# 基于关键词链的动态分面研究*

王 莉
（中国科学技术信息研究所 北京 100038）

【摘要】从关键词链入手，结合形式概念分析技术，提出一种基于关键词链的动态分面方法。该方法首先采用作者关键词描述文献，然后基于相似度计算判断并合并语义上几乎一致的关键词，形成粗细不同粒度的形式背景，最后利用格技术构造搜索结果的语义分面。实证分析证明该方法可行、有效。

【关键词】关键词链 动态分面方法 形式概念分析 概念格 语义相似度
【分类号】TP391

## Dynamic Faceted Method Based on Keyword Chains

Wang Li
(Institute of Scientific & Technical Information of China, Beijing 100038, China)

【Abstract】This paper proposes a dynamic faceted method based on formal concept analysis, which extracts author keywords of literature, then merges the semantic similar keywords based on similarity calculation, and generates different granularity formal contexts. Finally, a lattice - based semantic faceted navigation is given. The empirical analysis shows that the method is feasible and effective.

【Keywords】Keyword chains Dynamic faceted method Formal concept analysis Concept lattice Semantic similarity

## 1 引 言

传统科技文献检索以关键字匹配为技术基础，简单易用，但返回的检索结果数量庞大，信息过载日益严重。解决这一问题的有效途径之一就是对搜索结果动态分面。

与传统的静态分类目录相比，动态分面具有目录结构随查询结果动态变化，且类目非空的特点，能够有效提高浏览效率。然而，动态分面在文献检索领域并没有取得突破性进展，其关键在于文本数据具有高维、稀疏和歧义的特性，在特征表示和自动聚类等诸多方面难度较大。针对这一问题，本文从关键词链入手，将形式概念分析应用到文献聚类中，提出了一种基于关键词链的动态分面方法，为用户提供易于浏览的动态语义导航。

## 2 动态分面相关研究

动态分面相关研究涉及文本表示、聚类算法、个性化动态目录、浏览效率、可视化等多个方面，主要体现出以下特征：

（1）自动聚类是动态分面的主流技术，但是在算法和应用上都存在局限。

由于文本数据具有高维、稀疏和歧义的特点，许多在低维数据空间表现良好的聚类方法无法获得好的聚类效

收稿日期：2012-07-16
收修改稿日期：2012-08-15
*本文系国家十二五科技支撑计划项目"信息资源自动处理、智能检索与STKOS应用服务集成"（项目编号：2011BHA10B05）的研究成果之一。

果；现有的高维数据聚类方法（如基于超图的聚类和子空间聚类）由于缺乏具有一定通用性的数学模型，在算法和应用上都存在局限性[1]。对海量文献数据进行有效的聚类分析仍然是一个具有挑战性的问题。

（2）采用分面分析方法，从文本描述开始建立一个完整的分面体系。

Priss[2] 提出 FKR 分面知识表示框架，并以该框架为理论基础构建了 FaIR 系统。Kashyap 等[3]介绍了 BioNav 生物医学检索系统，采用 MeSH 词表类目组织检索结果集，基于对浏览代价的评估动态生成导航树。另一个非常值得关注的发展方向是分面元数据，以元数据的方式描述主题，支持面的灵活组配与链接，其典型代表是美国加州大学伯克利分校研发的 Flamenco 检索系统。

（3）结合用户行为提升用户体验。

Ling 等[4]将用户自定义关键字作为初始分面，在其基础上采用无监督学习方法自动构建分面目录。Krohn 等[5]将形式概念分析用于分析文档和查询词之间的关系，提出了一种基于格结构的检索机制。Koren 等[6]采用协同过滤机制和个性化定制的方法生成动态目录。Ben - Yitzhak[7] 和 Dash 等[8]将动态分面与 OLAP 相结合，强调探索式搜索。

（4）越来越关注分面目录的界面呈现形式。

何超等[9]将分面导航过程中的查询及其结果构成的二元组定义为浏览状态，采用层次概念格支持用户在不同浏览状态间灵活跳转。赵金海[10]利用主题图实现可视化的分面导航。Kuo 等[11]采用云图方式对检索结果进行分析汇总。此外，还出现了对移动环境的关注，例如 FacetZoom[12] 支持适合不同屏幕大小的分面导航。FaThumb[13] 通过对大规模数据集进行层次分类，利用迭代过滤的方法支持移动环境下的内容导航。

