



“十一五”国家科技支撑计划重点项目



# “十一五”文化遗产保护领域 国家科技支撑计划重点项目论文集

## 古代建筑保护技术及传统工艺科学化研究 (I)

科技部社会发展科技司 编  
国家文物局博物馆与社会文物司(科技司)

文物出版社

封面设计 周小玮  
责任印制 陆 联  
责任编辑 吴 湘

### 图书在版编目(CIP)数据

古代建筑保护技术及传统工艺科学化研究 / 科技部社会发展科技司, 国家文物局博物馆与社会文物司(科技司)编. —北京: 文物出版社, 2010. 11

(“十一五”文化遗产保护领域国家科技支撑计划重点项目论文集)

ISBN 978 - 7 - 5010 - 3080 - 4

I. ①古… II. ①科…②国… III. ①古建筑 - 文物保护 - 中国 - 文集②古建筑 - 建筑艺术 - 中国 - 文集  
IV. ①TU - 87②TU - 092. 2

中国版本图书馆 CIP 数据核字(2010)第 215617 号

“十一五”文化遗产保护领域国家科技支撑计划重点项目论文集

### 古代建筑保护技术及传统工艺科学化研究

科技部社会发展科技司 编  
国家文物局博物馆与社会文物司(科技司)

· \* ·

文物出版社出版发行

(北京东直门内北小街2号楼)

<http://www.wenwu.com>

E-mail: web@wenwu.com

北京君升印刷有限公司印刷

新华书店经销

787×1092 1/16 印张:50

2010年11月第1版 2010年11月第1次印刷

ISBN 978 - 7 - 5010 - 3080 - 4 定价:200.00元(全2册)

“十一五” 文化遗产保护领域  
国家科技支撑计划课题成果论文集  
编辑委员会

主任委员：

王伟中 宋新潮

委 员：（以姓氏笔画为序）

马 涛 马清林 马燕合 王旭东 刘华彬 闫 金  
宋永华 苏伯民 吴来明 吴顺清 陈建明 陈港泉  
周文生 苗建民 郑明燕 罗 静 段 勇 麻名更

# 序 言

历史悠久、弥足珍贵的中华民族文化遗产，既是不可再生、不可替代的深厚物质资源，更是博大精深、绵延不断的文化资源和精神资源，有着重要的历史、艺术和科学价值，对国家的统一、民族的团结、社会的和谐、人民的幸福具有重要的意义。

文化遗产保护科技是一个开放的复杂巨系统，包括人文社会科学、自然科学、工程技术科学等一切与文化遗产保护相关的科学和技术。作为多学科高度交叉综合的集成体，文化遗产保护科技已经在文化遗产价值的调查、认定、研究、展示、利用和传承，文化遗产本体的保存、保全和修复，以及对文化遗产相关环境的控制与治理中发挥着越来越重要的作用。文化遗产保护科技的进步对文化遗产事业的发展具有决定性影响，已成为推动着我国从文化遗产大国向文化遗产保护强国转变的核心要素，同时也将对国家科学和技术整体发展做出贡献。

“十一五”期间，在科技部的大力支持下，文化遗产领域有4个项目15项课题列入国家科技支撑计划第一批启动项目。包括“文化遗产保护关键技术研究”、“中华文明探源工程”、“大遗址保护关键技术与开发”、“古代建筑保护技术及传统工艺科学化研究”。随后，“石质文物保护关键技术及南京报恩寺地宫出土文物保护关键技术研究”、“中华文明探源工程及相关文物保护关键技术研究”又相继获得批复实施。国家文物局以组织实施国家科技支撑计划等重大科研项目为契机，努力推动体制机制创新，积极寻找建立跨学科、跨领域、跨行业、跨部门的合作机制与模式。通过重大科技计划的实施，统筹考虑行业的技术研发、装备升级、人才培养、基地建设和体制机制创新，实现了文化遗产保护科技的跨越式发展。

