

DOI:10.13196/j.cims.2014.01.wuhongtao.0079.10.20140110

## 基于概念推荐的专利查询方法

吴鸿韬<sup>1,2</sup>, 朱怀忠<sup>1</sup>, 马建红<sup>1,2</sup>, 檀润华<sup>2+</sup>

(1. 河北工业大学 计算机科学与软件学院, 天津 300130;

2. 河北工业大学 国家技术创新方法与实施工具工程技术研究中心, 天津 300130)

**摘要:** 为辅助设计人员实现概念空间拓展, 提出一种基于概念推荐的专利查询方法。该方法利用自然语言处理技术, 从设计原型专利中提取与设计目标、实现手段相关的多方面概念推荐给设计人员, 辅助设计人员更好地理解自己的信息需求。设计人员选择的多样性概念作为查询输入, 可以帮助设计人员获取多样的专利查询结果, 有助于拓展设计人员的知识空间, 激发设计人员的灵感。

**关键词:** 专利查询; 概念推荐; 概念提取; 概念空间拓展

**中图分类号:** TP391; TH164 **文献标识码:** A

### Concept suggestion based patent query method

WU Hong-tao<sup>1,2</sup>, ZHU Huai-zhong<sup>1</sup>, MA Jian-hong<sup>1,2</sup>, TAN Run-hua<sup>2+</sup>

(1. School of Computer Science & Engineering, Hebei University of Technology, Tianjin 300130, China;

2. National Engineering Research Center for Method and Tool of Technological Innovation, Hebei University of Technology, Tianjin 300130, China)

**Abstract:** To help designers expand concept space in the process of innovation design, a concept suggestion based patent query approach was proposed. Multi-facet concepts relevant to design objects and implementation means were extracted by using natural language processing technique. These concepts were suggested to designers to help them better understand their information requirement. The diverse concepts selected by designers were used to reformulate patent query and were expected to return diverse query results which were beneficial to extend their knowledge space and ignite designers' inspiration.

**Key words:** patent query; concept suggestion; concept extraction; concept space expansion

## 0 引言

概念空间拓展是借助功能所具有的层次性和语言所具有的可推理性, 变异和拓展原始设计目标, 使设计人员在更广阔的空间中寻找更多实现原始目的的功能的概念解, 从而有助于产品创新。概念空间拓展通过生成替代功能和变体功能来实现功能空间的拓展, 通过生成面向不同领域、具有不同抽象层次的多表达形式来描述设计目标, 进而产生多样性的

检索线索, 实现知识空间拓展。概念空间拓展已经被用来生成创新设想<sup>[1-2]</sup>和类比发现<sup>[3-4]</sup>, 取得了良好的效果。然而, 创新概念的形成需要以覆盖面广、多学科交叉的知识为基础, 要求设计者经验丰富, 能够付出大量时间。

计算机辅助概念推荐(简称概念推荐)就是借助计算机高效的知识获取和推理能力, 生成多样性的、与设计目标相关的概念, 显式或隐式地推荐给设计人员, 辅助设计人员实现概念空间拓展。通过提供

收稿日期: 2012-05-04; 修订日期: 2013-10-08。Received 04 May 2012; accepted 08 Oct. 2013.

基金项目: 国家自然科学基金资助项目(51275153, 51105128); 国家创新方法工作专项资助项目(2011IM010200, 2012IM020300, 2013IM030400); 河北省博士后基金资助项目(2012-14)。Foundation items: Project supported by the National Natural Science Foundation, China(No. 51275153, 51105128), the National Innovation Method Project, China(No. 2011IM010200, 2012IM020300, 2013IM030400), and the Postdoctoral Foundation of Hebei Province, China(No. 2012-14).

具有多样性、模糊性和不确定性的设计概念,帮助设计者打破思维定式,激发创新灵感,辅助设计人员生成创新设想和发现类比赛例。已有的概念推荐方法可归纳为:

(1)基于定性推理的方法 该方法主要利用物理量之间的因果关系进行定性推理,给设计人员推荐更多的原理解,如文献[2]提出了基于效应物理量推理的功能空间拓展,文献[5]利用基于物理效应的定性推理向设计人员推荐效应原理解。

(2)基于结构化词汇表的方法 该方法主要利用结构化词汇表(如功能基<sup>[6]</sup>、WordNet<sup>[7]</sup>)来获取原始功能的同义词、上位词、下位词,实现功能概念推荐。例如文献[8]利用功能基来扩展初始化查询关键词,文献[9]利用 WordNet 构造 WordTree 来辅助设计人员发现类比赛例。

(3)基于数据挖掘的方法 该方法利用数据挖掘技术获取和功能目标相关联的拓展词,实现功能概念推荐,如文献[3]通过关联规则方法挖掘相互关联的功能动词,得到所需功能动词的拓展词,再利用查找同义词和近义词的方法得到变异词,通过专利检索获取不同领域中具有相同功能的类比赛例。

(4)基于本体的方法 该方法基于本体论构造多样性、结构化的概念集,依据特定的推理规则向设计人员推荐和设计任务相关的、多样性的概念,如文献[10]提出的专利本体、文献[11]提出的基于本体的扩展功能关联模型等。

上述方法大多着重于某一方面概念的扩展,需要设计人员能够很好地理解自己的信息需求。然而,设计人员特别是经验不足的设计人员(voice designer),通常不能正确表达自己的信息需求<sup>[12]</sup>。很多时候,用户之所以要搜索,就是因为对所检索的话题知之甚少,这时用户很难构造有效的查询。研究文献表明,只有大约 25%的查询能清晰表达用户的意图<sup>[13]</sup>。更重要的是,经验不足的设计人员通常不知道设计任务需要什么样的知识,通过搜索也只能发现 35%所需要的知识<sup>[12]</sup>。让经验不足的设计人员明确设计任务的设计目标、拟解决的问题、应用环境与功能条件,有助于理解设计任务的知识需求<sup>[12]</sup>。

