Lei Zhang
Margaret E. I. Kipp
Hur-Li Lee
Information Organization Research Group
School of Information Studies, University of Wisconsin-Milwaukee

Genre Tag …Tag Genre

Abstract

This paper examines genre-related tags through a social bookmarking system CiteULike. Three types of genre-related tags were identified: "is" tag indicating a genre, "for" tag for creating another genre, and "about" tag dealing with a genre topic. An analysis of the frequency and functions of genre-related tags will enable us better understand tagging as well as genre.

Introduction

The concept of genre has a long existence in human history, having been used as a means of both producing and structuring knowledge. While categorization of literary genres in the West can be traced back to Plato (Frow, 2006), the ancient Chinese began applying genre as a textual organizing scheme before the time of Confucius (Zhang & Lee, 2012). Modern knowledge organization systems (KOS's) from library cataloging codes to metadata schemas to taxonomies applied in a variety of websites also include genre as an important element. However, the concept of genre has not been rigorously defined or studied in KO and its role in information systems and standards have so far been discussed or empirically examined only sporadically (Andersen, 2008; Crowston & Kwasnik, 2003; Zhang & Lee, 2013). Nevertheless, KO has in the past few years seen a steady increase in genre-related research, mostly resulting from users' demands for improved access to rapidly growing digital information collections and resources.

This study will investigate genre in social tagging from two prominent theoretical stances. First, modern genre theory developed in rhetoric, communication and other disciplines informs the study with an amended conceptual frame to perceive genre as not only textual categories but also purposeful social action (Miller, 1984). Second, the authors firmly believe that genre studies must incorporate the perspective of information system users as genre originates and functions in social environments. We thus attempt to gain a preliminary understanding of the intended purposes of tags associated with genre as assigned by taggers on a social bookmarking website, who are in fact both metadata creators and users.

Tags are descriptive labels users assigned to pictures, videos, texts and other online resources for sharing, discovering and recovering resources. Tags usually represent categories or classes to which resources belong. Key phenomena and questions addressed

in tagging include motivations of tagging, tagging service design and interface, relationships among taggers, tags, and tagged resources, and evaluation of tagging results (Furner, 2010). A review of the literature regarding tag types shows how tags are used: what (or who) it is about, identifying what it is, identifying who owns it, refining categories, identifying qualities or characteristics, self reference, and task organizing (Golder & Huberman, 2006). The most frequently used tags are identified from a qualitative study of Connotea, Delicious, and CiteULike: topic (61%), task (12%), format (10%), refining (7%), source (5%), reference (3%), and qualities (2%) (Park, 2010). Non-subject tags such as time, task, and emotion-related tags are commonly used in social bookmarking and frequently found with subject related tags (Kipp, 2008).

Genre is dealt with as a tag type, termed "what it is" or "format", with examples of blog, journal, article, tutorial, review, survey, lyrics, opensource, mp3, etc. (Golder & Huberman, 2006; Park, 2010), showing that genre has been addressed vaguely as an equivalent to form or format. However, genre—a triplet of content, form, and function—is an important complement to topic largely because of its functions (Crowston & Kwasnik, 2003; Kwasnik & Crowston, 2004). A genre has functions since a document usually suggests what it is supposed to be used for, and what users can expect to get from it or to do with it. Therefore user assigned tags are expected to reveal how users perceive genre functions. Our study intends to explore users' perceptions of genre by asking the following research questions:

  a. What tags are most commonly used to indicate genre?
  b. What are the functions of genre tags that might be intended by the taggers?

Methodology

To answer the above research questions, we selected CiteULike (http://www.citeulike.org/), a social bookmarking system designed for scholars to save and share academic papers, as a test site. First, we downloaded the dataset provided by CiteULike for September 25, 2012. This dataset contained all tags used in CiteULike associated with specific articles by their article ID (number of articles = 4,048,232; number of tags = 770,282). Second, we identified 144 potential genre tags with frequency over 500. Using this list of genre tags, we downloaded the full dataset again on January 25, 2013 (number of articles = 4,871,962; number of tags = 792,681). Then all entries associated with 144 genre tags were selected (number of articles = 390,626), from which 400 were randomly sampled for further analysis.

