

Keywords-based Recommendation System: A Case Study of Civic Chemical Literacy

Pei-Jung Lin¹, Chin-Cheng Chou^{2*}

¹ *Department of Computer Science and Information Engineering,
Hungkuang University, Taiwan*

² *Department of Science Education, National Taipei University of Education, Taiwan
e-mail: ccchou62@gmail.com*

Abstract

This study developed a keywords-based recommendation system for integrating the media cloud with education cloud, and is capable of assessing the correlations between horizontal and vertical curriculum integration. The system automatically links course outlines, newspaper media, and keywords, to understand the order in which the teaching of a concept is developed. The system can be used to search for previous and future teaching materials with regard to specific concepts. Linking everyday life to education and integrating interdisciplinary learning is the ideal orientation for curriculum design. However, identifying the concepts needed in daily life is a difficult task, particularly with rapid and continuous changes in the environment. This article is an interest research report, on the development of a Teaching Keywords-based Recommendation System (TKRS) that analyses the relationship between media cloud data and teaching keywords.

Keywords: Curriculum mapping, Newspaper Media

Introduction

Scientific literacy related to the media and public is a subject of continual interest and exploration. For non-major science students, general education courses in university are the most likely path to learn science. Over the past three years, universities of Taiwan have begun promoting school-wide curriculum mapping, the purpose of which is to more effectively integrate general education courses with professional courses in each department of study. Courses must also be highly integrated and computerized curriculum mapping system aids in curriculum design and development. Science curriculum in high schools is not only designed to prepare students for future studies in science, but also more importantly to provide a link between students in all fields and the science literacy of the general public. High school chemistry is the part of the national education responsible for the cultivation of civic literacy. The objective of high school chemistry courses should be to familiarize students with basic concepts related to chemistry literacy. Chemistry courses must be integrated with everyday life, media reports, and the requirements of chemical literacy in the

future. This study used data from the news media database to analyze the relationship between chemistry courses and news reports and investigate the correlation between textbook chemistry and chemistry in everyday life. For example, the school curriculum describes various causes of chemical pollution but fails to explain the actions required to deal with pollution. Likewise, the curriculum explains the principle of how water molecules are heated by microwaves, but does not outline the precautions required when using microwaves. Another example is the labeling of food products as trans fats; however, few students understand the significance of this label. Education authorities in Taiwan should link the national curriculum related to chemistry more closely to everyday practices and news reports, improve the integration of chemistry courses with other studies, and help students establish criteria for their actions when facing chemistry-related scenarios in daily life.

Linking everyday life to education and integrating interdisciplinary learning is the ideal orientation for curriculum design. However, identifying the concepts needed in daily life is a difficult task, particularly with rapid and continuous changes in the environment. For example, plasticizer is a term that has emerged only in recent years. Experts in different domains may differ in their selection of learning concepts that must be taken into account. Lexical analysis of news reports is applicable in the chemistry courses design and development. Pankong *et al.* (2012) combined lexical with semantic analysis to develop a system of analysis for use in social networks. Lee *et al.* (2011) employed lexical analysis to develop a recommendation system for social networks. By analyzing the words frequently used by every user in a social network, the system recommends other users with interests in similar topics. Cloud technology is currently applied in many domains. Raj and Mala (2012) developed a system for the extraction of news data from online media according to the interests of the user. In an exploration of the effects of cloud technologies on information engineering courses, Vaquero (2011) found that Platform-as-a-Service (PaaS) cloud technology made the greatest contribution to courses in information engineering. Alabbadi (2011) proposed a new concept for the combination of cloud technology with education, known as Education and Learning as a Service (ELaaS). This concept chiefly emphasizes the use of cloud technology to develop new education and learning services to enhance the efficiency of institutional education.

