# A Personalized Search Results Ranking Method Based on WordNet

*Abstract*-**Personalized search is more suitable to return search results for user based on different user's search history and improve the search efficiency and user's retrieval experience. In this paper, we put forward a search results ranking method that uses User Interest Model based on a semantic ontology WordNet, so that it returns the search results in accordance with user interestingness. Thereinto, different types of words are given by different score used different scoring mechanism, in order to obtain user interest model more accurately then provide a basis for measurement for ranking the original search results. The method advanced in this paper is proved to be practical and feasible. It improves the search effect and user's search experience to some extent.**

***Index Terms – Semantic, WordNet, User Interest Model, Personalized Search***

## I. INTRODUCTION

Information resources on the Web grow rapidly and people look for information by search engine. Although the traditional information retrieval technologies meet the most of people's needs, it can't satisfy some special queries from different users because of its universal property. It costs much time and resource to filter the exact results after search engine returning large numbers of retrieval results for user's queries. How to return more pertinence retrieval results to improve search efficiency and user's experience becomes a more pressing problem. Personalized Search is advanced in order to solve this problem, which can return the retrieval results more suitable for special user based on different user's search history, such as, query keywords, clicking states in results, visiting states in websites. Building User Interest Model (UIM) via exacting user's query history and resorting the retrieval results using UIM become the useful solution in personalized search. The paper put forward a algorithm for building user interest model based on the sematic ontology WordNet[1] then using UIM to resort the retrieval results by user's interestingness.

In recent years, there are many works about personalized or semantic search. Thereinto, Zhao[2] and Lu [3] use User Ontology and serious of semantic reasoning rules to return personalized search results. Zeng[4] implement a complete personalized retrieval system using vector space model and probability model. User Ontology or User profile advanced by the former is not nicety and the semantic relationships are so basic that the retrieval results are not satisfactory. While the system put forward by the latter is very large and complex relatively, in the search process, it only uses the rules of probability to return the result without using the keywords semantic relationships and analyzing the user's query history. The paper builds user interest model by analyzing the user's query history according to the semantic ontology WordNet and adopts scoring mechanism to express the user interestingness more smartly, which is used for scoring different types of words with different value furthermore provide a measurement for resorting the original retrieval results.

## II. WORDNET

WordNet is the English vocabulary semantic ontology is widely used. It is organized by a collection of synonym (synset). Besides of the words in synset, there are many relationships including Synonym, Hyponym/Hypernym, Meronym/Holonym, Antonym, and so on. The hierarchical description of vocabulary and the relationships between them play an important role in explaining the language behavior so that lots of psychologists and linguists explain this as a cognitive process.

The basic semantic relationship in the WordNet is synonymous relation. The synset is the basic constructing unit of WordNet with many different relationships as follows:

1) Synonymous relation, which is the set of synonyms. For example, the synonyms set of dog is domestic dog，Canis familiaris, etc;

2) Hyponym/Hypernym relation. For example, the hyponym of dog is like something is a kind of dog while the hypernym of dog is like dog is a kind of something;

3) Meronym/Holonym relation. For example, the hyponym of dog is like something is a part of dog while the Holonym of dog is like dog is a part of something;

4) Antonym relation. For example, the antonym of bad is good and the antonym of badness is goodness;

WordNet describe the relationships between words effectively which can map the relationships between words in natural language. So, we use WordNet as the basis for building user interest model. For brevity, we only consider the first three relations because a word may have many homologous words corresponding to different relations.

## III. USER INTEREST MODEL

In order to build user interest model, first of all, we should analysis user's history query keywords. For certain keyword, we extract the words which have the semantic relationships with the keyword and add them into the user interest model as nodes according to semantic relationships in WordNet. The user interest model which is similar to the vocabulary semantic web is formed updating words with the different scoring strategies in the final, whose scale is much smaller than WordNet. With new words added constantly, user is always interested in the kind of the words with higher score which standard for some type of knowledge.

