Indexing Historical, Political Cartoons for Retrieval†

Yejun Wu

School of Library and Information Science, Louisiana State University,
267 Coates Hall, LA 70803, U.S.A., <wuyj@lsu.edu>

Yejun Wu is an Assistant Professor in the School of Library and Information Science at Louisiana State University. He received the Doctor of Philosophy in Information Studies from the College of Information Studies, University of Maryland, College Park (2008). His research areas include knowledge organization, information retrieval systems, and information technologies.

ABSTRACT: Previous literature indicates that political cartoons are difficult to index because they have a subjective nature, and indexers may fail to understand the content of a cartoon or may interpret its content subjectively. This study aims to investigate the index-ability of historical, political cartoons and the variables that affect the indexing results. It proposes an indexing scheme for describing historical, political cartoons, and uses that indexing scheme to conduct indexing experiments. Through indexing experiments and statistical analysis, three variables, which affect the indexing results, are identified: indexers, indexing fields, and cartoons. There is a statistically significant difference in inter-indexer consistency on indexers, indexing fields, and cartoons. The paper argues that historical, political cartoons can be indexed if knowledgeable indexers are available, and the context of the cartoons is provided. It also proposes a mediated, collaborative indexing approach to indexing such materials.

1.0 Introduction

Many readers have seen cartoons in newspapers or on the Web, but few have searched them in a cartoon database. The cartoons need to be indexed before they can be searched in such a system. This paper addresses the issues in indexing historical, political cartoons, starting from the importance of such cartoons for humanities scholars.

1.1 Importance of cartoons for humanity scholars

The study of cartoons, graphic novels, and manga is of growing importance to humanities scholars, including political scientists, historians, sociologists, media studies scholars, and artists, as well as independent researchers. These scholars have been using cartoons for research and educational purposes. Cartoon art is of great intellectual significance for the humanities and represents complex political and social thought in many circumstances. Cartoons are considered a particularly attractive medium for teaching history because students are almost forced to think critically (Schnakenberg 2010). Editorial cartoons, also known as political cartoons, were used to measure Americans' perception of anti-Americanism (Long et al. 2009). They also were used to elicit opinions about political events (El Rafaie and Hörschelmann 2010). The Centre for the Study of Cartoons and Caricature at the University of Kent (England) has a collection of 110,000 political and social cartoons and strips. "From 1999 to 2003 almost 20,000 researchers registered to use its online cartoon catalog, and by the end of that period they were making 1400 research visits to the site every month .... The current total is closer to 2000 research visits a month" (Hiley 2006, 101). I choose historical, political cartoons as the study object because they are used by historians, political scientists, media researchers, teachers, and other humanities scholars for teaching and studying history, public opinions, news events, and comic arts (Kohut 1976; Hammett and Mather 2011; Langeveld 1981; Martinez-Fernandez 1998; Steinbrink and Bliss

284

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

1988). Historical, political cartoons are also harder to interpret and index than current political cartoons because knowledge of history and culture is needed for interpretation.

1.2. Challenges of indexing, searching, and understanding digital political cartoons

Managing cartoons for scholarly use presents two challenges: indexing and searching. People typically search cartoons by text surrogates. Searching is performed in an index. However, indexing cartoons using text is challenging. Short captions and scant text embedded in cartoon images may not effectively convey subject content. "Cartoon drawings, though they do often contain text, usually do not contain text that usefully describes their contents" (Bovey 1995, 115). Furthermore, "cartoons often represent a rather controversial perception of a historical event instead of an objective and impartial portrayal of the historical 'facts'" (Schnakenberg 2010, 32). Indexers may fail to reveal the content of a cartoon or may interpret its content subjectively.

Since searching is performed in an index, the difficulty of indexing naturally poses challenges to searching. Not knowing how different types of users want to retrieve historical, political cartoons from cartoon retrieval systems also presents challenges to searching. How users can translate their information needs to queries is restricted by the system's indexing capabilities and other search features. For example, if a user wants to find a historical cartoon with a Mexican man with a knife in it and is about the US-Mexico war, a cartoon retrieval system needs to allow the user to search for the subject of the cartoon, and the symbols and objects depicted in the cartoons. This asks for an appropriate indexing approach that responds to user's information needs. Understanding how to index historical, political cartoon is the starting point of building a cartoons retrieval system, which can be used to further solicit user's information needs.

Indexing political cartoons asks for understanding the cartoons. Interpreting cartoons can be challenging to not only indexers but also users. "Contrary to popular perceptions, political cartoons are complex and polysemous and require a particular form of literacy" because they "convey particular meanings within a specific social and cultural context" (El Rafaie and Hörschelmann 2010, 195). A particular difficulty with understanding cartoons "consists usually of the decoding of symbols, personifications and other stylistic devices frequently used by cartoonists" (Schnakenberg 2010, 32). 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 events or readers from a different culture.

Political cartoons are complex and polysemous, multimodal texts, and interpreting cartoons is a complex process that requires a whole range of different literacies. These include the ability to read visual grammar, a familiarity with the cartoon genre, broad knowledge of past and current events, a vast repertoire of cultural symbols, experience of thinking analytically about real-world events and circumstances, and the ability to convey particular meanings within a special social and cultural context (El Refaie 2009; El Rafaie and Hörschelmann 2010). "Cartoons are typically highly complex texts that are often hard to interpret; a particular difficulty consists usually of the decoding of symbols, personifications and other stylish devices frequently used by cartoonists" (Schnakenberg 2010, 32).

Carl (1968) polled adult Americans and found that only 15% of Americans fully understood the artist's intent of individual editorial cartoons, and 15% partially understood the artist's intent in an editorial cartoon. Therefore, 70% of Americans did not understand editorial cartoons. Junior/high school and undergraduate students also have difficulty interpreting such works. Bedient and Moore (1985) found that fifth, eighth, and eleventh graders were subject to misinterpreting political cartoons or not being able to interpret them at all and often had trouble identifying the actors in such works. DeSousa and Medhurst (1982) found that undergraduate communication students were surprisingly inaccurate in determining the correct meaning elements of political cartoons, perhaps due to visual illiteracy, cultural lag of the cartoon, and low socio-political awareness. Those studies indicate that the majority of the general public cannot interpret political cartoons correctly. Not understanding the meaning of the cartoons may cause an inability to identify the subjects of the cartoons correctly.

