CHAPTER 5

EFFICIENT TAG RECOMMENDATION

In this chapter, the research work has discussed how the constructed topic ontology plays a main role in efficient tag recommendation using Spreading Activation algorithm. In this recommendation approach, tags are recommended to the user are predicted from the extracted keywords from the existing blogs and the topics in constructed topic ontology. The activation score of the tags are determined using spreading activation algorithm depending upon the occurrence of the terms as topics in ontology. After computing the activation score for each extracted tags from the blogs, the extracted tags with high activation score are recommended to the user as a more relevant tags. Additionally, this proposed recommendation is useful in various real time applications such as tag popularity, sentiment analysis and spam reduction. It is clearly explained in this section.

5.1 TAG GENERATION MODELS

Tag generation models are mainly evolved to analyze and understand the tags based on the content of the web resources. It is also used to obtain the knowledge about the methods to generate tag recommendation and also provides the background knowledge about the content of web resources (Lipczak and Milios 2010). The current user has awareness about the tags suggested by the previous users. Basically, selection of tags is mainly influenced by the tags which are previously assigned by the user. Already
suggested tags are taken as main criteria to generate tags. Therefore only less amount of effort is needed to generate tags. Information foraging (Pirolli 2005) is a new technique that is developed to adapt the suggested tags according to the user behaviour. It is also used to achieve the faster retrieval of the information from web resources.

5.1.1 Polya Urn Generation Model

This model can express the variation of user constant prototype according to the time where already predetermined collective tags are applied for the resources. Tags are generated from the previously suggested tags from any other users in network. Therefore this approach could not generate new different tags. It is not practical to add generated tags into systems because it is supposed to be same as that of already suggested tags. This stochastic model is described using the experiment where an urn consists of two balls with two different colors (e.g. say a green and a blue). First, a ball of particular color is taken from the urn and put in to another urn which has ball as same color as first selected ball. This can be repeated for various numbers of steps until the fraction of balls in each urn will be stabilized into random limits. However, for each execution of the experiment, this fraction of balls congregates to random a limit which provides the different predictable outcome (Gupta et al 2010). This experiment resembles that tags which are generated to the web resources that are similar to the previously assigned one. Therefore, this model cannot add the newly selected tags into the system.

5.1.2 Yule-Simon Generation Model

In order to overcome the problem in poly urn generation model, Yule-Simon generation Model is proposed. In this approach, new tags are created for each step and it is combined to the tagging system with less
probability of $p$. Therefore, large number of different tags is raised to describe the web resources and also it can be dynamically vary over the time. But, the number of tags generated in each step will be declined. This model is described as follows: newly generated tags are added to the group of tags with probability $r$. The main important thing is that the newly added tag is not supposed to occur anywhere in the system. The tags are generated using the probability of $p$ from the tags that are already requested. It can follow the frequency rank distribution with a power law tail whose exponent is given by $\alpha = 1 - r$. The number of different tags ($N$) generated into the system is directly proportional to the number of newly generated tags over the time ($T$) (Gupta et al 2010). The generation rate of new tags decreases and finally, reaches zero. This type of sub linear growth is represented using Heap’s law.

5.1.3 Yulesimon Model With Long Term Memory

This model is also same as that of aforementioned Simon model but applicable with long term memory. In that approach, tags are generated using the copy of previously assigned tags with the same probability $r$. But, this model can generate the duplicate version of previously assigned tags using long term memory. Their model is described as follows: users of a collaborative tagging system can assign tags to resources where text is constructed based on $n$ number of words. The context of user is efficiently described using specified model that is used to identify the behavior of the user. If any new tags are generated to the system with the probability $p$, after the time step $t$, at the same time the one word from already assigned tags are copied into the newly generated tags with probability $1-p$. It can be performed for $n$ number of steps. The probability decays in the form of power law which is expressed using power law distribution function as $P(x) = n(t)/(x + t)$ where $n(t)$ is a normalization factor and $t$ is the time-scale factor over probabilities. This approach can produce appreciable characteristic value
along the slope in frequency-rank distribution of co-occurrence tag streams. But, it is failed to explain the distributive nature of resources among the tag streams and does not provide any information about the decay rate of the number of different tags (Cattuto et al 2007).