In the traditional search process, when users enter the specific keywords, the search engine search for the keywords and re-sort the initial results which are based on the User interested model after we get the initial results of the sequence, and in the final, we return the results to the user.

We must constantly update the user interested model after the users enter the new specific keywords. The whole process which is resorted based on the User interested model can be expressed by the Figure 1.

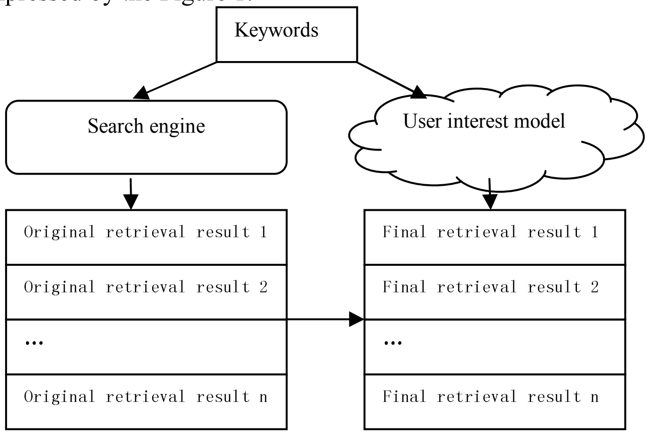


Figure 1 the whole process

### *A. The Expression of User Interest Model*

Retrieval system holds the keywords after the user queries which can be regarded as a data stream. User interest model is updated by the new keywords. This paper use the incremental updating strategy and give the related words the different score according to the relations which reflect their importance of different words in order to render the interestingness of the words. As a result, the more frequent words are, the higher their score are. Because history keywords have the order, that is the keywords which are inquired later always have more meaning than the keywords which are inquired earlier, it need multiply a factor of attenuation β when increasing the score. Because the keywords are added constantly and the scale of the user interest model becomes bigger, some old nodes must be removed in order to reduce user interest model. The main idea can be described as follows:

1)If a new keyword is exist in the original user interest model, we increase the score of the related nodes directly. That is, the node is given by five score after multiplying a factor of attenuation β. If it is not exist, we must create a new word node and give it five score.

2) Finding the following three relations between the keywords and inputted words based on the WordNet:

Synonymous relations: obtain the synonym set and insert every synonym into the original user interest model in turn. If the synonym is exist in the original user interest model, we increase the score of the related nodes directly. That is, the node is given by four score after multiplying a factor of attenuation β. Otherwise, create a new word node with four score and add a new undirected edge labeled synonym relation.

Hyponym or Hypernym relations: obtain the hyponym or hypernym set and insert every word into the original user interest model in turn. If the word is exist in the original user interest model, we increase the score of the related nodes directly. That is, the node is given by two score after multiplying a factor of attenuation β. Otherwise, create a new word node with two score and add a new directed edge labeled hyponym or hypernym relation.

Meronym or Holonym relations: obtain the meronymor holonym set and insert every word into the original user interest model in turn. If the word is exist in the original user interest model, we increase the score of the related nodes directly. That is, the node is given by one score after multiplying a factor of attenuation β. Otherwise, create a new word node with one score and add a new directed edge labeled meronym or holonym relation.

3) In order to reduce user interest model, the nodes which have the lower score must be removed after some time.

We can get a rich vocabulary of the semantic web, for example, in figure 2. In the figure, the factor of attenuation β is set to 1 in order to make the value look more clearly. But in practice, the factor of attenuation β can be set the value between 0 and 1. The nodes standard for some words expressed the keyword which user want to search while the edges represent the relationship between words. If another keyword searched by user in the future is corresponding to some node in user interest model, its score must be accumulated in order to offer an interestingness measurement when resorting the retrieval results. Along with the increasing history records, the user interest model becomes more enormous and the information becomes more and richer and it can reflect the user interestingness much better.

