International Journal of Database Theory and Application
Vol.9, No.7 (2016), pp.9-18
http://dx.doi.org/10.14257/ijdta.2016.9.7.02

Search Ranking Utilizing User's Opinion

Jun Jin Choong and Jer Lang Hong

School of Computing and IT, Taylor's University, Malaysia
junjin.choong@taylors.edu.my, jerlang.hong@taylors.edu.my

Abstract

The Internet is one of the most widely available services in the world today. With the Internet, people are now looking for reviews on the Internet; more specifically, the social networking services. Within the social network medium, we can identify a suitable service that describes more about a person's personality as the subject. The growth of social networking popularity has contributed to the in-crease in information available on social networking services. The flexibility of these services allows writing individual thoughts without restrictions. With the vast information available on social networking sites today, how is it possible to look all of these opinions? How do we know which opinion holds truth? How do we know if someone is not bias based on his writing? Hence, it is seen necessary to filter opinions. In this paper we look at the possibility of using search ranking as a medium of filter opinions by exploring opinion mining methods, social net-working candidates and search ranking methods. With existing sentiment analysis techniques, we can obtain opinions that are then ranked against a set of key-words.

Keywords*: Search Ranking, Opinion Mining, Web Search*

1. Introduction

The idea of searching, ranking, opinion mining and social networks are not new in today's research [5]. Generally speaking the possibility of opinion mining itself is endless. With many techniques that has been thoroughly studied and prototypes built, performing search ranking on a result data mined poses a challenge that ought to be considered. For example, given a carpooling system that uses the social net-work, how would it be possible to know who to trust from opinions that have been written by that particular person? Is that person good or bad at handling cars? Is he or she a cautious driver? Why should I carpool with him or her? These are questions that search engines cannot answer as they are subjective in nature. This may not be the case for data mining; in particular opinion mining. For instance, in Google search ranking has proven to be a useful suggestion to answers that relate to factual ques-tions. Most often than not, the first result from Google is usually what we are looking for. Should similar concept apply to opinion mining on social networks, we can per-form conclusive searches that answer the previous questions.

If Google can answer factual questions, what about subjective questions? If we ask Wolfram Alpha a mathematical question, there would always be an answer. Subjective questions may not be the same. Instead, there could be various possibilities when determining the process for answering such questions. Such complexity is defined as a complex problem in our case. A complex problem may require thorough investigation, researching before ending up on a decision. Most of the time, these are factored to scarce resources such as time and money. Assuming that most decision making can be simplified, perhaps these scarce resources could be put to better use. Therefore, we believe that the very first step for such achievement can be done with search ranking and opinion mining. Our domain of focus would be on social networks which we believe is the glue of most sparse data (opinions, reviews).

ISSN: 2005-4270 IJDTA

2. Related Work

2.1. Opinion Mining General Concept and Background

Opinion Mining or Sentiment Analysis is a Natural Language Processing technique (NLP) that analyses sentiment within a text [2]. Most technique revolves around the idea of sentiment polarity. This separates texts or words into positivity, negativity and sometimes neutrality. The process may look simple; but the process is complicated with many general challenges that have been investigated thoroughly by researchers. Opinion mining is subject of its own as there are many areas and possibility that needs to be covered to perform satisfactory segmentation when classifying opinions as positive or negative. Most opinion mining technique focuses on textual information, which could be classified broadly into two main categories, facts and opinions [3]. Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect people's sentiments or perceptions about the entities and events. Much of the existing research on text information processing has been (almost exclusively) focused on mining and retrieval of factual information, *e.g.*, information retrieval, Web search, and many other text mining and natural language processing tasks. Little work has been done on the processing of opinions until only recently.

Back then, the studies of opinions are less fare due to the lack of opinionated text before the World Wide Web. However, today's trend has change tremendously with information growing exponentially, mostly user generated content. One of the reasons is because the World Wide Web is an open medium. Anyone with an Internet access can create their own content (websites). With services like wordpress.com, users can focus less on the technicalities and instead focus on the content. One of the many use cases can be seen especially with the explosive growth of content. For instance, when an individual needs to make a decision, he/she typically asks for opinions from friends and families. When an organization needs to find opinions of the general public about its products and services, it conducts surveys and focused groups [4]. Alternatively, with the web, users would use search engine services like Google to find what they are looking for. Ideally, he/she would want to be able to search for reviews of products at merchants' sites or perhaps express views of the product. Usually, forums, discussion groups, blogs, which are collectively called user generated content. Opinion mining is a great technique to use hand in hand with social networks as the key concept is to analyze subjectivity within the objective of opinion mining. We take a step further by describing how opinion mining can be used in search ranking by taking key concepts from opinion mining techniques. Using these concepts, we can factor them into consideration by ranking results in a particular search query [12].