1.3. Status quo of intellectual access to political cartoons

Chapple-Sokel (1996) surveyed several cartoon museums and interviewed many librarians who were in charge of organizing political cartoons. She found that editorial cartoons were traditionally accessed via card or book catalogs, data sheets, or hunting through a bag of cartoons. Traditional print indexes have provided some limited access to editorial cartoons through an illustrator index, but subject access is rarely available. "Editorial cartoons have not customarily been indexed for inclusion in print or computer databases. The reasons for their omission seem to be due to little expressed interest in this information, and perhaps more importantly, the difficulty of indexing material as subjective as editorial cartoons .... Increased access to information about them may well create in-

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

285

creased interest" (Chapple-Sokel 1996, 29). As noted previously, the online cartoon catalog of the Centre for the Study of Cartoons and Caricature at the University of Kent (England) did attract many users.

Advances in information technology allow greater access to digital or digitized political cartoons. Cammarata (2008) provides a list of 19 editorial and political cartoon collections on the Web. Current digital cartoon projects either organize digital cartoons for browsing or manually catalog cartoons for searching. Many cartoonists organize their cartoons for browsing on their Websites (such as Daryl Cagle's website, www.cagle.com). The New Yorker Cartoon Bank (cartoonbank.com) provides a subject taxonomy for browsing, and manually describes the symbols in the cartoons, and records captions, embedded text in cartoons and keywords for searching. The Cartoon Research Library at Ohio State University (http://cartoons.osu.edu) 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 caption of each cartoon to facilitate searching. The University of Kent (England) converted their card catalog of 110,000 cartoons into an online database and developed a thesaurus (using the keywords recorded in the card catalog) to index the subjects of the cartoons.

Due to the challenges in the indexing of political cartoons, subject indexing has aroused wildly divergent opinions in the library community. Gomes et al. (2013) justified the need for prompting the discussions about political cartoons in the knowledge organization scope. The general purpose of this study is to gain insights into how to index historical, political cartoons using natural language. The specific purposes of this study are to develop an indexing scheme for describing historical, political cartoons, to investigate the variables that affect the indexing results, and then to propose an approach to indexing historical, political cartoons. Indexing here refers to describing cartoons using the indexing scheme, instead of supplying index terms. The indexing scheme is developed based on the literature review presented in Section 2.0. Following the literature review, I propose the indexing scheme and research questions in Section 3.0. Data collection and research method are presented in Section 4.0. Data analysis and findings are presented in Section 5.0. The last section concludes the paper with a discussion.

2.0 Related work

There is extensive literature on indexing visual materials, but very little literature on indexing cartoons. Here I aim to conduct a literature review that is sufficient enough to develop the indexing scheme proposed in Section 3.0.

2.1 Cartoon indexing

Indexing political cartoons is rarely studied in academic literature and has not been the focus of study in the field of library and information science (Landbeck 2008). Chapple-Sokel (1996) pointed out "the controversy as to whether editorial cartoons are appropriate subjects for indexing and inclusion in bibliographic databases" (21) because the subjective nature of cartoons makes them challenging to index.

Landbeck (2001) found that research subjects would universally use the words in a cartoon's caption, signs, or dialogue when searching a database of cartoon images. Landbeck (2008) assumed that the cataloging needs of political cartoons could be divided into three groups: bibliographic, descriptive, and subject. The needs of the bibliographic description can be categorized by the following fields: author (or cartoonist), date of publication, and media used. It is more important to include fields describing the cartoon and its subject (or subjects). Chapple-Sokel (1996) surveyed several indexers of cartoon collections, and her results indicated that no standard of indexing for editorial cartoons existed. The following cataloging fields, more or less, were used in the cartoon collections she surveyed:

- format (e.g., cartoon, caricature, comic strip),
- title,
- legend,
- cartoonist (or illustrator),
- medium,
- a copy of the cartoonist's signature,
- inscription,
- publication date,
- condition (physical description),
- source (e.g., newspaper, magazine, syndicate),
- abstract (verbal description of the graphics, delivering the message the creator intended),
- personalities or figures in the illustration (real or fictitious, depicted or alluded to),
- subject (a one-sentence visual description of the cartoon, or rarely available),
- subject headings.

Schnakenberg (2010) proposed the consideration of the following elements when studying a cartoon: the source of the cartoon; publication time; historical events; figures and objects in the cartoon; the meanings of the symbols; the purpose of the written comments; who is speaking to whom; all labels and caption; play on words (hidden or double meanings of the words); issue and speculated aboutness; the cartoonist's intention/aim/point of view and sentiment (expressed through satirical/critical genre, positive/negative presentation of the figures); and the reader's view on the cartoon's message.

286

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

Among the current cartoon projects, the University of Kent's metadata scheme, used in their cartoon database, is the most comprehensive. It includes as metadata fields: publication, date, artist, title or caption, format, series, collection, copyright holder, size, technique, embedded text, subjects, names of the real or fictional people caricatured in the drawing, notes (background information provided by artist, indexer or from publication media), implied text (any piece of text that is implied by the drawing but is not explicitly included—usually a quotation or a line from a well-known song or poem) (Hiley 2006).

When dealing with the conflict between the subjective meaning of the political cartoons and the art of indexing which is to objectively describe the aboutness of those images, Chapple-Sokel (1996) suggested the inclusion of multiple possible descriptions when assigning terms. For example, in a cartoon about Watergate, it is uncertain whether it was a grimace or a grin on Nixon's face. Therefore, an indexer might use both grin and grimace as index terms and indicate the ambiguity. Since cartoons are a kind of image, and image indexing has been studied extensively, the following section reviews image indexing.