作为搜索引擎中最关键的搜索策略之一,查询推荐(query suggestion)<sup>[14]</sup>可以有效地捕捉用户信息检索意图,帮助用户更好地进行信息检索,提高用户对搜索结果的满意度。因此,本文把查询推荐技

术引入工程设计领域知识检索中。

查询推荐根据所依赖数据的不同分为基于日志的方法和基于文档的方法两类<sup>[14]</sup>。基于日志的方法依靠分析搜索引擎的查询日志寻找过去出现过的相似查询,推荐给用户。然而,普通用户通常无法获取查询日志。此外,在工程设计任务中,专利搜索是设计任务相关的,大多数查询在日志中出现的次数不多,使得基于日志的方法面临严重的数据稀疏问题,查询推荐准确度较差。基于文档的方法主要通过处理包含查询的相关文档找出与查询相关的词或短语,然后利用这些词或短语构成要推荐的查询。因为 90%的工业设计活动是从已有设计出发进行变型设计(variant design)<sup>[15]</sup>,所以从设计原型专利出发,利用自然语言处理技术从专利文档中提取与设计目标、实现手段相关的概念并推荐给设计人员,是一种可行的方法。

综上所述,本文提出一种基于概念推荐的专利查询方法,该方法借助自然语言处理技术(Natural Language Processing, NLP),从设计原型专利中提取设计目标(包括拟解决的问题、用户需求、功能目标与性能目标等)、实现原理等多样性概念,推荐给设计人员,帮助设计人员更好地理解自己的信息需求;把设计人员选定的多样性概念提交到美国专利商标局(United States Patent and Trademark Office, USPTO)网站进行专利查询,以获取多样性的查询结果,实现知识空间的扩展。

### 1 查询专利获取

根据用户给定的查询专利号(即设计原型专利),本文实现的原型系统从 USPTO 网站下载 html 格式的查询专利,利用正则表达式对查询专利进行分割,获取专利的各个组成部分,如图 1 所示。

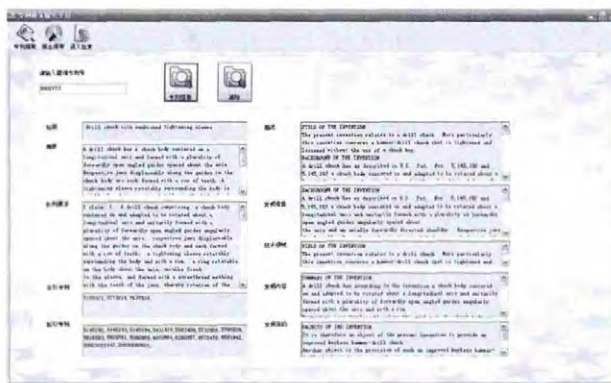


图1 专利自动获取与分割

## 2 专利内容选择

基于统计特征的方法是最常用的一种概念提取方法,但该方法不能很好地处理那些重要但出现频率很低(significant-rare)<sup>[16]</sup>的概念,如机械产品专利中关于拟解决问题的相关描述,在专利中通常只出现一次,但这些信息对设计人员理解专利设计目标、寻找具有相同设计目标的相关专利具有重要作用。因此,本文基于“问题—解”思想,从专利文本中选择信息量最丰富的组成部分进行概念提取。

英文专利通常由标题、摘要、技术领域、发明背景、发明目的、发明内容、附图说明、具体实施方案和权利要求等组成。专利的各个组成部分所要表达的内容各有侧重,从专利的各个组成部分能够提取到的知识也不尽相同。通过专利分析,本文拟从专利的如下组成部分提取不同类型的概念来辅助专利查询:

(1)标题 专利的标题通常有具体意义,且能明确、专业地表述专利发明所涉及的主题,便于从专利标题中提取专利的设计对象、特色属性、设计目标等相关信息。例如:专利 US6902172 的标题为“Drill chuck with cushioned tightening sleeve”,其中 Drill chuck 表明发明对象,cushioned tightening sleeve 表明该专利的特色属性;专利 US5031156 的标题“Method and apparatus for detecting and counting articles”中,detecting and counting articles 表明该专利的主要功能目标。

(2)摘要 摘要是对整个专利发明内容的概括,体现了专利的主要创新点,从中可以提取和设计目标、特色属性和实现手段等相关的概念。这些重要的概念可能出现多次,也可能仅出现一次。例如专利 US5031156 的摘要中 3 次出现的 air stream(含 stream of pressurized air)和 2 次出现的 detect the acoustic signal(含 The detected acoustic signal),表明了该专利实现的媒介和主要手段;专利 US7258351 的摘要中 2 次出现的 absorb a chucking reinforcement force 体现了该专利的特色功能;专利 US7243922 的摘要中 2 次出现的 adjustable depth stop 体现了该专利的特色属性;专利 US5339842 的摘要中仅出现一次的 utilize megasonic vibrations 表明了该专利的实现原理。

(3)技术领域 技术领域阐述了发明对象所属的领域,有时还会给出实现原理。以美国专利

US5031156 为例,技术领域描述“The present invention relates to article detection; and more particularly to a method and apparatus for detecting and counting articles or objects utilizing acoustics”,article detection 和 detecting and counting articles or objects 表明了该专利的功能目标,utilizing acoustics 表明了该专利的实现原理。

(4)发明背景 在发明背景部分,本文主要关注现有专利中存在的缺陷或专利中拟解决问题的相关描述,这些内容通常位于发明背景的最后一段,通常包含 however, disadvantage, drawback, difficult, remains an need, it is necessary, problem, desire, desirable 等提示词,因此可以从发明背景部分自动提取这部分信息。例如:专利 US6902172 发明背景部分“it has the disadvantage that, in particular when hammer-drilling, it vibrates excessively, making tightening and loosening difficult. Furthermore when tightened or loosened by hand it is difficult to control the maximum torque”,vibrates excessively 和 making tightening and loosening difficult 表明了该发明要解决的问题,可以看作是发明的性能目标;control the maximum torque 表明了该专利的功能目标。

