DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2019

Semantic Metadata as Meaning Making: Examining #hashtags and Collection Level Metadata

Hollie C. White
Curtin University, Australia
hollie.white@curtin.edu.au

Leisa Gibbons
Curtin University, Australia
leisa.gibbons@curtin.edu.au

Eileen Horansky
Yale University, United States
eileen.horansky@yale.edu

Abstract

Memory institutions and other organizations interested in preserving social media data are using a variety of collection level metadata to represent those materials. The aim of this paper is to start a dialogue within the metadata community about how metadata professionals can describe social media collections in better ways to ensure that the semantic complexity of hashtags remain intact at the collection level. This paper explores how hashtags manifest semantic metadata and how that expression is formally described. A study was conducted using two datasets. The first dataset on hashtags as defined by professional literature was examined and categorized using thematic analysis. The second dataset collected metadata from a selection of Document the Now Twitter datasets and was categorized using Gilliland's (2016) five categories of metadata. Findings delve into the use of collection level metadata to describe social media content.

Keywords: semantic metadata; meaning making; hashtags; hashtag datasets; collection level metadata; constructivism

1. Introduction

This short paper explores the meaning of hashtags and questions what happens to that meaning when they are put in large collections described by metadata. The paper begins by presenting literature on what is semantic meaning and meaning-making. For this study, two datasets were collected and analyzed. First, hashtags, as defined by scholarly peer-reviewed literature from 2006 to 2016, are examined and categorized using thematic analysis. Second, the metadata used to describe a selection of 2018 Twitter metadata collections is analyzed in relation to Gilliland's five categories of metadata. The discussion and conclusion emphasize the need for more conversation within the metadata community on how to best describe social media and hashtag content.

2. Meaning making and hashtags

In a social constructionism paradigm, meaning-making is inherently a social activity that draws from cultural knowledge, and is linked to a specific time and place. Meaning (as an outcome) is therefore socially constructed within specific contexts yet variable and interpreted according to individual assumptions and actions (Burr, 2003; Lock & Strong, 2010). The processes of meaning-making are contextual and evidential in that they provide information about how and why something was constructed (in its time and place). In this sense, semantic metadata plays multiple roles in organization, representation, and communication of meaning with the goal to emphasize and enrich the complex heritage of data's meaning or purpose.

Using meaning-making as a frame, expressions of individuality and uniqueness of the creator are central to the metadata creation process. While meaning-making can be unique and idiosyncratic, there are "regularities to the underlying structure of meaning-making systems" (Kegan, 1980, pp. 374). In regard to semantic metadata, especially those found in social media and the use of hashtags, this extends the function that the individual person plays within a larger social context.

https://doi.org/10.23106/dcmi.952141651

3. Methods

The purpose of this study is to explore how hashtags manifest semantic metadata and how that expression of semantic metadata is formally described in collection level metadata. A social constructionism interpretation of meaning-making underlies this study's definition of semantic metadata. The research question for this study is, *to what extent is hashtag-based semantic metadata represented by formally gathered collection level metadata?* A two-part mixed methods approach was used to help answer this question.

The first part of this study categorized the semantic metadata functions of hashtags as defined in scholarly articles. Social media research and studies are conducted by a variety of domains and disciplines (Costa et al., 2013; Highland & Leaver, 2015) and in order to get a multidisciplinary perspective of hashtag definitions, Google Scholar was used for data collection to identify either seminal and/or influential scholarly papers. Perspectives and papers from information, communication, media, and cultural studies disciplines were the main data collection focus. Seminal and/or influential papers were defined as articles that had more than 10 Google Scholar-identified citations. Each Google Scholar search was conducted using the specific discipline listed above and the word "hashtag" as keywords and then narrowed to a single year beginning with 2006 and expanding from that point year by year (e.g. 2006-2007, 2008-2009, etc.). The following data about article was captured: title; year; URL; keywords used by author (up to 6); discipline; the text used in the paper to define hashtags; and codes that categorized how hashtags were being defined. A total of 370 papers were found for this part of the study. Iterative thematic analysis was used to group the different ways that hashtags were defined in context within the various domains of knowledge to identify their semantic meaning. Data collection, verification, and thematic analysis was conducted from December 2016 to July 2017.

