FILTERING INSTAGRAM HASH TAGS THROUGH CROWD TAGGING AND THE HITS ALGORITHM

V. Sarala [1], Ch. Pooja Kiranmai[2],

[1]**Assistant professor , MCA DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-vedalasarala21@gmail.com
[2]**PG Student of MCA,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-poojakiranmai9@gmail.com

ABSTRACT

Instagram is a rich source for mining descriptive tags for images and multimedia in general. The tags-image pairs can be used to train automatic image annotation (AIA) systems in accordance with the learning by example paradigm. In previous studies we had concluded that, on average, 20% of the Instagram hash tags are related to the actual visual content of the image they accompany, i.e., they are descriptive hash tags, while there are many irrelevant hash tags, i.e., stop-hash tags, that are used across totally different images just for gathering clicks and for searchability enhancement. In this work, we present a novel methodology, based on the principles of collective intelligence, that helps locating those hash tags. In particular, we show that the application of a modified version of the well-known HITS algorithm, in a crowd tagging context, provides an effective and consistent way for finding pairs of Instagram images and hash tags, that lead to representative and noise-free training sets for content-based image retrieval. As a proof of concept, we used the crowd sourcing platform Figure-eight to allow collective intelligence to be gathered in the form of tag selection (crowd tagging) for Instagram hash tags. The crowd tagging data of Figure-eight are used to form bipartite graphs in which the first type of nodes corresponds to the annotators and the second type to the hash tags they selected. The HITS algorithm is first used to rank the annotators in terms of their effectiveness in the crowd tagging task and then to identify the right hash tags per image.

1 INTRODUCTION

Social media are online communication channels dedicated to community-based input, interaction, content sharing and collaboration. These media give the users the opportunity to share their content such as, text, video and images [31]. Users usually accompany the content they post with text such as comments or hashtags. That alternative text (comment, hashtags etc.) provide valuable information about the users posts and other information. Preece et al. [32] to construct a Sentinel platform that can enhance social media data in order to understand different situations they based also in YouTube video comments. Sagduyu et al. [33] present a novel system that can present

large-scale synthetic data from social media. In their system they use textual content (hashtags and hyperlinks in tweets) to produce topics and train n-gram model. The users in several of those media, e.g. Twitter, Instagram and Facebook, use hashtags to annotate the digital content they upload. Hashtags are, usually, words or non-spaced phrases preceded by the symbol # that allow creators / content contributors to apply tagging that makes it easier for other users to locate their posts.

2 RELEATED WORK

Topic modelling on Instagram hashtags: An alternative way to automatic image annotation:

Automatic Image Annotation (AIA) is the process of assigning tags to digital images without the intervention of humans. Most of the modern automatic image annotation methods are based on the learning by example paradigm. In those methods building the training examples, that is, pairs of images and related tags, is the first critical step. We have shown in our previous studies that hashtags accompanying images in social media and especially the Instagram provide a reach source for creating training sets for AIA. However, we concluded that only 20% of the Instagram hashtags describe the actual content of the image they accompany, thus, a series of filtering steps need to apply in order to identify the appropriate hashtags. In this paper we apply topic modelling with Latent Dirichlet Allocation (LDA) on Instagram hashtags in order to predict the subject of the related images. Since a topic is composed by a set of related terms, the identification of the visual topic of an Instagram image, through the proposed method, provides a plausible set of tags to be used in the context of training AIA methods.

3 IMPLEMENTATION STUDY

Existing System:

- Mitry et al. [28] compared the accuracy of crowdsourced image classification with that of experts. They used 100 retinal fundus photography images selected by two experts. Each annotator was asked to classify 84 retinal images while the ability of annotators to correctly classify those images was first evaluated on 16 practice - training images. The study concluded that the performance of naive individuals to retinal image classifications was comparable to that of experts. Giuffrida et al. [15] measured the inconsistency among experienced and non-experienced users in that task of leaf counts in images of

Arabidopsis Thaliana. According to their results everyday people can provide accurate leaf counts.

