

## Information Retrieval from Social media using Hashtags: Survey

Krutarth Anand Bhatt<sup>1</sup>, Harsh Nayan Bhatt<sup>2</sup>, Rutvij Bharatkumar Darji<sup>3</sup>, Shivani Kishorsinh Rana<sup>4</sup>, Dharmistha R. Chaudhari<sup>5</sup>

<sup>1,2,3,4,5</sup> Author Computer Science & Engineering R. N. G. Patel Institute of Technology Bardoli, India

\*\*\*

**Abstract** - Social media since its inception has been a prevalent way for us to communicate, but due to overabundance of information & the plethora of social platforms available to the end user, it has caused some serious social media fatigue among the people. Users can search for information regarding particular topics such as research, interview, food, fashion or travel. The total amount of information retrieval tools has increased exponentially and as a result people can use either one which is relevant to their needs. The amount of information accessible via the Internet and the amount of topics covered have grown by many folds. In this paper, we focus on different platforms and methods which can effectively and efficiently filter out the unnecessary post from users feed/search result. This would also allow the users to de-clutter their feed by filtering the posts to just the couple of tags that they are interested in, or even search for everything about certain event on all the platforms at once.

**Key Words:** Social media, Hashtags, Informational retrieval

### I. Introduction

The rapid advancement and growth of internet made it more accessible for the general public to access, retrieve and obtain information. The internet can be excellent as it stores a lot of information, but it is also very important for people to know the right method in retrieval the right information [1]. An increase in number of many more complex tools have been developed in order to make the process of information retrieval from the Internet much simpler and efficient, due to the enormous volume of information available on the internet, there are a few disadvantages that we need to be aware of, such as information overload and the problem in finding relevant data. By providing people with a better way of sorting through information will reduce the chances of data overabundance.

As a brief terminological summary of Twitter; a hashtag is a metatag that starts with # sign and being used in social media in order to write about, and easily search of a topic. A mention is a tweet which contains username, and it is used to refer to a user in Twitter. Moreover, Twitter does not require reciprocal relationships between users. Hence, users follow news media accounts, celebrities, and

other people who they want to read about, without being followed by them [2].

### History of Hashtag

The symbol “#” could be trace back to early 17<sup>th</sup> century but the recent usage of the hashtag is much more reminiscent of twitter’s implementation in 2007.

Chris Messina was a Google employee who was an early adopter of twitter while the platform was still in its infancy, Chris was also an avid user of flickr where he noticed that the platform used keywords and tags in order to group and categorize posts, so he thought of a similar approach that could be implemented in twitter to make topical interactions and intergroup linking much more easier, Chris pitched this idea to twitter as he thought hashtag has staying power as it allowed the use of hyperlinks.

The hashtag feature implemented by twitter blew up, since then all the major social media platforms uses hashtags in some form or other in order to categorize the data in particular groups. As we can see nowadays almost all social media platforms offers this facility to increase the reach/interaction of a particular post, and use of hashtag has only increased as time passes.

### Hashtag on Social media

Hashtags are a crucial part of social media marketing. It is a great way of labelling and finding social media updates [3]. Hashtags can boost impressions, improve the search ability of your content, and increase reach of your content to the peoples [4]. The hashtag allows you to connect with and engage other social media users based on a common theme or interest. Indeed, hashtags help you to increase traffic on Facebook, Instagram and Twitter.

Hashtags can increase awareness, getting your content seen by more people than just followers and increase your social reach. By simply incorporating hashtags into your social media content, you can expand your reach and increase your audience [3].

The user can find information easily with the existence of hashtag. A hashtag can be classified depending on the intention of the user to use the hashtag. The hashtag is used when the users want to express their comment, praise or criticize when to rise the event or people and to promote brands and also to provide updates on breaking news item and bring them to the content that easily findable and searchable across the social web.

As per our survey we found that displaying the relevant hashtags on the screen during a program increases the number of tweets related to the particular program, and therefore bringing in new viewers and increasing the overall reach and engagement of the program, this is called “Word of mouth”,

this process brings in audience free of cost with the help of social media market.

A hashtag can be divided into two categories which are Topical hashtag and Non-topical hashtag [1], the topical hashtag is where the hashtag name is built for long-standing term and represent of the name of event or occasion. For example, the government launches a program or initiative to stop smoking, so they will build a hashtag like #StopSmoking or #SmokingNotGood and along the program, the government will use the hashtag for long period [2], while the non-topical hashtag built for only for temporary period and nobody bother with the hashtag since the hashtag built for emotive marker. For example, he felt angry and started to update the status with hashtag #FeelingAngry; the hashtag will cease only at this stage.