2.2. Search Ranking Using Sentiment Analysis

When information is plenty, but the search results returned are irrelevant, the search engine has failed to perform. The idea of search ranking has been created by founders of Google during their days as a PhD student; although Jon Kleinberg proposed his algorithm about the same time. Search ranking is widely used in their product, Google which is popularized by the term "Page Rank" [1], [11]. It describes the relevancy of results against the search keyword. Therefore, improving search result accuracy and plausibility; improving overall searching experience for their users. Likewise, the same concept applies on top of opinion mining. If we can identify which opinions are relevant search, we can obtain similar results. In more detail, it performs a simple filter that describes more about opinions. Taking our previous ex-ample, a carpooling system that identifies opinions of its driver and their passengers; we can identify if a driver is trustworthy or the passenger is civilized. Notice that we used terms that describe about people. Another form of great example would be the use of search ranking to filter off irrelevant opinions. For instance, one would write reviews on a product but instead of criticizing the product, most complaints are about the publisher of the book. Such cases could still contribute to decision making, but it should not influence the outcome should someone asks “Is this book worth reading?” [8-9].

2.3. Decision Making Support

Decision making is a complex problem in most cases. However, in our research, we would classify these problems into a much broader category; namely simple problems and complex problems. The goal of decision making support is to assist users with their decision making process. For instance, “iPhone vs Android, which is better?” would rule out to be a complex problem. To simplify things, most of our discussion focuses on a small subset of complex problems. In computer science, decision making would probably be classified under fuzzy systems or fuzzy logic. In this case however, it’s not the decision of agents that we would like to assist, but rather humans themselves. If one would look up for something he/she is interested to buy and has limited resources (money), he/she would spend countless of hours before making this final crucial decision. If a bad choice is made, obviously money gone down the drain. Otherwise, the money is well spent.

One would argue, such feature resembles closely to the word ‘suggestions’. However, we would say that it’s more than just suggestions. It is technically more than that because suggestions would come by looking at a history of possibilities. Alternatively, the decision making support would merely be a mathematical formula that takes into account various factors. For instance, we could aggregate results and show a simple graph with the number of positive votes and negative votes. Perhaps if the user requires more information, a list of opinions can be shown to the users. In short, decision making support should ideally save you time and money. With the amount of information on the Internet these days, it could easily take a person countless hours of research to come to a conclusion to purchase information. With that, one of the goals of using search ranking is to provide a possibility of decision making.

3. Motivation of Using Social Network as Domain

The rise of social networking has increase the surge of interest in sentiment analysis. It is no doubt that social network has been the catalyst to sentiment analysis. Social networking services such as Facebook, Blogger, Youtube, Twitter *etc.* are widely used by users to share information. The flow of information propagates faster in social network services than any other comparable mediums such as printed newspapers. With speed comes at price. Information that propagates the social medium is not moderated. Hence, information written on the social network tends to be subjective in nature. Notably, this information is there to describe on a topic and share points of view with their respective audience. Due to its un-moderated nature, it is an ideal candidate for sentiment analysis. Often social networking services do not promote text searching. For instance, in Facebook searching focuses on looking up people and pages. Youtube on the other hand focuses on video title and description. There is no need to search on comments or other information. To business people, these are important factors that reflect upon their marketing strategy. “How do people like my product?” are often questions raised. Unlike reviews on niche oriented website, reviews are listed based on its usefulness to another buyer (eBay, Amazon), social networking services does not have such similarities. With exceptional cases, crowd-sourcing social networking sites such as Stackoverflow and Youtube or any other sites within the Stack Exchange network has voting ranking. Unfortunately these sites are restricted to factual answers and not opinions.

