Research on Intelligent Semantic Search Based on HowNet

Feng Wang[1,*], Hui Zhang[2], Yizhen Wang[3] and Guanghua Zhang[2]

[1]School of Mathermatics and Statistics Science, Ludong University.Yantai 264025, China, and Key Laboratory of Language Resource Development and Application of Shandong Province; Yantai 264025, China; [2]College of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China; [3]Tilburg University, Tilburg, the Netherlands

Abstract: HDWiki is similar to the famous Wikipedia but as being a major encyclopedia in Chinese, it includes vast investment of manual effort and judgment. This structured information is divided into three categories: concept, relationship and entity. This paper describes a novel symbol oriented search that facilitates the user to format unambiguous query, especially to format complex logical query which needs several queries to get answer on traditional search but once incorporated here, it builds a bridge between what the user knows and what the user wishes and between the intent of user query and the understanding of search engine. This facilitation and understanding based on the ternary relationships are everywhere in HDWiki and HowNet powered semantic similarity measure for knowledge matching. A logical semantic oriented search system is realized and a detailed user study is conducted with users and it is found that this logical semantic oriented search and its underlying knowledge base provide significant advantages over conventional search. It offers assistance to almost every query especially complex logical queries, making the query entry more clear and efficient for user as well as for search engine, broadening the processing scope of complex logical query in the search, improving the relevance of the searched documents and increasing the probability of achieving what the user wishes to get.

Keywords: Concept, HDWiki, HowNet, Information retrieval, Logical symbol, Query formulation, Semantic similarity, Ternary relationship, Wikipedia.

1. INTRODUCTION

Whenever there is a need to acquire new knowledge and people resort to the ubiquitous search engines, the struggle is always with the same fundamental paradox: how can one describe the unknown? This diachronic question still relevant on today's internet era is precisely what must be done to form a query [1]. What makes matters worse is that search engines have insufficient capacity to comprehend and deal with these logical descriptions as people do; they instead treat a query as nothing more than an excerpt: a few words or phrases – from a relevant document. The search engines are likely to provide good relevant documents only when a query contains multiple topic-specific keywords that accurately describe the information a user needs. In short, in order to find or acquire relevant information, one must already know a great deal of what is being sought.

For knowledge seekers, there are ravines between what they know and what they wish to know, between their vague initial query and the concrete concepts and entities available, between the intent of the query and the direction to which the search engine leads to. One possible bridge to what knowledge seekers need is provision of several semantic logics and an encyclopedia: a map of semantic relations between terms and phrases.

Knowledge seekers who cannot identify what terms could impel the query to be clear and concrete as concepts even have not realized that the ambiguity of their query could benefit from an encyclopedia that automatically ensures the specific concepts by the logical inferring based on its terminology of relationships, concepts and entities, together with recommending a list of the most possible concepts to them. Those who cannot formulate a specific query, especially a complex logical query also could use the structured knowledge of an encyclopedia and semantic logical syntax to straighten out the terms to make it precise for them as well as for the search engine. This could have the potential to greatly advance the art of information retrieval. There are freely available large encyclopedias like Wikipedia, HDWiki (http://www.baike.com/), *etc*., and this research is based on HDWiki, the largest Chinese online encyclopedia like Wikipedia.

HDWiki articles contain, besides free text, various types of structured information in the form of wiki markup. These various kinds of information can be divided into three categories: concepts, relationships and entity. With this vast knowledge base, a way should be found to make the knowledge structured to be used precisely and in order to make searching more facile. The next section describes the HDWiki knowledge and its extraction way, section 3 presents our methodology to construct a bridge -- our logical semantic search process and section 4 introduces the semantic similarity measure powered by HowNet-- the indispensable concatenation at the end of the bridge. In section 5, the search system is built

*Address correspondence to this author at the School of Mathermatics and Statistics Science, Ludong University, Yantai 264025, China;
Tel: +8615066387536; E-mail: wangfenglw@126.com

and the experiments are analyzed and the context of research surrounding this work is described in section 6. The paper concluded with a summary and the pointing of issues needed to be addressed in future research (section 7).

2. HDWIKI KNOWLEDGE

Similar to Wikipedia, HDWiki contains a vast investment of manual effort and judgment and it is an open, constantly evolving encyclopedia. In a general circumstance that binary relations are present everywhere even in one article, they are included at both sides of a colon or between subheads and the below text content. Furthermore, the title is relative to all the binary relations in an article. These can be regarded as ternary relationships and each of them contains the article title.

The title of an HDWiki article is defined as a concept. A binary relation abstracted has the left relationship being a ternary relationship, and the other one which is also the last one of a ternary relation can be entities or concepts. So HDWiki knowledge is divided into three categories: concept, relationship and entity.

1. Entity is usually composed of some terms; it can be a concrete human name, an organization, a book, a game, a place, a concrete event, etc. You cannot make sure who he is just according to a name, because there are many people. Similarly, it cannot be determined what something is according to an entity. In addition, entity could contain a few concepts (defined below).
2. Concepts stand for definitive information. For example, there is one “王菲” who is China's top diva, just like there is only one article about this “王菲” in HDWiki.
3. Concepts or entities are associated with the relationships of this concept. For example, the relationships of China's top diva “王菲” include the blood type, the relations of characters, musical works, anecdotes related to her, etc.

The logical semantic oriented search relies on - HDWiki knowledge base. This information should be made practicable and available on the web under an open license. The DBpedia extraction framework adopted from Wikipedia is quite good [2], but the knowledge it does not cover Chinese language. According to the regular DOM architecture of HDWiki web, the ternary relationships are abstracted through THML parsing and this can be carried out in real time.

