

# Feed Forward Based ANN Approach for Hash Tag Recommendation over Micro blogging

Priyanka Shrivastava<sup>#1</sup>

<sup>#</sup>CS Department, RGPV University  
India

<sup>1</sup>pspriyankacse@gmail.com

**Abstract**— In this era of science we all are using several kinds of technologies also several kinds of social networking sites that provide us the facility to communicate to the several people at a time. But here not everything gets over. The technologies we have using has a lot of data which will also get pile-up somewhere. So, for this information mining is an interdisciplinary field of software engineering and is alluded to as extricating or mining learning from expansive databases. It is the way toward performing mechanized extraction and creating the prescient data from a substantial database. Information mining comprises of the different specialized methodologies including machine learning, measurements, and database framework and so on.

**Keywords**—: Information Mining, ANN Approach, SDLC, Databases, FPFPA.

## I. INTRODUCTION

In the current years, a gigantic measure of information is being accumulated and put away in databases wherever over the globe, which is primarily originating from data industry and person to person communication destinations. There is a need to remove and order valuable data and learning from such information gathered. Information mining is an interdisciplinary field of software engineering and is alluded to as extricating or mining learning from expansive databases. It is the way toward performing mechanized extraction and creating the prescient data from a substantial database. It is really the way toward finding the concealed data or examples from the stores. The fields that utilization Data mining strategies incorporate medicinal research, advertising, media transmission, and securities exchanges, social insurance et cetera[1].

Information mining comprises of the different specialized methodologies including machine learning, measurements, database framework and so forth. The objective of the information mining process is to find learning from extensive databases and change into a human reasonable configuration. The DM and information disclosure are basic parts to the association because of its basic leadership technique. Arrangement, relapse and bunching are three methodologies of information mining in which occasions are assembled into recognized classes. Arrangement is a well known undertaking in information mining particularly in learning revelation and

future arrangement. It gives the insightful basic leadership. Arrangement not just examinations and inspects the current example information yet in addition predicts the future conduct of that specimen information. It maps the information into the predefined class and gatherings. It is utilized to foresee aggregate enrollment for information occurrences[8].

## II. RELATED WORK

In year 2007, Michael J Pazzani in the "Substance Based Recommendation System" examines content-based proposal frameworks, i.e., frameworks that prescribe a thing to a client in light of a depiction of the thing and a profile of the client's advantages [25].

In year 2006, K S Esmaili in the paper "Looking at Performance of proposal Technique in the Blogosphere" portray Weblogs one of crucial segments of Web experience issues in finding applicable online journals. Recommender 2.0 and there are a considerable measure of incompetent bloggers and guests systems' identity an answer for the data overburden issues. In this paper a weblog recommender framework in light of connection structure of weblog diagram is presented. Here we regard connects between weblogs as some sort of rating. The techniques are actualized on a genuine dataset [26]

In year 2001, J. Ben Schafer in the paper "Internet business Recommendation Applications" look at how recommender frameworks help E-business locales increment deals and break down the recommender frameworks at six market-driving destinations. In light of these cases, make a scientific classification of recommender frameworks, including the sources of info required from the buyers, the extra learning required from the database, the ways the proposals are exhibited to customers, the advances used to make the suggestions, and the level of personalization of the recommendations[27].

In year 2004, Jonathan L. Herlocker in the paper "Assessing cooperative sifting recommender frameworks" survey the key choices in assessing synergistic separating recommender frameworks: the client errands being assessed,

the sorts of examination and datasets being utilized, the courses in which expectation quality is measured, the assessment of forecast characteristics other than quality, and the client based assessment of the framework overall [28].

In Year 2008, Kleanthi Lakiotaki played out a work, "UTA-Rec: A Recommender System in light of Multiple Criteria Analysis". UTARec, a consolidates Multiple Criteria Analysis techniques is introduced. The framework's execution and capacity of tending to specific deficits of existing Recommender Systems is exhibited on account of film suggestions. UTARec's precision is measured as far as Kendall's tau and ROC bend examination and is additionally contrasted with a Multiple Rating Collaborative Filtering (MRCF) approach[29].

In Year 2008, Juan A. Recio-García played out a work, "Prototyping Recommender Systems in jCOLIBRI". Introduced objective is to emotionally supportive network designers in quick prototyping recommender frameworks utilizing Case-Based Reasoning (CBR) procedures. In this paper Author depict how jcolibri can serve to that objective. jcolibri is a question arranged structure in Java for building CBR frameworks that enormously profits by the reuse of already created CBR systems[30].

