Subject Perception Semantic Model for Information Retrieval in Tourism

Lingling Zi, Xin Cong
School of Electronic and Information Engineering, Liaoning Technical University, Huludao, China
Email: linglingziltu@126.com, chongzi610@163.com

Yaping Zhang
China Faculty of Computer Science & Information Technology, Yunnan Normal University, Kunming, Yunnan, China
Email: zhangyp.cs@gmail.com

Abstract—**With the continued growth of tourism multimedia information made available through the World Wide Web, more efficient and accurate retrieval approaches are leading to an increasing demand. Most existing information retrieval approaches are based upon keywords and therefore provide limited capabilities to capture the query requirements. However, a complete understanding of search requirements is essential for improving the effectiveness of retrieval. To achieve this goal, we propose a novel subject perception semantic model (SPSM) for searching tourism information, which allows the customized query threshold to retrieve tourism information. In this model, we present the definition of Subject Hierarchy Graph to support semantic search capabilities, and propose computing methods of subject perception to quantify the semantics of query requirements. The experiments show that SPSM obtained encouraging performances.**

Index Terms—**information retrieval; subject perception; tourism; quantification; query requirements**

I. INTRODUCTION

During the last decade, the rapid advance of search technology has made people acquire multimedia information easily. However, due to complex data formats of tourism information and the ambiguity of query requests, people have to spend much time to find out from query results exactly what they want. So how to understand accurately query requirements is a challenge for the development of information retrieval system. Nowadays, many technologies have been proposed, including ontology, schema summarization, semantic search and query expansion, and these technologies provide intelligent retrieval services [1-2]. For example, Avatar Semantic Search was used to express user semantic information through extracting facts and concepts [3]. KIM was provided to complete the service of automatic semantic annotation [4]. XSEarch was presented as a semantic search engine for XML [5]. A web search engine using page counts and texts was developed to measure semantic similarity between words [6]. A type of relevance feedback was proposed and more expansion words were associated with the user query [7]. A new pseudo-relevance feedback method was presented to retrieval information, and the words with the highest degree of query word were identified to new queries [8]. Nevertheless, there is still a problem of understanding difficulties. In order to better analyze query intention, we propose a subject perception semantic model (SPSM) from the perspective of requirement quantification. SPSM allows customized query thresholds to retrieval tourism information, stressing on the one hand the computation of subject perception, and on the other hand the consideration of label documents, which extract from multimedia resources as the target search space.

The development of SPSM presents the solutions to two key problems: one is how to quantify users' implication requirements and the other is how to integrate multimedia resources. With regard to the first problem, the key issue is to identify appropriate semantic information according to the requirements from users' vague queries. To address this problem, we introduce the definition of Subject Hierarchy Graph and subject perception computing to choose appropriate query subjects, and propose a new measuring method reflecting the ambiguity of users' requirements. The integration of multimedia resources is the second problem. It is important to make sure that multimedia search results be displayed in a comprehensive ranking. To address this, we propose the label documents method for tourism information retrieval. In conclusion, the novel contributions of this paper are shown as follows.

- We propose the definition of Subject Hierarchy Graph (SHG) to quantify the semantics of the query keywords, and it is the building block of Subject Perception.
- We present the computations of Subject Perception for measuring implication queries. And the application of Subject Perception in SPSM can also be demonstrated.
- Multimedia query results are displayed in a precise and comprehensive way, namely, not only text results, but also image results. In order to unify these multimedia resources, the label documents method is presented.

doi:10.4304/jsw.9.11.2844-2850

- Better solving of the semantic search problem, the implementation of SPSM is proposed for enhancing semantic retrieval. This system includes four parts 1) semantic annotation module 2) index module 3) subject perception module 4) result display module. SPSM is capable of improving the traditional problem of keyword search, so as to provide more accurate multimedia query results.

The rest of the paper is structured as follows. Section II discusses the related work. Section III shows the method of Subject Perception. Section IV illustrates the implementation of SPSM. Section V presents experimental works to demonstrate the effectiveness of SPSM. Section VI concludes the paper.

