SLOB: SECURITY LEARNING BY ONTOLOGY BROWSING - COMPREHENSIVE CYBER SECURITY LEARNING RESOURCES IN A WEB PORTAL *

Soon A. Chun
Information Systems and Informatics
CUNY-College of Staten Island
Staten Island, NY 10314
718-982-2931
soon.chun@csi.cuny.edu

James Geller, Ankur Taunk, Karthik Sankaran and Tushar Swaminathan
Department of Computer Science
New Jersey Institute of Technology
Newark, NJ 07102
973-596-3383
{james.geller,at375,ks584,ts336}@njit.edu

ABSTRACT

Ontologies represent knowledge of a domain that can be used for data annotation, natural language processing, data integration, etc. In this project, an ontology is used to support learning and teaching. We present the SLOB (Security Learning by Ontology Browsing) Web Portal that brings together many learning resources for college-level cyber security classes at one central location. SLOB provides simple access to multi-media data, including videos, PowerPoint presentations, book pages, images and scientific papers in a unified framework. These resources are indexed by a Cyber Security Ontology created in this project, which provides definitions and hierarchical context for domain concepts. In addition, SLOB integrates social media cyber security sources, both targeted towards specific ontology concepts and general purpose

cyber security. The methodology and approach are transferable to any knowledge-based teaching domain for which an ontology has been created.

INTRODUCTION

Students often want to augment what they hear in their lectures and see in their classrooms by additional learning materials. This might be motivated by a desire to learn more than is on offer in the classroom, or, on the contrary, by falling behind or by having missed the classroom lectures. There are many learning resources available on the Web, but a search with one of the standard search engines often returns material that is irrelevant in that it is not geared towards educational purposes. For example, many videos in Computing and its subfields are company-sponsored, product-oriented, and naturally "one-sided." Furthermore, a student is often searching for a specific modality of teaching material. If a lecture was missed and the student has sufficient time, a video-recorded lecture may be desired. On the other hand, if there is an impending exam, then a simple PowerPoint presentation on the topics may be preferred. Finding *bona fide* teaching materials in the desired modality is not supported by the widely-used search engines. Furthermore, the goal of a student is typically to find "the one best learning resource," while a search engine returns many Web pages that are ranked by criteria and algorithms that are not made public and that might not work according to the best interests of the student.

Background

Cyber security has become an important topic in Computer Science education [1]. The WWW provides a rich set of resources for studying topics in Computing. For example, the Web site http://www.w3schools.com/ provides access to excellent short courses on many Web-related topics, such as HTML, XML, Google Maps, AJAX, etc., as well as ancillary topics such as SQL. However, the courses are text-based, and the student advances through the different lessons by clicking on forward arrows. Learning preferences [7] are still a topic of ongoing research, but it is widely assumed that some students would prefer a spoken lecture instead of or in addition to the visual presentation of the material.

Searching for learning resources with a search engine such as Google is based on key words, and in spite of Google's work on knowledge graphs [8] is not based on concepts. Thus a student who does not know the correct search term will not find what he is looking for. This situation is, by the definition of "learning," more common in an educational environment than in a "normal" Web search. On the other hand, many search terms are "multi-homonymous." A Google search for "security" will return Web pages about cyber security, but also about airport security, locksmiths, homeland security, etc. A "security" may also be an investment instrument.

Ideally, a Web search should be a concept search, not a term search. As this goal is difficult to achieve, the SLOB project does the next best thing. Concepts from the cyber security domain are organized in an ontology [5]. Learning materials are then categorized by modality and indexed by concepts from the ontology. Ontologies with the topic of

cyber security have been presented by a number of researchers, e.g., [2,3]. The ontology used in the project was developed based on prior work of [6] and is described in more detail in [4].

Project Goals

The goal of the SLOB project is to collect learning materials or links to learning materials about cyber security in one single location, indexed by the concepts of a Cyber Security Ontology. Students browse the ontology starting from a root concept down to the specific concepts that they are interested in. Clicking on a specific concept brings up videos, lectures (as PowerPoint presentations), scientific articles, a concept definition, more general concepts, more specific concepts (if available), related images, etc. Watching a video lecture is a considerable time investment, thus, videos are ranked according to their relevance for a specific search concept. For students who are interested in how a cyber security concept relates to current affairs, tweets mentioning the specific concept are also displayed. Students who are interested in cyber security but "don't know what they are looking for" are presented with two social media sources on cyber security as a general topic. This is of interest, for example, when a new computer virus is attacking the computing infrastructure of the United States. Users with a general interest in cyber security need to be alerted to such events.

