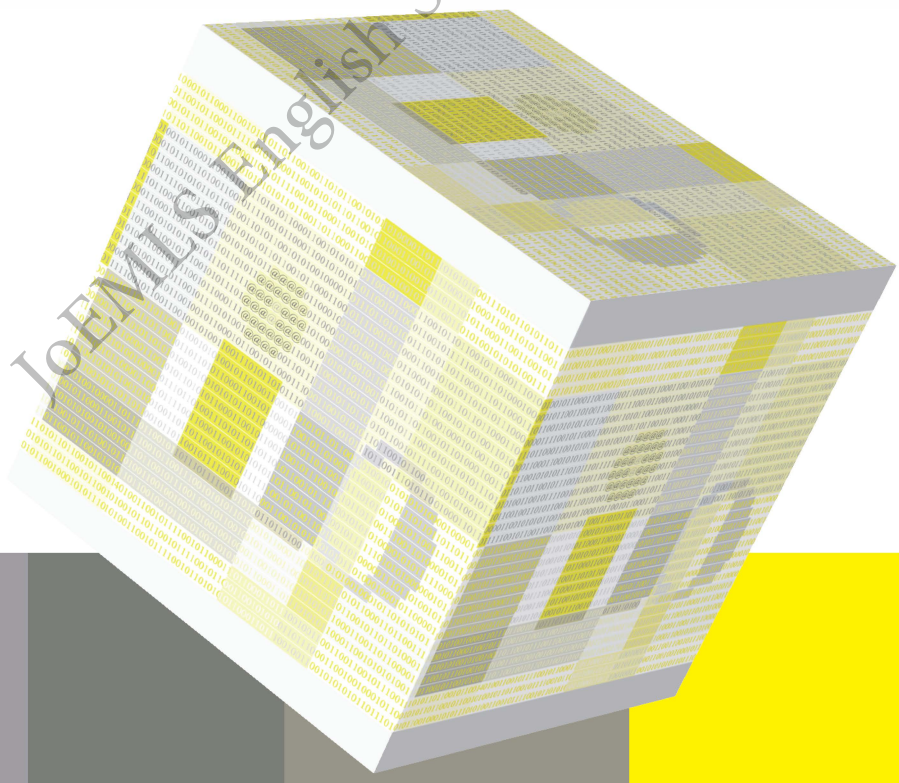
JoEMLS
DOI:10.6120/JoEMLS
CODEN: CYTHD5 ISSN 1013-090X

教育資料與圖書館學

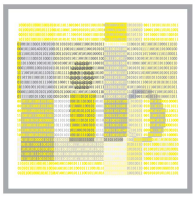
JOURNAL OF EDUCATIONAL MEDIA & LIBRARY SCIENCES

第五十七卷 第一期 二〇二〇年
Vol. 57, No. 1, 2020



淡江大學出版中心

教育資料與圖書館學 JOURNAL OF EDUCATIONAL MEDIA & LIBRARY SCIENCES Vol. 57, No.1, 2020



教育資料與圖書館學，始於1970年3月創刊之**教育資料科學月刊**，其間於1980年9月更名為**教育資料科學**，並改以季刊發行。自1982年9月起易今名。另自2016年11月起，改以一年出版三期（3月、7月、11月）。現由淡江大學出版中心出版，淡江大學資訊與圖書館學系和覺生紀念圖書館合作策劃編輯。本刊為國際學術期刊，2008年獲國科會學術期刊評比為第一級，2015年獲科技部人文社會科學研究中心評定為教育學門專業類一級期刊。並廣為海內外知名資料庫所收錄(如下英文所列)。

The JOURNAL OF EDUCATIONAL MEDIA & LIBRARY SCIENCES *(JoEMLS)*, published by the Tamkang University Press and co-published with the Department of Information & Library Science (DILS) and Chueh Sheng Memorial Library, was formerly the ***Bulletin of Educational Media Science*** (March 1970 – June 1980) and the ***Journal of Educational Media Science*** (September 1980 – June 1982). In 2015, The *JoEMLS* is acknowledged as the first class scholarly journal in Taiwan by Ministry of Science and Technology (MOST). Since November 2016, the *JoEMLS* has been changed from quarterly to a tri-annual journal, published in March, July, and November.

The *JoEMLS* is indexed or abstracted in

Cabell's Directory of Publishing Opportunities
Chinese Electronic Periodicals Service (CEPS)
Directory of Open Access Journal (DOAJ)
H.W. Wilson Database
Index to Chinese Periodicals
Library, Information Science & Technology Abstract (LISTA)
Library & Information Sciences Abstracts (LISA)
Library Literature & Information Science (LLIS)
Public Affairs Information Services (PAIS)
Scopus
Taiwan Social Sciences Citation Index (TSSCI)
Ulrich's Periodicals Directory

教育資料與圖書館學 封面意義：躍升於紙本印象上的數位與網路化圖書資訊圖騰。
The cover design of ***JoEMLS*** signifies:
L (Librarianship); **I** (Information Technology); **B** (Bibliophile and the Book trade)

教育資料與圖書館學
JOURNAL OF EDUCATIONAL MEDIA & LIBRARY SCIENCES

JoEMLS 編輯政策

本刊係採開放存取（Open Access）與商業資料庫付費途徑，雙軌發行之國際學術期刊，兼具電子版與紙本之平行出版模式。本刊除秉持學術規範與同儕評閱精神外，亦積極邁向InfoLibrary寓意之學域整合與資訊數位化理念，以反映當代圖書資訊學研究趨勢、圖書館典藏內容與應用服務為本；且以探討國內外相關學術領域之理論與實務發展，包括圖書館學、資訊科學與科技、書業與出版研究等，並旁及符合圖書資訊應用發展之教學科技與資訊傳播論述。

Open Access 典藏政策

*JoEMLS*向來以「綠色期刊出版者」（Green Publisher / Journal）自居，同意且鼓勵作者將自己投稿至*JoEMLS*之稿件，不論同儕評閱修訂稿與否，都能自行善加利用處理，但希望有若干限制：

⑴勿將已刊登之修訂稿（post-print）再自行轉為營利目的之使用；

⑵典藏版以期刊排印之PDF檔為首選；

⑶任何稿件之典藏版本皆須註明其與*JoEMLS*之關係或出版後之卷期出處。

JoEMLS Editorial Policy

The *JoEMLS* is an Open Access (OA) Dual, double-blind reviewed and international scholarly journal dedicated to making accessible the results of research across a wide range of Information & Library-related disciplines. The *JoEMLS* invites manuscripts for a professional information & library audience that report empirical, historical, and philosophical research with implications for librarianship or that explore theoretical and practical aspects of the field. Peer-reviewed articles are devoted to studies regarding the field of library science, information science and IT, the book trade and publishing. Subjects on instructional technology and information communication, pertaining to librarianship are also appreciated. The *JoEMLS* encourages interdisciplinary authorship because, although library science is a distinct discipline, it is in the mainstream of information science leading to the future of **InfoLibrary**.

Open Access Archiving

The *JoEMLS*, as a role of "OA green publisher/journal", provides free access onlined to all articles and utilizes a form of licensing, similar to Creative Commons Attribution license, that puts minimal restrictions on the use of *JoEMLS*'s articles. The minimal restrictions here in the *JoEMLS* are:

(1) authors can archive both preprint and postprint version, the latter must be on a non-commercial base;

(2) publisher's PDF version is the most recommend if self-archiving for postprint is applicable; and

(3) published source must be acknowledged with citation.

JOURNAL OF EDUCATIONAL MEDIA & LIBRARY SCIENCES

Volume 57 Number 1 2020

Contents

Journal of Educational Media & Library Sciences 57 : 1 (2020) : 1-5
DOI:10.6120/JoEMLS.202003_57(1).editorial

EDITORIAL

Innovations and Social Responsibilities of Scholarly Journals

Our journal (*JoEMLS*) has been continuously working on innovations, and dedicating to taking social responsibilities of academic societies. Insights from running scholarly journals could be flexibly applied to a wide range of areas, and our innovative management policies could contribute to scholarly publishing fields of various academic subjects. When readers, manuscript contributors and journal managers all have a comprehensive understanding of meanings and contents of journal management policies, it is beneficial to academic communication. In terms of organizational structures and editorial policies of scholarly journal publishing, all of the involved staff members are fundamental supporters, and the key to the success of an scholarly journal is a set of clear and comprehensive policies on journal organization and editing, which is also the baseline of various domestic and international scholarly journal ratings. However, at present, most chief editors and editorial board members in Taiwan focus only on the authoritative of related subject fields, and many journal management teams have overlooked the expertise of scholarly journal publishing, due to rapid personnel turnover and insufficient prerequisite knowledge and professional training of journal editing. Only when the expertise of academic editing, publishing and managing receives an equal attention as the authoritative of academic fields, can a scholarly journal have a continuous development and progress.

It is the 50th year of *JoEMLS*, which has built a solid foundation of scholarly journal publishing. Our several changes and measures of digital transformation regarding journal organization and editorial policies have been acknowledged and praised by many domain experts. Readers, manuscript contributors and journal managers from various subject fields have sent us mails for asking about issues regarding referencing formats, manuscript submitting, publishing and managing of scholarly journals. There have been several scholarly journals in Taiwan adopting the value-added services of academic communication created by our journal; however, it is a pity that the superficial measures might be observed and imitated, but the embedded meanings and essence cannot be easily understood and captured. Therefore, we look forward to engaging in information communication and sharing through our varied platforms of scholarly communication. It not only helps permanently preserve the experiences, but also facilitates the goal of sharing accumulated knowledge and thoughts of scholarly publishing and managing among a comprehensive academic research society. It is expected to

accumulate more powerful academic energy and bring about deeper influences on issue development of scholarly journal publishing and future studies on academic development in Taiwan. We have been constantly sharing new concepts regarding scholarly publishing, managing and editorial practices of *JoEMLS*, actively developing diverse application modules for innovative publishing, and providing a platform for other academic journal management teams to observe and learn. We also expect to meet the publication knowledge needs of the academic research community, to attend to the development of scholarly journals, and to promote correct and positive ideas and policies of journal management. We would like to become the bridge between any public-funded journal evaluation systems and editorial teams of journals, to facilitate the collaborations between these two groups. We aim to further help improve the academic productivity in Taiwan, build a quality environment of scholarly publishing, and contribute to the journal publishing and management in the research community of Chinese language societies of the world.

In Taiwan, these has not yet been any learned societies or associations like Council of Science Editors or Committee on Publication Ethics, for gathering scholarly journal publishing peers, providing academic services and professional consulting, and acting as a platform for communicating with domestic and international groups. However, a continuous and quality environment for scholarly journal publishing requires a set of "systemized" editorial works, leading chief editors to adopt correct and appropriate measures, for enhancing the overall quality of journal manuscripts and for providing comprehensive and useful advice on editorial work. Temporal expedients would never solve the true problems. To achieve the goals mentioned above, scholarly journal managers in Taiwan should reach a consensus, and work together to form a collaborative alliance of journals, and gradually expand its influence to set up a professional consortium. A mechanism like this will help and guide more journal publishing units to be equipped with sufficient capabilities for enhancing the publishing quality in their subject fields. It will also provide references for more scholarly journal publishers and invoke discussions. It is truly beneficial to the scholarly publishing environment.

Our journal has provided explanations on "Acting Together to Promote Innovation and Creativity in *JoEMLS*" in our Volume 55, Issue 2 (July 2018), and we have already realize these changes step by step. In the beginning of this Volume 57 Issue 1 (March 2020), we introduce our brand new design of web pages, and we also try to meet the browsing needs of various devices in our electronic version for reading on smart phones. The responsive web information is highlighted. We also demonstrate our passion for providing academic services

through our Facebook Fan Page, and the Cite2Style (https://cite2style.blog/), a website for introducing and recommending references styles of academic writing. As to the Open Peer Review (OPR) system, which was launched in *JoEMLS* of 2019, among the 13 articles of three volumes in 2019 (56-2, 56-3, 57-1), authors and reviewers of eight manuscripts have agreed on making the review comments open. We have produced six general reports on OPR, with an additional individual report. In almost every issue, we worked hard to encourage part of our authors and reviewers to agree on OPR. At present, we adopt several flexible and combination modes, including anonymous, non-anonymous, comment disclosure, and comment partial disclosure, to realize the ideal of OPR step by step. In the long run, we look forward to seeing the OPR system in full operation.

In this Issue1 of Volume 57, we have accepted four manuscripts and rejected five, with a rejection rate of 55.6%. Several manuscripts are still in the review process. The four manuscripts published in this issue include "Study on Library and Information Science Master Alumni Employment and Master Program Value at Taiwan" by Mei-Ling Wang and Jing-Yu Chang; "A Study on MARC21 Transformation and Application for Linked Data" by Ya-Ning Chen and Dar-Maw Wen; "Follow-Up Study of Inquiry-Based Integrated Information Literacy Curriculum in Junior-High Schools Level" by Lin Ching Chen, and "The Feasibility of Automated Topic Analysis: An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Chinese Text Classification" by Yuen-Hsien Tseng. We would like to thank these scholars for their excellent contribution and generous permission for making the peer review's comments and rebuttal open.

Jeong-Yeou Chiu
JoEMLS Chief Editor

Journal of Educational Media & Library Sciences 57 : 1 (2020) : 7-34
DOI:10.6120/JoEMLS.202003_57(1).0036.RS.AM

Research Article

Study on Library and Information Science Master Alumni Employment and Master Program Value at Taiwan[ψ]

Mei-Ling Wang[a*] Jing-Yu Chang[b]

Abstract

This article discusses the employment status of the LIS master graduates and evaluates the value of the master programs in Taiwan. The study adopted survey method, from April 24 to May 31, 2018 and sent the questionnaires to master graduates from the six LIS master's programs graduated during the period of 2011-2018. A total of 495 questionnaires were sent, and 182 responses were collected with a response rate of 36.36%. This study constructed competency model consisting of five dimensions with 27 indicators assessing the work application of graduates. The study shows that the employment market of LIS master graduates at Taiwan has changed, with library workers accounting for 50.5% and graduate employment has expanded to more information institutions. Overall, LIS graduates rated program provision of knowledge and skills with score 3.87 (Likert 5 points), and most highly in the following areas: research and planning, personal management, information skills, and LIS foundation and service knowledge. The overall satisfaction of graduate work is 3.66, and the satisfaction of library work graduates is higher than that of non-library institutions. The overall satisfaction of the graduates in the evaluation of the master programs is 4.15, and the overall agreement of the graduates on the value of the master's program is 3.89. The value of the LIS master programs in Taiwan is related to the master program satisfaction, competency application in work and job satisfaction.

Keywords: *Library and information science education, Employment survey, Master programs evaluation, Job satisfaction, Outcome evaluation, Taiwan*

[ψ] Part of this article came from Jing-Yu Chang's master thesis "Study on Alumni Career Tracking and Library and Information Science Master Programs Evaluation" under Meiling Wang's guidance. The article was rewritten by Meiling Wang, and was co-named by Meiling Wang and Jing-Yu Chang.

[a] Professor, Graduate Institute of Library, Information and Archival Studies, National ChengChi University, Taipei, Taiwan

[b] Master Student, Graduate Institute of Library, Information and Archival Studies, National ChengChi University, Taipei, Taiwan

[*] To whom all correspondence should be addressed. E-mail: meilingw@nccu.edu.tw

SUMMARY

Introduction

Library and information education is the main way to train library and information professionals, promote library and information science research, and improve library and information agency services. The Master of Library and Information Science (MLIS) education is a form of education in which higher education trains professionals to help learners acquire their abilities of professional practice and behavior. There are currently 6 MLIS programs in Taiwan, including the Department of Library and Information Science of National Taiwan University, the Institute of Library and Information Science of National Taiwan Normal University, the Institute of Library, Information and Archival Studies of National Chengchi University, the Institute of Library and Information Science of National Zhongxing University, Tamkang University Institute of Information and Library Science, and Institute of Library and Information Science at Fu Jen University.

