Image Indexing Through Hashtags in Instagram

Isabelle Dorsch
Dept. of Information Science, Heinrich Heine University, Germany. isabelle.dorsch@hhu.de

Wolfgang G. Stock
Dept. of Information Science, Heinrich Heine University, Germany. stock@phil.hhu.de

Franziska Zimmer
Dept. of Information Science, Heinrich Heine University, Germany. franziska.zimmer@hhu.de

ABSTRACT

Image indexing and knowledge representation on Instagram are organized by folksonomy-oriented hashtags. What kinds of hashtags do Instagram users apply for different picture categories? We distinguish between food, pets, selfies, friends, activity, art, fashion, quotes (captioned photos), landscape and architecture as image categories, as well as content-related (ofness, aboutness, iconology), emotiveness, isness, performativeness, fakeness, "Insta"-tags and sentences as hashtag categories. Are there any differences in relative frequencies of hashtags in the image categories? What hashtag categories dominate users' indexing activities? Given an image category, what is the distribution of hashtag categories? Given a hashtag category, what is the distribution of image categories? We analyzed 1,000 pictures on Instagram with all-in-all 14,649 hashtags deploying content analysis.

KEYWORDS

image indexing, Instagram, knowledge organization, folksonomy, hashtag, picture, user behavior, tagging behavior

INTRODUCTION

Besides indexing through controlled concepts (Stock & Stock, 2013), folksonomies (Peters, 2009) obtain a huge impact on indexing especially on social media. Folksonomies allow for the free allocation of keywords, called "tags" or "hashtags," by everyone. In Instagram, a mobile social networking application for sharing photos and videos, image indexing is realized through hashtags, e.g., #Loersfeld, #Castle and #Kerpen in Figure 1. The creator of a post has the possibility to add a description text and up to 30 distinct hashtags.

How do Instagram users tag their pictures? Our model of tagging behavior on Instagram is depicted in Figure 2.

Based upon this theoretical model, our research questions (RQs) are:

Figure 1. *Tagged image on Instagram*

(**RQ1**) Are there any differences in relative frequencies of hashtags in the image categories?

(**RQ2**) Given an image category, what is the distribution of hashtag categories?

(**RQ3**) Given a hashtag category, what is the distribution of image categories?

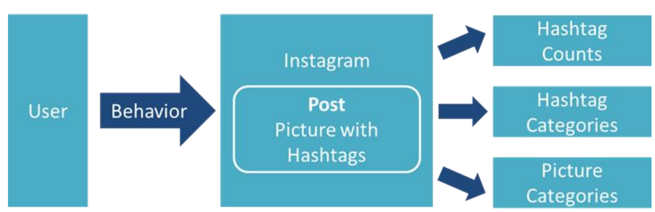

Figure 2. *Theoretical model of users' tagging behavior on Instagram*

80th Annual Meeting of the Association for Information Science & Technology, Washington, DC | Oct. 27-Nov. 1, 2017

METHODS

To analyze the tagging behavior of Instagram users, a content analysis (Krippendorff, 2014) of image postings was conducted. Applying a codebook, two coders independently categorized 1,000 pictures and 14,649 hashtags each. There were ten image categories (food, pets, selfies, friends, activity, art, fashion, quotes, landscape and architecture). Here, we follow predominantly Hu et al. (2014). Additionally, we worked with seven hashtag categories. Content-related tags contain *ofness* and *aboutness* (Shatford, 1986) which in turn are based on Panofsky's levels of meaning in art (Panofsky, 1955). Besides the known keyword categories *emotiveness* (emotional hashtags as #sad) (Knautz, 2012), *isness* (describing properties of the post as #photo) (Ingwersen, 2002), *performativeness* (requests of actions as #like4like) (Peters & Stock, 2007), *complete sentences* (#Ifeelgood), the hashtag categories fakeness (tags with no relation to the image) and *"Insta"-tags* (e.g., #instagood) were exclusively designed for this study. The data collection took place between November 2016 and January 2017.

