Searching Digital Political Cartoons

Yejun Wu
School of Library and Information Science
Louisiana State University
Baton Rouge, LA 70803
wuyj@lsu.edu

Abstract—**The study of cartoons, manga, and graphic novels is of growing importance to humanity scholars. Managing cartoons for scholarly use presents two challenges: searching and understanding. Current cartoon archiving and searching projects either organize cartoons for browsing or manually record the symbols and the text in cartoons for searching. We propose to automatically detect the text lines in the cartoons, recognize the text using OCR, expand the scant text evidence in cartoons using related external text (such as news) to facilitate searching, process the text into character trigrams, and use the related external text as a context to facilitate the understanding of the cartoons. Preliminary experiment results are reported.**

Keywords-**cartoon, search, retrieval, document expansion**

I. INTRODUCTION

The study of cartoons, manga, and graphic novels is of growing importance to humanity scholars, including historians, sociologists, and media studies scholars as well as independent researchers. Tools facilitating searching and understanding these materials would be useful to humanity scholars and archivists. Managing cartoons for scholarly use presents two challenges: searching and understanding. Short titles and text embedded in cartoon images makes searching them a challenge. The symbols and words in cartoons often make an allusion to news events and politics, or popular culture at that time, which might make understanding cartoons a problem, especially for readers who are not aware of the related news events or readers from a different culture.

Current digital cartoon projects either organize cartoons for browsing or manually record symbols, captions and/or text in cartoon images for searching. Most cartoonists organize their cartoons for browsing on their Websites (such as Daryl Cagle's website www.cagle.com). New Yorker Cartoon Bank (cartoonbank.com) manually describes the symbols in images and record the text in images for searching and provides digital cartoons for sale. The Cartoon Research Library at Ohio State University[1] is the largest academic research facility in the United States devoted to cartoon art. They have digitized a portion of their cartoon collection and have manually recorded the captions of these cartoons for searching. We propose to develop techniques to facilitate better searching and understanding of digital

political cartoons. We elected political cartoons for the following three reasons.

(1) Automatically detecting text in cartoons can be a difficult problem because the text embedded in the cartoon images can be in any orientation, the text sometimes is presented as an art, and the text sometimes mingles with the image components in the cartoons.

(2) The scant text in cartoons can be difficult to recognize by computers; some cartoons do not include text at all, which calls for manual annotation of the content and meaning of the cartoons. Political cartoons may have sarcastic meaning, which is different from general scenic images.

(3) Political cartoons are used by historians, political scientists, cartoon researchers, and other humanity scholars for studying public opinions, news events, and comic art. Historians view political cartoons to have great significance. For instance, historians can now use the cartoons in the 1920's to truly capture the minds of the American people during the great depression. The cartoonists with their drawings could say more in one cartoon than could be said by a speech given by a politician. "As an American historian, I am fascinated by the way in which it is possible to follow an event through the cartoons which were produced as the event unfolded, often over years. With cartoons produced concurrently to the event, we have none of the hindsight which can cloud much historical interpretation - we see the situation as it was seen by the people of the time" [9].

Although we elected to use political cartoons for our experiments, the approaches and techniques we aim to develop can be applied to the searching and understanding of the broader graphic plus text combination materials, such as handwritten notes, graphic novels, figures, graphic slides, and maps. For example, the techniques can be used to build a digital cartoon searching system to supplement the Chronicling America's online digital newspaper system, which is sponsored by the National Endowment for the Humanities and the Library of Congress.

we are at the initial stage of the project, and so report only preliminary experiment results here. The paper is organized as follows: Section 2 introduces related work, Section 3 addresses experiment design, Section 4 presents our preliminary results, and Section 5 concludes the paper.

---

[1]http://cartoons.osu.edu

II. Related Work

The first step to automatically recognize text in cartoon images is zoning to find the text region in the images. University of Massachusetts created TextFinder, a system to automatically detect and extract printed text from images [10], [11].

ICDAR 2003 robust reading competition and ICDAR 2005 text locating competition have promoted the research on locating text in images captured with digital cameras [7], [8]. The aim of the text locating competition was to find the system that can most accurately identify the word rectangles in an image. The two leading systems for the 2005 competition were the Hinnerk Becker System and the Alex Chen System [3].

University of Maryland has been working on detecting text line in handwritten documents. Their method, which is script independent, achieves high accuracy for detecting text lines in both handwritten and machine printed documents [1], [4], [6].

Once the text lines in the images are detected, they can be passed through an OCR system for recognition. The most popular OCR software are ABBYY FineReader, Omnipage, Readiris, Presto OCR, PrimeOCR. OCRopus is an open-source OCR software sponsored by Google.