Results

What tags are most commonly used to indicate genre?

Table 1 below presents a list of the 144 identified genre tags, with similar ones grouped together, and each tag followed by its frequency number. These genre tags may consist of full words, alternatives, or abbreviations (i.e., bibliography, bibliographic, biblio, bib), both singular and plural forms (i.e., review, reviews), and different languages (i.e., bibliography, bibliografia). Genre tags in the list may include a genre or its subgenre (i.e., bibliografia,

bibliografia_trasplantes), a single genre term or a combination of genre and other terms (i.e., ewd-bib, wise-bib-09-08-29, robotica-bib, wei-xing-bib-import-10-08-14, cj-bib, biblio-extracted-import, bm_pitman_bib, rcb-bibfile, ma_biblio_mehdi_juin07c, file-import-10-10-14-jkbibl, combinedbib). Terms beginning with an asterisk or other punctuation characters (e.g., *all-articles) are designed to sort before all other terms, and indicate article, endnote, bibliography, and paper are of priority to taggers. Among the 144 tags related to genre, the most frequently occuring tag is "review" (61,990), with a frequency more than double that of the next three, "methods" (27,633), "meeting-abstract" (27,388), and "article" (22,482).

Table 1. A preliminary list of genre tags

Genre Tag	Tag Label & Frequency
abstract	meeting_abstract,27388; gazette_abstract,5662; abstract,1887
ad	ad,1713
article	article,22482; paper,5441; papers,3894; staffpaper,3586; all-articles,3380; articles,2775; *papers-import-11-10-09,683; import-from-papers,602; paper1,599; *all-articles,559
bibliography	ewd-bib,6310; reference,4942; bib,2054; wise-bib-09-08-29,2051; robotica-bib,1992; wei-xing-bib-import-10-08-14,1967; references,1961; bibliography,1459; citation,1243; cj-bib,1118; biblio-extracted-import,1104; bibliografia,1030; bm_pitman_bib,883; rcb-bibfile,856; misra-references,855; citations,777; ma_biblio_mehdi_juin07c,720; bibliographic,635; ref,611; bibliografia_trasplantes,607; *file-import-10-10-14-jkbibl,568; combinedbib,529; biblio,500
bio	bio,1559; biography,695
blog	blog,2241; blogs,2032; weblogs,685; weblog,513
book	book,11057; dlbook,4722; books,3883; textbook,1706; ebooks,1034; boek,722
chapter	chapter,594
collection	collection,1204
conversation	dialogue,873; conversation 509
critique	crit,1012; critique,532; commentary,784
cv	cv,2336
database	database,12991; databases,6703; db,1368
drama	drama,523
editorial	editorial,1119
endnote	endnote-20091118ris,7650; endnote,2508; endnote_2005-01-05,1250; *import-endnote-lib-18122010,763; endnote-import-10-02-17,729; *gully-endnote-file-original-10-29-10,643; endnote-merged-home-trl-2009-01-19,528
essay	composition,3536; prose,1096; essay,529
exercise	exercise,5085
fiction/nonfiction	fiction,1345; nonfiction,727; novel,601; non-fictie,530
form	form,1229
guide	guidelines,2877; instruction,1599; guide,1065; guidance,717; guideline,612
humor	humor,518

index	index,4021; indices,511
introduction	introduction,1668; intro,758
journal	journal,3566; periodicals,818; journals,739
list	thelist,2231; inventory,1294; list,1089
literature	literature,3532; literatura,868
mail	email,1532; correspondence,952; project—email,689
manual	manual,1209
methods	methods,27633; method,12222
music	music,6114
narrative	narrative,2119
news	news,1771
play	play,896
poetry	poetry,1592
poster	poster,871
presentation	presentation,2586
program	program,2749; programs,1117; schedule,720
proposal	proposal,1349
protocol	protocol,2097; protocols,1412
recommendation	recommendation,2006
regulation	regulation,12702; rules,1170; rule,816
report	report,2370; case-reports,536
results	results,3920
review	review,61990; reviews,2211; disslit,1012; bookrev,887
speech	speech,5648
standard	standards,3552; standard,1674; c-standard_the_new,1566
survey	survey,8247; questionnaires,2006; surveys,1744; questionnaire,1096
thesis	thesis,16010; dissertation,6898; diplomarbeit,2022; *thesis,675
tutorial	tutorial,2645
website	website,1123
wiki	wiki,3303; wikipedia,2984; wikis,607
workflow	workflow,2451

It should be noted that these genre tags were identified out of context. Some terms may have more than one meaning, and user tags are uncontrolled vocabularies and thus personal and potentially inconsistent. The sampled entries describing the documents should provide us with additional context for how these genre tags are actually used.