II. Related Work

Ontology has a good ability of semantic representation and it is widely used in the field of information retrieval [9-10]. So many models and methods using ontology technique have been proposed for efficient retrieval. Semantic retrieval based on concept designing [11] is proposed and it makes full use of ontological concepts. A concept map learning system based on domain ontology is presented [12] and it plays a very important role for education. A service search engine for industrial digital ecosystems [13] is developed and it achieves accurate semantic retrieval. In order to provide implication query results, graph-based query rewriting for knowledge sharing is proposed [14]. Inspired by the idea of ontology, we construct Subject Hierarchy Graph for tourism multimedia information, so as to complete the search task.

Schema summarization is of great help to concise overview of searching results, and an important advantage is that a user can determine query results in a short time. So this technology has been widely applied in system development. Dynamic summarization of bibliographic-based data is developed to accommodate relevant information [15]. A visual tool called VIREX for producing XML schema is designed and one of the attractive features is to support XML views update in a form of summary obtained from web-based database [16]. In the SPSM model, all the query results can be summarized in the form of subject and shown clearly to the users.

III. The Subject Perception Method

In this section, we elaborate the subject perception method to obtain query semantics. This method contains two parts, one is to construct Subject Hierarchy Graph (SHG) and the other is to calculate the values of subject perception based on SHG. Finally, we present the application of subject perception which can be used in SPSM.

A. Subject Hierarchy Graph

A subject hierarchy graph contains three parts: the subject layer, the concept layer and the instance layer. The subject layer is composed of subject nodes SN, which is denoted by $< sid, h, n_c, n_s >$. The concept layer is composed of concept nodes CN, which is denoted by $< cid, sort, n_i >$. And the instance layer is composed of instance nodes IN, which is an instance of a concept associated with the given subject. Table I shows the definition of each notation in SHG.

Consider the SHG in Fig. 1. Subject nodes in the subject layer are organized in the form of a tree. Take subject S5<S5,3,4,0> as an example. We get the following information: the subject id is S5, the subject lever is 3, the subject type is a leaf node, and four concept nodes containing the subject are placed at the concept layer, where c5 and c6 belong to the basic concept (BC for short), c7 belongs to the comment concept (CC for short) and c8 belongs to the association concept (AC for short). So the corresponding concepts can be represented as <C5,BC,1>, <C6,BC,1>, <C7,CC,3>, and <C8,AC,1> respectively. The third part of above angle brackets represents the number of instances associated with the concept. For example, c7 connects three instances (i.e. i8, i9 and i10) in the instance layer.

TABLE.I
NOTATIONS OF SHG

Notation	Definition
sid	The identity of SN
h	The level of SN
n_c	The concept number associated with SN
n_s	The number of child nodes of SN, including leaf nodes and connection nodes
cid	The identity of CN
sort	The type of CN, including basic concepts(BC), association concepts(AC), and comment concepts(CC)
n_i	The instance number associated with CN

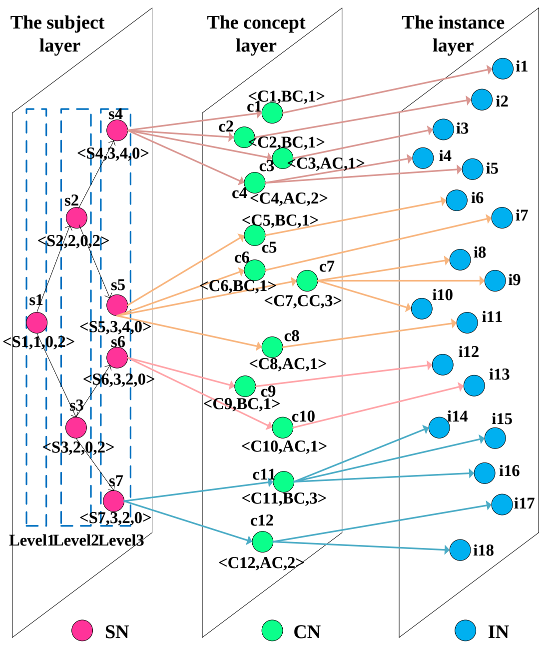

Figure 1.An example of SHG.

This work was supported by applied basic research programs of Yunnan Province (2010CD047) and the National Natural Science Funds (61262070).

B. The Computing of Subject Perception

An important contribution of this paper is to quantify the semantics of the query keywords. In order to achieve this goal, we compute the values of subject perception for each type nodes. Due to the type of nodes in SHG, the corresponding three methods are presented respectively, including the perception computing in the subject layer (P_S), the perception computing in the concept layer (P_C) and the perception computing in the instance layer (P_I).