截止2010年初，第一批启动的4个项目15项课题已全部通过了结项验收，据不完全统计，共研发新技术（工艺）21项，新产品、新材料、新装置36项，获得自主知识产权和专利179项，制定技术标准40项，培养博

士、硕士研究生 301 名，发表文章 513 篇，出版专著 15 本。一些科研成果已广泛应用于第三次文物普查、长城资源调查、重点文物保护单位、大遗址保护工程、灾后文化遗产抢救性保护、馆藏文物保存环境改善、博物馆展示提升等重大工程和重点工作，文化遗产保护科技含量大幅提升，行业自主创新能力得到显著提高。

为进一步做好科技成果的推广工作，我们就文化遗产领域科技支撑计划的部分成果汇编成册，这既是文化遗产保护科技成果的展示，也是向所有关心文化遗产保护的社会各界的回报。

值此，向勇于实践、不断创新的科技工作者，向文化遗产的保护者和守卫者致以崇高的敬意。

编 者

# 目 录

## 第一部分 古代建筑保护技术信息系统研发

- Ontology Based Query Expansion in Vertical Search Engine  
..... Ma Liangjun Chen Lin Gao Yibo Yang Yiping ( 3 )
- 基于知识树的领域知识组织和应用 ..... 于海涛 高一波 杨一平 ( 14 )
- Classified Vector Quantization and Population Decoding for Pattern Recognition  
..... Zhang Bailing Guan ShengUei ( 24 )
- An Image Indexing And Searching System Based Both on Keyword And Content  
..... Zhang Nan Song Yonghua ( 48 )
- Historic Chinese Architectures Image Retrieval by SVM and Pyramid Histogram  
of Oriented Gradients Features  
..... Zhang Bailing Song Yonghua Guan ShengUei Zhang Yanchun ( 57 )
- A Framework for the Massive Knowledge Web ... Sun Xiaoping Zhuge Hai Li Qing ( 75 )
- Scene Text Extraction Based on HSL  
..... Yao Jinliang Gao Yibo Ma Liangjun Yang Yiping ( 99 )
- Computing Parallel Speeded-Up Robust Features (P-SURF) via POSIX Threads  
..... Zhang Nan ( 109 )

## 第二部分 古代建筑虚拟修复及 Web 表现技术研究

- An Object Removal Method Based on Improved Texture Synthesis on Photographs  
..... Wang Fei Zheng Nanning Zhao Jizhong Niu Yanzhi ( 123 )
- A Fast, Memory-efficient and Parallelizable Arc/Circle Segmentation Algorithm  
..... Wang Fei Yuan Zejian Zheng Nanning Liu Yuehu ( 128 )
- An Exact Data Mining Method for Finding *Center Strings* and All Their Instances  
..... Lu Ruqian Jia Caiyan Zhang Shaofang Chen Lusheng Zhang Hongyu ( 140 )
- Some Cognitive Aspects of a Turing Test for Children  
..... Lu Ruqian Liu Hongge Zhang Songmao Jin Zhi Wei Zichu ( 170 )

从基于知识的软件工程到基于知件的软件工程 ..... 陆汝钤 金 芝 (183)

动画自动生成系统中情节知识库的改进策略 ..... 宫丽环 刘椿年 冀俊忠 (207)

动画自动生成中的定量动作计算 ..... 冯 佳 王 妍 刘椿年 (215)

Monitor Approach to Cumulated Anomaly in Databases  
..... Yi Junkai Lu Gang Lu Kevin (226)

Building the Knowledge Base to Support the Automatic Animation Generation of Chinese  
Traditional Architecture  
..... Wei Gongjin Bai Weijing Yin Meifang Zhang Songmao (249)

Full Life-Cycle Automatic Animation Generation Of Chinese Traditional Architecture  
..... Zhang Songmao Sun Kai (264)

面向任务的软件过程控制模型 ..... 万 静 何月娇 易军凯 (277)

基于双路索引的 XML 查询优化研究 ..... 万 静 姜 蓉 易军凯 (283)