## 3　基于关键词链的动态分面方法

### 3.1　方法概述

在对动态分面相关理论、技术及应用调研的基础上，本文利用关键词链描述文献之间潜在的语义关系，然后采用 WordNet 语义词典与模糊匹配相结合的方法计算词间语义相似度，判断并合并语义上几乎一致的关键词，生成粗细不同的多粒度形式背景，最后采用造格算法构建搜索结果的语义分面，以期为用户提供检索结果的可视化分面功能。

关键词是描述文献主题内容的词语，一个或多个相同的关键词在两篇或多篇文献中出现，反映出这两篇或多篇文献内容或作者之间存在一种潜在的关系，这种关系即关键词链[14]。通过关键词链不仅能完成文献的主题聚类，同时保留了文献之间潜在的语义关联，能够有效地帮助人们理解文献。

选择概念格技术实现动态分面，主要基于两点原因：形式概念分析应用到文档聚类上具有较好的聚类特性，同时采用格结构组织聚类结果，可以为用户提供更丰富灵活的浏览路径；与其他数据分析方法不同，形式概念分析不会人为减少信息，由形式背景构建的概念格包含了所有的数据细节。因此，采用概念格技术能够完整重现关键词链，进而揭示文献之间潜在的语义关系，有助于对文献的理解和判断。

本方法的实现流程如图 1 所示：

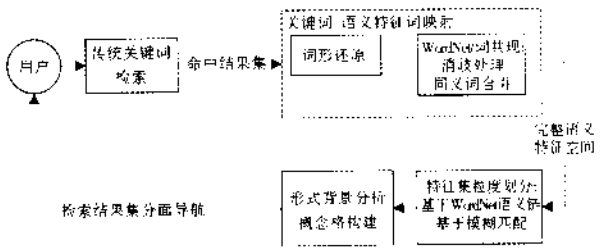


图 1　基于关键词链的动态分面模型

需要注意的是，从分面效果和效率考虑，本方法的应用需要具备两个前提条件：

（1）应用于检索结果集的分面。检索结果集是基于传统的关键词匹配技术，采用查询表达式命中的文献集合，其规模相对较小。

（2）假设用户是基于主题内容提出的查询表达，命中文献具有内容相近的特征（不考虑误检情况）。

### 3.2　文献描述

作者关键词作为科技论文写作的基本要素，在论文发表时由作者本人提出，鲜明而直观地表述了作者对其个人作品的认识，有助于读者清晰理解文献内容。虽然作者关键词没有经过主题规范，但在词义概括的灵活性和新颖性上具有自身的优势，而且易于获得。因此，本方法直接采用作者关键词（简称关键词）描述文献。多篇文献的关键词列表构成一个集合，基于字面比较进行简单去重之后，形成初始属性集。

### 3.3　特征空间的降维处理

结合本体知识或语义词典对关键词进行语义分

析，通过相似度比较发现并合并那些语义上几乎一致的关键词，从而实现维度的压缩。这种基于语义的降维方式可以显著地减少概念格的规模，同时最大程度地保留原始语义关系。

由于可利用的成熟的本体资源较少，本文的方法研究和实证研究都是基于 WordNet 展开。

(1)相似度计算

给定两个关键词 $W_1$ 和 $W_2$，如果 $W_1$ 或 $W_2$ 是短语形式，则相似度定义为词义、词形、词长和词序 4 个方面相似度的加权和[15]，如下所示：

$$Sim(W_1,W_2)=\alpha_1\times simSence(W_1,W_2)+\alpha_2\times simFormat(W_1,W_2)+\alpha_3\times simLen(W_1,W_2)+\alpha_4\times simOrder(W_1,W_2) \quad (1)$$

其中，$\alpha_1$、$\alpha_2$、$\alpha_3$ 和 $\alpha_4$ 是可调参数，分别控制词义、词形、词长和词序对关键词之间相似度的影响。具体取值要求满足条件：$\alpha_1+\alpha_2+\alpha_3+\alpha_4=1$ 且 $\alpha_1>\alpha_2>\alpha_3>\alpha_4>0$。

