Language Technologies for Enhancement of Teaching and Learning in Writing

Tak Pang Lau, Shuai Wang, Yuanyuan Man, Chi Fai Yuen, and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
{tplau,wangs,yyman,cfyuen,king}@cse.cuhk.edu.hk

ABSTRACT

Writing is a vital issue for education as well as a fundamental skill in teaching and learning. With the development of information technologies, more and more professional writing tools emerge. As each of them mostly concentrates on addressing a specific issue, people need a one-stop platform, which could integrate multiply functions. In addition, with the supported concept of e-learning ecosystem for future education, a comprehensive platform will be more promising. Therefore, we introduce VeriGuide Platform, which provides a professional writing toolbox to promote the enhancement of teaching and learning in writing. It contains six vertical components, which could be split into two groups. The first group, Editing Assistance, facilitates students write papers and point out grammar and spelling errors. While the second group, Text Analysis, offers document analysis results, which enables students to achieve further writing improvement with explanatory feedbacks. Furthermore, we could do education data analytics to enhance the efficiency of teaching and learning. Specifically, the Editing Assistance contains well-organized writing and formatting, and grammar and spelling checking, while readability assessment, similarity detection, citation analysis, and sentiment analysis are included in the Text Analysis.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*text analysis*; I.7.1 [**Document and Text Processing**]: Document and Text Editing—*spelling*; K.3.0 [**Computers and Education**]: General

Keywords

Text analysis; Education; Language technologies

1. INTRODUCTION

Writing has been regarded as a vital issue for education [15, 23], also a fundamental skill in teaching and learning. Especially, when the standards of student literacy are falling [15], corrective feedback is needed for enhancing writing skills [23]. Meanwhile, a number of multi-disciplinary studies [4, 10, 14, 21] have shown that professional writing plays an important role in various countries, domains and specific subjects. Thus, how to achieve enhancement in writing ability becomes a concerned topic.

Nowadays, with the rapid development of education technologies, more and more writing tools emerge to provide advanced services to address this issue. In terms of the writing tools, they are mostly text-processing related programs and each of them mainly concentrates on addressing a specific issue. To list some of them, (1) LaTeX[1] is a document editing tool for typesetting and formatting for common user; (2) WhiteSmoke[2] provides the writing correction and suggestion service including grammar and spelling checking; (3) Lexalytics[3] focuses on text analysis such as keyword extraction and sentiment analysis.

Despite the diversity of text processing tools, which satisfies most of the demand for professional writing, a one-stop service platform could further improve the working efficiency, operation convenience and user experience. The idea that an e-learning ecosystem will be the next generation for e-learning has been supported by recent research community [2, 6, 17]. So it is important to develop a learning platform to satisfy various requirements of users in professional writing.

In this paper, we introduce the VeriGuide Platform, which makes contributions to integrating and providing comprehensive language technologies to enhance teaching and learning in writing. On one hand, those language technologies are involved as a type of feature in our VeriGuide Platform, equipping our platform with more functionality modules. On the other hand, we can particularly utilize those language technologies to set up a unified and integrated writing environment to promote enhancement in writing for users.

The VeriGuide Platform has implemented two main components with applications of six language technologies for enhancement in writing. It includes well-organized writing and formatting, grammar and spelling checking in Editing Assistance component, readability assessment and similarity detection, citation recognition and sentiment analysis in Text Analysis component. With the continuous development of our platform, more language technologies will be added in future.

[1]http://www.latex-project.org/
[2]http://www.whitesmoke.com/
[3]http://lexalytics.com/

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.
WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2580058.