3. LOGICAL SEMANTIC ORIENTED SEARCH

How can the HDWiki knowledge base be used to facilitate search? – What methodology do we take to build the bridge for the user and the search engine? Here, five logical symbols are defined to guide both the user and the search engine. These are not the existing logical operators (AND, OR, NOT) but can be used together. Guide symbols described are “#C#”, “#I#”, “#P#”, “#E#” and “#S#”. The two previous symbols are used to ascertain the concept of an entity. “#C#” can be used with concept words directly, or “#I#” can be employed with specific entity or relationship which can then logically deduce the concrete concept based on the ternary relationship knowledge. Among concept, relationship and entity, the third element can be obtained if the two elements of this ternary relationship are already present. If there is a concept, then the entities can be obtained corresponding to a special relationship with “#P#”. And any third element using “#E#” can be obtained with two other elements. However, there may be a fundamental factor which cannot be ignored and then the question arises as how the terms can be matched which the user inputted to the element in the ternary relationship; this is provided in detail in section 4. Though it is possible that the above search based on ternary relationship can be carried out through the general search but there is a blurry logic that could not be comprehended by this. For example, the resemble logic. Here, a symbol “#S#” is introduced which is used to deal with those resembling logics with the help of semantic similarity measure.

Generally, the first step is to ascertain the concepts of user query to clear the search subject and then it can be helped to straighten out the search ideas for some complex logical queries. All of these are relevant for both the users and the search engines as the ravines are between them, and both of them should strive.

3.1. Semantic Concept Oriented Search

With respect to the term based query, binding a special semantic concept to a term can help to deal with word sense disambiguation problem in the information retrieval. For example, “王菲” has three concrete semantic concepts: “中国女歌手”, “四川省广元市代市□”, “大同大学教授” in HDWiki articles, and it cannot be made sure that who is “王菲” just according to these two characters.

Here, a binding symbol “#C#” is defined; the format is:

Entity#C#Concept

In the formulation above, “Entity” means the term with a vague search target and “Concept” means a concrete and unique semantic concept, abstracted from HDWiki.

For binding a semantic concept to an entity, the first direct way is binding the concept words of HDWiki, for example: “王菲#C#中国女歌手”. But most of the time, users do not know what are the special concept word, for example: “王菲” is a “歌手”, and she is a “女歌手”, she also is a “歌□天后” and there would be many more honors for her. So it is difficult that the concept terms bounded exactly match to the concept of HDWiki; this is a serious mismatch problem.

According to the semantic similarity between concept words what the user binds and HDWiki concepts, selecting the maximum semantic similarity as the right semantic concept would be a good solution to the mismatch problem mentioned above. The semantic similarity considers the co-occurrence and semantic correlation of the words; its detail is described in section 4.

In addition, it is known that a concept is always correlated to corresponding entities with some relationships. It means that if a concepts list is obtained, and with a special relationship and its corresponding entity, the concept can be ascertained which has this ternary relationship. Using the same symbol above, with the format: Entity#C# (Relationship Entity).

For example, query “王菲#C#(女儿李嫣)” means what user required. “王菲” is a concept that she or he may have a “女儿” named “李嫣”. The simple way to get this concept is to match “李嫣” to “女儿” of all concepts, and the concept of this can be found “王菲” as “中国女歌手”.

The logic shown above also needs semantic similarity calculation during the matching process, which is feasible, but two match processes should be executed; the first for relationship and then the entity should be matched in the second one. In consideration of the unnecessary process procedure, the entity of each concept can be straightforwardly checked using a new binding symbol “#I#”:Entity#I#Entity.

With this symbol, we can bind“Entity” (the first one above) with entity or relationship (the second “Entity”) for the semantic concept binding. For example, the second “Entity” can be a relationship (*e.g.* “□影”) or an entity (*e.g.* “李嫣”). In other words, “#I#” means the concept has a specific entity or a specific relationship.

In a nutshell, on the basis of HDWiki, with binding symbol “#C#” and “#I#”, the concept can be ascertained which the user required according to the bound concept or relationship or entity.

The interaction with users can strengthen the performance of semantic concept binding and this can make the semantic similarity calculation matching more better. It is that the concepts what the terms matching process procreant except the decided one can be recommended to users just in case; the semantic similarity measure process can narrow the selection scope for users to enhance the efficiency and convenience for them. A concise and visualized recommendation widget is designed as shown in section 5.

3.2. Logical Semantic Search

In many cases, what users know is just related or similar to what they wish to know. But mostly, the traditional search engines could not deal with this “related or similar” logical. And this may be worse when the dealing is with any complex logical query. The user may have to make judgments in many search processes and even have no idea how to get what they wish through the search engine. These are the issues which can be solved below by orienting the power of HDWiki knowledge base.

3.2.1. Combined Logical Search

If a user wants to know all the movies of a star, the existing powerful search engine may show a list of movies of this star and may be some other information related to him. For example, inputting “古天□□影”, Google would show the movie list and brief information of “古天□”. Now, the user needs to search an actress who is the heroine in a film of star “古天□” and does not know the film name and just thinks that this name is similar to “甜蜜”. The query what the user inputs is “古天□□影名似甜蜜女主角”, and the content (Fig. **3**) what the search engine returns with contains the query terms, which has nothing to do with what the user wished to find.

In order to find the actress, the user would have to keep trying various queries by paying a lot of time and patience. Finally, the answer may be found in a “smart” way that if the user has queried for the movies list of “古天□”, then the user would have to check them one by one according to “甜蜜” spending a lot of time, then the user may have to try the second query with a probable movie, then the third query, … and then may find the correct film name “甜言蜜□”.

Now, it is known that HDWiki has movies list of “古天□”. The semantic similarities between each movie and “甜蜜” can be computed to get the most probably right film name, and at the meantime, some other movies can be recommended to user with high semantic similarity. This seems very feasible. What needs to be done is to make the search engine understand the logical operation: Get the movies list first, and then calculate all the semantic similarities between movies and “甜蜜”, select the maximal one to take the place of “甜蜜” in the query, and display other film names which have high semantic similarity as substitution.

Here, two binding symbols “#P#” and “#S#” are defined to guide the logical operation. They are in the following formats:

Concept#P#Relationship

Entities#S#Terms

Symbol “#P#” refers to getting all corresponding entities in the “Relationship” of “Concept”. *e.g.* “古天□#P#□影” means to get all the movies of “古天□”.

