A web-based system for connecting undergraduate students and researchers via textbooks for multi-disciplinary, research-oriented education

Steven B. Kraines[1,*], **Weisen Guo**[1], **and Takaki Makino**[1]
[1]Science Integration Programme – Human, Division of Project Coordination, the University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa-shi, Chiba Pref. 277-8568 Japan
[*]sk@scint.dpc.u-tokyo.ac.jp

Abstract – *An application of the University on Textbooks (UoT) system to a web-based platform for supporting organic growth of an knowledge network related to sustainability science is described. The UoT system, which aims to help university students learn about cutting-edge research at their university through the medium of textbooks, uses a description logics reasoner to generate semantic links between ontology-based descriptors of topics in textbooks and of research articles. Those semantic links enable students to navigate between well-structured introductory materials provided by textbooks and advanced research on specific topics presented in research articles. The new UoT system maps sections from the book "Vision 2050: A Roadmap for a Sustainable Earth" to research articles and web pages of researchers at the University of Tokyo studying issues related to sustainability, energy and materials. The system is described, and implications for supporting education in highly multidisciplinary subjects such as sustainability science are discussed.*

Keywords: Expert knowledge, knowledge network, logical inference, semantic matching, undergraduate education, sustainability science

1 Introduction

Research universities are mandated to both 1) provide world-class undergraduate education and 2) advance world-class research. In theory these two tasks could work together, e.g. by effectively communicating knowledge produced through university research to students via the education system. However, in reality effective communication between researchers and students is rare, mainly because of the way in which researchers publish their knowledge in research articles. We identify three major obstacles preventing undergraduate students from accessing advanced research at their universities. First, undergraduate students do not know how to search for research articles in the literature [17]. Second, undergraduate students do not have the background knowledge needed to read and understand research articles [36]. Third, undergraduate students do not have sufficient knowledge breadth to understand how knowledge in a single research article connects with the overall knowledge in the field [30].

In order to overcome these obstacles, the research articles need to be arranged or "structured" in some way so that undergraduate students can easily access and learn about them within the context of their normal educational activities. We suggest that textbooks could be used to provide this structure. Textbooks are comprehensive overviews of specific fields in a well-structured form, which make textbooks well-suited as platforms from which students can explore research articles and other knowledge resources. The structure that is provided by textbooks, which enables students and teachers to share a broad common understanding of a field of science, is a consequence of the relatively slower rate in which the content of textbooks are updated in comparison to the research literature. We suggest that by using semantic matching techniques to position research articles in the knowledge structure provided by the textbooks, those articles could be made more accessible to students. In this sense, the textbook is like a slowly growing tree upon whose branches researchers hang their new research findings.

As an implementation of this idea, the University on Textbooks system has been developed that links research articles to a textbook in the domain of life sciences. Descriptions in a formal knowledge representation language grounded in a description logic, which we call semantic statements, were created for 100 passages from a life sciences textbook used by second-year undergraduates at the University of Tokyo and for 400 research articles in life sciences authored by researchers at the University of Tokyo [26]. The University on Textbook system is comprised of a web-accessible textbook browser, which we call the UoT textbook browser, and a web-based semantic statement editing and matching system called EKOSS for Expert Knowledge Ontology-based Semantic Search. The UoT textbook browser, which is directly accessible by a web browser with no need for plug-ins or any other software installation, provides a paginated and customizable layout showing the text and figures from the textbook. Pagination

and layout control is believed to be important for making textbooks easy to read on a web browser [4], [45]. The UoT textbook browser and the EKOSS system are connected via a set of web services that pass information about which researchers have written research articles that are semantically related to passages from the textbook, as soon as that information is updated.