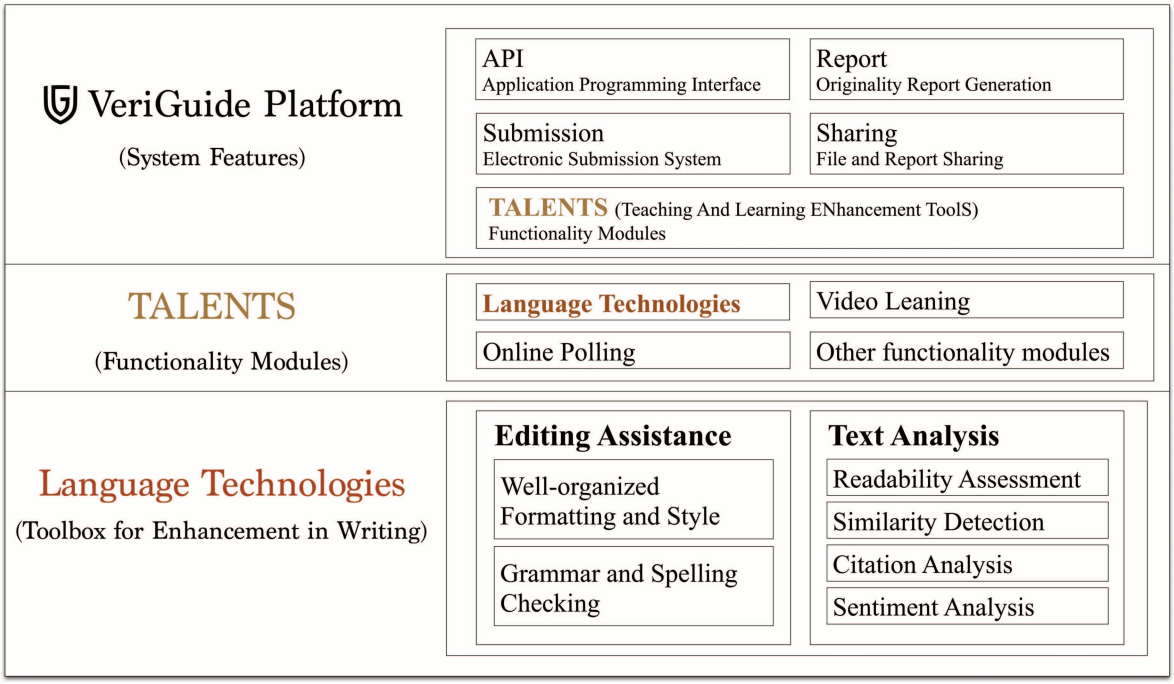

Figure 1: VeriGuide platform addressing the issue of enhancement in writing with language technologies

The organization of the paper is as follows. We present the related work in Section 2. Then we give a brief introduction of the architecture of VeriGuide Platform in Section 3. The six language technologies of our platform would be specifically discussed in Section 4. Finally, we draw the conclusion in Section 5.

2. RELATED WORK

There are many text analysis and writing tools available in the market. TopTenReviews[4] has reviewed services and products in spelling and grammar checkers[5], plagiarism detections[6], and writing enhancement tools[7]. Taking one of the software being reviewed, Whitesmoke[8] as an example, this kind of software generally covers typical grammatical errors (e.g., subject verb agreement, incorrect use of prepositions, sentence structure), and also gives suggestions and feedbacks on writings (e.g., alternate word and grammar choices, wordiness etc.). TextAnalysis[9] and Wikipedia[10] list a rich collection of tools for advanced text analysis and text mining. This software, such as Lexalytics[11], process unstructured texts and turn them into structured ones, which are then used for analytic operations like topic clustering, sentiment analysis, and document summarization.

[4]http://www.toptenreviews.com/
[5]http://online-grammar-check-review.toptenreviews.com/
[6]http://plagiarism-checker-review.toptenreviews.com/
[7]http://writing-enhancement-software-review.toptenreviews.com/
[8]http://www.whitesmoke.com/
[9]http://www.textanalysis.info/
[10]http://en.wikipedia.org/wiki/Text_mining
[11]http://www.lexalytics.com/

3. PLATFORM ARCHITECTURE

The architecture of the VeriGuide Platform and the particular toolbox for enhancement in writing with language technologies are shown in Figure 1. The key concept is elaborated as follows: (1) the top layer indicates that VeriGuide Platform is a unified platform with multiple features, such as holding a submission system, generating originality reports, providing API services. (2) The middle layer further illustrates that the TALENTS (Teaching And Learning ENhancement ToolS), as one feature in VeriGuide Platform, is consist of various independent functionality modules, like the language technologies, online polling and videos. (3) The bottom layer demonstrates that, based on those language technologies related modules, a particular toolbox for enhancement in writing is supported. Especially, this paper focus on the third layer, i.e., this article aims to address the topic of the enhancement of teaching and learning in writing through the language technologies provided by the VeriGuide Platform.

4. LANGUAGE TECHNOLOGIES

We have implemented six language technologies and they are split into two groups to provide services for enhancement in writing. The first group, Editing Assistance, facilitates users when they are generating professional documents, while the second group, Text Analysis, analyzes the written documents in order to further improve their writing skills with valuable feedbacks.

