ACADEMIA | Letters

# *Instagram hashtags as image metadata for Automatic Image Annotation*

Stamatios Giannoulakis, Nicolas Tsapatsoulis
Stamatios Giannoulakis, Department of Communication and Internet Studies, Cyprus University of Technology, Limassol, Cyprus
Nicolas Tsapatsoulis, Department of Public Communication, Cyprus University of Technology, Limassol, Cyprus

---

Around 3.2 billion digital images are shared on the Internet and social media every day[1]. Locating those images is a challenging task. It is essential to develop effective and efficient methods that allow users to retrieve images according to their needs. Different image retrieval methods have been proposed for image retrieval. Automatic Image Annotation(AIA) is the latest development in image retrieval that has drawn huge researchers' attention. AIA methods use pair image-tag paradigms to learn the visual representation of semantic concept models to tag new images automatically.

Creating image-tag pairs is not easy because it is essential to create excellent and representative models. Manual annotation is challenging and time-consuming since many images are necessary to create effective concept models. In addition, human judgment has the drawback of errors and subjectivity. So, it is crucial to locate a methodology that automatically creates pairs of images and tags. In modern social media, such as Instagram, we can locate images and associated hashtags that we can exploit to create image-tag pairs. With the help of Instagram, we can create image-tags pairs for AIA. Instagram is a photo-oriented social media platform where users upload images and describe them with hashtags; thus, we can use Instagram as a source for image-tags models in the AIA framework.

*Academia Letters, June 2022* 

**Corresponding Author:** Stamatios Giannoulakis, s.giannoulakis@cut.ac.cy
**Citation:** Giannoulakis, S., Tsapatsoulis, N. (2022). Instagram hashtags as image metadata for Automatic Image Annotation. *Academia Letters,* Article 5786. https://doi.org/10.20935/AL5786

We could use other social media platforms for AIA purposes, such as Facebook, YouTube, Pinterest, Twitter, Snapchat, or Flickr, but these platforms are not ideal for our study. Hashtags on Facebook are not very popular; YouTube focuses on video, and users are not familiar with hashtags. On Twitter, the portion of tweets containing text only is far more extensive than those consisting of images, videos, or gifs. Pinterest is a catalog of ideas, and hashtag functionality is a relatively new feature, so it is not ideal for AIA purposes. The role of hashtags is not the same in Snapchat, and Flickr is not very popular anymore. So, we can easily conclude that Instagram is ideal for AIA purposes.

The first step is to examine if Instagram hashtags and images are appropriate for AIA. It is essential to investigate if Instagram hashtags are suitable as image tags and calculate a rough estimation of the percentage of Instagram hashtags that describe the visual content of accompanying images. To achieve the goal mentioned above, we conducted two-stage research. In the first stage, we collected 30 images, and in the second stage, we increased the images to 1000. For each image, we collected 1 to 4 hashtags that better describe the visual content of each image. Then, we created an online questionnaire, gave the users relevant and irrelevant hashtags, and asked them to annotate those hashtags that better describe the image's content. From the results, we concluded that approximately 20% of Instagram hashtags are related to Instagram images' visual content [2,3].

Since only 20% of Instagram hashtags are related to the visual content of images, it is crucial to locate methodologies to filter irrelevant hashtags and keep only those that describe the visual content of the image. We focused on locating stophashtags, meaningless hashtags that appear in irrelevant image categories. We collected images and hashtags from 30 different subjects (e.g. #dog, #cat), and we identified Like4Like, instagood, love, l4l, follow4follow, picoftheday, instalike, instagramers, vscocam, instamood, like4like, likes4likes, L4L, beautiful, f4f, follow, instalikes, Instagram, followme, vsco, like, likeforlike, likesforlikes, l4like, followforfollow, likes[4]. Another approach we examined for hashtag filtering is the use of HITS algorithm. HITS is a ranking algorithm that we could use to filter Instagram hashtags and locate the most relevant. So, we apply HITS algorithm to identify image tags in a crowdsourcing environment. A two-stage search was conducted. In the preliminary research, a subset of 100 Instagram images from the set of 1000 images we used in our previous experiment was used in the current study. For each of the 100 images, we have manually selected 1-4 hashtags that better describe its visual content according to our interpretation. These hashtags consist of the ground truth, which we used to evaluate the proposed method. The primary research implied the algorithm in a real crowdtagging environment facilitated by the Figure-eight. A set of 50 Instagram images and their hashtags were automatically crawled with the aid of Python.

 

**Corresponding Author:** Stamatios Giannoulakis, s.giannoulakis@cut.ac.cy

**Citation:** Giannoulakis, S., Tsapatsoulis, N. (2022). Instagram hashtags as image metadata for Automatic Image Annotation. *Academia Letters,* Article 5786. https://doi.org/10.20935/AL5786

The collected Instagram images were uploaded to Figure-eight for crowdtagging in the form of tag selection. Five hundred annotators annotated every image for experimentation purposes. We implied the HITS algorithm to locate the most relevant hashtags in both stages. From the results, we can conclude that with the help of the HITS algorithm, we can locate hashtags related to the visual content of images[5,6].

