

Mining Trending Hash Tags for Arabic Sentiment Analysis

Yahya AlMurtadha

Department of Computer Science

Faculty of Computing and Information Technology, University of Tabuk
Tabuk, Kingdom of Saudi Arabia

Abstract—People text millions of posts everyday on microblogging social networking especially Twitter which make microblogs a rich source for public opinions, customer's comments and reviews. Companies and public sectors are looking for a way to measure the public response and feedback on particular service or product. Sentiment analysis is an encouraging technique capable to sense the public opinion in a fast and less cost tactic than traditional survey methods like questionnaires and interviews. Various sentiment methods were developed in many languages, such as English and Arabic with much more studies in the first one. Sometime, hash tags are misleading or may have a title that does not really reflects the subject. Tweets in trend hash tags may contain keyword or topics titles better represent the subject of the hash tag. This research aims at proposing an approach to explore Twitter Hash tag trends to retrieve tweets, group retrieved tweets to learn topics' profiles, do sentiment analysis to test the subjectivity of tweets then develop a prediction model using deep learning to classify a new tweet to the appropriate topic profile. Arabic hash tags trends have been used to evaluate the proposed approach. The performance of the proposed approach (clustering topics within hashtag trend to learn topics profiles then do sentiment analysis) shows better accuracy than sentiment analysis without clustering the topics.

Keywords—Arabic sentiment analysis; twitter; opinion mining; trending hashtags; text analysis; deep learning

I. INTRODUCTION

With the development of information technology, everyone is interacting with electronic devices of various sizes. Different users from different age groups are dependent on the Internet to collect information, exchange news and jokes. The technological development has strong effect in the transformation and the impact on social life, which prompted all social, cultural and political figures and even government agencies to create accounts in various means of social communication to disseminate what is going on in their minds and to influence the followers.

With the rapid increasing in the number of social networks and the awareness of the crucial role social networks play through the posts published by users, opinion miners are interested in discovering automated ways to retrieve, analyse and report users response so that decision makers can draw their policies to develop services to the community. Previously, public opinion poll methodologies survey relevant people by

distributing questionnaires or interviews which are considered as traditional methods. Mentioned previously traditional survey methods are based on small samples and may not be prepared accurately to cover what the public really feel on specific domain, which may result in inaccurate results, not to mention the cost. While the use of automated artificial intelligence methods may result in larger sample scans and actually measure their wishes indirectly by analysing their emotions and expressions that represents their feelings. In addition, the little cost that may be negligible compared to the cost of traditional survey methods mentioned above.

Data mining professionals use artificial intelligence techniques to develop methodologies that can identify the public's response, measure their opinions and interests, and classify them into different categories that can be used to promote products that are related to them. AI based techniques depend on the analysis of written posts according to the natural language in which they are written and the methodologies of natural language processing. Social networks such as Twitter can be used by specialists to sense the pulse of the public and to measure the extent of their impact or interest in a particular subject via opinion mining. Companies use opinion mining technology to measure the interest and admiration of the public to specific products for better marketing and guide users with ads lining to increase profits.

Opinion miners investigate and develop methodologies for better understanding of natural languages to improve the results of automated retrieval and understanding of desired written texts. Proposed techniques face glitches in natural language processing sine each language needs its own methodology and rules which vary from language to language such as Arabic language which is rich in synonyms, morphology and diversity of dialects [1]. Furthermore, when referring to web sites that provide a broadcasting service for Twitter hash tags that reaches trends as in Fig. 1, for example, clarify that most of popular trends may be obsessed with uncommon words that may not reflect the Twitter trend contents. Hence, different topics that might be found in a trend may lead commercial companies to post in unwanted hash tag trend Therefore, knowing the key words of Twitter within the trends may solve raised confusion. In addition, identifying the accurate search query to accurately measure the public response to specific topic is crucial for the interested individuals or companies.

45 minutes ago	1 hour ago
#التعاون_النصر#	#الاتحاد_احد#
#OTDirecto30E	#FelizMartes
#AntManandtheWasp	#OTDirecto30E
#FelizMartes	#AntManandtheWasp
#الاتحاد_احد#	#TuesdayThoughts
John Wall	Berkshire Hathaway
Mahrez	Toisón de Oro
Berkshire Hathaway	Raquel Espinosa
Raquel Espinosa	Aymeric Laporte
Dourado	양상블

Fig. 1. Twitter worldwide trends (<https://trends24.in/>)
Retrieved: 30-01-2018.

