ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

Automatic Extraction of Policy Networks Using Snippets and Social Networks

Vidhya R[1], Ajitha A[2]
M.E Computer Science Engineering[1], Assistant Professor/ Computer Science Engineering[2]
Mount Zion College of Engineering and Technology, Pudukottai[1, 2]

Abstract: Policy networks used to make a partnership among the different entities from various level of governance. The analysis of policy networks demands a series of difficult and time-consuming manual steps including interviews and questionnaires. From collected snippet, we calculate the relation strength between the actors in the policy networks using metrics. Metrics include webpage counts, out links, and lexical information extracted from web documents or web snippets. The existing approach was automatic and did not require any external knowledge source. The existing system extracted the irrelevant documents that affect the performance and accuracy. The proposed system involves the process of investigating machine learning algorithms for selecting the most informative metrics like recovery degree, inlink, broken link, anchor text type and kl-divergence and filtering web data based on relevance and type of source, investigating the applicability of proposed metrics for social networks.

Keywords: Policy networks, relatedness metrics, Spatial proximity, Social Proximity

I. INTRODUCTION

In recent year, Network is the most familiar word in the entire scientific disciplines includes the political science. Like Microbiologists describe cells as information networks, ecologists conceptualize the living environment as network system, computer scientists develop neuronal networks with self-organising and self-learning capacities.

However the use of network varies depends upon the discipline. They all share a common understanding. But, Policy networks are a set of staying steady relationships which are not including a system where things or people are separated into levels of importance and there depend each other in the nature linking a variety of actors. Actors share the common interest with regard to a policy and there exchange resources to follow these shared interests acknowledging for co-operation. Policy networks is a mere metaphor to denote the fact that policy-making involves a large number and wide variety of actors, others acknowledge them to valuable analytical tool to analyse the relations between actors interacting with each other in a given policy sector[1].

Automatic extraction of the Policy networks is achieved by information which is collected from the web. The relationship between actors can be estimated by i) the frequency of co-occurrence for each pair of actors in web documents, ii) the lexical contextual similarity between snippets iii) the co-occurrence of hyperlinks. In policy networks, nodes are actors who are involved in a given policy field and their relationship (friendship, coauthorship) is represented in edges [2] [3].

After extracting the web documents, the web spam is detected using some technique like AIRWeb [link] competition. It is the definition of features which take different values for spam and non-spam pages. It has focused on quantitative features and also qualitative features to improve web spam detection.

The proposed algorithm is applied to web document's metrics and social networks metrics. It increases the efficiency and reduces the error rate and develop high scalable algorithm for social proximity and spatial proximity [4]. Social and Spatial Ranking Query (SSRQ) reports the top-k users in the social networks based on ranking function that incorporates social and spatial distance from the query user.

II. RELATED WORK

Analyzing large amount of data for the policy networks has flourished in past years. There are two computational methods are used text analysis and social analysis. Most of the text analysis is done by electoral campaigns for identifying voter's profilers, ideology positions, political interaction and conflict contents [5][6]. One of the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

approaches of text analysis is the hand coding of traditional and highly labor intensive techniques of content analysis. That should be replaced by computerized coding schemes by matching the text to coding dictionaries. Estimating the policy position between two parties can be done by calculating the word scores. It can be done by WORDSCORE[7]. This can calculate the economic policy dimension to particular party. WORDFISH is an algorithm to estimate policy position based on frequencies in text. Extracted positions reflect changes in the party system more accurately than existing time series estimates and also to examine which words are important for placing parities on the left and right [8].

Social network analysis is an important fact of policy network. It can make formal and informal interactions. Extracting the social networks is done by the following steps (i) relation identification (ii) relation labeling [10] and (iii) estimating strength. Relation labeling is based on link based ranking method. Link refers to the sites and listing the site is automated rather than manual work. Generally site is an organized collection of pages on a specific topic maintained by a person or group. Topic of site might be quite broad, it broad range of subject matter and provides a range of service. Site finding is a huge process and also it makes a problem describing them. Inside the site it has a more content and the common features like the frequency of co-occurrence of related pair of terms in web documents, but other features lexical content, key phrase, log files and e-mail information are used to identify the relation.

