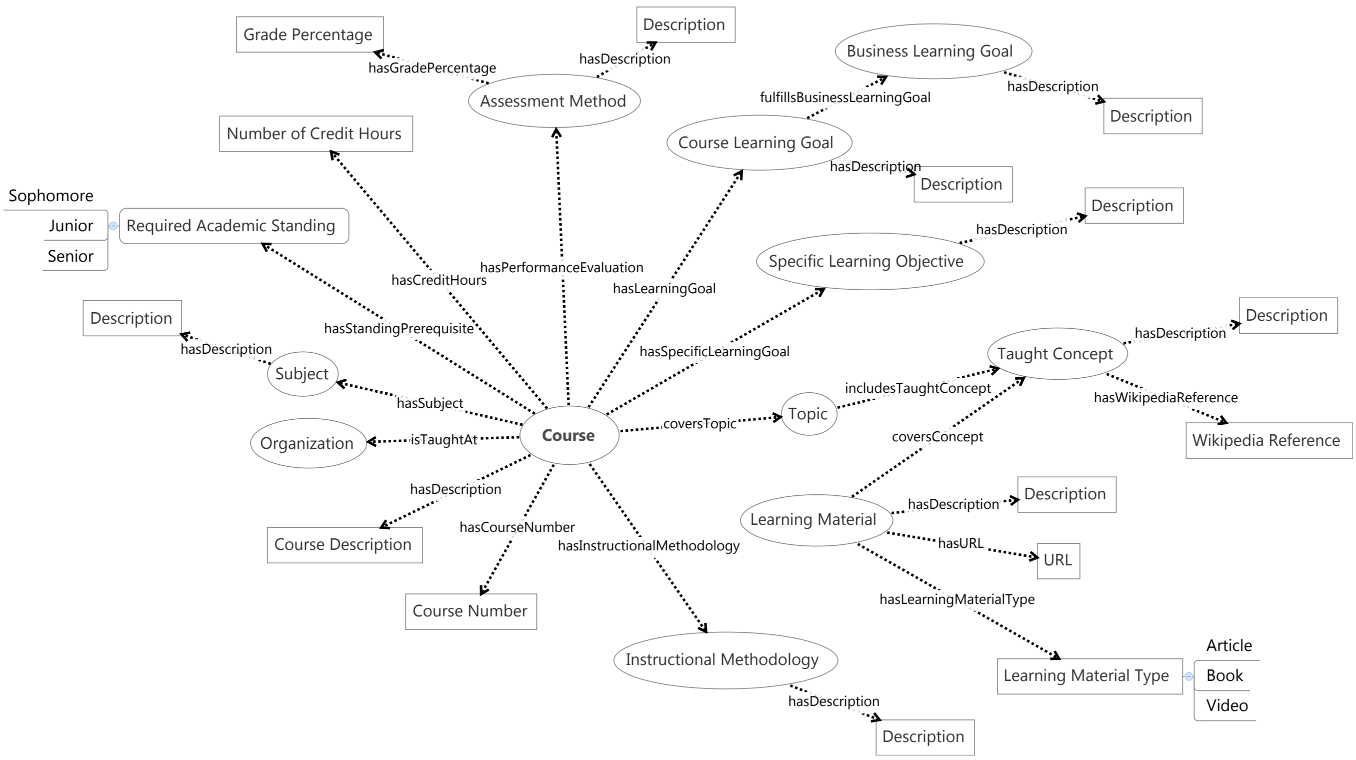

Figure 1: The Linked Data Graph.

fields to fill, and the vocabularies used can be predefined in the form when the RDF is generated.

Linking data is done by controlling the field content in the wiki. By adopting the wiki forms, the fields can guide users in reusing existing concepts from the wiki. For example when specifying the course prerequisite relation, the user is prompted with the list of courses available in the knowledge space that can be chosen from. This is a core feature used for the interlinking process. Another example is at the level of concepts covered in course topics. The same concept can be covered in one or more course topic, creating the links needed.

4.2 Steps Followed for Building the Linked Data Graph

Building the linked data graph was done in three phases. In the first phase, course syllabi are used as entry points, where the high level course information is entered. In the second phase, textbook materials used in the course are processed by the teaching assistants (TAs) to identify the concepts covered in the course. In the third phase, new external materials are added to the graph by instructors and students, using a semantic bookmarklet.