Social network is also an ideal candidate because of its potential to glue together thousands of information on a single service. In most cases, you can find content that you have never seen before on the internet and it's linked on the social network usually by sharing with friends. Indirectly, this has grouped sparse data into a single place where data can be searched. In addition to that, social networks are influential. Popularity goes with numbers of followers, friends or even likes. If we consider all these factors, we can use these variables to determine the strength of an opinion. For instance, opinions by celebrities would carry more weight than an average individual. Such cases could influence the outcome of decision making. Ironically, social net-works are not completely reliable. For instance, people may abuse social networks creating forging false information, posting fake reviews. Given that it's subjective in nature, it's understood that such behavior will occur especially on the Internet. Sometimes it's a matter of trust rather which coincides with reliability and credibility. In summary, subjectivity is the key to opinion mining and it can be widely found in social networking sites that are considered freedom of speech medium. People often put their thoughts to writing that relate topics such as humor, politics, products and many more.

4. Problem Formulation

From the concepts and definition of sentiment analysis, we get a glimpse of how sentiment analysis works. In addition, we discussed existing state-of-the-art techniques, double propagation as a method of identifying important features from a corpus. We also discussed the improvement of feature ranking mechanism used in the double propagation technique. From double propagation, we can see that the opinion lexicon is used to promote opinion mining towards a semi-supervised learning approach. Furthermore, double propagation allows us to identify the topic feature in a given corpus. The improvised feature on double propagation allows features to be more accurately extracted from the given corpus. The basis of the technique takes two different patterns namely, part-whole and "no" patterns to increase recall. It then ranks the extracted feature candidates by feature importance, which is determined by two factors: feature relevance and feature frequency. In terms of relevancy, a web page ranking algorithm was used (HITS) [6-7]. The frequency is applied using a computational algorithm to increase the ranking of feature. The improvised double propagation has successfully increased feature extraction precision.

Although feature extraction is an important aspect of opinion mining, it does not work on some social networking services. Facebook, for instance, consists of wall posts, comments that sometimes are complement with pictures. If wall posts contain no subject to talk about, feature extraction can be a practical way to reduce unwanted opinions (noisy data). However, what happens if wall posts contain meanings, but displayed as images? Unlike twitter, some short texts could be represented with a meaningful hash tag. Feature extraction from images is not easy. Recently in Google I/O 2013, Google+ has such feature built-in now. Given a photo with no text, Google+ is able to automatically assign hash tag to the related photo. In the next chapter, we propose an improvised HITS algorithm for social network search ranking. Before we introduce our proposed technique, we need to identify the problem with existing feature extraction techniques which create the need for search ranking. Therefore, to justify our proposed technique, we will first discuss on our discovery related to feature extraction.

5. Objectives

In our problem statement, we stated that search ranking is necessary to act as a filter when using opinion mining. The idea behind this allows people to go beyond asking Google factual questions such as "Who is the prime minister of Malaysia?" instead, questions like "How does the battery lifespan of the iPhone 5 perform?" can be answered using a search ranking algorithm. Imagine, answers given back in a statistical manner.

You would not need to spend time searching through thousands of newsfeed to conclude your decision for purchasing the iPhone 5. The potential of solving this problem would save each individual a lot of time for decision making. Although the domain could potentially be unlimited, to narrow down the scope, we would only discuss on social networking services. Hence, our technique needs to be as generic as possible. By the definition of generic, it should not only work on a single social networking service such as Facebook. It should ideally, be able to identify similar opinions across all social networking services and return an aggregate answer. To do so, one has to rely on method that could be plausible enough to distinguish opinion given are not fake or even bias. In addition, given that it has to work across different domains, we need to identify the common grounds between these social networking sites. We classify this as an extraction problem which will be discussed in the following section. However, in our proposed search ranking algorithm, we do not take this into account as there is only a single domain focus.

Within our objective, we will need to narrow our scope. There may be problems related to opinion mining which could affect our hypothesis, but most are out of our research scope. For instance, we need to identify a common problem within each social networking service simply because there are too many different types of social networking websites. Websites like Flickr socializes via photos and sites like Twitter socializes via text messaging. The next problem we discovered is the language barrier. What happens if someone posts an opinion which could be useful, but contain a mixture of languages? Therefore, our technique should realize such problem exist but not to solve it. Part of this problem relates to computer vision to identify objects. This is also an extraction problem which we classify it in next section. Assuming that we have solved the smaller problems of extraction and identification of a topic feature (such as using double propagation), we need formulate a suitable ranking algorithm. One way is to use existing solutions such as HITS or PageRank [10]. Alternatively, we could propose an improvement on existing solutions but adding additional vectors into consideration in the ranking algorithm. In the following section, we take a look at a few proposed examples that we think are important in addition to frequency and relevancy.