Social network is an important role in our life which includes knowledge management, information retrieval. That network should be extracted using some technique. One of the techniques is POLYPHONET which employs several advanced way of doing things to extract relations of people, to detect groups of people and also to obtain keywords for a people [9]. It has three main things to do, first to reduce the related method into simple, second to create algorithm for new things like classifying the categories, obtaining and utilizing the person-to-word relationship and finally all the modules are implemented into POLYPHONET. Automatically extracted keypharses are used to describe the relation between business such as connections/associations, roles, locations, part-whole and social relationships. Obtain a local context in which two entities co-occur on the web and accumulate the context of the entity pair, the key idea is clustering all entity pairs according to similarity of their collective contexts [8].

The influence between two users captures the probability that one user follows the other's actions. Stem from social networks can improve the marketing strategies, for instance, by recommending product to users based on purchases of their contacts [17]. In past graph scores were created for influencing users. But data scientists directly obtain the top-k most influential users from the historical data without intermediate step of constructing influence graph. Some of the techniques are, vertex neighborhoods (Shortest path distance between the users)[18] and the number of common friends [19].

Object associate with multiple domains attributes. Spatial and textual domains are used to querying the geo-tags which contains geo information and flick photos in web pages. Queried data has applied to IR tree structure(R tree with inverted file) which can be used to efficiently support novel types of spatio-textual queries [20]. [21] By observing location-based social networks, they build supervised learning framework based on predictions and common check-ins.

In graph search, the users build the graph with shortest path from source to target vertex. Dijikstra's algorithm starts from source and iteratively expands the network using priority heap until the target reached. Landmark concept were introduced which select a set of vertices as landmark in graph and pre-computes distance from every vertex to each landmark [22]. An approach to compute approximate distance between vertices in a graph is to construct oracles which provide constant query time while having linear space requirements.

III. RELATEDNESS METRICS

There are 3 kinds of metrics for computation between actors (i) Page count metric (i) Text based metrics and (iii) Link based metrics. The classical information retrieval techniques and natural language processing mainly consist of two stages (i) Extracting relevant information on a link and constructing complex queries and requesting to a search engine [10]. In the first stage, anchor text is an important source but it doesn't return enough information. So we take other sources like URL- the page that contains the link, the context of anchor text and cached page version of the analyzed sources that can be stored in digital library. Extraction can be done by two main approaches and they are based on frequency (Term Frequency- Inverse Document Frequency) and statistical language model (KL divergence). In the second stage, a link is considered to be recovered if the page pointed by the link is in the set of pages retrieved with some of the queries.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

Pagecount based metrics

Estimate the co-occurrence between the policy actors. In web documents Co-occurrence means capturing the variety of relation among the terms ranging that is both deal with common policy issues and serve with policy functions. For a set of documents indexed by search engine denoted by {D} and its cardinality be |D|. $\{D_{ai}\}$ is indicates the set of web document indexed by actor ai. $\{ D_{ai} , D_{aj}\}$ represents set of web documents that contains actor ai and aj and its cardinality $| D_{ai} , D_{aj} |$.[3]

Jaccard coefficient: This coefficient figure out the similarity between sets. The jaccard coefficient S^P_J between ai and aj is defined as:

$$S_J^P(ai,aj) = \frac{|D_{ai,aj}|}{|D_{ai}|+|D_{aj}|-|D_{ai,aj}|}$$

Dice coefficient: It is slightly same as jaccard coefficient and it defined as:

$$S_D^P(ai,aj) = \frac{2|D_{ai,aj}|}{|D_{ai}|+|D_{aj}|}$$

Mutual information: If the set of documents are random variable then point wise mutual information reflects the dependence between the occurrences of actors[11].

