

An Evaluation about the Queries for Opinions of Financial Market Professionals using Resources based on the Web 2.0 Technology

Bora UĞURLU^{1*}, Esmâ YENİSARI¹, Bahadır KARASULU¹ and Özcan Zafer AYAN¹

Department of Computer Engineering
Canakkale Onsekiz Mart University, Turkey

boraugurlu@comu.edu.tr

esmay@comu.edu.tr

bahadirkarasulu@comu.edu.tr

ozcanzaferayan@gmail.com

Motivation

- Recommendation of a knowledge discovery model using Turkish natural language processing (NLP) in order to response to meaningful queries and classify with a high accuracy.

Introduction

- The main goal of our study is to develop an user-friendly financial knowledge discovery system based on a knowledge discovery model.
- This system's aim is to explain to user the relationship between linguistic data from resources using Web 2.0 technologies and numerical data obtained by trend analysis of stock markets.

Related Work

- Asur and Huberman (2010) showed how social media content may be used to highlight the facts of real-world.
- Bollen et al. (2010) used text contents obtained from daily Twitter posts in order to measure mood state changes caused by individuals' behaviors in behavioral economy field.
- Drury (2013) proposed an approach that uses knowledge. It is able to infer when the prices will rise and drop.

Materials and Methods

- We implemented our study in five steps.
- Data are obtained from the web, and then, we assess the data about the membership degree of related category.
- In our experiments, our tool is fed with a pool of 200 tweets and 200 RSSs.
- Our tool can implement NLP indexing, scoring and ranking processes.
- The search results are obtained as a reference to the match between the keywords in the corpus and classification of the model.

- Two independent human experts marked the entries on the corpus that are grouped into four main categories (i.e., ground truth annotation).
- In the study, four main categories are determined. They are given as "döviz" (i.e., foreign currency), "altın" (i.e., gold), "borsa" (i.e., stock market), and "petrol" (i.e., oil).
- When a datum is not match with any above mentioned category, this datum is classified into “ekonomi” category (i.e., economy).

- Frequency changes in RSS feeds belonging to financial professionals are weekly periods from November 5, 2012 to March 5, 2014. However, due to instantaneous and frequent updates in Twitter resources, time interval of our Twitter data is from January 13, 2014 to March 10, 2014.
- We gathered the same number of Twitter and RSS feeds from prior mentioned professionals and extracted information from them for testing. 85 out of those 200 tweets and RSSs were taken into account by not considering their releases on Internet.