The job market for library and information science graduates has changed in recent years. In addition to library work, graduates of library and information science have also discovered jobs opportunities outside the library. All MLIS programs have potential to expand or reinvent themselves but lacks core values and long-term strategies. They have not yet developed an education system that supplies various topics to information service professionals in accordance with the market needs of the online information society. Chang (2008) pointed out that an imbalance occurred between the supply of library and information science education and the job market. The purpose of this study is to discuss an evaluation of MLIS programs with a focus on the graduates, and to explore the application of professional knowledge and ability and the evaluation of the master's degree after graduation.

Methods

This research is designed to explore the employment survey and evaluation of graduates in order to provide information for improving the master's programs. The main research purposes include: 1. Discuss the employment status, professional knowledge and job satisfaction of graduates of the Master of Science in Library and Information Science in Taiwan; 2. Exploring the satisfaction and value of graduates towards the master's degree; 3. Explore the correlation between graduate employment knowledge and the education value of master's programs; and 4. Discuss the differences of graduates' masters of library and information science education value from different universities. In this study, a questionnaire was used to construct a review model for the master's degree in

library and information science with references to relevant literature. This study focused on the graduates who graduates from one of the six MLIS programs in Taiwan from 2011 to 2018. The questionnaire used the Likert five-point scale and had two parts, graduate employment and employment value evaluation of the master's programs.

According to the above research purposes, this study proposes the following hypotheses:

*H*1: The degree of competency application at graduates' work is related how much they value the MLIS education.

*H*2: Job satisfaction is related to the educational value of the master's program.

*H*3: The degree satisfaction is related to the educational value of the master's program.

*H*4: The perceived value of MLIS education, competency application in work and job satisfaction of different working institutions (library and non-library) are different.

*H*5: The perceived value of MLIS education, competency application in work, job satisfaction and job satisfaction of graduates from different schools are different.

Results

A total of 495 questionnaires were sent and 182 responses were collected with a response rate of 36.36%. For the basic information of the responders, there were more women than men. Most participants were 20~30 years old, accounting for about 50%. The distribution of seniority was more than average, but about 80% of the respondents had less than 10 years. The ratio of respondents with an undergraduate degree in library and information science and those without was 1:1. That indicated that the number of non-books and information science college students attending the library and information science master's programs has increased. The respondents were graduates from the past seven years and most of whom have ten years of experience.

Regarding the employment status of graduates of Library and Information Science, 169 of the 182 subjects were employed, with an employment rate of 92.9%. This study showed that library and information science graduates working in libraries accounted for 50.5%. Comparing with 67% from the study by Ko and Wang (2007), less graduates worked in libraries. MLIS graduates learned what they could apply at work from the master's programs. This study used 27 indicators for five aspects of professional knowledge. The average score of the subjects was 3.87 which indicate that the knowledge that graduates

learned from the master's programs has not reached a high degree of agreement at work. Graduates from different working environment had different views on the education value of the master's degree in library and information science. Comparing library work graduates and non-library work graduates, the application of competencies and job satisfaction in MLIS programs, library work graduates are more positive; there is no difference in salary and job satisfaction between the two.

This study used correlation analysis to verify $H_1 \sim H_5$ research hypothesis and showed that the "Perceived Value of MLIS Education" was significantly correlated to "Competency application," "Job Satisfaction," and "Master Program Satisfaction." The highest correlation of job satisfaction was 0.653, which was close to high correlation. The degree of job satisfaction of graduates was highly correlated with the learned job application, see Table 1.

Table 1 Correlation between the Perceived Value of MLIS Education, the Competency Application in Work, Job Satisfaction and the Satisfaction of Master's Programs

	Intellectual application	Job satisfaction	Master program satisfaction	Master program education value
Competency application	1	0.561**	0.552**	0.549**
Job satisfaction	0.561**	1	0.423**	0.653**
Master program satisfaction	0.552**	0.423**	1	0.502**
Master program education value	0.549**	0.653**	0.502**	1

** $p < .01$

Conclusions and Suggestions

This study explored the employment status of graduates of Library and Information Science and the views on the educational value of the master's program, and put forward suggestions. This study showed that the job market for MLIS graduates has changed and expanded to more information institutions. Masters of graduates agreed that professional competency could be applied to work but they were not highly satisfied, and that library work graduates more than non-library work graduates have high job satisfaction. The Graduates were satisfied with the quality of the master's program and affirmed the value of the master's program. The acquired knowledge could be applied at work and could help to have good job performance.

This study showed that the "perceived value of MLIS programs" was highly correlated with the "competency application," "work satisfaction," and "satisfaction of the master's program," showing the competency application in work, job satisfaction, and satisfaction with the master's program. The three factors that affected the value of the master's program education. Therefore, the MLIS programs in Taiwan should pay attention to the design of course content in the future in order to improve the quality of MLIS programs, attach close links

with libraries and information institutions, and pay attention to the diversity of the job market and employability requirements, and help graduates to effectively apply professional competencies to work to improve work satisfaction.

ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT

王梅玲（2011）。臺灣圖書資訊學碩士教育回顧與展望。圖書館學與資訊科學，*37*(2)，20-40。【Wang, Mei-Ling (2011). Historical development and future trends of library and information science master's education in Taiwan. *Journal of Library and Information Science, 37*(2), 20-40. (in Chinese)】

王梅玲、劉濟慈（2009）。從圖書館價值探討臺灣圖書館員基本專業能力。圖書資訊學研究，*4*(1)，27-68。【Wang, Mei-Ling, & Liu, Chi-Tzu (2009). A study of the core competencies for Taiwan's librarians from the library values perspective. *Journal of Library and Information Science Research, 4*(1), 27-68. (in Chinese)】

吳美美、張淇龍（2017）。圖書資訊學教育。在國家圖書館（編），中華民國一〇五年圖書館年鑑（頁207-238）。國家圖書館。【Wu, Mei-Mei, & Chang, Chi-Lung (2017). Tushuzixunxue jiaoyu. In National Central Library (Ed.), *Zhonghua minguo yilingwunian tushuguan nianjian* (pp. 207-238). National Central Library. (in Chinese)】

林珊如（2016）。圖書資訊學教育。在國家圖書館（編），中華民國一〇四年圖書館年鑑（頁201-226）。國家圖書館。【Chang, Shan-Ju Lin (2016). Tushuzixunxue jiaoyu. In National Central Library (Ed.), *Zhonghua minguo yilingsinian tushuguan nianjian* (pp. 201-226). National Central Library. (in Chinese)】

邱子恒（編著）（2013）。我國公共圖書館工作人員專業知能研訂芻議。國家圖書館。【Chiu, Tzu-heng (Ed.). (2013). *Woguo gonggong tushuguan gongzuo renyuan zhuanye zhineng yanding chuyi*. National Central Library. (in Chinese)】

柯君儀、王梅玲（2007）。臺灣圖書資訊學碩士生就業與能力需求之研究。大學圖書館，*11*(1)，97-116。【Ko, Chun-Yi, & Wang, Mei-Ling (2007). Study on employment market and education of library and information science master program in Taiwan. *University Library Journal, 11*(1), 97-116. (in Chinese)】

胡歐蘭（1994）。臺灣圖書資訊人力資源現況之調查研究（*I*）（NSC 83-0301-H-004-102-M2）。國立政治大學圖書資訊學研究所。【Hu, Nancy Ou-Lan (1994). *The survey and study on human resources status in library and information development (I)* (NSC 83-0301-H-004-102-M2). Graduate Institute of Library, Information & Archival Studies, National Chengchi University. (in Chinese)】

胡歐蘭（1995）。臺灣圖書資訊人力資源現況之調查研究（*II*）（NSC 84-2413-H-004-011-M2）。國立政治大學圖書資訊學研究所。【Hu, Nancy Ou-Lan (1995). *The survey and study on human resources status in library and information development (II)* (NSC 84-2413-H-004-011-M2). Graduate Institute of Library, Information & Archival Studies, National Chengchi University. (in Chinese)】

國家圖書館（編）（2016）。我國大學圖書館館員專業知能指標研訂芻議。國家圖書館。【National Central Library. (Ed.). (2016). *Woguo daxue tushuguan guanyuan zhuanye zhineng zhibiao yanding chuyi*. National Central Library. (in Chinese)】

張淳淳（2008）。輔仁大學圖書資訊學系畢業生就業情況探討。圖書與資訊學刊，*67*，39-50。https://doi.org/10.6575/JoLIS.2008.67.04【Chang, Chwen-Chwen (2008). An exploration of the state of employment in library and information science graduates of Fu Jen Catholic University. *Bulletin of Library and Information Science, 67*, 39-50. https://doi.org/10.6575/JoLIS.2008.67.04 (in Chinese)】

楊美華（1997）。高級圖書館人才的生涯發展與人力之運用研究（NSC 86-2417-H-004-017）。國立政治大學圖書資訊學研究所。【Yang, Mei-Hwa (1997). *The career paths of senior library professionals and the utilization of human resources* (NSC 86-2417-H-004-017). Graduate Institute of Library, Information & Archival Studies, National Chengchi University. (in Chinese)】

American Library Association. (2009). *ALA's core competences of librarianship*. http://www.ala.org/educationcareers/sites/ala.org.educationcareers/files/content/careers/corecomp/corecompetences/finalcorecompstat09.pdf

American Library Association. (2019). *Directory of ALA-accredited and candidate programs in library and information studies*. http://www.ala.org/accreditedprograms/directory

Anderson, C. (2020). Professional education. In *Encyclopedia of business and finance* (2nd ed). http://www.encyclopedia.com/finance/finance-and-accounting-magazines/professional-education

CILIP. (2012). *Professional knowledge and skills base*. https://www.cilip.org.uk/page/PKSB

CILIP. (2019). *Accredited courses by level*. https://www.cilip.org.uk/page/AccreditedCoursesByLevel

Griffiths, J.-M., & King, D. W. (1986). *New directions in library and information science education*. Knowledge Industry Publications for the American Society for Information Science.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). Berrett-Koehler.

Marshall, J. G., Morgan, J. C., Rathbun-Grubb, S., Marshall, V. W., Barreau, D., Moran, B. B., Solomon, P., & Thompson, C. A. (2010). Toward a shared approach to program evaluation and alumni career tracking: Results from the workforce issues in library and information science 2 study. *Library Trends*, *59*(1-2), 30-42.

Mehdinezhad, V. (2011). *Criteria for educational program evaluation: The case of teacher education program*. LAP Lambert Academic.

Moore, N. (1987). *The emerging markets for librarians and information workers*. British Library Research and Development Department.

Warraich, N. F., & Ameen, K. (2010). Employment and learning outcomes of LIS graduates: A case of Pakistan. *Education for Information*, *28*(2-4), 315-324. https://doi.org/10.3233/EFI-2010-0914

Mei-Ling Wang ORCID 0000-0002-6748-7744
Jing-Yu Chang ORCID 0000-0002-8020-2694

Journal of Educational Media & Library Sciences 57 : 1 (2020) : 35-72
DOI:10.6120/JoEMLS.202003_57(1).0045.RS.AM

Research Article

A Study on MARC21 Transformation and Application for Linked Data

Ya-Ning Chen[a*] Dar-maw Wen[b]

Abstract

MARC has been accepted as a standard format for information interchange in libraries for decades. Owing to the outdated format, MARC is unknown and unused outside of libraries. Moving to the era of semantic web, the technology of linked data (LD) is regarded as a new approach to deconstruct library bibliographic data (LBD) into LD for libraries. It is deserved to examine what approach has been adopted to extend MARC into LD and its potential benefits. This study has analyzed MARC proposals and discussion papers related to LD as a basis to investigate what changes have been approved for MARC since 2006 of the LD initiative. Furthermore, eight use cases selected from two MARC records and an instance of one MARC proposal respectively were employed to address how MARC changes have been transformed MARC-based LBD into LD in practice by combining classes and properties of BIBFRAME and RDA bibliographic ontology. Consequently, it reveals that RDF's triplification has been integrated as part of MARC successfully. Therefore, MARC is not only a standard for communication and representation of bibliographic and related information, but also one for LD in libraries. Related issues to fundamental definition of bibliographic entity defined in MARC proposals for LD have also discussed.

Keywords: *MARC, Linked data, BIBFRAME, RDA ontology, RDFization*

SUMMARY

Introduction

MAchine Readable Cataloging (MARC) has been adopted as an international standard for information organization, especially for exchanging and sharing information between library automated systems. As information heads increasingly towards cyberization and digitization, search engines have become an essential tool for finding networked information resources on the Internet. Owing to an outdated format, MARC is not known in non-library domains and sectors. Most MARC-based information are embedded in proprietary library automated systems exists as an information silo owing to the isolation from coverage of

[a] Associate Professor, Department of Information and Library Science, Tamkang University, New Taipei City, Taiwan
[b] Chief Knowledge Officer, Flysheet Technologies Co., Ltd., Taipei, Taiwan
[*] To whom all correspondence should be addressed. E-mail: arthur@gms.tku.edu.tw

search engines (Lagace, 2014). On the other hand, Linked Data (LD), initiated by Tim Berners-Lee (2006), has been used as an approach to transform a web of documents into a web of data through URI naming and linking with related resources in an open networked environment. According to the investigation of Linked Open Data Cloud, "bibliography of publications" is one of the categories and shows the significance of library bibliographic information in the domain of LD. However, LD has gained attention from libraries to transform legacy library data into LD and explore its potential applications through the adoption of LD related technologies and tools.

Basically LD is data centric for data design (Di Noia et al., 2016). One of the key points of LD is to employ ontology as a basis for data modeling to delineate the relationships between individual LD (Hyland et al., 2014; Hyland & Villazón-Terrazas, 2011). It is encouraged to reuse existing authoritative vocabularies that are in widespread usage to describe common types of data (Villazón-Terrazas et al., 2011). Although the Functional Requirements for Bibliographic Records (FRBR) and the Bibliographic Framework (BIBFRAME) are conceptual models, actually they are regarded as ontologies for libraries in practice. For example, the National Library and Archive of IRAN (NLAI; Eslami & Vaghefzadeh, 2013), Biblioteca Nacional de España (BNE; Vila-Suero & Gómez-Pérez, 2013; Vila-Suero et al., 2012) and Bibliothèque nationale de France (2018) have used FRBR as an ontology for LD transformation, whereas cases of Linked Data for Production (LD4P) have employed BIBFRAME as an ontology to address issues related to LD transformation. Furthermore, vocabularies and their relationships of BIBFRAME and RDA ontology have been assigned URI maintained by the LC and RDA Registry, respectively. Therefore, these two bibliographic ontologies FRBR and BIBFRAME both have conformed to the requirements of ontology defined by Berners-Lee et al. (2001) for the semantic web.

There is no doubt that MARC is still employed to organize information by many library automated systems around the world. As a matter of fact, libraries have encountered the hybrid requirements for MARC and LD at the same time. Meaning that libraries must not only to transform MARC into LD, but also include external LD resources into library automated systems to migrate user's information navigation into LD driven resource discovery. It is of interest to know what changes have made to MARC and their applications in practice in accordance with the aforementioned hybrid requirements for inclusion of LD.

Literature Review

Totally 18 MARC documents (14 proposals and four discussion papers) published since the term LD was coined in 2006 were selected to investigate the revisions of MARC for LD implemented applications, including subfields $0,

$1, $2, $4, $e, $i, and tag 758. Furthermore, in this study, we checked against two online documents (MARC21 Format for Bibliographic Data (MFBD) and MARC21 Format for Authority Data (MFAD) to collate related MARC subfields and tags for LD applications.