5.2 TYPES OF TAG RECOMMENDATION SYSTEMS

There are four divisional categories of tag recommendation systems such as content, hybrid, collaborative and Flickr based tag recommendation systems. Content based techniques exclusively depend on textual metadata which is associated with the resource. Hybrid systems are the integration of two input methods such as content based systems and collaborative based system. Collaborative based recommendation system mostly deals with the collaborative filtering method (Melville and Sindhwani 1997).

5.2.1 Content-Based Tag Recommendation Systems

In this content based tag recommendation system (Cantador et al 2010), tags are extracted from the content of the blogs using artificial neural network. This type of neural network is trained using statistical data about the word frequencies and semantic relationships are extracted from the lexical relation of WordNet. This type of tag recommendation is dealt in Gaussian framework. In this approach, multi label classifier is trained according to the content of web resources. It can extract the title and short description from the web resources. Therefore, the class which specifies the topic is called tags. Tags derived from the various topics are combined together that performs the final recommendation task. The main disadvantages of this recommendation is already available tags are reused. Therefore, uniqueness of resources is limited and hence, it cannot be used to recommend large number of unlimited resources (Pazzani and Billsus 2007).
5.2.2 Collaborative-Based Tag Recommendation Systems

In Collaborative based recommendation system, most popular tags are recommended based on the social (group of users) interests, knowledge and goals within a community. It is social based recommendation approach as the tags are recommended by considering the opinion of the social people on web resources (Marinho and Thieme 2005). In this approach, user profile is constructed to each user according to their interest and similarities between the users are also computed. It also deploys Singular Value Decomposition (SVD) technique to derive the relation between the user, web resources and their corresponding tags. The certain probability value is assigned according to the relationship between user and their accessing resources. As a result, most probable tags are returned as per the user requirement. This type of recommendation systems are categorized into two systems such as:

- **Model-based approaches**

  In this approach, the probabilistic model is constructed to predict the similarity between the users based on the user browsing history and their future rating assignments.

- **Memory-based approaches**

  In this approach, statistical techniques are used to identify the users (neighbor) with similar behavior. After the identification of neighbor from this approach, list of tags that is to be recommended is generated from the user feedbacks (Dattolo et al 2010).
5.2.3 Hybrid Tag Recommendation Systems

Hybrid tag recommendation system is the combination of both content based and collaborative based recommendation approach (Burke 2007) and (Burke 2002). It takes only the strength from the two approaches and also overcomes its limitations. It is a most practical approach and also allows processing the wide variety of posts more efficiently. In this approach, tags are extracted from the web resources and also the profiles maintained by the user. Extracted tags are extended through NLP techniques. Finally, extended tags are combined with the tags that are extracted from content based approach. The similarity score is computed between the content of tags and it is the measure of content of presently posted tags. Tag recommendation from the file is constructed using the title associated with the content of the resources. Scalability of search engine and usage of source tags is mainly appreciable. There are four strategies available in this recommendation process such as

- **Weighted Hybrid Recommender System**

  In this system, a tag with high score is recommended to the blog users. The score for the resource is computed from the weighted sum of scores calculated by different recommender systems.

- **Cascade Hybrid Recommender System**

  In this system, filters, a starting set of resources use a first recommender system. The ranking of these resources is then refined by a second recommender system.
• **Switching Hybrid Recommender System**

  According to the some specific criterion, suitable recommender system is used to recommend the tags to the users.

• **Feature combination hybrid recommender system**

  In this system, both collaborative and content based approach is combined together to perform recommendation that considers features of collaborative as a data and then context based recommendation is performed (Dattolo et al 2010).

5.2.4 **Flickr Tag Recommendation Systems**

  Tagging is defined as the action of assigning keywords to provide useful description about the objects. It is mainly used to annotate the web resources such as academic journals, bookmarks and multimedia objects etc. Tagging allows users to manage a large amount of information with proper indexing. Flicker tag recommendation is used to assign tags to multimedia objects (Sigurbjornsson and van Zwol 2008). But it is so hard to assign tags due to the absence of textual content in multimedia objects. Flicker is generally defined as the community network that allows sharing photos with friends, family members etc within the network. Different users can post large number of photos under the same category or subject. Therefore, various tags are applicable to describe the photos based on their perspective and context. It can describe the content of photo using manual annotations that provide the semantic information and additional context. The fundamental thing for this recommendation relies on the characteristics of flicker tag. Users can describe the photo according to their own context and perspective. Therefore, many users provide different description for the same photo. This type of extended
recommendation approach can provide more enriched description about the uploaded multimedia objects. It also allows retrieval of photos based on the keyword queries.

