THE RESEARCH AND APPLICATION OF ACADEMIC ETHIC AND COPYRIGHT OF THE REUSED FIGURE FOR THE PUBLICATION BASED ON NTLTD IN TAIWAN

Han Wang[1] , LiShu Ho[2]

[1]Han Wang, Institute of Information Science, Academia Sinica

E-mail:hollywang@iis.sinica.edu.tw

[2] PhD Student, Department and Graduate Institute of Library and Information Science National Taiwan University

E-mail: lishuho@gmail.com

Abstract: Getting more understanding about users' queries in images/figures retrieval on the platform of NTLTD in Taiwan will be helpful for setting a access point when request those images/figures as references. Especially, to develop an automatic detection model for examining if any inappropriate retrieval or cite of image would be establishing the recognitions of copyright statement. In order to achieve the goal of detecting the reused figure from the uploaded publication via NTLTD in Taiwan, the checking and inspecting mechanism is necessary. In this paper, we followed the steps mentioned above to develop the mechanism. Anaconda provides a sturdy develop environment with virtual environment manager and open-source data packages. We utilize Anaconda to build up a virtual environment for Tensorflow, an open-source software library for dataflow programming. Through the feature comparison between targeted figure and the labeled figure, the similarity could be easily computed. In this paper we followed the steps mentioned above to develop the mechanism.

Keywords: detection reused figure, figure citation, NTLTD in Taiwan, image-seeking behavior, academic integrity

Extended Abstract

1. INTRODUCTION

Thanks to years of Internet advancement, searching for the images through any web browser is not a difficulty anymore. It makes the published image more accessible and more convenient for the free rider. Image abusing is usually be neglected by student and readily causes crucially jural issue. At some points, it attributes to that there is no appropriate image search system on existing Taiwan's thesis digital library. Author has no general conception to cite or refer the image/figure from the other papers, so that causes recognition gap between the authors.

According to the latest news from National Central Library in Taiwan (NCL, Taiwan), it becomes official member of NDLTD on Oct. 25, 2010. Up today, NCL is not the only member in the national, but also the only NDLTD member in Asia Pacific (includes Taiwan, Japan, China,

Korea, Singapore, etc.). Moreover, the National Digital Library of Theses and Dissertations in Taiwan would base on the research and supporting platform to offer the global research resource to all the researcher in Taiwan. This platform not only can shorter the barrier between city and country, but also can promote and exposure our academic research and influence [1]. From the record shown by NCL, Taiwan, it is obviously found the growing number of most cited and downloaded by the top five institutions is increase gradually in the past half year. (See Figure 1 and Figure 2) [2]When users download the full text of articles from the NDLTD in Taiwan and cited it for their publication references, whether the imagines would be referring for their citation? Or if users knew they have to make footnotes if they do cite those imagines for referencing. There are some limitations to get the evidence to show how many imagine has ever downloaded or cited from the platform even if any reused figure from the uploaded publication.
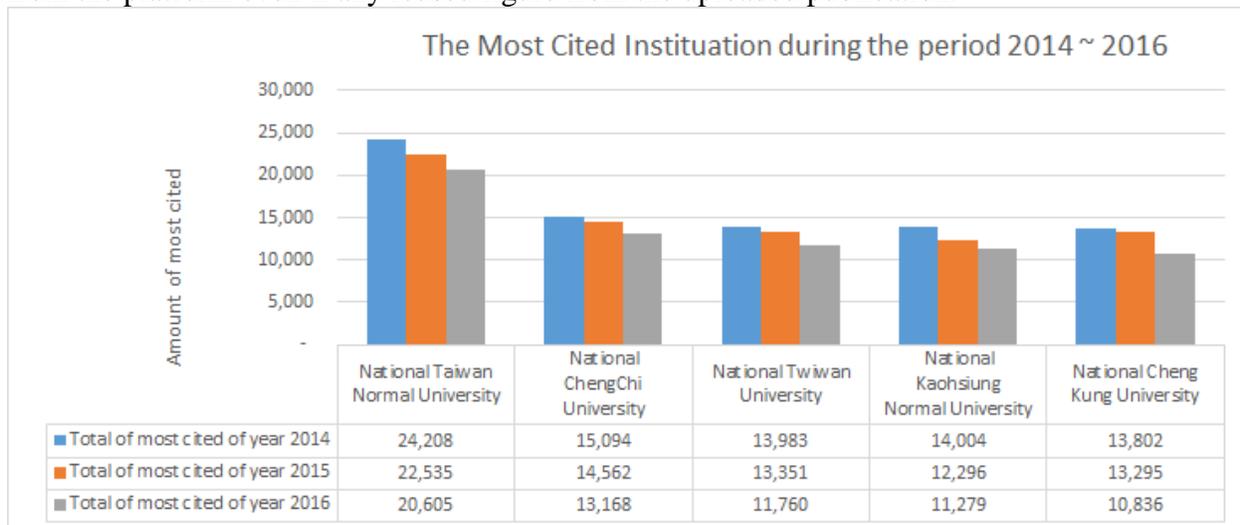
The Most Cited Instituation during the period 2014~2016