Methodology

First, MFBD and MFAD were selected as target subjects to examine how MARC implements related LD subfields and tags in practice. Then RDF triplification was performed for MARC. In other words, subfield a of tag 245 in MFBD and subfield a of tag 110 in MFAD were regarded as the subject of RDF, $4 was regarded as the predicate of RDF, and $0 or $1 both of MFBD and MFAD were regarded as the object of RDF. Conversely, $0 or $1 both of MFBD and MFAD were regarded as the subject of RDF, subfield a of tag 245 in MFBD and subfield a of tag 110 in MFAD as the object of RDF, and $4 still as the predicate of RDF. Third, vocabularies defined by BIBFRAME and RDA ontology were used as the predicate of RDF during transforming MARC to LD. Eight use cases derived from two MFBD records offered by the University of Michigan Ann Arbor Library and the University of Pennsylvania Libraries WebPACs, as well as instances of the aforementioned MARC documents addressed in the literature review section were employed to investigate how $0, $1, $2, $4, $e, $i and tag 758 were used to extend MARC to LD in detail. The eight use cases included the following relationships: authorship, work's uniform title, publisher, content/media/carrier, translator, subject, instance/manifestation, and organization and individual person. Lastly, each use case was provided with a summarized table to illustrate the distinction between the original MARC and RDFized MARC instance with vocabularies of selected bibliographic ontology (i.e., BIBFRAME and RDA ontology) in accordance with RDF's triple statement and their RDF graphs respectively.

Discussion

MARC is addressed from the following perspectives:

- In terms of LD linkage, MARC can be enriched through by internal enrichment to aggregate external LD resources.
- In terms of information exchange, MARC21 is not only a format for information interchange and sharing, but also an exchange format for sharing MARC-based LD information between library automated systems.
- In terms of application of ontology, MARC21 has become a data container of bibliographic ontology (such as BIBFRAME and RDA ontology), and is also a carrier to reify bibliographic ontology into practice.
- In terms of use cases, one of RDF's triplification approaches was used by MARC, that is, subfield a of tag 245 in MFBD and subfield a of tag

110 in MFAD are regarded as the subject of RDF, and $0 or $1 both of MFBD and MFAD as the object of RDF. On the contrary, it will be worth knowing whether the opposite RDF' triplification approach and syntax (i.e., $0 or $1 both of MFBD and MFAD are regarded as RDF's subject, and subfield a of tag 245 in MFBD and subfield a of tag 110 in MFAD as RDF's object) is a workable approach for MARC in the future.

- According to examination of eight use cases in this study, the 'bibliographic entity' of subfield a of tag 245 in MFBD has stood for various entities including work and instance in BIBFRAME, or work, expression and manifestation in RDA ontology. It has revealed there is a need for a reasonable definition for subfield a of tag 245 in MFBD when libraries adopt LD related MARC subfields and tags. In terms of structure of BIBFRAME and RDA ontology, it often needs more than two RDF triples statements to complete the semantic relationships between two individual LD resources. According to the illustration of eight use cases, one may find that MARC has employed one RDF triple statement to delineate the semantic relationships rather than a complete set of RDF triples, for example the relationships between BIBFRAME's instance/RDA's manifestation and publisher. Indeed a practical guideline is needed to direct libraries about how to select the appropriate BIBFRAME or RDA vocabularies to build up the semantic relationships between LD resources.

Conclusion

According to an analysis of MARC proposals and discussion papers focused on LD and eight use cases, it can be seen that related MARC subfields and tags have been revised to integrate the RDF data model and syntax. Thus external LD resources can be aggregated into part of MARC by enrichment. Furthermore, MARC is not only an international format for sharing bibliographic information, but also a container for exchanging MARC-based LD information in libraries. It would be interesting to know whether RDF-based MARC subfields and tags will be applied to other ontologies in addition to BIBFRAME and RDA ontology.

ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT

Berners-Lee, T. (2006). *Linked data: Design issue*. https://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*(5), 35-43.

Bibliothèque Nationale de France. (2018). *Open data*. data.bnf.fr. https://data.bnf.fr/en/opendata

Casalini, M. (2017, August 15-17). *BIBFRAME and Linked Data practices for the stewardship of research knowledge* [Paper presentation]. IFLA Satellite Meeting 2017: Digital

Humanities, Berlin, Germany. https://dh-libraries.sciencesconf.org/132918/document

Chen, Y.-N. (2017). A review of practices for transforming library legacy records into linked open data. In E. Garoufallou, S. Virkus, R. Siatri, & D. Koutsomiha (Eds.), *Metadata and semantic research: 11th International Conference, MTSR 2017 Tallinn, Estonia, November 28 – December 1, 2017 Proceedings* (pp. 123-133). Springer. https://doi.org/10.1007/978-3-319-70863-8_12

Cole, T. W., Han, M.-J., Weathers, W. F., & Joyner, E. (2013). Library marc records into linked open data: Challenges and opportunities. *Journal of Library Metadata*, *13*(2-3), 163-196. https://doi.org/10.1080/19386389.2013.826074

Deliot, C. (2014). Publishing the British National Bibliography as linked open data. *Catalogue & Index 174*, 13-18. http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf

Deliot, C., Wilson, N., Costabello, L., & Vandenbussche, P.-Y. (2016, October 13-16). *The British National Bibliography: Who uses our linked data?* [Paper presentation]. International Conference on Dublin Core and Metadata Applications 2016, Copenhagen, Denmark. http://dcpapers.dublincore.org/pubs/article/download/3820/2005

Di Noia, T., Ragone, A., Maurino, A., Mongiello, M., Marzoccca, M. P., Cultrera, G., & Bruno, M. P. (2016). Linking data in digital libraries: The case of Puglia Digital Library. In A. Adamou, E. Daga, & L. Isaksen (Eds.), *Proceedings of the 1st Workshop on Humanities in the Semantic Web co-located with 13th ESWC Conference 2016* (pp. 27-38). CEUR-WS. http://ceur-ws.org/Vol-1608/paper-05.pdf

Eslami, S., & Vaghefzadeh, M. H. (2013, August 17-23). *Publishing Persian linked data of national library and archive of Iran* [Paper presentation]. IFLA World Library and Information Congress: 79th IFLA General Conference and Assembly. Singapore. http://library.ifla.org/193/1/222-eslami-en.pdf

Hyland, B., Atemezing, G.A., & Villazón-Terrazas, B. (2014). *Best practices for publishing linked data*. W3C. https://dvcs.w3.org/hg/gld/raw-file/cb6dde2928e7/bp/index.html

Hyland, B., & Villazón-Terrazas, B. (2011, 15 March). *Linked data cookbook*. W3C. https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

Lagace, N. (2014). Pre-standards initiatives: Bibliographic roadmap and altmetrics. *Information Standard Quarterly*, *26*(3), 23-26. https://doi.org/10.3789/isqv26no3.2014.06

Lampert, C. K., & Southwick, S. B. (2013). Leading to linking: Introducing linked data to academic library digital collections. *Journal of Library Metadata*, *13*(2-3), 230-253. https://doi.org/10.1080/19386389.2013.826095

Library of Congress. (1998). *MARC proposal no. 1998-10*. Library of Congress. https://www.loc.gov/marc/marbi/1998/98-10.html

Library of Congress. (2009). *MARC proposal no. 2009-06/1*. https://www.loc.gov/marc/marbi/2009/2009-06-1.html

Library of Congress. (2010). *MARC proposal no. 2010-06*. https://www.loc.gov/marc/marbi/2010/2010-06.html

Library of Congress. (2015). *MARC proposal no. 2015-07*. https://www.loc.gov/marc/mac/2015/2015-07.html

Library of Congress. (2016a). *MARC discussion paper no. 2016-DP06*. https://www.loc.gov/marc/mac/2016/2016-dp06.html

Library of Congress. (2016b). *MARC discussion paper no. 2016-DP18*. https://www.loc.gov/marc/mac/2016/2016-dp18.html

Library of Congress. (2016c). *MARC discussion paper no. 2016-DP19*. https://www.loc.gov/marc/mac/2016/2016-dp19.html

Library of Congress. (2016d). *MARC discussion paper no. 2016-DP21*. https://www.loc.gov/marc/mac/2016/2016-dp21.html

Library of Congress. (2017a). *MARC proposal no. 2017-01*. https://www.loc.gov/marc/mac/2017/2017-01.html

Library of Congress. (2017b). *MARC proposal no. 2017-02*. https://www.loc.gov/marc/mac/2017/2017-02.html

Library of Congress. (2017c). *MARC proposal no. 2017-03. https://www.loc.gov/marc/mac/2017/2017-03.html*

Library of Congress. (2017d). *MARC proposal no. 2017-06*. https://www.loc.gov/marc/mac/2017/2017-06.html

Library of Congress. (2017e). *MARC proposal no. 2017-08*. https://www.loc.gov/marc/mac/2017/2017-08.html

Library of Congress. (2017f). *MARC proposal no. 2017-09*. https://www.loc.gov/marc/mac/2017/2017-09.html

Library of Congress. (2018a). *MARC discussion paper no. 2018-DP07*. https://www.loc.gov/marc/mac/2018/2018-dp07.html

Library of Congress. (2018b). *MARC proposal no. 2018-FT01*. https://www.loc.gov/marc/mac/2018/2018-ft01.html

Library of Congress. (2019a). *MARC advisory committee*. https://www.loc.gov/marc/mac/advisory.html

Library of Congress. (2019b). *MARC proposal no. 2019-02*. https://www.loc.gov/marc/mac/2019/2019-02.html

Library of Congress. (2019c). *MARC proposal no. 2019-03*. https://www.loc.gov/marc/mac/2019/2019-03.html

Library of Congress. (2019d). *MARC 21 to BIBFRAME 2.0 conversion specifications*. https://www.loc.gov/bibframe/mtbf/

Linked Data for Production. (2017). *LD4P grant proposal*. https://wiki.duraspace.org/display/LD4P/LD4P+Grant+Proposal

Malmsten, M. (2008, September 22-26). *Making a library catalogue part of the semantic web* [Paper presentation]. International Conference on Dublin Core and Metadata Applications 2008, Berlin, Germany. http://dcpapers.dublincore.org/pubs/article/view/927/923

Malmsten, M. (2009). Exposing library data as linked data. In *Proceedings of IFLA WLIC 2009*. http://disi.unitn.it/~bernardi/Courses/DL/Slides_10_11/linked_data_libraries.pdf

McCrae, J. P. (2019). *The linked open data cloud: Subclouds by domain*. https://lod-cloud.net/#about

Network Development and MARC Standards Office. (2006). Mapping of MARC data elements to FRBR and AACR. In *Functional Analysis of the MARC 21 Bibliographic and Holdings Formats* (Rev. ed.). http://www.loc.gov/marc/marc-functional-analysis/source/table3.pdf

Penn Libraries. (n.d.). *Jane Austen's Pride and prejudice / edited by Claudia L. Johnson, Susan J. Wolfson*. https://franklin.library.upenn.edu/catalog/FRANKLIN_9939511983503681

Possemato, T. (2018). How RDA is essential in the reconciliation and conversion processes for quality Linked Data. *JLIS.it*, *9*(1), 48-60. https://doi.org/10.4403/jlis.it-12447

RDA Steering Committee. (2019). *Frequently Asked Questions*. RDA Registry. https://www.rdaregistry.info/rgFAQ

Santos, R., Manchado, A., & Vila-Suero, D. (2015, August 15-21). *Datos.bne.es: A LOD service and a FRBR-modelled access into the library collections* [Paper presentation]. IFLA World Library and Information Congress: 81st IFLA General Conference and Assembly. Cape Town, South Africa. http://library.ifla.org/1085/1/207-santos-en.pdf

Schreur, P. (2015). *Implications of a linked data transition: Stanford University's projects and plans*. http://www.lib.berkeley.edu/catalog_dept/sites/drupal7.lib.berkeley.edu.catalog_dept/files/Implications%20of%20a%20Linked%20Data%20Transition.docx

SHARE-VDE. (n.d.). http://share-vde.org/sharevde/docBibframe/Work/139617-12

Simon, A., Wenz, R., Michel, V., & Di Mascio, A. (2013). Publishing bibliographic records on the web of data: Opportunities for the BnF (French National Library). In P. Cimiano, O. Corcho, V. Presutti, L. Hollink, & S. Rudolph (Eds.), *The Semantic Web: Semantics and Big Data. ESWC 2013: 10th International Conference, ESWC 2013, Montpellier, France, May 26-30, 2013. Proceedings* (pp. 563-577). Springer. https://doi.org/10.1007/978-3-642-38288-8_38

Smith-Yoshimura, K. (2016). Analysis of international linked data survey for implementers. *D-Lib Magazine*, *22*(7-8). https://doi.org/10.1045/july2016-smith-yoshimura

Smith-Yoshimura, K. (2018a). Analysis of 2018 international linked data survey for implementer. *Code4lib Journal, 42*. https://journal.code4lib.org/articles/13867

Smith-Yoshimura, K. (2018b). *What metadata managers expect from and value about the research library partnership*. Hanging Together. http://hangingtogether.org/?p=6683

Southwick, S. B. (2015). A guide for transforming digital collections metadata into linked data using open source technologies. *Journal of Library Metadata*, *15*(1), 1-35. https://doi.org/10.1080/19386389.2015.1007009

Suominen, O., & Hyvönen, N. (2017). From MARC silos to linked data silos? *O-Bib. Das Offene Bibliotheksjournal*, *4*(2), 1-13. https://doi.org/10.5282/o-bib/2017H2S1-13

University of Michigan Library. (2018). Ao man yu pian jian / Zhen, Aositing zhu ; [Xia Yinghui yi]. 傲慢與偏見 / 珍・奧斯汀著 ; [夏穎慧譯]. https://search.lib.umich.edu/catalog/record/014616392?query=Ao+man+yu+pian+jian+Xia+Yinghui+yi&library=U-M+Ann+Arbor+Libraries

Vila-Suero, D., & Gómez-Pérez, A. (2013). datos.bne.es and MARiMbA: An insight into library linked data. *Library Hi Tech*, *31*(4), 575-601. https://doi.org/10.1108/LHT-03-2013-0031

Vila-Suero, D., Villazón-Terrazas, B., & Gómez-Pérez, A. (2012). datos.bne.es: A library linked data dataset. *Semantic Web*, *4*(3), 307-313. https://doi.org/10.3233/SW-120094

Villazón-Terrazas, Vilches-Blázquez, L. M., Corcho, O., & Gómez-Pérez. (2011). Methodological guidelines for publishing government linked data. In D. Wood (Ed.), *Linking government data* (pp. 27-49). Springer. https://doi.org/10.1007/978-1-4614-1767-5_2

Wenz, R. (2013). Linked open data for new library services: The example of data.bnf.fr. *JLIS.it*, *4*(1), 403-415. https://doi.org/10.4403/jlis.it-5509

Ya-Ning Chen ORCID 0000-0001-7598-1139
Dar-maw Wen ORCID 0000-0003-1525-4815

Journal of Educational Media & Library Sciences 57 : 1 (2020) : 73-120
DOI:10.6120/JoEMLS.202003_57(1).0039.RS.CM

Research Article

A Follow-Up Study of Inquiry-Based Integrated Information Literacy Curriculum in Junior-High Schools Level

Lin Ching Chen

Abstract

The purpose of this study was to follow up the experiences and performance of junior high students who received the elementary inquiry-based integrated information literacy curriculum for six years. Furthermore, how this curriculum was implemented in junior high schools was another focus for this study. A longitudinal research method was used to gather data from 30 subjects for one and half years since 7th to 8th grades. The instruments included information literacy follow-up surveys for students and teachers, interview questions, and feedback from parents. Data was gathered through in-depth interviews, focus group interviews, surveys, and document analysis. The results showed that regardless of prior academic achievements, most students continuously applied the learned information literacy in learning and living of the junior high school level. However, the elementary inquiry-based integrated information literacy curriculum was not practiced well in the junior high schools, in the aspects of library literacy, media literacy, and inquiry learning. They were not deeply integrated into the curricula at the junior high school level.

***Keywords**: Information literacy, Inquiry learning, Multiple literacies, Follow-up study, Junior-high school*

SUMMARY

Introduction

Information literacy encompasses both the inquiry process and multiple literacies of library, media and computer. Information literacy instruction should start in elementary schools, then systematically be implemented through the secondary level, then to higher education. It had better be integrated across the contexts of school curriculum using inquiry-based learning, so that students can internalize information literacy (Chen et al., 2017; Chu et al., 2011; Kuhlthau et al., 2012). Furthermore, studies find that students of different academic

Professor, Department of E-learning Design & Management, National Chiayi University, Chiayi, Taiwan
E-mail: lingin@mail.ncyu.edu.tw

achievements may perform differently in integrated information literacy instruction (Chen et al., 2017; Ben-David & Zohar, 2009; Cuevas et al., 2005). However, most related studies are conducted in a short term and few research investigates the effects of information literacy instruction in a longer period of time. Therefore, the real effects of information literacy on student's performance are still not clear yet.