In Year 2007, Paul Resnick played out a work, "The Influence Limiter: Provably Manipulation Resistant Recommender Systems". Creator depict an influence-constraining calculation that can turn existing recommender frameworks into manipulation-safe frameworks. Creator portray both as far as possible and the data misfortune brought about because of those points of confinement as far as data theoretic concepts of misfortune capacities and entropies[31].

### III. PREVIOUS APPROACH

In this chapter, previous approach and proposed approach result comparison is performed, as per the monitored results from implementation which is obtained is compared. The proposed algorithm is presented and compared with existing solution. This chapter gives a comparison graph and statically analysis. As per observed, finally it shows the proposed approach is efficient in terms of total net flows, malicious net flows, accuracy, detection rate as well in the implementation analysis.

### IV. PROPOSED METHODOLOGY

A proposed approach of level upgrade of FP-Growth calculation with the given proposed neural system engineering sustain forward approach alongside the FP Growth is utilized. An improvement of beforehand characterized approach named FPFANN (FP Feed forward ANN) calculation is given by our framework.

1. Using the ANN bolster forward calculation for the hash label age and discovering importance between the words input.

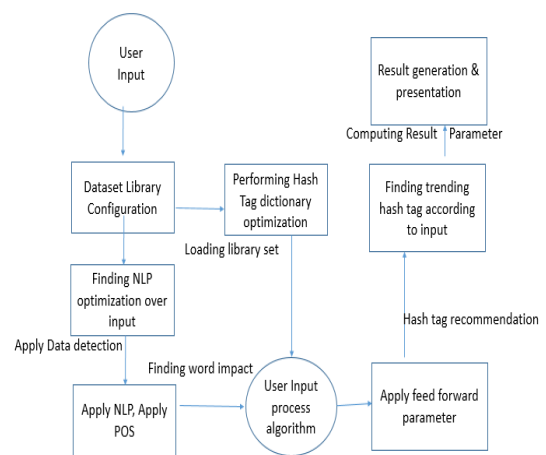
2. Algorithm additionally utilizes the utilization of regular example development approach for finding the word significance and their recurrence. As per their utilization and development score, the calculation is executed.

3. It is working with the substantial database and calculation which works with the beforehand given information approach.

4. This calculation creates a hashtag with high rate of exactness and in addition preferable outcome execution over already utilized ordinary FP-development calculation for the hash label age.

5. A info dataset from the given microblogging stage is taken and process by different sub process library

In proposed calculation a half and half approach working with FP and Feed forward layer based approach of ANN is utilized. This proposed calculation is productive while finding the information pertinence and getting hash label age over the substantial dataset.



**Figure 1. Architecture of Proposed System.**

According to our perception about the past method and their impediment in various terms and situation 's . Our work display another approach which is profoundly secure and devours low computational time and along these lines computational cost over the extensive number of organized accessible dataset. Our work propose another calculation PROPOSED calculation with all the more quick calculation FPFANN is performed alongside Map usefulness and Similarity measure score as more steady approach. Our calculation likewise checks for appropriate repetition, clamor utilizing more secure and dependable parameters. The proposed calculation is depicted underneath :

#### **Algorithm Pseudo Code :**

FPFANN approach:

Input : Input Tweets, Input data post, FP parameters.

Output : Communication process, feed forward outcome, .

Steps :

```

Foreach{
  Tweet file listing{t1,t2....tN};
  Input postrequest();
  Loading vocabData();
  Process FP over Input & Data dictionary;
  NLP performance();
  POS performance();
  Performing Frequence word();
  Finding relevancy();
  Finding feed forward outcome();
  Hashtag= TagGen();
  If(Hashtag==Accepted)
  {
  Substitute Hashtag();
  Set status=Active; generate statistics ();
  Generate result();
  Plotting outcomes();
  }else
  {
  Status=exit;
  generating data for request;
  }
  Return Computation time;
}
End.

```

According to our perception about the past method and their disservice in various terms and scenario's. Our work show another approach which is exceptionally secure and devours low computational time and consequently computational cost over the substantial number of organized accessible dataset.

Our work propose another calculation , Enhanced FPFANN calculation with all the more constant information handling algorithm is performed . Our calculation additionally checks for appropriate access control utilizing more secure and solid parameters.