Here we describe a new University on Textbooks system that we have developed, which uses the book "Vision 2050: A Roadmap for a Sustainable Earth" to provide a foundation for accessing advanced research related to sustainability science. The contents of the book have been adapted to the textbook browser, and we have created semantic statements to describe about 70 specific passages from the book varying in length from one sentence to a few paragraphs. Using a semantic matching algorithm developed in previous work, we evaluate the semantic similarity between each textbook statement and about 200 semantic statements that we created based on research articles and web pages written by researchers at the University of Tokyo to report research work related to the topics of sustainability, energy and materials. We describe the system and discuss the effectiveness for supporting undergraduate education in the highly multidisciplinary field of sustainability science.

2 Background and Related Research

Most recent work in e-learning and web-based education is focused on "stand-alone" modules of interactive learning material. These e-learning modules can be helpful supplementary teaching tools; however, they do not have sufficient structure and connection to provide students with a comprehensive learning environment. Even highly structured web resources, such as Wikipedia, lack the clarity and integrity that is needed in an undergraduate course. Textbooks, on the other hand, contain knowledge of the target domain that has been carefully structured, often through many revisions [1], making them suitable media for introducing the basic concepts and tools in specific academic fields to students who have very little background knowledge. Moreover, textbooks, particularly those that have been around for awhile, are often used as references by practicing researchers and teachers, e.g. to recall previously learned but partially forgotten knowledge.

The use of books as a framework for organizing knowledge resources in a particular field is certainly not a new idea [43], [33]. Many publishers provide web-accessible supplemental content for textbooks, such as problem sets, graphics, simple software applications, multimedia materials, and even web-accessible textbook readers giving free or prepaid access to the actual textbook (e.g. [40]). With web-based textbook readers, this supplemental content can be provided to readers "on the fly" [10], [4]. And books on the web can also support movies, animations, three dimensional graphics, and interactive learning applications [31], [32], [30].

The problem of how to map resources to specific sections of a textbook is more difficult. Some researchers have tried to map parts of a textbook to research articles or even other parts of the same textbook using glossary terms [4], [43]. However, the accuracy is limited by the ability of computer to match the glossary terms to the research articles, which is usually done in a typical "bag-of-words" fashion. Although other researchers have tried to use natural language processing to extract relationships between glossary terms [13], semantically rich relationships cannot be extracted reliably due to ambiguities of natural language.

The task of effectively matching knowledge resources has received considerable attention recently [24], [38], [12], [46]. Matching techniques can be differentiated into automatic approaches and manual approaches. Automatic approaches include semantic methods from simple keyword indexing to more sophisticated natural language understanding, and statistical methods, e.g. the use of collaborative filtering or inverse document term frequency measures. Manual approaches involve a human in the matching process, often to provide some form of computer-interpretable descriptors, which range from simple lists of keywords to descriptive statements in some formalized knowledge representation knowledge [35]. The computer-interpretable descriptors can then used by some computational algorithm for matching, as well as for classifying resources and extracting information of potential interest [18]. Semi-automatic methods that use automatic natural language processing techniques to facilitate manual creation of computer-interpretable descriptors have been proposed [20].

We further contrast different type of manual approaches 1) by the form of the descriptors created to represent the knowledge resources and 2) by the person who creates those descriptors. Most existing approaches utilizing descriptors that are more sophisticated than simple tag lists depend on expert knowledge curators to create those descriptors. However, it has been suggested that currently available web technologies could be used to "wrap" computer-interpretable knowledge representation languages with intuitive user-friendly authoring tools [11]. The ontologies that have emerged in the context of the Semantic Web could be used as formal knowledge representation languages for humans to author descriptors that enable computers to act as more effective and "intelligent" matchmakers for exchanging and integrating different forms of scientific knowledge [6], [44], [41], [29], [2], [15]. This could enable the original knowledge creators to author their own computer-interpretable descriptors in a reliable and semantically rich manner. And if computer interpretable descriptors of knowledge resources were provided by the original human

creators, then it should be possible to achieve a more effective and accurate forms of computer-aided knowledge sharing [19], [42], [7], [14], [37], [3], [8], [9]. Engaging the entire research community in the creation of computer-understandable descriptors is a scalable solution to the problem of curating all of the research articles related to a particular knowledge domain such as sustainability science [5], [16], [11]. To make this kind of "crowd sourcing" happen, researchers would have to be provided with some form of incentive for creating their own computer-understandable descriptors, perhaps by making their creation a part of the process of submitting research articles or grant proposals.