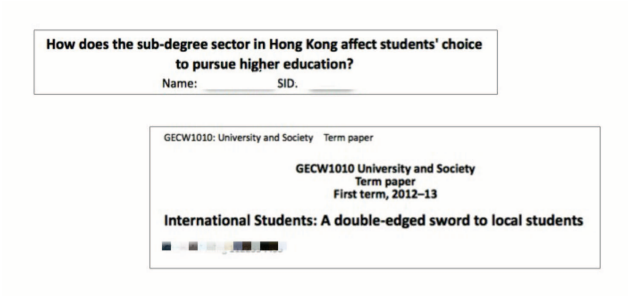

Figure 2: Submitted papers with various format styles

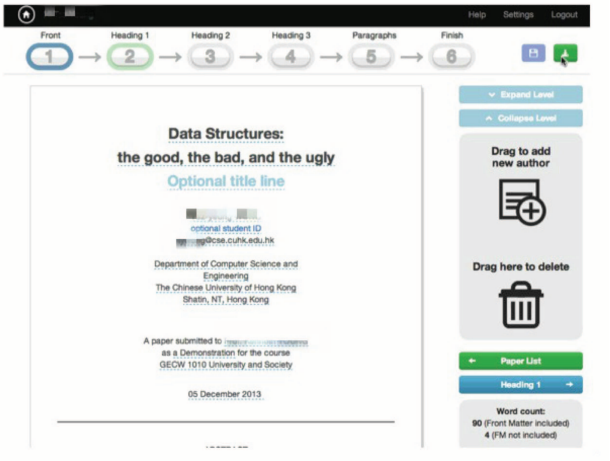

Figure 3: A screenshot of the well-organized writing and formatting module showing a standardized format

4.1 Editing Assistance

4.1.1 Well-organized Writing and Formatting

Generally in writing, novices encounter many problems when writing articles. (1) Few have a good habit of systematically organizing the article before writing, such as writing an outline and organizing sections. (2) They easily make many errors in their articles like grammar errors, formatting errors, bibliographic listing and citation errors. Specifically, these problems are obviously among students, especially for freshmen. They need a suitable software tool to help them learn how to systematically organize an article and make a correct formatting.

We develop a web-based module to help novices write an article. Currently, it is used in the university to guide students to write an academic term paper. Novices could benefit from this module to improve their writing skills and the quality of their papers: (1) the writer could be guided towards systematic organization. They could learn to write an outline before writing, from organizing sections and subsections of the paper, to write down a topic sentence for each paragraph. So the outline could be submitted to teachers for guidance. (2) The paper could be automatically produced to a professional-looking format. Users do not need to worry about how to adjust the formatting, so they could just focus on the content.

The module guides novices to complete an academic paper in three steps. (1) It will guide novices to write title, author information, paper information and abstract. It will automatically organize the information in a unified way. Users do not need to bother the font style and formatting. (2) It will guide users to firstly write sections and subsections of the paper and then scaffold sections with topic sentences. It will guide users to organize the paper systematically, so users can captivate a good habit of writing. (3) Users can select a specific document style and then output the document. They can also output a file for further editing in Microsoft Word[12].

Figures 2 and 3 illustrate the differences before and after using the system, in the aspect of format and style. Figure 2 shows papers that are produced by freshmen with different styles. Figure 3 shows how this module organizes unified format and styles. From the comparison, we can see that it could guarantee the unified format and styles, and improve the quality of the paper. Likewise, this type of problems is common in students' submission, so the well-organized formatting and style module in VeriGuide Platform assist students write high quality paper.

It further helps teachers in teaching and reviewing documents. (1) The system saves teachers' time as it has already doing their job in guiding students in writing. Teachers will receive and guide students improving the organization of the paper as the outline could be submitted to teachers by the system. As the system will do data analytics based on students' writing behavior, teachers will learn how to guide students improving their writing ability. (2) The system will improve teachers' work efficiency in reviewing documents, as they will get well-organized and standard format papers.

4.1.2 Grammar and Spelling Checking

We made involuntary mistakes while transforming our thoughts into textual expression. These mistakes could be classified into three categories of errors, and presented with different traits in different languages [12]. Among them, grammatical and spelling errors are two ubiquitous and concerned issues, which substantially influence the readability and comprehension of a written document. In order to solve these problems, some professional software like Microsoft Word[13] provides a powerful function of spelling and grammar checking. With the development of open source and the advancement of technology, applicable resources like LanguageTool[14], AfterTheDeadline[15] and Graviax[16] enable developers to implement their stand-alone applications.