Searching is the process of matching queries and cartoon representations (which are text extracted from cartoons). In a typical document retrieval setting, documents are long and queries are short. Query expansion using relevance feedback has shown good improvement in retrieval effectiveness [2]. In cartoon searching, the text documents representing the cartoons are short, therefore we can use related external text (such as news) to expand the documents to improve query-document matching. OCR may introduce errors in text. Indexing terms with n-grams and n-gram query formulation can improve retrieval results [5].

III. Experiment Design

Unfortunately neither the Hinnerk Becker System nor the Allex Chen System is available now. TextFinder is not maintained either. Therefore we plan to acquire Doermann's text line detector to locate the text in cartoons, run open-source or commercial OCR software to recognize text, then expand the recognized text for searching using n-gram indexing. Before we can do any information retrieval experiment, we have to collect some cartoons and represent the cartoons for retrieval purpose.

A. Data Collection

A graduate assistant at our school collected 103 digital political cartoons and their related external text (i.e., news articles) by searching and browsing Chronicling America's online news database [2], which provides America's historic

[2]http://chroniclingamerica.loc.gov/

Figure 1. Liberty Bond
http://chroniclingamerica.loc.gov/lccn/sn83030214/1918-04-06/ed-1/seq-1/

newspaper pages from 1860 to 1922. The newspapers are in the public domain. Figure 1 shows an example cartoon with text lines on staves in almost any orientation. Figure 2 shows another example with artistic text inside the image. A typical digital cartoon in the collection has a caption (usually printed), text inside the image (very often artistic), symbols in the image, and sometimes a text footnote (usually printed).

The graduate assistant also annotated the cartoons with 21 topics. Here are some example topic statements:

- Q01: bolshevism
- Q04: capital and labor conflict, such as strike, and its effect on public welfare
- Q08: United States and Mexico relationship, and United States' role in Mexican revolution
- Q11: President Roosevelt memorial

A topic statement of a cartoon tells the topicality of the cartoon. We have two types of cartoon topics - descriptive and semantic, depending on the level of the abstraction of the description. A descriptive topic (such as Q01 and Q11) tells specific concepts, objects, and events presented in the cartoons, whereas a semantic topic (such as Q08) provides high-level semantic meaning of a cartoon. There may be
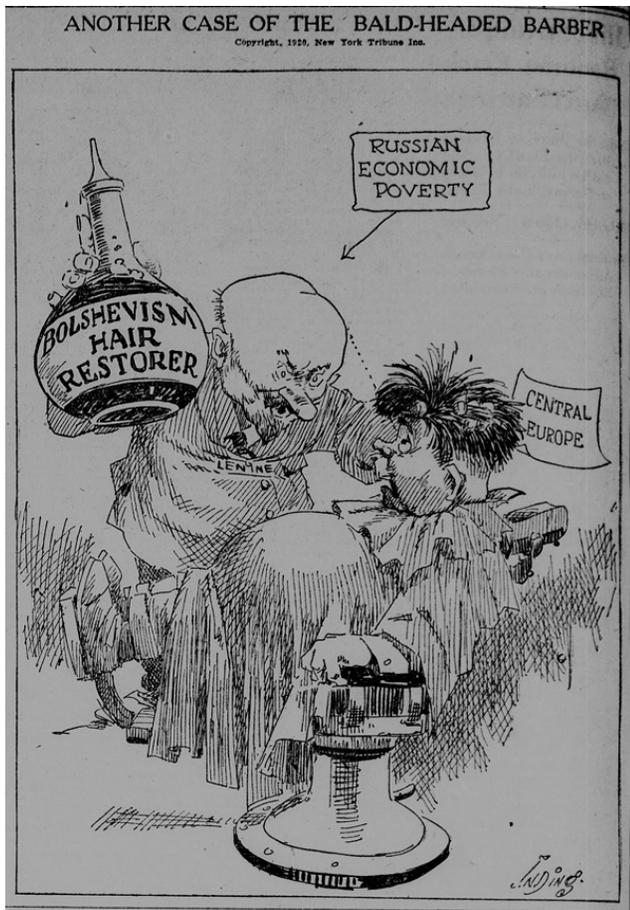
Figure 2.   Russian Economy Poverty
http://chroniclingamerica.loc.gov/lccn/sn83030214/1920-08-11/ed-1/seq-8/

mixed topics (such as Q04). Since the topics were developed along with the cartoon collection, the topic statements were very likely influenced by the text in the images.

B. Annotation Experiment

The descriptive information of a cartoon may include its caption, text inside the image, symbols inside the image, and a footnote. A political cartoon may express rich political, historical, and even sarcastic meaning, but its descriptive information is mostly scant. Users search information based on their information needs, which are expressed by queries. The queries must match the information about the cartoons in some way to retrieve the cartoons. The information about the cartoons may include descriptive information, topic/theme, and high level political and historical meaning, depending on how the cartoons are described and interpreted.