$$S_I^P(ai,aj) = \log \frac{\frac{|D_{ai,aj}|}{|D|}}{\frac{|D_{ai}|}{|D|}\frac{|D_{aj}|}{|D|}}$$

Google-based semantic relatedness: It is another similarity metric and proposed in [12][13]

$$S_R^P(ai,aj) = \frac{\max\{\log|D_{ai}|,\log|D_{aj}|\}-\log|D_{ai,aj}|}{\log|D|-\min\{\log D_{ai}|,\log|D_{aj}|\}}$$

Text-Based metrics

The proposed text based metric computes the lexical similarity between political actors who appear in snippet. The lexical similarity may occur in syntactic, semantic and topical features for example if two actors share their political activities, then it is expected that their activities mentioned in their lexical surroundings. To extract the lexical features for actor ai, text –based metrics apply a contextual window W (containing W words preceding and W word following the actor)

$$|f_{W,L}\ldots\ldots\ldots f_{2,L}, f_{1,L}|ai|f_{1,R}, f_{2,R}\ldots\ldots\ldots f_{W,R}|$$

Where f_j is the jth feature that exist left and right to the context of actor ai and the feature vector is built as $V_{ai,w}=(v_{ai,1}, v_{ai,2},\ldots\ldots\ldots, v_{ai,N})$. $v_{ai,j}$ is an non negative integer and N is an vocabulary size. Context-based metric S_W^T computes the cosine similarity between the actors ai and aj as follow as

$$S_W^T(ai,aj) = \frac{\sum_{l=1}^{N} v_{ai,l} v_{aj,l}}{\sqrt{\sum_{l=1}^{N}(v_{ai,l})^2}\sqrt{\sum_{l=1}^{N}(v_{aj,l})^2}}$$

Link based metrics

Examine the number of hyperlinks which is commonly shared between the two actors that contains the term of interest. Common links denote that political actors share the same interest or point to common links in the networks. It indicates the hyperlinks will point to topically relevant web sites and documents. Generally the outlinks are represented in two form either full form where whole path is specified or base form where only the main website address is indicated.

The set of outlinks for actors is represented by $\{O_{ai}\}$ that appears in the web documents. In this metrics the similarity between the actors can be calculated based on the overlap between the members of their outlink sets.

Google-based semantic relatedness using outlinks(S_G^L)

$$S_G^L(ai,aj) = \frac{\max\{\log|O_{ai}|,\log|O_{aj}|\}-\log|O_{ai,aj}|}{\log|O|-\min\{\log O_{ai}|,\log|O_{aj}|\}}$$

Where O_{ai} , O_{aj} and $\{ O_{ai,} O_{aj}\}$ are set of outlinks for actor ai and aj and jointly for both (intersection of O_{ai} and O_{aj})

Recovery Degree

In system each page tries to retrieve all their links and result as, three values are obtained: (i) the number of recovered links (top ten of search) (ii) the number of unrecovered links and (iii) the difference between both previous values. The degree of recovered links can be understood as a coherence measure between the analyzed page, one of its links, and the page pointed by this link [14].

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

Incoming-Outgoing links

This is link from spam pages to non-spam pages but non spam pages do not link spam pages. Taking advantages of the possibilities of the system to submit queries to search engine, to include new query to request the search engine, how many sites point to the analyzed pages (incoming pages).

Broken Links

It is common problem for both spam and non-spam pages, even when this sort of link has a negative impact in the PageRank. The number of spam pages is higher in almost the whole range of numbers of broken links considered.

Anchor text typology

The anchor text of many links is usually generated thinking in the context of the search engines instead of the users. We have to selected four features in order to measure the number of links that are formed by (i) punctuation marks (ii) digits (ii) a URL and (iv) an empty chain [14].

Linear fusion of relatedness metrics

Combine all the features using late integration, relatedness scores from the types of metrics. Linear fusion (S) between two actors is given by

$$S(ai,aj) = \lambda_P S^P(ai,aj) + \lambda_T S^T(ai,aj) + \lambda_l S^T(ai,aj)$$

Where λ_P ,λ_T λ_L are corresponding weights for each metrics.