5.3 PURPOSE OF RECOMMENDING TAGS

In order to overcome the aforementioned obstacles, researches have led to the development of techniques to automatically recommend the rich set of tags to the blog users. The prime purpose of recommending tags is to automatically suggest the popular tags to the users in order to direct the user towards the information they request. Recommender systems need to suggest the related tags for user untagged resources (Rae et al 2006). Tag recommendation for blogs could help users on giving the more suitable tags for a resource and improve the process of resource annotation. In social networking sites or blogs, tags act as a mediator between users and resources handling the user preferences. The main purpose for recommending tags is to get accurate user preferences via tags and exploit it for efficient recommendation.

In Collaborative tagging systems, Community services such as Flickr, Delicious and Bibsonomy are common types of folksonomy. It is popular for content management and sharing of resources among millions of users. These community services allow users to add tags on online items such as photos, videos and text. Folksonomies allow a set of freely selected text keywords as tags and these tags are imprecise, irrelevant and misleading (Jaschke et al 2007). It does not follow any formal guidelines to assist tag generation and tags are assigned to resources based on the idea of users. Hence, a tag recommender system may provide better tag suggestions in which users select tags for any specific resource.
5.4 FEATURES OF TAG RECOMMENDING SYSTEMS

The tagging process can overcome the problem in predefined categorization which does not consider user preference. Therefore, user is recently shifted to the tag recommending system in which a group of more expressive tags about the resource is combined together. The prime scope of tagging system is to intend relevant useful tags and to provide large potential solutions. Tag recommender system is the system that can recommend relevant tags for a user provided untagged resource. Tags are predicted as relevant according to the point of view of social (collaborative) experts in the corresponding domain and individual users who owns the resources. It mainly focuses on two aspects such as personalized and collaborative tag recommendation. Personalized recommendation approach helps individual users to annotate their content in order to manage and retrieve their own resources. Second, Collective tag recommendation aims at making resources more visible to other users by recommending tags that facilitate browsing. The three features of tag recommendation system are follows (Lipczak et al 2009).

5.4.1 Generality

Generality is defined as the specific own characteristics of personal or social character about the post in collaborative tagging system. Difference between the tags and their unique characteristics are mainly used to make decision about the resources and then it is also considered in order to design the tag recommendation system. If characteristics of tags manually adapt to the parameters of recommendation, they lead to various limitations. Therefore, the characteristics of tags can automatically adapt according to the system parameters due to the efficient recommendation system. Thus,
efficient learning algorithm is mainly deployed that can perform automatic parameter tuning in order to optimize the results.

5.4.2 Adaptability

Tag recommendation process is performed dynamically depending on the user preferred tags. Therefore, recommendation process also performed sequentially according to the tags instantly followed by the user tags. New valuable information also brings into the recommendation process because of the constantly varying feedback loop from the user. The recommendation can dynamically update the information of resources according to the content suggested by the user through post, user profiles and also associations exist between words and the preferred tags. This adaptability feature of tag recommending system allows us to provide online content adaptation. It improves the quality of tags extracted from a user profile and also dynamically adapt to the interest preferred by the user.

5.4.3 Efficiency

The efficiency of tag recommendation system is also highly appreciable because it can manage a large amount of information in user created repositories. It can produce more precise results even when unlimited vocabulary of tags is available. Moreover, tag recommendation system is extended from the collaborative tagging system and also can operate even with the limited resources. It can achieve high efficiency with limited resources because additional cache layer is deployed with text indexing engine (Lipczak and Milios 2011).
5.5 EFFECTIVE TAG RECOMMENDATION

In this proposed recommendation approach, more relevant tags are recommended to the blog user. It performs following recommendation tasks such as i) The most important keywords and terms in the popular blogs are extracted as tags ii) Topic ontology for the content of blog is efficiently constructed using the two external knowledge base source called Wikipedia and WordNet. iii) Spreading activation algorithm computes the activation score of each extracted tags depending upon the similarity relevance between the extracted tags and topics in topic ontology. Typically, Activation score of each extracted tags are computed. iv) Tags with high activation score are filtered out as highly activated tags. v) Tags which are more relevant to the blog are determined and set of relevant tags are separated from the highly activated tags. Finally, set of more relevant tags are recommended to the blog users. Therefore, user can assign the most popular and relevant tags to the blog. Figure 5.1 shows the flow of tag recommendation process.

