Twitter Content Categorisation: A Public Library Perspective

Ali Shiri* and Dinesh Rathi†

School of Library and Information Studies
University of Alberta, Edmonton, AB, Canada

*ashiri@ualberta.ca
†drathi@ualberta.ca

Published 27 December 2013

Abstract. With the rise of social media, many library and information services have begun to incorporate a wide variety of social media and social networking applications into their systems and services. Among the mainstream social networking applications, micro-blogging, in general, and Twitter, in particular, have gained increasing popularity. This paper reports the results of an exploratory study of the application of Twitter in the context of a large public library system. Specifically, this study has sampled, content analysed and categorised a select number of tweets created by a public library system in order to identify and document the ways in which Twitter can be used for various information services and knowledge management practices in public libraries. One of the main outcomes of this study is a tweet categorisation scheme that has a specific focus on the information services offered by public libraries.

Keywords: Social media; Twitter; Web 2.0; public libraries; tweet classification; knowledge management; tweet categorisation.

1. Introduction

“Social media refers to a set of online tools that supports social interaction between users” and “the term is often used to contrast with more traditional media such as television and books” (Hansen et al., 2011, p. 12). The social media tools and technologies which are also interchangeably used with the term Web 2.0 include such tools as blogs, wikis, microblogging (also known as tweets), RSS, podcast, social tagging and bookmarking and YouTube. Social media is currently used by many different organisations for multiple reasons (e.g. knowledge management, collaboration, dissemination, sharing, marketing, networking, etc.) (Aharony, 2010; Tripathi and Kumar, 2010; Redden, 2010; Kaplan and Haenlein, 2010). Microblogging is a newer form of blogging and unlike blog postings, it usually limits the postings and conversation within 140 characters (Aharony, 2010). Microblogging is more widely associated with Twitter and is one of the popular social media tools used not only by individuals but also by organisation in different domains such as airlines, fast food businesses, public and academic libraries for multiple reasons, namely conversion, reporting, sharing information, etc. (Aharony, 2010). Studying the flow of knowledge in Twitter is particularly interesting as it provides a new perspective on how social media facilitate knowledge creation and use. More specifically, in order for that knowledge to be accessible and useful, research should examine the ways in which knowledge inherent in tweets can be organised, represented and managed.

There is emerging interest among researchers in exploring the use of social media tools including Twitter in organisations, as their use and adoption is growing at tremendous pace, and it would not be hyperbole to claim that it is exploding as demonstrated by the following example facts. According to the Telegraph newspaper [published online on March 21, 2013], there are over 500 million users of Twitter; it took “3 years, 2 months and 1 day” for Twitter to have its first tweet to the billionth tweet; and in June 2012, there were 400 million tweets and it hit 500 million mark in October 2012 [http://www.telegraph.co.uk/technolog/twitter/9945505/Twitter-in-numbers.html]. These facts point out the need to regularly undertake research into social media as the membership and use are increasing with time and there is an evident need to continuously evaluate social media usage in order to contribute to the growing research work in these areas. Thus, this study explores the use and implementation of
Twitter in a large public library and extends upon the work of Aharony (2010), Cahill (2011) and Dann (2010) to develop a tweet categorisation scheme.

2. Literature Overview

Researchers study the emergence and the use of social media tools and technologies in different organisational contexts. A recent review of the literature of Web 2.0 technologies in libraries showed that between 2007 and 2011 more than 200 articles have been published in 13 leading LIS journals published by Emerald (Singh and Gill, 2013). In the following paragraphs, a review recent research on social media is provided. McLean (2008) explored the implementation of social media tools in multiple public libraries (e.g. Princeton Public Library; Darien Library; Public Library of Charlotte & Mecklenburg County, etc.) in the USA. The author found that the public libraries were using different tools such as wikis, allowing users to write reviews, podcast, create blogs, and make use of social networking sites like MySpace, etc. to provide virtual services. The paper concluded that “Virtual services are becoming a core part of services that public libraries offer” [...] and [...] “Web 2.0 tools provide a range of options for opening up our websites to collaboration with our users, enabling them to communicate with us in ways that were never possible before” (p. 446).