RESULTS

RQ1. The average number of hashtags per image post on Instagram varies from nearly 11 to about 19 hashtags with an average of 15 hashtags. The person-related categories *selfie* (⌀ 10.9 tags per picture) and *friends* (⌀ 11.7 tags per picture) received the lowest average values. *Pet* (⌀ 18.6 tags), *fashion* (⌀ 17.6 tags) and *landscape* (⌀ 16.8 tags) are the categories with the highest average hashtag count.

RQ2. With 60.20%, the majority of all hashtags were classified into the category *content-related*, followed by *isness* with 14.87%, *"Insta"-tags* (7.32%) and *performativeness* (7.20%). Only a minority of hashtags was classified into the categories *emotiveness* (about 4.38%) and *sentences* (0.99%). 5.03% of all tags are *fake* keywords.

RQ3. The highest value of relative frequency of *content-related tags* is in the category *activity* (74.29%), the lowest is *selfie* (51.05%). Most emotions are shown in *friends* (8.74%) and *pets* (8.02%), least of all in *art* (1.05%). *Landscape* images call at most for actions, as 10.75% of all landscape-tags are performative. High values for "Insta"-tags in the category *pet* (20.24%) and *fakeness* in *quotes* (11.80%) are striking.

All in all, after a chi-square test of independence, there is a statistically significant association between hashtag categories and image categories on Instagram.

	Content-related	Emotiveness	Fakeness	"Insta"-Tags	Isness	Performativeness	Sentences	%
#activity	74.29	4.35	3.75	1.72	9.90	5.25	0.75	100.00
#architecture	63.20	2.54	2.54	6.67	15.41	9.35	0.28	100.00
#art	68.54	1.05	5.62	4.19	14.68	5.69	0.22	100.00
#fashion	59.51	3.52	7.38	5.91	16.70	6.02	0.97	100.00
#food	51.34	1.81	6.71	6.31	25.84	7.11	0.87	100.00
#friends	57.75	8.74	3.94	7.88	14.40	6.17	1.11	100.00
#landscape	61.68	3.98	1.66	5.47	16.16	10.75	0.30	100.00
#pet	53.93	8.02	3.82	20.24	7.59	4.31	2.10	100.00
#quote	61.67	5.16	11.80	3.46	7.33	8.75	1.83	100.00
#selfie	51.05	4.57	2.38	7.96	23.70	9.06	1.28	100.00
Total	60.20	4.38	5.03	7.32	14.87	7.20	0.99	100.00

Table 1*. Relative frequency of hashtag categories by picture categories (N = 1,000 posts; 100 posts per picture category)*

REFERENCES

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In *Proceedings of the 8th International Conference on weblogs and social media, ICWSM 2014* (pp. 595-598). Palo Alto, California: AAAI Press.

Ingwersen, P. (2002). Cognitive perspectives of document representation. In H. Bruce, et al. (Eds.), *Emerging Frameworks and Methods: CoLIS 4* (pp. 285-300). Greenwood Village, CO: Libraries Unlimited.

Knautz, K. (2012). Emotion felt and depicted: Consequences for multimedia retrieval. In D. R. Neal (Ed.), *Indexing and retrieval of non-text information* (pp. 343-375). Berlin, Boston, MA: De Gruyter Saur.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Panofsky, E. (1955). *Meaning in the visual arts.* Garden City, NY: Doubleday Anchor Books.

Peters, I. (2009). Folksonomies: Indexing and retrieval in web 2.0. Berlin: De Gruyter Saur.

Peters, I., & Stock, W. G. (2007). Folksonomy and information retrieval. In Joining research and practice: Social computing and information science. Proceedings of the 70th ASIS&T Annual Meeting (pp. 1510-1542).

Shatford, S. (1986). Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, *6*(3), 39-62.

Stock, W. G., & Stock, M. (2013). *Handbook of Information Science*. Berlin, Boston, MA: De Gruyter Saur.