![Figure 5.1 Tag Recommendation Process](image)

**Figure 5.1 Tag Recommendation Process**
5.6 TOPIC HIERARCHY

In topic ontology construction process, corpus based approach constructs the topic hierarchy for the content of blogs using two external knowledge base sources called Wikipedia and WordNet. Topic hierarchy is structured in hierarchical manner where more relevant topics are occurred in the top level categories of hierarchy. This topic hierarchy is mainly organized to identify the similar topics and semantic relationships among the huge amount of topics (Maguitman et al 2010). Therefore, topics which are more relevant to the content of the blog are identified through knowledge base system.

Identifying and classifying topics in a well-organized structure is used to obtain valuable information from these topics hierarchy. Finally, a valuable resource that helps identify the topical categories of web pages has been developed. Later on, topic ontology is used to categorize the topics in which user is interested (Stamou and Ntoulas 2009). This is potential in creation of topic ontology because it maintains reliability and scalability. Finally, underlying topic based ontology is generated successfully.

5.7 KEYWORD EXTRACTION FROM EXISTING BLOGS

From the existing blogs, keywords which are relevant to the core concept of the blog are extracted as tags. A large number of keywords are extracted as tags but it is not more relevant to the content of the existing blog. Because, keyword based matching is not appreciable in the case large number of blogs that describe the same topic. Therefore, it is practically hard to extract suitable tags for the existing blogs. Therefore, keywords are extracted from the most popular blogs which are more related to the existing blogs. Popular blogs are blogs that are accessed from any other blogs using
hyperlinks and also blogs are retrieved using search engines. Popular blogs may have interesting content which is required by the large number of users.

Blogs are effectively searched through blog popularity. It is evaluated through RSS feed as a positive pointer of blog quality. Other positive pointers comprise click through ratio, blog-rolls, tagging and Page rank. Popularity ranking approach employs several social interrelations among bloggers. Blogosphere provides a rich collection of social facets relevant to user expressions, blog contents, interests, purposes and interaction with others. In order to express the content of blogs, user assigns tags, blog posts, comments and bookmarks etc. Furthermore, blog popularity is mainly computed from the bookmarks that replicate the individual user interest on particular blogs. All practical blog search engines exploit blog popularity as a major criterion to perform blog search (Goncalves et al 2010).

5.8 SPREADING ACTIVATION ALGORITHM

Spreading activation is introduced to elucidate the semantic model based suggestions in which activation paths are stored and utilized to lead a good recommendation (Hussein and Neuhaus 2010). Later on, these paths may be exploited to produce both verbal details and relevance feedback forms. Recommender systems are the most thriving ways to help user's selection process since recommender system conceals unrelated information and provides only user likings. Recommender systems consequently provide tailored information in which user is interested. Therefore spreading activation technique is used to identify the user interested topics. Spreading activation runs and spreads activation power to all interrelated concepts and items. Weight is assigned more and more to the user interests in iteration. Branch and bound approach describes a spreading activation run. Network should start prior to the commencement of spreading activation execution.
Link weight is assigned based on user background and activation values are set for each node in a network. Activation begins with certain value which is received by initial nodes and the level of activation is computed by assigning activation. Primary nodes are added in a priority queue arranged with downward activation.

Once initialization is performed, highest weighted node is removed from priority queue and node activation is spread to all nearest nodes. These nearest nodes are also added into the queue if they are not labeled as processed. Finally, node which spreads its activation to the nearest nodes is labeled as processed. After getting suggestions, the user can have more choices to treat with it so as to discover it. Spreading is triggered to all items and displayed in final suggestion list. Activation paths are examined and initial node paths are shown to be more accurate. In order to avoid spreading activation streams from items, items and their associations are set on the stop list (Hussein and Neuhaus 2010).

5.9 COMPUTATION OF ACTIVATION SCORE

Tag recommender systems recommend tags to the resources based on possible ratings. The user interested or weighted tag hierarchy, based on propagation of highest activation scores and categorizes the blog contents into the best matching tags according to pre determined topic ontology, has been developed. Therefore, the set of tags or keywords have been extracted from existing blogs and apply spreading activation algorithm to activate scores on the tags leading to an effective recommendations (Sieg et al 2007). The dataset from Folksonomy systems are collected. Folksonomy contains tags, resources and users where keywords are used to annotate the resources are also referred to as tags. Tags are collected for resources and activated with interest scores in a folksonomy. Folksonomy supports various resources.
Publications and bookmarks are shared in Bibsonomy. The weight of the tags is computed from its activation score.