UNESCO proposed for teachers a program on media and information literacy (Wilson, 2012). The most important step prior to implementing the program is to analyze the needs of students. This program includes: concepts and organizational framework, product and use information, media text and media sources, evaluation and analysis, media audiences, democratic dialogue and social participation, the orientation of curriculum adaptation, and classroom teaching methods. Gutierrez and Tyner (2012) pointed out the five areas of media and information literacy as defined by UNESCO (understanding, critical thinking, creativity, cultural awareness, and civic literacy), emphasizing the importance of conventional education curriculum in media integration. But, Klosterman *et al.* (2012) noted that few studies have researched the use of mass media by science

teachers to highlight socio-scientific and sustainability issues, emphasizing how the relationship between socio-scientific issues and education should be improved. Mizuno *et al.* (2010) pointed out that during the outbreak of the Serratia infection, news frequency (number of articles and word count) gradually increased, reaching a peak after three days before beginning to decline. The content of initial news reports on sporadic events is often erroneous, but further research would be required to determine how subsequently corrected reports have influenced the judgment of readers with regard to these events. In relation to information and media literacy, Operti (2009) claimed that the educational sector should devise ways to empower students to take the initiative in social participation and try to make school curricula more relevant to real life scenarios. That study described the use of a “glo-local” curriculum to encourage students to give back to the earth and their communities. Lau and Cortes (2009) indicated that the use of information is a primary resource in personal development and well-being, emphasizing the need for technological advancement and a modern educational model based on interdisciplinary learning. Masud *et al.* (2012) and Mohammed *et al.* (2010) claimed that applying cloud technology to e-learning could improve learning effectiveness and reduce the costs associated with developing educational platforms. Kim *et al.* (2011) developed a content-oriented smart e-learning model using cloud computation as its core environment to provide a learning services platform capable of facilitating individualization and customization.

Proposed Teaching Keywords-based Recommendation System (TKRS)

This study developed a teaching Keywords-based Recommendation System (TKRS) to analyse the relationship between media cloud data and teaching Keywords. This integrated search system comprises two sub-systems: An integrated search system for hard copy news archives, and an integrated search system for course outlines. Integrating the two sub-systems results in a comprehensive literacy-based keyword search and analysis system for news media and course outlines.

Integrated system structure

This study analysed the corresponding relationship between teaching keywords and the frequency with which these keywords have appeared in domestic and international media during the past five years. We developed an integrated system for searching and analysing keywords associations among news media and course outline. The TKRS can help curriculum designers to understand, within a short period of time, the relevance of a concept to the daily news, course outlines, and textbooks. Figure 1 outlines the system architecture. In the following, each module is described with regard to its operational methods and functionality. This integrated system for combined TKRS, Education Cloud, and Media Cloud in which data related to keywords of interest

are retrieved from major domestic and international media through the clouds.