3 Methods

The basic concept of connecting students to researchers at their university via textbooks is shown in figure 1. Semantic statements are created both for passages from the textbook and for research articles and other shared resources created by researchers to express their expert knowledge. The EKOSS semantic matching algorithm is then used to create semantic links between the textbook statements and the researcher statements, as shown by black curved arrows (figure 1). The links can be traversed in either direction, enabling a person reading a research article that has been linked to a textbook to refer to basic knowledge in the textbook related to the topics mentioned in the article.

For our application of the University on Textbook system to the domain of sustainability science, we have chosen the textbook "Vision 2050: A Roadmap for a Sustainable Earth" [23]. We selected this textbook in part because it is freely accessible from SpringerLink and in part because it covers many of the topics of research that are the focus of researchers at the University of Tokyo. The book, which is written in English, has 162 pages, 8 chapters, and 33 black and white figures. We created semantic statements for about 70 passages from the book ranging from one sentence to several paragraphs in length. We then determined the semantic similarity between those semantic statements and about 200 semantic statements describing research articles and web pages by researchers studying issues related to sustainability science. We calculated the similarities by using a semantic matching algorithm and scoring system developed in previous work, which utilizes logic and rule based inference together with inverse term and document frequencies [22].

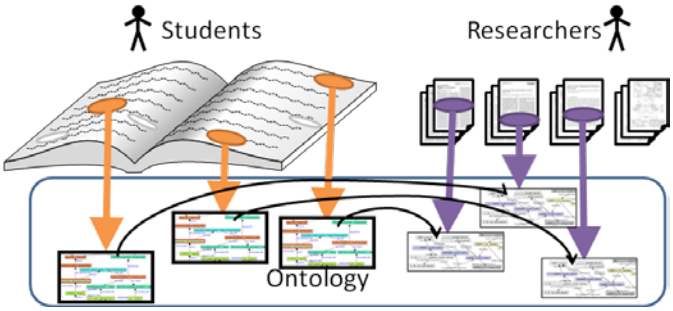

Figure 1: Bridging the knowledge gap between undergraduate students and university researchers by using ontologies to link textbooks and research articles