What are the usages of genre tags that might be intended by the taggers?
All tags from the 400 sampled articles were examined carefully, and their abstracts and the available full-texts were browsed for better understanding of these tags. It shows that a total number of 197 tags were used to indicate a genre ("is" tag), 62 tags to indicate the creation of another work of a different genre ("for" tag), and 250 tags indicating an article about

genre topic ("about" tag). The following figures illustrate the distribution of three types of genre-related tags—"is", "for", and "about", respectively.

Twenty tags as indicators of genre are shown in Figure 1 below. Tag "review" is the most frequently used, followed by "article", "abstract", and "book". For example, the "article" tag includes terms in the form of "article" (22), "articles" (4), "all-articles" (4), "paper" (1), "papers" (8), "paper1" (1), and "staffpaper" (3); the "abstract" tag includes terms "abstract" (2), "gazette_abstract" (6), and "meeting_abstract" (15); and the "book" tag includes terms "book" (10), "books" (5), "boek" (2), "dlbook" (3), and "ebook" (1). Genre addressed by these tags may be general (e.g., article) or specific (e.g., review, introduction, critique), whole (e.g., book, article) or fragmentary (e.g., chapter, abstract).

Figure 1. Distribution of "is" tags
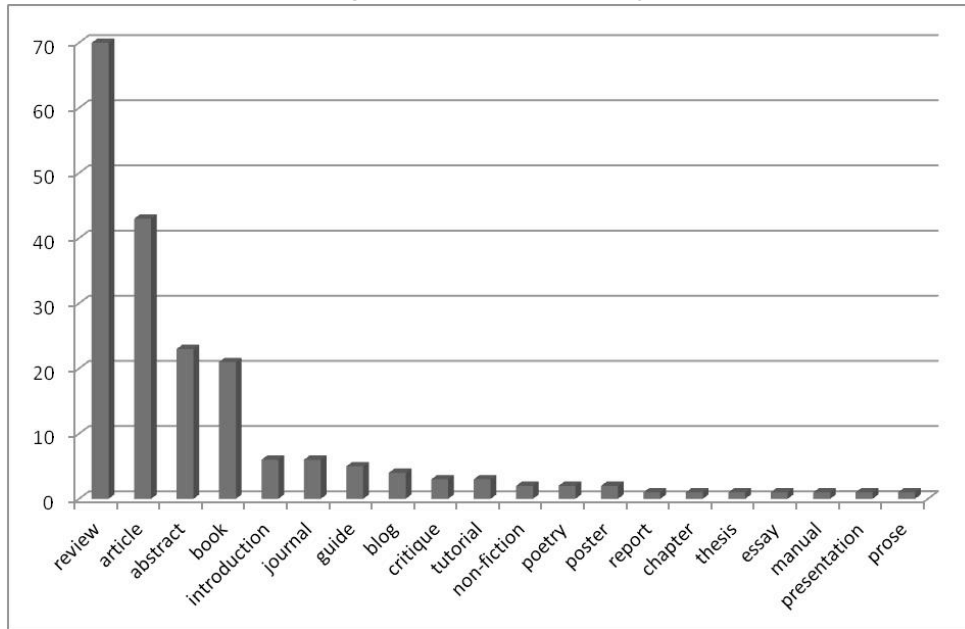
More interestingly, some genre tags are not intended to indicate the form of a work itself, but for potential use of the work as expressed in the form of a genre. For example, a "thesis" or "bibliography" tag assigned to a work that is not a thesis or a bibliography may be an indication that the tagger intends to use the work for writing a thesis or creating a bibliography. That was corroborated through a casual talk with an anonymous tagger, who said that she used the "thesis" tag to manage citations for her dissertation. Among nine tags in this category (Figure 2), the tag "thesis" includes the terms "thesis" (17) and

"dissertation" (4); and the tag "bibliography" includes the terms in the form of "bib" (2), "bibliografia" (3), "cj-bib" (2), "ewd-bib" (1), "wise-bib-09-08-29" (3), "reference" (3), "references" (1), and "misra-references" (1).