### 第三部分 古代建筑琉璃构件保护技术及传统工艺科学化研究

古代建筑琉璃构件剥釉机理内在因素研究  
..... 苗建民 王时伟 段鸿莺 窦一村 丁银忠  
李 媛 李 合 康葆强 赵 兰 (291)

清代剥釉琉璃瓦件施釉重烧的再研究  
..... 苗建民 王时伟 康葆强 段鸿莺 李 媛  
李 合 窦一村 赵 兰 丁银忠 (307)

Research on Protection of the Architectural Glazed Ceramics in the Palace Museum, Beijing  
..... Zhao Jing Li Weidong Luo Hongjie Miao Jianmin (329)

WDXRF 对古代建筑琉璃构件胎体主次量元素定量分析方法研究  
..... 段鸿莺 梁国立 苗建民 (348)

EDXRF 无损测定琉璃构件釉主、次量元素  
..... 李 合 丁银忠 段鸿莺 梁国立 苗建民 (356)

紫禁城清代建筑琉璃构件显微结构研究 ..... 李 媛 张汝藩 苗建民 (365)

X 射线衍射法对紫禁城明清琉璃构件中脱水叶腊石的判定研究  
..... 康葆强 窦一村 吕光烈 苗建民 (373)

黄瓦窑琉璃构件胎釉原料及烧制工艺研究  
..... 康葆强 段鸿莺 丁银忠 李 合 苗建民 赵长明 富品莹 (387)

故宫建筑琉璃瓦的保护研究 ..... 赵 静 李伟东 鲁晓珂 王海圣 罗宏杰 (402)



- 用于古建筑琉璃瓦保护的高分子聚合物材料的分析测试研究  
 ..... 赵 静 李伟东 罗宏杰 (414)
- Synthesis and Properties Analysis of a New Core-Shell Silicon-Containing Fluoroacrylate  
 Latex ..... Liang Junyan He Ling Li Weidong Luo Hongjie (426)
- Preparation and Comparative Evaluation of Well-Defined Fluorinated Acrylic Copolymer  
 Latex and Solution for Ancient Stone Protection  
 ..... He Ling Liang Junyan Zhao Xiang Li Weidong Luo Hongjie (441)

#### 第四部分 古代建筑油饰彩画保护技术及传统工艺科学化研究

- 中国传统古建彩画的保护修复研究进展 ..... 齐 扬 (459)
- 在国际文物建筑修复理念影响下近百年中国传统彩画修复方法的转变  
 ..... 纪立芳 (466)
- 中国古建油饰彩画保护技术及传统工艺科学化研究  
 ..... 马 涛 白崇斌 齐 扬 (473)
- 山西省芮城县永乐宫彩画初步分析 ..... 常亚平 刘宝兰 郑庆春 (488)
- 江苏省古建筑无地仗层彩绘传统制作工艺研究 ..... 何伟俊 (496)
- 徽州地区明代古建筑彩画传统制作工艺研究 ..... 何伟俊 (509)
- 清代苏南民居彩画研究 ..... 纪立芳 (517)
- 大理白族地区传统建筑彩绘亟待保护 ..... 钱 钰 (524)
- 一种基于形状和颜色特征的建筑彩画检索方法  
 ..... 毕学刚 黄 华 王官振 马 涛 齐 扬 (531)
- 古代建筑油饰彩绘传统工艺的科学化研究  
 ..... 胡道道 李玉虎 李 娟 李会云 胡晓琴 马 涛 (543)
- 古代建筑彩绘地仗用桐油熬制作用研究 ..... 李会云 胡道道 李玉虎 (570)
- 古建彩画胶料取代与耐久性系统评价研究  
 ..... 胡晓琴 李玉虎 胡道道 李 娟 (581)
- 颐和园的彩画与彩画的保护 ..... 刘 瑗 李粮企 (588)
- 少林寺白衣殿木构彩画样品的 XRD 分析研究 ..... 王 展 (594)
- 蒙古国博格达汗宫建筑木构件彩绘工艺及材料分析  
 ..... 杨秋颖 于群力 Rocco Mazzeo (600)
- 古代壁画、文物彩绘病害调研与治理研究  
 ..... 李玉虎 邢惠萍 汪娟丽 曹 明 柏红英 (609)