IV. METHODOLOGY

The proposed algorithm and metrics have been evaluated on the policy networks. In policy networks, documents, snippets and number of hits are mined from web for pair of actors and relatedness metrics are calculated in the form of correlation. The selection of most descriptive terms or passages from the text is crucial for several tasks. In major cases, ranking of all keywords or sentences are done and then top ranked items are selected as features, Semantic rank is a graph based ranking algorithm where it constructs a semantic graphs using implicit links, based on semantic relatedness of text nodes. It also ranks the consequent nodes using different ranking algorithms [15].

Algorithm Semanticrank

1. Input: a text document collection D, Mode flag
2. Output: a ranking R of the semantic graph nodes for every document $d_j \epsilon$ D
3. Execute(D, mode)
4. If mode is keyword then
5. Identify composite term of length up to 5 words
6. End if
7. Compute and index TF-IDF values for all terms
8. For all $d_j \epsilon$ D do
9. G: an initially empty graph
10. G: Construct_Semantic_Graph($d_j \epsilon$ D)
11. R= Rank_nodes(G)

Construct Semantic Graph($d_j \epsilon$ D)

12. G: an initially empty graph
13. If mode is keyword then
14. Initialize G with K_{dj}
15. Else
16. Initialize G with Sen_{dj}
17. End if
18. For all pairs of verticea(v_i , v_j)do
19. If mode is keyword then
20. $W_{i,j}= W_{j,i}=\lambda_{vi,vj}.SRT(v_i, v_j)$
21. Else
22. $W_{i,j}= W_{j,i}=SRS(v_i, v_j)$
23. End if
24. End for
25. Return G

Rank nodes G

26. Execute weighted page rank in G
27. R= rank vertices of G in descending order of pagerank values
28. Return R with their page rank values

According to the divergence value the pages should characterize the relationship between two linked web pages. These values are obtained by calculating the Kullback–Leibler (KL) divergence between one or more sources of information from each page. KL divergence can be applied to the anchor text of a link and the title of the page pointed

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

by this link. KL measures the probability of more relevant terms in terms of segment texts and documents [14].

SVM is supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. It is used to identify the most informative metric more efficiently. In this module, we introduce a machine learning algorithm for selecting the most informative metrics based on the frequent occurrences, as well as, for their fusion, and filtering web data based on relevance. It is used to classify and estimate the weights for the linear fusion of the additional metrics. In policy network, this algorithm is efficient for classification of social users [16].

Social and spatial ranking query (SSRQ) reports the top-k users in the SN based on a ranking function that incorporates social and spatial distance from the query user.

Algorithm top-k users

1. Input: N training samples for classification,
2. Output: predicted class
3. For each sample do
4. Input[i] =sample
5. Foreach i in neural metwork do
6. Output [i] = module. ForwardPropagate(input[i])
7. Input[i+1]=output[i]
8. End
9. Predictedclass = criterion (output)
10. If training then
11. Check error attributes
12. For each [k-i] in neural network do
13. Output = module. BackwardPropagate
14. Input[i+1]=output[i]
15. End
16. End
17. End

In each input sample, it can be positive, negative or neutral in the data sample. For the number of input training sample we have to perform the analysis of reviews. For each input sample the neuron network perform the output sample based on the prior knowledge. It is used to search the more accurate results along with hidden layer and this layer is used to map the semantic results for corresponding input sample. Hence it avoids the error values also it progress the speed of process more efficiently.

The above mentioned algorithm is used to classify and estimate the weights for the linear fusion of the additional metrics. The NN algorithm achieves better efficiency in finding the most informative metrics. The other metrics are fused using linear fusion technique. In policy network, this algorithm is efficient for classification of social users.