Figure 2. Distribution of "for" tags
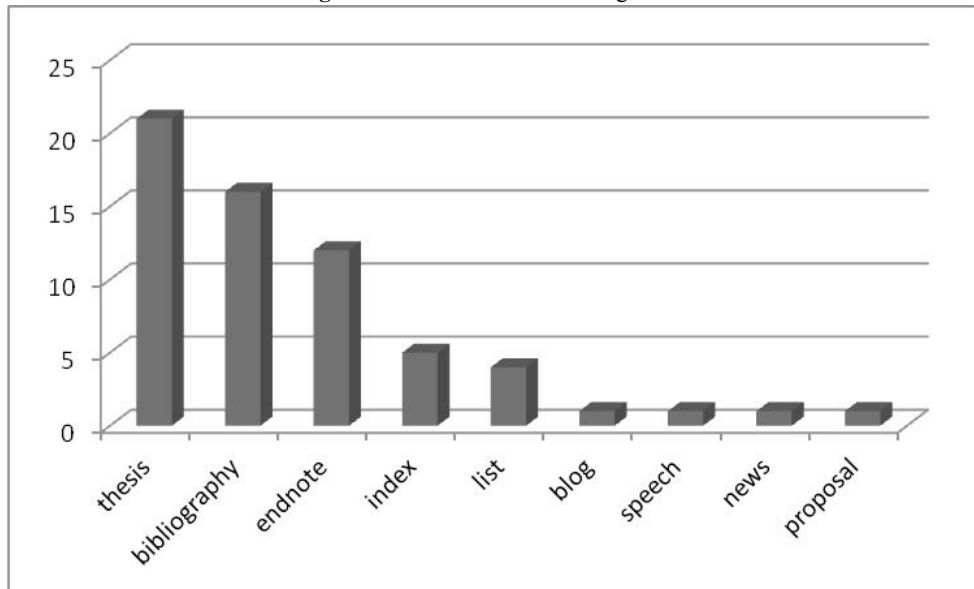
There is an association between a document genre and the genres it can be used for, as shown in Figure 3, illustrating the co-occurrence of "is" tags and "for" tags. For example, an examination of the sampled articles shows that when a document was tagged as a "review", it was also tagged with "thesis" (twice) or "bibliography" (three times) to indicate that the creation of a "thesis" or "bibliography" are the potential uses to which the review article will be put.

Figure 3. Co-occurrence of "is" tags and "for" tags
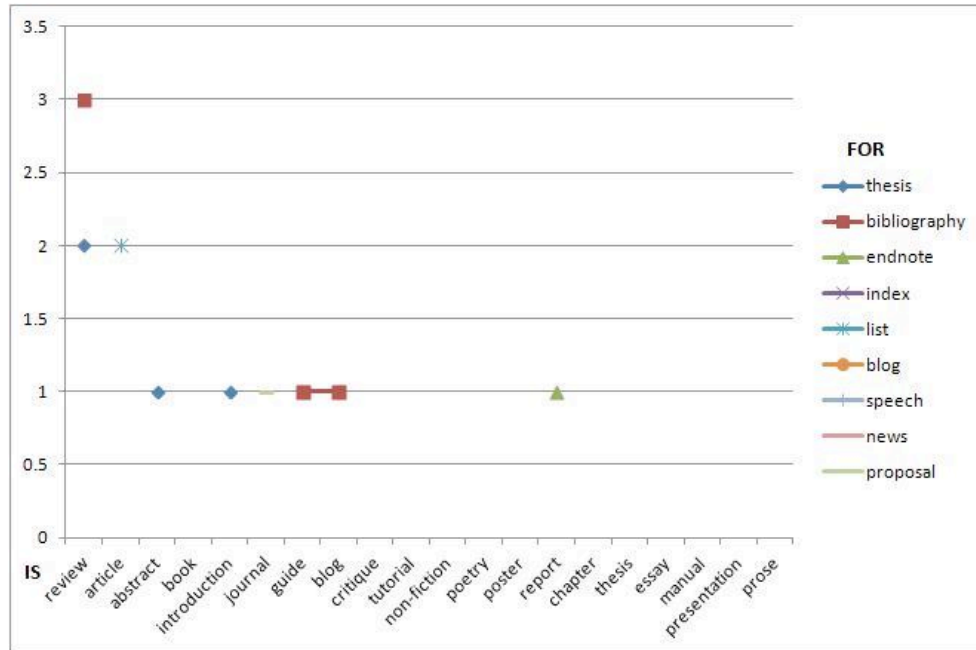
Quite a number of tags initially identified as genres actually refer to topics or approaches rather than genres. For example, the term "cv" which normally stands for curriculum vitae refers to cardiovascular; "survey" can stand for a research instrument or a method or an overview addressed in an article; "methods" and "results" can stand for two sections of an article; and some typical genre terms like "journal", "book", "article", etc. are addressed as topics here. Figure 4 shows a variety of genre-related topics covered in the CiteULike documents in the sample.