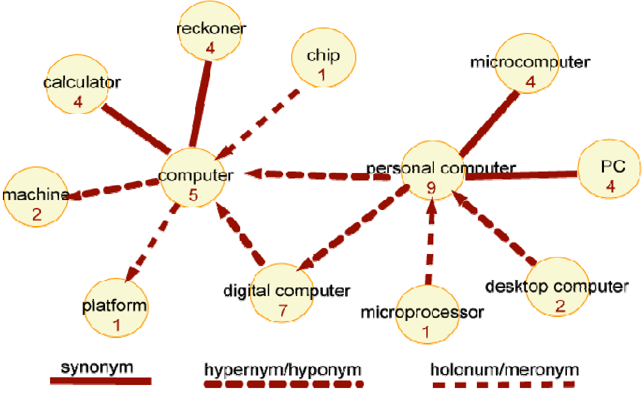


Figure 2 User Interest Model

### *B. Build User Interest Model s*

According to the idea in section 3.2, we advanced the updating algorithm of building user interest model:

**Algorithm 1:** UpdateUIM(keyword, U, β, N)
**Input:** keyword - new query word; U – original user interest model; β - attenuation coefficient; N – the number of words to delete;
**Output:** U' - the updated user interest model

```
# Add the new keyword into the original user
interest model U
If keyword is not exist in U:
    Create a new node with 5 score
Else:
        # already has the keyword node
        Update the node's score, that is, new score = old score * $\beta$ + 5
    # lookup the correlative words in WordNet and update
    synonym_set = GetSynonymSet() # get the synonym set of keyword
    For every synonym w in sysnonym_set:
        If w is not exist in U:
            Create a new node with 4 score
            Add a new undirected edge: keyword—w and label this edge synony relation
        Else:
            # already has the w node
            Update the node's score, that is, new score = old score * $\beta$ + 4
    hyponym_hypernym_set = GetHyponym_HypernymSet() # get the hyponym/hypernym set of keyword
    For every hyponym/hypernym w in hyponym_hypernym_set:
        If w is not exist in U:
            Create a new node with 2 score
            Add a new directed edge: w->keyword and label this edge hypony relation if w is a hyponym word; keyword->w if w is a hypernym word and label this edge hyperny relation
        Else:
            # already has the w node
            Update the node's score, that is, new score = old score *$\beta$ + 2
    meronym_holonym_set = GetMeronym_Holonym Set() # get the meronym/holonym set of keyword
    For every meronym/holonym w in meronym_holonym_set:
        If w is not exist in U:
            Create a new node with 1 score
            Add a new directed edge: w->keyword and label this edge meronym relation if w is a meronym word; keyword->w if w is a holonym word and label this edge holonym relation
        Else:
            # already has the w node
            Update the node's score, that is, new score = old score *$\beta$ + 1
    After some time, for example, a month later, filter the N words with lower score and remove these nodes and correlative edges from the user interest model U
    Finish the update and return the new user interest model U'
```

According to the algorithm, we can update the user interest model with new keywords constantly. If user changes his interestingness, this model also can reflect dynamically via the word's score in UIM. The construct of user interest model is the foundation of the next work.

## IV. Resorting the retrieval results

When user inputs the keyword, then submits it to the retrieval system, there are lots of the results returned by system which are general and ordinary results. So, we can filter and resort these results by user interest model and return results more purposefully. The following figure 3 shows the whole process:

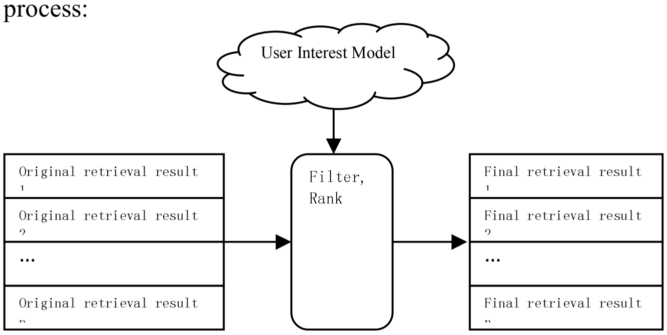


Figure 3 resorting process

Thereinto, the pivotal problem is how to weigh user's interestingness. Suppose one record X can be regard as a series of word ($x_1$, $x_2$, … , $x_n$), if one word xi is the node word in UIM, it's score can be added to the interestingness of record X denoted by Interest(X). Obviously, if the UIM is large enough, words can easily hit the network nodes and their score can be added to the interestingness. Interest(X) is initialized with 0.