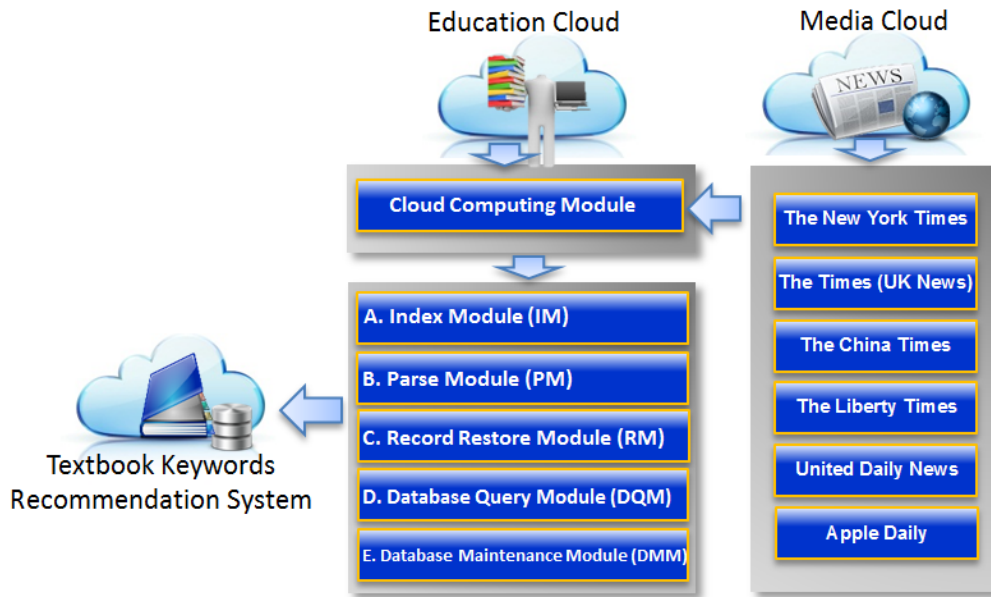


Figure 1. Integrated System Architecture

A. Index Module (IM)

The Index Module is used to input keywords collected by users into the TKRS from the news database of the Media Cloud. To improve the efficiency of indexing, the TKRS first determines whether each word has previously been searched. The Module IM is configured to retrieve a substantial volume of data from credible historic media organizations. These media sources include “The New York Times”, “The Times (UK News)”, “The China Times”, “The Liberty Times”, “United Daily News”, and the “Apple Daily”. After the TKRS has imported vocabulary into these media cloud, a fuzzy search of the keywords from internal databases is conducted. The source of indexing is the articles that have been published annually by each media organization. This module conveys to the TKRS how frequently these keywords have been used in the news according to the appearance of particular vocabulary. After comparing the retrieval results of a single word from multiple news organizations, similar trends were observed with regard to the percentage of news articles in which keywords appear. From this we can deduce that the search results of this system have an acceptable degree of accuracy and reference value.

B. Parse Module (PM)

The Parse Module (PM) converts and integrates the data extracted from each media source in the TKRS. Because the cloud of data each domestic and international media source differs in its format, the information transmitted back to the TKRS from each source also differs, thereby necessitating the analysis of HTML data from each media source. First, e-media data from each source is analysed to filter out redundant data from HTML or XML documents. This prevents the data from growing too large and reducing search efficiency. Finally, the data is deconstructed and

converted into a code format identifiable to the TKRS. An information table is produced for each media source using the coded data, whereupon the keywords and year are sequentially converted into system.

C. Record Restore Module (RM)

This module is responsible for saving the analysis results according to data content and format into the data tables for each media source. The search values for keywords and year are incorporated into these tables, which sequentially record all keywords inputted by the user to date, to facilitate future searches. The required keyword search results can be computed and compared once they have been uploaded to the database. The data search function converts the keywords into SQL search commands. The TKRS determines whether the keyword has been previously queried. If not, the RM Module connects to the cloud end of each media source and extracts information from cloud. If the keyword has previously been queried, the TKRS extracts data directly from its own and converts the SQL search results into HTML to allow users to view the data.

D. Database Query Module (DQM)

The keywords input by users are displayed according to their annual frequency of appearance. The TKRS automatically calculates the proportional appearance of keywords in any given year by dividing the total number of appearances for a designated year by 365. For example, a search for the term “nuclear energy” appeared 719 times in 2011. Dividing this number by 365 indicates that this keyword appeared approximately 2 times per day during that period, as an indication of its popularity. We can also determine how a keyword or term has changed in recent years. Comparing data obtained from the same media source over different years can reveal the relative popularity of a keyword at any given time. The statistical results of this study show that despite differences in the data obtained from different media clouds in a single year, a comparison of media sources over several years can reveal trends within these changes.

E. Database Maintenance Module (DMM)

To facilitate management, all keywords that have been searched for can be modified or deleted from cloud. The content values in the data tables of each media source corresponding to a certain keyword can also be deleted and internal data can be statistically compared to produce results reports. This DMM Module operates through cloud computing, which refers to the online accessing of data and computation of resources without the need for users to set up their own work platform. Simply inputting work through the execution end of the TKRS enables users to obtain calculation results using a browser or specific interface. In addition, users do not need to deal with maintenance or system update, which makes operations simpler and reduces the cost of the IT environment. For example, users can utilize mobile devices to connect to the cloud end for obtain relevant data.