P_S reflects the extent of subject concerned by tourists and the bigger the value of P_S is, the more attention this subject attracts. The computing formula of P_S is shown as (1).

$$P_S = w_1 * \vartheta(h) + w_2 * \frac{1}{n_s + 1} + w_3 * \frac{n_c + 1}{N_{\max} + 1} + w_4 * \varsigma * \varpi \quad (1)$$

In (1), $w_i (i = 1,2,3,4)$ is a weighting factor, which satisfies $w_1 + w_2 + w_3 + w_4 = 1$, $\vartheta(\mu) = (11 - \mu)/10$, $N_{\max}$ is the maximum number of concepts contained by the same subject, ϖ is the ratio of the subject resources to total resources, and ς is an amplification constant, here $\varsigma = 10$.

P_C is related to the concept type and the instance number n_i, shown as (2).

$$P_C = \vartheta(r) * \frac{n_i + 1}{\mathrm{I}_{max} + 1} \quad (2)$$

Where r is the ranking number of concept type, and the descending order is BC, AC and CC. I_{max} is the maximum number of instances with any concept contained by the same subject.

P_I indicates the attention degree of the instance, shown as (3).

$$P_I = \varepsilon_1 * P_C + \varepsilon_2 * \eta \quad (3)$$

Where ε_1 and ε_2 are adjustment coefficients, and η is the function of the linear conversion, shown as (4).

$$\eta = \frac{n_l - n_{\min}}{n_{\max} - n_{\min}} \quad (4)$$

Where n_l is the number of multimedia resources contained by the given IN, n_{min} is the minimal number of multimedia resources contained by any IN. Similarly, n_{max} is the maximal number.

C. The Application of Subject Perception

The proposed subject perception method is used to measure the implication query requests and this method is divided into leaf node measure (LNM) and connection node measure (CNM) according to the type of subject nodes in the SHG. The input parameters of each measure method contain subject keywords matched (*key* in short) and threshold σ (σ>0). Measure result is denoted by MR (*id*, *EK*), where *EK* represents expansion keywords of *key* and *id* is its corresponding sequence number. Furthermore, larger the value of σ (σ>1) is, the wider the range of subject is extended and with σ closer to 1, it indicates that *EK* is more important to the given *key*. Table II and Table III show the application of LNM and CNM respectively. The details of application way, including the method, condition and example, are demonstrated in the corresponding table.

TABLE.II
LNM

Case 1: 0<σ<1	Method	Find all the IN which satisfies the condition, rank IN, and output MR
	Condition	$P_I > \sigma$
	Example	Input: Imperial Palace and 0.9 Procedure: We find that the name of SN is the Imperial Palace and instance nodes satisfying P_I>0.9 are Hall of Supreme Harmony, Palace of Heavenly Purity and Palace of Earthly Tranquility. Output:(1,Hall of Supreme Harmony) (2,Palace of Heavenly Purity) (3,Palace of Earthly Tranquility)
Case 2: 1<σ	Method	Search all the SN whose parent node is the same with the parent node of *key*, find the SN which satisfies the condition, rank the SN, and output MR
	Condition	$P_{S_{key}} + \frac{1-\sigma}{e_1} < P_S < P_{S_{key}} + \frac{\sigma-1}{e_1}$, where $P_{S_{key}}$ denotes the value of P_S of *key* and e_1 is an amplification factor
	Example	Input: Great Wall Badaling and 1.5 Procedure: Subject nodes are Great Wall Badaling, Fragrant Hill and Xiayunling. Based on the computing of the above condition, results are Great Wall Badaling and Fragrant Hill. Output: (1,Great Wall Badaling) (2,Fragrant Hill)