Jansen et al. (2009) investigated the use of microblogging i.e. Twitter in electronic word-of-mouth (eWOM) publicity and branding as well as the overall structure, expressions and sentiments of the tweets posted. The authors suggested that Twitter is a viable medium for marketing campaigns, branding efforts and customer feedback and relationship. Also, Twitter was found to be a good communication channel that would affect brand awareness and image, and could be used to generate brand awareness at low cost. The research in social media is not limited to the evaluation and understanding the use and impact of social media tools to meet their goals such as brand building and awareness, creating community and increasing traffic onto their organisational websites. (Dankowski, 2013).

Social media are considered to be low cost tools and libraries of different size and type are adopting and adapting social media tools to meet their goals such as brand building and awareness, creating community and increasing traffic onto their organisational websites. (Dankowski, 2013).
conducted focusing on its [Twitter] use in libraries” (p. 336). The research explored the extent of Twitter use in public and academic libraries, the linguistic and content differences between the tweets created by the two types of libraries. One of the key outcomes of the research was the development of a set of categories (e.g. library events, reference questions, etc.). Cahill (2011) studied the use of social media by the Vancouver Public Library (VPL), one of the top three large public library systems in Canada. She noted that VPL has significant presence on social media sites and services, including sites such as Delicious, Facebook and Twitter. The library has made effective use of social media to “engage patrons, develop community and maximise resources in a time of constrained budgets” (p. 259). She also highlighted the ways in which VPL (a) adopted and implemented different social media tools in conjunction with their organisational website, (b) the learning experience VPL gained from the social media tools implementation (e.g. issues with Delicious), and (c) the success VPL got through their effort especially by the use of Twitter. For instance, the author cited a few success stories that VPL experienced through Twitter interaction such as Word-of-mouth marketing for a first aboriginal storyteller in Residence and customer service feedback.

Chu and Du (2013) noted that social networking tools such as Twitter are used for “marketing and publicity, enhancing reference services and knowledge sharing among staff” (p. 67). Further the authors suggest that social networking tools have the potential to be used for such purposes as communication, enhanced interactions, engagement and feedback seeking. Also, Cahill (2011) noted that “the VPL Twitter account has become a core mechanism for communication and relationship building with patrons” (p. 268). These exemplify instances of the use of Twitter in knowledge management activities.

Building on previous research and in particular the work of Aharony (2010), Cahill (2011) and Dann (2010), our study reported in this paper aimed to explore the use of Twitter by the Edmonton Public Library (EPL), one of the large public library systems in Canada. In particular, this study examined the nature and analysed the content of the tweets created by the Library in order to gain a better understanding of how public libraries can benefit from the widely used micro-blogging service, namely Twitter, to extend and enhance their information and communication services.

3. Methodology

Classification of tweets is an interesting line of research as it provides insight into the ways in which Twitter is used by different people, organisations and communities. In this study, we focused on the tweets that were generated in the context of a large public library system. Previous studies have proposed a number of classification and categorisation methods for tweet analysis. Sriram et al. (2010) propose a classification of incoming tweets into categories such as News (N), Events (E), Opinions (O) and Deals (D) based on the author information and features within the tweets. Aharony (2010) compared the tweets created by academic and public libraries and presented a categorisation of tweets for both types of libraries. The author created six major categories (within public libraries) and five major categories (within academic libraries), new subcategories were generated according to the tweets’ contents. The tweets were initially categorised into formal and informal language. The public library tweets in this study were divided into six categories: library, information about, miscellaneous, general information, general recommendations and technology. Academic libraries’ tweet content was divided into five: library, information about, miscellaneous, technology and general information. These general categories had sub-categories. For instance, in the “library in general” category, there were specific categories for public library tweets, namely library events, book recommendations, the library collection, library services, references and the library, in general. Dann (2010) presented a useful review of Twitter content classification studies and puts forward six broad categories, namely conversational, status, pass along, news, phatic and spam. He then provided specific instances and examples for each category status which was divided into temporal, physical, mechanical, work, location or automated.