- 浅谈颐和园建筑彩画的保护与修复 ..... 丛一蓬 刘 瑗 (623)
- 常熟赵用贤宅无地仗层彩绘的保护研究  
 ..... 何伟俊 杨啸秋 蒋凤瑞 龚德才 郑冬青 章忠民 王鸣军 (629)
- 传统材料及方法在江苏古建筑彩绘保护中的应用 ..... 龚德才 王鸣军 (641)
- 用于古建筑油画颜料层加固材料的合成研究 ..... 李明华 马政生 黄岳元 (647)
- 西安地区大气降尘对古建油饰彩画的影响研究 ..... 何秋菊 王丽琴 边巴顿珠 (652)
- 铅丹颜料变色因素及机理之探讨 ..... 王丽琴 樊晓蕾 王 展 (659)
- 关于铅丹变色的研究 ..... 巩梦婷 辛小虎 韩 飞 王丽琴 (666)
- 中国古代建筑油漆彩画地仗传统制作工艺及材料剖析 ..... 周文晖 王丽琴 (678)
- 湿度对古建油饰彩画的影响 ..... 何秋菊 王丽琴 严 静 (688)
- 中国北方古建油饰彩画中绿色颜料的光谱分析  
 ..... 王丽琴 严 静 樊晓蕾 马 涛 (696)
- 氨基酸分析法研究蒙古国博格达汗宫建筑彩画的胶料种类  
 ..... 杨 璐 王丽琴 黄建华 马 涛 李晓溪 (704)
- 北京颐和园古建筑上红色颜料的分析研究 ..... 严 静 王丽琴 李 立 (712)
- 文物颜料中胶结材料的分析研究 ..... 王丽琴 杨 璐 曹雪筠 李晓溪 (720)
- 古建油饰彩画的制作技术分析研究 ..... 周文晖 王丽琴 (725)
- 西安鼓楼油饰彩画主要病害分析  
 ..... 王丽琴 何秋菊 周文晖 杨 璐 马 涛 (737)
- 博格达汗宫古建柱子油饰制作工艺及材料研究  
 ..... 周文晖 王丽琴 樊晓蕾 齐 杨 马 涛 (746)
- 甘肃嘉峪关戏台建筑彩画胶结材料的氨基酸分析鉴定  
 ..... 杨 璐 王丽琴 黄建华 马 涛 (753)
- 中国古代彩绘类文物常用胶料的红外光谱特性研究  
 ..... 杨 璐 王丽琴 黄建华 李晓溪 曹雪筠 马 涛 (760)
- 两种新型文物保护材料的红外光谱及机械性能研究  
 ..... 杨 璐 王丽琴 黄建华 徐海斌 (766)
- 漫反射光谱法定量评价文物保护材料的“不改变原状”性能研究  
 ..... 杨 璐 王丽琴 黄建华 徐海斌 (771)
- 古代建筑油饰彩画组成材料及制作工艺的研究方法之探讨  
 ..... 王丽琴 杨 璐 周文晖 何秋菊 严 静 樊晓蕾 马 涛 齐 扬 (777)

# Ontology Based Query Expansion in Vertical Search Engine

Ma Liangjun<sup>1</sup>, Chen Lin<sup>1</sup>, Gao Yibo<sup>1</sup>, Yang Yiping<sup>1</sup>

(1. Institute of Automation, China Academy of Sciences, Beijing, 100190, China)

**Abstract:** Queries to search engine on the Internet are usually short, and cannot provide enough information for effective retrievals. Researchers have developed query expansion to cope with the problem and proved its usefulness. But previous researches have mainly on the general search engine and do not give the semantic enough attention. In this paper, we introduced a novel algorithm especially for the vertical search engine, which makes full use of the character that knowledge in special domain can be described more available and powerfully than that in the open domain. In the algorithm, we utilize the knowledge, formalized by ontology, to generate semantic diagraph for combinations of words in one query. And then according to the semantic distance between the vertexes in the diagraph, we selected the candidates to be added. Terms added into the initial query are obviously related in semantic with the initial one. And we experimentally show that it can improve the search result clearly.