Chen and Chen (2019) completed a six-year research project of integrating information literacy into elementary instruction using the Super3 and Big6 inquiry models in an elementary school. The researchers are curious about how these elementary-school graduates, who have received the inquiry-based integrated information literacy curriculum for the past six years will perform at the junior high school level. Will information literacy help them become better problem solvers compared to their peers? What are their experiences of using information? What are the teachers' opinions about these subjects' performance? Furthermore, how is the inquiry-based integrated information literacy curriculum implemented at the junior high school level? Are there any problems needed to be addressed? The above questions are worth conducting further investigation.

Methods

This study used a longitudinal research method to gather data from 30 subjects; that is, the data were collected over an extended period of time (Cohen et al., 2007). The criteria choosing the subjects included the completeness of the six-year data in the elementary school level, and their desire to participate in the follow-up study in their junior high school years. Of the 30 subjects, 8 low-, 8 medium-, and 14 high-academic achievement students participated in this study. Started in the 7th grade they were interviewed and answered surveys twice for one and half years. Besides the student subjects, 31 junior high school teachers and 3 elementary school information literacy teachers were also interviewed and answered the surveys one or twice.

The instruments included information literacy follow-up surveys for students and teachers, interview questions, and feedback from parents. Student subjects were asked to evaluate their own 15 information literacy competencies on a 5-point scale in the survey. Both junior high and elementary school teachers evaluated the subjects' competencies too.

Data was gathered through in-depth interviews, focus group interviews, surveys, and document analysis. All qualitative data were organized, coded, reviewed and analyzed multiple times. The quantitative data were analyzed using descriptive statistics.

Results

Experiences and Performance of Integrating Information Literacy into Junior-High Schools Learning

According to the survey results, the average scores of students' information literacy competencies evaluated by the 30 students themselves twice were 3.49 and 3.69 respectively, which were above the normal level. The elementary school teachers gave the students a higher average score (3.96), which was near the good level and meant that they had strong confidence in the subjects. However, the first average score given by the junior high school teachers was low (2.73); the second one was higher (3.28). The reason for such a difference between the two scores was that teachers did not understand the information literacy concept in the beginning. Then, through communicating with elementary school information literacy teachers, their own close observation of student's performance, and researcher's explanation, the junior high school teachers gradually understood the essence of information literacy.

Based on the elementary academic achievement levels, the researcher selected one to two subjects from the low, medium and high groups, to investigate their learning experiences of using information literacy at the junior high school level.

Low academic achievement group: No. S0-7

When in elementary lower-grade, S0-7 didn't have confidence and her academic achievement was behind peers due to learning Chinese phonics late; until fifth grade, her academic performance gradually improved (S0-7 elementary teacher interview). The average information literacy scores given by her elementary and junior high school teachers (twice) were 3.4, 3.67 and 3.8, which were close to the good level. Her own evaluation scores were 3.1 and 4.2 in grades 7th and 8th respectively, which meant that S0-7 gained self- confidence on learning. In an interview in 8th grade, S0-7 said that her geography was weak because the concepts were hard to understand; she overcame these problems through using the strategies of drawing concept maps and taking notes in detail, which she learned in the elementary information literacy courses (S0-7 interview). Her geography teacher gave excellent grades for her concept map and notes assignments.

Medium academic achievement group: S1-11

Her elementary teacher described her as a student with strong learning motivation and good reading habits; she was awarded the best orator in the 6th-grade debate (S1-11 elementary teacher interview). Her mother said she was interested in all inquiry projects designed in the elementary information literacy curriculum (S1-11 parent feedback). However, after becoming a junior

high student, inquiry projects were seldom conducted in curriculum; yet, S1-11 continued her interest in inquiry. She listed many subject-related inquiry topics in the survey like "stem cell" in biology, "Confucius influences" in Chinese, etc. (S1-11 survey). She also listed 8 useful experiences of using information: drawing concept maps, selecting suitable information seeking strategies, note-taking, comparing information, proposing critics, using reading strategies, integrating information with concept maps, and reflection. The junior high school teacher commended that S1-11 was one of the top three students in the class without going to a cram school, owing to these competencies and dispositions (S1-11 junior high teacher interview).

High academic achievement group

No. S2-26

Regardless of the elementary and junior high school teachers, they all were impressed by S2-26's competencies and attitude of using information literacy. She took good notes, arranged things logically, liked to try new learning strategies, and had good reflection ability. The junior high teacher identified S2-26's six learning advantages including using resources in the library, asking good questions, taking notes, selecting & integrating information, and problem solving (S2-26 teacher survey).

No. S2-17

Though S2-17 performed well in the elementary lower-grade, she was used to reciting information, rather than employing information literacy competencies. However, after becoming an 8th grader, she expressed in the survey that she studied hard but still couldn't understand main ideas in texts (S2-17 survey). Her junior high teacher noted that she seldom used concept maps because she thought it was an extra homework. S2-17 depended on the scaffolding given by teachers, instead of using independent learning competencies she learned in the elementary school.

In sum, regardless of students' prior academic achievements, if they would like to employ the learned information literacy continuously, all of them can become effective learners at the junior high school level.

Implementation of the Inquiry-Based Integrated Information Literacy Curriculum in Junior High Schools

According to the collected data, the elementary inquiry-based integrated information literacy curriculum was not practiced well in the junior high schools. We can discuss this issue from the following three aspects.

Little library literacy has been infused into junior high school curriculum

Both public and private junior high schools, the library resources were not sufficient nor comprehensive. Library resources and library instruction were seldom infused into school curriculum. Most school libraries were busy with

conducting a variety of promotion activities, such as book fairs, reading reports, searching contests etc., which were one-shot experiences, rather than inquiry-based learning with meaningful interactions with the resources.

Visual & media literacy were seldom taught

Newspaper and television channels have been introduced to junior high schools recently. However, the abilities and attitudes of critically consuming and producing information were not taught in the school instruction.

Inquiry learning was rarely implemented

Many students noted in the survey that they hoped for more inquiry projects that could be developed in the junior high curriculum in order to deepen their subject content learning. But teachers teaching fast and students going to cram schools were two popular conditions in junior high schools in Taiwan. Therefore, inquiry learning was rarely implemented in class.

Conclusion

If inquiry-based integrated information literacy curriculum can be implemented systematically in elementary schools, students internalize information literacy gradually. Then when the students enter junior high schools, they most would apply information literacy in their learning and daily lives. Despite of their prior academic performance, most students can overcome problems in subject content and learning performance using information literacy competencies. These competencies include concept mapping, note-taking, searching information, asking relevant questions, integrating various perspectives, and self-reflection.

ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT

Vaughn, S., Schumm, J. S., & Sinagub, J.（1999）。焦點團體訪談：教育與心理學適用（王文科、王智弘譯）。五南圖書出版。（原著出版於1996年）。【Vaughn, S., Schumm, J. S., & Sinagub, J. (1999). *Focus group interviews in education and psychology* (Wen-Ke Wang & Chi-Hung Wang, Trans.). Wu-Nan Book. (Original work published 1996) (in Chinese)】

李榮茂（2019年7月3日）。嘉縣溪口拍微電影 為達完美NG 數十次。國語日報，16版。【Li, Rong Mao (2019, July 3). Jiaxian Xikou pai weidianying Weida wanmei NG shushici. *Mandarin Daily News*, 16. (in Chinese)】

林生傳（2003）。教育研究法：全方位的統整與分析。心理出版社。【Lin, Sheng-Chuan (2003). *Jiaoyu yanjiufa: Quanfangwei de tongzheng yu fenxi*. Psychological Publishing. (in Chinese)】

林育慈、吳正己（2016）。運算思維與中小學科技課程。教育脈動，6。【Lin, Yu-Tzu, & Wu, Cheng-Chih (2016). Yunsuan siwei yu zhongxiaoxue keji kecheng. *Pulse of Education*, 6. (in Chinese)】

林菁（2011）。資訊素養融入國小一年級「校園生物大搜索」主題探究：以Super3模式為例。教育資料與圖書館學，*48*(4)，539-570。【Chen, Lin Ching (2011). Integrating information literacy into first-grade inquiry learning: An example of creature investigation on campus using the Super3 model. *Journal of Educational Media & Library Sciences*, *48*(4), 539-570. (in Chinese)】

林菁（2018）。國小探究式資訊素養融入課程之研究：理論與實踐。教育資料與圖書館學，*55*(2)，103-137。https://doi.org/10.6120/JoEMLS.201807_55(2).0004.RS.CM【Chen, Lin Ching (2018). Elementary inquiry-based integrated information literacy curriculum: Theory & practice. *Journal of Educational Media & Library Sciences*, *55*(2), 103-137. https://doi.org/10.6120/JoEMLS.201807_55(2).0004.RS.CM (in Chinese)】

林菁、沈桂枝、賴秀珍（2016）。公民行動取向之資訊素養課程：以國小六年級「台美生態學校夥伴計畫」為例。教育資料與圖書館學，*53*(2)，211-244。https://doi.org/10.6120/JoEMLS.2016.532/005.RS.CM【Chen, Lin Ching, Shen, Kui-Chi, & Lai, Hsiu-Chen (2016). Civic action-oriented information literacy curriculum: An example of sixth-grade "US-Taiwan Eco-Campus Partnership Program." *Journal of Educational Media & Library Sciences*, *53*(2), 211-244. https://doi.org/10.6120/JoEMLS.2016.532/005.RS.CM (in Chinese)】

林菁、顏仁德、黃財尉（2014）。探究式資訊素養融入課程成效之四年長期研究。教育資料與圖書館學，*51*(4)，561-595。https://doi.org/10.6120/JoEMLS.2014.5140/0629.RS.CM【Chen, Lin Ching, Yan, Ren-De, & Huang, Tsai-Wei (2014). The effects of inquiry-based integrated information literacy instruction: Four-year trends. *Journal of Educational Media & Library Sciences*, *51*(4), 561-595. https://doi.org/10.6120/JoEMLS.2014.5140/0629.RS.CM (in Chinese)】

吳齊殷、張明宜、陳宜蒨（2008）。尋找機制與過程：長期追蹤研究的功用。αβγ量化研究學刊，*2*(1)，1-26。【Wu, Chyi-In, Chang, Ming-Yi, & Chen, I-Chien (2008). Looking for the mechanism and the process: The function of panel study. *αβγ Journal of Quantitative Research*, *2*(1), 1-26. (in Chinese)】

高雪瑩、陳品華（2014）。引導式筆記教學對提升國小五年級學生社會領域學習表現與學習態度之成效。教育研究與發展期刊，*10*(4)，79-106。https://doi.org/10.3966/181665042014121004004【Kao, Hsueh-Ying, & Chen, Pin-Hwa (2014). The effects of guided notes in enhancing fifth graders' learning performance and learning attitude in social studies. *Journal of Educational Research and Development*, *10*(4), 79-106. https://doi.org/10.3966/181665042014121004004 (in Chinese)】

陳品華（2013）。大學生課堂筆記策略教學方案之成效。教育研究集刊，*59*(1)，73-112。【Chen, Pin-Hwa (2013). The effectiveness of an integrated lecture note-taking intervention program. *Bulletin of Educational Research*, *59*(1), 73-112. (in Chinese)】

陳怡芬、林育慈、翁禎苑（2018）。運算思維導向程式設計教學—以「動手玩音樂」模組化程式設計為例。中等教育，*69*(2)，127-141。https://doi.org/10.6249/SE.201806_69(2).0023【Chen, Yi-Fen, Lin, Yu-Tzu, & Weng, Chen-Yuan (2018). Modular programming with computational thinking. *Secondary Education*, *69*(2), 127-141. https://doi.org/10.6249/SE.201806_69(2).0023 (in Chinese)】

教育部（2008）。國民中小學九年一貫課程綱要總綱。教育部。【Ministry of Education. (2008). *Guomin zhongxiaoxue jiunianyiguan kecheng gangyao zonggang*. Ministry of Education. (in Chinese)】

教育部（2014）。十二年國民基本教育課程綱要總綱。教育部。【Ministry of Education. (2014). *Direction governing for the 12-year basic education curricula*. Ministry of Education. (in Chinese)】

劉佩雲（2018）。教出讀寫素養。心理出版社。【Liu, Pay-Yun (2018). *Jiaochu duxie suyang*. Psychological Publishing. (in Chinese)】

羅世宏（2018）。關於「假新聞」的批判思考：老問題、新挑戰與可能的多重解方。資訊社會研究，*35*，51-86。【Lo, Shih-Hung (2018). A critical thinking on 'Fake News': Old problems, new challenges and possible solutions. *Journal of Information Society*, *35*, 51-86. (in Chinese)】

Accardi, M. T., Drabinski, E., & Kumbier, A. (Eds.). (2010). *Critical library instruction: Theories and methods*. Library Juice Press.

Agosto, D. E. (Ed.). (2018). *Information literacy and libraries in the age of fake news*. Libraries Unlimited.

American Association of School Librarians. (2009). *Standards for the 21st-century learner in action*. American Association of School Librarians.

Ben-David, A., & Zohar, A. (2009). Contribution of meta-strategic knowledge to scientific inquiry learning. *International Journal of Science Education*, *31*(12), 1657-1682. https://doi.org/10.1080/09500690802162762

Boyle, J. R., & Forchelli, G. A. (2014). Differences in the note-taking skills of students with high achievement, average achievement, and learning disabilities. *Learning and Individual Differences*, *35*, 9-14. https://doi.org/10.1016/j.lindif.2014.06.002

Bruce, C. (1997). *The seven faces of information literacy*. Auslib Press.

Bruce, C., Davis, K., Hughes, H., Partridge, H., & Stoodley, I. (Eds.). (2014). *Information experience: Approaches to theory and practice*. Emerald Group Publishing.

Bui, D. C., & McDaniel, M. A. (2015). Enhancing learning during lecture note-taking using outlines and illustrative diagrams. *Journal of Applied Research in Memory and Cognition*, *4*(2), 129-135. https://doi.org/10.1016/j.jarmac.2015.03.002

Chang, C.-H. (2019). Does the learning of computational thinking concepts interact with the practice of digital curation in children? *Journal of Educational Media & Library Science*, *56*(1), 45-68. https://doi.org/10.6120/JoEMLS.201903_56(1).0033.RS.CE

Chen, L. C., & Chen, Y.-H. (2019). The six frames in schools: Practices from Taiwan. In K. Ranger (Ed.), *Informed learning applications: Insights from research and practice* (pp. 5-22). Emerald Publishing. https://doi.org/10.1108/S0065-283020190000046003

Chen, L. C., Huang, T.-W., & Chen, Y.-C. (2017). The effects of inquiry-based information literacy instruction on memory and comprehension: A longitudinal study. *Library & Information Science Research*, *39*(4), 256-266. https://doi.org/10.1016/j.lisr.2017.11.003

Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2017). *21st century skills development through inquiry-based learning: From theory to practice*. Springer. https://doi.org/10.1007/978-9811024795

Chu, S. K. W., Tse, S. K., Loh, E. K. Y., & Chow, K. (2011). Collaborative inquiry project-based learning: Effects on reading ability and interests. *Library & Information Science Research*, *33*(3), 236-243. https://doi.org/10.1016/j.lisr.2010.09.008

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.

Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching*, *42*(3), 337-357. https://doi.org/10.1002/tea.20053

Downey, A. (2016). *Critical information literacy: Foundations, inspiration, and ideas*. Library Juice Press.

Duran, L. B., & Duran, E. (2004). The 5E instructional model: A learning cycle approach for inquiry-based science teaching. *The Science Education Review*, *3*(2), 49-58.

Eisenberg, M. B., & Berkowitz, R. E. (1999). *Teaching information & technology skills: The Big6 in elementary schools* (7th ed.). Linworth Publishing.

