2010 2nd International Conference on Information and Multimedia Technology (ICIMT 2010)
IPCSIT vol. 42 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V42.10

Index Weight Decision Based on AHP for Information Retrieval on Mobile Device

Yuanyuan Wu[1], Ye Tian[1], Wendong Wang[1], Xirong Que[1], Xiangyang Gong[1], Jian Ma[2], Canfeng Chen[2], Xiaogang Yang[2]

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications Beijing, China

[2]Nokia Research Center, Beijing, China

yuanyuan326601@sina.com, humanty@163.com, wdwang@bupt.edu.cn, rongqx@bupt.edu.cn, xygong@bupt.edu.cn, jian.j.ma@gmail.com, canfeng-david.chen@nokia.com, ext-xiaogang.1.yang@nokia.com

Abstract—Index weight decision is important to the ranking result of information retrieval on mobile device. Nowadays many methods for determining the index weights are subjective and complicated. So an index weight decision technique based on the AHP method is introduced so that a better retrieval performance can be obtained. We introduce how the AHP method is applied to get the proper index weights in the mobile application in details, and then a prototype project is implemented to test the availability of the technique. The experimental results show that the index weight decision technique is effective in improving the performance of searching contents on mobile device.

Keywords-AHP; Search Engine; Mobile Application; Index; Weights

1. Introduction

With the rapid development of mobile market, content retrieval on mobile device is playing a more and more important role in our daily life. Normally there are several kinds of objects in mobile device, such as conversation, photo, calendar, email and contact. Retrieve relevant objects by taking certain mobile application object as an entry is a hot research point nowadays. When one is browsing a photo, he may probably wish to find the related information with the photo. He may want to know the contact information of the people who took the photo, or the conversations during the time the photo was taken. So the requirements of effective indexing and searching relevant contents in mobile device are growing rapidly.

Actually, the created index directly affects the ranking of searching results, while setting the field weights affects the created index. That is, changing the weight values will result in the changed scores of the fields, and then the result set is sorted accordingly so that the documents where the fields stay can be positioned in advance or by post. When calculating the scores, it is irrelevant between the document score and the weight value in default case. However, once the default weight value is changed, the relationship between the document score and the weight value becomes relevant. The larger the weight value is, the higher the document score is. Thus, setting field weights in the process of indexing is significant.

Nowadays, there are several ways of setting field weights. But in most cases they are subjective and merely upon the experience of the developers. To solve the problem, in the paper we exploit an AHP-based method to set the field weights such that objective and reasonable weights can be obtained.

The rest of the paper is organized as below. In Section 2, we simply introduce the related work with the paper. Then we describe AHP method and how we apply the method to set field weights in Section 3. In

Section 4, a prototype project of setting field weight with AHP method is developed and the experimental results are analyzed. Finally, we conclude our work and describe the future work in Section 5.

2. Related Work

Content retrieval on mobile device has been a hot topic nowadays, and effective indexing and searching act as important roles in the field. In this case, setting field weights in the process of indexing has been a meaningful research point.

Hye-Jin Jeong [4] proposes a technique that determines the weight of the index. The technique updates the weight of the index on the basis of the weight of terms, and calculates the weight for the term in each paragraph.

Ning Zhou and others [6] introduces a method to create index in Lucene and utilization of the index, such that the weighted frequent items can be found. They also implement a weighted association rule mining algorithm based on apriori algorithm, which searches the weighted frequent items until a satisfactory result can be obtained.

But the methods described above are mainly used in web search applications, and they are complicated and need to frequently updating the index, which are not suitable for the searching on mobile device. Therefore we propose a novel approach which can be well applied in the mobile application and produces effective performance.

3. System Framework

3.1 AHP Introduction

The Analytic Hierarchy Process (AHP) by Saaty [1] is an analysis method that combines qualitative and quantitative aspects. It is a method that decomposes a complex problem into various component factors, and groups those factors according to their dominating relations in order to form a progressive structure. Through the comparisons of pair-wise factors, the relative importance of factors is determined. And through the comprehensive decision judgment, the total order of the relative importance in the decision solution is determined. Figure 1 [2] depicts the model of AHP.

The characteristic of the method is that the weight of each factor is calculated by the relative importance of various factors, thereby it is scientific by better avoiding the subjectivity and arbitrariness.