“#S#” means selecting an entity which has the maximum semantic similarity with “Terms” from “Entities”. E.g. find a movie which has the maximum semantic similarity with “甜蜜” from the cinematographic work of “古天□”, the movie “甜言蜜□” will be found.

With the ternary relationships of HDWiki structured knowledge and combining the two symbols, the query “古天□#P#□影#S#甜蜜女主角” is processed to be the new query “古天□□影甜言蜜□女主角”. And the right web results can be obtained even through sending this processed query to Google. This is shown through our practical search system in section 5.

Furthermore, symbols “#P#” and “#S#” can be used separately. For example: Query “王菲#P#女儿” can get “女儿” of “王菲” from HDWiki; And we can get “美式足球” by the query “球□运□#S#足球” from HDWiki.

Without doubt, the users could expediently format other various complex logical queries with these logical semantic symbols based on the above concepts, relationships and entities of HDWiki knowledge, as well as format simple queries distinctly.

Just like the recommendation what when we bind concept using the symbol “#C#” do, in case that what symbol “#S#” selected by semantic similarity measuring probably be not user required, the recommendation widget would play a role again.

3.2.2. Query by Example

May be you would run into a situation that you wished to be similar to another you knew but had no direct link to it. For example, you know that Obama is a leader of USA but you cannot make sure that he is the president, and you want to know what the country Putin leads is, besides

Obama and Putin have the same leadership. The symbol"#E#" is defined to get the one left along with knowing other two in a ternary relationship. Syntax format is shown below:

#E#(Concept Entity)

Here, both "Concept" and "Entity" can be any two concepts, relationships and entities in all kinds of left-to-right orders. So, the query can be formatted: Putin #E# (USA Obama). This could be a process to be the new query: Putin president.

3.3. Priority for Processing

The query can be carried out using the above-defined five binding symbols in many compound modules. Each symbol has a big different function, so their priority is defined to be dealt with.

Generally, the query with logical operator is processed in the left to right order, and this rule is also applicable to our logical semantic symbols, but the semantic concept should be made sure at the beginning so the binding semantic concept symbol is set "#C#" and "#I#" having the first priority. Query by example can be analyzed anytime during the searching, as well as content of a specific relation can be obtained selecting the most correlated entity. The three "#P#", "#S#" and "#E#" should be dealt with in the general left to right order.

In addition, these logical semantic symbols can be used together with the normal logical operators including AND, OR, NOT. There is no conflict between the two groups operation.

4. SEMANTIC SIMILARITY MEASURE

As stated earlier in section 3, the required information cannot be obtained from HDWiki knowledge just by simply matching the terms. The simple example is that the user cannot get the concept "中国女歌手" through what HDWiki knowledge has by just matching the term "女歌手" to it, and it is more difficult to match "甜言蜜□" with "甜蜜". This consequentially impacts the deployment of logical semantic search based on HDWiki, so what is needed is the indispensable concatenation at the end of the bridge with user terms.

Most often, there may be same characters between the HDWiki terms user wished and the terms user bound, *e.g.* "女歌手", and further there are some uniform elements in between their semantic expressions. These two relations are defined in terms of similarity and semantic correlation. What is needed is the semantic similarity summed by these both.

It is important to note that the purpose of this semantic similarity calculation is to get the terms which have the most similar meaning with the contrasted terms. And this is different from what many other researches do.

Each of the concepts, relationships and entities consists of several terms which are here considered as phrases. And the semantic similarity is measured between the two phrases.

The similarity of terms mainly represents the influence of the same characters on the two phrases composed by the terms. Further, it should be considered that same terms have larger weight than the same characters. So the similarity of terms is measured as follows:

$$\mathbf{TS}(p,p_i) = \alpha.\frac{Sch(p,p_i)}{Lch(p)} + \beta.\frac{Ste(p,p_i)}{Lte(p)}, \qquad \boldsymbol{\alpha} + \boldsymbol{\beta} = \mathbf{1} \quad (1)$$

Where

Phrase p is the contrasted phrase and p_i is one of the selection sets,

$Sc\boldsymbol{h}(p,p_i)$ is the number of the same characters between two phrases p and p_i,

$Lc\boldsymbol{h}(p)$ is the length of phrase p in character,

$Ste(p,p_i)$ is the number of same terms between the two phrases,

$Lte(p)$ is the length of phrase p in the term,

α and β stand for two weights.

General semantic dictionary WordNet simply consists of the concept down to a tree-level system of concepts, but HowNet here is trying to describe every concept with a series of sememes [3, 4] and each term can be expressed by a number of concepts. In this work, the semantic correlation of two phrases is based on HowNet.

First of all, a phrase is divided into a list of terms using the Chinese segmentation system ICTCLAS [5]. Corresponding sememes of each term are obtained from HowNet official provision, that is, each term is defined as a vector $\mathbf{t} = [\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3} \dots \mathbf{e}_m]$, $\mathbf{e}_i$ is the i^th^ sememe of term $\mathbf{t}$ and each sememe in the vector is one dimension.

DAI`s [6, 7] modify strategy is adopted for calculating the similarity between two sememes in HowNet:

$$\mathbf{Sim}(\mathbf{e_1}, \mathbf{e_2}) = \frac{\alpha}{d+\alpha} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta \boldsymbol{h}}} \quad (2)$$

$\mathbf{Sim}(\mathbf{e_1}, \mathbf{e_2})$ is the similarity between sememe $\mathbf{e_1}$ and $\mathbf{e_2}$, d is the distance between $\mathbf{e_1}$ and $\mathbf{e_2}$ in the sememe tree, α is a parameter which means the distance when the similarity is 0.5, $\boldsymbol{h}$ is the depth of the first common parent node of the two sememes and β is a smoothing factor. α is set to be 1.6 and β is set to be 0.16 as defined in [15].