METHODOLOGY

The SLOB application is built using a Service Oriented Architecture (SOA). The architecture is designed to integrate and enhance features with agility keeping in mind the scalability aspect. The architecture is divided into two layers which make the UI (User Interface) layer loosely coupled with the business logic layer. The design follows the MVC (Model/View/Controller) design pattern. Thus, this separates the UI and the business logic layer and the development is less error-prone. The design also takes into account the factor of configurable code, which helps tweaking the logic on the fly without any code changes.

The Java J2EE framework was used along with the MVC design pattern to develop the UI layer. The implementation makes extensive use of JavaScript and jQuery to make the UI fast, robust and responsive. The Ontology is stored as an OWL (Web Ontology Language) file. By parsing this OWL file, the program generates a JSON (JavaScript Object Notation) file, which is fed as an input to a pre-existing tree view program to display the ontology in SLOB. As and when the user selects a class name or types the class name in the search box, an asynchronous call is made to an API (Application Programming Interface) for fetching and screening the data.

The program relies heavily on calls to several APIs that provide the source data. Scientific papers are access through the DBLP API. DBLP (Databases and Logic Programming) is a widely used bibliographic database that tracks publications in many areas of Computer Science, going well beyond the original two topics, "databases" and "logic programming." The DBLP API (http://www.dblp.org/search/api) exposes a GET() method that is used in SLOB.

The YouTube API (http://gdata.youtube.com/feeds/api/videos) also provides a GET() method with several useful query string parameters. PowerPoint presentations are accessible through the SlideShare API (https://www.slideshare.net/api/2/search_slideshows?q=ask&api_key=api). The API for Google Books is available at https://www.googleapis.com/books/v1/volumes?q=Technology&key=api_key.

The data repository of SLOB is implemented using an open source database system called MySQL. As it is both impossible and unnecessary in the existing computing environment to store many YouTube videos, the MySQL database contains links to videos that have been determined to be important and relevant to cyber security, as well as metadata about these videos. Metadata contains information about the length of each video, the number of "likes" it received from other users, the caption under the video, etc. Users may browse SLOB with or without creating their own user names. If a user creates a user name then s/he can store a private collection of videos. If a user does not find what s/he needs in the pre-stored collection of videos, then SLOB will perform a dynamic search of YouTube, using the selected ontology concept and the additional term "security" as search terms.

The most complex part of the SLOB program is the video ranking algorithm. The algorithm operates on the concatenation of the title, description and caption of each video. This block of text will be referred to as "snippet." A video will be ranked higher if the ontology concept and the word "security" appear more often in the snippet. Ranking is only done for the pre-stored videos. The problem solved here is that many cyber security terms are also used in other areas with a different meaning. Thus the term "asset" may describe a computer network in the cyber security context, but will refer to very different entities in a military context or a financial context. Thus the ranking needs to keep in mind (1) the absolute number of times the ontology concept requested by the user occurs, (2) the absolute number the term "security" occurs, and (3) the balance between those two numbers. A video will be preferred if it has high absolute numbers as well as a good balance. A detailed description and pseudocode display of the algorithm would go well beyond the scope of this paper.

RESULTS

A prototype of the SLOB Web portal was implemented as described above (http://isecurelab.info:8080/WASlob/index.html). Figures 1 to 3 are screen dumps from this prototype. Figure 1 shows the entry screen of the system where users see trending on Reddit the (self-proclaimed) cover page of the Web and Twitter on general cyber security subjects. In the upper left corner is a pull-down menu for an indented view of the ontology (on black background). The top level concepts, such as Asset, Threat, Defense Strategy, etc. are visible in Figure 1.