Eisenberg, M. B., Lowe, C. A., & Spitzer, K. L. (2004). *Information literacy: Essential skills for the information age* (2nd ed.). Libraries Unlimited.

Farmer, L. (2007). The school library media teacher as information literacy partner. In S. Curzon & L. Lampert (Eds.), *Proven strategies for building an information literacy program* (pp. 279-294). Neal-Schuman.

Flanagan, S. M., & Bouck, E. C. (2015). Mapping out the details: Supporting struggling writers' written expression with concept mapping. *Preventing School Failure: Alternative Education for Children and Youth*, *59*(4), 244-252. https://doi.org/10.1080/1045988X.2014.933400

Gretter, S., & Yadav, A. (2016). Computational thinking and media & information literacy: An integrated approach to teaching twenty-first century skills. *TechTrends*, *60*, 510-516. https://doi.org/10.1007/s11528-016-0098-4

Harvey, S., & Goudvis, A. (2007). *Strategies that work: Teaching comprehension for understanding and engagement* (2nd ed.). Stenhouse Publishers; Pembroke Publishers.

International Society for Technology in Education. (2016). *ISTE standards for students*. https://www.iste.org/standards/for-students

Kenney, B. F. (2008). Revitalizing the one-shot instruction session using problem-based learning. *Reference & User Services Quarterly*, *47*(4), 386-391.

Kuhlthau, C. C. (2004). *Seeking meaning: A process approach to library and information services*. Libraries Unlimited.

Kuhlthau, C. C., Maniotes, L. K., & Caspari, A. K. (2012). *Guided inquiry design: A framework for inquiry in your school*. Libraries Unlimited.

Latham, D., & Gross, M. (2008). Broken links: Undergraduates look back on their experiences with information literacy in K-12 education. *School Library Media Research*, *11*, 1-17.

Montiel-Overall, P., & Grimes, K. (2013). Teachers and librarians collaborating on inquiry-based science instruction: A longitudinal study. *Library & Information Science Research*, *35*(1), 41-53. https://doi.org/10.1016/j.lisr.2012.08.002

Peterson, E. (2009). Problem-based learning as teaching strategy. In M. T. Accardi, E. Drabinski & A. Kumbier (Eds.), *Critical library instruction: Theories and methods* (pp. 72-80). Library Juice Press.

Rockman, I. (2004). *Integrating information literacy into the higher education curriculum: Practical models for transformation*. Jossey-Bass.

Rubin, V., & Delellis, N. (2019). Educators' perceptions of information literacy and skills required to spot "fake news." *FIMS Publications*, *257*. https://doi.org/10.1002/pra2.2018.14505501116

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.

Wilson, C., Grizzle, A., Tuazon, R., Akyempong, K., & Cheung, C. K. (2011). *Media and information literacy curriculum for teachers*. United Nations Educational, Scientific and Cultural Organization.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35. https://doi.org/10.1145/1118178.1118215

Lin Ching Chen ORCID 0000-0002-6759-0645

Journal of Educational Media & Library Sciences 57 : 1 (2020) : 121-144
DOI:10.6120/JoEMLS.202003_57(1).0047.RS.CE

Research Article

The Feasibility of Automated Topic Analysis: An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Chinese Text Classification[ψ]

Yuen-Hsien Tseng

Abstract

Text classification (TC) is the task of assigning predefined categories (or labels) to texts for information organization, knowledge management, and many other applications. Normally the categories are topical in library science applications, although they can be any labels suitable for an application. Thus, TC often requires topical analysis which relies on human knowledge. However, in recent decades, machine learning (ML) techniques have been applied to TC for efficiency, as long as a sufficient number of training texts are available for each category. Nevertheless, in real-world cases, the number of texts (documents) for each category is often highly skewed for a certain TC task. This leads to the problem of predicting labels for small categories, which is viable for humans but challenging for machines. Deep learning (DL) is an emerging class of machine learning (ML) which was inspired by human neural networks. This study aims to evaluate whether DL techniques are feasible for the mentioned problem by comparing the performance of four off-the-shelf DL methods (CNN, RCNN, fastText, and BERT) with four traditional ML techniques on five skew-distributed datasets (four in Chinese, and one in English for comparison). Our results show that BERT is effective for moderately skewed datasets, but is still not feasible for highly skewed TC tasks. The other three DL-aware methods (CNN, RCNN, fastText) do not show any advantage in comparison with traditional methods such as SVM for the five TC tasks, although they captured extra language knowledge in the pretrained word representation. To facilitate future study, all of the Chinese datasets used in this study have been released publicly, together with all of the adapted machine learning and evaluation source codes for verification and for further study at https://github.com/SamTseng/Chinese_Skewed_TxtClf.

Keywords: *Text categorization, Real-world corpus, Deep learning, Performance evaluation*

[ψ] Part of manuscript has been published in Yuen-Hsien Tseng, "An Empirical Evaluation of Deep Learning Techniques Applied to Skew-Distributed Text Classification," Proceedings of the International Conference on Library and Information Science (ICLIS), pp. 303-310, Taipei, Taiwan, July 12-13, 2019 (with a total of 2,915 words and different conclusions). The current paper is a substantial revision with 8,414 words and different conclusions.

Distinguished Professor and Deputy Chair, Graduate Institute of Library & Information Studies, National Taiwan Normal University, Taipei, Taiwan
Sub-Investigator, MOST AI Biomedical Research Center, Taipei, Taiwan
E-mail: samtseng@ntnu.edu.tw

2019/10/05 received; 2020/03/08 revised; 2020/03/12 accepted

Introduction

Deep learning (DL) is a class of machine learning (ML) algorithms that use multiple layers of connected nonlinear processing units inspired by neural networks in the human brain (LeCun et al., 2015). The characteristics that distinguish DL from traditional ML is that DL techniques can automatically learn the knowledge representation features, either spatial or temporal, needed for a task. In recent years, DL techniques have been successfully applied to such areas as speech recognition, computer vision (Russakovsky et al., 2015), and natural language processing (Devlin et al., 2019). The success of DL in end-to-end applications can be attributed to the breakthrough of learning algorithms, feature representation (e.g., word embedding [Mikolov, Sutskever, et al., 2013]), and network architectures (e.g., convolutional neural networks, or CNN [Alex et al., 2012], recurrent neural networks, or RNN [Hochreiter & Schmidhuber, 1997], and transformers [Vaswani et al., 2017]). Many fields have attempted to apply DL to better deal with traditional tasks or to innovate new applications.

Text classification (TC) is the task of assigning predefined labels (or categories) to texts or documents for data analysis/mining, information browsing/searching, knowledge organization/management, and many others. A number of studies have applied deep learning to text classification (Chen & Lee, 2017; Lai et al., 2015; X. Zhang et al., 2015). Most studies have evaluated their approaches on datasets with relatively balanced document distribution, that is, the number of documents for each category is fairly even (X. Zhang et al., 2015). However, in real-world cases, texts often follow Zipf's law: the distribution of the frequencies of terms is highly skewed, as is the distribution of the documents to the categories to which they belong. In other words, most documents belong to a few categories, and most categories have very few documents. A good machine classifier should correctly predict a document's category not just for a few large categories, but also for many more small categories with limited training data in the same TC task.

The problem of correct prediction for small categories is challenging and is a long-standing issue in TC study. As DL has been successfully applied to many tasks, even better than humans in some cases, the question that follows is: can the emerging DL techniques solve this problem better than traditional methods to the extent that automated topic analysis with minimal human involvement becomes feasible. To answer this question, in this study we compared the performance of four off-the-shelf DL methods with four traditional machine learning techniques on five skew-distributed datasets.

In library science, this is an important question to answer, as information organization and topic analysis is one of the main technical services in library/

institute practices. In this age of the resurgence of Artificial Intelligence (AI), whether the recent techniques are mature enough to augment human power to an unprecedented level is worth examining and exploring. Therefore, this study is not technique-oriented (i.e., it does not aim to devise more advanced methods for further performance improvement), but rather, it examines existing tools to test their capability for immediate application.

The rest of the paper is organized as follows: The next section briefly reviews related work to further motivate this study. Section 3 clarifies the research questions and possible contributions of this work. Section 4 introduces the machine learning methods for performance comparison. In Section 5, the five real-world corpora for text classification are described. Section 6 details the experiment results. The conclusions are summarized in the final section.

Related Work

Automated text classification has been studied for decades. This section reviews some representative studies based on both the traditional ML and the DL methods. Because there is a huge body of literature on automated text classification, including research issues on document representation, feature selection, classifier construction, parameter tuning, evaluation metrics, and various applications (e.g., classification by topic, source, language, emotion, spam, etc.), our emphasis is only on those studies which experimented on topical classification for skewed datasets.

In the late 1990s, various text classifiers were proposed. Many different studies experimented on datasets of different variations, leading to slightly inconsistent conclusions. Y. Yang and Liu (1999), therefore, compared the effectiveness of five classifiers, namely SVM (Support Vector Machine), kNN (k-Nearest Neighbors), Linear Least Square Fit (LLSF), Neural Networks (NNet), and Naïve Bayes (NB), based on the real-world (skew-distributed) Reuters-21578 dataset (described later), with the results verified by a statistical significance test. Their experiments showed that SVM, kNN, and LLSF significantly outperformed NNet and NB when the number of positive training instances per category was small (less than 10), and that all the methods performed comparably when the categories were sufficiently common (over 300 instances). Therefore, a genuinely effective text classifier should perform well for categories no matter whether there are sufficient positive examples or not.

Lewis et al. (2004) extensively described a new large Reuters dataset called RCV1-v2 for text categorization research. The RCV1-v2 dataset[1] contains

[1] RCV1-v2 is available at: http://www.ai.mit.edu/projects/jmlr/papers/volume5/lewis04a/lyrl2004_rcv1v2_README.htm, accessed on 2020/02/18.

804,414 documents which were split into 23,149 training documents and 781,265 test documents (this split is called the LYRL2004 split). There are in total 103 topic categories, the distributions of which are extremely skewed: the largest category has 374,316 texts while the smallest has only five. The authors benchmarked the dataset with SVM, kNN, and Rocchio classifiers and confirmed the findings of past studies: SVM is dominant, kNN is competitive, and the Rocchio algorithm is only plausible. The best SVM performance was 81.6% for MicroF and 60.7% for MacroF (these metrics will be described later). It is noted that the MacroF value is lower than the MicroF value, meaning that many small categories were not classified well, compared to the large categories. The authors did not particularly study the skewed category problem, that is, they did not try to improve the MacroF for this dataset.

The review paper of Sebastiani (2002) summarized the TC studies around 2000. He pointed out that the evaluation of text classifiers is typically conducted experimentally, rather than analytically, due to its subjective characteristics. Therefore, the more datasets (with diverse domains) that are used in TC experiments, the more reliable and consistent the insights drawn from the experiments will be. This can be observed when sentiment analysis was studied, beginning in 2002 (Pang et al., 2002; Turney, 2002), an increasing number of datasets for sentiment analysis (most from Twitter) were made publicly available. There were at least nine by 2013, as described by Saif et al. (2013). However, these sentiment datasets have few categories, normally only three: positive, negative, and neutral. Even with so few categories, Pang et al. (2002) still stated that "studying the effect of skewed class distributions was out of the scope of this study," which confirms that skew-distributed TC is another issue that needs to be addressed in addition to those mentioned earlier (such as feature extraction, parameter tuning, etc.).

In addition to Reuters datasets, there are only a few commonly used datasets for topic-based TC study. Sebastiani (2002) introduced only three more datasets: one is proprietary (the AP collection)[2], and the other two are publicly available. These two public datasets are described as follows: 1. OHSUMED[3] is a set of 348,566 bibliographic records from 270 medical journals over a 5-year period (1987-1991) from the online medical information database MEDLINE. The available fields are title, abstract, MeSH indexing terms, author, source, and publication type. The categories are the MeSH terms or the clusters of MeSH terms under certain diseases, depending on how researchers use this dataset

[2] Details of the AP collection can be found at: http://www.daviddlewis.com/resources/testcollections/trecap/, accessed on 2020/02/18.

[3] OHSUMED is available at: https://trec.nist.gov/data/t9_filtering.html, accessed on 2020/02/18.

(Joachims, 1998). 2. The 20 newsgroups (20NG)[4] contains messages posted to 20 Usenet newsgroups. Each group has about 1,000 texts and thus there are a total of approximately 20,000 texts, which make it a balanced dataset. Considering the characteristics of these datasets, it can be seen that there are insufficient real-world datasets for experiments on skew-distributed TC tasks.

With the resurgence of AI research momentum (more data and more computation power have become easily accessible), various DL techniques equipped with pre-trained word embedding, sequence recognition, and attention mechanisms for semantic understanding have been proposed for better performance in various TC-related tasks. Example DL techniques include CNN, RCNN, fastText, and BERT, as described below.

Johnson and Zhang (2015) proposed Convolutional Neural Networks (CNN) for sentiment prediction of movie reviews (dataset name: IMDB), Amazon's electronic product reviews (dataset name: Elec), and for topical classification of news (dataset name: RCV1-v2). They reported the error rates (number of texts with incorrect category prediction divided by the total number of texts) for these three datasets. Additionally, they reported MicroF and MacroF measures for RCV1-v2 with 84.0% and 64.8%, respectively. This performance is obviously better than that of Lewis et al.'s (2004) results mentioned above. Therefore, CNN is a strong classifier that is worth testing on more topic-based TC tasks in different domains.

Lai et al. (2015) proposed Recurrent Convolutional Neural Networks (RCNN) and experimented with this new approach on four datasets, namely, 20NG for topical discussion group classification, the Fudan set for Chinese TC with 20 topical categories, the ACL Anthology Network for the prediction of five native languages of the authors of scientific papers, and the Stanford Sentiment Treebank for predicting five sentiment levels for movie reviews. Although they do not report the skewedness of these datasets, their use of accuracy (number of correctly classified texts divided by number of all texts) as the main performance metric (for the last three datasets) suggests that the datasets are balanced (otherwise the use of accuracy would be an improper measure). They did use MacroF for the 20NG, which is a balanced dataset. Therefore, how well RCNN will perform on skewed datasets is unknown.

Another approach, fastText (Joulin et al., 2016) reported accuracy on eight sentiment analysis datasets and reported precision of 1 (denoted as Prec@1, i.e., the number of correct predictions in the first place when ranking the tags) on a huge tag prediction dataset called YFCC100M where there are 312,116 unique

[4] 20NG is available at: http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html, accessed on 2020/02/18.

image tags for prediction (each tag occurs more than 100 times) based on the image titles and captions with 91,188,648 training examples and 543,424 testing examples. It was found that fastText performed better or was often on a par with various DL classifiers in terms of accuracy, and many orders of magnitude faster for training and evaluation. The accuracy or Prec@1 did not reveal how well fastText could boost the performance of low-frequency categories.

BERT (Devlin et al., 2019) obtained state-of-the-art results on 11 natural language processing tasks, most of which were related to sentence pair inferencing (the task names are: QQP, QNLI, STS-B, MRPC, and WNLI) or entailment prediction (MNLI, RTE, and SWAG). Two tasks were to predict a single sentence's sentiment (SST-2) or linguistic acceptability (CoLA), and two were about a question answering test (SQuAD v1.1 and SQuAD v2.0). In other words, the 11 tasks consisted of a variety of semantic classification tasks, but none were related to topical classification which also needs semantic understanding for better text classification.

