International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

Methods for Determination of the Weight of Documents in Electronic Resources

Khujaev O.K., Sh.B.Yusupova, M.R.Allaberganova

Abstract: ***This paper devoted to the issue of ranking electronic documents for e-government web resources. In the paper also analyzed related works for calculating resource weights in the ranking of web resources in the electronic government system and proposed the use of PageRank algorithm for calculating weight coefficients of web resources. Then is structured calculation steps for PageRank algorithm in e-government environment.***

Keywords**: Document ranking, ranking methods, normalization of weights.**

I. INTRODUCTION

Electronic Government (E-Gov.) - a way to provide information and assistance already formed a set of public services to citizens, businesses, other branches of government and state representatives, in which the personal interaction between the state and the applicant is minimized and use of information technology is as much as possible [1].

The main objectives of e-government are: 1.optimization of the delivery of government services to people and businesses; 2.support and empowerment of the Self-service of citizens; 3.growth of technological knowledge and skills of citizens; 4.Increased participation of all voters in governance and management of the country;
5.reducing the impact factor of geographical location. Over the years, various attempts were made by different countries to implement e-government, including the Republic of Uzbekistan. Only 15% of the implemented programs in the world to create successful e-government were recognized. The rest of the countries, having spent a lot of effort and financial resources, could not boast of increased efficiency and economy of time and money. In turn, the experience of countries that have achieved success in this field started being to be learned from all sides. [2] Standard e-government includes the development of the South Korean government, which in addition to the development of tools EG just shared with experience in this area.

Revised Manuscript Received on January 30, 2020.
***** Correspondence Author
Khujaev Otabek Kadambayevich*, Information Technologies Department, (Urgench Branch Of Tashkent University Of Information Technologies Named After Muhammad Al-Khwarizmi), Urgench, Uzbekistan, Email: Otabekhujaev@Gmail.Com
Yusupova Shokhida Botirovna, Information Education Technologies Department, (Urgench Branch Of Tashkent University Of Information Technologies Named After Muhammad Al-Khwarizmi), Urgench, Uzbekistan, Email: Gratifikus@Gmail.Com
Allaberganova Muyassar Allaberganova Muyassar Rimberganovna, Information Education Technologies Department, (Urgench Branch Of Tashkent University Of Information Technologies Named After Muhammad Al-Khwarizmi), Urgench, Uzbekistan, Email: Amuyassar83@Gmail.Com

Initiatives to create e-government in South Korea received legal registration in 2001. The developed vision of e-government included improving the efficiency of the administrative authorities and consisted of three phases [3].
Along with other countries, the actions of Uzbekistan should be noted. In improving the information sphere the resolution of the President Islam Karimov "On measures for further development of national information and communication system of the Republic of Uzbekistan" dated June 27, 2013, has a great importance, according to which the program was approved by the development of telecommunication technologies, networks and communications infrastructure in the Republic of Uzbekistan for 2013- 2020 [4].
As part of this decision at the State Committee of communication, information and telecommunications technology created new structures - the Centre of the "electronic government" and the Center for information security. Supported several projects in the field of e-government which include: gov.uz; nis.uz; e-kommunal.uz; my.gov.uz; id.uz etc. [5] [6].

II. STATEMENT OF THE PROBLEM.

To develop a search engine to work independently with all the resource base of e-government and providing access to the documents in EG sphere with the expansion of uz, as well as the application of methods for determining the ranking of the weight of the received documents.

III. SOLUTION OF THE PROBLEM.

To solve the problem, a number of methods have been examined of searching and ranking algorithms to provide a list of documents relevant to the user's search query. Ranking applied to the search engines is called sorting sites in the search results. As a rule, there are many factors to rank, among which are the site's ranking, number and quality of inbound links, the relevance of the text to the search query, and many others, on the basis of which search engine generates a list of sites in the search results [7] [10].

Ranging frequency words in the document.