如果 $W_1$ 和 $W_2$ 都是单词形式，则只需要考虑单词之间的语义相似度，如下所示：

$$Sim(W_1,W_2)=simSence(W_1,W_2) \quad (2)$$

①词义相似度

假设关键词 $W_1$ 包含 x 个单词，记作 $A_1,A_2,\cdots,A_x$，关键词 $W_2$ 包含 y 个单词，记作 $B_1,B_2,\cdots,B_y$，则关键词 $W_1$ 和 $W_2$ 之间的语义相似度可以通过分别计算这些单词之间的语义相似度得到，如下所示[16]：

$$simSence(W_1,W_2)=\left(\frac{|\sum_{i=1}^{x}a_i|}{x}+\frac{|\sum_{j=1}^{y}b_j|}{y}\right)/2 \quad (3)$$

其中，$a_i$ 表示单词 $A_i$ 与 $W_2$ 中所有单词之间的最大语义相似度，假设用 s(A,B)表示两个单词之间的语义相似度，则 $a_i=\max_{j=1\cdots y}s(A_i,B_j)$。同理，$b_j$ 表示单词 $B_j$ 与 $W_1$ 中所有单词之间的最大语义相似度，记作 $b_j=\max_{i=1\cdots x}s(B_j,A_i)$。

当关键词 $W_1$ 和 $W_2$ 都是单个词汇时，即 x=1，y=1，公式(3)演变为公式(4)。

$$simSence(W_1,W_2)=s(A,B) \quad (4)$$

设单词 A 有 m 个词义：$S_{A1},S_{A2},\cdots,S_{Am}$，单词 B 有 n 个词义：$S_{B1},S_{B2},\cdots,S_{Bn}$，则可以将单词 A 和 B 的相似度定义为各个词义之间相似度的最大值，如下所示[17]：

$$s(A,B)=\max_{i=1\cdots m,j=1\cdots n}s(S_{Ai},S_{Bj}) \quad (5)$$

其中，$s(S_{Ai},S_{Bj})$ 表示两个词义之间的语义相似度，可以采用文献[18]中基于路径长度的计算方法得到。

②词形相似度

采用 Dice 系数计算方法，以单词为基本单位计算关键词 $W_1$ 和 $W_2$ 在词形上的相似性，如下所示[15]：

$$simFormat(W_1,W_2)=2\times\frac{same(W_1,W_2)}{len(W_1)+len(W_2)} \quad (6)$$

其中，$len(W_1)$、$len(W_2)$ 分别表示 $W_1$ 和 $W_2$ 包含的单词个数，$same(W_1,W_2)$ 表示 $W_1$ 和 $W_2$ 间相同单词的个数。

③词长相似度

以单词为基本单位计算关键词长度，关键词间的词长相似度定义如下所示[15]：

$$simLen(W_1,W_2)=1-\left|\frac{len(W_1)-len(W_2)}{len(W_1)+len(W_2)}\right| \quad (7)$$

其中，$len(W_1)$、$len(W_2)$ 分别表示 $W_1$ 和 $W_2$ 包含的单词个数。

④词序相似度

词序是指单词在短语中的位置，利用共有单词在两个关键词 $W_1$ 和 $W_2$ 中所处的位置信息计算其相似性。具体计算方法如下：

计算 $W_1$ 和 $W_2$ 中都出现且仅出现一次的单词，其个数用 n 表示。若 n=0，即两个关键词之间没有共有单词，则 simOrder($W_1$,$W_2$)=0；若 n=1，即两个关键词之间共有 1 个单词，则 simOrder($W_1$,$W_2$)=1；若 n>1，即两个关键词之间包含多个共有单词，则进一步比较共有单词的位置信息。

查找共有单词在关键词 $W_1$ 中的位置，构成位置向量 PFirst；用 PFirst 中的分量按对应单词在关键词 $W_2$ 中的次序排序生成向量 PSecond；进而计算 PSecond 各相邻分量的逆序数，用 RevCount 表示，则词序相似度如下：

$$simOrder(W_1,W_2)=1-\frac{RevCount}{n-1}$$

以上不同条件取值如下所示[15]：

$$simOrder(W_1,W_2)=\begin{cases}0 & \text{if } n=0\\ 1 & \text{if } n=1\\ 1-\dfrac{RevCount}{n-1} & \text{if } n>1\end{cases} \quad (8)$$