TABLE.III
CNM

Case 3: 0< α <1	Method	Find all the SN whose child node is *key* and P_S of the obtained SN need satisfy the below condition, and output MR
	Condition	$\frac{P_S - P_{S_{\min}}}{P_{S_{\max}} - P_{S_{\min}}} > \sigma$, where $P_{S_{\min}}$ and $P_{S_{\max}}$ are the minimal and maximal values of P_S of subject nodes contained by the parent node *key*
	Example	Input: natural scenery and 0.7 Procedure: We find all the child nodes of natural scenery and get the subject node with the minimal value is Changping Huyun and the subject node with the maximum value is Great Wall Badaling. So the results returned are natural scenery, Great Wall Badaling and Fragrant Hill. Output：(1,natural scenery) (2,Great Wall Badaling) (3,Fragrant Hill)
Case 4: 1< α	Method	Search all the SN whose parent node is the same with the parent node of *key*, find the SN which satisfies the condition, and output MR
	Condition	$\lvert P_{S_{key}} - P_S \rvert * e_2 < (\sigma - 1)$, where e_2 is an amplification factor
	Example	Input: Old Town and 1.2 Procedure: To be similar with case 2, the range of P_S is computed under the setting of e_2=10. So the results are Old Town and street scene. Output：(1,Old Town) (2,street scene)

IV. The Subject-Aware Semantic Retrieval Model

In this section, we demonstrate the subject perception semantic model (SPSM). The model architecture is presented in Fig. 2, and it consists of four parts: semantic annotation module, index module, subject perception module, and results display module. It is emphasized that the subject perception module is the core of SPSM. This module is established according to the information of semantic annotation module and the output results of this module serve for the results display module. Next we take each part of the SPSM model and describe its corresponding function within the infrastructure.

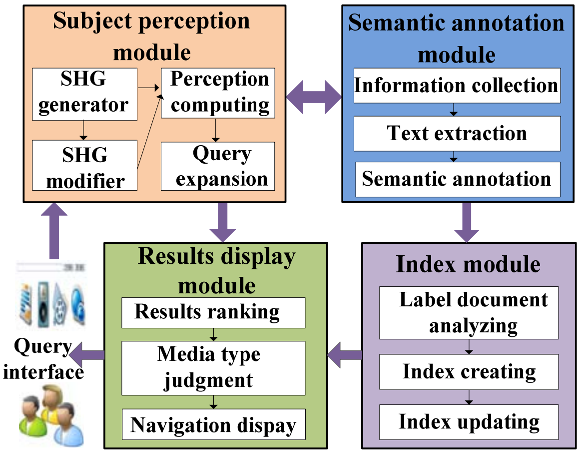

Figure 2 The architecture of SPSM.

A. Semantic Annotation Module

This model is responsible for collecting the massive tourism information from multiple sources of information, such as portal sites, travel forums, and blogs. It contains the following three parts. 1) Information collection. Firstly, we use a Meta search engine to search web sites which have a high correlation with tourism information and these web sites are put in the queue of URL, as a source set of information collection. Then a metadata judgment method using the technique of semantic analysis is used to extract new web links in this set, and keyword-based vector space model is adopted to exclude the web pages with useless information, so as to improve the accuracy of collected web pages. Finally, valid URL websites can be acquired through link filtering, and the corresponding images and texts can be captured. 2) Text extraction. Web crawler automatic captures multimedia information from URL websites and the obtained information are saved in the corresponding database. Then we adopt a series of operations, including feature extraction, structural analysis, and duplicate content elimination, to get information semantic [17]. And information semantic can be recorded in a file, which contains subject tags, the concept tags, instance tags, and label texts. 3) Semantic annotation. We establish a label document for each information file and the creation of this method is static, which is independent of the process of searching. The contents of label documents mainly contain document property information, resource collection information and semantic information. Through the method of label documents, images and texts can be unified from the perspective of semantic level.

B. Index Module

The index module is responsible for creating index fields according to the label documents, so as to lay the foundation for the searching of tourism information. This process mainly contains three steps (see Fig. 3):

Step1: The contents of label documents are analyzed and the corresponding index terms are extracted.

Step2: According to the obtained index terms, the index fields are constructed and the inverted index is created.

Step3: Both batch updating and incremental updating are adopted to complete constant renewal of index files.

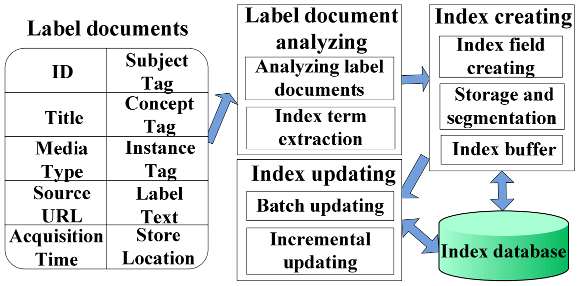

Figure 3 Index module.