We develop and integrate the grammar and spelling checking module into our platform, in order to offer a comprehensive environment for writing. We make use of off-the-shelf tools with modification and adaptation to implement this module. It currently supports not only English, but also Chinese. In addition, we apply particular rules to detect typical and distinct errors in Hong Kong for localized customization. With implementation of proper presentation, like Figure 4, users can see the highlight of their errors,

[12]http://office.microsoft.com/word/
[13]http://office.microsoft.com/word/
[14]https://www.languagetool.org/
[15]http://afterthedeadline.com/
[16]http://sourceforge.net/projects/graviax/

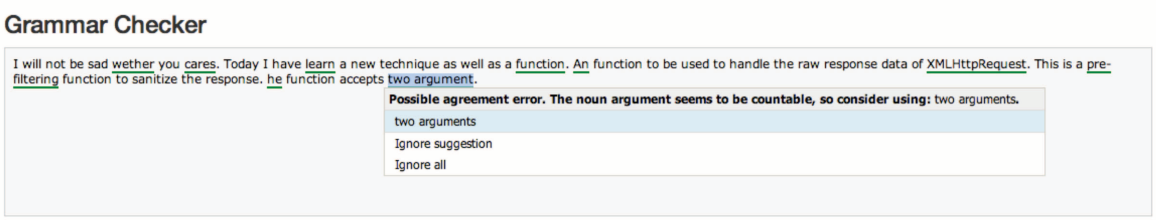

Figure 4: A screenshot of the grammar and spelling checker module suggesting corrections

which assist them in correcting their mistakes and thus improving the whole quality of their works. Grammar and spelling checking module enables users to be aware of the possible errors and make according corrections in time.

The system could further tracks students' performance in writing and help to detect plagiarism with the help of data analytics. (1) By measuring errors in papers, the system helps to reflect the variation of students' writing ability across the entire academic years. If they make fewer errors as time goes on, it turns out they make a progress in writing papers. (2) It provides another way to detect plagiarism. If two students tend to make similar errors in their papers, one student may copy others' materials or find others help him writing the document.

4.2 Text Analysis

4.2.1 Readability Assessment

Readability assessment is a method of estimating the level of difficulty of a piece of writing. Klare [13] describes the term "readability" in the following three ways: (1) To indicate the legibility of a document. (2) To indicate the ease of reading due to the interest-value or pleasantness of writings, and (3) To indicate the ease of understanding due to the writing style.

While there is a long history of assessment for English [5, 7, 8, 9, 18], there are not much similar researches for Chinese. The study [20] proposes a Chinese readability assessment approach based on language factors, Chinese text processing, and machine learning techniques. The mechanism is implemented in the VeriGuide system.

For the use of readability analysis, in addition to the difficulty assessment, we can also use it for checking originality and tracking learning progress. For originality check, if a student has written a piece of texts which its readability is found inconsistent to its education level, teacher may need to pay attention to the work to see if the text is originated from other sources. For tracking learning progress, we can record the readabilities of a student's writings throughout the learning period and see if there is any improvement on this.

4.2.2 Similarity Detection

Plagiarism has become a bigger problem since everyone could access information from the Internet at a few low cost. The fast growth of Internet allows people to have more effective way to search information for writing. However, it also causes side effect on academic dishonesty. The rich information available provides opportunity for writers to copy the ideas or even the exact contents from others, and hence result in plagiarism. "Plagiarism is defined as the use, without giving reasonable and appropriate credit to or acknowledging the author or source, of another person's original work, whether such work is made up of code, formulas, ideas, language, research, strategies, writing or others form(s)." [22] There are five types of plagiarism, namely copy and paste plagiarism, word switch plagiarism, style plagiarism, metaphor plagiarism, and idea plagiarism [1]. Neither of any types should be tolerant, especially in educational sector.

It is difficult to review suspected plagiarism cases with limited human resources, especially when the major tasks for educators is to focus on improving academic quality. In higher education sectors, educators demand the need to have automatic tools to uncover potential plagiarism cases. An automated, scalable and flexible detection tool can assists markers and professors to identify suspected cases by computerize similarity comparison among texts without time consuming manually review. In fact, it is almost impossible for human review against huge media sources, either from Internet, database or previous students' archive.

Thus, we develop a similarity detection module in the VeriGuide Platform to tackle with such tasks. Apart from Internet sources and journal databases, previous students documents and even within class submissions are also the major sources of plagiarism cases. Among several detection approaches, substring matching and bass of word detection give higher detection rate and performance[17]. A web-based platform with course management module, i.e., VeriGuide has been developed at the Chinese University of Hong Kong (CUHK) since 2005 to satisfy such detection purpose.