To compare the performance of different classifiers, Sebastiani (2002) also pointed out that different sets of experiments may be used for cross-classifier comparison only if the experiments have been performed: 1. on exactly the same collection (i.e., the same documents and same categories); 2. with the same "split" between training set and test set; and 3. with the same evaluation measure. If these protocols are not followed, it is likely that different authors in their individual papers may report incomparable results because the classifier implementation details are often insufficient to reproduce the same result (e.g., how exactly features are selected, stop words are used, stemming is performed, parameters are chosen, etc.). Therefore, to claim that a new technique performs better than previous ones whether on new or old datasets, three options are viable: 1. reporting an **obvious** improvement on the same datasets over previous techniques without the need for statistical significance tests (due to the lack of previous example-wise results); 2. cooperating with the authors of previous techniques to use their exact techniques on the same datasets for comparison; and 3. re-implementing previous techniques to the extent that it approximates the previous performance on the same datasets, and then comparing the performance of the re-implementation with that of the new technique. Option 1 has been widely used; one example is the study by Johnson and Zhang (2015) mentioned above. Option 2 is relatively rare, but it encourages research cooperation, such as the study by Tseng and Teahan (2004). Option 3 is less reliable, as pointed out by Sebastiani (2002), but it is somewhat inevitable when novel techniques or new datasets are introduced. Option 3 was used in X. Zhang's (2015) study, in which they re-implemented a multinomial logistic regression classifier, a word-based CNN, and

an RNN based on Long-Short Term Memory (LSTM; Hochreiter & Schmidhuber, 1997) to compare their proposed character-level CNN. In addition, they collected eight large-scale balanced datasets for their experiments, ranging from hundreds of thousands to several millions of samples. In our work, Option 3 was used, because four more real-world Chinese datasets were introduced.

To sum up this brief review, four important notes can be re-stated:

1. Skew-distributed TC is an issue worth studying.
2. A genuine effective text classifier should perform well on all categories, no matter whether there are sufficient positive examples or not.
3. Because of the empirical nature of TC study, more publicly available real-world datasets for experiments on skew-distributed TC tasks are needed, especially for traditional Chinese.
4. When comparing classifiers it should be borne in mind that the datasets, preprocessing, and parameters should be transparent (detailed enough) for future comparison or verification to facilitate the advancement of TC research.

Research Questions

Based on the above introduction and review, we clarify our research questions as follows:

RQ1. Do the latest deep learning techniques perform better than traditional machine learning methods in real-world, topic-based, and skew-distributed text classification tasks?

RQ2. Are there any real-world datasets for which certain deep learning techniques exhibit better performance for skewed category distribution, and in what context?

In the process of pursuing empirical answers to the above questions, the possible contributions of this work are to:

1. Introduce more real-world datasets, especially in Chinese, for text classification research, and to release the datasets and TC codes publicly for future comparative studies.
2. Provide empirical reports on how well deep learning techniques perform comparatively over these datasets, and especially their effectiveness versus their efficiency, in order to suggest a guideline and to reveal the feasibility of the current deep learning techniques for TC practitioners and researchers.

Machine Learning for Automated Text Categorization

The ML application to text classification follows these pipeline steps: 1.

dataset preparation, 2. feature extraction, 3. model training, and 4. performance evaluation. This section introduces the second and third steps, leaving steps 1 and 4 to the next sections.

For ease of comparison, we adapted the Python code from https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/ for the traditional machine learning methods: Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), and single hidden-layer neural networks (NN), and also for the deep learning methods: Convolutional Neural Networks (CNN; Johnson & Zhang, 2015) and Recurrent Convolutional Neural Networks (RCNN; Lai et al., 2015). For fastText (Joulin et al., 2016), we used the open-source library released by Facebook's AI Research (FAIR) to learn text representations and text classifiers. For BERT (Bidirectional Encoder Representations from Transformers; Devlin et al., 2019), which is effective for a wide array of natural language processing tasks, we used the publicly available text classification code example released by Google Research in November 2018. All codes were modified for Chinese and enhanced for our experimental flow and analysis.

Feature Extraction

The texts in each dataset were first cleaned and standardized by tokenization, segmentation, lowercasing, and stop-word removal for all the machine classifiers other than BERT (BERT only needs original text input and lower-case English words). Each text was then transformed into a feature vector with each element representing a term in the dataset. The term's value can be the term's occurring frequency (term frequency, TF) in a text, or the normalized TF multiplied by the logarithm of the inverse document frequency (IDF) of the term in the whole dataset (TFxIDF; Salton, 1989). The term here can be a normal word, N consecutive words, or N consecutive characters. With these different values for a term and different ways to denote a term, there are four feature vectors (or feature sets) commonly used: 1. Word Count: the term represents a normal word and its value is the word's TF; 2. TFxIDF: the term is a normal word and its value is the word's TFxIDF; 3. Word N-grams: the term is an N consecutive word in a text and its value is the term's TFxIDF; and 4. Char N-gram: the term is an N consecutive character in a text and its value is the term's TFxIDF.

In contrast to the above large sparse feature vectors (e.g., larger than 10,000 dimensions) where semantically similar terms may have no intersection in their vector representation, Word2Vec word embedding (Mikolov, Chen, et al., 2013) is a special way to transform a word into a relatively small dense vector (e.g., 300 dimensions), where semantically similar terms are also similar in their embedding vectors (e.g., similar cosine similarity between the corresponding

word embedding vectors). Word embeddings can be directly trained using the input corpus or can be obtained from pre-trained word embeddings (such as those provided by Google's word embedding vectors or Facebook's fastText trained with very large text corpora) followed by fine-tuned training from the input corpus for text classification.

Machine Learning Models

Four traditional ML techniques were used, namely Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), and the single hidden-layer neural network (NN). These are all mature ML methods widely used before the DL methods were introduced. Their technical details can be found on the Web (e.g., Wikipedia) or in related textbooks (Witten et al., 2011).

For the DL techniques (please refer to the above-mentioned URL, from which we have fixed some bugs and added some code for dealing with Chinese), the first one is based on Convolutional Neural Networks (CNN), where a pretrained Word2Vec word embedding with trainable weights for the classification task is the first layer. This is followed by a 1-d convolution layer to convolve the embedding vectors to extract local contextual information of the input word embeddings. This becomes the input to a max pooling layer to summarize the local contextual information. A dense hidden layer is followed to map the summarized information to a category prediction layer, which uses softmax as its activation function and is thus called the softmax output layer. In sum, this CNN deep neural network has 5 layers. The pretrained Word2Vec files were downloaded from https://fasttext.cc/docs/en/pretrained-vectors.html, which was released by Facebook and is one of many Word2Vec files freely available on the Web.

For the second DL technique, Recurrent Convolutional Neural Networks (RCNN), the first layer is the same pretrained word embedding layer with trainable weights for the application task, which is followed by a 1-dimensional convolution layer. The next layer is a bi-directional GRU (Gated Recurrent Unit) to further extract longer dependent information in the text, which is followed by another 1-dimensional convolution layer, a max pooling layer, and then a dense hidden layer. The final output layer is a softmax layer. In sum, this RCNN has six layers.

The third DL technique is fastText. Although it may not be a full DL method, it combines some of the most successful concepts of natural language processing, such as word embedding and machine learning. It uses a hierarchical classifier instead of a flat structure, in which categories are organized into a tree. FastText exploits the fact that classes are imbalanced by using the Huffman

algorithm to build the tree used to represent the categories. The depth of the tree for very frequent categories is therefore less than that for infrequent categories, leading to further computational efficiency.

The Word2Vec word embedding vectors mentioned above have two deficiencies: 1. the homograph problem: a word with various different meanings has the same embedding vector, leading to incorrect meaning representation of a homographic word; 2. the Out-of-Vocabulary (OOV) problem: if a word in a text classification corpus does not belong to a pretrained Word2Vec vocabulary, the OOV word would have no correct embedding vector to use. The emergence of the BERT technique is able to overcome these two deficiencies, by outputting different embedding vectors for a homographic word depending on its context and by learning the meaning of the sub-word representation of a word (which alleviates the OOV problem). In addition, one of the major techniques used in BERT is the attention mechanism (Vaswani et al., 2017) which allows BERT to focus on the important words for the application task. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as text classification. However, due to its complexity using 12 layers of transformers with a total of 110 million parameters for fine-tuning during training, the currently released BERT model takes as input a document with a maximum length of only 512 English words or 512 Chinese characters (no Chinese segmentation is needed). Despite this limitation, the document length is mostly sufficient as one can often tell the topics of a document by looking at only the first few sentences. We slightly modified the text classification Python code released by Google to accommodate our datasets for topic training and for prediction by BERT.

For all of the above classifiers, we adopted the default values as far as possible. The training epoch was set to 20 (the classifier sees each of the training texts 20 times) for all the DL techniques except fastText. With these settings, the classifiers achieved over 95% accuracy for classifying the training set at epoch 20. We believe that this training epoch number is large enough to obtain good results and to prevent the classifier from overfitting by restraining its training time. For fastText, we followed the tutorial instructions and tried out various settings until we could not obtain better results.

The classifiers were tested and verified on a Chinese binary classification corpus (we call it the CnonC dataset). Its text per document is merely a project title, and the corresponding labels are either Construction or Non-Construction. The training set has only 232 examples and the testing set has 100 examples. The category distribution is balanced: half of the examples belong to Construction and the others belong to the Non-Construction category. All the classifiers performed

normally, with MicroF1 ranging from 0.87 to 0.93. This verification process suggested that all of our eight classifiers were valid programs (without bugs) for the later experiments and that the training set did not need to contain a large number of examples to achieve high performance (the consistent characteristics between the training and testing set is more important than the number of training examples). This CnonC corpus is also an ideal small dataset for early prototyping of the classifier design (either for choosing the parameters or architecture).

Datasets

In this study, five text datasets were used. Four of them were in Chinese and one in English for comparison. Their basic statistics are shown in Table 1.

Table 1 Basic Statistics of the Five Datasets

Dataset	Train docs.	Test docs.	Categories	Avg. chars./ words	Diversity	Divratio %
WebDes_1686	1,190	496	26	75	9.4	36.30
News_914	644	270	12	502	5.1	42.32
Joke_3414	2,389	1,025	9	118	7.4	81.94
CTC_27320	19,266	8,054	81	800	17.4	21.78
Reuters_9130	6,561	2,569	52	134	3.9	7.57

Each of the five datasets was split into a training subset and a testing subset such that the testing subset contained 30% of the total documents for each category. The number of classes in the dataset is shown in the Categories column. The average length of the texts for each dataset is shown in terms of the number of Chinese characters or English words. The Diversity and DivRatio columns indicate the skewedness of the dataset. Diversity was defined by Simpson (1949; and is also known as the Herfindahl–Hirschman Index, or HHI; Calkins, 1983; Hirschman, 1964) as:

$$HHI = \sum_{i=1}^{n} s_i^2$$

where s_i denotes the share of category i in a dataset (percentage of number of texts belonging to category i) and n is the total number of categories. For example, suppose there are totally three (n = 3) categories with 70, 20, and 10 texts in a dataset, in each dataset, their shares would be 0.7, 0.2, and 0.1, respectively. HHI is proportional to the average share, weighted by individual share. Therefore, for this example, the HHI is 0.49 + 0.04 + 0.01 = 0.54. As such, it ranges from 1/n to 1.0. Interestingly, the reciprocal of the index (e.g., 1/HHI), which corresponds to the Diversity column, indicates the "equivalent" number of dominant categories (Liston-Heyes & Pilkington, 2004). For the example with HHI = 0.54, it means that there are equivalently only 1.85 (= 1/0.54) categories that are used to label the texts in the dataset. Since each dataset has a different category number (n),

we divide the Diversity index by the number of categories to yield the value in the DivRatio column to indicate the percentage of categories used to label the texts in the dataset. This DivRatio is, in our view, a reasonable quantitative value to indicate the skewness of the categories, where a 100% DivRatio indicates that each category has an equal number of texts.

WebDes_1686: This dataset contains Chinese webpage descriptions from an internet portal. It was used to evaluate machine classifiers before their deployment for daily use. It contains 1,190 texts for training and 496 for testing, with a total of 1,686 texts; thus, it is named WebDes_1686. This naming convention also applies to the datasets described below. This dataset has an average of 75 Chinese characters (or English words) for each text. It has 26 categories, but on average only 9.4 categories were used to label the 1,686 texts, which is equivalent to having 36.30% categories mostly used, while the other 63.70% of categories have relatively few texts.

News_914: This is a Chinese news dataset from the same portal mentioned above. Hence, the purpose is the same: to seek assistance from machine learning to guide daily incoming news to desired categorical webpages for news browsing or subscription by the portal users.

Joke_3414: This dataset containing humorous texts was collected over the past 2 years from over 40 sources including 27 websites, 11 joke books, and three joke apps for possible use in humor generation applications. A joke is humorous only when it is applied in a proper context. To help decide the right context, the corpus was manually classified into nine categories by at least two people (all majoring in Library and Information Science). Two annotators classified each joke independently based on the category definition. When there was inconsistency (62 jokes had inconsistent categories), a third annotator helped label these jokes. The majority of the categories among the three was then assigned to the joke. The inter-rater agreement for the two major annotators had a Cohen's kappa coefficient as high as 0.97.

CTC_27320: The source of this dataset was the news broadcasts of Mainland China's radio stations between 1966 and 1982. These broadcasts were transcribed manually and labeled by domain analysts. The purpose of the labeling is document organization, browsing, and retrieval. The validity of this Chinese corpus was verified by Tseng and Teahan (2004).

Reuters_9130: This English dataset contains Reuters newswires from 1987. It was originally collected and labeled by Carnegie Group, Inc. and Reuters, Ltd. during the course of developing the CONSTRUE text categorization system, and is publicly available at: http://www.daviddlewis.com/resources/testcollections/reuters21578/.

The original CTC and Reuters datasets are both multi-labeled (i.e., a text can belong to multiple categories), although most documents are single-labeled. To exclude the effect of prediction uncertainty in this study, we removed the multi-labeled documents, which reduced the number of categories from 82 to 81 and the number of total documents from 28,013 to 27,320 for the CTC_27320 dataset, and from 90 to 52 and 10,788 to 9,130 for Reuters_9130. To distinguish the original datasets from those used in this study, we appended the number of used documents to the name of the dataset.

The number of texts for each category is depicted in Figures 1 to 5 to visually show their skewness. As can be seen from the figures, the document distributions of the categories in all the datasets are skewed, with the last two being highly skewed (the small categories have only one or two documents compared to the largest categories with thousands of documents).