Disadvantages:

- There is no Hash Tag based technique to implement HITS Algorithm.
- An image retrieval is very slow due to absent of HITS Algorithm

3.2 PROPOSED SYSTEM:

In the proposed system, the system presents a novel methodology, based on the principles of collective intelligence that helps locating those hash tags. In particular, we show that the application of a modified version of the well-known HITS algorithm, in a crowd tagging context, provides an effective and consistent way for finding pairs of Instagram images and hash tags, that lead to representative and noise-free training sets for content-based image retrieval.

4.1 Advantages:

Image tagging and image retrieval is fast due to bi-layer clustering framework to locate relevant tags to social images.

The system is more effective due to HITS algorithm for hash tags through crowd tagging.

IMPLEMENTATION

4.1 MODULES

Server: In this module, the admin has to login by using valid user name and password. After login successful he can perform some operations, such as View All Users, All Friends Status, View All Images, View Recommended Images, View Image Reviews, View Dislikes, and View Image HITS Results.

Friend Request & Response:

In this module, the admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user photo, requested user name, user name request to, status and time & date. If the user accepts the request, then the status will be changed to accept or else the status will remains as waiting.

User: In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password and can perform some operations like View Profile, Search Friends, View Friend Requests, Add Image, Search Images, View My Images, View Friends Images, View Image Recommends, View Friends or Others Reviews on Images.

Searching Users to make friends:

In this module, the user searches for users in Same Network and in the Networks and sends friend requests to them. The user can search for users in other Networks to make friends only if they have permission.