	National Taiwan Normal University	National ChengChi University	National Twiwan University	National Kaohsiung Normal University	National Cheng Kung University
■ Total of most cited of year 2014	24,208	15,094	13,983	14,004	13,802
■ Total of most cited of year 2015	22,535	14,562	13,351	12,296	13,295
■ Total of most cited of year 2016	20,605	13,168	11,760	11,279	10,836

Figure 1. The most cited institution during the period 2014 - 2016

The Most Downloaded by Institution

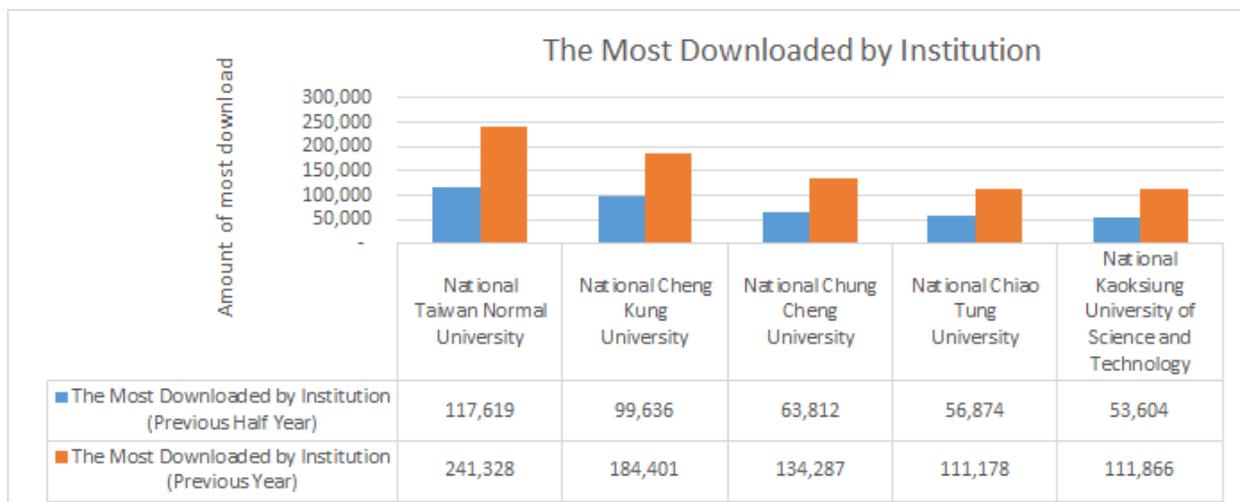|	National Taiwan Normal University	National Cheng Kung University	National Chung Cheng University	National Chiao Tung University	National Kaoksiung University of Science and Technology
■ The Most Downloaded by Institution (Previous Half Year)	117,619	99,636	63,812	56,874	53,604
■ The Most Downloaded by Institution (Previous Year)	241,328	184,401	134,287	111,178	111,866