V. EVALUATION

To execute the proposed technique and generate various results we use java language in this environment. In this section, the analysis has been done for existing and proposed research work by using algorithms. In this evaluation part, the performance metrics are executed by using existing and proposed method. The performance metrics are such as correlation, MSE values, graph distance and accuracy. The proposed system is shown the higher performance in terms of high correlation, lower MSE values, efficient graph distance and greater accuracy.

The process can be evaluated based on correlation and Mean Square Error (MSE). Correlation is calculated depends on some relatedness scores. That has H= (h_1, h_2,.., h_M)and K =(k_1,k_2,…. k_M) indicates human rated and automatically computed relatedness scores, respectively, where M is the total number of relations. k_i is calculated from the section 3 mentioned metrics.

It scaled as

$$e_i = \frac{2(k_i - k_{\min})}{k_{\max} - k_{\min}} + 1$$

Where k_{min} and k_{min} is minimum and maximum score of metrics and is normalized relatedness score that takes continuous value [3]. Correlation is calculated between human rate and normalized relatedness score by using Pearson coefficient,

$$r_{H,E} = \frac{\sum_{i=1}^{M}(h_i - \overline{H})(e_i - \overline{E})}{\sqrt{\sum_{i=1}^{M}(h_i - \overline{H})^2 \sum_{1=1}^{M}(e_i - \overline{E})^2}}$$

Where $\overline{H}$ and $\overline{E}$ denoted the sample mean of H and E. In Mean Square Error (MSE) is used to measure the distance between the human ratings and normalized relatedness averaged over all investigating relations.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

The correlation value should be high in the policy network. If the correlation value is high then system is performed in an effective manner.

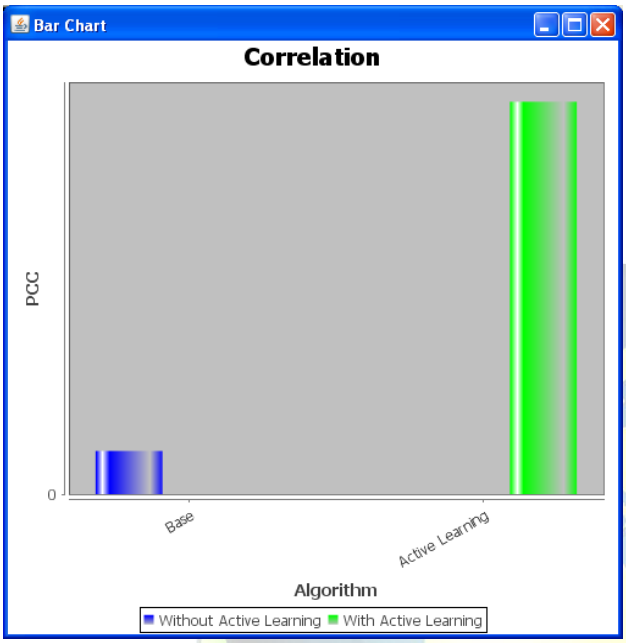
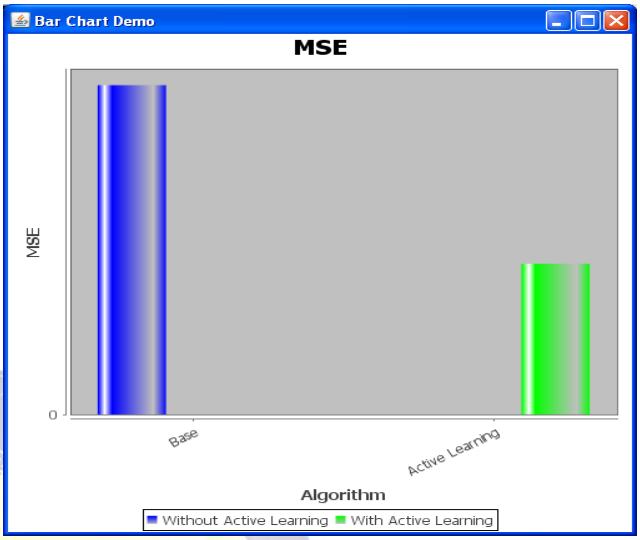


Fig. 1 Correlation among non-active learning and active learning algorithm

From the above graph we can observe that the comparison of existing and proposed system in terms of correlation metric. In x axis we plot the algorithms and in y axis we plot the correlation values. The correlation values are lower by using existing algorithm. The correlation value is higher by using the proposed algorithm of active learning. From the result, we conclude that proposed system is superior in performance.