Phase 1: Creating Course Information. This phase was straight forward, as course syllabi follow a predefined structure. Some difficulties were faced at the level of identifying learning goals, as some of courses' learning goals were described as text, without a clear structure. Another complexity at this level was that course learning goals were at two levels (as depicted in Figure 1): courses with specific learning goals, which are linked to a broader list of business learning goals identified within our School of Business. Another challenge faced at this level is when the topics of courses change with time, due for example to changes in textbook editions, or in the delivery of content across semesters. For instance, in the Foundations of Information Systems course, the Social Media topic covered in the "Experiencing MIS" [5] textbook has changed from the 3rd to 4th edition. In this case, while the topical coverage has changed, some of the concepts that were covered in the previous edition were still there. We handled such cases by archiving topics and removing the link to the course. This way we were able to preserve the concepts related to the old topic, and reuse them if needed in the new topic. The evolution of changes at this level can be better managed in the future. One possible improvement can be done through capturing temporal changes, coupled with the type of changes performed (e.g. adding or removing concepts from the concept graph). While tracing such evolution patterns can be done at the data entry level, a post analysis of changes occurring on the data graph can be possibly performed. Such features are beneficial for analytics applications, and can be further explored as part of our future research. We have processed so far the 19 core courses offered at the School of Business, leaving the elective courses to be represented at a later stage.

Phase 2: Identifying Concepts Taught in Courses. This phase was the most extensive and time consuming phase. TAs were trained to identify concepts covered in the topics of the courses. Based on the model we created (cf. Figure 1), concepts are linked to the topics of a course, and not directly to the course itself. This choice of design enables grouping concepts by topic, rather than by course. This has a practical implication in filtering concepts covered in specific topics (as we discuss later in the paper when we integrate data in moodle). Another implication is at the level of program analytics. Overlap among course topics can be highlighted easily, which proved to be useful for the program review and design exercises. At this level, the TAs were going into each topic covered in the textbook to identify the main concepts, adding definitions from the book, and linking to a Wikipedia[9] reference when found. When applicable, the TAs were instructed to reuse existing concepts in the repository. The *concepts* field in the *topic* form automatically provides the TAs with the list of existing concepts to choose from. One challenge at this level is when concepts from different topics share the same name, however are semantically different. For example, the *Optimization* concept covered in the *Managerial Decision Making* (i.e. operations research), is semantically different than the optimization concept in the *Managerial Economics* field. This is where Wikipedia is used as an external reference for disambiguating such cases. While linking to the concepts to Wikipedia entries is currently done manually, this task can be performed (semi)-automatically by querying DBpedia [1] and aiming to find overlap between the concepts' definition and the description in the DBpedia page. In addition to text matching, the graph of the concepts can be used as a context to highlight the degree of matching. Then the user can browse the proposed matching concepts to select the most appropriate one. Currently concepts identified are not related through an explicit relation. We are planning to capture in the future relations such as sub-class and prerequisite relations among concepts. This can also have an impact on the analysis of course sequencing in the program. So far we have identified 2680 concepts covered in the core courses of the School of Business.

Phase 3: Semantically Anchoring Learning Material to Courses. While the previous phases were mainly focusing on reorganizing internal knowledge sources, this phase is more about linking new materials to the program, by anchoring them to concepts covered in courses. Here the aim was to enable students and instructors to link any interesting online material to the graph. We developed a simple bookmarklet that can be installed in any browser. When the user links a learning material, clicking the bookmarklet will automatically extract the page link, title and description. The user is then prompted to a page pointing to our wiki platform, where the concepts covered in the material can be entered by reusing existing concepts from the graph[10]. As mentioned earlier, links between a learning material and the program are done through the concepts. This somehow enables educators and students to think more around the relevance of the material around concepts. For example we witnessed a student who bookmarked an article related to Big Data[11], relevant to the Foundations of Information Systems course she was enrolled in. While anchoring this article, she reused concepts[12] from the graph that indirectly spread to two other courses. Starting from an information systems' related material, the student indirectly linked to a management and operations management courses.