**Figure 5.2 Spreading Activation Algorithm**

```
Input : Interest score and tags of existing blogs
Output : Recommendations with updated tags
Tags = (T1, ..........., Tn), tags with interest scores
Interest score (Ti),
Interest score (Ti) = 1, no interested tags
I = (b1, ..........., bn), blogs
for each bi ∈ I do
  Initialize queue;
  for each Ti ∈ Tags do
    Ti. Activation = 0;
    end
    for each bi ∈ I do
      Compute sim (bi, Ti);
      if sim (bi, Ti) > 0 then
        Ti. Activation = Interest score (Ti) * sim (bi, Ti);
        Queue. Add (Ti);
      else
        Ti. Activation = 0;
      end
    end
    While Queue. Score > 0 do
      Order queue // Activation values
      (Ascending)
      Ts = Queue [0]
      if pass limitations (Ts) then
        Relevant Tags = Get Relevant Tags (Ts);
        for each TR in Related Tags do
          TR. Activation +=
          TS. Activation * TR. Weight;
          Queue. Add (TR);
        end
      end
    end
end
```
All related tags are mapped into graph as nodes with a particular activation level and these tag relations are denoted by links. Activation scores are spread to the neighboring nodes to a particular threshold value. Based on the tags with highest activation scores, weight is computed. Relation weight $R_w$, tag $t$ and sub tags $s$ is calculated using vector $n$.

$$R_w = \frac{nt*ns}{nt*nt}$$

(5.1)

Where $R_w$ is the Relation weight,

$T$ is the tag,

$S$ is the Sub tag,

$nt$ is the number of tags and

$ns$ is the number of sub tags.

After calculating the weight, sum of tags and sub tags relations weight should be equal to 1. Tags that are activated to the queue and ordered based on the greatest activation values. Interest score for the tags in topic ontology is updated and added to its queue (Lee and Chun 2007).

5.10 SET OF RELEVANT TAGS

Tags are significant components of folksonomy for surfing and getting the resources. Tag recommendation aims at improving the user’s knowledge by presenting relevant tags to the resources and users’ requirements. Co occurrences of tags are mainly used to discover the relationship among the tags in WordNet. Frequent tag co occurrences are related with each other. Spreading activation algorithm demonstrates that how much related tags are recommended out of all the tags that returned to the users. Interested topic as a set of relevant tags is calculated based on the frequency of tags. Basic similarity of tags is calculated in order to suggest the
tags. Furthermore, Similarity calculation makes use of relational hierarchy for suggesting highly related tags (Mabroukeh and Ezeife 2011). Table 5.1 shows the most popular topics with frequent tags.

Table 5.1 Most Popular Topic with Frequent Tags

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Topic</th>
<th>Frequent Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Download</td>
<td>bittorrent, p2p, torrent, torrents, latex, anime, tex</td>
</tr>
<tr>
<td>2</td>
<td>Programming</td>
<td>Microsoft, C++, C, dotnet, c#, JAVA</td>
</tr>
<tr>
<td>3</td>
<td>Social Network</td>
<td>social, community, collaboration, networks, network, research, web, social software</td>
</tr>
<tr>
<td>4</td>
<td>Security</td>
<td>hacking, hack, wifi, hacks, firewall, wireless</td>
</tr>
<tr>
<td>5</td>
<td>Entertainment</td>
<td>Shopping, Cinema, Playing, fashion</td>
</tr>
<tr>
<td>6</td>
<td>Awards</td>
<td>Nobel, Oscar, Miss world, Academy</td>
</tr>
<tr>
<td>7</td>
<td>Sports</td>
<td>Cricket, Football, Tennis, Golf, Olympics</td>
</tr>
<tr>
<td>8</td>
<td>Design</td>
<td>webdesign, portfolio, graphic, portal, graphics, illustration, graphicdesign</td>
</tr>
<tr>
<td>9</td>
<td>Scripting</td>
<td>html, web, standards, xhtml, accessibility, w3c, webdesign</td>
</tr>
<tr>
<td>10</td>
<td>News</td>
<td>daily, magazine, imported, magazines, media, newspaper, newspapers, TV</td>
</tr>
</tbody>
</table>