After obtaining the interestingness of every result, we can resort the retrieval results according to the size of interestingness. If two results have the same interestingness, they can be arranged in the original order. Thus, the bigger interestingness the result has, the higher its rank is. According to this idea, algorithm 2 is advanced as follows:

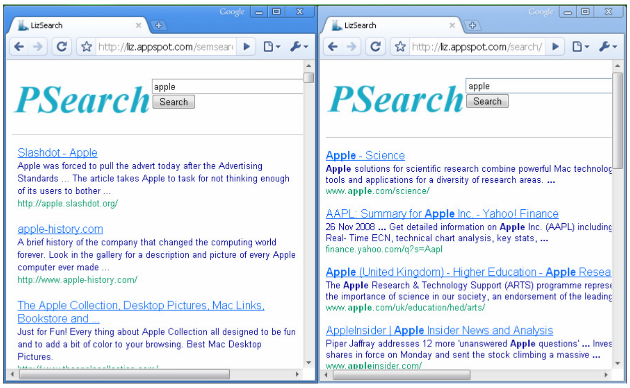
**Algorithm 2:** ReSort(the original retrieval results, U)
**Input:** ($X_1$~$X_n$) - the original retrieval results; U – user interest model
**Output:** ($Y_1$~$Y_n$) - the retrieval results after resorting

```
# Travel all original retrieval results in turn
For $X_i$ in ($X_1$~$X_n$):
    Interest ($X_i$) = 0
    # Compute the interestingness of $X_i$($x_1$, $x_2$, ··· , $x_n$)
    For $x_j$ in $X_i$ (($x_1$, $x_2$, ··· , $x_n$):
        # If $x_j$ is exist in U, add the score of $x_j$ to Interest ($X_i$)
        If U contains $x_j$:
            Interest ($X_i$) += Score ($x_j$)
    # Inserts $X_i$ into Y
    For $Y_k$ in Y:
        If Interest ($X_i$) > $Y_k$:
            # Inserts $X_i$ into the font of $Y_k$
            Break
    If the interestingness of $X_i$ is lower than all $Y_k$, put $X_i$ in the end of the Y directly, that is, in order to insert it In accordance with the original order
Y returns the result Y
```

The result of algorithm 2 is that the retrieval result contains the keywords or a correlative word user often queries has a high rank. Therefore, it can be more suitable for user and improve the user experience.

## IV. Experimental Analyses

Because personalized search depend on user interest model, the user interest model in the paper is constructed according to WordNet2.0[5] to obtain the semantic relation between user's history keywords in Google Web History. The detail procedure is implemented in algorithm 1. Then one hundred retrieval results obtained from Google Search API are returned to user after resorting according to user interest model. The whole experiment use Google App Engine[6] as deployment platform and we did some tests in this personalized search engine, shown in Figure 4.



Slashdot - Apple
Apple was forced to pull the advert today after the Advertising Standards ... The article takes Apple to task for not thinking enough of its users to bother ...
http://apple.slashdot.org/

apple-history.com
A brief history of the company that changed the computing world forever. Look in the gallery for a description and picture of every Apple computer ever made ...
http://www.apple-history.com/

The Apple Collection, Desktop Pictures, Mac Links, Bookstore and ...
Just for Fun! Every thing about Apple Collection all designed to be fun and to add a bit of color to your browsing. Best Mac Desktop Pictures.