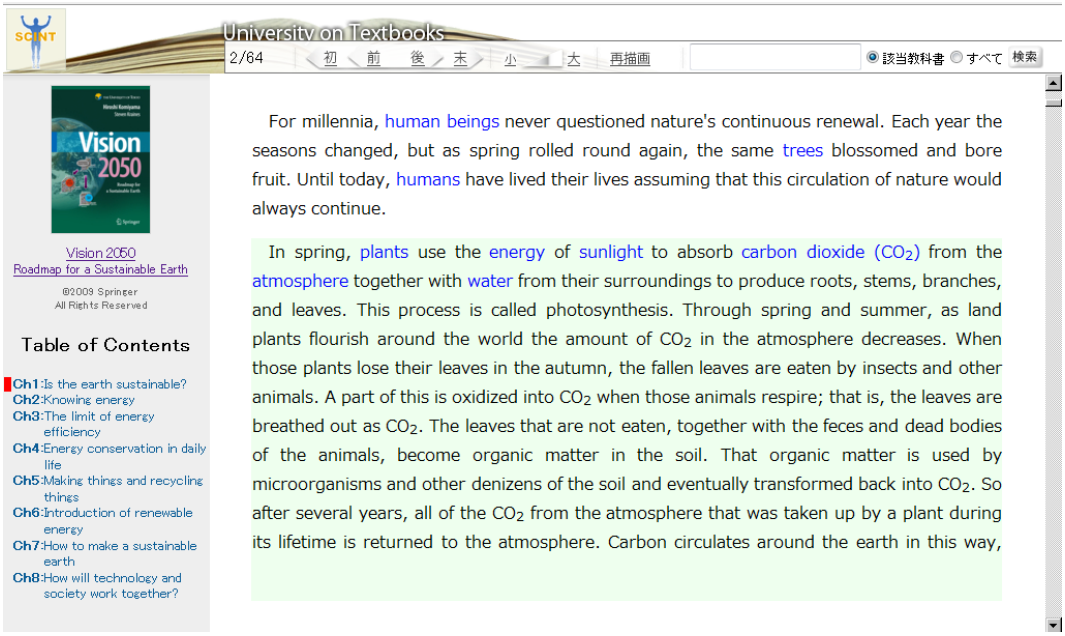

Figure 2: A view of the UoT Textbook Browser with a large font size selected showing the passage from chapter 1 of "Vision 2050". Terms in blue font are keywords that have been mapped to classes from the ontology. The light green background indicates that the passage has an associated semantic statement.

4 Results

The University on Textbook system uses a web-based multi-column paginating textbook browser, which we call the UoT Textbook Browser, to provide the user interface for reading the selected textbook and accessing researchers via the semantic links. The text flow control, search features, and popup information windows are all implemented using Javascript, so that the UoT Textbook Browser can function on standard web browsers. Examples of the basic UoT Textbook Browser views are shown in figures 2 and 3.

The UoT textbook browser supports layout arrangement, text justification and hyphenation, handling of dangling punctuations, and a pop-up window management system for showing information on ontology classes and researchers that are obtained via web services provided by the EKOSS system. The EKOSS system provides two main web services for giving information on classes and researchers to the UoT textbook browser. The first web service provides information about a class from the ontology, including a short definition of the class, the number of times the class has been used in the semantic statements handled by the EKOSS system, the number of researchers using that class, the number of research articles in which instances of that class appear, and contact and affiliation information for each researcher who is an author of one of the research articles. The last type of information is helpful for assisting undergraduate students in finding thesis advisors or graduate school laboratories to which to apply. An example is shown in figure 4.

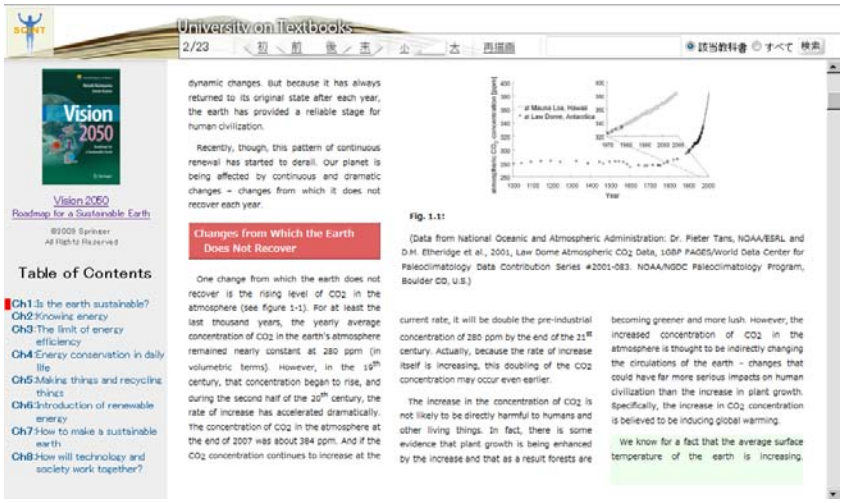

Figure 3: A view of the UoT Textbook Browser with a small font size selected. The textbook browser automatically repaginates to multiple columns and repositions figures.