VeriGuide intends to provide a simple, multi-languages, accurate web platform, which also adapted to the existing teaching and learning methodology. The major components inside include front-end web interface, parameter module, core engine, and report generator [3]. The primary detection languages of VeriGuide include English, Traditional Chinese and Simplified Chinese. Users submit text-based documents to system together with checking parameters such as minimum length of sentence, leniency control, etc. The core engine will start the checking procedure and forward the results to report generator and finally display to users in well-organized side-by-side web interface which users could export it to the portable document format (PDF). One major difficulty of similarity detection is paraphrasing as most students will not directly copy the contents, i.e. trying to avoid being identified by tools like VeriGuide. To address this point, it will perform text extraction, and normalization to retrieve keywords and further computes the similarity result

[17]http://en.wikipedia.org/wiki/Plagiarism_detection

Figure 5: A screenshot of the VeriGuide originality report showing the similarity checking results in side-by-side comparison

by advanced paraphrased detection algorithms, particularly in Chinese. In addition to parameters like leniency, our similarity detection module also implements filters to prevent false alarms, e.g. false positives, group submissions, duplicate submissions, etc. Nevertheless, there may still contain referenced materials and reference lists, which will require further reference recognition (Refer to 3.5 Reference recognition for details). Apart from identifying the suspected sentences, the adoption of manually review on results are significant to improve academic integrity as markers are major player who could provide feedback or even punishment to students suspected to have plagiarized. Hence a simple and clear side-by-side comparison layout could facilitate markers to review generated results and take further actions if necessary, as shown in Figure 5.

4.2.3 *Citation Analysis*

Citation is important in, or even the basis of academics [11]. Academics, that is the pursuit of knowledge, is a process of (1) base on the currently established facts, information, descriptions, or skills (2) to create new ideas using logical and innovative reasoning skills. Making accurate citation is thus essential in tracing the knowledge development and providing proofs of the legitimacy and correctness of the ideas.

Because of the importance, citation analysis, that is to examine, analyzing, and mining the references provided in academic writings, can (1) help to verify the citations have been made correctly, (2) provide valuable intra- and inter-documental information. For intra-documental information, we can understand the robustness and novelty of the ideas proposed. For inter-documental information, we can discover those significant writings or ideas, understand the current trend of knowledge development, and inter-connecting educators or researchers through the underlying citation networks.

Citation analysis involves a number of text processing and machine learning procedures. Generally citation analysis consists of the following sub-tasks: (a) reference sections and entries recognition; (b) parsing of a reference entry into various fields; (c) data mining and statistical analysis to discover the underlying knowledge.

In VeriGuide Platform, citation analysis helps improving the accuracy of plagiarism detection. For example, (1) by identifying the reference section, similarity detection over this section can be omitted so as to avoid false-positive on matching reference titles; (2) if the system knows that a segment of similar texts actually has been cited in the reference section, the system can also inform the teachers about this for considerations.

4.2.4 *Sentiment Analysis*

Sentiment analysis targets at identifying the underlying viewpoint in a sentence or a document, to achieve automated opinion discovery [16, 19]. One particular type of sentiment analysis is the detection of sentiment polarity, also known as opinion orientation, which is for indicating whether the overall orientation of a certain text is Positive or Negative. Moreover, a third type of orientation is Neutral, which points out the absence of opinion or subjectivity. Another type of Sentiment Analysis is to provide more specific statistics, for example, a score or percentage of Positive and Negative.

In terms of enhancement in writing, people need sentiment analysis to make sure they make correct expression. Following are some scenarios for illustration: (1) we might sometimes abuse the chosen phrases in our written works to express wrong feeling; (2) when writing comments about something we might be too harsh and critical to unintentionally deliver pessimistic viewpoint than expected; (3) we might need to double-check whether the tone presented in our report is too sentimental to disorient our objective description or factual statement. In all similar cases like these, sentiment analysis contributes to detect the misdirection and thus insures our correct expression.

Since sentiment analysis could improve our writing, we develop a simple sentiment polarity detection module in VeriGuide Platform. This module cast the inputted text into three possible opinion orientations, as Positive, Neutral and Negative, meanwhile, with demonstration of proportion. So the workflow is easy to understand, our platform processes the imported text and then returns the polarity to help the author to evaluate whether it is the expected.