5.11 RECOMMENDATIONS

Tags with highest activated scores or weights are returned as recommendations. Based on the relationship among the tags of interest for the user, recommendations are available to users or resources. Finally, the top most tags are presented to the user, who chooses the tags to attach to the post. An user can give tags to their bookmarks. This approach can make use of the most exact tags and merge for the absolute recommendation. Furthermore, bloggers utilize tags for their resource which results in greater precision and
recall of tag recommendation. Internal data from folksonomies is used for the recommendation process. In addition, each tag score is adjusted using spreading activation technique which provides more precise recommendations. By getting actual user interests wrong suggestions can also be avoided. Finally, system presents most suggested tags for the particular domain. The created tags are related to the target blog content and document as well. As a result, user is satisfied with tag recommendations (Lu et al 2009). Figure 5.3 shows the snapshot of tag recommendation of the delicious bookmarking.

![Snapshot of Tag Recommendation of the Delicious Bookmarking](image)

**Figure 5.3** Snapshot of Tag Recommendation of the Delicious Bookmarking

### 5.12 APPLICATION OF PROPOSED APPROACH

The proposed efficient tag recommendation approach deals well with the various applications such as

a) Spam reduction

b) Sentiment analysis

c) Tag popularity
5.12.1 Spam Reduction

Normally, spams are located in search engines and it adds extra content while tagging for the resource. This spam affects the performance of tag recommendation system. Delicious dataset gets affected from spam based on the links. Therefore, many companies use some spam blocking measures. Preventing messages from unknown address or unnecessary message receiving is called spam prevention. Spammers get benefits from frequently occurred tags despite of their meaning in the perspective of Folksonomy. All datasets are observed for the behavior of spamming. This research work focuses on spam relationship between tags and resources since spam damages the system, tag measure and identical resources by producing irrelevant tags or creating artificial links. The main goal of this research work is to recommend tags with reduced spam. Spreading activation score for a tag, to identify spammers from the blogs, has been used. Spam blogs are called splogs. Interest scores of tags are handled as a vector and are incremented and decremented as well. Top scored tags are recommended first then the lower score tags in a tag profile considered as spam. This can be determined as follows:

1. If interest score (Is) > \( \sigma \) (threshold value), it can be recommended to the resources.

2. If interest score (Is) < \( \sigma \) (threshold value), it can be considered as spam in the suggestion process.
Figure 5.4 Spam Reduction Algorithm

It is assumed that a set of tags $T$ is already determined and algorithm returns a set of reduced spam $R$ as output. Tag scores can be updated and maintained constantly. Tags $T$ with high scores are then recommended to the resource. Tags occurred less than the threshold value is considered as spam. If newly added tags exist, it can be securely removed with respect to the given threshold. The proposed approach provides the list of popular tags and reduces the spam at the end of the recommendation task. To avoid the automatic creation of account or annotation of tag spam, CAPTCHAs are used. Irrelevant tags for the posts are captured and reported as spam (Krause et al 2008), (Jin et al 2011), (Liu et al 2009), (Zhu et al 2011) and (Koutrika et al 2008). Existence of spam blogs (splogs) and fake blog pages include advertising or other unrelated content used to support associated sites are also detected and eliminated. Spams are most probable to individually discover the tagged content and more probable for a user to exploit it (Halpin et al 2007). This approach prevents the spammers and publishing the spam post, spam blogs and also it is adequately perfect (Attardi and Simi 2006). The reduction in spam goes a long way in enhancing the system performance.
5.12.2 Sentiment Analysis

Bloggers and blog users express their attitude, sentiments and communicate with others using blog posts in terms of comment writing or annotating each other’s blogs. Bloggers can express their sentiment via statements in determining the sentiment polarity of opinionated texts. Determining the overall opinion (Pang and Lee 2008) or attitude articulated in a fraction of text in a blog is called sentiment analysis.

Sentiment analysis is mainly used to distinguish the sentiments expressed in blogs. Many works are carried out on classifying sentiment polarity. The sentiment analysis task is a process of categorizing a blog articles (keywords) into positive (good) or negative (bad) based on the interest scores (Melville et al 2009). The sentiment analyzer is used to extract sentimental keywords from topic ontology using natural language processing systems to train and test a classifier. There are 1031 sentimental keywords that are retrieved with positive and negative orientations. Sentiment analysis focuses on classifying keywords based on their polarity (i.e.) positive or negative. Frequency of positive or negative words is computed based on the occurrences in topic ontology by applying interest scores.