The second part of this study examined the types of metadata being used to describe social media collections and hashtag datasets. Documenting the Now's Tweet Catalog (Tweet ID Datasets at https://www.docnow.io/catalog/) from 2018 were chosen as the collection used for part 2. Documenting the Now (https://www.docnow.io/), a major ongoing research project in the United States started in 2016, was designed to explore the needs, limitations and ethics around capturing and building social media datasets, collections, and archives. Data from Documenting Now mainly comes from Twitter and focuses on topics with international interests and themes. Datasets are added voluntarily to the Documenting the Now catalog and reflect a global range of institutions and all its metadata are licensed through Creative Commons.

All 25 social media datasets cataloged in Documenting the Now in 2018 were included in this study's second dataset. Collection level metadata from each dataset was recorded and then categorized. Each metadata element (labeled or unlabeled) was categorized according to Gilliland's (2016) main information object features (Content, Context or Structure) and main categories of metadata (administrative, descriptive, preservation, technical, and use). Data collection and analysis for the second part of this study was conducted from March to April 2019.

4. Findings

Study findings are presented in two parts. In Part 1 results, the semantic meaning of hashtags, are reported. In Part 2, results related to social media and hashtag collection level metadata, are reported.

4.1. Part 1 Results

https://doi.org/10.23106/dcmi.952141651

A total of 370 papers were identified in part 1 of this study as seminal and/or influential scholarly papers from information, communication, media, and cultural studies disciplines about hashtags. Of this group, 293 papers were used for the thematic analysis, however only 225 papers included clear definitions of hashtags. The thematic analysis revealed a total of 18 semantic meanings of hashtags. Themes are included below,

1. language
2. adoption of hashtags
3. engagement
4. mediated/ mediatized politics
5. folksonomy/ taxonomy
6. classification
7. communication politics
8. microblogging
9. aggregated communication
10. geotagging
11. collaborative meaning making
12. innovation
13. diffusion
14. user-generated content
15. crisis informatics
16. networked communication
17. convergence
18. health informatics

The Appendix includes definitions.[1]

4.2. Part 2 Results

In 2018, 25 social media datasets were submitted to Document the Now's Tweet Catalog. Six different institutions/organizations held the datasets cataloged by Document the Now. These institutions were the Internet Archive, the University of Las Vegas (UNLV), the Harvard Dataverse/GW Libraries Dataverse (Harvard/GWU), the University of California at Riverside (UCR), the University of North Texas (UNT), and the University of Glasgow (UG). The datasets range in size from 364 to 65,000,000 tweets. Table 1 elaborates on the number of submission and number of fields used by each institution.

TABLE 1: Number of submission and fields by Collecting Institution

Collecting Institution	Number of 2018 submissions	Number of fields
Internet Archive	11	6-12
UNLV	4	21
Harvard/GWU	3	10
UCR	5	8-9
UNT	1	21
UG	1	15

The Document the Now Tweet Catalog has seven metadata elements: title, creator, tweets, published, date coverage, tags, and description. The title links to the collection level metadata found at the collecting institution/organization. Collection level metadata used to describe the datasets are unique to the institution that housed the dataset. The number of metadata elements used to describe the datasets ranged from six to twenty-one. Due to interface decisions some metadata fields displayed are repeated. Repeated metadata elements were not included in these calculations.

[1] To conform with submission page limits, the Appendix is not attached to this paper, but available online at: http://bit.ly/2PkOGxL

https://doi.org/10.23106/dcmi.952141651

A total of 273 metadata elements were used across all the examined collections. Each element from each collection was evaluated in two ways: by information object feature and by main categories. As mentioned in the Methodology section, these categories are based on Gilliland's (2016) main information object features (Content, Context or Structure) and main categories of metadata (administrative, descriptive, preservation, technical, and use). Table 2 shows the number of elements corresponding to Gilliland's (2016) main object features. Table 3 shows the number of elements corresponding to Gilliland's (2016) main metadata categories.