[9]www.wikipedia.org

[10]A video tutorial on how to use the bookmarklet can he accessed at: `http://linked.aub.edu.lb/docs/tutorial_material_bookmark`

[11]`http://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making`

[12]The full list of concepts highlighted by the student can be found in Figure 2.

5. CONSUMPTION OF LINKED DATA

We present in this section two scenarios where we used the linked data generated. We accessed the data through the semantic search feature of the wiki endpoint that provides querying functionalities with different output formats. The queries were formulated and passed through php to the wiki query endpoint, and results were returned in JSON for processing.

5.1 Interlinking Course Learning Material on Moodle

The data generated was used to enrich and connect the moodle course pages. Moodle is extensively used at the School of Business as a way to communicate course related information, and to interact around course deliverables.

The suggested design of moodle course pages at the American University of Beirut is to subdivide the page using the covered topics in the course, and add topic related material, assignments and others under each section. Course instructors design their own page at the beginning of each semester, or reuse an existing one if the course was already taught by the instructor. One trend that is observed is that instructors of the same course tend to share interesting materials that could be used in classrooms. However such insights are not usually captured, and have to be re-shared whenever a new instructor teaches the course. In addition to sharing constraints, another bottleneck observed is that courses are designed (even on moodle) in isolation. However material relevant to some courses, can potentially be relevant to others (as perceived with the Big Data article discussed earlier).

The aim of this application is to break out of the static and isolated nature of content shared within a course topic on moodle. The application offers the functionality to dynamically enrich moodle pages, without leaving it, with material relevant to the topics of the course.

A video tutorial was used as a guide for students and instructors to follow[13]. This application can be launched by pressing a bookmarklet in the browser, and the following sequence of steps is performed:

1. Scan the topics available in the moodle page: the executed code first will launch a javascript that will scan the moodle page for the topics header.

2. Enable the buttons on the moodle page: the buttons are dynamically added to the moodle page, with the corresponding topic and course code embedded inside the button links.

3. Send query to the wiki endpoint: when the user presses the button, a query is passed to the wiki endpoint with the topic and course code. As per our model in Figure 1, the link between the course and material is done at two levels, through the *topic* and then through the *taught concepts*. So the query is built to first fetch the concepts covered in a topic, and then filter the learning material based on the concepts in focus.

[13]`http://linked.aub.edu.lb/docs/tutorial_extract_material/`

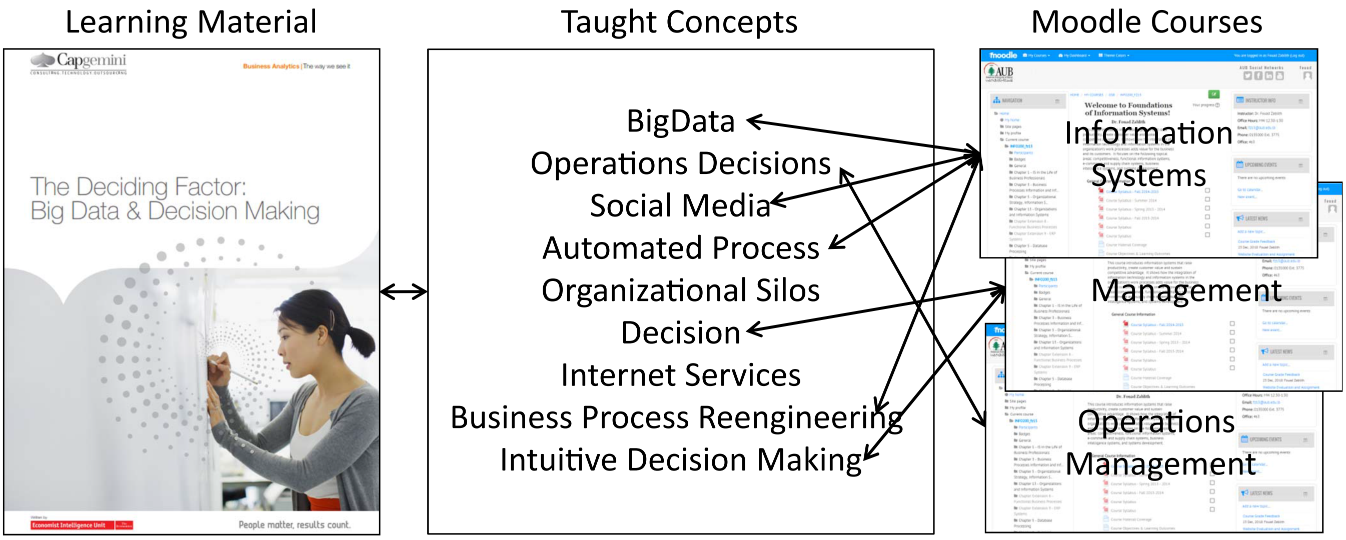

Figure 2: Learning Material Cross Connecting Courses.