Some genre tags appear with at least one other genre tag; this happens to 34% of "is" tags, and to 35% of "for" tags, though the other tag is not necessarily of the same type. Some articles also were tagged with genre-related tags with a frequency less than 500, e.g., "books general", "english-abstract", "handbook", "lecture4", "my papers", "proceedings", "research-paper", "superlist", etc. Among other types of tags for these genre-tagged articles, most are topic tags. However, a few task tags were observed, such as "to check refs", "todo", "toread", etc.

Figure 4. Distribution of "about" tags
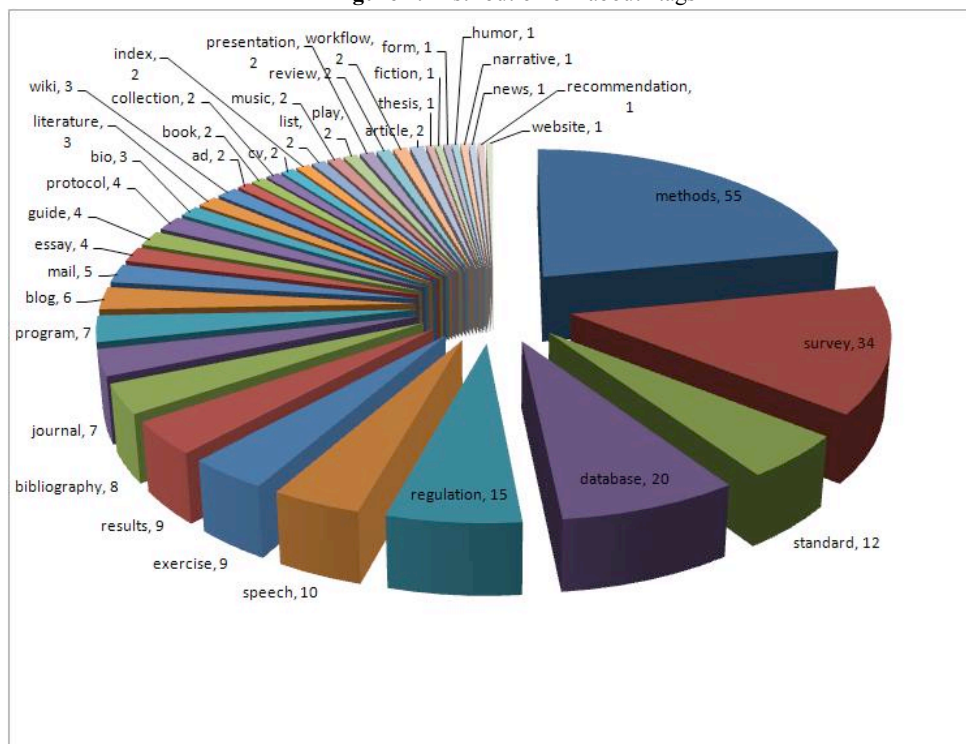
Discussion and Conclusion