C. Subject Perception Module

As a quantitative basis for semantic retrieval, the subject perception module is the core part of SPSM (see Fig. 4.). In this module, the values of subject perception are calculated according to SHG and on this basis, a list of expansion keywords are obtained according to the query keywords and the query thresholds. An overview of the process is shown as follows. 1) According to the keywords entered by users, preprocessing operations, such as null detection and Chinese word segmentation, are carried out. 2) The filtered keywords are matched in the SHG using the technique of word matching. 3) Using the proposed subject perception method, the appropriate query expansion lists are returned and saved in the hash tables.

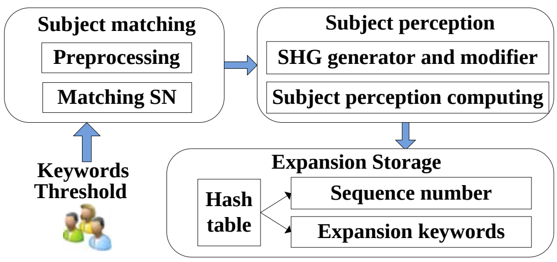

Figure 4 Subject perception module.

D. Results Display Module

As a large number of results obtained from the index module, the navigation method is used to display media results in an order way. Specifically, the expansion words are placed in the navigation bar, and all media results are classified and displayed according to the given expansion word. This module (see Fig. 5) consists of three parts. 1)

Results ranking. The type of ranking includes navigation ranking and content ranking. The former ranks expansion words according to the sequence number in the saving hash table and the latter ranks the results according to the correlation between the given expansion word and label texts in the label documents, i.e. the term frequency of label documents in the index file. 2) Media type judgment. Due to different display contents of media, the representation type of media results is determined. If the further details are need to be browed, users can only click the titles of returned results because the source URL is also saved the field of index file. 3) Navigation display. Using the navigation view, SPSM shows multi-faceted tourism information search results integrated with texts and images.

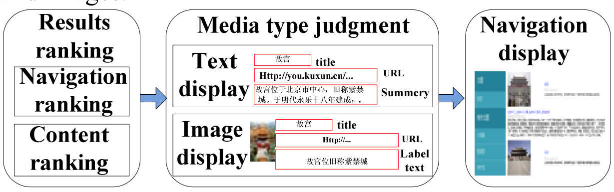

Figure 5 Results display module.

V. Experimental Results And Discussion

For the development of SPSM, we used Myeclipse8.5 platform, MySQL 5.1 and a PC with Intel Core(TM) 2 Duo T6570 processor and 2GB of main memory. Also, the open source full-text search engine Lucene and Web crawler Heritrix were also utilized. Using SPSM, we collected 6191 multimedia objects in the field of tourism, including 2934 texts and 3257 images. In this section, firstly we validate the efficiency and accuracy of the proposed approach by setting different weight values and then show the system performance from user's perspective. Finally, we conduct the comparison experiments.

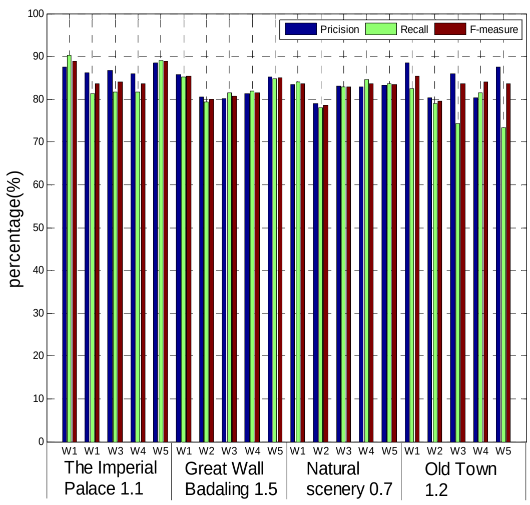

Figure 6 P/R/F results of different weights.