TABLE 2: Number metadata elements by of main object features

Main Object Features	Number of Metadata Elements
Content	125
Structure	60
Context	88

TABLE 3: Number metadata elements by of main categories of metadata

Main Categories of Metadata	Number of Metadata Elements
Administrative	64
Descriptive	155
Preservation	2
Technical	29
Use	23

5. Discussion and Future Questions

5.1. Use of Collection Level Metadata to Describe Social Media Content

Descriptive metadata is included more than all other metadata categories in collection level metadata and account for over 50% of the metadata elements used. This would imply that semantic meaning is an important component of collection level metadata. Yet, is all descriptive metadata semantic in nature? While descriptive metadata can be semantic in nature, not all descriptive metadata fields are central to emphasizing meaning and purpose. The relationships between descriptive metadata and semantic metadata is not explored in this paper, but will be the focus of future research.

Like descriptive metadata, the content object feature is more prevalent than other object features, accounting for around 45% of the collection level metadata elements used. This means that intrinsic features of the information object, in this case social media content, are represented the most in collection level metadata. Content, which focuses on extrinsic features, made up about 33% of the metadata elements. Structural object features, which focus on how information relate to each other, was represented with the least number of collection level metadata elements, made up 22% of total elements

5.2. Further Questions and Future Study

This paper is an early exploration that begins to question the semantic meaning of hashtags and what happens to that meaning once hashtags are put into collections and described using collection level metadata. Preliminary findings from this study introduce more questions for further study and consideration. Some of those questions are,

https://doi.org/10.23106/dcmi.952141651

- *Since hashtags serve a variety of functions when used in social media, how well are those functions preserved and maintained once they are put into collections?*
- *In Tweet-based and other social media collections, what aspects of those items are most often represented?*
- *Is the metadata used in collection level metadata sufficient for describing the semantic meaning of hashtags?*
- *Are there better types of metadata that can be used to make sure that semantic meaning of hashtags are maintained when used in collection level metadata?*

Some of these topics will be addressed in other papers and presentations, but may be best answered within the metadata community itself. Overall, this topic is broad, and a single paper cannot accurately represent the complexity of the relationships between metadata, semantic meaning, and hashtags. More and different types of research are needed to fully explore the semantic nature of hashtags and how that can be best represented in collection level metadata describing social media collections.

6. Conclusion

This paper aims to explore the semantic metadata function of hashtags used in social media and to see how well that same data is represented in formal collection level metadata. While attempting to archive social media content for long term preservation and use, the metadata used for describing social media may not fully convey some of the important semantic meaning expressed through hashtags that is central to meaning-making. The purpose of this paper is to start a dialogue within the metadata community about how metadata professionals can describe social media collections in better ways to ensure that the semantic complexity of hashtags remains intact at the collection level.

References

Burr, V. (2003). Social constructionism (2nd ed.). London, UK: Routledge.

Costa, J., Silva, C., Antunes, M., & Riberiro, B. (2013) Defining semantic meta-hashtags for Twitter classification, in M. Tomassini et al. (eds), *International Conference on Adaptive and Natural Computing Algorithms 2013*. 4-6 April, Lausanne, Switzerland. pp. 226-235.

Gilliland, A. (2016). Setting the Stage, in Baca, M. (ed). *Introduction to Metadata.* 3d ed. Los Angeles: Getty Research Institute.

Highland, T. & Leaver, T. (2015 January). A methodology for mapping Instagram hashtags. *First Monday, 20*(1-5).

Kegan, R. (1980 January). Making meaning: The Constructive-Developmental Approach to persons and practice. *Personnel & Guidance Journal*. pp. 373-380.

Lock, A., & Strong, T. (2010). *Social constructionism: Sources and stirrings in theory and practice.* Cambridge, UK: Cambridge University Press.

Postman, N. & Weingartner, C. (1969). *Teaching as a subversive activity.* New York: Delacorte Press.

https://doi.org/10.23106/dcmi.952141651