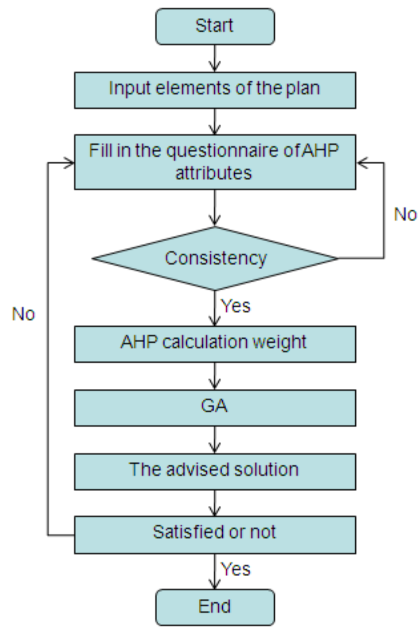

Figure 1. AHP flow chart

Due to its practicability and effectiveness in the complex decision-making problem, it is quickly applied in the scopes of economic planning and management, energy policy and distribution, behavioral science, transportation, agriculture, education, medical treatment, environment, etc. But it is rare in the field of computer science, especially in the scope of information retrieval.

3.2 Mobile Application Based on AHP

In this paper we exploit AHP model for mobile applications retrieval. The mobile application objects, such as photo, conversation, email, calendar and contact, can be taken as an entry to search relevant objects. Figure 2 shows the architecture of the mobile application which creates the index based on AHP.

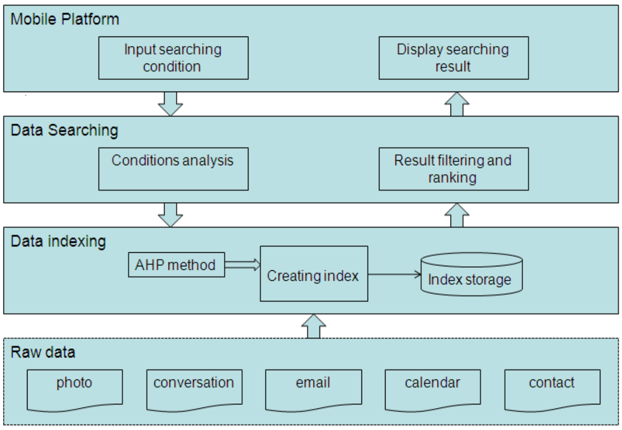

Figure 2. Architecture of mobile application based on AHP

The raw data including the photos, conversations, emails, calendars, and contacts are collected on mobile device. When indexing based on those data, the AHP method determines the weight of each field which will be stored into the index files. If indexing is done, users can input search conditions on mobile device, and the system will get data from the index after analyzing the conditions. At last, the search results are displayed on mobile device when the data obtained from the index is filtered and ranked by the system.

The architecture shows that the AHP method plays an important role in determining the field weights in the process of indexing. In the next section, we will discuss the implementation of setting the index weights based on AHP in details.

4. System Implementation

In this section, we introduce how AHP is applied to set index weights in details based on a prototype project.

4.1 Establish hierarchical analysis structure model

On the basis of thorough analysis of the practical problems, the factors influencing the content retrieval on mobile device are decomposed into several layers. The factors of the same layer influence that of the upper layer, and dominate that of the lower layer. When setting the index weights with AHP, the hierarchical structure model can be established as shown in Figure 3.

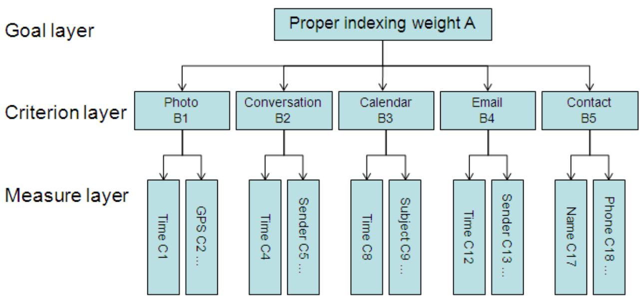

Figure 3. Hierarchical structure of mobile application

The top layer A is called goal layer, which aims to solve the problem using hierarchy process (AHP). Here using the AHP method aims at obtaining proper indexing weights.