The first method of ranking - ranking frequency words. Briefly, this metric can be described with the following words: the number of occurrences of words in the document specified in the query helps to determine its degree of relevance of the document. In other words, the documents in which the word specified in the query string is more common, and not to get documents in which the word is given through without affecting the search term [12].
We represent the set of documents as a set of X, the search query as a set Z, and rang the set storing the results of calculations rank document,

then the function calculating the rank for each document takes the form:

$$rang = \sum_{i}^{n} X_i \left[\sum_{l}^{m} Z_l \right] \quad (1)$$

where n - the number of documents in the calculation; m - number of words involved in a search query; Z_l- location keyword herein.

We briefly describe the algorithm of this metric:

1. The function takes an array of query results search.
2. Loops through each document listed in the array.
3. For each link count the number of repeated occurrences of the word in the request (1)
4. Add the results into a new array with the identifier URL.

This method differs from the simple use of a full-text search, as is a more logical way of sorting documents. As you can see, this method of ranking is relatively easy to implement. However, it is resource-intensive because it requires frequent reference to the data store, but this problem is easily solved if we competently approach the issue of constructing queries. At this stage of the experiment, we are not interested in the problems of this nature, because they separate us from the purpose of the experiment.

Ranking by the arrangement of the words in the document.

Next, consider a method which allows sorting the documents based on the location of words in the document. This method can theoretically provide high performance since in most cases a direct relevance of the document depends on the location of the desired keywords. That is if a keyword is situated close to the top or located in the header.

This method is not very different from that discussed above, but it has one major difference in the method of calculation. Here calculation function takes the following form:

$$rang = \sum_{i}^{n} X_i \left[\min \left(\sum_{l}^{m} Z_l \right) \right] \quad (2)$$

where the function min - returns only the minimum amount of occurrences of words in the document, and rang - provides a set of sums, the ranks of all documents. A closer look at the algorithm of this method:

1. The function takes an array of query results search.
2. In the loop through the array, summing the combination of occurrences of words, identifying the minimum amount (2).3. Further, the array is sorted in the ascending rate of occurrences.

4. Results are moving into a new array with the identifier URL.

It should be noted that in this case the more the rank, the less relevance (relevance) of the document. It is not clear that this method of ranking is better or worse than the previous one. It often happens that one method is better in some cases, but gives very poor results in others. Some methods are popular in the implementation and in the algorithm, and others, are very similar.

Consider the following ranking method that is somewhat similar to the previous method - ranking arrangement of words in the document.

Ranking by the distance between words.

It is often useful to rank documents based on how close to each other the words entered in the query. This method seems very logical because when a user enters query string keywords, he expects that in those documents, these words will occur close to each other, or be part of one sentence, thus as close as possible to the subject of the search. It should also be borne in mind that the order of the keywords does not have to be exactly this, as it is introduced.

Consider the algorithm of this method:

1. the function takes an array of query results search;
2. initialized array duplicated large values (eg: 1000000);
3. in the loop through the array and calculate the difference between the current location and the previous location of occurrences of words in each document:

$$rang = \sum_{i}^{n} X_i \left[\sum_{l}^{m} Z_l - Z_{l-1} \right] \quad (3)$$

4. the array is sorted in ascending order, ie, the less the value, so the better suited to us.

This method is also simple to implement and comprises a strong logical substrate to use just this method. However, the main disadvantage of this algorithm is it is completely useless in the case when the query string is specified only with one word.

Reference counting and algorithm PageRank.

The basis of the above methods of ranking is that they analyze the inner content of documents. These methods are still used by some search engines. Further, we are going to consider the method of calculating the rank of the document, taking into account, as other documents speak about this document. However, the rank of documents is calculated depending on how many resources refer to the document in question. This method is particularly useful when indexing pages of questionable content since it is unlikely that these pages have links to real sites. Using the links contained in a document, you can count the number of references to each document in the list and perform the function of normalization. Thus we obtain all pages containing search keywords, ordered by the number of external links to them. But this method of ranking somewhat correct and has some flaws, for example, it is possible to manipulate the results by creating a number of "empty" pages pointing to the rank that you would like to raise. Further, we are going to consider one of the most famous algorithms that changed the operation of all search engines and raised them to a new level - the algorithm PageRank. PageRank algorithm was invented by the founders of Google, and variations of this algorithm are now used in all the major search engines. This algorithm assigns each page rank assessing its significance. The significance of the page is calculated on the basis of relevance, referring to her page and the total number of links available on each of them. [9] [10]