In our study, the EPL was chosen as a case study since the library has successfully adopted a wide range of crowdsourcing technologies, including the use of Twitter and has effectively incorporated user-generated content into its search system as well as its library catalogue (Shiri and Rathi, 2012).

We downloaded around 2500 tweets from Twitter with the hashtag “epldotca” created between February 2009 and May 2012. We then created a plain-text format of the downloaded tweets. In order to create a categorisation scheme, we adopted a two-stage process using the grounded theory approach to develop categories and sub-categories as emerging themes (Glaser and Strauss, 1967). First, we analysed the first 100 tweets from the EPL account available on the Web. The objective of this stage was to conduct a pilot test and a pilot analysis of the content of the tweets in order to allow us to gain a perspective of the nature of the tweets created by the Library. This analysis drew upon the previous categorisation
A. Shiri and D. Rathi

schemes reported in the literature. Both researchers analysed the 100 tweets and created a tentative scheme for the proper analysis. The outcome of our initial analysis resulted in a categorisation scheme consisting of 13 categories [see Section 4].

In the second stage of our analysis, a systematic sampling of 250 tweets out of the downloaded 2500 tweets was conducted for this study. We selected the first 10 tweets from every 100 tweets out of the total of 2500 tweets. This sampling was conducted to allow us to manually analyse a manageable sample of the tweets. We evaluated the nature, content and context of each tweet and categorised them according to the devised categorisation scheme. These 13 categories were used to classify the 250 sampled tweets and to evaluate their suitability and their comprehensiveness of their coverage of tweets. This type of analysis led to the finalisation and confirmation of the 13 initially identified categories. In addition, the sampled tweets were quantitatively analysed to assess the distribution of the tweets across the 13 categories.

A further analysis was carried out on the nature of tweets to establish the sub-categories and to identify the specific topics or examples of discussion topics within these sub-categories. It should be noted that the sub-categories represented the more specific aspects or issues of main categories. These sub-categories and their specific topic examples allowed us to form a more granular view of the content and the context of each tweet. A more descriptive account of these sub-categories along with topic examples are provided in Section 4 of this paper.

4. Findings and Discussion

While our proposed categorisation scheme shares some of the categories proposed by previous research, it has some unique and library specific categories. Our aim was to present a more public library-specific categorisation of tweets. For instance, advisory services, feedback seeking, information sharing and library operations are developed with a view to the library and information services usually offered in public libraries.

4.1. Tweet categories: Definition and examples

As was mentioned in the previous section, our tweet content categorisation scheme consists of 13 main categories. The following section provides a detailed account of the definition and coverage for each category.

- **Acknowledgement** — An example of a tweet categorised under this category has to be a direct acknowledgment, usually of thanks, to another tweet and contain within the tweet the @ reply acknowledging the recipient. An example of such a tweet would read as follows: “Thanks — so glad you like them! @ . . . Check out the Edmonton Public Library’s clever library cards . . . ”

- **Advisory Services** — Tweets categorised in this category must be a reply to a services related query from a patron to an EPL staff, as revealed in the context of the tweet itself, and usually take the form of using an @ reply format. Example: “@ . . . direct message us your library card number and we can check it out.”