Supposing that the term t has M sememes $t_1, t_2, \dots t_M$ and the other term k corresponding to N sememes is $k_1, k_2, \dots k_N$, then the similarity between these two terms is:

$$\mathbf{Sim}(t,k) = \frac{\sum_{i=1}^{M} Max\left(\mathbf{Sim}(\mathbf{t}_i, \mathbf{k}_j)\right)}{M}, j = 1,2, \dots N \quad (3)$$

Assuming there are two phrases p and v, p is segmented into E terms $p_1, p_2 \cdots p_E$ and $\mathbf{v}$ is divided into F terms $v_1, v_2, \cdots v_F$. The similarity between p_i and $\mathbf{v}_j$ is $\mathbf{Sim}(p_i, v_j)$, where $i = 1,2, \dots E$, $j = 1,2, \dots F$ [8] and the semantic correlation between the two phrases p and v is:

$$\mathbf{SC}(p,v) = \frac{\sum_{i=1}^{E} Max\left(Sim(p_i, v_j)\right)}{E}, \qquad j = \mathbf{1}, \mathbf{2}, \dots F \quad (4)$$

Finally, the semantic similarity of the two phrases is:

$$\mathbf{SS}(p,v) = \boldsymbol{\gamma}.\mathbf{TS}(p,v) + \boldsymbol{\delta}.\mathbf{SC}(p,v)\,, \qquad \boldsymbol{\gamma} + \boldsymbol{\delta} = \mathbf{1}. \quad (5)$$

Where γ, δ represent weights respectively.

Segmentation system developed by the Chinese Academy of Sciences is used.

The phrases are sorted by the semantic similarity in the descending order; the maximum one is decided to be the right one and the other phrases whose semantic similarity is higher than a specific threshold value can be used as a recommendation for the users.

5. REALIZATION AND ANALYSIS

On the basis of the popular search engine Google, our own logical semantic search system is designed that embeds the processing module of logical semantic symbol based on HDWiki and HowNet. Users can compare the web search results between Google and our engine intuitively. Fig. (**1**) describes the architecture of this search system.

The concept bound search instance is described in Fig. (**2**).

And a big difference can be seen between the logical semantic search and general search with the query “古天□#P#□影#S#甜蜜女主角” as shown in Fig. (**3**).

And a concise and effective recommendation widget is designed like this in Fig. (**4**).

During the practice of realizing this search system, it was found that the logical semantic oriented process can be carried out before the first submit during the searching process. In other words, when the user finishes one map of binding phrase which is part of the whole query in the input box, this can be processed immediately. This may have greater efficiency and can reduce the second searching to a great extent.

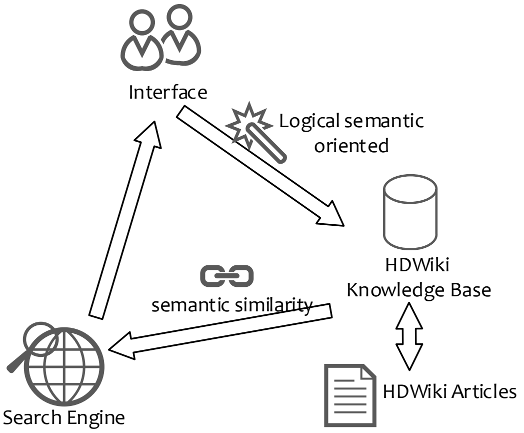
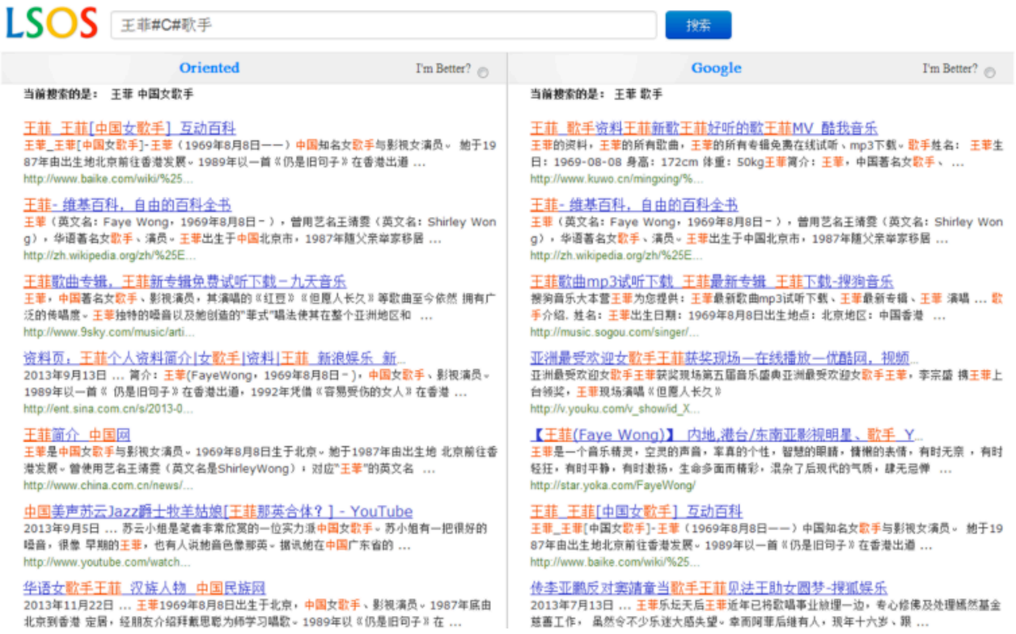


Fig. (1). Architecture of logical semantic oriented search.

6. RELATED WORK

The central idea of this research is to extract encyclopedic information from HDWiki and to use the structured knowledge to facilitate query formulation and oriented search semi-automatically. The interactive aspect is that the users are required to input the oriented logic with the five symbols initiatively and may also need to do some selecting work according to the recommendation. The automatic aspects involve abstracting of the HDWiki knowledge and processing of the query according to the logical semantic symbol with the semantic similarity measurement based on this structured information.

The most used encyclopedia in the information retrieval research is Wikipedia. The research focuses on the resources including hyperlinks, categories [9-11] and structured knowledge [1, 4]. Some users may wish to make the Wikipedia knowledge more convenient and effective; therein the structured knowledge of Wikipedia can be abstracted using DBpedia extraction framework in order to facilitate the com-



王菲#C#歌手

当前搜索的是： 王菲 中国女歌手

王菲_王菲[中国女歌手]_互动百科
王菲_王菲[中国女歌手]-王菲（1969年8月8日——）中国知名女歌手与影视女演员。她于1987年由出生地北京前往香港发展。1989年以一首《仍是旧句子》在香港出道 ...
http://www.baike.com/wiki/%25...