SCREENSHOTS

5.2.1 HOME PAGE

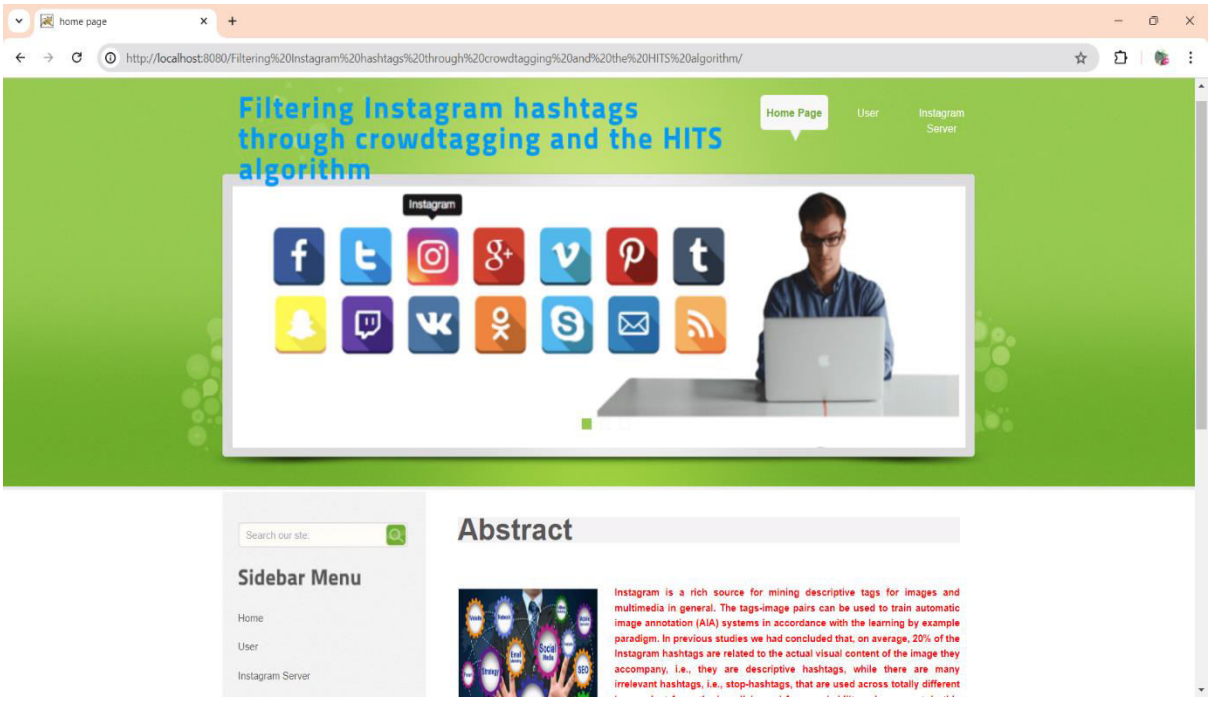

FIG 5.1 Home Page with Project Concept – SCREEN SHOT

5.2.2 Instagram Server Page

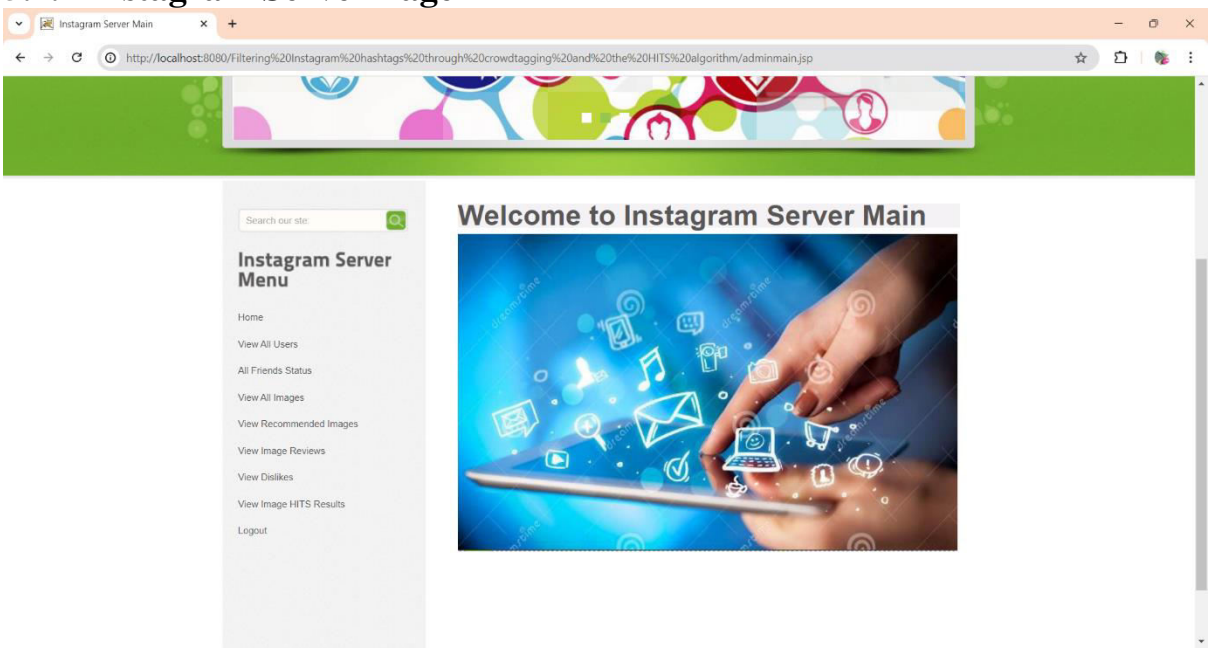

FIG 5.2 Instagram Server – SCREEN SHOT

5.2.3 User Registration Form

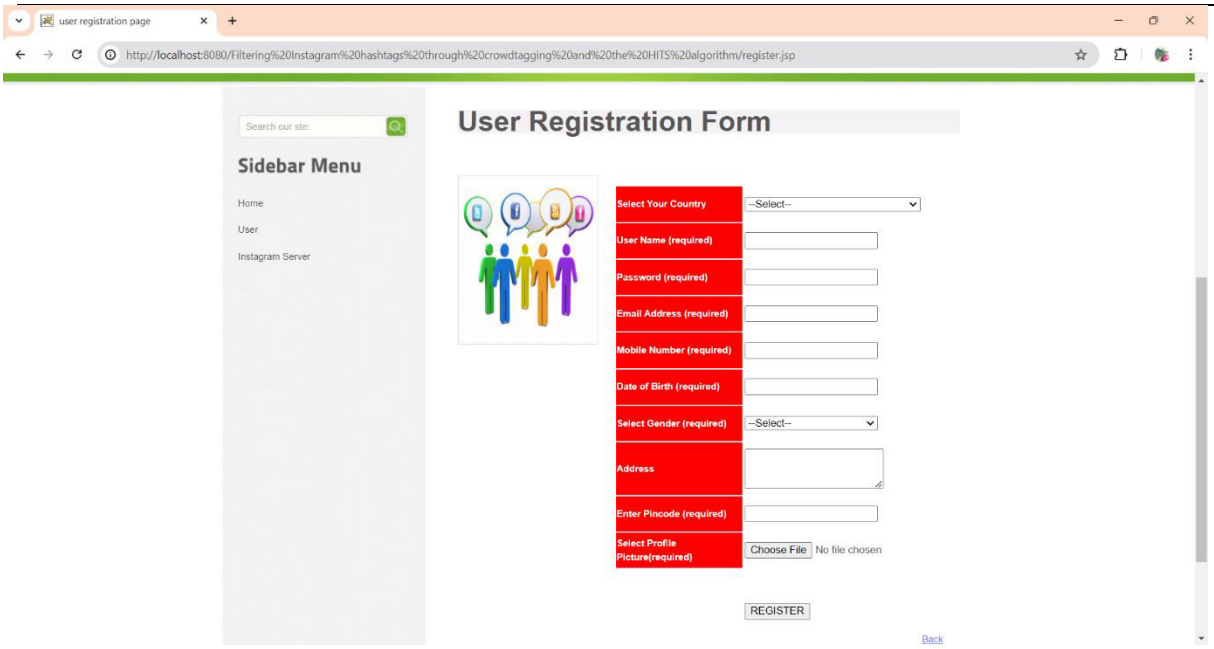

FIG 5.3 Registration page – SCREEN SHOT

5.2.4 User Login Page

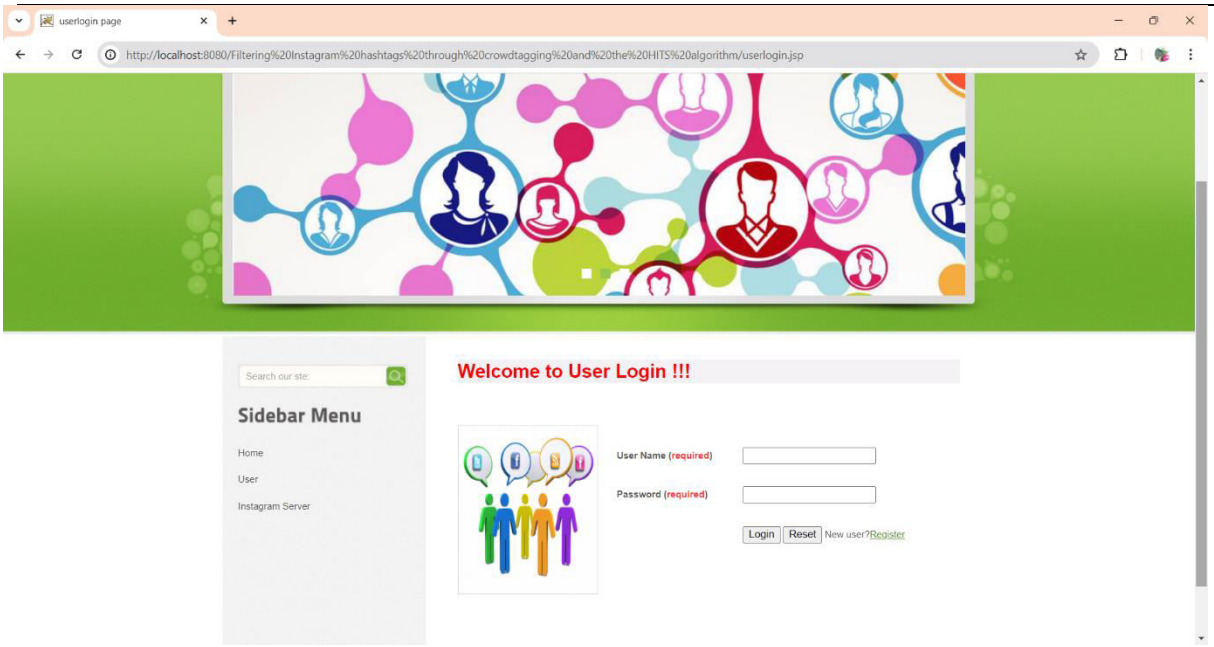

FIG 5.4 User Login – SCREEN SHOT

5.2.5 View All Images Reviews screen

FIG 5.5 All Images Reviews – SCREEN SHOT

5.2.6 Disliked images screen

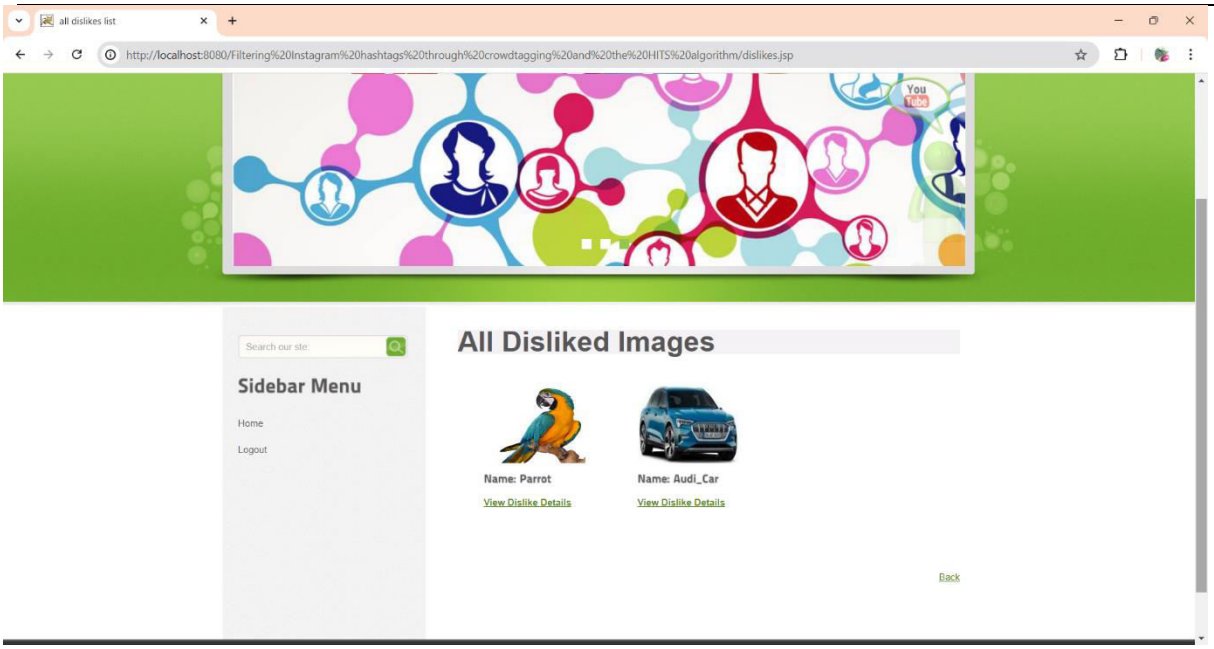

FIG 5.6 Disliked Images – SCREEN SHOT

6. CONCLUSION AND FUTURE WORK

CONCLUSION

In the current work, we have presented an innovative methodology, based on the HITS algorithm and the principles of collective intelligence, for the identification of Instagram hashtags that describe the visual content of the images they are associated with. We have empirically shown that the application of a two-step HITS algorithm in a crowd tagging context provides an easy and effective way to locate pairs of Instagram images and hashtags that can be used as training sets for content-based image retrieval