- **Announcement** — An announcement tweet is from either EPL staff or another party to the general public announcing a related activity or news piece pertaining to or associated with EPL that may or may not take the form of an official event. Example: “EPL opens Sundays all year round! . . . ”

- **Event** — Similar to an announcement tweet, but pertaining to a specific event with a specific time and date contained within the tweet itself. Example: “Double feature picture show! 2 25-minute films: Rocky Barstad + Fred McDonald: Aboriginal artists. Fri@7pm. Downtown EPL . . . ”

- **Feedback Seeking** — Tweets under this category are general inquiries presented to a general audience, either by EPL, or by another party, that present a query worded to receive a direct answer. They can take the form of general inquiries or formal surveys. Example: “Naked Stephen Harper on wall at Kingston Frontenac Public Library @ . . . http://t.co/6uk71C3B Art, absurd…”

- **Informal Conversation** — Informal Conversation, as based on context, is an ongoing conversation taking place between either staff and patron, or patron and patron. Informal conversations are usually denoted by use of the @ reply and the RT function. Example: “@ . . . Family time at EPL’s Whitemud location is great! I have brought my little ones there several times. Glad you enjoy too!”

- **Information Sharing** — Composed of tweets where the identity of the sharer is unclear but where the focus is centered on sharing specific information with others, either publicly or to specific recipients, on any number of subjects. Example: “Just bought 5 DVDs, 3 books . . . LOVE our book sale. New books/movies/music every hour, I’m going back!”

- **Library Operations** — Library Operations encompass tweets relaying information about library hours, changes in library hours, library closures due to any number of reason, or anything else pertaining to the operations of the library. These tweets can take the form of an
announcement but must be related to operations. Example: “Just a reminder to all Abbottsfield branch customers — the branch will be closed on Monday, June 27 for a scheduled power outage.”

- **News** — These tweets must report on a newsworthy item of interest that does not originate within the EPL cultural framework, though is still pertinent to EPL and its clientele. Example: “The Raven opens in 14 theaters tonight. Perhaps pondering pusillanimously perusing Poe, perfect panic!...”

- **Opinion** — An opinion consists solely of expressing a personal opinion on a matter, be it a book, film, event, news item, or anything else. It can be directed to the general public or in a specific @ reply. Example: “I do love audiobooks. Calms my nerves on a long commute. Just finished the Golden Compass. Worth a listen!...”

- **Query** — Tweets that fall into this category directly ask a question with an expectation of a specific answer. The question can be to the general public or to specific respondent. Example: “Looking for a new audiobook to help make my commute bearable. Any good ideas from our oh-so-well read/listened followers?...”

- **Recommendation/Suggestions** — Tweets in this category consist mainly of suggestions, presumably by EPL staff to the general public, designed to lead the reader to the EPL catalogue. They almost always include catchy and creative introduction and an URL Pointer to the page in the catalogue where one can then access and borrow the suggested material. With that said, this category can also consist of recommendations or suggestions either from EPL to the public or from the public to EPL regarding any matter. Example: “Steve Jobs was born on Feb 24, 1955. A remarkable life makes for a great read...”

- **Request** — Requests include Tweets that directly demand something. The request is different from a query in that it is not specifically seeking an answer, but giving a directive. Example: “Share the reasons you lae these books”

A quantitative analysis and categorisation of the number of tweets was carried out to identify the most frequently created categories. As Fig. 1 shows the top five categories include information sharing, recommendations/suggestions, announcement, advisory services and informal conversation.

To gain an insight into the specific details of each of the above categories, we manually analysed the content of each tweet that had been categorised using the above main categories. Table 1 shows an overview of a set of sub-categories that we developed as a result of the analysis of the tweets using main categories. For instance, one category is advisory services. We were interested in examining the nature and types of advisory services tweets to provide a more granular perspective of the use of Twitter for advisory services. Examples of an advisory service may include “library martial” and “E-services”.

In the following examples of each sub-categories are presented.