(5)发明目的 发明目的部分概括性地阐述了发明的目的,可以从该部分内容中提取设计对象、性能目标、功能目标等概念。如专利 US6902172 发明目的“It is therefore an object of the present invention to provide an improved keyless hammer-drill chuck. Another object is the provision of such an improved keyless hammer-drill chuck which overcomes the above-given disadvantages, that is which is easy to tighten and loosen and where vibration is less of a problem”,从中可以提取主要设计目标 provide an improved keyless hammer-drill chuck 和 easy to tighten and loosen。

(6)权利要求 权利要求分为独立权利和从属权利两类,本文当前仅关注独立权利项的主题部分(即 comprising 前面的部分),从中提取功能设计目标的多种表达形式。如从专利 US5031156 的独立权利项的部分主题“1. Apparatus for monitoring articles delivered along a path comprising”,“4. An apparatus for counting signatures conveyed in a first direction along a path, comprising”,“15. A

method for counting articles comprising the steps of”和“33. An apparatus for counting spaced and overlapping individual signatures conveyed in a first direction along a path, comprising”中可以获取功能设计目标 monitoring articles, counting signatures, counting articles 和 counting spaced and overlapping individual signatures。

### 3 概念提取

Stanford Parser<sup>[17]</sup> 是 Stanford 大学自然语言处理 (Natural Language Processing, NLP) 小组开发的开源自然语言处理工具, 能够对英文文本的词法、句法和语义依存进行标注, 为高级自然语言的理解提供基础。Stanford Parser<sup>[17]</sup> 通过 Stanford 依存关系 (Stanford dependencies)<sup>[18]</sup> (如图 2) 提供了词与词之间的语义依存关系, 使人们可以简单、有效地从文本中进行信息提取。Stanford 依存关系<sup>[18]</sup> 是由支配对象 (governor)、受动对象 (dependent) 及两者之间的语法关系 (grammatical relation) 组成的三元组。本文利用 Stanford Parser<sup>[17]</sup> 提取专利中与设计目标、实现原理相关的概念。

```
Dependencies for Bills on ports and immigration were
submitted by Senator Brownback, Republican of Kansas
nsubj_jpass(submitted-7, Bills-1)
prep_on(Bills-1, prots-3)
Prep_on(Bills-1, immigration-5)
Conj_and(ports-3, immigration-5)
auxpass(submitted-7, were-6)
nn(Brownback-10, Senator-9)
agent(submitted-7, Brownback-10)
appos(Brownback-10, Republican-12)
prep_of(Republican-12, Kansas-14)
```

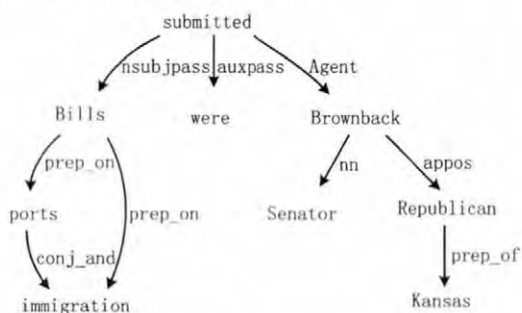


图2 Stanford依存关系

#### 3.1 功能提取

功能创新是产品创新设计的最终目的, 它贯穿于产品概念设计的各个阶段, 是概念设计最基本的元素<sup>[19]</sup>。在产品概念设计过程中, 设计人员经常需要查找具有相似功能的解决方案。在基于功能的解决方案检索中, 如何表达功能是一项非常重要的任务。

常见的功能表达方式包括: ① 功能动词<sup>[3]</sup>; ② Subject-Action-Object(SAO)三元组<sup>[20]</sup>; ③ Action-Object (AO)<sup>[21]</sup>。从平衡专利检索结果的准确率和召回率的角度来看, AO 这种功能表达方式更适合专利检索。相对于功能动词, AO 更能具体地表述用户的信息需求; 相对于 SAO, 把 AO 看作问题, 通过搜索 AO 可以获取不同的解决方案。在 AO 中: A 表达的是操作行为, O 表达的是操作对象, 从语法的角度来看, O 是 A 的直接宾语。因此, 可以利用直接宾语 (direct object, dobj) 依存关系和被动型名词主语 (passive nominal subject, nsubjpass) 依存关系进行功能提取。以美国专利 US7258351 的摘要为例, 利用 Stanford Parser<sup>[17]</sup> 生成的部分依存关系如图 3 所示。从依存关系 dobj (absorb-36, force-40) 可以提取功能 absorb force。

需要说明的是, 利用 dobj 和 nsubjpass 提取到的短语并不都可以用来表达功能, 需要对提取到的短语作进一步处理:

(1) 在提取到的 dobj 和 nsubjpass 依存关系中, 有些动词并不表达操作行为。图 3 中, dobj (has-12, element-15), dobj (includes-21, mechanism-25) 所包含的动词 has 和 include 表达的是“包含”的概念, 这些动词通常不会被设计者用来表达功能。可以通过构建非功能动词词典的方式过滤这类 dobj 关系; 而 dobj (locking-31, device-32) 中的动词现在分词 locking 在本质上是修饰名词 device 的, 功能性动词及其操作对象之间通常不包含 a, the, one 和 such 等类型的限定词, 可以依据这一事实构建相关规则, 过滤掉这类 dobj 关系。