(2)降维处理关键步骤

①词形还原

针对英文关键词执行词形还原操作，利用 WordNet Morphy 实现。

②同义词合并

利用 WordNet 开放接口函数 findtheinfo_ds 查找关键词词义，返回三类结果：没有命中，表示该关键词未被收录，保留原形态；词义数目为 1，表示该关键词是单义词，得到关键词到同义词集的映射对；词义数目大于 1，表示该关键词是多义词，得到关键词到同义词集的多个映射对。

通过以上简单查询可以发现关键词间的同义关系，对那些具有明确同义关系的单义关键词直接采取合并操作。

③多义词消歧处理

利用共词现象进行消歧处理，基本思路为：依次提取两个关键词，利用 Jaccard 相似性指数[19]计算方法得到它们的共现强度，强度越大表明词间关系越密切，反之则较为疏远；采用公式(4)和(5)计算两个关键词间的语义相似度，结合共现强度进行消歧判断。

对任意一个多义词 $W_1$ 而言，若能找到一个与之共现强度最大且语义明确的关键词 $W_2$，则可以计算 $W_1$ 不同词义与 $W_2$ 之间的语义相似度，利用 $W_2$ 的语义信息对 $W_1$ 的词义进行排歧。即公式(5)中 $m>1, n=1$ 的情况，通过计算得到满足最大相似度的 i 值($1\leqslant i\leqslant m$)，则 $W_1$ 采用第 i 个词义 $S_{Ai}$。

对任意一个多义词 $W_1$ 而言，虽然找到一个与之共现强度最大的关键词 $W_j$，但 $W_j$ 仍然是一个多义词，且无法通过其他共现词进行消歧，这时需要计算两个多义词不同词义间的相似度，取相似度最大的两个词义，分别作为两个词的含义。即公式(5)中 $m>1, n>1$ 的情况，通过计算得到满足最大相似度的 i 值和 j 值($1\leqslant i\leqslant m, 1\leqslant j\leqslant n$)，则 $W_1$ 采用第 i 个词义 $S_{Ai}$，$W_2$ 采用第 j 个词义 $S_{Bj}$。

经过消歧处理后，一个关键词被唯一地归入一个同义词集，在此基础上再次采用同义词合并方法降维，最终形成一个完整的、建立在语义分析基础上的特征空间。

④基于语义相似度构建不同粒度的特征集

至此形成的特征空间包含了所有细节，如果直接用于构建概念格，不仅造格过程复杂，同时生成的格非常具体，并不利于理解和分析。因此有必要进一步合并语义相近的特征项，形成不同粒度的集合。

1)对于 WordNet 登录词可以利用语义链进行处理。

WordNet 同义词集之间通过上下位关系连接形成一条语义链，同一链上上下相邻的两个同义词集显然具有较大的语义相似度。基于这一思想，采用语义距离度量方式进一步对特征集进行划分。

取两个具有直接上下位关系的关键词 $W_1, W_2$，计算其语义距离。由于关键词间具有直接父子关系，在计算语义距离时，引入概念层次树的区域密度，具体算法参见文献[20]。

设定相似度阈值 k 对父子节点进行筛选，构建多个不同粒度的属性集。当父子节点间的距离小于 k 时，表明节点间具有较高相似度，可以将子节点归入父节点，原特征集得到进一步约简。同时，原始父子节点生成一个新的集合。

2)采用模糊匹配方式处理未登录短语。

利用公式(6)词形相似做第一步筛选，直接过滤掉词形差异大的短语；在一个筛选出的小集合中，利用公式(1)进一步比较词间相似度。通过限制两种相似度的阈值，得到若干关键词簇，簇内关键词语义近似，执行合并操作，同时生成一个新的集合。