How users want to search cartoons remains a research question. User may want to search cartoons from their descriptive information, topics, or high level political and historical meaning. A starting point to this question, how-

ever, is to understand how people describe and understand cartoons.

Two graduate students at LSU, one from the Department of History (marked as Annotator1), and the other from the School of Library and Information Science with some history education in his/her undergraduate study (marked as Annotator2), were recruited to do a cartoon annotation experiment. They were instructed to read 30 digital political cartoons sampled from our small collection, and to record the topics/themes, captions, text inside the cartoons, footnotes, symbols, the meaning of the cartoons, and the most important elements contributing to their understanding (including whether historical knowledge is required). The annotation task took three hours.

C. Text Extraction and Recognition

We used Doermann's text line detection and text recognition system ([1], [4], [6]) in an effort to detect the text lines in a cartoon image (including the caption, text inside the image, and footnote), and to recognize the text.

D. Cartoon Retrieval Experiment

A cartoon uses images, artistic and printed text to deliver meanings. In our cartoon retrieval experiment, each cartoon is represented by its text surrogate. Four text surrogate collections were created:

- Caption: each cartoon is represented by its OCRed caption only. This collection has 99 documents since four cartoons have no captions.
- Description: each cartoon is represented by its caption, text inside the image, text footnote. The text description was manually recorded by the graduate assistant. This collection has 103 documents.
- Caption expanded by a news article: each cartoon is represented by its OCRed caption (if any) supplemented by a relevant OCRed news article. This collection has 103 documents.
- Description expanded by a news article: each cartoon is represented by its manually created description supplemented by a relevant OCRed news article. This collection has 103 documents.

Depending on the printing and digitization quality of the newspaper, the captions and news articles generated by OCR carry recognition errors. An unrecognized character is typically replaced by a question mark.

Out of the 21 topics developed by the graduate assistant, 17 well-defined topics were selected for this experiment. The four unused topics are too general and have only one cartoon document. Topic statements were used as queries.

The documents and topics were processed by removing typical English stopwords and punctuation marks except question marks (because question marks are placeholders of characters unrecognized by OCR). Due to the scant text in cartoons and OCR errors in captions and news articles, each

document and topic statement was processed into character trigrams. No stemming was performed on documents and topic statements because trigrams have already served the stemming purpose. The Indri search engine[3] was used to create the four information retrieval systems for the four collections.

IV. PRELIMINARY RESULTS

Here we report the preliminary results of the annotation experiment, the text detection and recognition experiment, and the cartoon retrieval experiment.

A. Annotation

One annotator annotated 21 cartoons and the other annotated 23, so here we analyze the 21 cartoons that were annotated by both annotators. The two annotators were able to record the captions, text inside image, and footnotes consistently, but interpreted the topics mostly differently. The topics of 2/3 of the cartoons were annotated differently between them; the remaining 1/3 were annotated with similar descriptions or some overlapped words. For instance, Figure 1 was annotated as "World War I" by Annotator1 and "comment on WWI finance" by Annotator2. Figure 2 was annotated as "communism/red scare" by Annotator1 and as "economic commentary/warning" by Annotator2.

There were also much disagreement in the identification and description of the symbols in the cartoons. Annotator1 consistently identified and recorded fewer symbols than Annotator2 did. Excluding the three cartoons whose symbols were not recorded by either annotator, 2/3 of the cartoons had few overlapped words in their descriptions. For instance, for Figure 2, the symbols were recorded by Annotator1 as "barber chair and costume, bottle of hair restorer" and by Annotator2 as "coat, torture table, giant bottle with the solution."

The two annotators reported the meaning of cartoons very differently. For 1/3 of the cartoons, the two annotators reported similar meaning but with different levels of understanding (descriptive vs. high level historical). For another 1/3, Annotator2 reported not understanding the cartoons well. For fewer than 1/3 of the remaining cartoons, they reported similar meaning with similar narratives.

B. Text Detection and Recognition

The text line detection and text recognition system was able to detect and recognize printed text in captions and footnotes very well, but could not detect the artistic text inside the images. For instance, the text output of Figure 1 is "TAKE A STAVE FROM YOUR BARREL AND GET IN LINE." The reasons why the system fails to detect the artistic text inside the images are that (1) the features of the text lines are too close to their image background (e.g., the text mingle with the image components, the text is presented

as an art), making the segmentation of the text lines difficult; and (2) the text embedded in the images can be in any orientation.

C. Cartoon Retrieval

	Capt	Desc	Capt+News	Desc+News
MeanRecall	0.512	0.852	0.952	0.980
MeanPrec	0.106	0.095	0.057	0.057
MAP	0.202	0.217	0.152	0.156
Topics-0-Docs	5	1	0	0

Note: MeanRecall: mean recall, MeanPrec: mean precision, MAP: mean average precision, Topics-0-Docs: numbers of topics without any relevant documents retrieved, Capt: Caption, Desc: Description, Capt+News: Caption expanded by news article, Desc+News: Description expanded by news article.