- det (chuck-3, A-1)
- amod (chuck-3, quick-action-2)
- nsubj (has-12, chuck-3)
- prep\_in (chuck-3, particular-6)
- det (tool-10, a-8)
- nn (tool-10, power-9)
- prep\_for (particular-6, tool-10)
- det (element-15, an-13)
- amod (element-15, actuating-14)
- dobj (has-12, element-15)
- det (device-19, a-17)
- nn (device-19, control-18)
- dobj (has-12, device-19)
- conj\_and (element-15, device-19)
- nsubj (includes-21, which-20)
- rcmmod (device-19, includes-21)
- quantmod (one-24, at-22)
- mwe (at-22, least-23)
- num (mechanism-25, one-24)
- dobj (includes-21, mechanism-25)
- prep\_for (includes-21, controlling-27)
- quantmod (one-30, at-28)
- mwe (at-28, least-29)
- dobj (controlling-27, one-30)
- xcomp (controlling-27, locking-31)
- dobj (locking-31, device-32)
- nsubj (serves-34, that-33)
- rcmmod (device-32, serves-34)
- aux (absorb-36, to-35)
- xcomp (serves-34, absorb-36)
- det (force-40, a-37)
- amod (force-40, chucking-38)
- nn (force-40, reinforcement-39)
- dobj (absorb-36, force-40)
- det (device-44, the-42)
- amod (device-44, locking-43)
- prep\_via (force-40, device-44)
- nsubj (controllable-47, that-45)
- cop (controllable-47, is-46)
- rcmmod (device-44, controllable-47)
- det (device-51, the-49)
- nn (device-51, control-50)
- prep\_by (controlling-27, device-51)

图3 专利US7258351摘要利用 Stanford Parser<sup>[17]</sup>生成的依存关系片段

(2) 有些 dobj 和 nsubjpass 的关系中并不包含

操作对象,如 *dobj*(*controlling-27, one-30*), *one* 并不是功能动词 *control* 的操作对象,它后面的名词短语才是 *control* 的操作对象。此时,可以从依存关系列表中搜索包含该受动对象的依存关系,直至搜索到的依存关系的受动对象为名词为止。如果不存在其他包含该受动对象的依存关系,则从该受动对象的下一个词(通过词后面的序号来标志)开始搜索,直至搜索到受动对象为名词的依存关系为止。在本例中,不存在包含 *one* 的其他依赖关系,只能从 *one* 后的 *locking* 开始搜索,得到依存关系 *dobj*(*locking-31, device-32*)。该依存关系中的受动对象 *device* 为名词时,搜索停止,从而得到 *control* 的操作对象 *device*。

(3)在某些 *dobj* 和 *nsubjpass* 依存关系中,组成依存关系的名词并没有完整表达出操作对象,因此需要通过搜索包含该名词的依存关系集合,补充完整操作对象。例如,对于 *dobj*(*absorb-36, force-40*),通过 *absorb* 和 *force* 各自的序号可以判断出两者之间存在其他单词,因此可以通过搜索包含 *force* 的依存关系 *amod*(*force-40, chucking-38*), *nn*(*force-40, reinforcement-39*)获取 *absorb* 的完整操作对象 *chucking reinforcement force*。

### 3.2 属性提取

Dewulf 认为属性(property)表达了产品的专有特征,通常用形容词来描述,通过属性能够发现直接类比,而通过功能只能发现间接类比<sup>[22]</sup>。一种功能可以通过多种属性来实现,而属性的变化也会产生新的功能。通过分析产品属性和功能的变化(range)和范围(scale or range),可以基于属性—功能的设计方法指导创新设计<sup>[22]</sup>。为了提高专利查询效率,本文利用“形容词+名词”的表达方式来表达专利的特色属性。Stanford Parser<sup>[17]</sup>提供的形容词修饰(adjectival modifier, *amod*)依赖关系表达了形容词及其修饰的名词之间的语义依赖关系,可以用来提取专利的特色属性。在提取的过程中,采用 3.1 节功能提取中的相似处理方法来获取完整的特色属性描述。如从专利 US7243922 摘要提取到的相关依赖关系为 *amod*(*stop-4, adjustable-2*), *nn*(*stop-4, depth-3*),从中可以提取特色属性 *adjustable depth stop*。

### 3.3 设计对象提取

专利的设计对象通常由名词短语来表达,常见的名词短语表达方式除了“形容词+名词”外,还包

括“名词+名词”、“名词+介词 of 短语”。因此,本文利用形容词修饰 *amod*、名词组合修饰(*noun compound modifier, nn*)、介词修饰(*prepositional modifier, prep*) (本文只考虑介词 *of*)来提取设计对象。以美国专利 6902172 为例,可以从标题“Drill chuck with cushioned tightening sleeve”中根据依赖关系 *nn*(*chuck-8, drill-7*),提取发明对象 *drill chuck*。

### 3.4 设计目标提取

专利的技术领域、发明背景、发明目的等组成部分中,从不同角度论述了专利的设计目标。有些专利的发明背景部分也存在与设计目标相关的描述。本文总结了如下设计目标表达形式:

(1)形容词+名词 如 *low reliability* 和 *bad security*,这类设计目标通常与性能相关,可以利用 *amod* 依存关系提取,用来检索具有相同性能设计目标的专利。

(2)动词+名词 如 *control maximum torque*,这类设计目标可以利用 *dobj* 和 *nsubjpass* 依存关系提取,用来检索实现同一设计目标的同类或异类专利。

(3)动词+副词 如 *vibrates excessively* 和 *re-assembled easily*。*vibrates excessively* 表达了专利拟解决的问题,本文也将其看作是设计目标。该类设计目标可以利用包含动词的副词修饰语依存关系(*adverbial modifier, advmod*)来提取。这类设计目标可以用来检索解决同一问题的同类或异类专利(如 *vibrates excessively*),或用来检索具有相同性能设计目标的同类专利(如 *reassembled easily*)。

(4)形容词+to+动词 如 *easy to tighten and loosen* 和 *easier to operate*,这类设计目标通常与性能相关,形容词和后面动词的依存关系可以用开放式分句补语(*open clausal complement, xcomp*)依存关系来表达。例如利用 *xcomp*(*easy-23, tighten-25*)和两者之间的介词 *to* 可以提取 *easy to tighten*。如果 *xcomp* 中的动词与其后的某个动词之间存在联合(*conjunct, conj*)依存关系,如 *conj\_and*(*tighten-25, loosen-27*),则可以提取完整的 *easy to tighten and loosen*。