6. Problem with Existing Techniques

6.1. Data Extraction

We define extraction as the need to classify texts in their respective category. For instance, given the sentence “Is Android better than iOS?” we need identify what we are interested in. In this particular sentence, we are interested in the opinions related to Android and iOS written across the user’s Facebook feeds and comments. Initially we identified the possibility of reusing Double Propagation state-of-the-art technique proposed by Bing Liu and Qiu. The attractive idea of being able to extract features from a given corpus seems to be ideal choice. However, that is not the case for certain websites (Imgur, Facebook, Reddit *etc.*). Social networking websites may not have a given corpus but may vary in content presentation but may still be classified as one. Double Propagation is a technique meant for text mining, which technically meant that it’s unsuitable to be used on social networks presented in the form of images. Unfortunately, the trend has changed dramatically since bandwidth is getting cheaper and affordable; images being used to deliver opinions are no longer uncommon. Such trend can be seen from websites like Imgur and the infamous ‘Share’ button on Facebook.

Statistically, the trend of such websites has been increasing such that opinions can be seen not only expressed as texts, but with the possibility of pictures as well. Logically, we can address such problem using computer vision feature detection. We can also apply machine learning techniques in a semi-supervised learning manner. For example, we could take sample images of superheroes from movies such as The Dark Knight. In a semi-supervised learning, the machine will identify the edges of a particular superhero and differentiate the features from another superhero. In most cases, feature detection is sufficient to be our topic feature, but may not be enough. Moreover, there is a tendency that social networking sites have unstructured gram-mar or sometimes, a mixture of languages. For example, we can observe that a commenter posts a Romanized Japanese word '何' which meant 'What' in English. Such instances tend to fall under Natural Language Processing problem but can also be classified as sub-problem for opinion mining. After all, opinion mining depends on natural language processing to identify the sentiment polarity of text. Such problem are difficult to be addressed, and is out of our report scope. In this section, we have emphasized problems related to feature extraction but we have yet to state the cause and reason for this. Using feature extraction from opinion mining, it would contribute as a parameter (vector) in our search ranking algorithm. The higher the relevancy between the search query and the subject of the opinionated text, the more likely the opinionated text is of high quality content. Due to the nature of subjectivity, quality of content can only be considered by identifying the strength of the subject. Though, this is subject to argument.

6.2. Searching and Ranking Data

Extraction alone is insufficient to determine the outcome of the search query. As a starting point, we need somewhere to begin looking for opinions. Perhaps, we could consider using Google or Yahoo as our starting point. In this section, we discuss viable alternatives to prevent ourselves from re-implementing the wheel. There are other forms of searching that we need to consider as well. There are full text searches, topic based searches or even graph search. Recently, Facebook has introduced Graph Search as a beta feature to English (US) users. Alternatively, developers are allowed usage via Facebook's API. The semantic search engine allows typical users to look for more friends, photos or even people that are not related to you. The idea behind it is similar to opinion mining; using Natural Language Processing power to understand words (semantics) and crafting them in to search algorithms. This differs from Google which tries to match keywords. One of the missing features in graph search relates to searching within a newsfeed or comment. Nevertheless this can still be done with lesser granularity *e.g.* searching public posts.

7. Proposed Solutions

7.1. Overview

With HITS as the basis for our search ranking algorithm, we include an additional vector that can be obtained from opinion mining. This vector includes the overall polarity (positivity, negativity or neutrality) of a given corpus. To obtain this vector, an initial run of sentiment analysis algorithm is done on the given corpus. The result returned is the polarity of the corpus. In relations to HITS which uses authority and hub to rank the importance of a given page, we need to translate such vector from opinion mining to HITS [13]. In our solution we would classify an opinionated text as hubs and the authority as the owner of the text. There is more than one way to do this. For most social networking sites, a single user may have different types of friends. Among these friends, some would be reputable, smart, or possibly unknown. If taken into consideration of such status, we can formulate a similarity between HITS algorithm and our opinion based search ranking algorithm. Alternatively, we can also consider the number of shares, likes *etc.* Although useful, there is a sign of bias vector being included in the algorithm. What happens if an opinionated text is written with honesty from an unknown individual compared to popular individuals who wrote short and biased opinions? Truthfully, the popular individual will be given more weightage in terms of authority. In most cases, PageRank works the same way but with additional vectors such as content quality being a consideration. PageRank would give higher ranking to well establish sites like bbc.co.uk and then consider the links in that site (hubs) as important.