The intermediate layer B is called criterion layer, which links goal layer A and measure layer C. In this system the layer is needed to determine the final weights in the goal layer, and it is influenced by the measure layer at the same time. Here photo, conversation, calendar, email and contact are considered in this layer

The bottom layer C is called measure layer, which stands for the policies or measures to solve the problem. And in the system the layer contains the detail factors of each object in criterion layer. Actually, those factors in layer C directly influence the result of layer B.

4.2 Establish the judgment matrix

From the goal layer A to the measure layer C of the hierarchical structure model, the judgment matrix is established through the comparison between the paired factors which are relevant in the same layer. The comparison result represents the importance degree of the lower layer to the upper one, and it is expressed with 1 - 9 scale definition, which is shown as Figure 4.

scale	definition
1	Represents the importance of a1 is equal to that of a2.
3	Represents the importance of a1 is more than that of a2 slightly.
5	Represents the importance of a1 is more than that of a2 apparently.
7	Represents the importance of a1 is more than that of a2 strongly.
9	Represents the importance of a1 is more than that of a2 absolutely.
2,4,6,8	Represents the middle values of above two adjacent scales.
reciprocal	Represents the importance degree of a2 compared with a1.

Figure 4. 1 - 9 scale and definition

In this system, each factor of the criterion layer B should be compared with the others in the same layer, so that the influence to the goal layer A can be obtained. That is, the relative importance among photo, conversation, calendar, email and contact is needed to calculate. And then the factors of the measure layer C should be made paired comparison in order to get the influence to the relevant factor of the criterion layer B. For example, photo has the attributes: time, GPS and keywords. The comparison among those attributes stands for the importance degree to the photo.

4.3 Determine the index weights

The steps of calculating the index weights are as follows.

- Based on the judgment matrix, calculate $\overline{W_i}$ of each row in the matrix.

$$\overline{W_i} = \sqrt[n]{\prod_{j=1}^{n} b_{ij}}\,(i = 1,2,3...n) \tag{1}$$

- Normalize the vector $\overline{W} = [\overline{W_1}, \overline{W_2}, ..., \overline{W_n}]^T$.

$$W_i = \overline{W_i} / \sum_{i=1}^{n} \overline{W_i} (i = 1,2,...,n) \tag{2}$$

$W = [W_1, W_2, ..., W_n]^T$ is the desired eigenvector.

- Calculate the maximum eigenvalue $\lambda_{\max}$ of the judgment matrix.

$$\lambda_{\max} = \sum_{i=1}^{n} \frac{(AW)_i}{nW_i}. \tag{3}$$

$(AW)_i$ represents i-th component of vector AW.

- In order to test the consistency of the judgment matrix, CI (consistency index) is needed to calculate.

$$CI = \frac{\lambda_{\max} - n}{n-1}. \quad (4)$$

when CI = 0, the matrix is concluded to have the complete consistency.

when CR = CI/RI < 0.10, the matrix is concluded to have satisfactory consistency. Otherwise the matrix is needed to adjust. To test whether the consistency of the judgment matrix is satisfactory, it is needed to compare CI with average random consistency index RI. For the matrix with 1 ~ 9 order, RI values are shown as Figure 5.

order	1	2	3	4	5	6	7	8
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41

Figure 5. RI values

- With the above steps, the index weights of layer B to layer A and layer C to layer B can be calculated. In this application, the weight of each factor in the criterion layer is calculated, and the results show that the relative importance of each factor of the criterion layer to the goal layer. For the goal of getting proper index weights, we can see that the conversation is more important than the photo when the weight of the conversation is larger than that of the photo. So is the measure layer to the criterion layer.
- Based on the weights of layer B to layer A and layer C to layer B, the final weights of layer C to layer A can be calculated and they are exactly what we desire. That is, the weight of each field is obtained to reach the goal of getting proper index weights in the application.

5. Performance Evaluation

Based on the implementation of the prototype project, we collect 1800 photos, 1800 conversations, 216 calendars, 1080 emails and 524 contacts on mobile device. And those data are gathered based on 50 scenes in two weeks. For example, we set a scene of inviting friends to watching flowers, in which we send 35 conversations and 20 emails, set 4 calendars and 10 contacts, take 40 photos. All of the data are collected to support the performance evaluation of the indexing weights with AHP.

5.1 Results of the application with AHP

1) Survey on the desired index weights: We distribute the questionnaires to 30 graduate students in the field of computer science. From the collected data of the 30 questionnaires designed on the basis of 1 - 9 scale, the judgment matrixes will be generated. Figure 6 shows the sample of the questionnaire, and '√' represents the student's selection of the raletive importance between the fields.