## 1 Introduction

The proliferation of the World Wide Web prompts the wide application of search engine. Especially, the emergence of the WEB2.0 is booming the vertical search engine [1]. However, short queries [2] and the incompatibility between the terms in queries and the index have infected the performance of the two kinds of searching engine. Many researchers have invented various algorithms of query expansion to solve the short query problem. And most of them focus on the automatic query expansion [3]-[7][13]. Generally, the automatic query expansion can be categorized into local strategy and global strategy.

A query method based on local strategy usually extracts expansion terms from part of the

initial query results by applying the relevance feedback technology, with the vector processing method and the probabilistic feedback method as basic ones. Early experiment [5] has shown improvement in precision for small test in practice it. Recently [3] [13] have improved the strategy by involving user logs and applying the minimal feedback separately. But the strategy consumes much time and is affected by the first query result badly in some case; and also the users may be reluctant to select the relevant files.

As in the global strategy, researchers like to make use of the thesauri, which is co-occurrence-based, automatically built or handed-crafted, to expand the initial query words. In [4] [6] [7], researchers have examine the method and developed it. This strategy is easy to apply and more timeless. Comparing with applying the co-occurrence-based thesauri, people like to study and apply the co-occurrence-based thesauri, but in the lasted result [4], Hui Fang showed using the manual crafted thesauri, for example *WordNet*, will have a better result than just using the co-occurrence-based thesauri. And the limitation of co-occurrence-based query expansion was mentioned even earlier in [9].

But the algorithms above are mostly for the general search engine, of course, they do not pay enough attention to the knowledge in domain. Terms added into the initial queries are always lack of clear semantic relationship. But the abundant, more available and describable domain knowledge is just the specialty of the vertical search engine versus the general one. What's more, mining the relationship between terms based on the domain knowledge becomes more possible.

And then taking account of the result of [3]-[9] and [13], the more related terms added with the initial query, the much better the result shows.

In this paper, we developed a novel method of query expansion, which is based on the ontology-described knowledge and focus on the semantic relationship of terms. The method can make sure terms added are obviously related with the initial query in semantics. Furthermore, it is developed especially for the vertical search engine. In this method, we firstly establish the domain knowledge database by adopting ontology as the describing tool; secondly parse the query string into terms, and then construct semantic diagraph with each term as the first vertex based on the domain knowledge; thirdly calculate the semantic distance between the first vertex and each vertex in the semantic diagraph, and then according to the threshold, select the expanded terms of each semantic diagraph; at last, we combine all the terms gotten from the semantic diagraph with logic operator and then obtain the result of query expansion.

## 2 Domain Knowledge Structure

In this section, we'd like to introduce the structure of the domain knowledge database. And the domain knowledge database is the basis of this method introduced in this paper.

Firstly, taking into account of the ontology's powerful ability to describe knowledge, we chose it as the core of the domain knowledge sources.

Besides that, we suggest the domain knowledge source includes, not limited, the ontology, the fact database and, for each language supported in the search engine, a lexicon that includes an onomasticon, a lexicon of names, a thesaurus that only includes the synonym, and some supporting category tree, especially a shallow semantic analyzer.