Integrated search system for hard copy news archives

Existing newspaper databases can aid researchers to identify terms that frequently appear in the news and the approximate frequency of usage in a given year. These databases also assist educators in the identification of key vocabulary that should be emphasized in curriculum design. This study compared the number of news articles on nuclear energy that have appeared in the New York Times and UK Times over the past five years with the number of similar articles in Taiwan’s Liberty Times, Apple Daily, United Daily News, and China Times. Our results indicated differences in the rate of change among these newspapers. The number of articles related to nuclear energy published by the four main newspapers in Taiwan declined slightly from 2008-2010, drastically increased in 2011 after the Fukushima nuclear disaster in Japan, and then declined slightly again in 2012. For example, the number of articles related to nuclear energy published in the Liberty Times was 85 in 2010, 719 in 2011, and 291 in 2012. Following the Fukushima nuclear disaster in Japan in 2011, newspapers in Taiwan published a significantly greater number of articles related to nuclear energy. In the New York Times and UK Times, the number of articles related to nuclear energy published in 2011 was slightly higher than the number in 2010 but the growth was relatively limited. We infer that this may be due to the relative proximity of Japan to Taiwan compared to the US and Europe. The number of articles related to nuclear energy in the UK Times was 612 in 2010, 635 in 2011, and 561 in 2012. Table 1 below shows the number with which articles on nuclear energy appeared during the period 2008-2012.

Table 1. Number of articles related to nuclear energy published in the past five years

News paper Year	The New York Times	The Times (UK News)	The Liberty Times	Apple Daily	United Daily News	The China Times
2008	3510	873	101	45	93	113
2009	3790	763	98	33	119	77
2010	3250	612	85	21	101	72
2011	3620	635	719	241	487	426
2012	3900	561	291	98	162	166

Inter-grade and inter-disciplinary course outline search system

The instruction of particular concepts is often scattered throughout different courses at different grade levels. Teachers often struggle to understand the relationship between a particular course concept and other subjects. The inter-grade and inter-disciplinary course outline search system assists teachers to understand concept-based K-12 curriculum distribution. As shown in Figure 2, the concept of nuclear energy is first introduced in relation to the concept of sustainable development in the science and life technology courses of 5th and 6th grade students. These courses outline the development and utilization of energy and describe how various power generation resources in Taiwan depend on imports (such as thermal power and nuclear power). These courses also discuss power generation capacity in recent years and the extent to which various energy sources are used to generate power in Taiwan (thermal power, nuclear energy, hydro power). As part of the 7th-9th

grade curriculum, sustainable development is explored and the advantages, disadvantages and uses of various power sources (thermal power, nuclear energy, hydro power, solar power, and gasoline) are discussed with regard to their influence on society, ecology, and the environment. Basic physics courses in the 10th grade explain the splitting of nuclei and the generation of nuclear energy and radiation safety. Basic chemistry courses at this level describe the uses of various energy resources such as light energy, solar energy, nuclear energy and biomass energy, in everyday life. The inter-grade and inter-disciplinary course search system can inform teachers of the content related to nuclear energy in the previous and following curriculum. This is meant to aid in the design of current lessons to ensure that courses are not overly complex, excessively easy, or unduly repetitive.