2.2. Image indexing

There is a large body of literature addressing image indexing (Rasmuseen 1997; Chu 2001). Yet, an extensive literature review on image indexing is not the focus of this paper. Instead, I will focus on previous literature that is central to shaping the indexing scheme that is to be proposed in Section 3.0. The previous works that have had significant impact on this study are Panofsky's theory defining three levels of meaning in a work of art (Panofsky 1974), and Shatford's interpretation of Panofsky's theory when analyzing the subject of pictures (Shatford 1986; Shatford Layne 1994). The interpretation is also discussed in Svenonius (1994).

Panofsky's (1974, 28) first level of meaning in a work of art is the "pre-iconographical description," which is "primary or natural subject matter, subdivided into factual and expressional." It requires the "practical experience ... of everyday familiarity with objects and events" to understand this level of meaning (Panofsky 1974, 27). Shatford (1986, 43) interpreted this as the "generic description of the objects and actions represented in the picture." She interpreted factual meaning as the ofness of the picture, and expressional meaning as the aboutness of the picture, using the example of "Dorothea Lange's Migrant Mother ... a picture of a woman and children, and probably about strength, or suffering, or determination" (Shatford 1986, 43). At this level, ofness is a "generic description of objects and event;" aboutness is a "description of the mood of the picture" (Shatford 1986, 45).

Panofsky's (1974, 29) second level of meaning is "iconography," which is the "secondary or conventional subject matter, viz., the world of specific themes or concepts manifested in images, stories and allegories." It requires "knowledge of literary sources (familiarity with themes and concepts)" to understand this level of meaning (Panofsky 1974, 41). Shatford (1986) interpreted the objective meaning as ofness, and mythical, abstract, or symbolic meaning as aboutness. She gave the example of Sir Joshua Reynolds' portrait of Mrs. Siddons as the Tragic Muse. The picture is "of Mrs. Siddons (pre-iconographically of a woman) but about the Tragic Muse" (Shatford 1986, 44). At this level, ofness is "a specific, or proper, appellation of those objects and events;" aboutness is "an identification of mythical beings that have no unique and verifiable concrete reality, of symbolic meanings and abstract concepts that are communicated by images of the picture" (Shatford 1986, 45).

Panofsky's (1974, 30) final level of meaning is "iconology," which is the "intrinsic meaning of content." It requires "synthetic intuition (familiarity with the essential tendencies of the human mind)" (Panofsky 1974, 41). "Iconological interpretation is based on accurate pre-iconographic description and correct iconographical analysis of a picture .... This level of meaning cannot be indexed with any degree of consistency" (Shatford 1986, 45), and does not adhere to an ofness-aboutness distinction (Shatford 1986, Svenonius 1994).

In Panofsky's categories, pre-iconography is a description, iconography is analysis, and iconography is interpretation (Panofsky 1974; Shatford 1986). They can be roughly mapped onto Chen and Rasmussen's (1999) hard and soft indexing, the former a description of the image's visual features, the latter its meaning. Even with these theories, it's a challenge to provide appropriate levels of descriptive access to cartoons.

Rafferty and Hidderley (2005) proposed a "levels of meaning" indexing template for still images, which contains the following levels:

A1: biographical: information about the image as a document, such as title, date, artist,
A2: structural contents: significant objects and their physical relationship with the picture,
B1: overall content: overall classification of the image, type of image,
B2: object content: classification of each object in the image (defined in A2),
C1: interpretation of whole image: overall mood; words or phrases to summarize the image,
C2: interpretation of objects: mood of individual objects (when relevant).

The higher levels of meaning (C1, C2) are similar to Panofsky's iconological level, and the lower levels (A2, B1, B2)

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

287

are similar to Panofsky's first two levels or Shatford's of-ness and aboutness.

3.0 Indexing scheme

There is no standard metadata scheme for describing digital, political cartoons. Dublin Core or the Library of Congress's Metadata Object Description Schema could be used, but the Visual Resources Association's Core 4.0 (whose intent is to describe artwork) serves as the best metadata schema for the description of political cartoons (Landbeck 2008; VRA Core 2007). The VRA Core can handle both bibliographic data of cartoons (such as author or cartoonist, date of publication, publisher, source, title), and administrative data (such as rights, measurements). VRA Core has an inscription element, which refers to "all marks or written words added to the object at the time of production or in its subsequent history" (VRA Core 2007). It can be a cartoon's caption, word spoken by the actors in the cartoon, a label that identifies an object or person, or a background text or footnote (Landbeck 2008). The description element in VRA Core is a "free-text note about the image, including comments, description, or interpretation, that gives additional information not recorded in other categories" (VRA Core 2007). VRA Core also has a subject element.

Considering the indexing fields used by various cartoon collections and proposed by cartoon researchers (see Section 2.1), the interpretation difficulty of political cartoons (see Section 1.2), and the three levels of meaning in images (see Section 2.2), I propose the indexing scheme (shown in Table 1) in an effort to enhance the access to, and facilitate the understanding of, historical, political cartoons. The scheme was developed using a top-down classification and bottom-up clustering approach. Panofsky's three levels of meaning in images were used to set up a framework with three types of indexing fields: general description fields (i.e., generic scene/aboutness; conceptual description of objects & events), historical/sarcastic de-

scription fields (i.e., specific, historical subject; historical/sarcastic interpretation of objects & events), and historical/sarcastic meaning. The indexing fields used by various cartoon collections and proposed by cartoon researchers were selected to fill up the framework using a bottom-up approach. Since bibliographic elements (such as publication date, cartoonist, title/caption, footnote, words in cartoon) can be adopted from current cataloging practice (see Section 2.1), the scheme focuses on description and interpretation of historical, political cartoons.