4. Parse and visualize query results: the query results are returned in a JSON output, and parsed to identify the different types of related material (so far we have video material, articles, and books). Finally the results are used to populate the page where users can read articles or play videos.

Figure 2 highlights how the Big Data article mentioned earlier is linked to various courses through the commonly taught concepts. This design can place the material automatically under the right topic on the moodle page.

5.2 Using Linked Data for Program Review

Another context where we used linked data was in the program review process at the School of Business. Every four years, the curriculum has to be reviewed for changes, where courses are studied to be added or removed from the program. Another task that is part of the review process is the course sequencing. To achieve this purpose, traditionally each course is studied on its own, and compared to other courses, and to the learning objectives in the program. However it was not possible to perform an in-depth analysis beyond the syllabi content, and hence it was hard to know exactly what is covered in each course, and what are the concepts that are repeated across different courses.

Course mapping was highlighted to be one of the major tasks required for adjusting course sequencing and coverage. Our platform provided an unprecedented view of how courses overlap, going to the concept level. We deployed a visualization showing how courses connect through topics and taught concepts. This visualization is dynamically generated based on the wiki content. A visualization example around the Foundations of Informations Systems course can be accessed online at: `http://linked.aub.edu.lb/collab/index.php/Learning_Concepts_Graph`. This visualization is created using the HyperGraph tool[14]. However this visualization proved to be a bit complex to analyze in a systematic way.

[14]`http://hypergraph.sourceforge.net`

Another view that we built on top of the data was a tabular representation of the courses, along with the topics and overlapping concepts. The table can be accessed online[15], and we show part of the table in Figure 3. The users were able to see the list of courses (course codes are on the top row of the table), and in each row we present the list of topics and their corresponding concepts. The “X” marks the number or occurrences of this concept in the whole program. When more than one “X” is in the cell, the user can roll the mouse over the table cell to see where this concept occurs in other courses. For example Figure 3 indicates that the concept *Audit* is mentioned in the Accounting 210 course, and in the Marketing 222 course. We also fetch the topics where the concepts are covered. This tool proved to be useful in the program review process, as it enabled the program coordinators and instructors to highlight the parts of the courses that need adjustment.

6. CONCLUSION AND FUTURE WORK

We presented in this paper our work on connecting higher education program information at a conceptual level using linked data. The data model we designed captures and connects courses information, going down to the topical coverage and concepts taught. The concepts are then used as anchors between learning materials and the higher education program.

We highlighted our continuous work on collaboratively building this linked data graph by involving people around a semantic Mediawiki. The wiki offers a platform where instructors can reach a consensus around what is taught in their courses, by having a controlled environment to manage the reuse of existing knowledge and appropriate vocabularies for creating linkages.

One advantage of linked data is the ease of building applications that make use of the data generated. In our work we showed how we can bring learning materials in the context of a course, and place them under its corresponding topic.