Therefore, a hot area of research is to develop a methodology that identifies key phrases and synonyms in a manner that is adapted to the change in the interests of the public that may retrieve more relevant and relevant texts then develop a mechanism to suit the written language processing.

This research is concerned with finding a mechanism that will quest for better key words and phrases (seed words) that can better retrieve the related tweets in a way that ensures continuous follow-up in the public opinion change. Twitter as the most famous social networking in Saudi Arabia will be used as the platform of analysis and experiment. Experiments of this research will test the proposed approach using seed words from Arabic language then report precision, recall and accuracy of classification.

The rest of the paper is organized as follows: Section II overviews the work related to sentiment analysis in Arabic. Section III describes the data collection and dataset construction. Section IV includes a description of the proposed methodology for mining the trending hashtags. Section V describes the experimental setup. Section VI presents the results of the evaluation of the proposed approach and Section VII concludes the paper.

II. RELATED WORKS

Sentiment analysis attracts researchers to focus on mine the tremendous amount of available information, discussions, texts, reviews and opinions in the digital form [2]. Recently, opinion mining or sentiment analysis has been used widely for various purposes due to its encouraging findings. Habitually, sentiment analysis is used in the business sector to measure the public response and points of view [3]. Government agencies, business sectors and research agencies are using sentiment analysis for better and indirect understanding of public opinion in specific service, product or suggestions. Sentiment analysis shows promising results to its efficiency and less-cost compared to traditional survey methods. Various articles

investigated the use of sentiment analysis for different social networks such Facebook and Twitter and apply sentiment analysis for different languages such as English and Arabic. Since microblogging such as Twitter has been introduced relatively recently, view researches have been investigated sentiment analysis with increasing attention [4]. Other research highlights the importance to do sentiment analysis for unstructured information in social network such as Facebook [5]. In [6] author develop a sentiment prediction model to investigate the added value of auxiliary data, including leading, lagging information, and traditional post variables on Facebook posts.

Each natural language has its own rules, lexicons and morphology and diversity of dialects. Hence, sentiment analysis should consider the difference in morphology and rules to get the desired results. Sentiment analysis has been investigated sufficiently in English language resulting in two derived methodologies: Corpus-based and lexicon-based [7]. Arabic language includes a huge number of words synonyms resulting of data sparsity [8]. Various methods for subjectivity and sentiment analysis for Arabic language has plotted in [9]

SemEval-2017 Task4 [10] describes the fifth year of the Sentiment Analysis in Twitter task as a competition for 48 teams in both English and Arabic languages. In [11] author presents NileULex, which is an Arabic sentiment lexicon containing thousands of Arabic terms and compound phrases. A semantic approach has been developed to extract the user outlook from social networks in Arabic language investigating both regular and Arabic dialects to announce an Arabic Sentiment Ontology (ASO) includes various terms that describes how robust extracted terms express the feelings [12]. Expressions/proverb phrases lexicon has been used to advance sentiment analysis polarity in Arabic sentences [13].

AdaMC was presented to boost the best accuracy of sentence-level negative/positive classification via adaptive Multi-Compositionality [14]. This research focus in finding an approach to find the best seed words that express the public opinion keeping in mind the frequent change over time.

III. DATA COLLECTION

Millions of tweets are posted every day in different languages as a response to popular hashtags. Twitter hashtags attract public attention immediately at particular times becomes trending hashtags. The main aim of the research is to mine the Arabic trending hashtags to sense public response to particular trend hashtag. A trending hashtags dataset is needed to assess the proposed approach. To construct Twitter data set that can be analysed to evaluate the proposed methodology, following steps followed:

1) For the purpose of obtaining a large number of tweets for experiments with the potential to produce clear results, a site <https://trends24.in/> was used – as suggested by SemEval-2017 [10] - to identify the trending topics of interest in the search interval of time. Further supporting information such as filtering based on the geographical location will be of extra benefit.

- 2) Subscribe to Twitter API Streaming service provided by Twitter. Such registration will give you access token that permit you to download the tweets containing the search query.
- 3) Using Python code, create a session to download tweets texts (along with extra supporting information) then save retrieved tweets to a file. The limit of the number of tweets to be inspected for each search query was set to 100000.
- 4) Do tweets pre-processing to remove duplications and URLs from the tweets contents.
- 5) Only topics with tweets more than 100 are retained for the dataset constructions.