Figure 2. The most downloaded by institution

Several previous studies have found there are more and more research on image-seeking behavior and visual information seeking behavior, for instance, Armitage and Enser [3]studied user

need in image archives around seven library; Frost [4] studies how users browse the online image database and observe their search patterns; Cunningham et. al[5]studies how user describe their image information needs when they looking for visual arts works; Yoon studies for what features of image retrieval system do users prefer as well as what kind of characteristics of users' image need for their tasks and goals [6]; Matusiak[7] studies what kind of research methods have been employed in studying information behavior of image users in the field of library and information science.

There are also picture researches create three primary attributes to perform the task for accessing images or visual information: collection knowledge, domain knowledge and would knowledge [8]. Recently, Acuna et al [9] point out there are some of scientific misconduct has been found that there are inappropriate reuse of figure or imagine on their science reports. In Acuna et al's [10]study show how they detect images by applying a copy-move detection algorithmic and a classification of possibility alike work as well as analyze potential illegal reuse of figures in the field of biological science literatures. Resnik [11] points out that the importance of ethical norms in each field of studies; for instance; to promote the purposes of research then avoid error and against falsifying. Those ethical norms can not only promote the values of fairness but also help researchers to have the great credit and contribution. The last but not the least is ethical norms construct public funding and promote the social responsibility.

When user cite the article from the platform of NDLTD in Taiwan, how well they conceptualize and how much they have acknowledgement related to copyright issue of using and citing images from others' works. As a result, in this paper we would design a prototype related to how to examine the inappropriate reuse imagine/figure on NDLTD in Taiwan.

2. METHODOLOGY

In order to strengthen the digital library of thesis and dissertation of National Center Library, we survey the foreign online systems and find several possibility. Image searching for digital library and detection of reused figure for thesis submit system are what NTLTD lacks. In this paper, these two issues are considered, and different approaches are proposed respectively.

2.1. Image Searching

Searching for image that inquired by text is widely used for search engines such as Google, Bing, Yahoo...etc. However, it is not common on thesis searching system. And, people usually has no basic concept of how to search an image through the digital library. Thus, getting deeper understanding of user behavior is our preliminary step. Questionnaires is usually a best way to study this issue. According to Goker, Ayse, and John Davies studied [12], they conducted seven classes of image use for studying users' image searching behavior. There are illustration, information processing, information dissemination, learning, generation of ideas, aesthetic value, and emotive/persuasive purposes. Each uses has its own function, such as illustration is used to convey some meaning, as primary important message, gaining knowledge, providing an inspiration, adding something more attractive, and being stimulate someone's imagination [13](Westman, 2009).

2.2 Detection of Reused Figure

In order to achieve the goal of detecting the reused figure from the uploaded publication via NTLTD in Taiwan, the checking and inspecting mechanism is necessary. General steps of image analysis include pre-processing, feature extraction, clustering or classification. Pre-processing aims to suppress the unwanted distortion and noise that could affect the accuracy.

Techniques of machine learning are always popular in the area of computer vision, and especially, deep neural network (DNN) is widely utilized for feature extraction from the image. For example, based on the DNN framework, Sun et al. [14](2014) proposed a novel image representation by using convolutional neural network (CNN) in the purpose of image retrieval and matching; Zhao et al. [15](2016) proposed the descriptor for image matching. Features are the representation of the image as well as the indicator. Through the feature comparison between targeted figure and the labeled figure, the similarity could be easily computed.

In this paper, we followed the steps mentioned above to develop the mechanism. Anaconda provides a sturdy develop environment with virtual environment manager and open-source data packages. We utilize Anaconda to build up a virtual environment for Tensorflow, an open-source software library for dataflow programming. This is also a popular library for deep learning algorithm. Further, OpenCV is exploited for image processing. OpenCV is a basic library for the state-of-art algorithm in the field of computer vision. Therefore, we develop a simple program for reused figure on thesis uploading system based on the help of these tools mentioned.