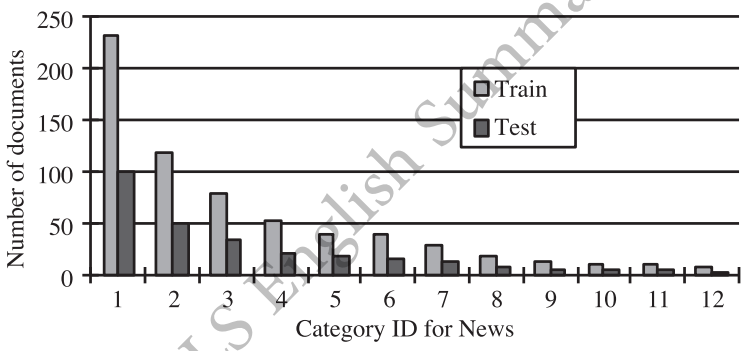


Figure 1 Number of Documents for Each Category in the News_914 Dataset

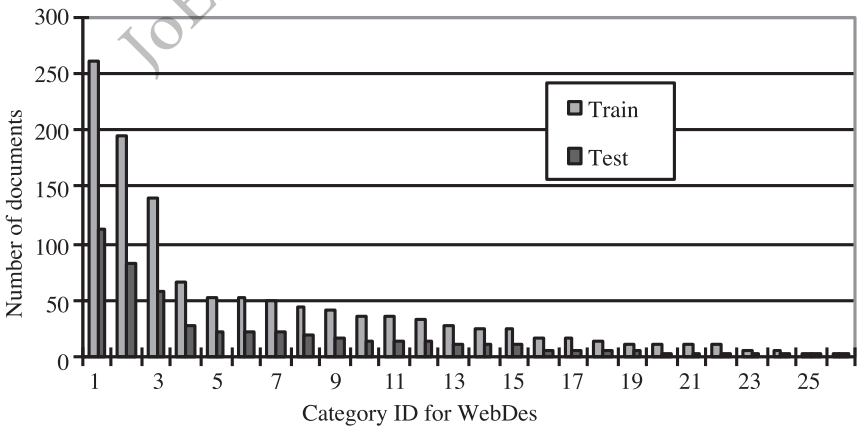


Figure 2 Number of Documents for Each Category in the WebDes_1686 Dataset

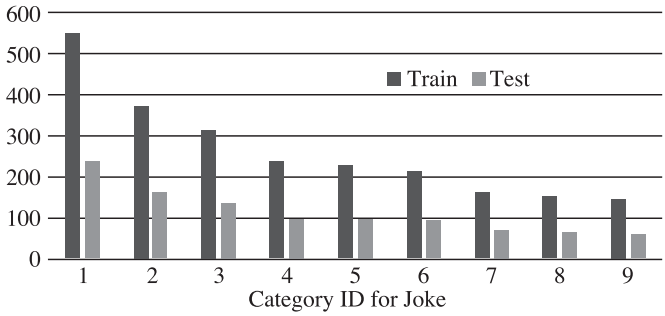


Figure 3 Number of Documents for Each Category in the Joke_3414 Dataset

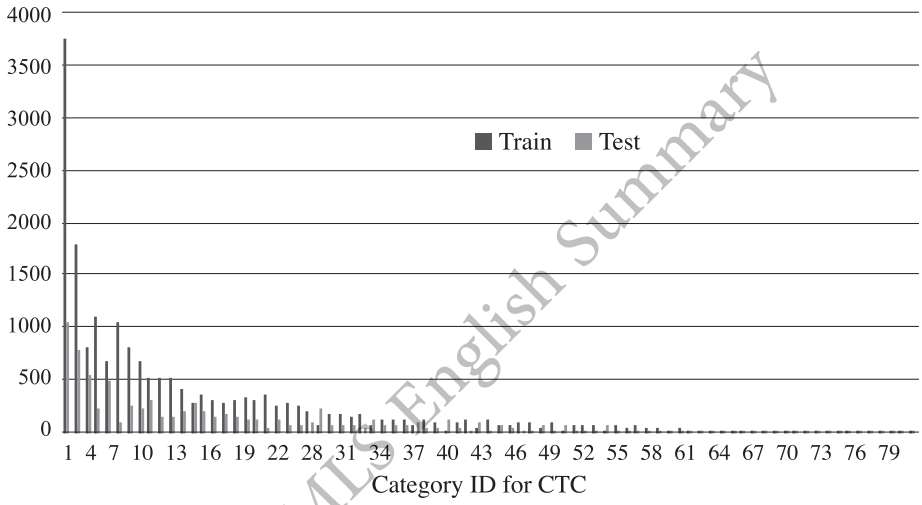


Figure 4 Number of Documents for Each Category in the CTC_27320 Dataset

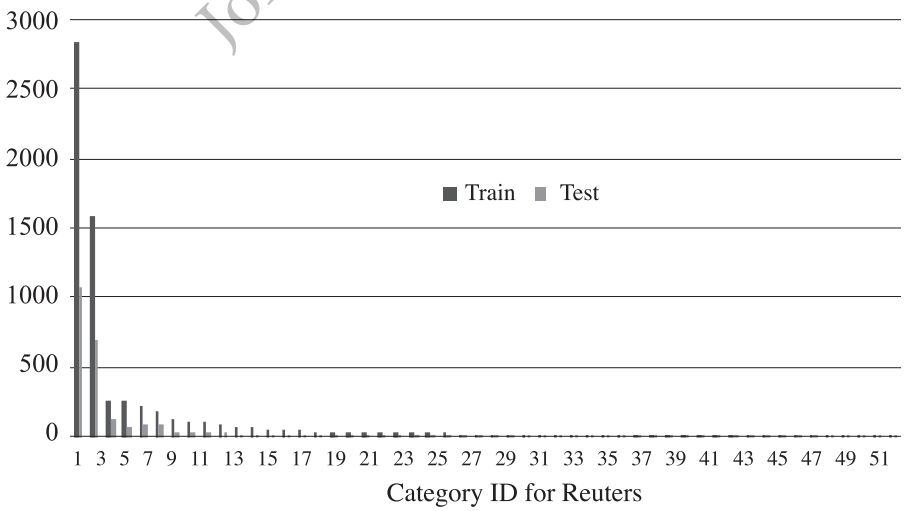


Figure 5 Number of Documents for Each Category in the Reuters_9130 Dataset

Evaluation Results

Two metrics were used for the performance comparison: MicroF1 and MacroF1. For both metrics, a confusion matrix like the one shown in Table 2 was computed for each category, where each cell represents the number of documents in the corresponding case (TP, FP, FN, or TN). The precision and recall were then calculated based on the matrix. The F1 score is the harmonic average of the precision and recall. The Micro- and Macro-averages were computed in different ways, and thus their interpretation differed. A macro-average will compute the metric independently for each category and then take the average (hence weighting all categories equally), whereas a micro-average will aggregate the contributions of all categories to compute the average metric.

Table 2 The Confusion Matrix for a Category

For a category *i*	Machine prediction as Yes	Machine prediction as No
Human labels as Yes	True Positive (TP_i)	False Negative (FN_i)
Human labels as No	False Positive (FP_i)	True Negative (TN_i)

Below are the detailed calculating equations for the micro-precision and micro-recall:

$$MicroPre = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i}$$

$$MicroRec = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FN_i}$$

assuming there are *n* categories. MicroF1 is then calculated by the equation:

$$MicroF1 = \frac{2 \times MicroPre \times MicroRec}{MicroPre + MicroRec}$$

In contrast, the macro-precision and macro-recall are:

$$MicroPre = \frac{1}{n} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i}$$

$$MicroRec = \frac{1}{n} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FN_i}$$

MacroF1 is then calculated by the equation:

$$MicroF1 = \frac{2 \times MicroPre \times MicroRec}{MicroPre + MicroRec}$$

Based on the above calculation, MicroF1 measures overall document classification effectiveness and thus reveals the effectiveness of a few major categories. In contrast, MacroF1 takes each category's effectiveness into

Table 3 Performance of Various Classifiers With Different Features on Five Datasets

Dataset	Models	Features	MicroF1	MacroF1	Time (seconds)
WebDes_1686	NB	Word Count	0.8004	0.6129	0.01
	SVM	TFxIDF	0.8346	0.7160	0.07
	RF	Word Count	0.8044	0.6203	0.06
	NN	Word Count	0.8245	0.6856	5.15
	CNN	Word Embedding	0.7943	0.6124	8.91
	RCNN	Word Embedding	0.7883	0.6001	8.56
	fastText	Word Bi-gram	0.8024	0.6281	< 1
	BERT		**0.8730**	**0.7684**	3,332.31
News_914	NB	Word Count	0.7444	0.6244	0.01
	SVM	Char N-gram	0.7851	0.7388	0.16
	RF	TFxIDF	0.6111	0.3648	0.07
	NN	Char N-gram	**0.8037**	0.7400	3.98
	CNN	Word Embedding	0.6444	0.4231	27.69
	RCNN	Word Embedding	0.6518	0.4135	27.94
	fastText	Word Bi-gram	0.7518	0.6186	< 1
	BERT		**0.7963**	**0.8283**	4,106.27
Joke_3414	NB	Word Count	0.5229	0.4297	0.77
	SVM	Char N-gram	0.5417	0.4705	0.84
	RF	Char N-gram	0.4692	0.3718	1.09
	NN	Char N-gram	0.5121	0.4526	33.96
	CNN	Word Embedding	0.4985	0.4266	922.02
	RCNN	Word Embedding	0.4720	0.4144	4,787.52
	fastText	Word Bi-gram	0.5036	0.4195	5.05
	BERT		**0.6429**	**0.6183**	13,108.68
CTC_27320	NB	Word Count	0.4243	0.1669	0.96
	SVM	TFxIDF	**0.4963**	**0.3706**	32.12
	RF	Char N-gram	0.3507	0.1560	18.37
	NN	Char N-gram	0.4559	0.3312	94.51
	CNN	Word Embedding	0.3605	0.1709	3,580.61
	RCNN	Word Embedding	0.3559	0.1699	3,579.44
	fastText	Word Bi-gram	0.4557	0.2838	< 5
	BERT		0.4217	0.3394	107,890.33
Reuters_9130	NB	Word Count	0.8824	0.4112	0.08
	SVM	TFxIDF	**0.9513**	**0.7573**	1.42
	RF	TFxIDF	0.8450	0.3556	0.94
	NN	TFxIDF	0.9482	0.7541	25.44
	CNN	Word Embedding	0.9396	0.6906	175.21
	RCNN	Word Embedding	0.9377	0.6885	175.20
	fastText	Word Bi-gram	0.9369	0.6784	< 3
	BERT		0.9124	0.4581	16,223.65

consideration and thus reveals the effectiveness of most minor categories. Table 3 shows the performance figures for all the classifiers on the five datasets.

Because DL techniques use numerous trainable parameters, we ran the classifiers on a virtual machine (VM) on the Google Cloud Platform. The VM has a NVIDIA Tesla K80 GPU with 12 GB RAM. The execution time for each

method is also reported in Table 3 to reveal the computation power required for each method. The fastText tool does not support GPU and is fast enough to run on a MacBook computer. Therefore, only its execution time is available for reference.

In Table 3, we show the performance metrics for each ML method for all five datasets. If there are multiple feature sets that can be applied to a classifier (e.g., NB, SVM, RN, NN, and fastText), only the one that leads to best performance is shown for clarity. The time spent on training and predicting the datasets is also shown to reveal the cost of applying a certain machine learning technique. Note that BERT may output slightly different results even with the same hyper-parameters (such as learning epochs) due to its randomness during its training and prediction process. Nevertheless, the output variation of BERT is insignificant compared to its performance gap with other classifiers.

In terms of efficiency, all the DL techniques are time-consuming, except for fastText. Generally, the more layers of the neural networks that are used, the more time is needed to train and test the datasets. As an example, BERT spent far more time than the others on the CTC_27320 dataset: it took about 30 hours (107,167 seconds) on a K80 GPU to train 19,267 documents and about 12 minutes (723 seconds) to predict 8,054 test documents. In terms of the cost of using the Google Cloud Platform (GCP), it took about US$30 to train and predict the CTC_27320 and Reuters_9130 datasets using GCP's VM and GPU. Therefore, only when more powerful and cheaper computation facilities to train BERT are available, can it be used to label the text categories, as its prediction time is still within the acceptable limit (predicting about 11 documents per second for the CTC_27320 dataset).

In terms of effectiveness, BERT excels for the Joke_3414, News_914, and WebDes_1686 datasets. For the Joke_3414 dataset with nine categories, both BERT's MicroF1 and MacroF1 are far better than the other classifiers, with an absolute improvement of 10.12% (0.6429-0.5417) for MicroF1 and an absolute 14.78% (0.6183-0.4705) improvement for MacroF1 compared with the second best SVM classifier. This is a great enhancement achieved by BERT. However, compared to the human annotators' consistency (inter-rater agreement is 0.97 in Cohen's kappa coefficient), all the machine classifiers have a great deal of room to improve because we hypothesize that the upper bound performance for a machine classifier would be higher than 0.64 in the Joke_3414 dataset (as the inconsistency gap between humans is very small, meaning that it is easy for humans to determine their topical category based on the texts).

For the News_914 dataset with 12 categories, BERT not only performs the best, but also its MacroF1 (0.8283) exceeds its MicroF1 (0.7963). This is another unparalleled performance achieved by BERT, as it is very difficult for MacroF1 to exceed MicroF1 in a skewed dataset. As can be seen from Table 4, the training

Table 4 Breakdown Effectiveness for Each Category in the News_914 Dataset by BERT

	Precision	Recall	F1-score	Num. of test texts	Num of training texts
004-Industry	0.7925	0.8485	0.8195	99	232
003-Finance	0.6304	0.5800	0.6042	50	117
001-Politics	0.8889	0.7273	0.8000	33	78
002-Society	0.8750	0.9545	0.9130	22	53
009-Life	0.6842	0.7647	0.7222	17	40
006-Entertainment	0.9375	1.0000	0.9677	15	38
008-Local	0.8462	0.9167	0.8800	12	29
005-Technology	1.0000	0.7143	0.8333	7	18
007-Sports	1.0000	1.0000	1.0000	4	12
010-Medicine	1.0000	1.0000	1.0000	4	10
013-Education	1.0000	1.0000	1.0000	4	10
012-Leisure	0.5000	0.3333	0.4000	3	7

examples for the smallest five categories have fewer than 20 texts (10 times less than that of the largest categories) and BERT can predict them very well, except for the last category. We hypothesize that 1. for the smallest four categories with F1 over 0.8, it is easy to tell their topical themes based on the news story; and therefore 2. BERT can perform well despite only a small set of training examples being available, because it was pre-trained with a large set of texts and was able to capture the contextual meanings of Chinese characters to some extent.

Actually, BERT was trained on very large corpora (BooksCorpus with 800 million words and English Wikipedia with 2,500 million words for the English model alone) for a very long time (4 days on 4 to 16 Cloud TPUs) using bidirectional contextual information based on a deep network architecture (12 layers of attention-based and position-aware transformers). The resulting pre-trained word representations are thus context-aware, which captures more precise semantics than the context-free word representations of Word2Vec. For example, the word embedding vector of "bank" in "bank corruption" and "river bank" is different for BERT but is the same for Word2Vec, which apparently led to better semantic processing for BERT over Word2Vec.

However, these word embeddings capture only vague (or latent) semantics, especially for those words that do not occur often enough in the training corpora. The embeddings are 300-dimensional (for Word2Vec) or 768-dimensional (for BERT) real-value vectors. One cannot tell the meaning only by looking at these vectors. They need to be compared or processed in a task such as text classification to show their usefulness.

For the highly skewed datasets, such as CTC_27320 and Reuters_9130, BERT performs more poorly than the traditional classifier SVM. One of the techniques to build the language knowledge in BERT is to mask an English word

(or a Chinese character) in a sentence and try to predict the masked word by its contextual words during its training. Although much word knowledge can be captured by this technique, still a considerable amount of language information other than this kind of co-occurrence is not easily captured, such as phrases and multi-word name entities. This may be the reason why BERT is not the optimal technique. Another possible reason is that the terms or sentences used in the texts of CTC_27320 and Reuters_9130 occur rarely in the pre-training data of BERT, because texts in CTC_27320 were collected between 1966 and 1982 and texts in Reuters_9130 contain many business news abbreviations from 1987, such as qtr (for quarter), dlrs (for dollars), etc. This vocabulary mismatch may have degraded BERT's performance in these two datasets. Even so, the exact reason why BERT failed on CTC_27320 and Reuters_9130 still needs further exploration. The commonly observed unstable performance of deep neural networks based on latent semantics trained in the embedding vectors (BERT seems not to be an exception) is a disadvantage of applying the current DL technology.

Based on Table 3, SVM is a competitive technique in TC tasks, even for Chinese, which confirms past studies that SVM is a dominant text classifier. Given a proper feature set, SVM is able to perform well because of its innovative idea of creating support vectors that maximize the margin to separate categories and therefore is capable of generalizing well to unseen examples. Although not shown here, we observed that even SVM may perform poorly if supplied with an incorrect feature set, such as word n-grams on these datasets. However, this disadvantage is relatively better than the word embedding based DL methods, because the feature sets supplied to SVM are controllable and interpretable, such that we know where to improve (e.g., by trying another feature set as we did in this experiment).

From Table 3, we also observe that: 1. When efficiency is a concern, SVM is generally the most effective classifier if a suitable feature set is used for a particular TC task. 2. A single hidden-layer neural network (NN) can compete with SVM at the cost of larger computation power. 3. The DL methods (CNN, RCNN, and fastText) did not show any advantage in these five tasks, even though they utilized pre-trained embedding vectors that embedded additional language knowledge. 4. NB is a method that is obviously only feasible for large categories, because the MacroF1 gap between NB and the others is much larger than its MicroF1 gap, which is quite reasonable due to the nature of NB. 5. This phenomenon of NB does not occur to the DL methods, at least not for fastText. 6. RF is not a good text classifier, because it performed worst on four datasets.