In the indexing scheme, fields a1 and b1 (b11, b12, b13) are general description fields, which correspond to Panofsky's pre-iconographical level of meaning; fields a2 and b2 (b21, b22, b23) are historical/sarcastic description fields, which correspond to Panofsky's iconographical level of meaning; field c corresponds to Panofsky's iconological level of meaning. We can imagine that general description fields may be interesting to the students who want to study the cartoon art, whereas the other fields may be interesting to historians.

4.0 Research questions and method

Previous studies show that political cartoons are difficult to index because they have a subjective nature, and indexers may fail to understand the content of a cartoon or may interpret its content subjectively. This paper aims to investigate some deeper issues regarding indexing historical, political cartoons by using the indexing scheme (shown in Table 1) to answer two research questions:

RQ1: Can historical, political cartoons be indexed? Specifically, what variables may affect the indexing results? What fields can be indexed reasonably well?

RQ2: What is a feasible approach to indexing historical, political cartoons?

Aboutness/Subject/Topic/Theme:			
(a1) Generic Scene (without historical knowledge)			
(a2) Specific, Historical Subject			
Objects & Events (be as specific as possible):			
	Who/What Objects	Where	What Action
(b1) Conceptual Description	b11	b12	b13
(b2) Historical or Sarcastic			
Interpretation	b21	b22	b23
Historical/sarcastic meaning the cartoonist wanted to express:			

Table 1. Indexing scheme

288

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

RQ2 can be answered by applying the insights gained from the answers to RQ1. In order to answer RQ1, we collected some historical, political cartoons, and conducted a three-step experiment. First, in a pilot study, two participants were recruited to index the cartoons using a simplified indexing scheme, aiming to understand the qualifications of the indexers. Second, five participants were recruited to index the cartoons using the proposed indexing scheme. In both studies, the participants were asked to use natural language rather than indexing terms to describe the cartoons because using natural language is easier. Third, two participants were recruited to judge the similarity of the indexing results created by the five participants. Inter-indexer consistency is calculated to measure the agreement of indexing terms between or among indexers. The indexing results and similarity judgment results were analyzed using IBM SPSS Statistics to test the following three hypotheses:

H1: Inter-indexer consistency is not the same between indexing fields,
H2: Inter-indexer consistency is not the same between indexer pairs,
H3: Inter-indexer consistency is not the same between cartoons.

Some of the method details are introduced in the next section because they are better explained in the context of specific experiments.

5.0 Experiment design

This section addresses data collection, the pilot study, the formal study, the similarity judgment experiment, and some details about the method.

5.1 Data collection

A graduate assistant in the School of Library and Information Science accessed the Chronicling America's online newspaper database (http://chroniclingamerica.loc.gov) in Spring 2010, and manually extracted 103 digitized historical, political cartoons from the newspaper Webpages. There are certainly many more historical, political cartoons in the database, but the small collection of cartoons is enough for this study. The database provides America's historical newspapers from 1880 to 1922. Figure 1 shows an example cartoon image.

5.2 Pilot study and findings

In order to learn what qualifications indexers needed to perform the indexing task, a pilot study was conducted

Figure 1. Russian Economic Poverty

Source: http://chroniclingamerica.loc.gov/lccn/sn83030214/1920-08-11/ed-1/seq-8/

(Wu 2010). Two graduate students at our university, one from the Department of History (Indexer-1) and the other from the School of Library and Information Science with some general history education in her undergraduate study (Indexer-2), were recruited for the pilot study. They were instructed to read 30 cartoons sampled from our small collection of 103 cartoons, and then index as many cartoons as possible in three hours using a less-complicated scheme, which included the following fields: topics/themes, captions, text inside the cartoon, footnotes, symbols in the cartoon, meaning of the cartoon, and the most important elements contributing to the understanding of the cartoon (including whether historical knowledge is required). Three hours is an appropriate amount of time for a graduate student to complete the experiment without scheduling difficulty and feeling fatigued. The indexing task here was actually describing the cartoons using natural language, rather than assigning descriptors from a controlled vocabulary, therefore a formal training on professional indexing was not provided.

Indexer-1 annotated 21 cartoons and Indexer-2 annotated 23. I analyzed the 21 cartoons that were annotated by both indexers. Both indexers were able to record the cap-

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

289

tions, text inside image, and footnotes consistently, but interpreted the topics often 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 overlapping words. For instance, Figure 1 was annotated as "communism/red scare" by Indexer-1 and "economic commentary/warning" by Indexer-2. There was also a significant amount of disagreement in the identification and description of the symbols in the cartoons. Indexer-1 consistently identified and recorded fewer symbols than Indexer-2. Excluding the three cartoons whose symbols were not recorded by either indexer, 2/3 of the cartoons had few overlapping words in their descriptions of symbols. For instance, for Figure 1, the symbols were recorded by Indexer-1 as "barber chair and costume, bottle of hair restorer," and by Indexer-2 as "coat, torture table, giant bottle with the solution." The two indexers also reported the meanings of cartoons very differently. For 1/3 of the cartoons, the two indexers reported different levels of meanings (descriptive vs. high level historical). For another 1/3, Indexer-2 reported not understanding the cartoons well. For the other 1/3 of the remaining cartoons, they reported similar meaning with similar narratives.

From the pilot study, we see two variables (i.e., indexers and cartoons) that affect the indexing results. Some cartoons are less subjective and easier to understand than others, contributing to the similar indexing results between the two indexers. Indexers' historical knowledge, use of words, and indexing practice (such as recording fewer symbols) also affect the indexing results. The pilot study sheds light on the design of the formal indexing experiment. First, it is necessary to use the more complicated indexing scheme in order to force indexers to provide different levels of description and interpretation. Second, in order to make sure the indexers understand the meanings of the cartoons, I recruited participants with necessary historical knowledge and provided context of the cartoons (i.e., a URL to the newspaper page where a cartoon is extracted).