Tick behind the more important field		Equally	Slightly		Apparently	Strongly	Absolutely
time √	GPS		√				
time	keywords √			√			
GPS	keywords √				√		

Figure 6. Sample of the questionnaire

From the sample, we can see that the user thinks time is more important than GPS slightly to the photo searching. Based on 1-9 scale, the value of time to GPS is 3, and the value of GPS to time is 1/3 in the judgment matrix.

2) Calculation of the index weights: On the basis of the collected data from the questionnaires and the caculation method introduced in Chapter IV, the weights of layer B to layer A are obtained shown as TABLE I. In the table, B1 represents the photo, B2 represents the conversation, B3 represents the calendar, B4 represents the email, and B5 represents the contact.

TABLE I. A – B JUDGMENT MATRIX AND WEIGHTS

A	B1	B2	B3	B4	B5	W
B1	1	1/5	1/3	1/6	1/6	0.0440
B2	5	1	4	1	2	0.3242
B3	3	1/4	1	1/4	1/4	0.0841
B4	6	1	4	1	1	0.2928
B5	6	1/2	4	1	1	0.2549

The weights of layer C to layer B are also obtained. Here we take the photo B1 for example, and the weights of time, GPS and keywords to the photo are shown as TABLE II. In the table, C1 represents time, C2 represents GPS, and C3 represents keywords.

TABLE II. B1 – C JUDGMENT MATRIX AND WEIGHTS

B1	C1	C2	C3	W
C1	1	1	1/4	0.2212
C2	1	1	1/4	0.2212
C3	4	4	1	0.5576

Finally, the weights of layer C to layer A are calculated shown as TABLE III. They reflect the importance degree of each field to the search results. The larger is the weight, the more important is the field. To enlarge the difference among the field weights, each weight is multiplied by 200 to get the final field weights.

TABLE III. A – C FINAL WEIGHTS

A	B	C	C to A	Field Weight
	B1(0.0440)	C1(0.2212)	C1(0.0097)	C1(1.94)
		C2(0.2212)	C2(0.0097)	C2(1.94)
		C3(0.5576)	C3(0.0245)	C3(4.90)
	B2(0.3242)	C4(0.0963)	C4(0.0312)	C4(6.24)
		C5(0.0973)	C5(0.0315)	C5(6.30)
		C6(0.3828)	C6(0.1241)	C6(24.82)
		C7(0.4236)	C7(0.1373)	C7(27.46)
	B3(0.0841)	C8(0.1041)	C8(0.0088)	C8(1.76)
		C9(0.3878)	C9(0.0326)	C9(6.52)
		C10(0.3878)	C10(0.0326)	C10(6.52)
		C11(0.1203)	C11(0.0102)	C11(2.04)
	B4(0.2928)	C12(0.0733)	C12(0.0216)	C12(4.52)
		C13(0.0572)	C13(0.0168)	C13(3.36)
		C14(0.1637)	C14(0.0479)	C14(9.58)
		C15(0.3852)	C15(0.1128)	C15(22.56)
		C16(0.3166)	C16(0.0928)	C16(18.56)
	B5(0.2549)	C17(0.5453)	C17(0.1390)	C17(27.80)
		C18(0.1653)	C18(0.0421)	C18(8.42)
		C19(0.0948)	C19(0.0242)	C19(4.84)
		C20(0.0763)	C20(0.0194)	C20(3.88)
		C21(0.1183)	C21(0.0302)	C21(6.04)

5.2 Evaluation of the results

Here, we use the most popular metrics of precision ratio and recall ratio to evaluate the search performance. And then the relevance ratio is used to evaluate the determination of the document ranking [3].

3) Precision Ratio and Recall Ratio Evaluation: Precision ratio is defined as the searched relevance items to the total searched items shown as (5) [4].

$$PR = \frac{SRD}{SD}. \tag{5}$$

where PR is precision ratio, SRD is the number of searched relevance documents and SD is the total number of searched documents [4].

Recall ratio is shown as (6) [4], which is defined as

$$RR = \frac{SRD}{RD}. \tag{6}$$

where RR is recall ratio, SRD is the number of searched relevance documents and RD is the total number of relevant documents [4].

Figure 7 shows the test results of precision and recall metrics. We can see that the index weights determined by the AHP method in the paper yield the improved search performance with the higher precision ratio and recall ratio.