Conclusion

In answer to our first research question, and also as a general guideline for

ML-based text classification, SVM and BERT can be the first choices for any new dataset based on the above experiments, discussions, and past studies. For SVM, the choice of feature sets is important. For BERT, only the first 512 characters (or words) will be considered currently. Therefore, if the texts in the new dataset are too long, care must be taken to see if the first 512 characters could be used to determine their topics. Also, be aware of the corpora used to pre-train the BERT model. If the characteristics of the training corpora are very different from those of the new dataset, BERT might not yield good results, although this is unlikely to happen with modern texts. Also, if computational resources are scarce, SVM is preferable to BERT, and vice versa.

In response to our second research question, the problem of correct prediction for very small categories (which is viable for humans) is feasible for BERT for some datasets with commonly seen topics such as those in the news, but it may be necessary to seek more varieties of deep learning methods for other older datasets such as CTC_27320 and Reuters_9130, or to resort to other solutions such as few-shot learning (Yan et al., 2018) or a combination of various approaches.

Actually, subsequent improvement of BERT has been made in ERNIE (Enhanced Representation through kNowledge IntEgration; Sun, Wang, Li, Feng, Chen, et al., 2019) released by Baidu, another ERNIE by Tsinghua University in Beijing (Z. Zhang et al., 2019), XLNet by Carnegie Mellon University/Google Brain (Z. Yang et al., 2019), and ERNIE 2.0 by Baidu (Sun, Wang, Li, Feng, Tian, et al., 2019). Because these improvements were made in recent months, small updates to their publicly released codes occur from time to time. Therefore, they were not used in this current study.

Note that if the explored datasets are not diverse enough, it would easily lead to a different or biased conclusion. An example biased conclusion might be that BERT is most effective for Chinese TC tasks if only the first three datasets (WebDes_1686, News_914, and Joke_3414) are used. Therefore, the five Chinese text classification datasets (the four in Table 1 and the CnonC for verification) together with the TC source codes have been released at https://github.com/SamTseng/Chinese_Skewed_TxtClf. They would facilitate future study to further examine the strengths and possible weaknesses of the new techniques (e.g., ERNIE, XLNet, etc.). In this way, we would be more knowledgeable to determine the feasibility of automated topic analysis in library/institute practice, and therefore to decide when and in what ways human effort should be involved in the tasks related to information organization and topic analysis.

Acknowledgement

This work was supported in part by the grant MOST 107-2221-E-003-014-

MY2, MOST-109-2634-F-002-023 and MOST AI Biomedical Research Center from Ministry of Science and Technology of Taiwan.

References

Alex, K., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, & L. Bottou (Eds.), *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems* (Vol. 1, pp. 1097-1105). Curran Associates. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Calkins, S. (1983). The new Merger Guidelines and the Herfindahl-Hirschman Index. *California Law Review*, *71*(2), 402-429. https://doi.org/10.2307/3480160

Chen, L., & Lee, C. M. (2017). Predicting audience's laughter using convolutional neural network. arXiv. https://arxiv.org/abs/1702.02584v2

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. https://arxiv.org/abs/1810.04805v2

Hirschman, A. O. (1964). The paternity of an index. *The American Economic Review*, *54*(5), 761.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In C. Nédellec & C. Rouveirol (Eds.), *Machine learning: ECML-98: 10th European Conference on Machine Learning Chemnitz, Germany, April 21-23, 1998 proceedings* (pp. 137-142). Springer. https://doi.org/10.1007/BFb0026683

Johnson, R., & Zhang, T. (2015). Effective use of word order for text categorization with convolutional neural networks. In R. Mihalcea, J. Chai, & A. Sarkar (Eds.), *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 103-112). Association for Computational Linguistics. https://doi.org/10.3115/v1/N15-1011

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. arXiv. https://arxiv.org/abs/1607.01759v3

Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In *AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 2267-2273). Association for the Advancement of Artificial Intelligence.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444. https://doi.org/10.1038/nature14539.

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, *5*, 361-397.

Liston-Heyes, C., & Pilkington, A. (2004). Inventive concentration in the production of green technology: A comparative analysis of fuel cell patents. *Science and Public Policy*, *31*(1), 15-25. https://doi.org/10.3152/147154304781780190

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv. https://arxiv.org/abs/1301.3781v3

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems* (Vol. 2, 3111-3119). Curran Associates.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (Vol. 10, pp. 79-86). Association for Computational Linguistics. https://doi.org/10.3115/1118693.1118704

Russakovsky, O. Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Li, F.-F. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211-252. https://doi.org/10.1007/s11263-015-0816-y

Saif, H., Fernández, M., He, Y., & Alani, H. (2013). Evaluation datasets for Twitter sentiment analysis: A survey and a new dataset, the STS-Gold. In C Battaglino, C. Bosco, E. Cambria, R. Damiano, V. Patti, & P. Rosso (Eds.), *Proceedings of the First International Workshop on Emotion and Sentiment in Social and Expressive Media: Approaches and perspectives from AI* (pp. 9-21). http://ceur-ws.org/Vol-1096/proceedings.pdf

Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley Longman Publishing.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, *34*(1), 1-47. https://doi.org/10.1145/505282.505283

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*, 688. https://doi.org/10.1038/163688a0

Sun, Y., Wang, S., Li, Y., Feng, S., Chen, X., Zhang, H., Tian, X., Zhu, D., Tian, H., & Wu, H. (2019). ERNIE: Enhanced representation through knowledge integration. arXiv. https://arxiv.org/abs/1907.12412v2

Sun, Y., Wang, S., Li, Y., Feng, S., Tian, H., Wu, H., & Wang, H. (2019). ERNIE 2.0: A continual pre-training framework for language understanding. arXiv. http://arxiv.org/abs/1907.12412

Tseng, Y.-H., & Teahan, W. J. (2004). Verifying a chinese collection for text categorization. In K. Jarvelin, J. Allen, P. Bruza, & M. Sanderson (Eds.), *Proceedings of Sheffield SIGIR - Twenty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 556-557). ACM Press.

Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In P. Isabelle (Chair), *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 417-424). Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073153

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio,

H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*. Neural Information Processing Systems Foundation. https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data mining: Practical machine learning tools and techniques (3rd ed.). Morgan Kaufmann Publishers.

Yan, L., Zheng, Y., & Cao, J. (2018). Few-shot learning for short text classification. *Multimedia Tools and Applications*, *77*(22), 29799.29810. https://doi.org/10.1007/s11042-018-5772-4

Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. In F. Gey, M. Hearst, & R. Tong (Chairs), *SIGIR '99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 42-49). Association for Computing Machinery. https://doi.org/10.1145/312624.312647

Yang, Z., Dai, Z., Yang, Y., Carbonell, J. G., Salakhutdinov, R., & Le, Q. V. (2019). *XLNet: Generalized autoregressive pretraining for language understanding* [Paper presentation]. 33rd Conference on Neural Information Processing Systems, Vancouver, Canada. https://papers.nips.cc/paper/8812-xlnet-generalized-autoregressive-pretraining-for-language-understanding.pdf

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems* (Vol. 1, pp. 649-657). MIT Press.

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., & Liu, Q. (2019). ERNIE: Enhanced language representation with informative entities. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1441-1451). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1139

Yuen-Hsien Tseng ORCID 0000-0001-8904-7902

JoEMLS 註釋（Notes）暨參考文獻（References）羅馬化英譯說明

2015年1月31日修訂

1. 本刊針對部分國外西文專業資料庫之引文索引建檔與中文辨讀之需求，凡屬中文稿件之英文摘錄末，特別增列中文羅馬化拼音之「註釋」（或「參考文獻」）一式。
2. 作者（含團體作者）、機構名稱（出版者）、地名（出版地）：依事實與習慣為英譯，如無法查證時，中國大陸地區作者以漢語拼音處理，台灣以威妥瑪拼音（Wade-Giles system）處理。
3. 出版品、篇名：採用（登載於原刊名、篇名等之正式英譯）照錄原則；若原刊文無英譯，則由本刊依漢語拼音音譯著錄之。
   e.g. 南京大學學報 *Journal of Nanjing University*
   e.g. 中國科學引文數據庫 *Chinese Science Citation Database*
   e.g. 玉山國家公園解說志工工作滿足之研究 *Yushan National Park jieshuo zhigong gongzuo manzu zhi yanjiu*
   e.g. 教育資料與圖書館學 *Journal of Educational Media and Library Sciences*
4. 混用狀況：地名、機構、人名與其他事實描述，交錯共同構成篇名之一部分時，為避免冗長拼音難以辨讀，可將該名詞中之「地名、機構、人名」依事實與習慣英譯，其餘字詞則由本刊補以漢語拼音處理。
   e.g.「中國科學院與湯姆森科技資訊集團聯手推出中國科學引文索引」
   "Chinese Academy of Sciences yu Thomson Scientific Lianshou Tuichu *Chinese Science Citation Database*"
5. 本刊文章註釋（Notes）或參考文獻（References）羅馬化英譯規則，仍遵循Chicago（Turabian）或APA之精神及原則，進行必要且相對應之編排處理。**此羅馬化作業屬權宜措施，不可取代原有正式之引文規範。**
6. 羅馬化範例：
   範例1 －註釋（Notes）
   林信成、陳瑩潔、游忠諺，「Wiki協作系統應用於數位典藏之內容加值與知識匯集」，教育資料與圖書館學 43卷，3期（2006）：285-307。【Sinn-Cheng Lin, Ying-Chieh Chen, and Chung-Yen Yu, "Application of Wiki Collaboration System for Value Adding and Knowledge Aggregation in a Digital Archive Project," *Journal of Educational Media & Library Sciences 43*, no. 3 (2006): 285-307. (in Chinese)】
   範例2 －參考文獻（References）
   林雯瑤、邱炯友（2012）。教育資料與圖書館學四十年之書目計量分析。教育資料與圖書館學，*49*（3），297-314。【Lin, Wen-Yau Cathy, & Chiu, Jeong-Yeou (2012) A bibliometric study of the *Journal of Educational Media & Library Sciences,* 1970-2010. *Journal of Educational Media & Library Sciences*, *49*(3), 297-314. (in Chinese)】

About Romanized & Translated Notes/References for Original Text

The main purpose of Romanized and Translated Notes (or References) at the end of English Summary is to assist Western database indexers in identifying and indexing Chinese citations. This Romanization system for transliterating Chinese cannot be a substitute for those original notes or references listed with the Chinese manuscript. The effect of Chinese Romanization for citation remains to be seen.

Notes for Contributors

1. The *JoEMLS* is a fully peer-reviewed and Open Access quarterly sponsored and published by the Tamkang University Press, Taipei, Taiwan.
2. It is a condition of publication that all or part of manuscript submitted to the *JoEMLS* has not been published and will not be simultaneously submitted or published elsewhere.
3. The Editors welcome submissions of manuscripts mainly on topics related to library science, information science and technology, the book trade and publishing. The other library related fields such as instructional technology and information communication are also accepted.
4. Contributions are accepted on the strict understanding that the author is responsible for the accuracy of all contents of the published materials. Publication does not necessarily imply that these are the opinions of the Editorial Board or Editors, nor does the Board or Editors accept any liability for the accuracy of such comment, report and other technical and factual information.
5. The authors of any submissions to this *JoEMLS* hereby agree that if any submission being accepted by the Journal, then the *JoEMLS*, Tamkang University Library, and Department of Information & Library Science (DILS) shall be authorized to duplicate, publicly transmit by the Internet, and publish by any other means for the purpose of non-profit use such as study and education etc.
6. The authors of any submissions to the *JoEMLS* hereby agree that if any submission being accepted by the Journal, then the *JoEMLS* shall be authorized to grant a non-exclusive license to National Central Library for collecting such a submission into the Remote Electronic Access/Delivery System (READncl System), or grant other database providers sublicense to collect such a submission into their databases, and to duplicate, publicly transmit by the Internet, downloaded, and printed by authorized users of those providers. In addition, the format of submissions may be changed in order to meet the requirements of each database.
7. Manuscript requirements:
   (1) Submissions should go through the online system, however articles submitted as email attachments in one of the following preferred formats, Word or Rich Text Format, are acceptable.
   (2) Three types of contributions are considered for publication: full & regular research article in IMRAD format should be between 6,000 and 12,000 words in length, brief communication of approximately 3,000 words, and observation report which tends to be a review article of more than 5,000 words.
   (3) Letters to the Editor should not exceed 1,500 words in length and may be: comments or criticisms of articles recently published in the *JoEMLS*; and preliminary announcements of original work of importance warranting immediate publications.
   (4) Both Chinese (if available) and English titles should be provided.
   (5) All manuscripts should be accompanied by an abstract of 300 words approximately. Chinese abstract can be optional. Up to six keywords should be provided, and should not exceed 12 tables and figures.
   (6) A brief autobiographical note should be supplied including full name, post & title, affiliation, e-mail address, and full international contact details.
   (7) Referencing style (notes or references): Authors should follow one of the forms, the Chicago style (Turabian Manual) or the APA format.
8. For Book Review column, the *JoEMLS* is looking for book recommendations as well as individuals willing to review them, you may contact the editor.
9. It is the author's responsibility to obtain written permission to quote or reproduce material that has appeared in another publication. This includes both copyright and ownership rights, e.g. photographs, illustrations, and data.
10. First Author should be the equivalent of the Principal Author. The Principal Author must clearly specify who are the Corresponding Author and co-authors in proper sequence. Submission of manuscripts previously published in conference proceedings or revision based on thesis should be clearly indicated in the front page of manuscripts.
11. Revision should be returned to the editor within 4 months for further peer review process. Revision behind the period could be rejected or treated as a new manuscript by the Journal.
12. Each author will receive 1 free copy of the *JoEMLS*. However, authors can find online full-text of PDF format via Open Access mechanism on the websites of *JoEMLS* and *DOAJ*.
13. Submissions of manuscripts in either Chinese or English and editorial correspondence please use the Online Submission & Peer Review Service (ScholarOne- JoEMLS) at http://joemls.dils.tku.edu.tw/, https://mc.manuscriptcentral.com/joemls, or mail to the editor:
    Professor Jeong-Yeou Chiu, Department of Information and Library Science, Tamkang University, Taipei, Taiwan. Email: joyo@mail.tku.edu.tw

About English Summary

A brief English Summary is a supplement to Chinese article. Authors who contribute to the *JoEMLS* in Chinese language would need to supply English Summaries themselves. Such English Summary will carry a disclaimer: "This English Summary is provided by the author(s) or translated by the *JoEMLS* editors, and the author(s) have certified or verified that the translation faithfully represents the Chinese version of their own in the journal. It is for convenience of the English users and can be used for reference and citation."

訂閱資訊（Subscription）

Address changes, subscriptions and purchase of back issues, exchanges should be addressed to: Journal of Educational Media & Library Sciences, Department of Information and Library Science, Tamkang University.
Address: 151, Ying-chuan Rd., Tamsui, Taipei 25137, Taiwan
Tel.: +886 2 2621 5656 ext.2382
Fax: +886 2 2620 9931
E-mail: joemls@mail2.tku.edu.tw
A crossed cheque should be made payable to "**TAMKANG UNIVERSITY**".

一年新臺幣1,200元（台灣地區）
Annual subscription （payable in advance）US$80.00（outside Taiwan）
國外航空郵費另加(Additional charge for airmail outside Taiwan)
US$15.00 (per year) for America, Europe, Australia & Africa
US$8.00 (per year) for Japan, Korea, Thailand & the Philippines
US$6.00 (per year) for Hong Kong & Macao
訂閱本刊，請以匯款郵局(局號2441285，帳號0388761，戶名：教育資料與圖書館學)或劃線支票，戶名抬頭請填寫《教育資料與圖書館學》匯寄訂費，謝謝。

本刊網頁：**http://joemls.tku.edu.tw**

教育資料與圖書館學 JOURNAL OF EDUCATIONAL MEDIA & LIBRARY SCIENCES Vol. 57, No.1, 2020

Tamkang University Press
Taiwan

本刊獲
科技部人文社會科學研究中心
補助編輯費用

ISSN 1013-090X
9 771013 090005