IV. METHODOLOGY

The research aims to provide a methodology that can help better understand and measure the public response or opinion on a particular topic by analysing their sensation expressed in social networks. Twitter is chosen for its popularity. Twitter streaming API was used to search and retrieve the trending hashtags. The proposed methodology can be described as in Fig. 2 and as follows:

- The search query (mother seed) used to extract the tweets of a hash tag trend to look for candidate seed words representing the public opinions of a desired topic. Additional factors may give a more subtle twist when used, such as the geographic location of Tweets, which may sometimes be mentioned in some tweets and can be added as key words.
- After that, tokenize the tweets into words, exclude the very common words that cannot distinguish the subject from another stop words. Microblogs differs from regular texts since microblogs includes noisy text blocks, therefore, TF-IDF proves ability to extract keywords from microblogs [15]. Stem the remaining words, then weight them using weighting techniques such as TF-IDF.
- Considering that the public feelings change over time and the trend might contains altered topics, cluster the tweets into groups to formulate topics profiles within a trend. Each cluster (topics profile) includes weighted words as mentioned in the previous step.
- Sort the weighted words according to their weights (highest to lowest) as Algorithm 1 directs, then the top five words are picked to be the surrogate seed words that better represent the tweets' topic according to the users' interests and change over time.
- Topic profiles are mined using sentiment analysis technique such as AYLIEN Text Analysis. Sentiment analysis classify the tweets into objective (contains factual information) or subjective (useful for describing the opinion or feelings about specific topic).
- Finally, apply several classification methods to calculate the precision, recall and accuracy. The classification method with the highest accuracy is chosen.

Algorithm 1: Build Surrogate Seed Words

Input: Main Keyword (Mother Seed Word k), Seed-List S

Output: Adaptive Seed Words

```
1:      Function Build-Adaptive-list ( $k$ )
2:       $C \leftarrow$  Extract candidate words related to  $k$ 
3:      for ( $w \in C$ ) do
4:          Weight words
5:       $C \leftarrow$  Sort words
6:       $S \leftarrow$  Choose highest 5 words
7:      Return  $S$ 
```

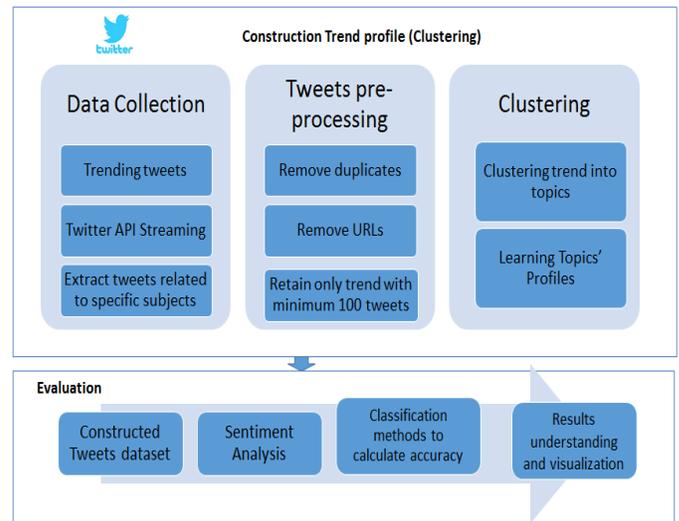


Fig. 2. The proposed approach.

V. EXPERIMENTS

This section describes the experimental set up and settings to give a clear understanding for the readers.

- 1) First, as proposed by SemEval-2017 [10] using <https://trends24.in/> to explore the recent hash tag trends while setting the region as Saudi Arabia.
- 2) Second, for each hash tag trend
 - a) Use Rapidminer studio to retrieve the tweets posted related to that trend using *Twitter search operator* which requires setting the access token sent by Twitter during subscription on Twitter streaming API. Write retrieved tweets to a file for later processing. The saved file contains the tweets in addition to some details {*created at, from user, from user ID, to user Id, language, source, text, ID, geo-location latitude, geo-location longitude, ID, retweet count*}.
 - b) Remove tweets duplication.
 - c) For each tweet text, remove retweet symbol and links (-http).
 - d) Use Support Vector Clustering technique to cluster the collected tweets of a specific trend into groups represent topics of related seed words.