(5)名词+of+名词 如 *convenience of operation*,这类设计目标本质上是“形容词+名词”的不同表达形式,可以通过 *prep* 来提取,例如从依存关系 *prep\_of*(*convenience-9, operation-15*)中提取 *convenience of operation*。

(6) make something + 形容词 如 making tightening and loosening difficult, make the torque transmission off-balance。当 make 后的宾语为名词短语时, Stanford Parser<sup>[17]</sup> 认为 make 和作为宾语补足语的形容词之间具有 xcomp 依存关系, 宾语和宾语补足语之间具有名词型主语 (nominal subject, nsubj) 依存关系, 因此可以依据 nsubj (off-balance-12, transmission-11) 和 xcomp (make-8, off-balance-12) 提取短语 make torque transmission off-balance。当 make 后面的宾语为动名词时, Stanford Parser<sup>[17]</sup> 认为动名词与形容词具有形容词补语 (adjectival complement, acomp) 依存关系, 因此可以从依存关系 acomp (loosening-11, difficult-12) 来提取 loosening difficult。再结合 make 即可得 make loosening difficult。如果 acomp 中的动词及其前面的某个动词之间存在 conj 依存关系, 如 conj\_and (tightening-9, loosening-11), 则可以把前面的动词和连词 and/or 集成到提取的短语中, 即 tightening and loosening difficult。

### 3.5 物理参数提取

物理参数对于理解设计任务、寻找潜在的解决方案具有重要的作用, 广泛应用于各种设计任务中。因为发明问题解决理论 (Theory of Invention Problem Solving, TRIZ)<sup>[23]</sup> 的效应知识库中包含了大量由各种物理参数组成的物理定律, 所以从专利中提取物理参数有助于利用效应知识库来扩展查询输入。本文基于效应知识库, 利用 Protégé<sup>[24]</sup> 预定义了一个物理参数词典 (thesaurus), 如图 4 所示。对于提取到的概念, 如果有单词出现在该词典中, 则认为该概念为物理参数。例如, 概念 Megasonic vibrations 中, 单词 Megasonic 出现在物理参数词典中, 因此, Megasonic vibrations 被归类为物理参数。

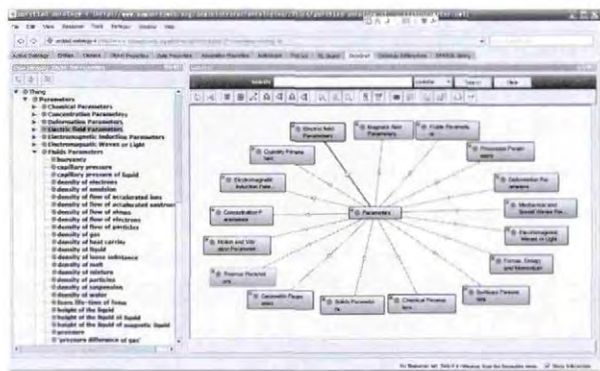


图4 物理参数词典

## 4 概念选择

### 4.1 停用词过滤

停用词是指那些不能反映主题的人类语言中包含的功能词, 如 the, a, each 之类的助词, 以及 however, thus 等只能反映句子语法结构的词语, 有必要将其滤除。本文利用 USPTO Stopwords<sup>[25]</sup> 来过滤候选短语中除 and, or 和 of 外的停用词, 以提高词频统计的精确性, 如 US7243922 摘要中, 把 The adjustable depth stop 和 an adjustable depth stop 经过停用词过滤后统一为 adjustable depth stop。

此外, 专利中经常会出现一些不能反映文献主题的短语, 如 present invention, above-given disadvantage, overcomes above-given disadvantages 等。本文依据专利分析结果扩展了 USPTO Stopwords<sup>[25]</sup> 停用词表, 使其能够过滤这些无意义的短语。

### 4.2 概念合并

概念之间并不都是彼此独立的, 有些概念之间通过某些表示具有 (如 having, with, without, capable of)、目的 (如 for)、手段 (如 via, by) 等含义的词联系在一起。如从专利 US5842527 的标题中提取到的概念 hammer drill 和 mode change-over mechanism 之间通过 with 联系在一起, 表达了专利所涉及的产品及其特色。又如从专利 US5842527 的发明领域可以提取到的概念 hammer blow mechanism 和 transmitting hammer blows 可以通过 for 合并为 hammer blow mechanism for transmitting hammer blows, 合并后的概念能够更详细地表达设计人员的信息需求。

此外, 如果两个概念之间具有部分与整体的关系 (如 chuking reinforcement force 和 absorb chuking reinforcement force)、特殊与一般的关系 (如 mechanical counter 和 mechanical signature counter), 则需要将两者合并, 保留长度较长的短语。由相同的动词和名词组成的两个概念 (即动词短语与名词短语, 如 monitor article 和 monitoring article) 仅保留动词短语。

### 4.3 概念选择

在查询推荐中, 为了获得多样性的查询结果, 推荐给设计人员的概念不能太少。同时, 如果把过多的概念推荐给设计人员, 则可能导致一些不重要的概念被包含进来。因此, 需要对提取的概念进行初

步选择,把最重要的概念推荐给设计人员。词频和位置在反映候选词和文献主题的关系上起着重要作用<sup>[26]</sup>。因此,本文利用词频来提取那些重要且经常出现(significant-frequent)的关键概念,基于位置构建规则提取那些重要但出现频率低(significant-rare)<sup>[16]</sup>的关键概念。