The genre tags identified in this study are characterized by the user community of CiteULike—mostly scholars—and the typical activities in the site—depositing and sharing scholarly articles for academic purposes. These tags depict what kinds of documents the taggers are reading (through "is" tags), what kinds of future documents they intend to create with the information obtained from the currently tagged documents (through "for" tags"), and what document topics are of interest to them (through "about" tags). The taggers use multi-media for scholarly communication, illustrated by digital genres (e.g., blog) and spoken genres (e.g., presentation). The three types of genre-related tags—"is", "for", and "about"—suggest the tripartite nature of genre. The same document can be labeled by the "is" tag (e.g., review) to indicate its form and by the "for" tag (e.g., bibliography) to suggest what functions it might have, providing evidence that form and function are integrated and intertwined constituents of a genre. We also observe that a document tagged by a particular "is" tag (e.g., review) will likely also be tagged by a

particular "for" tag (e.g., bibliography). Therefore the consistent co-occurrences of the "is" and "for" tags would suggest what sets of genres are routinely used and how multiple genres jointly mediate activities in a community like CiteULike. This finding has significant implications for research on genre repertoires and genre ecologies in a web 2.0 environment.

This study will help us understand tagging as well as genre. Unsurprisingly, "about" tags constitute a large proportion of the genre-related tags. The occurrence of task tags is not high among sampled articles probably because genre tags can more or less perform the role of task tags, such as to read or to keep as a reference, and in turn for creation of another work of a particular genre. Splitting genre tags into the "is" and "for" categories illustrates the functions of genre tags: either used to indicate what the document in question is or for task purposes by indicating what the document is for. It may reshape the conception of tags termed as "what it is" or "format". Tags have a popular usage because they provide a richer description far beyond subject and incorporate context and task associated with document use, which coincides with the nature of genre. Therefore, an examination of genre-related tags will help to uncover the usages of tags and the intentions of taggers.

Understanding how the general public uses genre concepts in labeling and categorizing online resources is a significant step toward improving application of genre in the broad area of knowledge organization. The study provides preliminary data about genre usages by a self-selected group of people. Such findings will assist the project researchers in designing a rigorous, large-scale, and user-focused study of genre in information organization and retrieval. Furthermore it is expected that the distribution of genre tags will vary with social bookmarking systems with a different focus from CiteULike because genre is essentially context- and task-oriented. Future studies must extend to tagging practices in other types of settings and user groups.

References

Andersen, J. (2008). The concept of genre in information studies. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (Vol. 42, pp. 339-367). Medford, NJ: Information Today.

Crowston, K., & Kwasnik, B.H. (2003). Can document-genre metadata improve information access to large digital collections? *Library Trends, 52*(2), 345-361.

Frow, J. (2006). *Genre*. London: Routledge.

Furner, J. (2010). Folksonomies. In M. Bates & M.N. Maack (Eds.) *Encyclopedia of library and information sciences*, 3rd ed. (pp. 1858-1866). Boca Raton: Taylor & Francis.

Golder, S., & Humberman, B. (2006). Usage patterns of collaborative tagging system. *Journal of Information Science*, *32*(2), 198-208.

Kipp, M.E.I. (2008). @toread and cool: Subjective, affective and associative factors in tagging. *Proceedings of the 36th conference of the Canadian Association for Information Science*, Vancouver, BC, June 5-7, 2008.

Kwasnik, B.H., & Crowston, K. (2004). A framework for creating a faceted classification for genres: Addressing issues of multidimensionality. *Proceedings of the 37th Hawaii International Conference on System Science*, Big Island, Hawaii, January 5-9, 2004.

Miller, C.R. (1984). Genre as social action. *Quarterly Journal of Speech, 70*(2), 151-167.

Park, H.-J. (2010). *Understanding folksonomy as a web classification.* Unpublished doctoral dissertation. University of British Columbia.

Zhang, L., & Lee, H.-L. (2012). Rethinking genre in knowledge organization through a functional unit taxonomy. In R.P. Smiraglia & H.-L. Lee (Eds.), *Cultural frames of knowledge* (pp. 135-149). Würzburg, Germany: Ergon-Verlag.

Zhang, L., & Lee, H.-L. (2013). The role of genre in the bibliographic universe. *Advances in Classification Research Online, 23*(1), 38-45. doi:10.7152/acro.v23i1.14236