The proposed calculation is portrayed beneath

1. Listing and stacking of all the parameter, part for the reproduction reason and arrangement of all the required situation system.

2. Creating a question of all required part.

#### IV. EXPERIMENT EVALUATION & RESULT ANANLYSIS

All the experiments were performed using an i3-4005U CPU @ 1.70 GHz processor and 4 GB of RAM running windows 10. The discussed feature selection algorithms were implemented using language Java and the platform Eclipse IDE.

#### V. RESULT ANALYSIS

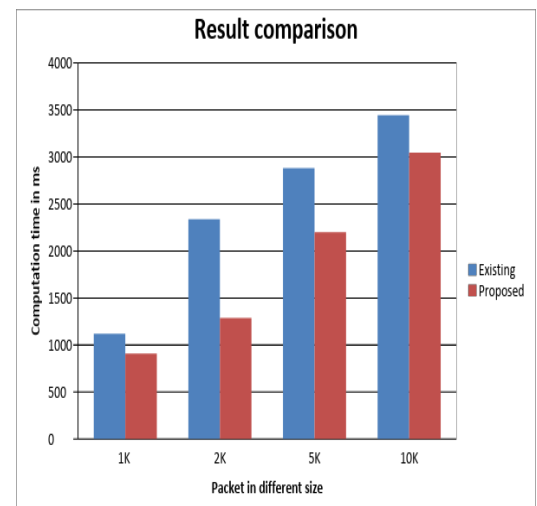
In result examination here is the framework resilience detail I have connected some arbitrary snap and watched my best investigation result. In the proposed framework, A twitter ongoing information bringing , mark area finding and plotting

| Technique Approach Number of Tweets / Computation time in ms | Existing technique (Computation time in msec) | Proposed Technique (Computation time in msec) |
|--|---|---|
| 1K   | 1121  | 910   |
| 2K   | 2339  | 1289  |
| 5K   | 2881  | 2201  |
| 10K  | 3443  | 3046  |
| 50K  | 8776  | 6989  |

over the guide and looking is performed. So it is essential that total procedure stream easily. In underneath figures, a concise review of the proposed venture is introduced.

The proposed and existing system is performed with the above various information estimate document, where the information is handled and following yield comes about were observed:

**Table 1: Statically analysis of obtained result.**



**Figure 2. Comparison Bar graph for technique analysis.**

In the above diagram drawn x hub as computational time for the question preparing for indicated dataset and structured presentation is printed utilizing the outline library gave by the Microsoft and further examination would easility be able to performed accordingly the PROPOSED outflank the best and low computational time with a similar inquiry number as the calculation time is diminished because of Proposed strategy. The diagram portrayal demonstrates the productivity of our proposed calculation work and it beat the low computational time, along these lines the low computational cost with the quantity of various question and information handling.

| Technique Approach<br>Tweets /<br>Computation time in<br>ms | Existing technique<br>(Computation time in<br>msec) | Proposed Technique<br>(Computation time in<br>msec) |
|---|---|---|
| 1K  | 12.2  | 14.36   |
| 2K  | 17.98   | 16.0  |
| 5K  | 23.12   | 20.9  |
| 10K   | 43.5  | 40.89   |
| 50K   | 143.8   | 132.6   |

**Computation Cost Comparison :**

The proposed and existing technique is performed with the above different data size file , where the data is processed and following output results were monitored:

Table 2: Statically analysis of obtained result

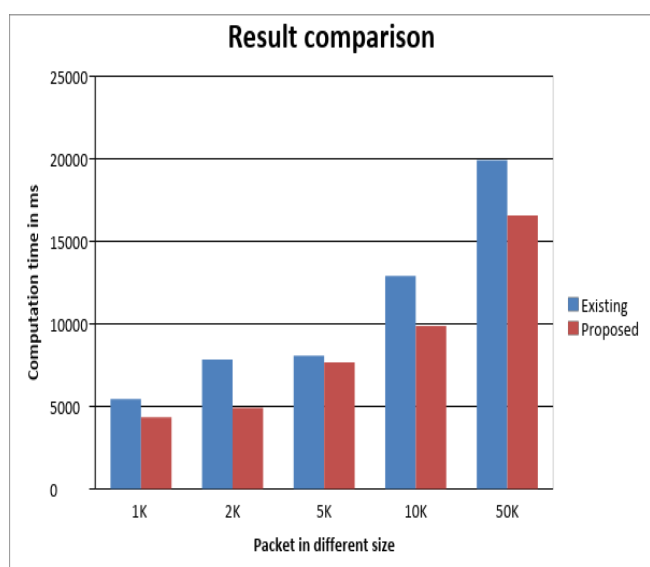


Figure 3. Comparison Bar graph for technique analysis

In the above diagram drawn x pivot as computational time for the question handling for indicated dataset and structured presentation is printed utilizing the outline library gave by the Microsoft and further examination would easility be able to performed in this way the PROPOSED beat the best and low computational time with a similar inquiry number as the calculation time is diminished because of Proposed procedure. Dealing with content based graphical calculation, for example, our proposed method validation will be effective and simple to imagine and so as to make it all the more simple for client to utilize, such QAP strategy will be productive to utilize.

**Computation Time Complexity :**

The proposed and existing technique is performed with the above different data size file , where the data is processed and following output results were monitored:

Table 3: Statically analysis of obtained result

| Technique Approach<br>Tweets /<br>Computation time<br>in ms | Existing technique<br>(Computation time<br>in msec) | Proposed Technique<br>(Computation time in<br>msec) |
|---|---|---|
| 1K  | 5454  | 4340  |
| 2K  | 7845  | 4908  |
| 5K  | 8081  | 7668  |
| 10K   | 12897   | 9880  |
| 50K   | 19912   | 16569   |

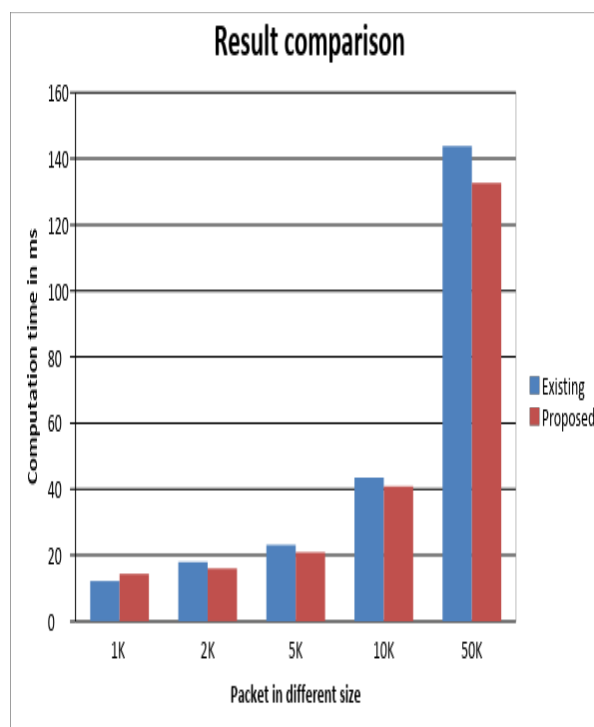


Figure 4. Comparison Bar graph for technique analysis

In the above diagram drawn x hub as computational time for the question preparing for indicated dataset and visual diagram is printed utilizing the graph library gave by the Microsoft and further examination would easility be able to performed in this manner the PROPOSED beat the best and low computational time with a similar inquiry number as the calculation time is diminished because of Proposed system.

IV. CONCLUSIONS

Accordingly to defeat these issues related with conventional paper approach. Our improvement calculation PROPOSED calculation with more proficient calculation is performed alongside ongoing tempest apparatus as more steady confirmation approach. Our calculation likewise checks for appropriate excess, clamor utilizing more secure and dependable parameters.

Our proposed calculation makes utilization of correlation parameter as calculation time and in addition calculation cost to register the examination investigation. The calculation is



engineer in Java dialect with Java net-beans apparatus setup utilizing intel i3 processor, 750 GB RAM. The examination investigation and execution result demonstrates that our proposed approach PROPOSED beat best while contrasting and existing calculation.

#### ACKNOWLEDGMENT

I am very thankful to my Guardians who has helped me throughout.