- **About library**: Canadian Library Association, Edmonton Public Library, Librarians, Branches, Philosophy.
- **Library Materials**: Books, Audio Books, Newspapers, Magazines, CDs, DVDs, Video Games, Biography, Autobiography.
- **Culture**: Canadian Content, Surveys, Cinema, Film, Art, Dance, Music, Radio, Novels, Photos, Reading, Writing, Stories, Poetry, Authors, Sports, Television, Movies, Aboriginal, Contests, Fundraising, Family.
- **E-Services**: E-content, E-books, E-readers, iPad, Apps, Tags, iPhone, Freegal, Freading.
- **Programs**: Freedom to Read, Spread the Words, Writer in Residence, Writer in Exile, Leader in Residence, Summer Reading Club, The Great Library Read-In, Books to Film, Books into Movies, Food for Fines, Teen

![Fig. 1. Category distribution of tweets.](image-url)
A. Shiri and D. Rathi

Table 1. Main categories and sub-categories of public library tweets.

<table>
<thead>
<tr>
<th>Main Categories</th>
<th>Sub-categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>About library</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>✓</td>
</tr>
<tr>
<td>Advisory Services</td>
<td>–</td>
</tr>
<tr>
<td>Announcement</td>
<td>–</td>
</tr>
<tr>
<td>Event</td>
<td>–</td>
</tr>
<tr>
<td>Feedback Seeking</td>
<td>–</td>
</tr>
<tr>
<td>Informal Conversation</td>
<td>✓</td>
</tr>
<tr>
<td>Information Sharing</td>
<td>✓</td>
</tr>
<tr>
<td>Library Operations</td>
<td>–</td>
</tr>
<tr>
<td>News</td>
<td>✓</td>
</tr>
<tr>
<td>Opinion</td>
<td>✓</td>
</tr>
<tr>
<td>Query</td>
<td>✓</td>
</tr>
<tr>
<td>Recommendation/Suggestions</td>
<td>✓</td>
</tr>
<tr>
<td>Request</td>
<td>–</td>
</tr>
</tbody>
</table>

- **Memberships**: Accounts, Passwords, Library Cards, Benefits.
- **General Library Services**: Genealogy, Library Guides, Requests/Holds, Interlibrary Loans, Press Display, Study Guides, Newsletters, Tutorial.
- **Library Catalogue**: Bibliocommons.
- **Miscellaneous**: Temporary Closure, Construction, Power Outage, Water Main Break, Branch Opening, Opening Hours.

The analysis of EPL tweets reveals a number of interesting trends. EPL’s Twitter interactions indicate that twitter is used largely to communicate traditional library/patron interests. In the following paragraphs, we provide an analysis and discussion of the some of the specific categories which had relatively higher number of tweets.

The “Announcement” category, at 45 tweets, was the category with the highest number of tweets. As shown in Table 1, this category contained all varieties (i.e. sub-categories) of tweets including the most common tweets or sub-categories found in this were about the “Culture” and “Programs”. One of the interesting outcomes of this analysis is that it seems EPL patrons enjoy the Book Sale phenomenon (Sale/Books2Buy) which EPL periodically runs.

The “Information Sharing” (main) category, at 37 tweets, was the second highest utilised category. The “Information Sharing” category is composed of tweets where the focus is centered on sharing specific information with others, either publicly or to specific recipients, on different subjects but primarily centered around books, movies and other media. The most prevalent sub-category in “Information Sharing” was the “Culture” sub-category where the most prevalent discussion was around movies, i.e. Twitter was used to exchange information on popular feature film culture.

In “Recommendation/Suggestions” with 35 tweets is the third highest number of interactions. Tweets in this category consist mainly of suggestions, most likely by EPL staff to the general public, designed to lead the reader to the EPL catalogue. They almost always include a catchy and creatively worded stories and an URL Pointer to the page in the catalogue where one can then access and borrow the suggested material. An example of creativity reads, “MT … The list of people outlived by Keith Richards is getting ridiculous…; Keith’s bio is fascinating http://t.co/qAQjaqTg”. Thus, this category is dominated by the sub-category, “Library Material” and a large number of tweets in this sub-category were related to Books. This highlights the fact that while patrons seek out newer medium on their own accord, they still rely on traditional librarianship to guide them to their books.