Unlike Google's PageRank, we cannot take quality of content as a vector; rather we consider the overall sentiment polarity as sufficient vector to denote that an opinion is of higher quality. Since polarity can be higher or lower, our next step is to normalize it. Consider that we have a total polarity equating to 1. We would consider 0.5 being neutral, < 0.5 being negative and > 0.5 being positive. The calculation to perform the normalization would be dependent on the type of opinion mining method used and its accuracy. Next, we would need to consider the scaling factor. Should the scale of opinion increase, we need another way to determine the existing standings of previous opinions. In this case, we would follow HITS algorithm. For each iteration, we will compute a normalization value for the authority and the hub. This would create a converging value. The iteration would stop upon reaching a converging value that is negligible of changes.

7.2. HITS Algorithm Revisited

To understand how our algorithm works, we need to reiterate the importance of HITS algorithm. Using HITS idea of authority and hubs, we can relate to which opinion carries higher importance over another opinion. The idea behind authority and hubs is a mutually reinforcing relationship. In the case of opinion mining, the terms are changed. A good hub is a comment/feed points to many good authorities; a good authority is a comment/feed that is pointed to by many good hubs. For example, consider the case where a comment on iPhone battery lifespan. John is a highly reputable user on the social network having many friends or followers comments on the iPhone battery lifespan. His comments are "retweeted" by many of his followers (Twitter) or shared by many friends (Facebook). The mutually reinforcing relationship states that such condition would provide John's opinion with higher weightage. Obviously, the vector to determine whether what is an authority and hub can differ in different social networking sites. In this case, Facebook could be number of shares, comments, likes or in the case of Twitter; it could be the number of retweets, or replies.

The method we used to determine the authority or hub is by using HITS algorithm. More precisely, we take the sum of authority and hub score from each node and compared them to see which is higher. The higher score for authority vector would be the authority node. By using HITS algorithm as our foundation, we have successfully connected the search term a relevancy. That is, given that a user has so many news feed or comments, we are likely to know which opinion we are looking for. In other words, we have performed a filter on all opinions that could be related to the search query. HITS algorithm has its limitations if we consider its usage in opinion based search ranking. We need to provide the algorithm with an extra vector that justifies this weakness. Hence, in the section, we propose simple strategy by first deducing the overall polarity of the opinionated text.

7.3. Sentiment Polarity as Vector

One of the benefits when using opinion mining is the underlying technique used. Opinion mining simplest technique could classify a given text in its overall sentiment polarity form given an opinionated text. In questions such as "Is the iPhone battery durable and lasting?" which we have used throughout our example, HITS alone was insufficient to justify a decision making. Looking back at the previous section, we conclude that an authority node is a highly trusted opinion. Subsequent shares, likes, comments from it would reinforce the relationship between the authority and hubs. What is missing is the weightage of a given opinion. With weightage taken into consideration, we can make decisions on opinions that are old, fake or even biased. Overall, we are only interested in the output value from the opinion. Additional factors would be included as considerations in an opinion mining technique; a filter in opinions itself. For instance, given the example text “iPhone has tremendous battery life!” we would run a sentiment analysis on it. Assume the function f(opinion, e) where e is the optional parameter for consideration such as date, length of opinion *etc.* The function should return a weightage in the form of normalized vector; between 0 and 1.

Equation 1:

$$f(opinion, date, length) = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Where x, y and z denotes the normalized value for opinion, date and length respectively such that x would denote the overall polarity of the text (x> 0.5 for positivity, x < 0.5 for negativity or 0.5 for neutrality). The remaining variables y and z could be a scale determine by the search engine. For example, saying y = 0.8 would mean that the date is of 80% accurate. Alternatively we could say that the importance of the date hereby can be determined by a scale of 0.8. The vector produced by the function would return 0.9 which denotes the positivity of the opinion, y which determines if the date is within exactly 2013-05-01 and finally z which tells the function that length can be anything with the input value being 0. This function provides a simple example on how we can go about identifying the weightage of the opinion with other factors into consideration. The assumption of implementing this function is such that there is a way to determine the polarity of the opinionated text. The method of determining is taken into consideration during opinion mining. Therefore, the additional overhead may be negligible.