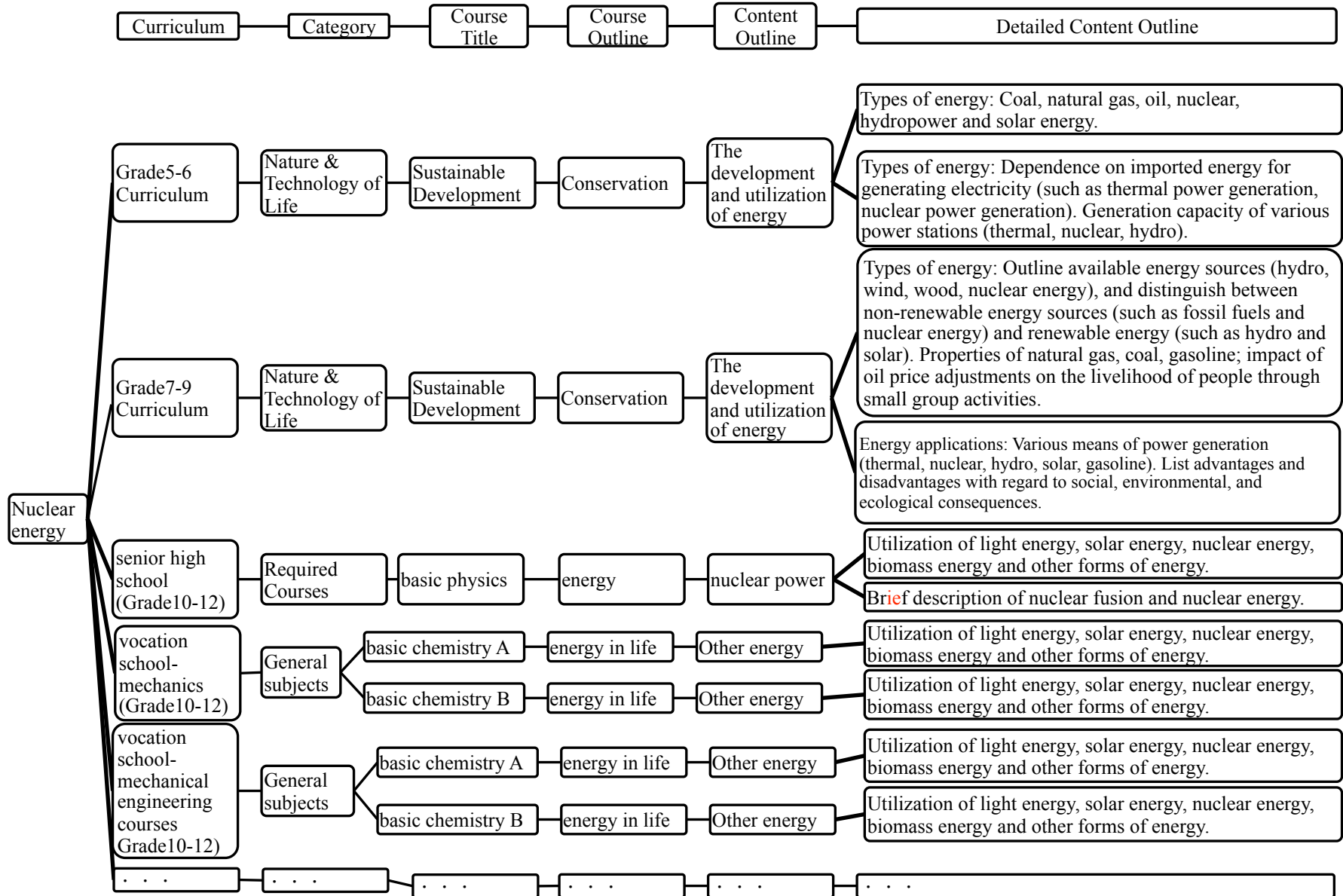


Figure 1. Nuclear energy-based integrated curriculum structure

Conclusions and Recommendations

The greatest challenge faced by individuals designing the unit themes of an integrated curriculum is unraveling the complex relationships between different grades in an environment of shifting curriculums and learning materials. Educators must be aware of the materials that have previously been studied and are yet to be studied in order to determine the appropriate depth and breadth of the current lesson. In particular, the link between a unit theme and popular news media influences how course designers plan teaching content and methods. General education courses taught in university have important implications for education for civic literacy and are crucial to integrating real life scenarios into teaching. As indicated by Shwartz, Ben-Zvi and Hofstein (2006), situational literacy in chemistry entails (1) understanding the importance of chemistry-related knowledge to interpreting daily life scenarios; (2) the ability to evaluate new products from the perspective of chemistry and participate in chemistry-related issues in society; (3) understanding the relevance of developments in chemistry with regard to social change. This study recommended a system for developing unit by integrating teaching keywords with those found in the media cloud and education cloud.