王菲- 维基百科，自由的百科全书
王菲（英文名：Faye Wong，1969年8月8日－），曾用艺名王靖雯（英文名：Shirley Wong），华语著名女歌手、演员。王菲出生于中国北京市，1987年随父亲举家移居 ...
http://zh.wikipedia.org/zh/%25E...

王菲歌曲专辑，王菲新专辑免费试听下载－九天音乐
王菲，中国著名女歌手、影视演员，其演唱的《红豆》《但愿人长久》等歌曲至今依然 拥有广泛的传唱度。王菲独特的嗓音以及她创造的"菲式"唱法使其在整个亚洲地区和 ...
http://www.9sky.com/music/arti...

资料页，王菲个人资料简介|女歌手|资料|王菲_新浪娱乐_新...
2013年9月13日 ... 简介：王菲(FayeWong，1969年8月8日－)，中国女歌手、影视演员。1989年以一首《 仍是旧句子》在香港出道，1992年凭借《容易受伤的女人》在香港 ...
http://ent.sina.com.cn/s/2013-0...

王菲简介_中国网
王菲是中国女歌手与影视女演员。1969年8月8日生于北京。她于1987年由出生地 北京前往香港发展。曾使用艺名王靖雯（英文名是ShirleyWong），对应"王菲"的英文名 ...
http://www.china.com.cn/news/...

中国美声苏云Jazz爵士牧羊姑娘[王菲那英合体？] - YouTube
2013年9月5日 ... 苏云小姐是笔者非常欣赏的一位实力派中国女歌手。苏小姐有一把很好的嗓音，很像 早期的王菲，也有人说她音色像那英。据讯她在中国广东省的 ...
http://www.youtube.com/watch...

华语女歌手王菲_汉族人物_中国民族网
2013年11月22日 ... 王菲1969年8月8日出生于北京，中国女歌手、影视演员。1987年底由北京到香港 定居，经朋友介绍拜戴思聪为师学习唱歌。1989年以《仍是旧句子》在 ...

当前搜索的是： 王菲 歌手

王菲_歌手资料王菲新歌王菲好听的歌王菲MV_酷我音乐
王菲的资料，王菲的所有歌曲，王菲的所有专辑免费在线试听、mp3下载。歌手姓名： 王菲生日：1969-08-08 身高：172cm 体重：50kg王菲简介：王菲，中国著名女歌手、 ...
http://www.kuwo.cn/mingxing/%...

王菲- 维基百科，自由的百科全书
王菲（英文名：Faye Wong，1969年8月8日－），曾用艺名王靖雯（英文名：Shirley Wong），华语著名女歌手、演员。王菲出生于中国北京市，1987年随父亲举家移居 ...
http://zh.wikipedia.org/zh/%25E...

王菲歌曲mp3试听下载_王菲最新专辑_王菲下载-搜狗音乐
搜狗音乐大本营王菲为您提供：王菲最新歌曲mp3试听下载、王菲最新专辑、王菲 演唱 ... 歌手介绍. 姓名：王菲出生日期：1969年8月8日出生地点：北京地区：中国香港 ...
http://music.sogou.com/singer/...

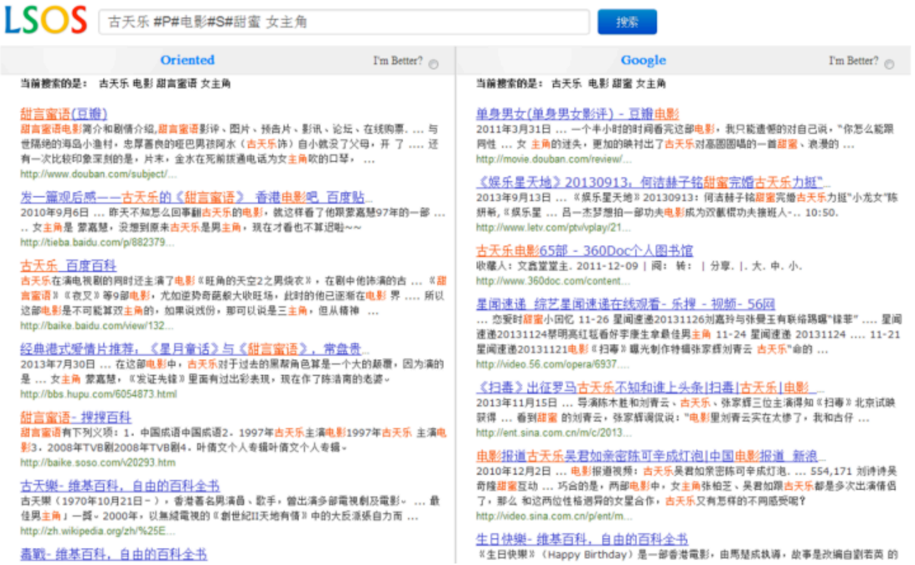
亚洲最受欢迎女歌手王菲获奖现场—在线播放—优酷网，视频...
亚洲最受欢迎女歌手王菲获奖现场第五届音乐盛典亚洲最受欢迎女歌手王菲，李宗盛 携王菲上台领奖，王菲现场演唱《但愿人长久》
http://v.youku.com/v_show/id_X...

【王菲(Faye Wong)】_内地,港台/东南亚影视明星、歌手_Y...
王菲是一个音乐精灵，空灵的声音，率真的个性，智慧的眼睛，慵懒的表情，有时无奈 ，有时轻狂，有时平静，有时激扬，生命多面而精彩，混杂了后现代的气质，肆无忌惮 ...
http://star.yoka.com/FayeWong/

王菲_王菲[中国女歌手]_互动百科
王菲_王菲[中国女歌手]-王菲（1969年8月8日——）中国知名女歌手与影视女演员。她于1987年由出生地北京前往香港发展。1989年以一首《仍是旧句子》在香港出道 ...
http://www.baike.com/wiki/%25...