Three experiments were carried out to investigate the proposed method and it was measured by Precision, Recall and F-measure (P/R/F in short). Firstly, we set four query keywords and corresponding query thresholds using different weight values to evaluate P_S. The types of subject nodes of these query keywords contain both leaf nodes and connection nodes. The following weight value sets were studied: W1=<0.25,0.25,0,25,0.25>, W2=<0.1, 0.1,0,7,0.1>,W3=<0.7,0.1,0,1,0.1>,W4=<0.1,0.7,0,1,0.1> and W5=<0.1,0.1,0,1,0.7>. This experiment was helpful to find appropriate weights in the P_S formula. Fig. 6. summarizes the impact of different weights by presenting P/R/F results. More precisely, Precision values range from 79% to 88.5%, Recall values range from 73.4% to 90.3% and F-measure values range from 78.5% to 88.9%. We observe that W1 and W5 were the best weighting schemes. But for the last query case, Recall was only 73.4% using W5. In conclusion, W1 obtains better P/R/F results and this was due to a reasonable balance for all the factors of P_S. The reason of poor results produced by the other weighting schemes is because of highlighting only one factor contribution in the P_S formula.

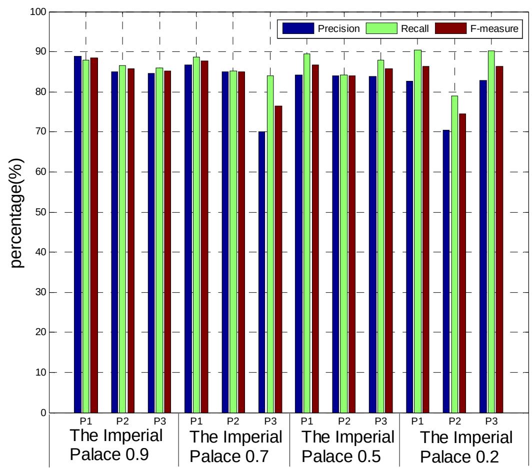

Figure 7 P/R/F results of different parameters.

Secondly, we investigated the P_I formula by setting different parameters. The following parameters were studied: P1: ε_1=0.5, ε_2=0.5; P2: ε_1=0.3, ε_2=0.7 and P3: ε_1=0.7, ε_2=0.3. Fig. 7 shows the P/R/F results in the case of the same query words under the different query thresholds and we can see the following three points. 1) Using the parameter P1, Precision values range from 82.7% to 88.9%, Recall values range from 87.9% to 90.5%, F-measure values range from 86.4% to 88.4%, and the average P/R/F values are 85.6%, 89.1%, 87.3% respectively. This shows that the results are relatively stable. (2) Using the parameter P2, Precision values range from 70.5% to 85%, Recall values range from 79% to 86.6%, F-measure values range from 74.5% to 85.8%, and the average P/R/F values are 81.1%, 83.8%, 82.3% respectively. It is noted that when query threshold is 0.2, Precision has the lowest value. Therefore, the results under this parameter are relatively unstable. (3) Using the parameter P3, Precision values range from 70.1% to 84.5%, Recall values range from 84.1% to 90.2%, F-measure values range from 76.5% to 86.3%, and the average P/R/F values are 80.4%, 87%, 83.4% respectively. It is also noted that when the query threshold is 0.7, Precision has the lowest value. Therefore, the results under this parameter are also relatively unstable. In view of these facts, we find the appropriate parameters in the P_I formula, namely ε_1=0.5 and ε_2=0.5.

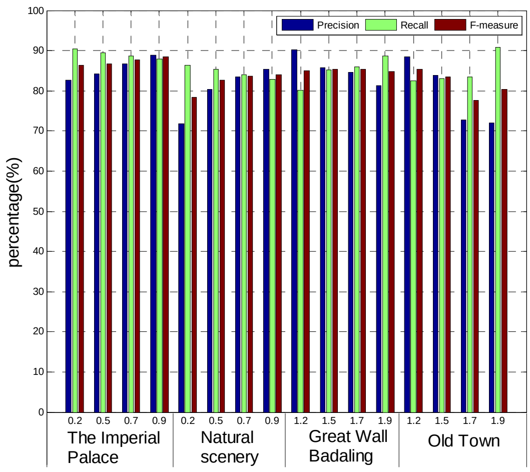

Figure 8 P/R/F results.

In the third experiment, four query keywords are selected and each query keyword contains four query thresholds in order to test the performance of SPSM. Fig. 8 depicts the results produced by W1 and P1. It can be seen from the figure that Precision values range from 71.8% to 90.3%, Recall values range from 80.1% to 90.8%, F-measure values range from 77.7% to 88.4% and the average P/R/F values are 82.6%, 85.9%, 84.1% respectively. It shows that SPSM has the encouraging results. Note that with regard to the same query keywords, we can find the following two points. 1) For the first two query keywords, as the decreasing of query thresholds, namely the value of query threshold gradually approaching zero, Precision values decrease whereas Recall values increase. (2) For the latter two query keywords, as the decreasing of query threshold, namely the value of query threshold gradually approaching one, Precision values increase whereas Recall values decrease. The reason of these facts is probably that different query thresholds get the different implicated query results.