Through course outlines and news media, this study developed an teaching keywords-based recommendation system for integrating the media cloud with education cloud, capable of assessing the correlations between horizontal and vertical curriculum integration. The aim was to determine how the words related to a specific concept are presented in the curriculum and integrate this with course outlines. Further, this study automatically linked course outlines, newspaper media, and keywords, in order to understand the order in which the teaching of a concept is developed. The system can be used to search for previous and future teaching materials with regard to specific concepts.

This study explored how to use news media to build importance indicators for chemistry-related concepts. In the future, we hope to analyse a greater number of chemistry-related keywords in terms of the relationship between course outlines and news reports. The proposed system can be used to design foundational chemistry courses that are more student-friendly. In addition, the links between basic chemistry-related concepts and everyday chemistry must be established to enable students to use their knowledge of chemistry in everyday decision making. Many students have misconceptions of everyday chemistry themes, and future chemistry-related general education courses must work to correct these misconceptions. A more advanced integrated curriculum search system is currently being researched and developed. In the future this system could be applied to other academic subjects as well.

The authors would like to thank the National Science Council of Taiwan for financially supporting this research under Contract No. NSC100-2511-S-241-006-MY2.

References

Alabbadi, M. M. (2011). "Cloud computing for education and learning: Education and learning as a service (ELaaS)," *Proceedings of the 2011 14th International Conference on Interactive Collaborative Learning (ICL)*, 589-594.

Gutierrez, A.; Tyner, K. (2012). "Media Education, Media Literacy and Digital Competence", *Comunicar*, **38**, 31-39.

Kim, S.; Song, S. M.; Yoon, Y. (2011). "Smart Learning Services based on Smart Cloud Computing", *International Journal Sensor*, **11**, 7835-7850.

Klosterman, M. L.; Sadler, T. D.; Brown, J. (2012). "Science Teachers' Use of Mass Media to Address Socio-Scientific and Sustainability Issues", *Research in Science Education*, **42** (1), 51-74.

Lau, J.; Cortes, J. (2009), "Information Skills: Conceptual Convergence between Information and Communication Sciences", *Comunicar*, **32**, 21-30.

Lee, S. S.; Chung, T.; McLeod, D. (2011). "Dynamic Item Recommendation by Topic Modeling for Social Networks," *Proceedings of the 2011 Eighth International Conference on Information Technology: New Generations (ITNG)*, 884-889.

Masud, A. H.; Huang, X. (2012). "An E-learning System Architecture based on Cloud Computing", *World Academy of Science Engineering and Technology*, **62**, 74-78.

Mizuno, Y.; Narimatsu, H.; Kishi, Y.; Kodama, Y.; Murashige, N.; Yuji, K.; Kami, M. (2010). "Structural problems of medical news reports in newspapers: a verification of news reports on an incident of mass nosocomial *Serratia* infection", *Journal of Infection and Chemotherapy*, **16** (2), 107-112.

Mohammed, A. Z.; Samir, A. E. S.; Mudasser, F. W. (2010). "Cloud Computing Based E-Learning System," *International Journal of Distance Education Technologies (IJDET)*, **8**(2), 58-71.

Opertti, R. (2009). "Curricular Contribution to Media Education: a Work in Progress", *Comunicar*, **32**, 31-40.

Pankong, N.; Prakanchaen, S.; Buranarach, M. (2012). "A combined semantic social network analysis framework to integrate social media data," *Proceedings of the 2012 4th International Conference on Knowledge and Smart Technology (KST)*, 37-42.

Raj, D. A. A.; Mala, T. (2012). "Cloud Press: A next generation news retrieval system on the cloud," *Proceedings of the 2012 International Conference on Recent Advances in Computing and Software Systems (RACSS)*, 299-304.

Shwartz, Y., R.; Ben-Zvi,; *et al.*. (2006), “Chemical literacy: What does this mean to scientists and school teachers?” *Journal of Chemical Education*, **83** (10), 1557-1561.

Vaquero, L. M. (2011). “EduCloud: PaaS versus IaaS Cloud Usage for an Advanced Computer Science Course,” *IEEE Transactions on Education*, **54** (4), 590-598.

Wilson, C. (2012). “Media and Information Literacy: Pedagogy and Possibilities”, *Comunicar* **39**, 15-22.