传李亚鹏反对窦靖童当歌手王菲见法王助女圆梦-搜狐娱乐
2013年7月13日 ... 王菲乐坛天后王菲近年已将歌唱事业放埋一边，专心修佛及处理嫣然基金慈善工作， 虽然令不少乐迷大感失望。幸而阿菲后继有人，现年十六岁、跟 ...

Fig. (2). Concept bound search practice.



古天乐 #P#电影#S#甜蜜 女主角

当前搜索的是： 古天乐 电影 甜言蜜语 女主角

甜言蜜语(豆瓣)
甜言蜜语电影简介和剧情介绍,甜言蜜语影评、图片、预告片、影讯、论坛、在线购票. ... 与世隔绝的海岛小渔村，忠厚善良的哑巴男孩阿水（古天乐饰）自小就没了父母，开 了 .... 还有一次比较印象深刻的是，片末，金水在死前拨通电话为女主角吹的口琴， ...
http://www.douban.com/subject/...

发一篇观后感——古天乐的《甜言蜜语》_香港电影吧_百度贴...
2010年9月6日 ... 昨天不知怎么回事翻古天乐的电影，就这样看了他跟蒙嘉慧97年的一部 ... .. 女主角是 蒙嘉慧，没想到原来古天乐是男主角，现在才看也不算迟啦~~
http://tieba.baidu.com/p/882379...

古天乐_百度百科
古天乐在演电视剧的同时还主演了电影《旺角的天空2之男烧衣》，在剧中他饰演的古 ... 《甜言蜜语》《夜叉》等9部电影，尤如逆势奇葩般大收旺场，此时的他已逐渐在电影 界 .... 所以这部电影是不可能算双主角的，如果说戏份，那可以说是三主角，但从精神 ...
http://baike.baidu.com/view/132...

经典港式爱情片推荐，《星月童话》与《甜言蜜语》，常盘贵...
2013年7月30日 ... 在这部电影中，古天乐对于过去的黑帮角色算是一个大的颠覆，因为演的是 ... 女主角 蒙嘉慧，《发证先锋》里面有过出彩表现，现在作了陈浩南的老婆。
http://bbs.hupu.com/6054873.html

甜言蜜语- 搜搜百科
甜言蜜语有下列义项：1．中国成语中国成语2．1997年古天乐主演电影1997年古天乐 主演电影3．2008年TVB剧2008年TVB剧4．叶倩文个人专辑叶倩文个人专辑。
http://baike.soso.com/v20293.htm

古天樂- 维基百科，自由的百科全书
古天樂（1970年10月21日－），香港著名男演員、歌手，曾出演多部電視劇及電影。 ... 最佳男主角」一獎。2000年，以無綫電視的《創世紀II天地有情》中的大反派張自力而 ...
http://zh.wikipedia.org/zh/%25E...

毒戰- 维基百科，自由的百科全书

当前搜索的是： 古天乐 电影 甜蜜 女主角

单身男女(单身男女影评) - 豆瓣电影
2011年3月31日 ... 一个半小时的时间看完这部电影，我只能遗憾的对自己说，"你怎么能跟同性 ... 女 主角的迷失，更加的映衬出了古天乐对高圆圆唱的一首甜蜜、浪漫的 ...
http://movie.douban.com/review/...

《娱乐星天地》20130913，何洁赫子铭甜蜜完婚古天乐力挺"...
2013年9月13日 ... 《娱乐星天地》20130913：何洁赫子铭甜蜜完婚古天乐力挺"小龙女"陈妍希,《娱乐星 ... 吕一杰梦想拍一部功夫电影成为双截棍功夫接班人-.. 10:50.
http://www.letv.com/ptv/vplay/21...

古天乐电影65部 - 360Doc个人图书馆
收藏人：文鑫堂堂主. 2011-12-09 | 阅： 转： | 分享. |. 大. 中. 小.
http://www.360doc.com/content...

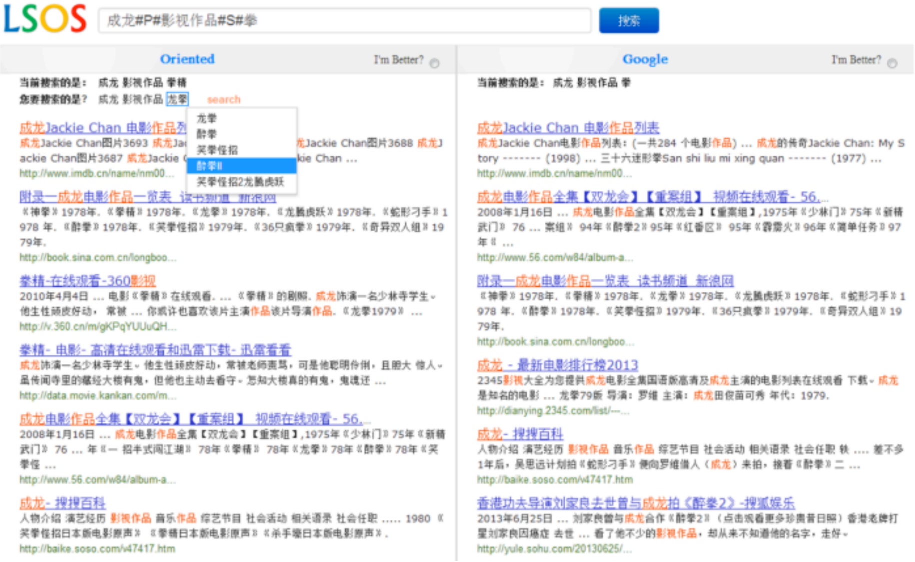
星闻速递_综艺星闻速递在线观看- 乐搜 - 视频- 56网
... 恋爱时甜蜜小回忆 11-26 星闻速递20131126刘嘉玲与张曼玉有联络踢爆"绯菲" .... 星闻速递20131124黎明高红毯看好李康生拿最佳男主角 11-24 星闻速递 20131124 .... 11-21 星闻速递20131121电影《扫毒》曝光制作特辑张家辉刘青云 古天乐"命的 ...
http://video.56.com/opera/6937....