In statistics, the mean squared error (MSE) of an estimator measures the average of the squares of the "errors". $MSE = \frac{1}{M}\sum_{i=1}^{M}(h_i - e_i)^2$

Note the MSE values lies between 0 and 4 [3].



The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

Accuracy can be calculated from formula given as follows

$$Accuracy = \frac{true\ positive + True\ negative}{TruePositive + True\ negative + False\ positive + False\ negative}$$

An accuracy of 100% means that the measured values are exactly the same as the given values.

By using machine learning algorithm the proposed system has shown the highest accuracy rather than previous system.

The social and spatial proximities are calculated based on the graph distance value. The proposed system is shown the aggregate proximity values efficiently and the nearest as well shortest distances are computed effectively.

VI. EXPERIMENTAL SETUP

In this section we experimentally evaluate the ranking algorithm and proximity values. All the methods are implemented in Java. We use publicly available dataset like INDIA, CHINA [other countries]. In experimental setup, first to retrieve the document from the web by entering the actors and then to listing the abbreviation from the web

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

documents. To tackle both data sparseness and term ambiguity problems and also a number of lexicalization form is manually selected for each actor in collaboration with actors. The machine learning algorithm is such as SVM which is used to identify the most informative metrics in more efficiently.

TABLE I. INFORMATIVE METRICS

ACTORS	PAGE COUNT BASED METRICS			
	Jacaard coefficient	***Dice coefficient***	***Mutual information***	***Google***
India,China	0.690	0.549	0.078	0.431
	TEXT BASED METRICS(COSINE SIMILARITY)			
	0.5211			
India, China	**LINK BASED METRICS**			
	0.2088			
	Linear Fusion			
	0.4627			

After calculate the metrics to select the most informative metrics and to grouping the actors according to the deadline. That dead line should be decided by developer based on required actors. For example the minimal value is 0.6 and who are all having above 0.6 in linear fusion they become group 1 or else they become group 2. Classification is use to easily identify the actor's relationship. Finally it calculate the correlation co-efficient and MSE among the active learning algorithm and manual calculation. That should be describe in previous section. MSE would be reducing and correlation would be increase.

The second part of this work is social network. Recently so many social network are available like Facebook, linkedIn, Flickr, Google+ etc., here I took Facebook. First step, to login the Facebook (Fb) Id in search engine. For retrieving the data from the Fb, want to get token from Fb. For that to go Facebook developer login and get token from fb to logged in account. After entering token, to get the friends details from the Facebook . Calculate the social and spatial proximity for each friend with log in person. Like who made a post to me or like my profile, made Comments etc,.

Graph distance should be calculated among the actor and all the users. Ecludean distance is used for calculate the graph distance. Then Support Vector Machine algorithm is find the top kth person to make interaction with user. Finally the accuracy should be returned.

VII. CONCLUSION

The proposed system is higher efficiency than the existing system. System is identifying automatically the actors participating in policy networks and their lexicalizations. And also in the proposed system, the semantic ranking function is introduced to estimate the semantic similarity higher. It is filtering web data based on relevance and type of source for instance nonprofit, government, corporate, blogs and news. Then apply the machine learning algorithms such as SVM to select the most informative metric for the additional features. Then apply the machine learning algorithms such as SVM to select the most informative metric for the additional features. In the proposed system, we apply aggregate index search and twofold search approaches for computing nearest neighbor spatial and social proximities in social networks. Also we enhance the graph search distance for finding shortest path distance in the given social network. Thus the proposed system methods are useful to retrieve the more semantic social information. The experimental result proves that the proposed system has highest performance rather than the existing system.