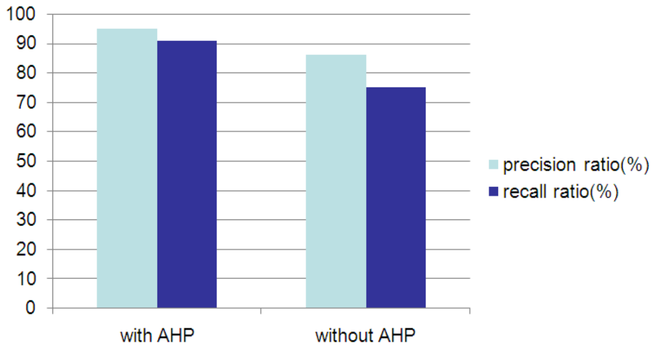

Figure 7. Precision and recall of results

4) Relevance Ratio Evaluation: Relevance ratio is suitable for evaluating the ranking performance determined by the index weights.

$$relevance = \frac{\sum_{i=1}^{n} R_{score}}{\sum_{i=1}^{n} R_{\max}}. \quad (7)$$

where R_{score} represents that the paper relevance as evaluated by the user is within the range 0-3. The meaing of each value is defined as: 0 = Not relevant, 1 = Normal, 2 = Relevant, 3 = Very relevant. $R_{\max}$ represents that the maximum rlevance value is 3. n represents the number of documents within the top % of ranked papers [4].

Relevance ratio of the tests results is drawn in and Figure 8. We can see that the ranking performance is improved significantly after using the AHP method, especially for the search precision of the photo, conversation and email.

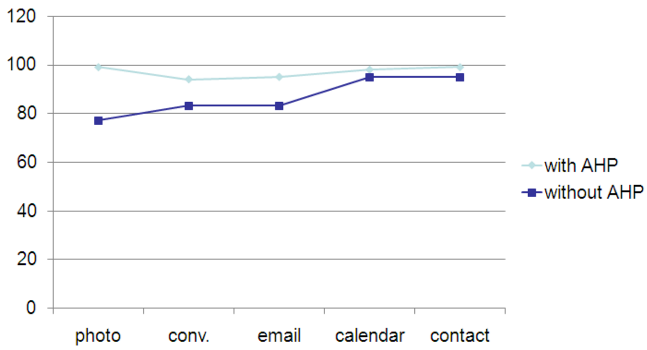

Figure 8. Relevance ratio of results

6. Conclusion and Future Work

In this paper, we introduce the AHP method to get the objective and scientific index weights. Firstly we describe AHP method and the mobile application based on AHP. Secondly, setting the field weights with AHP is introduced in the application in details. Finally, we develop a prototype project of setting field weight with AHP method and evaluate the performance of the application with AHP.

We can see that the index weights generated by AHP improve the ranking of search results, and obtain the satisfactory contents on mobile device. However, there are many other factors affecting the ranking of search results except of the index weights. So in the future, there is still a long way to go for the better ranking in the field of searching on mobile device.

7. References.

[1] T.L. Saaty, Decision Making with Dependence and Feedback: the Analytic Network Process, Pittsburgh: RWS, 2001.

[2] Li Yongxin, Wang Kaizhuo. Index Weight Technology in Threat Evaluation Based on Improved Grey Theory. Intelligent Information Technology Application Workshops, 2008.

[3] Seon-Mi Woo, Chun-Sik Yoo, Yong-Sung Kim. User - Centered Document Ranking Technique using Term Association Analysis. Journal of Korea Information Science Society, Vol.28, NO.2, pp.149-156, 2001.

[4] Hye-Jin Jeong. Index Weighting Based on the Weight of the Index of Each Tag. Future Generation Communication and Networking Symposia, 2008.

[5] Hye-Jin Jeong, Yong-Sung Kim. Index Weight Decision Technique for Search Reliable Documents. Future Generation Communication and Networking, 2008.

[6] Ning Zhou, JiaXin Wu, Shaolong Zhang, Hongqin Chen, XiangRong Zhang. Mning Weighted Association Rules with Lucene Index. Wireless Communications, Networking and Mobile Computing, 2007.

[7] Mei Yang, Meihong Deng, Rongyan Jia, Yue Liu. Research on index weight based on improved Grey Relational Analysis. Machine Learning and Cybernetics(ICMLC), 2010.