TABLE.IV
THE QUERY CASE

Query ID	Query keyword	Query threshold
Q1	The Imperial Palace	0.9
Q2	The Imperial Palace	0.2
Q3	Great Wall Badaling	1.5
Q4	Great Wall Badaling	1.8
Q5	Natural scenery	0.7
Q6	Natural scenery	0.2
Q7	Old Town	1.2
Q8	Old Town	1.6

We demonstrate the performance of SPSM from the perspective of the users. The query cases are shown in Table IV and according to the results returned, ranking accuracy and satisfaction scores are displayed in Fig. 9. The criteria of satisfaction scores is set as follow: $0 < score \le 20$ represents slight satisfaction, $21 \le score \le 40$ represents fair satisfaction, $41 \le score \le 60$ represents moderate satisfaction, $61 \le score \le 80$ represents substantial satisfaction and $81 \le score \le 100$ represents almost perfect satisfaction. From Fig. 9, we find that users are satisfied with the query results.

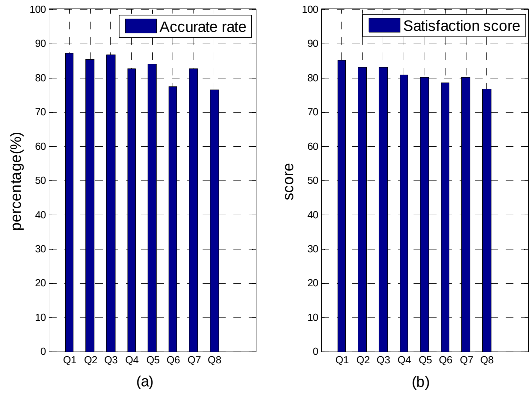

Figure 9 Performance evaluation by users.

Finally, we define the subject coverage measure to evaluate the integrity of query results. At the same time, the subject novelty measure is also presented to evaluate the expansibility of query results, shown as (5).

$$Coverage = \frac{N_{correct}}{N_{relevant}}, Novelty = \frac{N_{unknown}}{N_{known} + N_{unknown}} \quad (5)$$

In (5), $N_{correct}$ denotes the number of correctly subjects of returned results, $N_{relevant}$ denotes the number of relevant subjects of returned results, N_{unknow} denotes the number of unknown subjects of returned results, and N_{known} denotes the number of known subjects. Table V shows the comparison results of *Coverage* and *Novelty* using SPSM and Mediapedia [18].

TABLE.V
COMPARISON OF COVERAGE AND NOVELTY

Query ID	*Coverage* of SPSM	*Coverage* of Mediapedia	*Novelty* of SPSM	*Novelty* of Mediapedia
Q1	0.27	0.66	0.07	0.26
Q2	0.77	0.66	0.33	0.26
Q3	0.33	0.17	0.16	0.16
Q4	0.5	0.17	0.28	0.16
Q5	0.25	0.81	0.09	0.23
Q6	0.88	0.81	0.29	0.23
Q7	0.5	0.25	0.5	0.5
Q8	0.75	0.25	0.66	0.5

The obtained values of *Coverage* and *Novelty* using SPSM are generally higher than those of Mediapedia. But for Q1 and Q5, the above values are lower than those of Mediapedia. That is due to the restriction of query thresholds. In a word, experimental results show that SPSM obtains good performance.

VI. Conclusions

This paper presents a novel method of measuring user implicated query intention, aiming at accurate searching tourism multimedia information. Based on the proposed approach, we construct the SPSM model which can quantify the relations between user query intention and query results. The experiments show that SPSM has encouraging performances. Future research lines will focus on extracting image semantics using the technology of transfer learning [19], so as to better improve the accuracy of information annotation.

Acknowledgment

This work was supported by applied basic research programs of Yunnan Province (2010CD047) and the National Natural Science Funds (61262070).