8. Experimental Test

We conducted an experimental test on a wide range of datasets. We collected a sample of 500 opinion based pages. We use the term recall and precision to measure the effectiveness of our method. Recall and precision are calculated as follows:

Recall=Correct/Actual*100

Precision=Correct/Extracted*100

Correct depicts the number of pages correctly ranked. Actual is the actual number of pages to be ranked. Extracted depicts the number of pages ranked.

Table 1. Experimental Tests

Terms	Our method	PageRank	HITS
Actual	500	500	500
Extracted	494	485	481
Correct	488	386	347
Recall	97.60%	77.20%	69.40%
Precision	98.79%	79.59%	72.14%

We then benchmark our method with state of the art search ranking algorithm PageRank and HITS (Table 1). Experimental tests show that our method could achieve recall of 97.60% and precision of 98.79% while PageRank obtains recall of 77.20% and precision of 79.59%, and HITS obtains recall of 69.40% and precision of 72.14%. This is due to the fact that our method provides an additional parameter to rank the results. Furthermore, the input to rank our results is based on opinion based results, which is not supported by HITS and PageRank.

9. Conclusion

The aim of this paper is to hypothesize a possibility such that opinions can be ranked given the usage of opinion mining techniques either by indexing or real-time. The algorithm may or may not proof to be useful, it shows that there is a possibility of requiring ranking algorithms with the boom of user generated content on social networking mediums. Ideally, we would like to be able to ask the search engine and conclude our decision immediately (decision making support). By taking the very first step of linking both opinion mining and search ranking, we can see that there is a possibility to produce a system that could truly be of use to save time and money. In future possibilities, we look forward to investigating alternative search ranking algorithms that could potentially show its usefulness with opinion mining. Perhaps an algorithm that goes beyond considering authority and hubs or even probability of a user clicking a particular link.

Acknowledgement

This work is carried out within the framework of a research grant funded by Taylor's University Research Grant Scheme (Project Code: TRGS/1/2012/SOCIT/004).

References

[1] A. Alon and M. Tennenholtz, "Ranking Systems: The PageRank Axioms", (PDF). Proceedings of the 6th ACM conference on Electronic commerce (EC-05). Van-couver, BC. Retrieved 2008-02-05, **(2005)**.

[2] B. Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), **(2010)**.

[3] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), **(2002)**, pp. 79-86.

[4] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales", Proceedings of the Association for Computational Linguistics (ACL). **(2005)**, pp. 115–124.

[5] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine", Computer Networks and ISDN Systems, vol. 30, **(1998)**, pp. 107-117.

[6] F. Ayman, L. Faro, T. Miller, C. Joel, R. Gregory and A. W. Lesley, "Authority Rankings from HITS, PageRank, and SALSA: Existence, Unique-ness, and Effect of Initialization", SIAM Journal on Scientific Computing, vol. 27, no. 4, **(2006)**, pp. 1181-1201.

[7] K. Jon, "Hubs, Authorities, and Communities", Cornell University. Retrieved, 2008-11-09, **(1999)**.

[8] N. L. Amy and D. M. Carl, "Survey: Deeper Inside PageRank", Internet Mathematics, vol. 1, no. 3, **(2003)**.

[9] N. L. Amy and D. M. Carl, "Google's PageRank and Beyond: The Science of Search Engine Rankings", Princeton University Press. ISBN 0-691-12202-4, **(2006)**.

[10] L. Li, Y. Shang and W. Zhang, "Improvement of HITS-based Algorithms on Web Documents", Proceedings of the 11th International World Wide Web Conference (WWW 2002). Honolulu, HI. ISBN 1-880672-20-0, **(2002)**.

[11] P. Larry, "PageRank: Bringing Order to the Web", Stanford Digital Library Project, talk, (archived 2002), **(1997)**.

[12] P. Bo and L. Lillian, "Subjectivity Detection and Opinion Identification", Opinion Mining and Sentiment Analysis. Now Publishers Inc., **(2008)**.

[13] V. A. Luis, "Hubs and Authorities", (PDF). 15-396: Science of the Web Course Notes. Carnegie Mellon University. Retrieved 2008-11-09, **(2008)**.