**规则 1** 选择所有出现频率大于 1 的概念。经过内容选择后,后续处理的专利内容通常都不会太多,概念出现的频率相对较低。此外,设计人员可以对推荐给他的概念进行选择,因此本文选择所有频率大于 1 的概念推荐给设计人员。

**规则 2** 对于只出现一次的概念,如果该概念出现在标题,则选择该概念;如果该概念是设计目标概念、物理参数概念,以及出现在摘要和独立主权项部分的功能概念,则选择该概念。

## 5 基于形态变换的查询扩展

由于现在分词和过去分词的存在,使英语中动词的表达形式具有多样性,同一动词的不同表达形式,既可以用作动词又可以用作形容词。如动词 detect 在动词短语 detect signal 中用作动词,在名词短语 detected signal 和 detecting signal 中用作形容词。而不论 detected signal 还是 detecting signal 都隐式体现了 detect signal 这一功能,因此在利用 detect signal 作为查询关键词进行信息检索时,应该把包含 detected signal 和 detecting signal 的信息返回给用户,而 USPTO 网站提供的专利查询功能并没有实现这个目标。因此,对于用户选定的功能概念中的动词,本文对其进行形态变换,引入其现在分词、过去式和过去分词形式,以扩展查询输入。

此外,由动词派生出来的名词(deverbal noun)也可以隐式表达相应动词所具有的功能含义,如 signal detection,一个包含该短语的专利很显然具有 detect signal 的功能。利用 WordNet<sup>[27]</sup>可以查询动词所派生出来的名词,如图 5 所示。因此,本文利用 C# 语言,基于 WordNet.Net<sup>[27]</sup>实现从 WordNet 3.0<sup>[27]</sup>中根据给定动词检索其派生的名词,然后通过替换和颠倒次序来获取动词短语所对应的名词短语。

综上所述,初始查询关键词 detect signal 经过语法变换后得到新的关键词集合 {detect signal, detects signal, detected signal, detecting signal, signal detection, signal detector},辅助用户获取更多和功

能 detect signal 相关的检索结果,如图 6 所示。

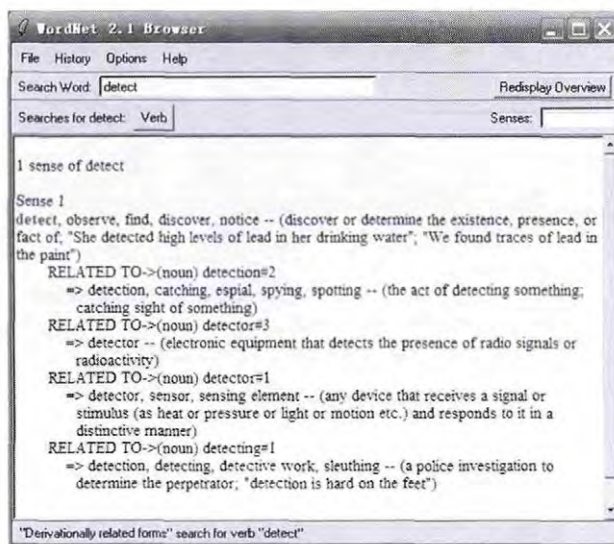


图5 利用WordNet检索动词派生的名词



图6 形态变换结果

## 6 应用实例

### 6.1 概念提取

本文以美国专利 US6902172 和 US5031156 为例,展示了概念提取算法的有效性。

**实例 1** US6902172 “Drill chuck with cushioned tightening sleeve”。

针对美国专利 US6902172,本文提取到的设计概念如图 7 所示。其中:axially elastically deformable spring 和 plurality of forwardly open angled guide 表明专利中用到的主要零部件;keyless hammer-drill chuck 和 drill chuck with tighten sleeve 表明该专利所涉及的发明领域;vibrate excessively, easy to assemble and use, make tightening and loosening difficult, easy to tighten and

loosen 和 difficult to control maximum torque 表明该专利所解决的问题。

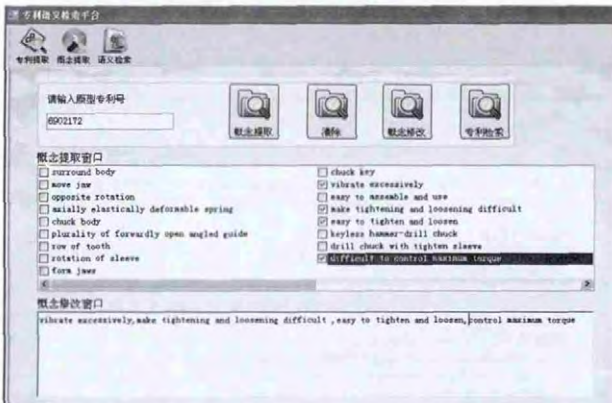


图7 US6902172设计概念提取结果

**实例 2** US5031156 Method and apparatus for detecting and counting articles.

针对美国专利 US5031156, 本文提取到的设计概念如图 8 所示。其中: detect and count article, article detection, detect and count signature 和 monitoring and counting article 表明该专利的设计目标; detect acoustic signal, direction of air stream, utilize acoustic 和 direct stream of pressurized air 从不同侧面表明该专利发明目标实现的手段; require reliable detection, counting of relatively fast moving, accumulation of dust 和 reduce effect of noise 从不同侧面表明了该专利的性能目标或拟解决的问题。

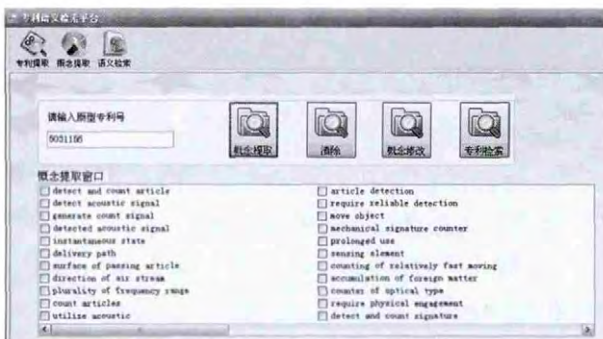


图8 US5031156设计概念提取结果

### 6.2 概念推荐

传统的 USPTO 提供的查询方法中, 设计人员提交关键词到专利检索系统中, 专利检索系统返回与该关键词相关的结果专利。如果设计人员的查询需求不是很明确, 则需要通过浏览查询结果产生新的搜索需求, 进行新一轮专利查询。经过数次反复