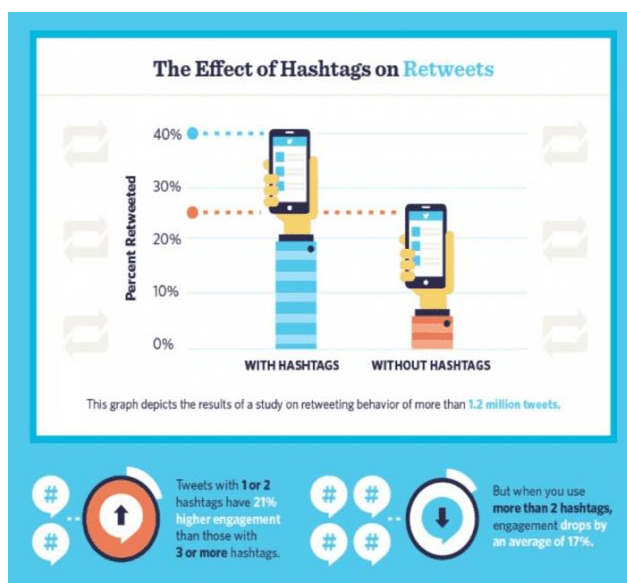


Fig.1: The Effect of hashtags on retweets [5]

Figure 1 is showing the effect of the hashtag on retweet. It is clearly showing that if user uses hashtags related to their tweets then it increases the reach of the tweet by 15% in compare to tweet without hashtags.

## II. Existing Platforms

### 1. Tagsfinder

TagsFinder [6] generates related, similar or combined hashtags based on your entered keywords. You have 3 options to find the right hashtags you need. You can easily copy the generated hashtags with the copying button. In related hashtag it generate top hashtags which are often used on your keyword and it can be on different topics and also popular hashtags, Generate similar hashtags based on your keyword.

### 2. Top-hashtag

Top-hashtag [7] provide with the top 100 hashtags on a particular platform like Instagram and Tiktok in order to provide better tools for increasing the reachability of a

particular post and in turn increase traffic on the user's profile. You can convert a group of text to hashtags easily and view the stats of each hashtag.

### 3. Ritetag

Ritetag [8] provide instant hashtag suggestions for images and texts on desktop and mobile, based on real time hashtag engagement. This would make the process of fetching relatable hashtags very easy and efficient. It is also providing hashtag suggestion related to its content.

### 4. All-hashtag

All Hashtag [9] is a website that will help you to create and analyse fast and easy top relevant hashtags for your social media content and marketing. You can generate thousands of relevant hashtags that you simply copy and paste into your social media posts.

It is providing some facility like Generator, Creator, Analytics and Charts. This help to generate different hashtag related to searched hashtag, it is suggesting similar hashtag and also showing analysis of specific hashtag and charts of hashtags. But main disadvantage of this platform is, it is not providing posts from social media which user is seeking for.

### 5. Hashtagify

Hashtagify [10] is one kind of analysis tool which helps you to find hashtag's popularity, recent popularity, month trend, week trend and also providing full analysis of searched hashtag. It is also providing that in which country that hashtag is in trending and also showing twitter posts of searched hashtag. User can also search for instagram posts but, this feature is locked behind a paywall.

It enhances the reach of tweets by providing the related hashtag suggestions, it can track any hashtag on Twitter and it will give in depth analysis of that particular hashtag user can also monitor relevant content.

### 6. Hashtag

This tool helps the user to track and analyse any particular hashtag of their choice, it also recommends all the popular/trending hashtags in your geographical location. Although some of its features are lock behind a paywall, the free tire is also very functional and useful. It also enables user to sort the hashtags alphabetically which makes it really easy and efficient to go through all the required hashtags [11].

Features	TagsFinder [6]	Top-hashtag [7]	Ritetag [8]	All-hashtag [9]	Hashtagify [10]	Hashtag [11]
Hashtag Suggestions	✓	✓	✓	✓	✓	✓
Hashtag Analysis	✗	✗	✗	✓	✓	✓
Retrieving data	✗	✗	✗	✗	✓	✓
Unification of data from multiple platforms	✗	✗	✗	✗	✗	✗

Table 1: Comparison of Various Platforms

### I. Comparison analysis

We have compared various existing platforms (TagsFinder, Top-hashtag, Ritetag, All-hashtag, Hashtagify, Hashtag) in order to gauge their features and functionalities. We have performed the comparison using standard feature such as Hashtag suggestions, Hashtag Analysis, Retrieving post, Unification of post from multiple platforms. After performing this comparison, we can conclude that, as shown in the comparison table some of the platform provides features like post retrieval and others provides features like hashtag analysis, the hashtag suggestion feature is common in each platform, but not a single platform provides unification of data from multiple platforms.

#### Algorithm: Data Collection Algorithm

Step I: Extraction and Storage of data (content) from different social networks.