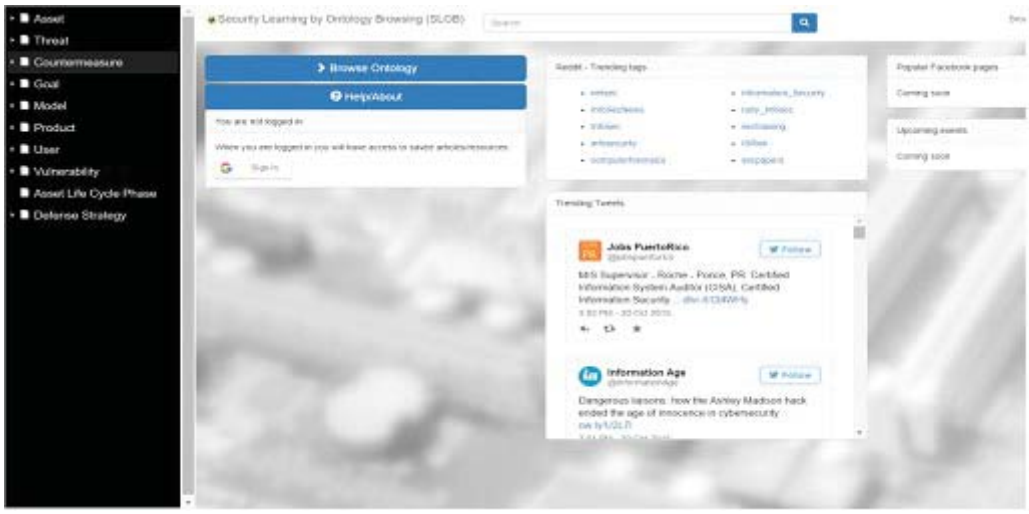

Figure 1: Entry screen of SLOB with Ontology opened

Figure 2 shows the main screen that appears after a concept has been chosen by the user with the mouse. The upper left corner shows the concept definition, superclasses (parents in the concept hierarchy), subclasses, etc. The other six subwindows show (from left to right) Google images, scientific articles from DBLP, YouTube videos, (and in the second row) PowerPoint presentations from the Slideshare Website, relevant pages from Google Books, and tweets mentioning the selected ontology concept.

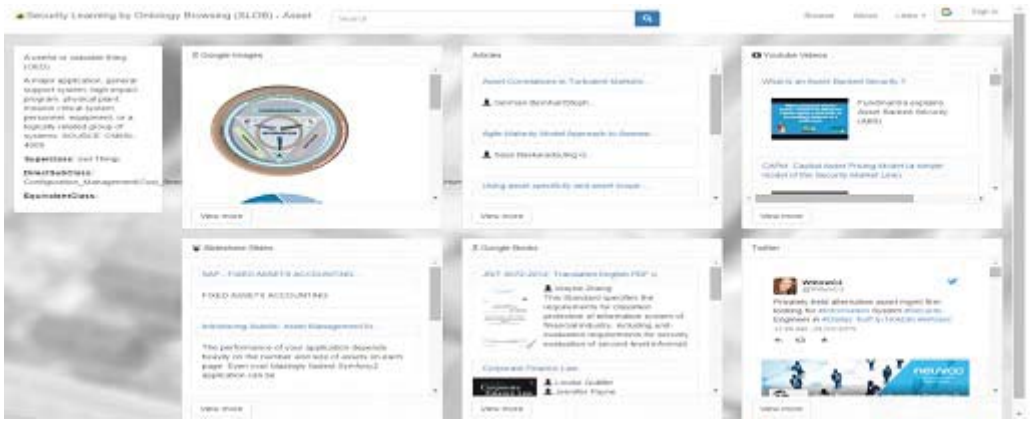

Figure 2: Main screen of SLOB

When a user is interested in getting more details in one of the six subwindows, s/he can click on the View more button under it. Figure 3 shows the result of pushing the View more button under the Google Books page.