The graph-based method is based on the crowd and, as a result, cannot be automated. Topic modeling is based on word probabilities. Words with higher probabilities in a corpus can give a good idea of what topics are discussed in that corpus. Assuming that a corpus can be derived by compiling all hashtags appended to Instagram images retrieved via a single hashtag query, we can use topic modeling to find relevant hashtags to the queried hashtag. This approach can be automated because we can collect hashtags from relevant images, imply topic modeling, and locate relevant hashtags. We constructed a dataset composed of 1000 Instagram images and their hashtags by querying 20 different subjects/hashtags (i.e., #airplane, #ring, etc.). The Latent Dirichlet Allocation (LDA) method was applied to the hashtags of each subject in an effort to create topic models for each one of the subjects. Topic coherence was used to evaluate topic models, and we concluded that topic model could produce relevant hashtags [7]. Moreover, we examined the accuracy of topic models with human interpretation. For each one of the topic model a word cloud was created. The token corresponding to the associated subject (query hashtag) was excluded in order to examine whether human would guess it correctly. The results showed that Instagram images of similar visual content share hashtags that are related to the subject. So the hashtags derived from topic model can describe an image[8]. Furthermore, we explored if we could achieve Automatic Image Annotation with the help of Instagram images and hashtags. So, we explore if we can bridge the semantic gap between image low-level features such as color histogram and high-level semantic content as hashtags. We concluded that color histogram and hashtag sets are complementary for image retrieval, especially for relevant posts, we have a strong indication to continue research to lower the semantic gap. Finally, we explored whether Instagram photos collected through a query hashtag (subject) can be used for adapting concept models with the aid of transfer learning. The main conclusion is that we can use pretrained models to classify with high recall images from Instagram.




**Corresponding Author:** Stamatios Giannoulakis, s.giannoulakis@cut.ac.cy
**Citation:** Giannoulakis, S., Tsapatsoulis, N. (2022). Instagram hashtags as image metadata for Automatic Image Annotation. *Academia Letters,* Article 5786. https://doi.org/10.20935/AL5786

## References


[1] Thomson, T., Angus, D. and Dootson, P., 2022. *3.2 billion images and 720,000 hours of video are shared online daily. Can you sort real from fake?*. [online] The Conversation. Available at: <https://theconversation.com/3-2-billion-images-and-720-000-hours-of-video-are-shared-online-daily-can-you-sort-real-from-fake-148630> [Accessed 10 March 2022].

[2] S. Giannoulakis and N. Tsapatsoulis, "Instagram Hashtags as Image Annotation Metadata," in *11th International Conference on Artificial Intelligence Applications and Innovations*. Cham: Springer, September 2015, p. 206–220. [Online]. Available: https://doi.org/10.1007/978-3-319-23868-5_15.

[3] S. Giannoulakis and N. Tsapatsoulis, "Evaluating the descriptive power of Instagram hashtags," *Journal of Innovation in Digital Ecosystems*, vol. 3, no. 2, pp. 114–129, December 2016. [Online]. Available: https://doi.org/10.1016/j.jides.2016.10.001.

[4] S. Giannoulakis and N. Tsapatsoulis, "Defining and Identifying Stophashtags in Instagram," in *2nd INNS Conference on Big Data*. Cham: Springer, October 2016, p. 304-313. [Online]. Available: https://doi.org/10.1007/978-3-319-47898-2_31.

[5] S. Giannoulakis, N. Tsapatsoulis and K. Ntalianis, "Identifying image tags from Instagram hashtags using the HITS algorithm," in *15th IEEE International Symposium on Dependable, Autonomic and Secure Computing*. Piscataway, NJ : IEEE, November 2017, p. 89-94. [Online]. Available: https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.29.

[6] S. Giannoulakis and N. Tsapatsoulis, "Filtering Instagram Hashtags through crowdtagging and the HITS algorithm," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 592 –603, June 2019. [Online]. Available: https://doi.org/10.1109/TCSS.2019.2914080.

[7] A. Argyrou, S. Giannoulakis and N. Tsapatsoulis, "Topic modelling on Instagram hashtags: an alternative way to Automatic Image Annotation?," in *13th International Workshop on Semantic and Social Media Adaptation and Personalization*. Piscataway, NJ : IEEE, September 2018, p. 61-67. [Online]. Available: https://doi.org/10.1109/SMAP.2018.8501887.


 

**Corresponding Author:** Stamatios Giannoulakis, s.giannoulakis@cut.ac.cy
**Citation:** Giannoulakis, S., Tsapatsoulis, N. (2022). Instagram hashtags as image metadata for Automatic Image Annotation. *Academia Letters,* Article 5786. https://doi.org/10.20935/AL5786

[8] S. Giannoulakis and N. Tsapatsoulis, “Topic Identification via Human Interpretation of Word Clouds: The Case of Instagram Hashtags,” in *17th International Conference on Artificial Intelligence Applications and Innovations*. Cham: Springer, June 2021, p. 283–294. [Online]. Available: https://doi.org/10.1007/978-3-030-79150-6_23

 

**Corresponding Author:** Stamatios Giannoulakis, s.giannoulakis@cut.ac.cy
**Citation:** Giannoulakis, S., Tsapatsoulis, N. (2022). Instagram hashtags as image metadata for Automatic Image Annotation. *Academia Letters,* Article 5786. https://doi.org/10.20935/AL5786