经过以上步骤，原特征空间演变为一个粗粒度的全局特征集和若干个细粒度的局部特征集。

### 3.4　形式背景分析与建格[21]

通过语义分析，原本由一组作者关键词表示的文献转由一组具有区别性的、语义明确的特征词来描述，生成形式背景：

$$K_{D/S}:=(D,(W_D,S_D),I_D) \tag{9}$$

其中，D 表示检索结果集，$(W_D,S_D)$ 是基于语义映射构建的关键词原型－语义特征词二元组。$I_D$ 表示两个集合之间的关系，一个关键词在一篇文献中出现，则称为这个关键词与该文献相关。

由全局特征集生成初始概念格，形成检索结果集的全局视图。局部特征反映了全局视图中某一特定节点下更多的文献细节，采用嵌套线图[21]的方式逐步构造，由用户在格节点上的点击行为触发。

整个模型具有以下特点：

(1)采用作者关键词结合语义分析方法，构建文献集合的形式背景。作者关键词来自作者的自然用语，更接近用户的查询表达，以此衍生的概念格利于人的理解与分析。基于语义的降维处理合并了那些语义相似度大的特征项，同时强化了稀有关键词对文献内容辨析的贡献作用。

(2)采用粗细两种粒度分别造格。粗粒度概念格帮助用户迅速了解检索结果的全局信息，是分面导航的起点；细粒度概念格支持用户对特定节点的深入探析，进一步扩展导航能力。

(3)采用概念格呈现方式，直观反映文献之间潜在的语义关系。在格中处于越下层的节点，文献之间的关键词链强度越大。

## 4　实例分析

以“information overload”为检索词，对国家工程技术数字图书馆(http://www.istic.ac.cn)馆藏期刊文献进行检索，命中 73 篇英文文献，以其作为分析对象。

### 4.1　特征分析

从 73 篇文献中提取字面上不重复的作者关键词共计 387 个，得到 25 组共现关系，数据稀疏问题严重。

通过 WordNet 简单查询命中 73 个关键词(包括 17

个短语),仅占总量的18.9%,可见仅采用WordNet难以达到较好的降维效果。将WordNet、关键词共现和模糊匹配相结合进行语义分析,最终得到245个语义特征词,59组共现,如表1所示:

表1 特征词片段

| 特征词编号 | 语义信息 | 文献编号 |
|---|---|---|
| W1 | information handling efficiency/information processing | 1,5,8 |
| W2 | risk judgments/risk taking | 4,8 |
| W3 | consumer behavior/consumer requirements | 5,34 |
| W4 | information pull/information push - delivery | 6,23 |
| W5 | structural equation modeling | 7,16 |
| W6 | governance/IT governance | 8,26 |
| W7 | information fatigue syndrome/information anxiety | 9,14 |
| W8 | satisfaction/satisficing/end - user satisfaction | 9,39,55 |

(说明:使用JWNL1.4-RC2工具包访问WordNet2.1;相似度计算采用公式(1),参数取值为$\alpha_1=0.5,\alpha_2=0.3,\alpha_3=0.15,\alpha_4=0.05$;在划分特征集粒度时,相似度阈值设为0.7,即当两个关键词之间的相似度>0.7时,在全局特征集中进行合并,同时衍生出局部特征。)

### 4.2 形式背景分析及建格

选取具有共现关系的特征词构建文献集合的初始形式背景,最终生成的概念格具有5层结构,包含86个概念,如图2所示:

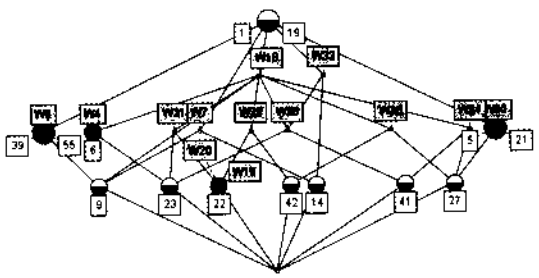

图2 概念格片段

图2给出采用ConExp1.3生成的格片段,带有深色上部半圆的节点表示该概念拥有一个属性即特征词,用深色方框标注;带有深色下部半圆的节点表示该概念拥有一个对象即文献,用白色方框标注。