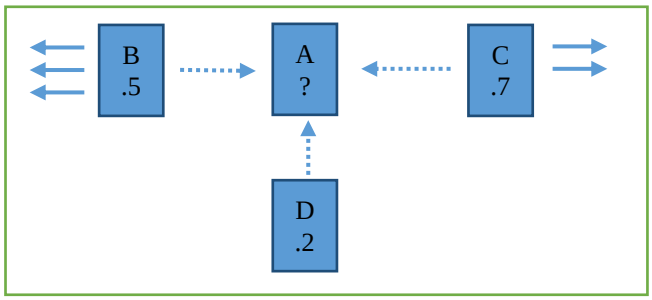

Figure 1. Page rank calculation

The diagram shows computation page rank A, wherein each of the pages B, C, and D refer to the page A. At B there are three more links of the page, and the page C - two. D refers only to A. To find the rank of A, take the ranks (PR) each referring to A page, divide them by the total number of links on this page, we add the resulting value is then multiplied by the attenuation coefficient of 0.85 and add a minimum value 0 15, for example:

Table-I:Calculation steps of Page Rank algorithm

Calculation steps	Calculation process
1-step	PR(A) = 0.15 + 0.85 * (PR(B) / link(B) + PR(C) / link(C) + PR(D) / link(D))
2-step	0.15 + 0.85*(0.5/4+0.7/3+0.2/1)
3-step	0.15+0.85*(0.125+0.233+0.2)
4-step	0.15+0.85*0.588
5-step	0.15+0.4998
6-step	0.6498

Note that D gives a great contribution to the rank A, or B then C, although its own rank is lower. This is due to the fact that D refers only to A and thus brings the result of his rank as a whole.

At first glance, a rather simple formula, but here we are faced with another problem, namely, how do we get the original page rank B, C, and D? To solve this problem, we developed a method of calculating the ranks of all the pages in our database. The accuracy of the calculation of the ranks with this method depends on the number of iterations. The experiment was carried out with 20 iterations; whereat the initial stage of each document is assigned the rank of 1.0. Further, we'll describe the work function of calculation of the ranks of all the pages available in the database:

Select all documents links from the data store and fill the array.

1. In the loop through the array with references.

1.1. For each document, find all documents that link to the document - the set Y (4).

1.2. Next, calculate the rank function of the document takes the form:

$$rang = 0{,}15 + \sum_{i=1}^{n} 0.85 * (PR(Y_i)/count_Y) \quad (4)$$

where n - number of pages linking to a document; PR (Y_i) - the rank of the referring page; $count_Y$ the number of links on this page.

2. Go to the next iteration, or finish the calculation.

Immediately it should be noted that this approach to calculating the page rank has two significant advantages: firstly, page rank is calculated beforehand, which allows not boot the system calculations rank; Secondly, a substantial reduction of the time between the receipt of a user request and return an answer.

IV. RESULTS

Testing the system according to the key query “tax amendments” for rank calculation using the PageRank algorithm.