- e) Process each cluster: tokenize the texts into words, remove stop list, stem the words and then weight them.
- f) Learn the topic profile by sorting the weighted words, then pick the top five words as the surrogate words best represent the topic profile.
- g) Do sentiment analysis. The following attributes added to the file {polarity, subjectivity, polarity_confidence, subjectivity_confidence}.
- h) Apply deep learning as a multi-layer feed-forward Artificial Neural Network (ANN) to build a prediction model helps to classify new tweet into the appropriate topic profile. Deep learning. Deep learning permits computational system represented by many processing layers to pick up data exemplifications with various stages of generalization [16].
- i) Use 10 k cross validation to test the prediction accuracy of the developed predict model. Cross validation splits the dataset into k subsets whereby each time one subset is used for testing and the remaining k-1 subsets are used for training. The average value for all the k experiments are used as the validation value.

VI. RESULTS AND DISCUSSION

Keeping in mind that the hash tag might contains more suitable seed words that better represent the topics, the evaluation of the proposed approach first cluster the tweets then do the sentiment analysis finally develop a prediction model to classify the new tweet to the class belongs to wither subjective or objective. Precision, recall and accuracy have been used to test the proposed approach efficiency. Precision, recall and accuracy are usually used to assess and relate various detection algorithms [17]. For better understanding of the evaluation method in relation to the sentiment analysis of tweets, precision, recall and accuracy has been defined as follows:

Precision measures the portion of relevant tweets among the truly positive and false positive retrieved tweets on a specific class:

$$\text{Precision} = \frac{TP}{(TP + FP)}, \text{ where } TP \text{ is truly positive and } FP \text{ is false positive.} \quad (1)$$

Recall measures the portion of retrieved relevant tweets over the total amount of relevant tweets:

$$\text{Recall} = \frac{TP}{(TP + FN)} \text{ where } TP \text{ is truly positive and } FN \text{ is false negative.} \quad (2)$$

Accuracy is the = $\frac{TP+TN}{(TP+TN+FP+FN)}$ where TP is truly positive, FP is false positive, TN is truly negative and FN is false negative. (3)

Test analysis classifies the texts into two categories: facts and opinion [18]. Facts are “objective expressions about entities, events and their properties” while opinion “are subjective expressions that describe people’s sentiments,

appraisals or feelings” [19]. Several experiments were conducted on various trending hashtags but only trends with more than 100 tweets were reported. Tables from I to VIII report the precision and recall for retrieved four highest hash tags trends investigated in Arabic language from the website <https://trends24.in/>. Sentiment analysis for tweets normally classify the tweet into either objective or subjective. Subjectivity feelings of opinions can be used to tell about two concerns: people’s feelings and language expressions used to describe that feelings [19] while objectivity describes some factual information. Tables illustrate the efficiency of the proposed approach (clustering the trending hashtags into profiles then do sentiment analysis for the tweets’ texts) against the usual sentiment analysis (do sentiment analysis for tweets’ texts without learning trends profiles first).

TABLE I. PRECISION FOR HASH TAG1

اي جهاز تفضل ايفون او سامسونج Which one you prefer, iPhone or Samsung	Class Precision (Normal Sentiment Analysis)	Class Precision (Proposed Approach)
pred. subjective	0.00%	0.00%
pred. objective	97.10%	91.18%

TABLE II. RECALL FOR HASH TAG1

اي جهاز تفضل ايفون او سامسونج Which one you prefer, iPhone or Samsung	True Subjective	True Objective
class recall (Normal Sentiment Analysis)	0.00%	100.00%
class recall (Proposed Approach)	0.00%	98.41%

TABLE III. PRECISION AND RECALL FOR HASH TAG2 (NORMAL SENTIMENT ANALYSIS)

يصلحه لعالم ما يفسده What the world destroy, name a thing that fix it	True Objective	True Subjective	Class Precision
pred. objective	149	12	92.55%
pred. subjective	26	6	18.75%
class recall	85.14%	33.33%	

TABLE IV. PRECISION AND RECALL FOR HASH TAG2 (PROPOSED APPROACH)

ما يفسده العالم يصلحه What the world destroy, name a thing that fix it	True Objective	True Subjective	Class Precision
pred. objective	154	12	92.77%
pred. subjective	21	6	22.22%
class recall	88.00%	33.33%	