“Informal Conversation”, at 28 Tweets, is the category that provides insight into the ongoing conversations taking place on Twitter. “Informal conversations” are usually denoted by use of the @ reply and the RT function. Within this category a considerable number of tweets are about some “cultural” aspects and in that a large number is around “movies”. This finding implies that most informal conversations on EPLs twitter exchanges involve interactions with popular feature film culture.
“Advisory Services”, with 25 tweets, is a category that is normally concerned with a reply to a services related query from an EPL staff to a patron, as revealed in the context of the tweet itself, and usually take the form of using an @ reply format. Within this broad category are included some “general library service” (the major sub-category) queries such as account renewal etc.

Tweets categorised under the “Acknowledgement” category (19 tweets) are associated with a direct acknowledgment to another tweet and contain within the tweet the @ reply acknowledging the recipient. Most common is an acknowledgment from staff to a patron regarding services offered by EPL, such as E-Books, Freegal and Book Sale/Books2Buy. E-Services prevalent under Acknowledgment is indicative of the large volume of staff assistance in the growing E-Services portion of library services, especially as it pertains to E-books. The Freegal service is new as of the collection of this sample of tweets, indicating a high initial interest in the use of the Freegal music service. By far, E-Books is the most sought after service through direct and ongoing exchanges between staff and patrons, indicated by the gratitude on either side when information is shared regarding the functioning of e-Books through EPL.

The opinion category, with 13 tweets, consists solely of expressing a personal opinion on a matter, be it a book, film, event, news item, or anything else. It can be directed to the general public or in a specific @ reply. The most commonly expressed opinions were associated with the library’s “culture” aspect. The tweets associated with this sub-category show that twitter is highly utilised by EPLs clientele in interacting over common interests both in expressing and receiving feedback through traditionally library related cultural factors, such as Cinema, Film, Art, Dance, Music, Radio and Photos but primarily the focus is on “Movies”.

The “Query” category covers the questions with an expectation of a specific answer. The question can be to the general public or to specific respondent. Queries were about library materials and electronic services, in particular, about books. The implications here suggest that while patrons of EPL use twitter to interact with the wider library community when it comes to sharing ideas and opinions, they still rely on traditional Librarian interaction exchanges when it comes to specific service related questions. This category also shows that there is a high demand and need through Twitter for information and assistance on how to use these emerging services.

The above analysis and findings provide a basis for the understanding of trends in the use of Twitter in public libraries as represented by parsed categories and examples discussed above.

5. Conclusion

This paper reported an exploratory study of the use of Twitter by the Edmonton Public Library. This paper extended upon the previous work by Aharony (2010), Cahill (2011) and Dann (2010) and thus, the paper contributed to the overall growing body of literature on social media and particularly on Twitter with a particular focus on the content categorisation. A general categorisation scheme was proposed for the analysis and classification of tweets within a public library context. Based on our analysis of the tweets, 13 main categories were developed (e.g., acknowledgement, advisory services, announcements, event, information sharing, news, opinions, etc.). Given that some these categories were broad in nature and required more granular analysis of content, a number of sub-categories were developed to more specifically elaborate on the content and nature of the tweets in a meaningful way.

One of the key findings of this study was that a significant number of tweets were associated with information sharing, recommendations/suggestions and advisory services. These tweets demonstrate some of the ways in which libraries, in general, and public libraries, in particular, can adopt and utilise Twitter for various user-centered and collection-centered services and activities. The categorisation scheme proposed in this paper can provide a useful framework for the analysis and classification of tweets created and conveyed within the context of libraries and information services. It complements previous research on Twitter classification through proposing a number of library-specific analytical categories and sub-categories.

One of the limitations of this study was that it utilised a small sample of tweets from only one public library. However, the study has generated a good categorisation scheme which can be used to analyse a larger sample of public library tweets and can be adapted to suit other types of libraries. Given the challenge of knowledge management and content archiving in public libraries, this categorisation scheme can provide an analytical framework for the organisation, representation and management of twitter content in libraries and also serve as a model.

References