systems in the learning by example paradigm. As a proof of concept, we have used 25000 evaluations (500 annotations for each one of 50 images) collected from the Figure-eight crowdsourcing platform to create a bipartite graph composed of users (annotators) and the tags they selected to describe the 50 images. The hub scores of the HITS algorithm applied on this graph, called hereby full bipartite graph, give us a measure of reliability of the annotators. The aforementioned approach is based on the findings of Theodosiou et al. [39] who claim that the reliability of annotators better approximated if we consider all the annotations, they have performed rather than the subset of Gold Test Questions. In a second step a weighted bipartite graph for each image is composed in the same way as the full bipartite graph. The weights of these graphs are the hub scores computed in the previous step. By thresholding the authority scores of the per image graphs, obtained by the application of the HITS algorithm on the weighted graphs, we can rank and then effectively locate the hashtags that are relevant to their visual content as per the annotator's evaluation. Some important findings of the current work are briefly summarized here. The first refers to the value of crowd tagging itself.

7. REFRENCES

[1] A. Argyrou, S. Giannoulakis and N. Tsapatsoulis,"Topic modelling on Instagram hashtags: An alternative way to Automatic Image Annotation?" in Proc. 13th International Workshop on Semantic and Social Media Adaptation and Personalization,2018, pp. 61-67.

[2] B. I. Aydin, Y. S. Yilmaz, Y. Li, Q. Li, J. Gao, and M. Demirbas,"Crowdsourcing for multiple-choice ques-tion answering" in Proc. 28th. AAAI Conference on Artificial Intelligence, 2014, pp. 2946–2953.

[3] C. D. Cabrall, Z. Lu, M. Kyriakidis, L. Manca, C. Dijksterhuis, R. Happee, and J. de Winter, "Validity and reliability of naturalistic driving scene categorization judgments from crowd sourcing," Accident Analysis & Prevention, vol. 114, pp. 25–33, 2018.

[4] Q. Cheng, Q. Zhang, P. Fu, C. Tu, and S. Li, "A survey and analysis on automatic image annotation," Pattern Recognition, vol. 79, pp. 242–259, 2018.

[5] N. Craswell, "Mean Reciprocal Rank," in Encyclopedia of Database Systems, London : Springer, 2009, pp. 1703-1703.

[6] H. Cui, Q. Li, H. Li, and Z. Yan, "Healthcare fraud detection based on trustworthiness of doctors," in Proc. Trustcom/Big-DataSE/I SPA, IEEE, 2016, pp. 74–81.

[7] A. R. Daer, R. Hoffman, and S. Goodman, "Rhetorical functions of hashtag forms across social media applications," in Proc. 32nd ACM Int. Conf. on the Design of Communication CD-ROM,

[8] E. Ferrara, R. Interdonato, and A. Tagarelli, "Online popularity and topical interests through the lens of instagram," in Proc. 25th ACM Conf. on Hypertext and Social Media, ACM, 2014, pp. 24–34.

[9] J. M. Fletcher and T. Wennekers, "From structure to activity: Using centrality measures to predict neuronal activity," InternationalJournal of Neural Systems, vol. 28, no. 02, p. 1750013, 2018.

[10] M. Gao, L. Chen, B. Li, Y. Li,W. Liu, and Y.-c. Xu, "Projectionbased link prediction in a bipartite network," Information Sciences, vol. 376, pp. 158–171, 2017.

[11] S. I. Gass and C. M. Harris, "Bipartite Graph," in Encyclopedia of operations research and management science, Boston: Springer, 2013, pp. 126.

[12] 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, 2016.

[13] S. Giannoulakis and N. Tsapatsoulis, "Defining and identifying stop hashtags in instagram," in Proc. INNS Conference on Big Data, Springer, 2016, pp. 304–313.