Table-II:Results for Page Rank algorithm

№	Rank	Url
1	1	http://e-gazeta.norma.uz/publisg/doc/text109747_platit_nado_vovremya
2	1	http://products.norma.uz/seminar
3	0.3568	http://norma.uz//sbscribe.html
4	0.3568	http://norma.uz//novoe_v_zakonadatelstve/v_perechen_importiruemoy_produkcii_do_bavleny_izdeliya_podlejashchie_obyazatelnoy_sertifikacii
5	0.3568	http://norma.uz//novoe_v_zakonadatelstve/odobren_zaym_na_stroitelstvo_turakurganskoy_elektrostancii
6	0.3568	http://norma.uz//novoe_v_zakonadatelstve/za_oformlenie_kadastrov_platnim_po_poryadku
7	0.3568	http://e-gazeta.norma.uz?paper=sbx
8	0.3568	http://norma.uz//profi_novosti
9	0.3568	http://norma.uz//product/more/7_maloe_predpriyatie_uchet_nalogi_pravo
10	0.3568	http://norma.uz//product/more/6_nalogi_voprosi_i_otveti

IV. CONCLUSION

As you can see, there is no perfect method of ranking documents. Each algorithm is based on a logical analysis of data, some of the contents of the document on other significance relative to other documents. The purpose of this experiment was the application of ranking and at the end, you can draw several conclusions:

1. Organized a system that combines a variety of resources of e-government in a unified database of links to documents.

2. Developed search engine autonomously conducting indexing of web pages and documents in EP.

3. When using a full-text search method used for calculating the rank of the document to obtain a ranked list of retrieved documents and links to resources.

4. Proposed the combined use of several methods of ranking, depending on different situations arise in the system. As the example of a search query consisting of one keyword, the method of "distance between the words" absolutely not suitable, and the best results can be expected from the method "of calculating the frequency of words" and vice versa otherwise.

Using the PageRank algorithm in finding documents can be considered as the best option for ranking documents, as the highest rank will receive the document referenced by others, which allows the user to get a list of primary source documents.

REFERENCES

1. V.Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
2. https://ru.wikipedia.org/wiki/Электронное_правительство
3. http://review.uz/ru/article/441
4. http://www.egovforum2014.kz/article/электронное-правительство-южной-кореи-выходит-на-уровень-30-бак-гюрнг-гуг
5. http://www.gov.uz/ru/press/technology/24029
6. http://uzinfocom.uz/ru/page/show?alias=egov
7. http://lex.uz/pages/getpage.aspx?lact_id=2237921
8. http://ornitos.blogspot.com/2009/03/pagerank.html
9. http://wseob.ru/seo/searchengine-anatomy
10. Samandarov Bunyod G'ayratovich, Matqurbanov To'lqin Alimboevich, Yangiboeva Madina Rustamovna, Developing methods of allocation resource in the servers of IMS subsystem., 'International Journal of Innovative Technology and Exploring Engineering (IJITEE)', ISSN: 2278–3075 (Online), Volume-9 Issue-1, November 2019, Page No. 4606-4609

AUTHORS PROFILE

Khujaev Otabek Kadambayevich, has graduated the Urganch Branch of Tashkent University of Information Technologies (UB TUIT) in 2009 on speciality Information technologies with an honours diploma. He actively participates at scientific and technical, scientifically practical and methodical conferences, seminars of republican and international levels. He received PhD degree in 2019 on topic "Development of a software tool for intellectual analysis for decision support systems" He is head of the department Information technologies at computer engineering faculty at Urgench branch of TUIT. He has more than 50 papers scientific journals and conference proceedings.

Yusupova Shohida Botirboyevna, In 2007 he graduated from the Tashkent Pedagogical University named after Nizami with a degree in "Informatics". She is a senior lecturer at Urgench branch of the Tashkent University of information technologies. Every year, several articles of him have been publishing in our republic and abroad scientific journals. She actively participates in scientific-technical, scientific-practical and methodical conferences, in national and international seminars. Her professional interests: Distance Learning Technologies and Pedagogical Web Design. Author of more than 30 scientific articles and conference papers.

Allaberganova muyassar Rimberganovna, graduated from the Tashkent University of Information Technology in 2011, majoring in Teaching Methods in Professional Science. She actively participates in scientific and technical, scientific-practical and methodical conferences, seminars of the republican and international level. She is an assistant professor at Telecommunication technologies and professional education faculty of UB TUIT. Her professional interests Technology of creating e-learning resources and web-design. She has more than 10 papers scientific journals and conference proceedings.