[15]`http://linked.aub.edu.lb/apps/tablebrowser/table.php`

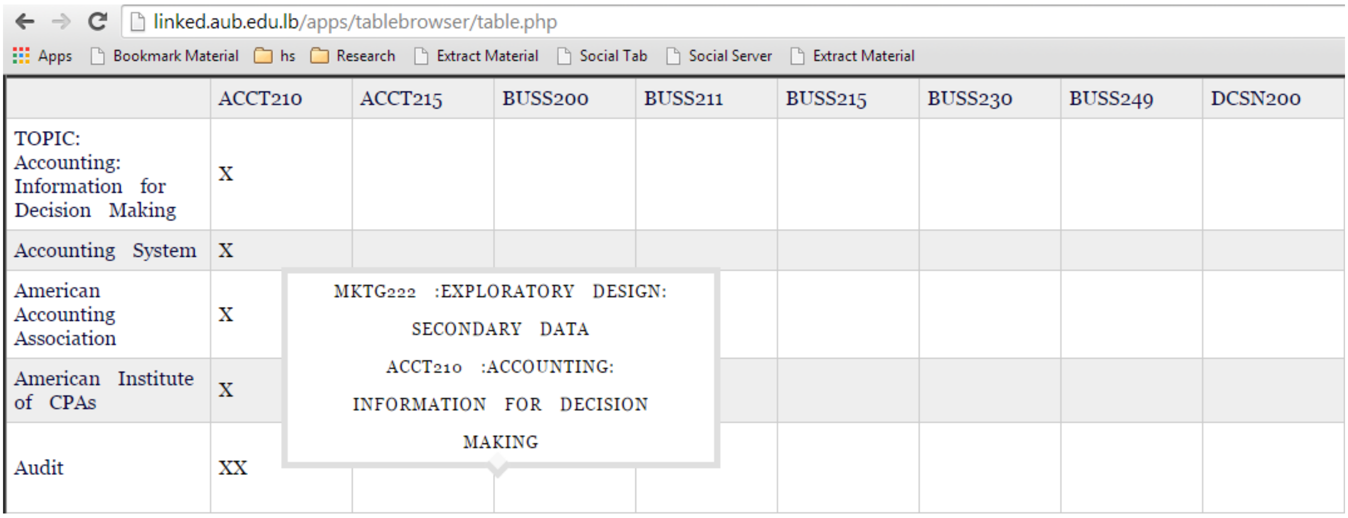

Figure 3: View of the Table Listing Concepts Covered in Courses with Overlap Detection.

The use of taught concepts as anchors between the materials and courses gave us a great flexibility in fetching material that stretch the boundaries of courses delivery, which can improve the learning experience of students and highlight how courses cross-connect in the degree program they are enrolled in.

Another scenario where we applied the use of linked data is at the level of program analysis and review. Program designers at the School of Business were offered a unique view that was not possible before on how courses conceptually connect. They were able to see how concepts are repeated in courses, enabling them to make better decisions around required changes in the program.

We believe that this work is only a glimpse of what can be done with this new linked data layer. We are planning to evaluate the impact of having such data layer in learning environments through a guided user study coupled with evaluation measures. We are currently working on capturing social interactions around our education program. By merging the social and education graphs, we anticipate that we will be able to granularly analyze how teachers and students interact around concepts delivered during their higher education journeys.

7. ACKNOWLEDGMENTS

This work was supported by the University Research Board grant at the American University of Beirut. I would like to thank my colleagues and students who contributed to the extension of the graph through the wiki, and supported me in the implementation of the tools. I would also like to thank the students who have used the bookmarklet to bookmark the interesting materials that triggered invaluable insights and observations.

8. REFERENCES

[1] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, 2009.

[2] M. d'Aquin. Linked data for open and distance learning. Commonwealth of learning report, 2012.

[3] T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. 2011.

[4] T. Heath, R. Singer, N. Shabir, C. Clarke, and J. Leavesley. Assembling and applying an education graph based on learning resources in universities. In *Linked Learning (LILE) Workshop*, 2012.

[5] D. Kroenke, D. Bunker, and D. N. Wilson. *Experiencing mis*. Prentice Hall, 2010.

[6] M. Krotzsch, D. Vrandecic, and M. Volkel. Semantic mediawiki. In *The Semantic Web-ISWC 2006*, pages 935–942. Springer, 2006.

[7] Stefan Dietze, Salvador Sanchez-Alonso, Hannes Ebner, Hong Qing Yu, Daniela Giordano, Ivana Marenzi, and Bernardo Pereira Nunes. Interlinking educational resources and the web of data. *Program*, 47(1):60–91, Feb. 2013.

[8] F. Zablith, M. d'Aquin, S. Brown, and L. Green-Hughes. Consuming linked data within a large educational organization. 2011.

[9] F. Zablith, M. Fernandez, and M. Rowe. Production and consumption of university linked data. *Interactive Learning Environments*, 23(1):55–78, 2015.