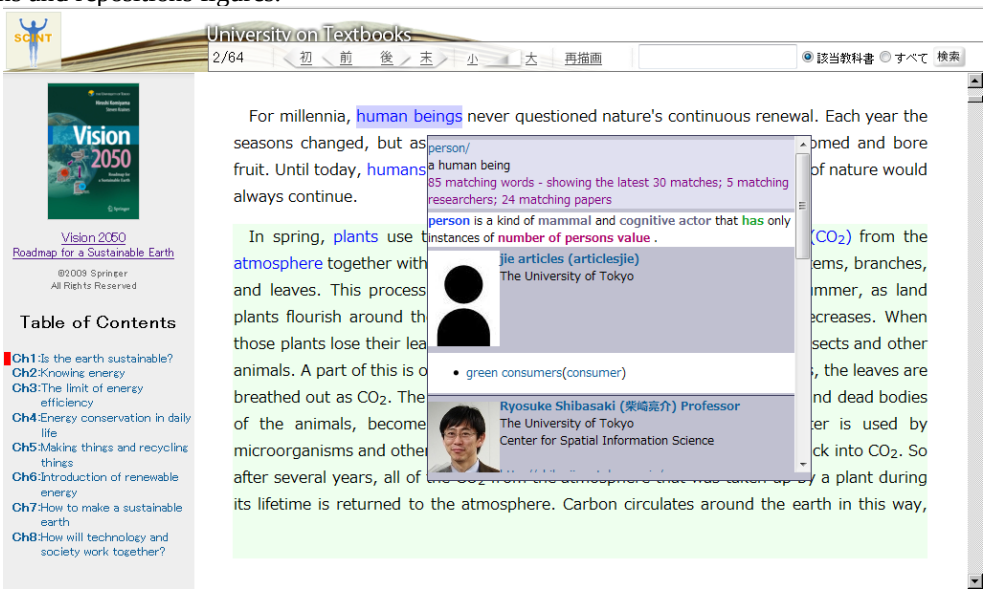

Figure 4: Popup window showing information provided by the first EKOSS web service in response to a query for a class from the ontology. A short definition and usage statistics of the class are given, followed by a list of researchers using the class or a subclass in their research articles. The researcher information includes links to a researcher information page giving contact information and lists of semantic statements, the EKOSS ontology class browser, and the actual semantic statement for the research article.

Figure 5: View of the UoT Textbook Browser showing the popup window containing information in three tags that is provided by the second EKOSS web service for a semantic statement describing the highlighted passage in the textbook. The tab that is shown lists the researchers who have written research articles whose semantic statements are semantically similar to the textbook statement. Each researcher description includes links to 1) a researcher information page giving contact information and lists of semantic statements, 2) the matching research article, and 3) the research article's semantic statement. The "subject verb object" triples from the research article statement that matched with the textbook statement are rendered in natural language.

The second EKOSS web service, shown in figure 5, provides information and semantically linked research articles for a passage from the textbook, for which a semantic statement has been created. The information includes an image file of the graph representation of the semantic statement and a formatted natural language representation generated by the EKOSS NLG algorithm [25]. The semantically linked articles are those that have semantic statements with sufficient semantic similarity to the textbook semantic statement, as calculated by the EKOSS semantic matching algorithm. The results of the semantic matching (the tab that is currently shown in figure 5) are given as a list of researchers who have authored one of the matched semantic statements. For each matching researcher, the web service provides affiliation information, the name and link of the matching research article and its semantic statement, and a rendering in natural language of the "subject verb object" triples from the semantic statement that were found to match with the textbook statement.

5 Discussion

There are a number of possible applications of the University on Textbook system in undergraduate education. The driving application for us has been to help second year undergraduate students decide which department they would like to join, to help third year students decide with which specific research professors they would like to do their undergraduate thesis work, and to help fourth year students select laboratories in which they would like to do their graduate study. For that reason, we have focused on providing information that enables students to access the research faculty at the university who have written research articles mapped to a particular passage from the textbook. However, because the UoT textbook browser is implemented in standard Javascript, it is possible to extend its functionality by integrating different multi-media, semantic web and web 2.0 technologies. For example, support could be added for multimedia such as animated movies, 3D images, and interactive "game" applications (e.g. [34]). Also, because it is possible to annotate text and use hyperlinks freely, social network system technologies could be used to provide interactive and user-customization features such as annotation, bookmarking, and "just in time teaching" mechanisms, e.g. to provide a feedback mechanism for the students to post questions or comments [30].