5.3 Indexing experiment and similarity judgment

Five graduate students, who have knowledge of the history of the USA from 1910 to 1922, were recruited from our university's School of Library and Information Science and Department of History to do the indexing experiment. The full indexing scheme (shown in Table 1) was used, and the context of the cartoons was provided. The five indexers (marked as Indexer-A, -B, -C, -D, -E) annotated 30, 13, 23, 20, and 30 cartoons in three hours, respectively. Table 2 and 3 show two indexing results created by two indexers.

Two extra graduate assistants in the School of Library and Information Science were recruited to judge the similarity of the five indexers' results, field by field. Since this is a very time-consuming task, they were instructed to compare each pair of indexers' results on the five fields only (i.e., a1, a2, b13, b23, c) because the results of other four fields (b11, b12, b21, b22) are short and easy to compare. The following five similarity categories were used:

(a)	Aboutness/Subject/Topic/Theme:
(a1)	Generic Scene (without historical knowledge): Bald barber attempting to sell hair restorer to man with hair.
(a2)	Specific, Historical Subject: Lenin is attempting to convince central

Europe to embrace bolshevism even as it is dooming the Russian economy

(b)   Objects & Events (Please be as specific as possible):

	Who/What Objects	Where	What Action
Conceptual Description	Bald Barber, patient, hair restorer	barbershop	A barber is attempting to sell a hair restoring tonic to a haired man
Historical or Sarcastic			
Interpretation	Lenin, Central Europe, Russian poverty, Bolshevism	barbershop	Lenin is attempting to convince Central Europe to embrace Bolshevism even as it has caused ruin to Russia.

(c)   Historical/sarcastic meaning the cartoonist wanted to express:

The cartoonist is trying to portray Lenin as someone who is trying to trick Central Europe to joining him. Lenin himself is bald and he is attempting to sell hair restorer in the form of Russian Bolshevism to a man who already has hair. The baldness represents poverty and the hair represents wealth. The author is trying to state that it is best for central Europe to stay the course and not jump to Bolshevism in order to avoid poverty.

Table 2. Indexer A's description of Figure 1.

290

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

(a)	Aboutness/Subject/Topic/Theme:	
(a1)	Generic Scene (without historical knowledge): Barber shop	
(a2)	Specific, Historical Subject: Central Europe does not want Bolshevism as their form of Government	
(b)	Objects & Events (Please be as specific as possible):	

	Who/What Objects	Where	What Action
Conceptual Description	Two men, barber's chair, hair treatment bottle.	In a barber's shop	A man wants to give hair treatment to another man.
Historical or Sarcastic			
Interpretation	Lenin, Bolshevism hair restorer, man sitting in chair	In a barber's shop	Lenin wants to force Central Europe to adopt Bolshevism in the form of hair treatment.

(c)   Historical/sarcastic meaning the cartoonist wanted to express:

The caption of the cartoon says, "Another case of the bald-headed barber." Lenin is forcing a man representing Central Europe to accept Bolshevism. The author is making fun of Lenin, because if Bolshevism is so good why does he not use it on himself to grow hair. The man in the chair is resisting because he does not want to be bald like Lenin.

Table 3. Indexer B's description of Figure 1.

– Very similar (VS): apparently similar
– Mostly Similar (MS): similar in most aspects
– S/D: between MS and MD, similar in some aspects, but different in other aspects
– Mostly Different (MD): different in most aspects
– Very different (VD): apparently different

They were able to record the similarity/difference for each of the five fields of the top 10 cartoons in a 5x5 table in six to ten hours. One judge spent more time than the other.

5.4 Data processing, analysis and findings

This section presents the procedure of processing and analyzing the experimental data and the findings. Inter-indexer consistency is calculated to measure the agreement (or variability) of indexing terms between indexers directly. Inter-judge agreement is calculated to measure the agreement between the two judges, and it also measures the similarity (or variability) of indexing results between indexers indirectly but holistically.

5.4.1 Index similarity judgment data analysis

For the ease of computing the inter-rater agreement, I consolidated the five categories into three categories: Similar (VS/MS), S/D, Different (MD/VD). There are 468 valid data points out of the 500 in total. Cohen's Kappa measures inter-rater agreement and examines the reliability of ratings (Howell 2002). The statistic is calcu-

lated using IBM SPSS Statistics: Kappa = 0.366, statistically significant at $p<0.01$, N=468. This is a fair amount of agreement (Landis & Koch 1977; TexaSoft 2008). Table 4 shows the agreement and disagreement between the two judges.

		Judge 1			
		D (Different)	S/D	S (Similar)	Total
Judge 2	D (Different)	295	3	60	358
	S/D	9	0	1	10
	S (Similar)	39	2	59	59
	Total	343	5	120	468

Table 4. Agreement and disagreement between the two judges

Table 4 shows that most inter-rater agreement cases (63%) occur on the D category, which indicates that the indexers mostly made different descriptions of the five fields. Much fewer agreement cases (13%) occur on the S category. There are 24% of cases the two judges could not agree with each other, resulting in the fair amount of consistency in similarity judgment.

5.4.2 Processing of indexing results

The most common measure of inter-indexer consistency is AB/(A+B), where A represents the terms assigned by indexer a, B represents the terms assigned by indexer b, AB represents the terms assigned by both indexers (Lancaster 1991). However, in our experiment, the indexers describe the cartoons using natural language rather than supplying

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

291

indexer terms. In order to calculate inter-indexer consistency, I wrote a Perl program to process the index results of the 30 cartoons (created by the five indexers) to word stems. Word stems are often used by information retrieval systems to build an index based on the "bag of terms" model (Baeza-Yates and Ribeiro-Neto 2013). The program transforms upper case letters to lower case; removes common stop words and several commons words used by the indexers to describe the cartoons (such as cartoon, cartoonist, author, represent, represents, represented); runs the Porter Stemmer (Porter 1997) to get word stems; then calculates inter-indexer consistency for each indexing field using the word stems. The inter-indexer consistency data does not follow normal distribution, and is skewed toward 0.