1. If interest score for positive keywords (Is) > \( \sigma \) (threshold value), positive document.

2. If interest score for negative keywords (Is) < \( \sigma \) (threshold value), negative document.

3. Otherwise, neutral.
Subjective words are utilized as tags to categorize the opinionated sentences. Index is enhanced with tags for words to help in discovering opinionated blogs. Index, not only has words but also contains a cover of tags for each word (Attardi and Simi 2006). Tags are checked with WordNet in order to differentiate objective tags. Above all, the proposed approach to sentiment analysis provides two major things. Collection of particular emotional tags appears as the most important ones for obtaining the user sentiment headed for a particular resource. Each tag is activated with a score that conveys the weight of the sentiment. WordNet lexicon is exploited to express the users’ sentiment (Baldoni et al 2012).
Table 5.2 Sentiment Analysis

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affectionate</td>
<td>Positive</td>
</tr>
<tr>
<td>Happy</td>
<td>Positive</td>
</tr>
<tr>
<td>Calm</td>
<td>Positive</td>
</tr>
<tr>
<td>Unaffectionate</td>
<td>Negative</td>
</tr>
<tr>
<td>Sad</td>
<td>Negative</td>
</tr>
<tr>
<td>Anger</td>
<td>Negative</td>
</tr>
<tr>
<td>Console</td>
<td>Neutral</td>
</tr>
<tr>
<td>Interest</td>
<td>Neutral</td>
</tr>
<tr>
<td>I want</td>
<td>Positive</td>
</tr>
<tr>
<td>I don’t want</td>
<td>Negative</td>
</tr>
</tbody>
</table>

From the large number of keywords, 10 sample keywords have been selected, 4 of which are actively posting negative and rest of the keywords indicate more positive or neutral (Table 5.2). The sentiment classification accuracy is much higher when compared with other process.

5.12.3 Tag Popularity

All tags occurred in a blog post are counted collectively with the resources and most frequent tags are used for recommendations. For this purpose, Interest score is applied for each tag to find how many times it occurred in blog posts, based on the highly activated scores, tags are represented as most popular for the resource.

Popularity calculation

Spreading activation algorithm provides the top tags (top-p) with parameters for users based on tag scores. Therefore, spreading activation
algorithm maximizes the tag popularity (giving a value of 1 for this parameter). Tag popularity can also be called weight of the tags which is computed by applying interest scores on the tag occurrences. Most popular tags are secured to the most resources. Based on the occurrence of tags in a blog post, tags are used as recommendations or suggestions. Interest scores (tag weights) are automatically modified based on the occurrences of tags and activation levels in topic ontology (Jaschke et al 2007), (Sieg et al 2007) and (Durao and Dolog 2009).

Figure 5.6 Tag Popularity
After activating the interest scores to the topics, collaborative filtering is applied to filter the less scored tags in order to provide the highly activated tags as most popular for user resources (Wang et al 2008). Collaborative Filtering (CF) is one of the broadly executed approaches for recommendations. CF can be divided into tag based and user based recommendations in which the CF is used to assume that users are interested in relevant tags. Tags that are used by high quantity of people for a specific object have less probability to be a spam (Halpin et al 2007).

5.13 RESULTS AND DISCUSSION

Popular tags provide an accurate valuable retrieval results. Looking at tags in Bibsonomy and Delicious, they provide most popular tags with high interest scores. Tags are important in ranking multiple keywords and ensure that the tags are related with the blog post. From users’ perspective, use of most popular tags is fine grained for recommendation process. Delicious page provides bookmarks in reverse sequential order and tags that user assigned to the bookmarks since any one can filter the list of bookmark of anyone’s.