This linking of "person to person" is also in line with the underlying goal of the University on Textbook project that we touched on in the introduction, which is to motivate researchers to create their own semantic statements describing their research. The notion is that although

research faculty at universities are eager to attract students to their laboratories, they are often overwhelmed by research-oriented tasks and do not have time to communicate with the undergraduate students directly. The University on Textbook system enables those faculty members to use their regular research publications as advertising material to the undergraduate student population for the (ideally small) added cost of creating a semantic statement. The semantic statement could then be used in other "intelligent" computer services, such as semantic search [28], knowledge mining [27], natural language generation [25], recommender systems [21] and matching with peer reviewers.

In future work, we will attempt to increase the number of semantic statements describing research articles to link with the University on Textbook system. A critical part of this work is the development of authoring tools that utilize natural language processing techniques to help researchers to create semantic statements describing their research publications easily and accurately. We also plan to add additional textbooks to the University on Textbook system from life sciences, sustainability science, and other domains of science. Finally, we are planning a study to evaluate the effectiveness of the system in helping students at the University of Tokyo identify research faculty with whom they would like to do their undergraduate thesis and graduate research work.

6 Acknowledgements

Funding support for this research was provided by the President's Office of the University of Tokyo and the Japan Office of the Alliance for Global Sustainability. The UoT textbook browser was implemented by NalaPro Technologies.

7 References

[1] Alberts, B. (2008): Shortcuts to medical progress? Science 319(5871), 1733.

[2] Allenby, B. (2006): The ontologies of industrial ecology? Progress in Industrial Ecology. An International Journal 3(1): 28-40.

[3] Androutsopoulos, I., Oberlander, J., and Karkaletsis, V. (2007): Source authoring for multilingual generation of personalised object descriptions. Natural Language Engineering, vol 13(3) pp. 191-233

[4] Appleton, L. (2005): Using electronic textbooks: promoting, placing and embedding. The Electronic Library 23(1), 54–63.

[5] Attwood, T.K., Kell, D.B., McDermott, P., Marsh, J., Pettifer, S.R., Thorne, D. (2009): Calling International Rescue: knowledge lost in literature and data landslide! Biochemical Journal vol 242, pp. 317-333

[6] Berners-Lee, T. and Hendler, J. (2001): Publishing on the Semantic Web. Nature, 410: 1023—1024.

[7] Biller, O., Elhadad, M., Netzer, Y., Univerisy, B. G. (2005): Interactive authoring of logical forms for multilingual generation. In Proc 10th European Workshop on Natural Language Generation, pp. 24-31

[8] Bizer, C., Heese, R., Mochol, M., Oldakowski, R., Tolksdorf, R., Berlin F. U., and Eckstein, R. (2005): The Impact of Semantic Web Technologies on Job Recruitment Processes. In Proceedings of the 7th International Conference Wirtschaftsinformatik, pp. 1367-1383

[9] Bradley, J. (2008): Pliny: A model for digital support of scholarship. Journal of Digital Information, 9(1)

[10] Brusilovsky, P., Chavan, G., and Farzan, R. (2004): Social adaptive navigation support for open corpus electronic textbooks. Adaptive Hypermedia and Adaptive Web-Based Systems 3137, 24-33.

[11] Ceol, A., Chatr-Aryamontri, A., Licata, L., and Cesareni, G., (2008): Linking Entries in Protein Interaction Database to Structured Text: the FEBS Letters Experiment. FEBS letters, 582(8), 1171-1177.

[12] Chua, A.Y.K., Lam, W., and Majid, S. (2006): Knowledge reuse in action: the case of CALL. J. Information Science 32(3), 251-260.