In order to examine whether the word stems can replace indexing terms for calculating inter-indexer consistency, the correlation between the inter-indexer consistency and each judge's similarity judgment is calculated using Spearman's rho (rank correlation), which is a nonparametric test (Dallal 2000; Howell 2002). Spearman rank correlation with one judge ranges from 0.236 to 0.563, where nine out of ten are statistically significant at p=0.05 (two-tailed test, N=43~50). The correlation with other judge ranges from 0.117 to 0.426, where six out of ten are statistically significant at p=0.05 (two-tailed test, N=42 ~50). This indicates that the word stem overlap of two index fields between two indexers are correlated to the judges' index similarity judgment, therefore the words stems can be used as approximate surrogates of indexing terms when calculating inter-indexer consistency, and inter-indexer consistency data can be used to replace the judges' similarity data to investigate the variables that may affect the indexing results.

In order to get a rough idea of the inter-indexer consistency data by fields, the range, average, and standard deviation of inter-indexer consistency is provided in Table 5. Table 5 shows that half of the fields have a wide range (0-

1) of inter-indexer consistency and the average consistency is low (ranging from 0.07 to 0.18). Not surprisingly, the general description fields (a1, b11, b12, b13) have higher average inter-indexer consistency than the historical/sarcastic description fields (a2, b21, b22, b23, c). This leads to the first hypothesis we want to test.

5.4.3 Hypotheses testing

In this section, we test three hypotheses. The indexing results of the top 10 cartoons are used because they are annotated by all the five indexers and are sufficient for the statistical tests. Since the inter-indexer consistency data does not follow normal distribution, the Friedman's rank test for several related samples (two-tailed, p=0.05), which is the nonparametric analogue of the one-way repeated-measures analysis of variance (Howell 2002), is used. For post-hoc tests, Wilcoxon's matched-pairs signed-ranks test on the 45 pairs of inter-indexer consistency (p=0.001) is performed (note that the significance level for post-hoc tests needs to be p/45=0.05/45=0.001).

(1) *Hypothesis H1: inter-indexer consistency is not the same between indexing fields.*

The Friedman's rank test for several related samples is performed on the inter-indexer consistency data for the 10 fields of the top 10 cartoons. The dependent variable is inter-indexer consistency while the independent variable is the indexing field. The test shows a statistically significant difference in the inter-indexer consistency on indexing fields (Chi-square=134.2, df=9, p<0.05 (p=0.001), N=100, two-tailed). A post-hoc Wilcoxon's matched-pairs signed-ranks test on the 45 pairs of inter-indexer consistency (at p=0.001) is performed to examine the difference between field pairs; the statistically significantly different cases are shown in Table 6.

Field	a1	a2	b11	b12	b13	b21	b22	b23	c	t
Range	0-1	0-0.3	0-0.71	0-1	0-1	0-0.47	0-1	0-0.63	0-0.23	0-0.4
Average	0.17	0.09	0.18	0.14	0.16	0.12	0.12	0.11	0.07	0.16
Standard deviation	0.21	0.09	0.14	0.31	0.15	0.11	0.25	0.12	0.06	0.07

Table 5. Inter-indexer consistency by fields for top 10 cartoons (N= 100), t=total or all fields.

a1	a2	b11	b12	b13	b21	b22	b23	c	t
a1>c	a2<b11	b11>a2		b13>a2	b21<b11	b22<b11	b23<b11	c<a1	t>a2
	a2<b13	b11>b21		b13>c	b21>c	b22<t	b23<t	c<b11	t>b21
	a2<t	b11>b22			b21<t			c<b13	t>b22
		b11>b23						c<b21	t>b23
		b11>c						c<t	t>c

Table 6. Inter-indexer consistency on *indexing* fields; statistically significantly different cases.

292

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

Table 6 shows that inter-indexer consistency on field b11 is statistically significantly higher than many other indexing fields (i.e., a2, b21, b22, b23, c). Inter-indexer consistency on field c is statistically significantly lower than many other fields (i.e., a1, b11, b13, b21, and t). Field t is a comprehensive field which includes all the other nine fields. Inter-indexer consistency on all the subjective fields (i.e., a2, b21, b22, b23, c) is statistically significantly lower than the comprehensive field t.

Inter-indexer consistency on a1 (generic scene) is higher than a2 (specific, historical subject) (p=0.002) even though not statistically significant at p=0.001, indicating that the more subjective field (a2) is more difficult to agree upon. Examining every indexer's work of the 10 cartoons on b12 (general description of location) and b22 (historical description of location) shows that most of the descriptions of the two fields are the same, indicating that the two fields can be merged.

(2) *Hypothesis H2: inter-indexer consistency is not the same between indexer pairs.*

The Friedman's rank test for several related samples is performed on the inter-indexer consistency data for the five fields (i.e., a1, a2, b13, b23, c) of the top 10 cartoons. The five fields were used in the similarity judgment experiment, and the data of the five fields is sufficient for this test. The dependent variable is inter-indexer consistency while the independent variable is the indexer-pair. The test shows statistically significant difference in the consistency of indexer pairs (Chi-square=27.4, df=9, p<0.05 (p=0.001), N=50, two-tailed). A post-hoc Wilcoxon's matched-pairs signed-ranks test on the 45 pairs of inter-indexer consistency (at p=0.001) is performed to examine the difference between inter-indexer consistency pairs. Inter-indexer C-D consistency is statistically significantly lower than inter-indexer A-B consistency. Inter-indexer C-E consistency is statistically significantly lower than inter-indexer A-B consistency; all other 43 comparisons are not statistically significantly different.

(3) *Hypothesis H3: inter-indexer consistency is not the same between cartoons.*

The Friedman's rank test for several related samples is performed on the inter-indexer consistency data for the 10 fields of the top 10 cartoons. Dependent variable is inter-indexer consistency while independent variable is cartoon number. The test shows statistically significant difference in the inter-indexer consistency on different cartoons (Chi-square=83.2, df=9, p<0.05 (p=0.001), N=100, two-tailed). A post-hoc Wilcoxon's matched-pairs signed-ranks test on the 45 pairs of inter-indexer consistency (at

p=0.001) on different cartoons shows that 17 pairs are statistically significantly different. For example, inter-indexer consistency on Cartoon no. 01 is statistically significantly higher than that on Cartoon nos. 03, 04, 08, 09, and 10. Table 1 and 2 show that the descriptions of Cartoon no. 01 (as shown in Figure 1) made by two indexers are similar.