**Apple** - Science
**Apple** solutions for scientific research combine powerful Mac technolog tools and applications for a diversity of research areas. **...**
www.**apple**.com/science/

AAPL: Summary for **Apple** Inc. - Yahoo! Finance
26 Nov 2008 **...** Get detailed information on **Apple** Inc. (AAPL) including Real- Time ECN, technical chart analysis, key stats, **...**
finance.yahoo.com/q?s=Aapl

**Apple** (United Kingdom) - Higher Education - **Apple** Resea
The **Apple** Research & Technology Support (ARTS) programme represe the importance of science in our society, an endorsement of the leading
www.**apple**.com/uk/education/hed/arts/

AppleInsider | **Apple** Insider News and Analysis
Piper Jaffray addresses 12 more 'unanswered **Apple** questions' **...** Inves shares in force on Monday and sent the stock climbing a massive **...**
www.**apple**insider.com/

Figure 4 the screenshot of personalized search engine
(Note: the results of left figure is with UIM; the results of the right figure is without UIM)

In order to prove the validity of the methods in this paper, we compare the 20 results from Google Search API to our retrieval results after algorithm 2 and count the number of results which user are satisfied with. The final result is shown in figure 5.

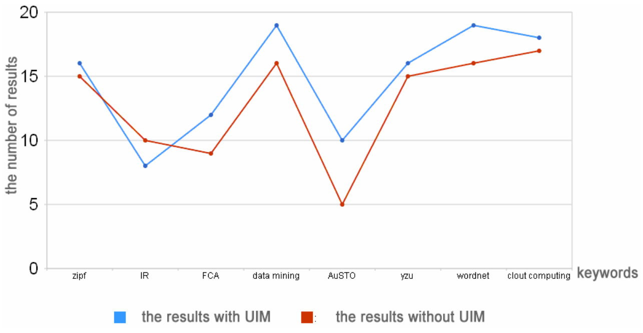


Figure 5 comparing result

From the figure 5, the results with UIM are better than the results without UIM in most case. But the satisfaction is measured based on the user's subjective will, test results in the figure above only reflect some advantages to a certain extent. However, along with updating of user interest model constantly, the results will be more tallies with the needs of user.

## V. Summaries and Outlook

The paper analyze the user history keywords to build user interest model using WordNet, then resort the original retrieval results according to user interest model, finally return the results which is more tally with the needs of user. The experiment shows the method used in the paper is effective. However, since WordNet is a English vocabulary ontology, it can't deal with Chinese vocabulary. This is a main problem. On the other hand, the construction of user interest model is based on the existed ontology, so for some search tasks in special domains, such as book search, we can construct book domain ontology then update user interest model according to the book ontology, rank the search results more suitable in the final.

## Acknowledgment

This research was supported in part by the Chinese National Natural Science Foundation under grant No. 60575035 and 60673060 , Natural Science Foundation of Jiangsu Province under contract BK2008206, and Natural Science Foundation of Education Department of Jiangsu Province under contract 08KJB520012.

## References

[1] Hirst and D. St-Onge. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In C. Fellbaum, editor, WordNet: An electronic lexical database, pages 305–332. MIT Press.

[2] Zhao Zhongmeng, Yuan Wei, He Shili. Research on the Intelligent Adjustive Algorithm for User Profile in Personalized Search Engine [J]. Computer Engineering and Applications, 2005, 41(24):184-187.

[3] Lu Linlan, Li Ming. Construction of user ontology and its application in personalized retrieval. JOURNAL OF COMPUTER APPLICATIONS, 2006, 11, Vol.26, No.11

[4] Zeng Chun, Xing Chunxiao, Zhou Lizhu. Personalized Search Algorithm Using Content-Based Filtering. Journal of Software,2003,14(05)0999:1000-9825.

[5] WordNet 2.0: http://wordnet.princeton.edu/wn2.0.shtml

[6] Google App Engine: http://code.google.com/appengine/