Table I
RETRIEVAL EFFECTIVENESS FOR FOUR COLLECTIONS.

Since the text inside an image is difficult to be recognized, and Chronicling America's online news database has provided OCRed captions and news, we decided to use their OCRed captions and news directly for our cartoon retrieval experiment.

Since the topics were developed along with the cartoons, it was known which cartoon was relevant to which topic. Since we have a small cartoon collection, the number of relevant documents for each topic is known. We use recall, precision, and mean average precision to measure the retrieval effectiveness. Recall is the fraction of the relevant documents which has been retrieved[2]. Mean recall is calculated by averaging each topic's recall across all the 17 topics. Precision is the fraction of the retrieved documents which is relevant. Mean precision is calculated by averaging each topic's precision across all the 17 topics. "Most standard among the Text Retrieval Conference community is Mean Average Precision (MAP), which provides a single-figure measure of quality across recall levels. Among evaluation measures, MAP has been shown to have especially good discrimination and stability"[4]. Table I shows mean recall, mean precision, MAP, and the number of topics without any relevant documents retrieved for each of the four collections.

Table I shows that, the Caption system retrieved no relevant documents for five topics, indicating caption only is not a sufficient text surrogate of a cartoon. The MAP of the Description system is numerically higher than that of the Caption system, because the Description collection was manually created and the text of each description is longer than (or at least as long as) its corresponding caption. Expanding a caption or a description by a news article improves recall, but dramatically hurts precision and MAP, indicating that the trigram representation of a news

---

[3]http://www.lemurproject.org/indri/

[4]http://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-ranked-retrieval-results-1.html

article has introduced more noise than signals. Therefore, the trigram representation of caption and description can be effective for retrieval, but the trigram representation of expanded caption and description may not be an effective approach.

V. Conclusion and Future Work

The scant text in a cartoon brings its retrieval challenging. To make the retrieval task even more challenging, people describes the symbols inside the cartoon images differently, interpret the topics/themes differently, understand the meaning differently or at a different level, and the current text line detection and text recognition system cannot detect and recognize the artistic text inside a cartoon image well.

Fortunately we can detect printed text in captions and footnotes well. A trigram representation of a caption or description can be effective for retrieval. Furthermore, the news articles surrounding a cartoon on newspaper may help understand the cartoon although how to use them for retrieval remains further exploration.

Since we are at the initial stage of the project, the conclusion made here is suggestive rather than affirmative. In the future, we plan to collect more cartoons, have real users to develop the topics, investigate an approach to consistently describing cartoons for retrieval purpose and a better approach to retrieving expanded captions and descriptions, and find a way to represent metaphorical and sarcastic meanings of cartoons.

References

[1] Agrawal, M. and Doermann D., 2010. Context-Aware and Content-Based Dynamic Voronoi Page Segmentation. http://lampsrv02.umiacs.umd.edu/pubs/abstract.php?id=546

[2] Baeza-Yates, B. and Ribeiro-Neto, B., 1999. *Modern Informtion Retrieval*. ACM Press.

[3] Chen, X. and Yuille, A. L., 2004. Detecting and reading text in natural scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004. 366-373.

[4] Li, Y., Zhang, Y., and Doermann, D., 2006. Detecting text line in handwritten documents. *International Conference on Pattern Recognition*, 2006. 1030-1033.

[5] Harding, S. M., Croft, W. B., and Weir, C., 1997. Probablistic retrieval of OCR degraded text using n-grams. In Peters, C. (ed.), *Lecture Notes in Computer Science*, 1324, 345-359.

[6] Li, Y., Zheng, Y., Doermann, D. and Jaeger, S., 2006. A new algorithm for detecting text line in handwritten documents. *10th International Workshop on Frontiers in Handwritting Recognition*, 2006. 35-40.

[7] Lucas, S. M. et al., 2005. ICDAR 2003 robust reading competitions: engries, resutls, and future directions. *International Journal of Document Analysis and Recognition*, 2005. p. 1-48

[8] Lucas, S. M., 2005. ICDAR 2005 text locating competition results. In *Proceedings of the 2005 Eighth International Conference on Document Analysis and Recognition* (ICDAR'05). 2005. 1-5.

[9] Walker, J. A., 2009. Dry humor: the American cartoonists' view of Prohibition. http://www.politicalcartoon.co.uk/html/history2.html (accessed on 3/20/2009)

[10] Wu, V., Manmatha, R., and Riseman, E. M., 1997. Finding text in images. In *Proceedings of 2nd ACM Conference on Digital Libraries*, 1997. 3-12.

[11] Wu, V., Manmatha, R., and Riseman, E. M., 1999. TextFinder: An automatic system to detect and recognize text in images. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (11). November 1999.

[12] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.