#### REFERENCES

- [1]. Mir Saman Tajbakhsh, Jamshid Bagherzadeh, "Microblogging Hash Tag Recommendation System Based on Semantic TF-IDF", IEEE 2016 4th International Conference on Future Internet of Things and Cloud Workshops.
- [2]. Otsuka, E., Wallace, S.A., and Chiu, D.: 'Design and evaluation of a Twitter hashtag recommendation system'. Proc. Proceedings of the 18th International Database Engineering & Applications Symposium, Porto, Portugal2014 pp. Pages.
- [3]. Gupta, M., Li, R., Yin, Z., and Han, J.: 'Survey on social tagging techniques', SIGKDD Explor. Newsl., 2010, 12, (1), pp. 58-72
- [4]. Givon, S., and Lavrenko, V.: 'Predicting social-tags for cold start book recommendations'. Proc. Proceedings of the third ACM conference on Recommender systems, New York, New York, USA2009 pp. Pages.
- [5]. Li, R., Wang, S., Deng, H., Wang, R., and Chang, K.C.-C.: 'Towards social user profiling: unified and discriminative influence model for inferring home locations'. Proc. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, China 2012 pp. Pages.
- [6]. Y. Matsuura, Y. Sakurai, B. A. Prakash, L. Li, and C. Faloutsos. Rise and fall patterns of information diffusion: model and implications. KDD, 2012
- [7]. Zangerle, E., Gassler, W., and Specht, G.: 'On the impact of text similarity functions on hashtag recommendations in microblogging environments', Social Network Analysis and Mining, 2013, 3, (4), pp. 889-898
- [8]. Godoy, D., Rodriguez, G., and Scavuzzo, F.: 'Leveraging Semantic Similarity for Folksonomy-Based Recommendation', IEEE Internet Computing, 2014, 18, (1), pp. 48-55.
- [9]. Lops, P.; Gemmis, M.; de Semeraro, G. Content-based Recommender Systems: State of the Art and Trends. In Recommender Systems Handbook; Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., Eds.; Springer US: New York, NY, USA, 2011.
- [10]. Oxana Ye. Rodionova, Paolo Oliveri, and Alexey L. Pomerantsev. 2016. Rigorous and compliant approaches to one-class classification. Chemometrics and Intelligent Laboratory Systems 159, Complete (2016), 89–96.
- [11]. Choi, K.; Yoo, D.; Kim, G.; Suh, Y. A hybrid online-product recommendation system: combining implicit rating-based collaborative filtering and sequential pattern analysis. Electron. Commer. Res. Appl. 2012, 11, 309–317
- [12]. Ukrit Marung 1 , Nipon Theera-Umpon 1,\* and Sansanee Auephanwiriyakul," Top-N Recommender Systems Using Genetic Algorithm-Based Visual-Clustering Methods", Symmetry 2016, 8, 54; doi:10.3390/sym8070054.
- [13]. Sigurbjornsson, B., and Zwol, R.v.: 'Flickr tag recommendation based on collective knowledge'. Proc. Proceedings of the 17th international conference on World Wide Web, Beijing, China 2008 pp. Pages
- [14]. Lipi Shah, Hetal Gaudani, Prem Balani," Survey on Recommendation System", International Journal of Computer Applications (0975 – 8887) Volume 137 – No.7, March 2016.
- [15]. Carlos A. Gomez-Urbe and Neil Hunt. 2015. The Net-ix Recommender System: Algorithms, Business Value, and Innovation. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13 (Dec. 2015), 19 pages.
- [16]. Yandi Xia, Giuseppe Di Fabbriozio," A Content-based Recommender System for E-commerce Oers and Coupons", In Proceedings of SIGIR eCom 2017, Tokyo, Japan, August 2017, 7 pages. DOI: 10.1145.
- [17]. J. Bobadilla, F. Ortega, A. Hernando, and A. GutiéRrez. 2013. Recommender Systems Survey. Knowledge Based Syst. 46 (July 2013), 109–132.
- [18]. Yang, X.; Steck, H.; Liu, Y. Circle-based Recommendation in Online Social Networks. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012, Beijing, China, 12–16 August 2012.
- [19]. Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. 785–794
- [20]. W. Feng, J. Wang, W. Zhang. We Can Learn Your #Hashtags: Connecting Tweets to Explicit Topics. ICDE 2014, Chicago, IL, USA.
- [21]. Li, C.; Liang, C.-Y. Cold-start eliminating method of collaborative filtering based on n-sequence access analytic logic. Syst. Eng. Theory Pract. 2012, 7, 1537–1545.
- [22]. Kim, H.-N.; El-Saddik, A.; Jo, G.-S. Collaborative error-reflected models for cold-start recommender systems. Decis. Support Syst. 2011, 51, 519–531.
- [23]. Godoy, D., Rodriguez, G., and Scavuzzo, F.: 'Leveraging Semantic Similarity for Folksonomy-Based Recommendation', IEEE Internet Computing, 2014, 18, (1), pp. 48-55.
- [24]. S. Chung, D. McLeod. Dynamic Pattern Mining: An Incremental Data Clustering Approach. University of Southern California
- [25]. Sigurbjornsson, B., and Zwol, R.v.: 'Flickr tag recommendation based on collective knowledge'. Proc. Proceedings of the 17th international conference on World Wide Web, Beijing, China2008 pp. Page.