《扫毒》出征罗马古天乐不知和谁上头条|扫毒|古天乐|电影_...
2013年11月15日 ... 导演陈木胜和刘青云、古天乐、张家辉三位主演得知《扫毒》北京试映获得 ... 看到甜蜜 的刘青云，张家辉调侃说："电影里刘青云实在太惨了，我和古仔 ...
http://ent.sina.com.cn/m/c/2013...

电影报道古天乐吴君如亲密陈可辛成灯泡|中国电影报道_新浪...
2010年12月2日 ... 电影报道视频：古天乐吴君如亲密陈可辛成灯泡. ... 554,171 刘诗诗吴奇隆甜蜜互动 ... 巧合的是，两部电影中，女主角张柏芝、吴君如跟古天乐都是多次出演情侣了，那么 和这两位性格迥异的女星合作，古天乐又有怎样的不同感受呢？
http://video.sina.com.cn/p/ent/m...

生日快樂- 维基百科，自由的百科全书
《生日快樂》（Happy Birthday）是一部香港電影，由馬楚成執導，故事是改編自劉若英 的

Fig. (3). Logical semantic search instance.



成龙#P#影视作品#S#拳

当前搜索的是： 成龙 影视作品 拳精
您要搜索的是？ 成龙 影视作品 龙拳

成龙Jackie Chan 电影作品列
成龙Jackie Chan图片3693 成龙Ja 龙Jackie Chan图片3688 成龙Jackie Chan图片3687 成龙Jackie kie Chan ...
http://www.imdb.cn/name/nm00...

附录一成龙电影作品一览表_读书频道_新浪网
《神拳》1978年. 《拳精》1978年. 《龙拳》1978年. 《龙腾虎跃》1978年. 《蛇形刁手》1978 年. 《醉拳》1978年. 《笑拳怪招》1979年. 《36只疯拳》1979年. 《奇异双人组》1979年.
http://book.sina.com.cn/longboo...

拳精-在线观看-360影视
2010年4月4日 ... 电影《拳精》在线观看. ... 《拳精》的剧照. 成龙饰演一名少林寺学生。他生性顽皮好动， 常被 ... 你或许也喜欢该片主演作品该片导演作品. 《龙拳1979》 ...
http://v.360.cn/m/gKPqYUUuQH...

拳精- 电影- 高清在线观看和迅雷下载- 迅雷看看
成龙饰演一名少林寺学生。他生性顽皮好动，常被老师责骂，可是他聪明伶俐，且胆大 惊人。虽传闻寺里的藏经大楼有鬼，但他也主动去看守。怎知大楼真的有鬼，鬼魂还 ...
http://data.movie.kankan.com/m...

成龙电影作品全集【双龙会】【重案组】_视频在线观看- 56...
2008年1月16日 ... 成龙电影作品全集【双龙会】【重案组】,1975年《少林门》75年《新精武门》 76 ... 年《一 招半式闯江湖》 78年《拳精》 78年《龙拳》78年《醉拳》78年《笑拳怪 ...
http://www.56.com/w84/album-a...

成龙- 搜搜百科
人物介绍 演艺经历 影视作品 音乐作品 综艺节目 社会活动 相关语录 社会任职 ..... 1980 《笑拳怪招日本版电影原声》 《拳精日本版电影原声》《杀手壕日本版电影原声》.
http://baike.soso.com/v47417.htm

当前搜索的是： 成龙 影视作品 拳

成龙Jackie Chan 电影作品列表
成龙Jackie Chan电影作品列表：(一共284 个电影作品) ... 成龙的传奇Jackie Chan: My Story ------- (1998) ... 三十六迷形拳San shi liu mi xing quan ------- (1977) ...
http://www.imdb.cn/name/nm00...

成龙电影作品全集【双龙会】【重案组】_视频在线观看- 56...
2008年1月16日 ... 成龙电影作品全集【双龙会】【重案组】,1975年《少林门》75年《新精武门》 76 ... 案组》 94年《醉拳2》95年《红番区》 95年《霹雳火》96年《简单任务》97年《 ...
http://www.56.com/w84/album-a...

附录一成龙电影作品一览表_读书频道_新浪网
《神拳》1978年. 《拳精》1978年. 《龙拳》1978年. 《龙腾虎跃》1978年. 《蛇形刁手》1978 年. 《醉拳》1978年. 《笑拳怪招》1979年. 《36只疯拳》1979年. 《奇异双人组》1979年.
http://book.sina.com.cn/longboo...

成龙 - 最新电影排行榜2013
2345影视大全为您提供成龙电影全集国语版高清及成龙主演的电影列表在线观看 下载。成龙是知名的电影 ... 龙拳79版 导演：罗维 主演：成龙田俊苗可秀 年代：1979.
http://dianying.2345.com/list/---...

成龙- 搜搜百科
人物介绍 演艺经历 影视作品 音乐作品 综艺节目 社会活动 相关语录 社会任职 轶 .... 差不多1年后，吴思远计划拍《蛇形刁手》便向罗维借人（成龙）来拍，接着《醉拳》二 ...
http://baike.soso.com/v47417.htm

香港功夫导演刘家良去世曾与成龙拍《醉拳2》-搜狐娱乐
2013年6月25日 ... 刘家良曾与成龙合作《醉拳2》（点击观看更多珍贵昔日照）香港老牌打星刘家良因癌症 去世 ... 看了他不少的影视作品，却从来不知道他的名字，走好。
http://yule.sohu.com/20130625/...

Fig. (4). The recommendation widget.

plex query of Wikipedia. A new search engine Koru would help guide the user interactively and would expand query automatically based on the structured knowledge, however the user would need relatively more operation time during this reciprocal process.

The work follows this theme but differs in the way that it considers that all the structured knowledge of HDWiki can be used as ternary relationships and that this information can be abstracted according to the HTML structure in real time and what is recommended to users needs less thinking and action.