And Figure 1 shows the sketch structure of the domain knowledge:

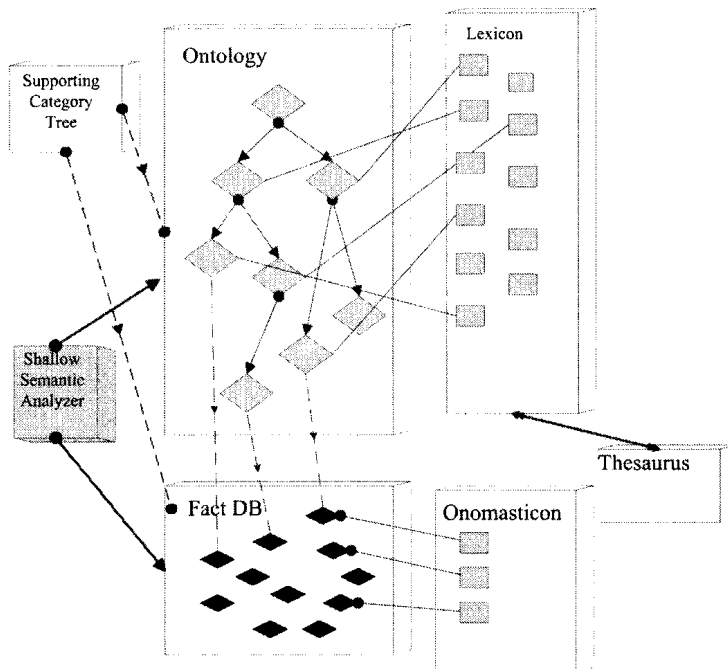


Fig. 1 the sketch structure of the domain knowledge

The ontology can be divided into two classes, according to the fact they describe, one is the main ontology which is the basic standard that most material classified according to; the other is the supporting ontology in which the items are candidates for values of the instances' of the main ontology or the ontology's slots.

For the values in the instances' or the ontology's slots which cannot be described in ontology but can be expressed in hierarchy, we store the values in hierarchy which are named after

the supporting category tree.

And the shallow semantic analyzer is used to get some special attributes from user's query. For example, a query "1940 dollars, computer", the analyzer will obtain a pair of attribute "price; 1940 dollars". In our method, it simply utilizes the regular expression to obtain the attribute pairs. e. g. the regular expression:  $[一二三四五六七八九]? [百千万]? [元] | \ d \ d * [百千万]? [元]$  to match the pattern of "三千元" (which means three thousand yuans) in Chinese, and get the attribute pair of price: 3000yuan.

### 3 Semantic Diagram

In this section, we will post a structure to express the semantic relationship among words or phrase based on the domain knowledge. Referring other researchers' result, we also give an equation to calculate the semantic distance in a diagraph, which quantifies the relationship between vertexes in the diagraph.

#### 3.1 The definition of the Semantic Diagram

Semantic diagraph is created to describe the semantic relationship among words, phrases and strings, partly referred to semantic net mentioned in [6], with its vertexes expressing the words, phrases or strings, whose edges labeled with relationship of the Tail and Head and each edge is assigned with a weight. In the semantic diagraph, we use the distance among the vertexes, the relationship between any of the two vertexes, and the weight of the edge to mark the semantic distance of the two vertexes.

So, a semantic diagraph can now be defined as follows:

*Definition 1.* A Semantic Diagraph structure is a couple

$$SD: = (V, E),$$

Where

- SD: a semantic diagraph;
- V: the vertexes of a semantic diagraph. It can be a word, phrase or string.
- E: the edges of a semantic diagraph. And it is defined as the definition 2.

*Definition 2:* An edge of a semantic diagraph between vertex  $v_i$  and vertex  $v_j$ :

$$E_{ij}: = \{v_i, v_j, r_{ij}, w_r\}$$

Where:

- $v_i$  and  $v_j$  are the two vertexes that the edge  $E_{ij}$  connected; and  $i$  is the distance of the vertex  $v_i$  to the initial vertex  $v_0$ . And the distance of the  $v_i$  and  $v_0$  is always regarded as the

jumps in nearest path of the two Vertices.