迭代后, 设计人员才能完全明确自己的查询需求, 获取自己所需要的专利。

本文提出的基于概念推荐的专利查询方法中, 从设计人员给定的原型专利中提取多样性的设计概念, 再把提取到的设计概念推荐给设计人员, 设计人员从中选择符合自己查询意图的概念, 帮助设计人员明确自己的查询需求。设计人员可以对选定的概念进行修改, 也可以添加由已有概念刺激产生的新概念(如图 7)。选中的概念经过基于形态变换的语义扩展后, 依次提交到 USPTO 中进行专利查询, 查询结果显示在基于 WebBrowser 控件开发的结果显示程序中。以查询专利 US6902172 为例, 把如图 7 所示提取到的概念推荐给设计人员, 设计人员选择 vibrate excessively, make tightening and loosening difficult, easy to tighten and loosen, difficult to control maximum torque 作为检索异类专利的查询关键词, 作适当修改后提交到 USPTO 中, 查询结果如图 9 所示。图 9 中, 左侧树状列表是设计人员选择的概念, 右侧是该概念作为关键词, 提交到 USPTO 网站后进行专利查询后得到的结果。

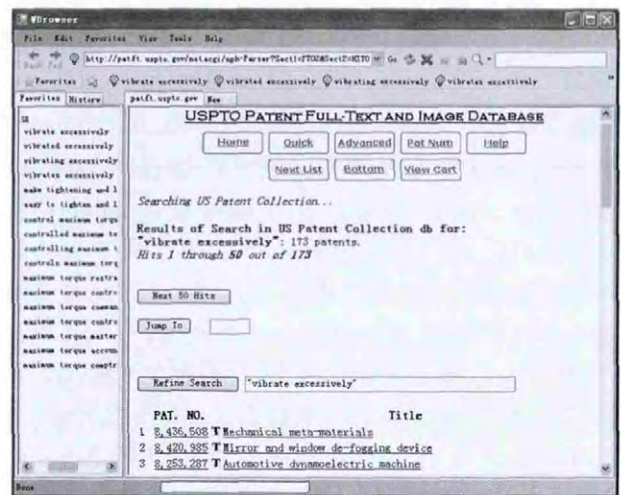


图9 专利检索结果

### 6.3 基于形态变换的查询扩展

动词的形态变换也有助于设计人员获取更多的相关结果, 表 1 给出了关键词 control maximum torque 及其形态变换在 USPTO 中的查询结果。与单纯的利用关键词 control maximum torque 进行检索相比, 形态变换后查询到的相关专利大幅增加, 有效解决了“不同的专利撰写者对同一问题可能采用不同表达形式这一现象对专利查询所造成的不良影响”。



表 1 control maximum torque 及其形态变换查询结果

查询关键词	数量	结果专利(前 5 个)
control maximum torque	56	8450955,8441220,8432121,8390223,8384327
controls maximum torque	9	8496229,7540334,7328752,7121358,7093668
controlled maximum torque	24	8450962,8447477,8180542,7918690,7869912
controlling maximum torque	19	8310177,8021293,7913594,7709990,7430945
maximum torque control	101	8450955,8441220,8427088,8423217,8390223
maximum torque controller	2	6836087, 5549185

## 7 结束语

为了辅助设计人员进行概念空间拓展,更好地理解信息搜索需求,提供给设计人员多样性的专利查询结果,本文提出一种基于概念推荐的专利查询方法。该方法以设计原型专利作为查询输入,利用自然语言处理技术对专利文本进行语义分析,从中提取与设计目标、实现手段相关的、多方面的设计概念,推荐给设计人员进行专利检索,辅助设计人员实现知识空间扩展。应用实例表明,本文工作有效解决了现有专利查询网站不能进行查询推荐这一不足。

今后将进一步研究概念提取方法,提高概念推荐的准确率与多样性,进而利用概念提取方法进行:

### (1) 专利族分析

通过提取专利族中与设计目标、实现手段相关的概念,以图或网络的方式显示给用户,以有助于用户快速了解现有产品的创新点、解决的问题和实现手段。

### (2) 类比发现

通过提取查询结果中专利文档中的概念并对其进行聚类,以辅助设计者发现解决相同问题、应用相同原理的类比专利。

## 参考文献:

- [1] FANTONI G, TAVIANI C, SANTORO R. Design by functional synonyms and antonyms : a structured creative technique based on functional analysis[J]. Proceedings of IMechE Part B: Journal of Engineering Manufacture, 2007, 221(4): 673-683.
- [2] HE Bin, FENG Pei'en, PAN Shuangxia. Functional space expansion process for product innovation[J]. Journal of Mechanical Engineering, 2009, 45(11): 189-196(in Chinese). [何斌, 冯培恩, 潘双夏. 面向产品创新的功能空间拓展进程[J]. 机械工程学报, 2009, 45(11): 189-196.]
- [3] ZHANG Hui, QIU Qingying, FENG Pei'en, et al. Method to trigger design inspiration from heterogeneous product patents [J]. Computer Integrated Manufacturing Systems, 2010, 16(3): 484-490(in Chinese). [张惠, 邱清盈, 冯培恩, 等. 异类产品专利激发设计灵感的方法[J]. 计算机集成制造系统, 2010, 16(3): 484-490.]
- [4] LINSEY J S, MARKMAN A B, WOOD K L. Design by analogy: a study of the word tree method for problem representation [J]. ASME Journal of Mechanical Design, 2012, 134(4): 041009.
- [5] CAO Guozhong, TAN Runhua, SUN Jianguang. Process and realization of functional design based on extended-effect model [J]. Journal of Mechanical Engineering, 2009, 45(7): 157-167(in Chinese). [曹国忠, 檀润华, 孙建广. 基于扩展效应模型的功能设计过程及实现[J]. 机械工程学报, 2009, 45(7): 157-167.]
- [6] HIRTZ J, STONE R B, MCADAMS D, et al. A functional basis for engineering design: reconciling and evolving previous efforts[J]. Research in Engineering Design, 2002, 13(2): 65-82.
- [7] MILLER A. WordNet: a lexical database for english[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [8] CHEONG H, STONE R B, MCADAMS D A, et al. Biologically meaningful keywords for functional terms of the functional basis[J]. ASME Journal of Mechanical Design, 2011, 133(2): 021007.
- [9] ORIAKHI E V, LINSEY J S, PENG X. Design-by-analogy using the word tree method and an automated wordtree generating tool[C]//Proceedings of the 18th International Conference on Engineering Design. Scotland, UK: The Design Society, 2011: 394-403.
- [10] CHEN Jixi, GU Xinjian, CHEN Guohai, et al. Ontology based patent retrieval technologies [J]. Journal of Zhejiang University: Engineering Science, 2009, 43(12): 2213-217, 2224(in Chinese). [陈苡熙, 顾新建, 陈国海, 等. 基于本体的专利检索技术[J]. 浙江大学学报: 工学版, 2009, 43(12): 2213-217, 2224.]
- [11] YANG Bo, GAO Changqing, YIN Xiaoling, et al. Research of extended product function modeling for conceptual design[J]. Journal of Mechanical Engineering, 2013, 49(15): 153-162(in Chinese). [杨波, 高常青, 尹晓玲, 等. 面向概念设计的

- 扩展功能建模方法[J]. 机械工程学报, 2013, 49(15): 153-162.]
- [12] AHMED S, WALLACE K M. Identifying and supporting the knowledge needs of novice designers within the aerospace industry[J]. Journal of Engineering Design, 2004, 15(5): 475-492.
- [13] STROHMAIER M, KRÖLL M, KÖRNER C. Intentional query suggestion: making user goals more explicit during search[C]//Proceedings 2009 Workshop on Web Search Click Data. New York, N. Y., USA: ACM Press, 2009: 68-74.
- [14] LI Yanan, WANG Bin, LI Jintao. A survey of query suggestion in search engine[J]. Journal of Chinese Information Processing, 2010, 24(6): 75-84(in Chinese). [李亚楠, 王斌, 李锦涛. 搜索引擎查询推荐技术综述[J]. 中文信息学报, 2010, 24(6): 75-84.]
- [15] GAO Y, ZEID I, BARDEZ T. Characteristics of an effective design plan to support re-use in case-based mechanical design[J]. Knowledge Based Systems, 1998, 10(6): 337-350.
- [16] LI Y R, WANG L H, HONG C F. Extracting the significant-rare keywords for patent analysis[J]. Expert Systems with Applications, 2009, 36(3): 5200-5204.
- [17] KLEIN D, MANNING C. Accurate unlexicalized parsing[C]//Proceedings of the 41st Meeting of the Association for Computational Linguistics. Stroudsburg, Pa., USA: The Association for Computational Linguistics, 2003: 423-430.
- [18] MAMEFFE M, MACCARTNEY B, MANNING C. Generating typed dependency parses from phrase structure parses[C]//Proceedings of the 5th International Conference on Language Resources and Evaluation. Paris, France: European Language Resources Association, 2006: 449-454.
- [19] SUN Qiyang, LI Yan, LI Wenqiang, et al. Strategy and application of product innovation based on function incentive[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(3): 197-202(in Chinese). [孙其英, 李彦, 李文强, 等. 基于功能激励的产品创新策略[J]. 农业机械学报, 2011, 42(3): 197-202.]
- [20] PARK H, YOON J, KIM K. Identifying patent infringement using SAO based semantic technological similarities[J]. Scientometrics, 2011, 90(2): 515-529.
- [21] KIM H, KIM K. Causality-based function network for identifying technological analogy[J]. Expert Systems with Applications, 2012, 39(12): 10607-10619.
- [22] DEWULF S. Directed variation: variation of properties for new or improved function product DNA, a base for connect and develop[C]//Proceedings of 2006 TRIZ Future Conference 2006. Procedia Engineering. Amsterdam, the Netherlands: Elsevier, 2011, 9: 646-652.
- [23] TAN Runhua. TRIZ[M]. Beijing: Science Press, 2004 (in Chinese). [檀润华. 发明问题解决理论[M]. 北京: 科学出版社, 2004.]
- [24] Stanford Center for Biomedical Informatics Research. Protégé[EB/OL]. [2013-10-07]. <http://protege.stanford.edu/>.
- [25] United States Patent and Trademark Office. StopWords[EB/OL]. [2013-10-07]. <http://www.uspto.gov/patft/help/stopword.htm>.
- [26] ZHENG Jiaheng, LU Jiaoli. Study of an improved keywords distillation method[J]. Computer Engineering, 2005, 31(18): 194-196 (in Chinese). [郑家恒, 卢娇丽. 关键词抽取方法的研究[J]. 计算机工程, 2005, 31(18): 194-196.]
- [27] SIMPSON T, CROWE M. Word Net. Net[EB/OL]. [2013-10-07]. <http://opensource.ebswift.com/WordNet.Net>.

### 作者简介:

吴鸿韬(1977—),男,辽宁昌图人,博士研究生,讲师,研究方向:计算机辅助创新、信息检索, E-mail: hongtaowu@hebut.edu.cn;

朱怀忠(1978—),男,河北肃宁人,硕士研究生,讲师,研究方向:计算机辅助创新、软件工程;

马建红(1965—),女,河北保定人,教授,博士,研究方向:软件工程、软件设计、计算机辅助创新;

檀润华(1958—),男,河北任丘人,教授,博士,博士生导师,研究方向:创新设计、面向大规模定制的产品设计、技术创新管理等,通信作者, E-mail: rhtan@hebut.edu.cn。