Word sense disambiguation (WSD) is a natural and well known approach to the vocabulary problem in information retrieval [12] and is traditionally considered an AI-hard problem [13]. Machine learning and statistical techniques to WSD have been proved to have inherent limitations. Actually, many WSD processes are powered by WordNet [14]. Compared with these methods, the concept can be obtained directly from HDWiki with the semantic concept binding symbol, and in the future, some steps can be carried out to get it based on the structured HDWiki knowledge without any symbol.

There are a great deal of similarity measures based on WordNet [15-18]. Most of them concentrate on the similarity between two concepts mapped in the semantic network and emphasize the horizontal and vertical correlations. Yet, the similarity what we focus on is pursuing the same meaning and a simple semantic similarity measuring powered by the HowNet is chosen [19] to resolve it. In addition, phrase semantic similarity could be helpful.

CONCLUSION

This paper introduced a neoteric logical semantic oriented search that makes advantage of the manual structured knowledge of HDWiki. Our intuition is that the information in this encyclopedia could be abstracted in ternary relationships combined with the semantic similarity measuring powered by HowNet for knowledge matching and these together could provide a lot of assistance in facilitating users to format unambiguous query and complex logical query to get answer even more quickly. This can allow the users to apply the knowledge found in HDWiki to their retrieval process easily, effectively and efficiently. It has been experimented with varied query requirements on a realized search system and it was able to lend assistance to almost all queries especially the complex logical query which needed several queries on traditional search but once obtained, this significantly improved the retrieval performance. Our goal in the future is to reduce the oriented symbols to simplify the formatting as much as possible, try to give the result directly to seekers and to make good use of the concept of HDWiki and its corresponding article to improve the information retrieval in the automatic query expansion.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This research is supported by the National Natural Science Foundation of China (61300120), the University Scientific Research Foundation of Hebei Province of China (YQ2014036, YQ2013032).

REFERENCES

[1] D. N. Witten, and I. H. Nichols, “A knowledge-based search engine powered by Wikipedia”, In: *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, Lisbon, Portugal, 2007, pp. 445-454.

[2] R. Hahn, C. Bizer, C. Sahnwaldt, C. Herta, S. Robinson, M. Bürgle, H. Düwiger, and U. Scheel, “Faceted wikipedia search”, *Business Information Systems*, vol. 47, 2010, pp 1-11.

[3] X. He, L. Liu, and J. Wu, “Semantic similarity calculation based on sememe set”, *Artificial Intelligence and Computational Intelligence (AICI)*, vol. 1, pp. 423-428, 2010.

[4] D. Z. Dong, *“HowNet and Computation of Meaning”,* Singapore: World Scientific Press, 2006, pp. 197-206.

[5] ICTCLAS (Institute of Computing Technology, Chinese Lexical-Analysis System). http://ictclas.org/

[6] L. Dai, B. Liu, Y. Xia, and S. K. Wu, “Measuring Semantic Similarity between Words Using HowNet”, In: *International Conference on Computer Science and Information Technology,* Singapore, 2008, pp. 601-605.

[7] Q. Liu, and S. J. Li, “Word Semantic Similarity Computationbased on HowNet”, In: *3rd Chinese Lexical and Semanticproseminar*, Taipei, 2005. (in Chinese)

[8] Y. Liu, C. Li, P. Zhang, and Z. Xiong, “A query expansion algorithm based on phrases semantic similarity”, In: *International Symposiums on Information Processing (ISIP)*, Moscow, Russia, 2008, pp. 31-35.

[9] J. Hu, G. Wang, F. Lochovsky, and Z. Chen, “Understanding user's query intent with Wikipedia”, In: *Proceedings of the 18th International Conference on World Wide Web*, Madrid, Spain, 2009, pp. 471-480.

[10] A. Krizhanovsky, “Synonym search in Wikipedia: Synarcher”, available on: http://arxiv.org/abs/cs/0606097v2

[11] M. Völkel, M. Krötzsch, D. Vrandecic, H. Haller, and R. Studer, “Semantic Wikipedia”, In: *WWW '06 Proceedings of the 15th international Conference on World Wide Web*, Edinburgh, Scotland UK, 2006, pp. 585-594.

[12] C. Carpineto, and G. Romano, “A survey of automatic query expansion in information retrieval”, *ACM Computing Surveys,* vol. 44, no. 1, 2012.

[13] R. Navigli, and P. Velardi, “Structural semantic interconnections: a knowledge-based approach to word sense disambiguation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1075-1086, 2005.

[14] S. Liu, F. Liu , C. Yu, and W. Meng, “An effective approach to document retrieval via utilizing WordNet and recognizing phrases”, In: *SIGIR '04 Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information retrieval*, University of Sheffield, UK, 2004, pp. 266-272.

[15] A. Budanitsky, and G. Hirst, “Evaluating wordnet-based measures of lexical semantic relatedness”, *Computational Linguistics*, vol. 32, no. 1, pp. 13-47, 2006.

[16] D. Yang, and D. M. W. Powers, “Measuring semantic similarity in the taxonomy of WordNet”, In: *ACSC '05 Proceedings of the Twenty-eighth Australasian Conference on Computer Science,* Darlinghurst, Australia, vol. 38, pp. 315-322, 2005.

[17] P. Resnik, “Using information content to evaluate semantic similarity in a taxonomy”, *Proceedings of IJCAI*, vol. 1, pp. 448-453, 1995.

[18] G. Varelas, E. Voutsakis, and P. Raftopoulou, “Semantic Similarity Methods in WordNet and theirApplication to Information Retrieval on the Web”, In: *WIDM '05 Proceedings of the 7th Annual ACM International Workshop on Web Information and data Management*, Bremen, Germany, 2005, pp. 10-16.

[19] J. Hu, L. Dai, and B. Liu, “Measure Semantic Similarity between English Words”, In: *ICYCS,* Huan, China, 2008, pp. 1689-1694.

Received: June 10, 2015 Revised: July 29, 2015 Accepted: August 15, 2015