- [26]. Ferragina, P., Piccinno, F., and Santoro, R.: 'On Analyzing Hashtags in Twitter' (2015. 2015)
- [27]. Bart P. Knijnenburg, Martijn C. Willemsen, Alfred Kobsa" A Pragmatic Procedure to Support the User-Centric Evaluation of Recommender Systems", RecSys'11, October 23–27, 2011, Chicago, Illinois, USA. ACM 978-1-4503-0683-6/11/10 (pp 321-324)
- [28]. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. International Conference on Learning Representations.
- [29]. Ayan Bandyopadhyay, Kripabandhu Ghosh, Prasenjit Majumder, and Mandar Mitra. 2012. Query expansion for microblog retrieval. International Journal of Web Science, 1(4):368–380.
- [30]. Erich Christian Teppan," Implications of Psychological Phenomenons for Recommender Systems", RecSys'08, October 23–25, 2008, Lausanne, Switzerland. ACM 978-1-60558-093-7/08/10 (pp 323-326)
- [31]. A. S., and Yang, Y. (2008). Personalized active learning for Collaborative filtering. In SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on the research and development in the information retrieval, Singapore (pp. 91–98). New York: ACM.
- [32]. Zhuoye Ding, Qi Zhang, and Xuanjing Huang. 2012. Automatic hashtag recommendation for microblogs using topic-specific translation model. In Proceedings of COLING 2012: Posters, pages 265–274, Mumbai, India, December. The COLING 2012 Organizing Committee.
- [33]. G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng., 17(6):734–749, 2005.
- [34]. Volodymyr Mnih, Nicolas Heess, Alex Graves, et al. 2014. Recurrent models of visual attention. In Advances in Neural Information Processing Systems, pages 2204–2212.
- [35]. Roman Dvogopoli and Matt Nohelty," Twitter Hash Tag Recommendation", arXiv:1502.00094v1 [cs.LG] 31 Jan 2015.
- [36]. E. Zangerle, W. Gassler, and G. Specht. Recommending #-tags in Twitter. Proceedings of the Workshop on Semantic Adaptive Social Web, 2011
- [37]. Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In Advances in Neural Information Processing Systems, pages 1693–1701.
- [38]. Yang Li†, Ting Liu†, Jing Jiang ‡, Liang Zhan," Hashtag Recommendation with Topical Attention-Based LSTM", Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 3019–3029, Osaka, Japan, December 11-17 2016.
- [39]. Joeran Beel," It's Time to Consider 'Time' when Evaluating Recommender System Algorithms", 2016.
- [40]. Sebastian Nanz • Carlo A. Furia," A Comparative Study of Programming Languages in Rosetta Code", arXiv:1409.0252v4 [cs.SE] 22 Jan 2015.
- [41]. Richard P. Gabriel," The Structure of a Programming Language Revolution", October 19–26, 2012, Tucson, Arizona, USA. Copyright © 2012 ACM 978-1-4503-1562-3/12/10.
- [42]. <http://adilmoujahid.com/posts/2014/07/twitter-analytics/>
- [43]. <https://www.ethz.ch/content/dam/ethz/special-interest/gess/computational-social-sciencedam/documents/education/Spring2016/Datascience/twitter.pdf>
- [44]. <http://www.oracle.com/technetwork/articles/javase/jdk-netbeans-jsp-142931.html>
- [45]. Nanhay Singh<sup>1</sup>, Achin Jain<sup>1</sup>, Ram Shringar Raw," COMPARISON ANALYSIS OF WEB USAGE MINING USING PATTERN RECOGNITION TECHNIQUES", International Journal of Data Mining & Knowledge Management Process (IJDKMP) Vol.3, No.4, July 2013.