5. CONCLUSIONS

Writing plays a vital role in education as well as a fundamental skill in teaching and learning. With the advancement of technology, a number of writing tools have been developed for learners to improve their writing skills. As they concentrate on addressing specific issues, people need a one-stop platform, which could integrate multiply functions. In addition, with the concept of e-learning ecosystem for future education, a comprehensive platform will be more promising. Therefore, we introduce VeriGuide Platform, which makes use of various language technologies to promote the enhancement in writing. It provides a writing toolbox with integrated language technologies for all students and teach-

ers. The toolbox contains six vertical components, which could be split into two groups. The first group, Editing Assistance, facilitates users to generate documents. It contains well-organized writing and formatting, and grammar and spelling checking. The second group, Text Analysis, offers document analysis results. Readability assessment, similarity detection, citation analysis and sentiment analysis are included in the Text Analysis. Nevertheless, the e-learning ecosystem for education are multi-disciplinary and kept updating, so the VeriGuide Platform will continuously add more new language technologies in the future.

6. REFERENCES

[1] Cecilia Barnbaum. Plagiarism: A student's guide to recognizing it and avoiding it. *Valdosta State University*, 2002.

[2] Vanessa Chang and Christian Guetl. E-learning ecosystem (eles)-a holistic approach for the development of more effective learning environment for small-and-medium sized enterprises (smes). In *Digital EcoSystems and Technologies Conference, 2007. DEST'07. Inaugural IEEE-IES*, pages 420–425. IEEE, 2007.

[3] Chi Hong Cheong, Tak Pang Lau, and Irwin King. Policy and issues in deploying automated plagiarism detection systems in academic communities: A case study of veriguide. *Ict Ethics and Security in the 21st Century: New Developments and Applications*, page 172, 2011.

[4] Patricia Clarke, Daniel Schull, Glen Coleman, Rachael Pitt, and Catherine Manathunga. Enhancing professional writing skills of veterinary technology students: linking assessment and clinical practice in a communications course. *Assessment & Evaluation in Higher Education*, 38(3):273–287, 2013.

[5] Edgar Dale and Jeanne S Chall. A formula for predicting readability. *Educational research bulletin*, 27:11–28, 1948.

[6] Bo Dong, Qinghua Zheng, Jie Yang, Haifei Li, and Mu Qiao. An e-learning ecosystem based on cloud computing infrastructure. In *Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on*, pages 125–127. IEEE, 2009.

[7] Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3):221, 1948.

[8] Edward Fry. Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, 21(3):242–252, 1977.

[9] Robert Gunning. *The technique of clear writing*. McGraw-Hill New York, 1968.

[10] Mahabalagiri N Hegde. *A coursebook on scientific and professional writing for speech-language pathology*. CengageBrain. com, 2009.

[11] Judy Hunter. The importance of citation. *URL: http://web. grinnell. edu/Dean/Tutorial/EUS/IC. pdf (12.04. 2007.)*, 2006.

[12] Carrie A Keen. inguistic development in english, chinese and persian children: a longitudinal comparison. 2004.

[13] George R Klare. The measurement of readability. *Iowa State University Press*, 1963.

[14] Christine L Latham and Nancy Ahern. Professional writing in nursing education: Creating an academic-community writing center. *The Journal of nursing education*, pages 1–10, 2013.

[15] Mary R Lea and Brian V Street. Student writing in higher education: An academic literacies approach. *Studies in higher education*, 23(2):157–172, 1998.

[16] Bing Liu. Sentiment analysis: A multi-faceted problem. *IEEE Intelligent Systems*, 25(3):76–80, 2010.

[17] UDEN LORNA and ERNESTO DAMIANI. The future of e-learning: Elearning ecosystem. In *Proceedings of the first IEEE International Conference on Digital Ecosystems and Technologies, Cairns, Australia*, volume 100, pages 113–117, 2007.

[18] G Harry McLaughlin. Smog grading: A new readability formula. *Journal of reading*, 12(8):639–646, 1969.

[19] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics, 2004.

[20] LAU Tak Pang. Chinese readability analysis and its applications on the internet. Master's thesis, The Chinese University of Hong Kong, 2006.

[21] Lucy Rai and Theresa Lillis. 'getting it write'in social work: exploring the value of writing in academia to writing for professional practice. *Teaching in Higher Education*, (ahead-of-print):1–13, 2012.

[22] Ethan T Vishniac. What is plagiarism? In *American Astronomical Society Meeting Abstracts# 219*, volume 219, 2012.

[23] Mohammad Zohrabi and Parvin Rezaie. The role of form-focused feedback on developing students' writing skill. *Theory and Practice in Language Studies*, 2(7):1514–1519, 2012.