3. FINDING

Our aim of this paper is to develop the mechanism which help NTLTD in Taiwan to apply the function of imagine detection. The primary focus of our study is to analyze what the general or specific themes or domains requested by users. In their imagine retrieval task, how they conduct their workflow even how much if they knew about the importance of citing images is the same requirement with citing text. To develop an automatic detection model for examining if any inappropriate retrieval or cite of image would be establishing the recognitions of copyright statement.

4. CONCLUSIONS

In this paper, we propose both qualitative research and quantitative research on the academic ethic and copyright of the reused figure. Several works are surveyed to obtain hints about what to strengthen NTLTD from the existing digital library. With the aim of improving NTLTD, we first survey the user's behavior on image searching and figure citation by conducting questionnaires. The target volunteer is college student who is working on the project or thesis. Then, we analyze and discuss the result to lean the preference. On the other hand, a simple program of reused image/figure detection is developed to alert the users whether he/she abuses the image when uploading the thesis through the system. We hope this study can raise lots awareness of importance of choosing and citing images/figures from platform of NTLTD in Taiwan as well as maintain academic integrity.

REFERENCES

[1] National Central Library, "National Digital Library of Theses and Dissertation in Taiwan", https://etds.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/ccd=7zKbSk/webmge?switchlang=en (accessed August 2, 2018).

[2] National Central Library, "Latest News", https://etds.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi/ccd=7zKbSk/newsresult?search=1&qs0=id=%22108%22&qf0=_hist_#result (accessed August 2, 2018).

[3] Armitage, Linda H. "Analysis of User Need in Image Archives." *Journal of Information Science*23, no. 4 (1997): 287-99.

[4] Frost, C.O., Taylor, B., Noakes, A., Markel, S., Torres, D., & Drabenstott, K.M. (2000). Browse and Search Patterns in a Digital Image Database. *Information Retrieval, 1*, 287-313.

[5] Cunningham, Sally Jo, David Bainbridge and Masood Masoodian. "How people describe their image information needs: a grounded theory analysis of visual arts queries." *Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries, 2004.*(2004): 47-48.

[6] Yoon, JungWon. "Searching Images in Daily Life." *Library & Information Science Research*33, no. 4 (2011): 269-75.

[7] Matusiak, Krystyna. K. "Studying Information Behavior of Image Users: An Overview of Research Methodology in LIS Literature, 2004–2015." *Library and Information Science Research*1, no. 39 (2017): 53-60.

[8] Enser, Peter. "Visual Image Retrieval: Seeking the Alliance of Concept-based and Content-based Paradigms." *Journal of Information Science*26, no. 4(2000):199-210.

[9] Acuna, Daniel E., Paul S. Brookes, and Konrad P. Kording. "Bioscience-scale Automated Detection of Figure Element Reuse." *BioRxiv*, 269415. (2018). doi:https://doi.org/10.1101/269415.

[10] Ibid.

[11] Resnik, David B. "What Is Ethics in Research & Why Is It Important?" National Institute of Environmental Health Science. December 1, 2015. Accessed July 30, 2018. https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm.

[12] Ibid. 5

[13] Westman, Stina. (2009). Image Users' Needs and Searching Behaviour. In Information Retrieval (eds A. Göker and J. Davies) (2009). doi:10.1002/9780470033647.ch4. Cited in Conniss, Lynne R., A. Julie Ashford, and Margaret E. Graham. *Information Seeking Behaviour in Image Retrieval: VISOR I Final Report*. 2000. Library and Information Commission Research Report 95.

[14] Sun, S., Zhou, W., Li, H., & Tian, Q. (2014). Search by Detection: Object-Level Feature for Image Retrieval. *ICIMCS*.

[15] Zhao, Y., Li, Y., Shao, Z., & Lu, H. (2016). LSOD: Local Sparse Orthogonal Descriptor for Image Matching. *ACM Multimedia*.