[13] Crestani, F., and Melucci, M. (2003): Automatic construction of hypertexts for self-referencing: the hyper-textbook project. Information Systems 28(7), 769–790.

[14] Davis, B., Iqbal, A. A., Funk, A., Tablan, V., Bontcheva, K., Cunningham, H., and Handschuh, S. (2008): RoundTrip Ontology Authoring. Intl Semantic Web Conf, ed. Amit P. Sheth et al., LNCS vol 5318 pp. 50-65

[15] Davis, C., Nikolic, I. and Dijkema, G. P. J. (2009): Integration of Life Cycle Assessment Into Agent-Based Modeling. Journal of Industrial Ecology, 13: 306-325.

[16] DeRose, P., Shen, W., Chen, F., Doan, A., and Ramakrishnan, R. (2007): Building structured web community portals: a top-down, compositional, and incremental approach. In VLDB '07: Proc 33rd Intl Conf on Very Large Data Bases, Vienna, Austria, pp. 399-410.

[17] Flaspohler, M. R., Rux, E. M., and Flaspohler, J. A. (2007): The annotated bibliography and citation behavior: enhancing student scholarship in an undergraduate biology course. CBE Life Sci. Educ. 6, 350-360.

[18] Garfield, E. (2001): A retrospective and prospective view of information retrieval and artificial intelligence in the 21st century. JASIST 52 (1), 18-21.

[19] Gerstein, M., Seringhaus, M. and Fields, S. (2007): Structured digital abstract makes text mining easy. Nature, 447: 142.

[20] Goldstein-Stewart, J. and Winder, R.K. (2009): Designing a system for semi-automatic population of knowledge bases from unstructured text. In Proceedings of the Intl Conf of Knowledge Discovery and Information Retrieval (KDIR 2009) Madeira, Portugal, 2009. pp. 88-99

[21] Guo, W. and Kraines S.B. (2010): Semantic content-based recommendations using semantic graphs, Arabnia, Hamid R. (ed.) Advances in Comp. Biology, Advances in Experimental Medicine and Biology, 2010, Vol. 680, Part 7, pp: 653-659, Springer Science+Business Media, LLC 2010.

[22] Guo, W. and Kraines S.B. (2008): Explicit scientific knowledge comparison based on semantic description matching. ASIST Annual meeting 2008, Columbus, Ohio.

[23] Komiyama, H. and Kraines, S.B. (2008): Vision 2050: A Roadmap for a Sustainable Earth. Springer.

[24] Kostoff, R.N. (2002): Overcoming specialization. BioScience 52(10), 937-941(5).

[25] Kraines, S.B. and Guo, W. (2009): Using Human Authored Description Logics ABoxes as Concept Models for Natural Language Generation, ASIST 2009 Annual Meeting, Vancouver, British Columbia, Canada.

[26] Kraines, S.B., Makino, T., Guo, W., Mizutani, H. and Takagi, T. (2010): Bridging the Knowledge Gap between Research and Education through Textbooks, In: X. Luo et al. (Eds.): ICWL 2010, LNCS 6483, pp. 121-130.

[27] Kraines, S.B., Guo, W., Hoshiyama, D., Mizutani, H., Takagi, T. (2010): Generating Literature-based Knowledge Discoveries in Life Sciences Using Relationship Associations, Intl Conf. on Knowledge Discovery and Information Retrieval, KDIR 2010, Valencia, Spain, pp: 35-44.

[28] Kraines, S.B., Guo, W., Kemper, B. and Nakamura, Y. (2006): EKOSS: A knowledge-user centered approach to knowledge sharing, discovery, and integration on the Semantic Web. LNCS, Vol 4273, ISWC 2006: 833-2091.

[29] Kraines, S.B., Batres, R., Koyama, M., Wallace, D.R. and Komiyama, H. (2005): Internet-based integrated environmental assessment: using ontologies to share computational models. J. Industrial Ecology, 9: 31-50.