TABLE V. PRECISION AND RECALL FOR HASH TAG3 (NORMAL SENTIMENT ANALYSIS)

اسم اللي تحبه بالرموز Name the one you love with symbols	True Objective	True Subjective	Class Precision
pred. objective	207	15	93.24%
pred. subjective	67	4	5.63%
class recall	75.55%	21.05%	

TABLE VI. PRECISION AND RECALL FOR HASH TAG3 (PROPOSED APPROACH)

اسم اللي تحبه بالرموز Name the one you love with symbols	True Objective	True Subjective	Class Precision
pred. objective	229	18	92.71%
pred. subjective	45	1	2.17%
class recall	83.58%	5.26%	

TABLE VII. PRECISION AND RECALL FOR HASH TAG4 (NORMAL SENTIMENT ANALYSIS)

الشباب الاهلي Ahli_Youth	True Objective	True Subjective	True?	Class Precision
pred. objective	344	35	93	72.88%
pred. subjective	18	58	98	33.33%
pred. ?	142	42	269	59.38%
class recall	68.25%	42.96%	58.48%	

TABLE VIII. PRECISION AND RECALL FOR HASH TAG4 (PROPOSED APPROACH)

الشباب الاهلي Ahli_Youth	True Objective	True Subjective	Class Precision
pred. objective	495	36	93.22%
pred. subjective	9	99	91.67%
class recall	98.21%	73.33%	

Table IX reports the accuracy calculated for the four hash tags illustrating the number of tweets retrieved for each hash tag and the number of topics profiles (clusters) for each hash tag. The following observations obtained from Table IX:

1) The performance of the normal sentiment analysis is higher than the proposed approach only for the first hash tag. The reason beyond that can be explained due to the small number of tweets retrieved for that hash tag (96 tweets) compared to the remaining three hash tags. Hence, results prove the assumption mentioned in the methodology that we retain only hash tag with a minimum of 100 tweets.

2) Notably as illustrated by hash tag from 2 to 4, the more tweets we retrieve in a hash tag the better accuracy we get.

TABLE IX. ACCURACY FOR THE 4 HASH TAGS

#Trend	Normal Sentiment Analysis Accuracy	The proposed Approach Accuracy	No. of Tweets	No. of Clusters
اي جهاز تفضل ايفون او سامسونج Which mobile phone you prefer, iPhone or Samsung	97.14 %	90.00 %	69	5
ما يفسد العالم يصلحه What the world destroy, name a thing that fix it	80.32 %	82.84 %	194	13
اسم اللي تحبه بالرموز Name the one you love with symbols	72.08%	78.57 %	294	14
الشباب الاهلي Ahli_Youth	61.05 %	92.96 %	1100	42

VII. CONCLUSION

Mining the tremendous amount of text, information, posts and customers' comments is essential to extract the desired knowledge. Companies might survey customers via traditional methods like questionnaires and interviews which is time consuming, costly to attract large sample size and people might respond incorrectly for different reasons. Sentiment analysis is a technique helps to understand and measure the targeted folk's opinions. The proposed approach aims at learning topics profiles helps to better understand the public response to a particular service, product or feedback by analysing recent Twitter hash tags trends. Occasionally, some hash tag titles is not understandable or misleading to tweets that represents different topics. To solve such problem, clustering of tweets was proposed to learn topics profiles. Recent Arabic hash tag trend-as listed by trends website announcer- were retrieved then the proposed approach was tested on the tweets of popular hash tags that becomes trends. Results show that the more tweets retrieved for a hash tag, the more groups or cluster (topics profiles) leading to enhanced sentiment analysis. Applying deep learning, findings show that the accuracy of the proposed approach is better than the normal sentiment analysis. As a future work, further investigation in using the proposed approach to automatically use the learned topics profile in each hashtags to retrieve similar topics.