References

[1] G Huang, S Wang, X Zhang. "Query expansion based on associated semantic space," *Journal of Computers*, vol. 6, no. 2, pp.172-177, 2011.
[2] X Peng, Z Niu, S Huang, et al. "Personalized web search using clickthrough data and web page rating," *Journal of Computers*, vol. 7, no. 10, SPL.ISS., pp 2578-2584, 2012.
[3] E Kandogan, R Krishnamurthy, S Raghavan, et al. "Avatar semantic search: a database approach to information retrieval," Proceedings of the 2006 ACM SIGMOD international conference on Management of data. ACM, 2006: 790-792.
[4] A Kiryakov, B Popov, I Terziev, et al. "Semantic annotation, indexing, and retrieval," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 2, no. 1, pp. 49-79, 2004.
[5] S Cohen, J Mamou, Y Kanza, et al. "XSEarch: A semantic search engine for XML," Proceedings of the 29th international conference on Very large data bases, Germany, 2003: 45-56.
[6] D Bollegala, Y Matsuo, M Ishizuka. "A web search engine-based approach to measure semantic similarity between words," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 977-990, 2011.
[7] J H Su, W J Huang, P S Yu, et al. "Efficient relevance feedback for content-based image retrieval by mining user navigation patterns," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 3, pp. 360-372, 2011.
[8] V Jalali, M R M Borujerdi. "Information retrieval with concept-based pseudo-relevance feedback in MEDLINE," *Knowledge and information systems*, vol. 29, no. 1, pp. 237-248, 2011.
[9] J Li, P Shi, M Cheng. "Ranking Ontologies based on formal concept analysis," *Journal of Computers*, vol. 9, no. 1, pp.215-221, 2014.
[10] L. Lin, Z. Xu, Y. Ding, "OWL Ontology Extraction from Relational Databases via Database Reverse Engineering," *Journal of Software*, vol. 8, no. 11, pp. 2749-2760, 2013.
[11] R. Setchi, Q. Tang, I. Stankov, "Semantic-based information retrieval in support of concept design," *Advanced Engineering Informatics*, vol. 25, no. 3, pp. 131-146, 2011.
[12] K. K. Chu, C. Lee, R.S. Tsai, "Ontology technology to assist learners' navigation in the concept map learning system," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11293-11299, 2011.
[13] D. Hai, F.K. Hussain, E. Chang, "A service search engine for the industrial digital ecosystems," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 6, pp. 2138-2196, 2011.
[14] B. Qin, S. Wang, X. Du, Q. Chen, Q. Wang, "Graph-based query rewriting for knowledge sharing between peer ontologies," *Information Sciences*, vol. 178, no. 18, pp. 3525-3542, 2008 .
[15] T.E. Workman, J.F. Hurdle, "Dynamic summarization of bibliographic-based data," *BMC Medical Informatics and Decision Making*, vol. 11, no. 6, pp. 1-10, 2011.
[16] L. Anthony, O. Tansel, T. Radwan, K. Keivan, J. Jamal, A. Reda, "XML materialized views and schema evolution in VIREX," *Information Sciences*, vol. 180, no. 24, pp. 4940-4957, 2010.
L. Kallipolitis, V. Karpis, I. Karali, "Semantic search in the world news domain using automatically extracted metadata files," *Knowledge-Based Systems*, vol. 27, pp. 38-50, 2012.
[17] H. Richang, Z. ZhengJun, G. Yue, C. Tat-Seng, W. Xindong, "Multimedia encyclopedia construction by mining web knowledge," *Signal Processing*, vol. 93, no. 8, pp. 2361-2368, 2013.
[18] S. J. Pan, Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

Lingling Zi was born in Liaoning Province, China. She is currently a Ph.D. candidate in the school of Computer Science. Also she is a lecturer in School of Electronic and Information Engineering, Liaoning Technical University. Her research concentrates on multimedia systems and intelligent information systems.

Xin Cong was born in Liaoning Province, China. He is currently a Ph.D. candidate and a lecturer in School of Electronic and Information Engineering, Liaoning Technical University. His research concentrates on multimedia processing and converged network.

Yaping Zhang was born in Yunna Province, China. She received the MS degree in computational mathematics from Yunnan University in 2005, and PhD degree in Computer Science from Zhejiang University in 2010, China. She is currently an associate professor in the Computer Science Department of Yunnan Normal University. Her research concentrates on real-time computer graphics and parallel computing.