[30] Marrs, K. A., and Novak G. (2004): Just-in-time teaching in biology: creating an active learner classroom using the Internet. Cell Biol. Educ. 3, 49-61.

[31] Mayer, R. E., and Moreno, R. (2002): Animation as an aid to multimedia learning. Educ. Psychol. Rev. 14(1), 87-99.

[32] McClean, P., Johnson, C., Rogers, R., Daniels, L., Reber, J., Slator, B. M., Terpstra, J., and White, A. (2005): Molecular and cellular biology animations: development and impact on student learning. Cell Biol. Educ. 4, 169-179.

[33] McFall, R. (2005): Electronic textbooks that transform how textbooks are used, The Electronic Library 23(1), 72-81.

[34] Miller, L., Moreno, J., Willcockson, I., Smith, D., and Mayes, J. (2006): An online, interactive approach to teaching neuroscience to adolescents. CBE Life Sci. Educ. 5, 137-143.

[35] Mons, B., Ashburner, M., Chichester, C., van Mulligen, E., Weeber, M., den Dunnen, J., van Ommen, G.J., Musen, M., Cockerill, M., Hermjakob, H., Mons, A., Packer, A., Pacheco, R., Lewis, S., Berkeley, A., Melton, W., Barris, N., Wales, J., Meijssen, G., Moeller, E., Roes, P.J., Borner, K., and Bairoch, A. (2008): Calling on a million minds for community annotation in WikiProteins. Genome Biol. 9(5), R89.

[36] Porter, J. R. (2005): Information literacy in biology education: an example from an advanced cell biology course. Cell Biol. Educ. 4, 335-343.

[37] Power, R. (2009): Towards a generation-based semantic web authoring tool. In Proc 12th European Workshop on Natural Language Generation, Athens, Greece, pp. 9-15

[38] Ruttenberg, A., Clark, T., Bug, W., Samwald, M., Bodenreider, O., Chen, H., Doherty, D., Forsberg, K., Gao, Y., Kashyap, V., Kinoshita, J., Luciano, J., Marshall, M.S., Ogbuji, C., Rees, J., Stephens, S., Wong, G.T., Wu, E., Zaccagnini, D., Hongsermeier, T., Neumann, E., Herman, I., and Cheung, K.-H. (2007): Advancing translational research with the Semantic Web, BMC Bioinformatics 8(Suppl 3), S2.

[39] Razmerita, L., Firantas, R., and Jusevicius, M. (2009): Towards a new generation of social networks: merging social web with Semantic Web. 4th AIS SigPrag Intl Pragmatic Web Conf. Track. http://i-semantics.tugraz.at/2009/papers/owards_a_new_generation_of_social_networks.pdf

[40] Safari Books Online (2009): Safari Books Online – Home. Available at http://www.safaribooksonline.com.

[41] Shotton, D., Portwin, K., Klyne, G., and Miles, A. (2009): Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. PLoS Comput Biol, Vol. 5 pp. e1000361

[42] Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E. and Ciravegna, F. (2006): Semantic annotation for knowledge management: requirements and a survey of the state of the art. Web Semantics: Science, Services and Agents on the World Wide Web, 4 (1): 14—28.

[43] Vassiliou, M., and Rowley, J. (2008): Progressing the definition of “e-book”. Library Hi Tech 26(3), 355-368.

[44] Wang, X., Gorlitsky, R. and Almeida, J. S. (2005): From XML to RDF: how semantic web technologies will change the design of ‘omic’ standards. Nature Biotechnology, 23 (9): 1099-1103.

[45] Wilson, R., Landoni, M., and Gibb, F. (2003): The WEB Book experiments in electronic textbook design. Journal of Documentation 59(4), 454-477.

[46] Weikum, G., Kasneci, G., Ramanath, M., Suchanek, F. (2009): Database and Information-retrieval Methods for Knowledge Discovery. Comm. of the ACM, 4, 56-64.