Individuals can view and filter any other’s personal page using tag. User can get knowledge from others popular page. The main aspect of Delicious is that user would obtain interested websites and interested people. Table 5.3 represents top 15 most popular tags in Bibsonomy and Delicious with their interest scores.
### Table 5.3  Top 15 most popular tags in BibSonomy and Delicious with their interest scores

<table>
<thead>
<tr>
<th>Tags</th>
<th>BibSonomy Interest Scores</th>
<th>Tags</th>
<th>Delicious Interest Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Art</td>
<td>1302</td>
<td>Computing</td>
<td>2640</td>
</tr>
<tr>
<td>Sports</td>
<td>2281</td>
<td>Internet</td>
<td>1484</td>
</tr>
<tr>
<td>About Me</td>
<td>1500</td>
<td>Schools</td>
<td>1098</td>
</tr>
<tr>
<td>Awards</td>
<td>1248</td>
<td>Software</td>
<td>2495</td>
</tr>
<tr>
<td>Academic</td>
<td>2253</td>
<td>Academic</td>
<td>1974</td>
</tr>
<tr>
<td>Tips</td>
<td>2326</td>
<td>Books</td>
<td>1717</td>
</tr>
<tr>
<td>Books</td>
<td>2456</td>
<td>Networking</td>
<td>118</td>
</tr>
<tr>
<td>Software</td>
<td>1974</td>
<td>Web</td>
<td>178</td>
</tr>
<tr>
<td>Sceneries</td>
<td>1328</td>
<td>Technologies</td>
<td>1258</td>
</tr>
<tr>
<td>Web</td>
<td>297</td>
<td>Music</td>
<td>1196</td>
</tr>
<tr>
<td>Schools</td>
<td>1233</td>
<td>News</td>
<td>1015</td>
</tr>
<tr>
<td>Music</td>
<td>2645</td>
<td>Awards</td>
<td>963</td>
</tr>
<tr>
<td>Photos</td>
<td>1429</td>
<td>Travels</td>
<td>1375</td>
</tr>
<tr>
<td>News</td>
<td>861</td>
<td>Interview tips</td>
<td>1212</td>
</tr>
<tr>
<td>Cosmetics</td>
<td>1258</td>
<td>Cartoons</td>
<td>1634</td>
</tr>
</tbody>
</table>
Figure 5.7 and 5.8 shows the interest scores for BibSonomy and Delicious.

Figure 5.7 Interest Scores for BibSonomy

Figure 5.8 Interest Scores for Delicious
Table 5.4 shows top 15 frequent tags with their interest score in corpus. Figure 5.9 shows the frequent tags with their interest score.

**Table 5.4 Top 15 frequent tags with their interest score in corpus**

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Tags</th>
<th>Interest Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blog</td>
<td>20,543</td>
</tr>
<tr>
<td>2</td>
<td>Software</td>
<td>19,847</td>
</tr>
<tr>
<td>3</td>
<td>Design</td>
<td>18,745</td>
</tr>
<tr>
<td>4</td>
<td>Web</td>
<td>18,621</td>
</tr>
<tr>
<td>5</td>
<td>Program</td>
<td>17,923</td>
</tr>
<tr>
<td>6</td>
<td>Music</td>
<td>17,612</td>
</tr>
<tr>
<td>7</td>
<td>Video</td>
<td>16,821</td>
</tr>
<tr>
<td>8</td>
<td>References</td>
<td>16,567</td>
</tr>
<tr>
<td>9</td>
<td>Tools</td>
<td>15,734</td>
</tr>
<tr>
<td>10</td>
<td>Web 2.0</td>
<td>15,142</td>
</tr>
<tr>
<td>11</td>
<td>News</td>
<td>14,873</td>
</tr>
<tr>
<td>12</td>
<td>Photography</td>
<td>14,564</td>
</tr>
<tr>
<td>13</td>
<td>Business</td>
<td>13,191</td>
</tr>
<tr>
<td>14</td>
<td>Tutorial</td>
<td>13,012</td>
</tr>
<tr>
<td>15</td>
<td>Free</td>
<td>12,745</td>
</tr>
</tbody>
</table>
Figure 5.9 Interest Score for frequent tags in corpus

5.14 CONCLUSION

In this chapter, the topic ontology from Wikipedia and WordNet has been constructed. Spreading activation algorithm has been applied to existing blog tags in order to activate the tags with interest scores. Weight of the tags is also computed based on the highest activation score on the tags. Every time a new tag is added and the user is interested in being handled using spreading activation, interest scores for the tags are updated. This work investigate experiments that permits the users to observe and guarantee the incremental updates to the interest scores precisely causes changes in user interest. The tag suggestion benefit from the better performance of the most popular tags on datasets and it is efficient in reducing the quantity of spam. Finally, the proposed approach also dealt with applications such as spam reduction, sentiment analysis and tag popularity.