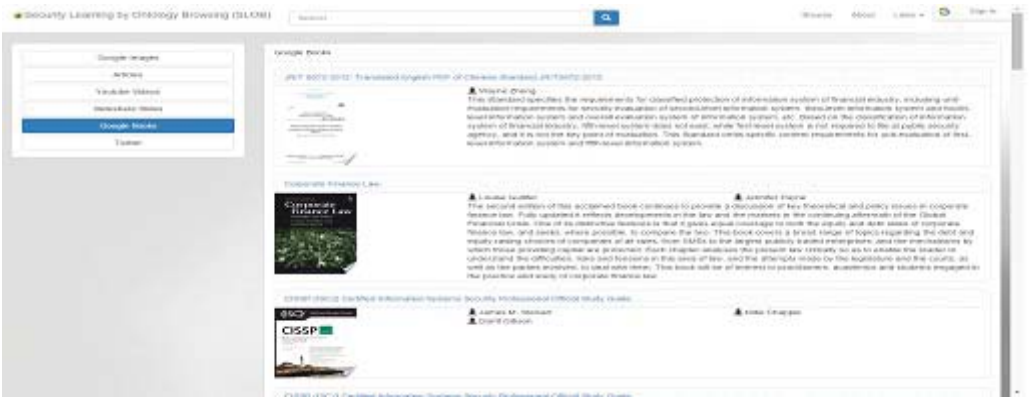

Figure 3: Expanded view of Google Books relevant to the topic.

The video ranking algorithm was evaluated by comparing the results of the ranking algorithm with the determinations of three human raters (evaluators). Three ontology terms were chosen: Phishing, SSL (Secure Socket Layer) and Symmetric Cryptography. For each of these keywords, 15 videos were chosen and ranked by the algorithm and the three human raters according to a five point Likert scale. Evaluators were graduate CS students who were otherwise not involved in this research. The choices were: Strongly Relevant, Relevant, Not Sure, Not Relevant, Strongly Not Relevant. Comparisons between human raters and the algorithm were computed using "Cohen's ?" (Kappa). We performed pairwise comparisons between the three human raters and between each one of the three human raters and the algorithm. These comparisons were performed for the three evaluation terms, phishing, SSL and Symmetric Cryptography. Subsequently all human/human values of ? were averaged (0.2705), and all algorithm/human values were also averaged (0.3027). The results show that there is better agreement between the algorithm and the human raters than between pairs of human raters. This indicates that the ranking algorithm produces a satisfactory result.

CONCLUSIONS

SLOB makes it easier to access cyber security learning resources in several different modalities. Students can search for relevant materials in a limited environment and select concepts based on an Ontology. YouTube videos were pre-stored and their relevance to "security" and to the 861 ontology terms in the Cyber Security Ontology was algorithmically determined. The videos were ranked based on this algorithm and the results were compared with three human raters. Overall, the agreement between the algorithm and human raters was slightly higher than the agreement between pairs of human raters. Future work will include an evaluation of the system with several undergraduate students taking a cyber security course.

ACKNOWLEDGMENT

This work is partially funded by NSF grant DUE1241687. We thank Ashneel Sharma and Michel Mansour for their work on the ontology and prior prototypes.

REFERENCES

[1] Bai, Y., Wang, X., Teaching Offensive Security in a Virtual Environment, *Journal of Computing Sciences in Colleges*, 31(1), 140-142, 2015.

[2] Bajec, M., Eder, J., Souag, A., Salinesi, C., Comyn-Wattiau, I., Ontologies for Security Requirements: A Literature Survey and Classification. *Proceedings Advanced Information Systems Engineering Workshops*, 2012.

[3] Fenz, S., Ekelhart, A, Formalizing information security knowledge. *Proceedings 4th International Symposium on Information, Computer, and Communications Security*, Sydney, Australia, 2009.

[4] Geller, J., Chun, S.A., Wali, A., A hybrid approach to developing a cyber security ontology. *Proceedings of DATA 2014, 3rd International Conference for Data Management Technologies and Applications*. Vienna, Austria, 377-384, 2014.

[5] Guarino, N., Oberle, D., Staab, S., What is an Ontology?, *Handbook on Ontologies*, 1-17. Springer Berlin Heidelberg, 2009.

[6] Herzog, A., Shahmeri, N., Duma, C., An Ontology of Information Security. *International Journal of Information Security and Privacy*, 1(4), 1-23, 2007.

[7] Rezler, A. G., Rezmovic, V, The Learning Preference Inventory, *Journal of Allied Health*, 10(1), 28-34, 1981.

[8] Singhal, A., Introducing the Knowledge Graph: things, not strings, *Google Blog*, http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html. 2012. Accessed 12/11/2015.