Step II: Aggregation of the extracted data into a single file.

Step III: Classification of data into different classes based on their language.

Step IV: Translation of classified data of different languages into standard language (English).

Step V: Aggregation of standard language contents into

Fig.2: Algorithm for Data Collection from MSN [12]

#### Algorithm: Information Retrieval from Database

Step I: Design interface where user input query.

Step II: Apply document similarity models.

Step III: Apply ranking algorithms to rank documents.

Step IV: Ranked result list.

Step V: Result Evaluation.

Fig.3: Information Retrieval from Multiple Social network [12]

### III. Methodology

We have studied different aspects of social network based information. Content based information (textual content) provides valuable input for information organization and its utilization. Social network based contents are multilingual (people prefers their native language to share information), ambiguous (short nature post may create multiple context of same content), fragmented (scattered on different platform based content, some are structured other are unstructured) into different format. To handle these challenges, we outline a procedure for content based information retrieval from multiple social networks. The procedure for Content based Information retrieval from multiple social networks is given below. This procedure consists of two phases as outlined by the algorithms presented in the Fig 3 and 4 [12].

- Collection of user's data from multiple social networks and then form a database
- Information Retrieval from the database built in the first phase.

In the first phase, extract data from social networking services: Facebook, Instagram and Twitter using their respective application program interfaces (APIs). Put the extracted data into either text based file or document based file format for further processing. In next step, merge the contents of different files into a single file to study the varieties in the contents (languages). Then classify similar types of contents using supervised learning method initially. After classification of similar contents of the single file discussed above, put homogeneous (based on language) information in separate files for further translation. Use English as standard language in which different language contents would be translated. During translation process, all standard language contents are placed into unified database. This unified database consists of multiple social networks based information [12].

Second phase basically describes stepwise information retrieval process. In this phase, an interface for writing query

is designed that describes users need. Then a document similarity models are used to measure the similarity among documents and query. Most commonly used models are vector space model [13], statistical language model [14] and Boolean model [15]. Similarity measure (cosine similarity measure) used to calculate the relevant score among documents and query. The documents can be put in order of their decreasing relevant scores. To evaluate the results of the system, evaluation metrics [15], like precision and recall can be used.

#### IV. Conclusion

In this paper, we addressed different methodologies through which the various social media platforms use hashtag in order to categorise data/post, hashtags are mainly used to categorise post into various subsections. We have mentioned Plethora of tools and services used by people in order to analyse specific hashtag or information retrieval about particular hashtag. We performed a survey of various types of currently used social networks and also the types of information shared by users on that platform and categorized content based information into textual content based and visual content based information. Specifically, we outlined a procedure for content based information extraction, management and retrieval.

#### V. References

- [1] Farleen Azrina Zamberi, Nor Hafifah Hassan Adli, Norhayati Hussin, Masitah Ahmad, "Information Retrieval via Social Media" in International journal of Academic Research in Business & Social Sciences by hrmars ISSN: 2222-6990
- [2] Zeynep Zengin, "Extracting Topical Information of Tweets Using Hashtags" Alp Institute of Science and Technology Istanbul Technical University Maslak, Istanbul, Turkey  
Sule Gunduz oguducu Department of Computer Engineering Istanbul Technical University Maslak, Istanbul, Turkey 34469 in 2015 IEEE 14th International Conference on Machine Learning and Applications
- [3] Jimit Bagadiya, "How To Use Hashtags Effectively In Social media marketing?"
- [4] [Katie Lundi](#), "How Hashtags Can Help You Improve Your Small Business Marketing"
- [5] How, when and why you should be using hashtags  
<https://www.smartinsights.com/social-media-marketing/how-when-and-why-you-should-be-using-hashtags/>
- [6] TagsFinder (Website)  
<https://www.tagsfinder.com/en-us/>
- [7] Top-hashtag  
<https://top-hashtags.com/instagram/>
- [8] Ritetag  
<https://ritetag.com/>
- [9] All-hashtag (Website)  
<https://www.all-hashtag.com/index.php>
- [10] Hashtagify (Website)  
<https://hashtagify.me/hashtag/covid19>
- [11] Hashtag  
<https://www.hashtags.org/>
- [12] Waseem Ahmad1 and Rashid Ali2, "Information Retrieval from Social Networks: A Survey" Department of Computer Engineering AMU, Aligarh, India
- [13] G. Salton, A. Wong and C. S. Yang, "A vector space model for automatic indexing," communication of the ACM, Vol. 18, Issue 11, pp. 613-620, 1975.
- [14] Ponte, M. Jay and w. B. Croft, "A language modeling approach to information retrieval," In Proceedings of the 21st Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 275-281, 1998
- [15] B. Liu, "Web Data Mining," Springer, Berlin Heidelberg, New York, 2007.