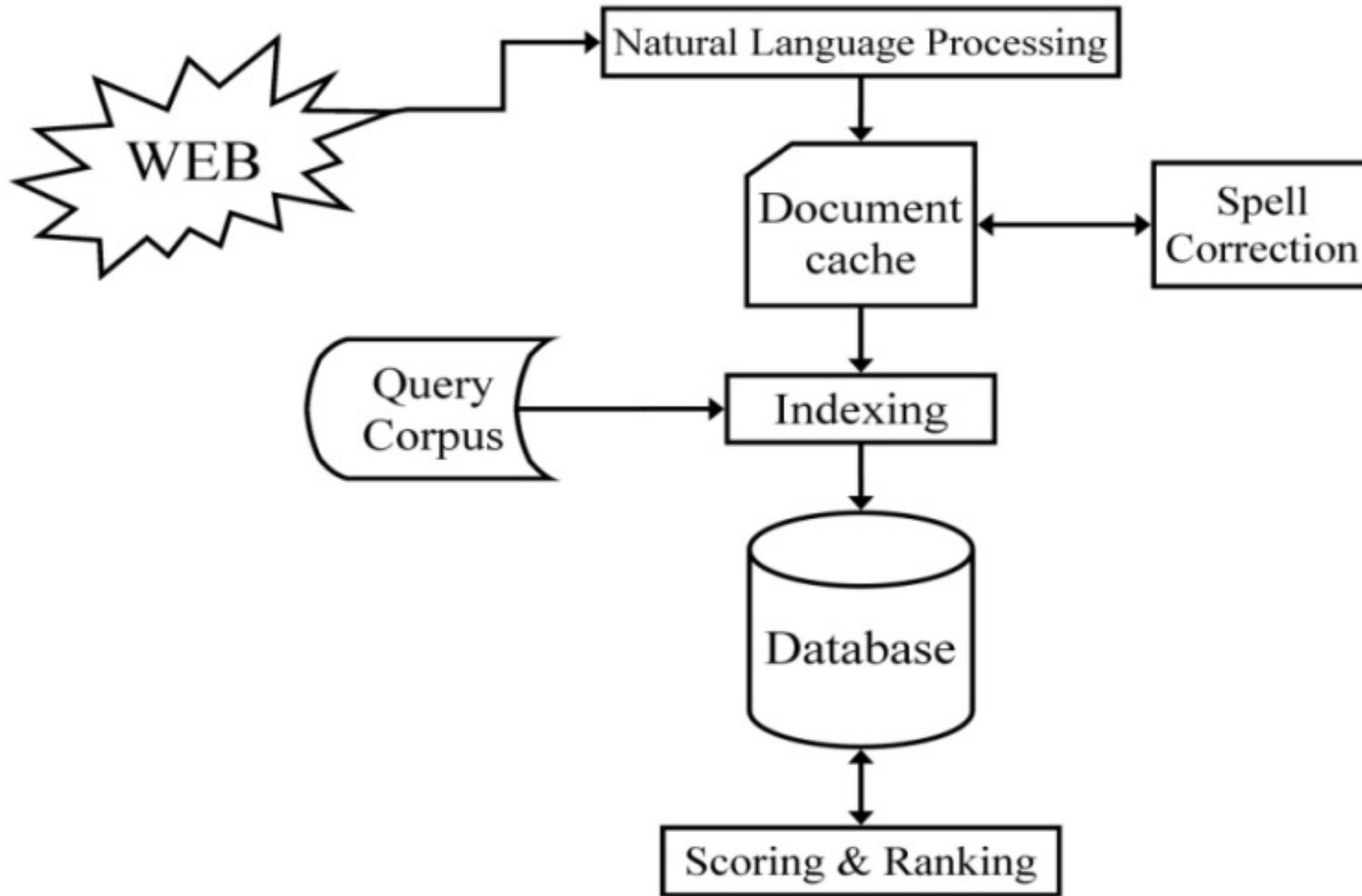


Figure 1. The data flow chart of proposed model.

- According to the steps in Figure 1, the generated inputs from the data set is parsed using NLP into wordy phrases through application program interface (API) called NZemberek (2010).
- Data set is stored as a "document cache".
- This corpus consists of some keywords which are distributed five categories.

Table 1. Selected 15 Turkish keywords from query corpus.

Enflasyon, kur, döviz, finans, BIST, dolar, tahvil, kriz, lira, ekonomi, petrol, bütçe, yatırım, vergi, sermaye
--

- In the indexing step, according to the query results and separated data in "document cache", indexing process is done according to date and categorical identifications (IDs).
- Then, indexing results and their dataset are stored to the database.
- Finally, the most suitable match for related categories are obtained, and also, the related results are reflected to the database.

Results and Evaluation

- In the scope of information retrieval, we used the balanced accuracy in our quantitative evaluation.
- The ground truth annotation is based on the marking made by two independent human experts.

Table 2. Statistical assessment of query results.

Categories	RSS	Percentage	Twitter	Percentage
Borsa	35	41.17%	20	23.52%
Döviz	22	25.88%	10	11.76%
Altın	17	20.00%	9	10.58%
Petrol	0	0.0%	4	4.70%
Ekonomi	11	12.94%	42	49.41%

Table 3. Quantitative evaluation based on performance measurements.

Metrics	Precision		Mean	Recall		Mean	F-measure		Mean	Balanced Accuracy		Mean
Categories	RSS	Twitter		RSS	Twitter		RSS	Twitter		RSS	Twitter	
Borsa	0.76	0.28	0.52	1.00	1.00	1.00	0.86	0.44	0.65	0.79	0.94	0.86
Döviz	0.81	0.90	0.85	0.69	0.75	0.72	0.75	0.81	0.78	0.68	0.81	0.74
Altın	0.5	1.00	0.75	0.5	0.62	0.56	0.5	0.76	0.63	0.63	0.75	0.69
Petrol	0.0	0.75	0.37	0.0	0.75	0.37	0.0	0.75	0.37	0.45	0.73	0.59
Total Average	0.62			0.66			0.60			0.72		

Conclusion

- The main contribution of this study is to develop a knowledge discovery model for querying Turkish financial market professionals' opinions.
- There is not any match in the query corpus which is generated by considering 16 months period of RSS feeds for "petrol" category due to weekly columns of professionals. Consequently, the proposed model implemented via our software tool cannot identify the financial situation of “petrol” category.

- The main reason for this that, the underlying algorithm is not able to categorize the phrases in the RSS feeds as “petrol”.
- According to the comparison based on the performance metrics, best performance results are obtained for “döviz” and “borsa” categories.
- While the number of financial professionals and of tweets and RSS feeds is not sufficient for a large number of categories, it is large enough for limited number of categories to build an expository model for knowledge discovery.

References

- Asur, S., & Huberman, B.A.(2010). Predicting the Future With Social Media, 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.
- Bollen, J., Mao H., & Zeng, X. (2010). Twitter Mood Predicts the Stock Market, Journal of Computational Science, vol. 2, no. 1, (pp. 1-8).
- Drury, B.(2013). A Text Mining System for Evaluating the Stock Market's Response to News, Phd. Dissertation, University of Porto, Faculty of Science, Doctorate Dissertation, 162p.
- Hammerley, B. (2003). Contents Syndication with RSS, O'Reilly Media Publisher,224p., ISBN:978-0-596-00383-8.
- Manning, C.D., Raghavan, P., & Schütze, H. (2009). Introduction to Information Retrieval, Cambridge University Press, New York,NY, 2009, Draft Online Copy (2009.04.01). Retrieved March 13, 2014 from <http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>
- NZemberek - Turkish Natural Language Processing. Net API, (2014), Retrieved March 13, 2014 from <https://code.google.com/p/nzemberek/>

Thanks!
Questions?