-  $r_{ij}$  is the relationship between the two vertexes, generally it can be categorized into the following categories: *hyperonymy* (Mount Tai is-a Mountain, donated with @), *hyponymy* (hyperonymy's reverse, donated with anti@), *meronymy* (computer has-a hardware, donated with %), *holonymy* (meronymy's reverse, donated with anti%), *similarity* (love similar-to like, donated with \$). Including the relationship above, there are still some a little more complicated ones. (1) *Attribute*: Mount Tai is a node of ontology WORLD-HERITAGE, with an attribute named LOCATION whose value is Tai'an, a city of China. So Tai'an is *attribute-of* Mount Tai. And Attribute is donated with #; (2) *HasAttributeValue*: In opposition to the Attribute, Mount Tai *HasAttributeValue* Tai'an, which is denoted with anti#. (3) *Instance*: to describe the relationship from class in ontology to its instances; (4) *Class*: to describe the relationship from an instance to its class in the ontology.

-  $w_r$  is the weight assigned to the relationship r.

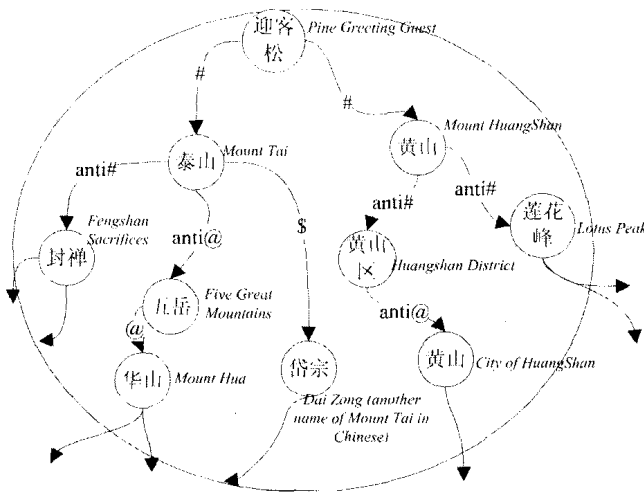


Fig. 2 part of a semantic diagram with the initial vertex “迎客松” (means the Pine Greeting Guest)

### 3.2 Semantic Distance

Semantic Diagram is established directly on the connection in the lexicon, named entity database and the ontology. The difference between two vertexes in semantic is mainly related with the two variables: the distance of the two vertexes in the nearest path, the relationship between every two connected vertexes in the nearest path.

We use the semantic distance to express the semantic difference and relation between vertexes. Consulting the study in [10] on the similarity computing based on *HowNet* that the fur-

ther the distance of the two vertexes is, the more dissimilar the two vertexes are. So we define semantic distance between the initial vertex and any other vertex  $v_i$  in the semantic diagram as follows:

*Definition 3:* The semantic distance between the initial vertex and any other vertex  $v_i$ :

$$\begin{aligned} S_i &= (-1) \cdot \log_2 \left\{ \prod_{n=1}^i T_n \cdot \left( \frac{k}{i + \lambda} \right)^m \right\} + 1 \\ &= - \sum_{n=1}^i \log_2 T_n - m \cdot \log_2 \left( \frac{k}{i + \lambda} \right) + 1 \end{aligned} \quad (1)$$

Where  $T_n$  is the weight of the relationship between vertex  $v_{n-1}$  and  $v_n$  in the nearest path from  $v_0$  to  $v_i$ , which is various and confined within  $[0, 1]$ , especially when  $i = 0$ ,  $T_n = 1$ .  $k$  and  $\lambda$  are integer and various.  $m$  is an attenuation, which is integer, various, and  $m > 2$ .

For any of the two vertexes in the diagram, assuming  $v_i$  and  $v_j$ , the semantic distance is defined as follows:

*Definition 4:* Assuming  $i > j$ , the semantic distance of the vertex  $v_i$  and  $v_j$  in the semantic diagram:

$$S_{ij} = S_i - S_j = \sum_{n=1}^j \log_2 T_n + m \cdot \log_2 \frac{k}{i + \lambda} - \sum_{m=1}^j \log_2 T_m + m \cdot \log_2 \frac{k}{j + \lambda} \quad (2)$$

Where  $T_n$  is the weight of the relationship between vertex  $v_{n-1}$  and  $v_n$  in the nearest path from  $v_0$  to  $v_j$ .