### 4.3 结果分析

从图2可以看到,与互联网(W18)相关的文献有7篇。其中,文献22讨论了内容挖掘、知识处理和语义网;文献42的作者关键词多达20个,涉及信息检索、自然语言、人工智能、神经网络等领域,这两篇文献在知识管理(W30)方面具有更强的关联(详见图3)。文献6和文献23都论及了电子商务;同时文献23还提到协同技术(详见图4)。可见,概念格支持用户从多个角度观察事物,具有较好的分面导航能力。

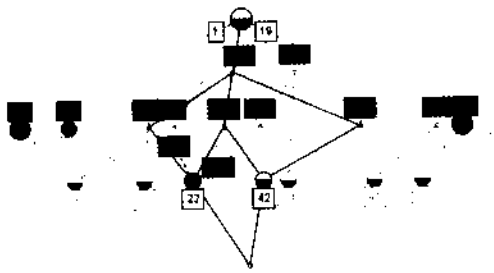

图3 结果分析-文献22与42关系展示

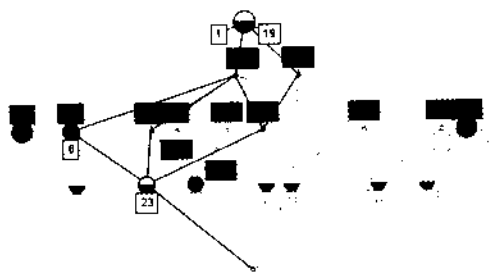

图4 结果分析-文献6与23关系展示

## 5 结 语

本文围绕搜索结果动态分面展开研究,提出一种基于关键词链的动态分面方法,特征项选取与造格是本方法的两个关键环节。前者直接关系到最终产生的分面形式是否有效,后者则是影响本方法是否可行的主要因素。

在选择特征项方面,作者关键词与WordNet语义分析相结合的方法简单灵活,容易实现,但也存在一些不足,主要体现在两个方面:作者关键词表达了作者的主观认识,存在很多不确定因素,关键词能否准确而完整地概括文献的主题,将对最终的分面效果产生直接影响;词汇的多义性非常复杂,缺少有效的词义规范手段支持,难以得到满意的降维效果。

在概念格的构造方面,本文将分面限定在检索结果集范围内,提出可以通过同义合并、相似映射、按需触发等多种方式降低格的复杂度,以此提高模型的可用性。但是,初始检索可能命中大量文献,随着文献数量的增加,计算开销和概念格的规模将迅速增长,造格的时间复杂度和空间复杂度仍然是影响其应用的主要因素。

本文提出的方法试图在可行和可用之间寻找一个平衡点。实证研究证明按照关键词链的思路构建分面

具有可行性，对检索结果的分析也是有效的。在语义分析方面，模型还有很大的改进空间，这也是未来深入研究的内容。


[1] 李伯阳. 文本聚类方法研究及其应用[D]. 厦门：厦门大学，2008.（Li Boyang. Research on Text Clustering Methods and Their Applications [D]. Xiamen: Xiamen University, 2008.）

[2] Priss U. Faceted Information Representation[C]. In: *Proceedings of the 8th International Conference on Conceptual Structures*. Germany: Shaker Verlag Aachen, 2000: 84 - 94.

[3] Kashyap A, Hristidis V, Petropoulos M, et al. Effective Navigation of Query Results Based on Concept Hierarchies [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 23(4): 540 - 553.

[4] Ling X, Mei Q Z, Zhai C X, et al. Mining Multi - faceted Overviews of Arbitrary Topics in a Text Collection [C]. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*. New York: ACM Press, 2008: 497 - 505.

[5] Krohn U, Davies N J, Weeks R. Concept Lattices for Knowledge Management [J]. *BT Technology Journal*, 1999, 17(4): 108 - 116.

[6] Koren J, Zhang Y, Liu X. Personalized Interactive Faceted Search [C]. In: *Proceedings of the 17th International Conference on World Wide Web (WWW'08)*. New York: ACM Press, 2008: 477 - 486.

[7] Ben - Yitzhak O, Golbandi N, Har'El N, et al. Beyond Basic Faceted Search [C]. In: *Proceedings of the International Conference on Web Search and Web Data Mining (WSDM'08)*. New York: ACM Press, 2008: 33 - 44.