REFERENCES

[1]. Tanja A.Borzel, "An exploration of the concept and its usefulness in studying European governance," European Integration online papers, vol. 1, pp. 016.

[2]. R.A.W. Rhodes,"Policy network analysis", John Donne [1611], 'The First Anniversary. An Anatomy of the World', 1985 edition, 335 line 213.

[3]. Theodosis Moschoulos,Elias Iosif, "Towards automatic extractiojn of policy networks usinf web links and documents," in IEEE, vol. 25, No. 25, 2013, pp. 2403-2417.

[4]. Kyriakos Mouratidis, Jing Li, Yu Tang and Nikos Mamoulis, "Joint search by social and spatial proximity," in IEEE transactions on knowledge and data engineering, Vol 27, No. 3 March 2015.

[5]. Burt L.Monroe, "introduction to the special issue: The statistical analysis of political text,"in Advance access publication, February 4, 2009.

[6]. L.Zhu, Computational Political Science Literature Survey, http://www.personal.psu.edu/luz113, 2013.

[7]. Michael laver and Kenneth benoit, "Extracting policy positions from political texts using words as data,"American political science review, vol. 97, No. 2,May 2003

[8]. Jonathan B. Slapin and Sven-oliver Proksch,"A scling model for estimating time-series party positions from texts" Trinity college, Dublin and University of california, Los Angeles.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 5, May 2016

[9]. Nick Craswell, David Hawking and Stephen Robertson " Effective site finding using link anchor information" .

[10]. Yutaka Mastsuo, Junichiro Mori and Masahiro Hamasaki " POLYPHONET: An advanced social network extraction system from the web" in National institude of Advanced industrial science and technology, United Staes, University of Tokyo, Japan.

[11]. D. Bollegala, Y. Matsuo, and M. Ishizuka, "Measuring Semantic Similarity between Words Using Web Search Engines," Proc. 16th Int'l World Wide Web Conf., pp. 757-766, 2007.

[12]. R. Cilibrasi and P. Vitanyi, "The Google Similarity Distance," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 370-383, Mar. 2007.

[13]. P. Vitanyi, "Universal Similarity," Proc. Information Theory Workshop Coding and Complexity, pp. 238-243, 2005.

[14]. Lourdes Araujo and Juan Martinez- Romo, " Web Spam Detection: Mew Classification Features Based on Qualified Link Analysis and Language Model" in IEEE Trans. On Information Forensics and Security, Vol. 5, No. 3 September 2010.

[15]. George Tsatsaronis and Iraklis Varlamis, " Semantic Rank: Ranking Keywords and Sentences using Semantic Graphs, in Norwegian University of Science and technology

[16]. Priyanka N Guttedar and Pushpalata S, " Scene Text Recognition in Mobile Application using K- Mean Clustering and Supporrt Vector Machine", In IJARCET, Vol. 4 5 May 2015.

[17]. L. Adamic, O. Buyukkokten, E. Adar, "Social network caught in the Web",First Monday 8 (6).

[18]. J. Golbeck, J. Hendler, Accuracy of metrics for inferring trust and reputationin semantic Web-based social networks, in: Proceedings of the EKAW 2004, 2004.

[19]. P. Mika, Flink, Semantic Web technology for the extraction and analysis of social networks, J. Web Semantics 3 (2) (2005) 211–223.

[20]. M. Harada, S. Sato, K. Kazama, Finding authoritative people from theWeb,in: Proceedings of the Joint Conference on Digital Libraries (JCDL2004),2004.

[21]. P. Turney, Mining theWeb for synonyms: F PMI-IR versusLSAon TOEFL,pp. 491–502, in: Proceedings of the ECML-2001, 2001.

[22]. J. Goecks, E.D. Mynatt, Leveraging social networks for information sharing, pp. 328–331, in: Proceedings of the ACM CSCW, 2004.