Through these hypothesis tests, we see that there is a statistically significant difference in inter-indexer consistency on indexers, indexing fields, and cartoons. It might be interesting to know the interaction effects between indexers, cartoons, and indexing fields. However, a non-parametric two-way ANOVA test is currently not available in IBM SPSS Statistics Version 21. Finding a way to run this test using other software packages can be on the list of our future work.

6.0 Discussion and conclusion

Previous literature indicates that political cartoons are difficult to index because they have a subjective nature, and indexers may fail to understand the content of a cartoon without the knowledge of history, politics, culture, and visual literacy. This study aims to investigate the index-ability of historical, political cartoons, and the variables that affect the indexing results. It proposes an indexing scheme for describing historical, political cartoons, and used that scheme to conduct indexing experiments. It uncovers three basic variables that affect the indexing results: indexers, indexing fields, and cartoons. There is statistical significantly difference in inter-indexer consistency on indexers, indexing fields, and cartoons.

The pilot study indicates that indexers without the historical knowledge of the cartoons cannot perform indexing without the context of the cartoons. In the formal experiment, the indexers have the historical knowledge and are given the context of the cartoons, therefore they can perform indexing reasonably well. As indicated by previous literature, interpreting and indexing political cartoons can be subjective. Rafferty (2011) asked the question of whether the preferred meaning of an image is determined and controlled by the encoder or whether the meanings(s) might be ambiguous, fluid, and open to interpretation by the readers. She tends to argue that meanings of an image are historically contingent and constructed by human interpreters. Similarly, a cartoon can have two types of meaning—its cartoonist's intended meaning, and its reader's interpretation of its meaning. Both types of meaning can contribute to the index of the cartoon. However, misinterpretation of the cartoon should generally be avoided in the index. Qualified indexers of historical, political cartoons are expected to have sufficient historical and cultural knowledge of the persons and events de-

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

293

picted in the cartoon, have visual literacy of interpreting cartoons, and are given the context of the cartoon. Historical, political cartoons can be indexed if the right indexers are available and the context of the cartoons is provided.

The proposed indexing scheme includes four general description fields, five more subjective, historical/sarcastic description fields, and an all-inclusive comprehensive field. Inter-indexer consistency on some fields is statistically significantly different from the others. On average, the four general description fields have higher inter-indexer consistency than the five historical/sarcastic description fields. The higher inter-indexer consistency in the general description fields indicates that those fields are easier to index than the historical/sarcastic fields, therefore, the differentiation of general description fields from historical/sarcastic description fields may help to increase the index-ability of the cartoons.

These findings can be useful to find a feasible approach to indexing historical, political cartoons. The basic objective cataloging fields can be filled by a professional indexer in a cartoon library or museum. Indexing the subjects and other subjective fields can be open to multiple indexers in order to provide a higher recall in cartoon retrieval. However, hiring multiple indexers to index one cartoon can be expensive. Rafferty and Hidderley (2005) proposed the democratic indexing approach as an alternative, as it allows users to record their private indexes for retrieval. This approach is similar to collaborative tagging, which has economic advantages (Macgregor and McCulloch 2006). Collaborative indexing can be done by users, thereby "reducing the potential costs associated with the indexing performed by the intervention professionals" (Ménard and Smithglass 2012, 301). However, Rafferty's (2011) study suggests that leaving image annotation and indexing solely to users could lead to the loss of historically contingent information over time unless conscious efforts are made to preserve it. Furthermore, the findings of our pilot study imply that the indexing of historical, political cartoons should be limited to knowledgeable users who have the historical, political, and cultural knowledge of the persons and events depicted in the cartoons. This would require a validation of the user's qualification (such as membership of a historical, political association in an online information system), and intervention of a professional indexer. The professional indexer, and the less knowledgeable users, can index the general description fields, whereas the historical/sarcastic description fields can be open to the more knowledgeable users for indexing. Some cartoons are harder to interpret, and hence index, than others, so the professional indexer can invite more knowledgeable users to index certain cartoons. The indexer will then approve the user-generated

index, in order to preserve historically contingent information and prevent purposeful pollution of the index. This approach can be called mediated collaborative indexing. Once the user-generated index is collected and approved, the professional indexer can also add controlled vocabularies to improve retrieval effectiveness (Muddamalle 1998). For instance, controlling the various forms of a person's names (e.g., President Roosevelt, Teddy Roosevelt, Theodore Roosevelt) may increase recall.

Inspired by the professional indexer's need to choose cartoons, as well as the number of knowledgeable users for collaborative indexing, the next step of the research proposes to investigate the features of the cartoons that are difficult to index. It is also interesting to see whether collaborative indexing of political cartoons can converge. Furthermore, the purpose of indexing cartoons is to provide access to them. Investigating how different types of users want to retrieve historical, political cartoons from cartoon retrieval systems may also shed light on how we should index those cartoons. However, building such an index is only the starting point of understanding how users want to use it.

References

Baeza-Yates, Ricardo and Ribeiro-Neto, Berthier. 2013. *Modern information retrieval: The concepts and technology behind search*, 2nd ed. New York: ACM Press.

Bedient, Douglas and Moore, David M. 1985. Student interpretations of political cartoons. *Journal of visual/verbal languaging* 5n2: 29-35.

Bovey J. D. 1995. Building a thesaurus for a collection of cartoon drawings. *Journal of information science* 21: 115-22.

Cammarata, Paul. 2008. Editorial cartoons on the web: Picturing politics. *College research library news* 69: 536-9.

Carl, LeRoy M. 1968. Editorial cartoons fail to reach many readers. *Journalism quarterly* 45: 533-5.