[8] Dash D, Rao J, Megiddo N, et al. Dynamic Faceted Search for Discovery - driven Analysis [C]. In: *Proceedings of the 17th International Conference on Information and Knowledge Management (CIKM'08)*. New York: ACM Press, 2008: 3 - 12.

[9] 何超，程学旗，郭嘉丰. 面向分面导航的层次概念格模型及挖掘算法[J]. *计算机学报*，2011，34(9)：1590 - 1602.（He Chao, Cheng Xueqi, Guo Jiafeng. Mining Hierarchical Concept Lattice for Faceted Navigation [J]. *Chinese Journal of Computers*, 2011, 34(9): 1590 - 1602.）

[10] 赵金海. 基于分面主题图探索式搜索研究[J]. *情报杂志*，2012，31(1)：175 - 179.（Zhao Jinhai. Research on the Facet - based Exploratory Search in Topic Maps [J]. *Journal of Intelligence*, 2012, 31(1): 175 - 179.）

[11] Kuo B Y - L, Hentrich T, Good B M, et al. Tag Clouds for Summarizing Web Search Results [C]. In: *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*. New York: ACM Press, 2007: 1203 - 1204.

[12] Dachselt R, Frisch M, Weiland M. FacetZoom: A Continuous Multi - scale Widget for Navigating Hierarchical Metadata [C]. In: *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. New York: ACM Press, 2008: 1353 - 1356.

[13] Karlson A K, Robertson G G, Robbins D C, et al. FaThumb: A Facet - based Interface for Mobile Search [C]. In: *Proceedings of the 24th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*. New York: ACM Press, 2006: 711 - 720.

[14] 罗式胜. 篇名关键词链特征的统计分析及应用[J]. *中国图书馆学报*，1995，21(1)：27 - 29.（Luo Shisheng. The Statiscal Analysis of the Characteristics of Title Keywords [J]. *Journal of Library Science in China*, 1995, 21(1): 27 - 29.）

[15] 裴婧，包宏. 汉语句子相似度计算在 FAQ 中的应用[J]. *计算机工程*，2009，35(17)：46 - 48.（Pei Jing, Bao Hong. Application of Chinese Sentence Similarity Computation in FAQ [J]. *Computer Engineering*, 2009, 35(17): 46 - 48.）

[16] 何娟，高志强，陆青健，等. 基于词汇相似度的元素级本体匹配[J]. *计算机工程*，2006，32(16)：185 - 187.（He Juan, Gao Zhiqiang, Lu Qingjian, et al. Element Level Ontology Matching Based on Lexical Similarity [J]. *Computer Engineering*, 2006, 32(16): 185 - 187.）

[17] Wu Z, Palmer M. Verb Semantics and Lexical Selection [C]. In: *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics (ACL'94)*. Stroudsburg: Association for Computational Linguistics, 1994: 133 - 138.

[18] 刘宇鹏，李生，赵铁军. 基于 WordNet 词义消歧的系统融合[J]. *自动化学报*，2010，36(11)：1575 - 1580.（Liu Yupeng, Li Sheng, Zhao Tiejun. System Combination Based on WSD Using WordNet [J]. *Acta Automatica Sinica*, 2010, 36(11): 1575 - 1580.）

[19] 谢彩霞，梁立明，王文辉. 我国纳米科技论文关键词共现分析[J]. *情报杂志*，2005，24(3)：69 - 73.（Xie Caixia, Liang Liming, Wang Wenhui. Co - Keyword Analysis in the Filed of Nanotechnology in China [J]. *Journal of Intelligence*, 2005, 24(3): 69 - 73.）

[20] Agirre E, Rigau G. A Proposal for Word Sense Disambiguation Using Conceptual Distance [C]. In: *Proceedings of International Conference on Recent Advances in Natural Language Processing (RANLP'95)*, Tzigov Chark, 1995.

[21] Priss U E. A Graphical Interface for Document Retrieval Based on Formal Concept Analysis[C]. In: *Proceedings of the 8th Midwest Artificial Intelligence and Cognitive Science Conference*. 1997: 66 - 70.


（作者 E - mail：wangli@istic.ac.cn）