## 4 Ontology based Query Expansion in Vertical Search Engine

In this section we'd like to introduce a novel algorithm that applying the domain knowledge and the semantic diagram to select the expanded terms. Shortly the algorithm is as follows:

Assuming  $Q$  is the initial query,

(1) Shallow Semantic Analysis: utilizing the shallow semantic analyzer, we obtain the attribute pairs from  $Q$  and then delete the matched items from  $Q$ . Let the attribute pairs are  $A$ .

(2) Based on *onomasticon* in the domain knowledge source, we obtain the named entities  $N$  from the remaining of  $Q$ , and delete the matched items.

(3) Let  $Q'$  be the remaining  $Q' = \{w_1, w_2, w_3, \dots, w_n\}$  is the initial query, and  $w_i$  is the  $i$ th of all possible word combination in  $Q'$  (stop words are pruned as usual)

(4) For each attribute pair in  $A$ , and each item in  $Q'$ , we construct the semantic diagrams according to the definition of the semantic diagram in Section 3.

And then we obtain semantic diagram collection:

$$SD = \{sd \mid \text{invertex}(sd) = i, i \in A \text{ or } i \in Q'\} \quad (3)$$



Where  $inivertex(sd) = i$  means that the semantic diagraph  $sd$ 's initial vertex is  $i$ .

When building the semantic diagraph, rule below should be followed:

**Rule 1:** to generate the sub-nodes of  $v_i$  in a semantic diagraph based on the relationship named *Attribute* mentioned in Section 3, if  $v_i$  is a node in the supporting ontology or in the supporting category tree, all the nodes in the ontology that have the attribute's value equal to  $v_i$  or its sub-nodes should be added into the semantic diagraph as the sub-vertexes of  $v_i$ ;

As showed in Figure 2, it is part of a semantic diagraph created in our test environment with an initial vertex “迎客松” in Chinese (means the Pine Greeting Guests).

(5) Calculate the semantic distance between any vertex and the initial vertex in each semantic diagraph, according to the equation (1) and (2), and compare it with the threshold set before, select the vertexes, whose semantic distance to the initial vertex is smaller than the threshold, as the items to be added.

For each semantic diagraph  $sd_i$ , we can obtain the candidates:

$$EP_i = \{v_j \mid S_j < S', v_j \in sd_i\} \quad (4)$$

Where  $EP_i$  means candidates gotten from the semantic diagraph  $sd_i$ ;  $S_j$  means the semantic distance between  $v_j$  and the initial vertex of  $sd_i$ .  $S'$  means that the threshold set ahead; and the  $v_j$  means the vertex.

(6) Combine all temporal result to get the final result EPQ:

$$EPQ = \cap EP_i \quad (5)$$

When applying the combining, some special case should be pay attention to:

**Rule 2:** when  $v_i \cap v_j$  and  $v_i$  and  $v_j$  are both nodes in the main ontology or in the supporting ontology:

If  $v_i$  is sub-node of  $v_j$ ,  $v_i \cap v_j = v_i$ , else  $v_i \cap v_j = v_j$

**Rule 3:** when  $v_i \cup v_j$  and  $v_i$  and  $v_j$  are both nodes in the main ontology or in the supporting ontology:

If  $v_i$  is sub-node of  $v_j$ ,  $v_i \cup v_j = v_j$ , else  $v_i \cup v_j = v_i$

**Rule 4:** when  $v_i \cap v_j$  and  $v_i$  and  $v_j$  are not both nodes in the main ontology or in the supporting ontology:

If  $v_i$  is the same to  $v_j$ ,  $v_i \cap v_j = v_i = v_j$ ; else null.

(7) Submit the last result of the query expansion, EPQ, to the search engine. And get the result.

## 5 Experiments

In this section, we introduced the environment of our experiments, and evaluate the effec-