Chapple-Sokol, Angie. 1996. Indexing editorial cartoons. *Special libraries* 87n1: 21-31.

Chen, Hsin-Liang and Rasmussen, Edie M. 1999. Intellectual access to images. *Library trends* 48n2: 293.

Chu, Heting. 2001. Research in image indexing and retrieval as reflected in the literature. *Journal of the American Society for Information Science and Technology* 52: 1011-8.

Dallal, Gerard E. 2000. *Nonparametric statistics*. Available http://www.jerrydallal.com/LHSP/npar.htm.

DeSousa, Michael A. and Medhurst, Martin J. 1982. The editorial cartoon as visual rhetoric: Rethinking Boss Tweed. *Journal of visual/verbal languaging* 2n2: 43-52.

El Refaie, Elisabeth, 2009. Multiliteracies: How readers interpret political cartoons. *Visual communication* 8: 181-205.

294

Knowl. Org. 40(2013)No.5
Y. Wu. Indexing Historical, Political Cartoons for Retrieval

El Rafaie, Elisabeth and Hörschelmann, Kathrin, 2010. Young people's readings of a political cartoon and the concept of multimodal literacy. *Discourse: Studies in the cultural politics of education* 31: 195-207.

Gomes, Thulio Pereira Dias, Guedes, Vânia Lisbôa da Silveira and Santos, Maria José Veloso da Costa, 2013. Temas e questões em análise documentária de charge (Topics and questions on subject analysis applied to political cartoons). In *II Congresso Brasileiro em Representação e Organização do Conhecimento, 27-29 May 2013 Rio de Janeiro, Brazil.* Available http://isko-brasil.org.br/wp-content/uploads/2013/05/cadernoderesumos_isko20131.pdf.

Hammett, Daniel and Mather, Charles. 2011. Beyond decoding: Political cartoons in the classroom. *Journal of geography in higher education* 35: 103-19.

Hiley, Nicholas. 2006. Indexing cartoons. *The indexer* 25: 100-4.

Howell, David C. 2002. *Statistical methods for psychology*, 5th Ed. Pacific Grove, CA: Duxbury.

Kohut, Sylvester. 1976. Political cartoons for teaching United States history. *Teaching political science* 3: 343.

Lancaster, F. W. 1991. *Indexing and abstracting in theory and practice.* Champaign, IL: University of Illinois.

Landbeck, Christopher Ryan. 2001. *The organization and categorization of political cartoons: An exploratory study* Master thesis. Tallahassee, FL: Florida State University. Available http://diginole.lib.fsu.edu/etd/3301/.

Landbeck, Christopher Ryan. 2008. Issues in subject analysis and description of political cartoons. In *19th Annual ASIS SIG/CR Classification Research Workshop.* Available http://journals.lib.washington.edu/index.php/acro/article/view/12854.

Landis, J. Richard and Koch, Gary G. 1977. The measurement of observer agreement for categorical data. *Biometric* 33: 159-74.

Langeveld, Willem. 1981. Political cartoons as a medium of political communication. *International journal of political education* 4: 343-71.

Long, Mark, Bunch, Rick L. and Lloyd, Robert E. 2009. *Social science quarterly* 90: 652-73.

Macgregor, George, and McCulloch, Emma. 2006. Collaborative tagging as a knowledge organisation and resource discovery tool. *Library review* 55: 291-300.

Martinez-Fernandez, Luis. 1998. The birth of the American empire as seen through political cartoons (1896-1905). *OAH magazine of history* 12: 48-54.

Ménard, Elaine and Smithglass, Margaret, 2012. Digital image description: A review of best practices in cultural institutions. *Library hi tech* 30: 291-309.

Muddamalle, Manikya Rao. 1998. Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics. *Journal of the American Society for Information Science* 49: 881-7.

Panofsky, Erwin. 1974. *Meaning in the visual arts: Papers in and on art history.* Woodstock, New York: The Overlook Press.

Porter, M. F. 1997. An algorithm for suffix stripping. In Jones, Karen Sparck Jones and Willet, Peter, eds., *Readings in information retrieval.* San Francisco, Calif.: Morgan Kaufman, pp. 313-6.

Rafferty, Pauline. 2011. Informative tagging of images: The importance of modality in interpretation. *Knowledge organization* 38: 283-98.

Rafferty, Pauline and Hidderley, Rob. 2005. *Indexing multimedia and creative works: The problems of meaning and interpretation.* Burlington, VT: Ashgate Publishing Company.

Rasmussen, Edie M. 1997. Indexing images. *Review of information science & technology* 32: 169-96.

Schnakenberg, Ulrich. 2010. Developing multiperspectivity through cartoon analysis: Strategies for analyzing different views of three watersheds in modern Germany history. *Teaching history* 139: 32-9.

Shatford, Sara. 1986. Analyzing the subject of a picture: a theoretical approach. *Cataloging & classification quarterly* 6n3: 39-62.

Shatford Layne, Sara. 1994. Some issues in the indexing of images. *Journal of the American Society for Information Science* 45: 583-8.

Steinbrink, John E. and Bliss, Donna, 1988. Using political cartoons to teach thinking skills. *Social studies* 79: 217-220.

Svenonius, Elaine. 1994. Access to nonbook materials: The limits of subject indexing for visual and aural languages. *Journal of American Society for Information Science* 45: 600-6.

TexaSoft. 2008. *Interrater reliability (Kappa) using SPSS.* Available http://www.stattutorials.com/SPSS/TUTORIAL-SPSS-Interrater-Reliability-Kappa.htm.

VRA Core. 2007. *VRA Core 4.0 Element Description.* Available http://www.loc.gov/standards/vracore/VRA_Core 4_Element_Description.pdf.

Wu, Yejun, 2010. Searching digital political cartoons. In *Proceedings of the 2010 IEEE International Conference on Granular Computing, August 14-16, 2010, San Jose, California, USA*, pp. 541-5.