REFERENCES

- [1] Sallab AA Al, Baly R, Hajj H. "Deep Learning Models for Sentiment Analysis in Arabic". ANLP Work. 2015;9-17. Available from: <http://www.aclweb.org/anthology/W15-32#page=21>
- [2] Liu B. "Sentiment Analysis and Opinion Mining". Synthesis Lectures on Human Language Technologies.2012;5:1-167. Available from: <http://www.morganclaypool.com/doi/abs/10.2200/S00416ED1V01Y201204HLT016>.
- [3] Neuendorf KA. "The Content Analysis Guidebook" (Second Edition). 2017. Available from: <http://academic.csuohio.edu/kneuendorf/SkalskiVitae/SkalskiNeuendorfCAjigas17.pdf>
- [4] Patodkar VN, IR S. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining". IJARCCCE [Internet] 2016;5:320-2. Available from: [http://ijarccce.com/upload/2016/december-16/IJARCCCE 74.pdf](http://ijarccce.com/upload/2016/december-16/IJARCCCE%2074.pdf)
- [5] Dasgupta SS, Natarajan S, Kaipa KK, Bhattacharjee SK, Viswanathan A. "Sentiment analysis of Facebook data using Hadoop based open source technologie's". In: Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015. 2015.
- [6] Meire M, Ballings M, Van den Poel D. "The added value of auxiliary data in sentiment analysis of Facebook posts". Decis. Support Syst. 2016;89:98-112.
- [7] Assiri A, Emam A, Aldossari H. "Arabic Sentiment Analysis: A Survey". Int. J. Adv. Comput. Sci. Appl. 2015;6:75-85.
- [8] Abdul-Mageed M, Diab M, Kübler S. "SAMAR: Subjectivity and sentiment analysis for Arabic social media". Comput. Speech Lang. 2014;28:20-37.
- [9] Korayem M, Crandall D, Abdul-Mageed M. "Subjectivity and Sentiment Analysis of Arabic: A Survey". In: Communications in Computer and Information Science. 2012. page 128-39.
- [10] Rosenthal S, Farra N, Nakov P. "SemEval-2017 Task 4: Sentiment Analysis in Twitter". In: Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017). 2017. page 502-18. Available from: <http://aclweb.org/anthology/S/S17/S17-2088.pdf>
- [11] El-Beltagy SR. "NileULex: A Phrase and Word Level Sentiment Lexicon for Egyptian and Modern Standard Arabic". to Appear Proc. Lr. 2016 2016;2900-5.
- [12] Tartir S, Abdul-Nabi I. "Semantic Sentiment Analysis in Arabic Social Media". J. King Saud Univ. - Comput. Inf. Sci. 2017;29:229-33.

- [13] Ibrahim HS, Abdou SM, Gheith M. "MIKA: A tagged corpus for modern standard Arabic and colloquial sentiment analysis". In: 2015 IEEE 2nd International Conference on Recent Trends in Information Systems, ReTIS 2015 - Proceedings. 2015. page 353–8.
- [14] Li Dong FWMZKX. "Adaptive Multi-Compositionality for Recursive Neural Models with Applications to Sentiment Analysis". 2014;1–7.
- [15] Simsek A, Karagoz P. "Sentiment enhanced hybrid TF-IDF for microblogs". In: Proceedings - 4th IEEE International Conference on Big Data and Cloud Computing, BDCLOUD 2014 with the 7th IEEE International Conference on Social Computing and Networking, SocialCom 2014 and the 4th International Conference on Sustainable Computing and Communications, SustainCom 2014. 2015. page 311–7.
- [16] Lecun Y, Bengio Y, Hinton G. "Deep learning". Nature 2015;521:436–44.
- [17] Zhu M. "Recall, precision and average precision". Dep. Stat. Actuar. Sci. 2004;1–11. Available from: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Recall+Precision+and+Average+Precision#0>
- [18] Pang B, Lee L. "Opinion Mining and Sentiment Analysis". Inf. Retr. Boston. 2008;2:271–8. Available from: <http://www.aclweb.org/anthology/P/P04/P04-1035.pdf> <http://portal.acm.org/citation.cfm?id=1218990> <http://www.cs.cornell.edu/home/llee/omsa/omsa-published.pdf> <http://arxiv.org/abs/cs/0409058v1>
- [19] Liu B. "Sentiment Analysis and Subjectivity". Handb. Nat. Lang. Process. 2010;1–38. Available from: <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf> [http://people.sabanciuniv.edu/berrin/proj102/1-BLiu-Sentiment Analysis and Subjectivity-NLPHandbook-2010.pdf](http://people.sabanciuniv.edu/berrin/proj102/1-BLiu-Sentiment%20Analysis%20and%20Subjectivity-NLPHandbook-2010.pdf) <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>