

# 基于疑问句句型识别的问题理解研究

刘朝涛<sup>1,2</sup> 李祖枢<sup>1,3</sup>

(重庆大学智能自动化研究所 重庆 400044)<sup>1</sup> (重庆交通大学机电与汽车工程学院 重庆 400074)<sup>2</sup>  
(重庆工学院人工智能研究所 重庆 400050)<sup>3</sup>

**摘要** 问题理解是问答系统中的一个关键步骤。问题理解的过程包括疑问词分类、问句分词及词性标注、疑问句句型识别、问题分类及问题中心的识别等。提出了对疑问句句型的识别方法,即通过对疑问句的短语句法树的分析,得到疑问句的句型结构,通过计算句型结构与句型模式之间的相似度来识别句子的句型。结合疑问句的语法分析,通过识别疑问句的句型实现了问题理解。实验结果表明,该方法提高了问题理解的准确度。

**关键词** 问答系统,问题理解,句型识别,相似度,句法树

## Research on Question Analysis Based on Sentence Type Recognition of Interrogative Sentences

LIU Chao-tao<sup>1,2</sup> LI Zu-shu<sup>1,3</sup>

(Institute of Intelligent Automation, Chongqing University, Chongqing 400044, China)<sup>1</sup>  
(School of Mechatronics and Automotive Engineering, Chongqing Jiaotong University, Chongqing 400074, China)<sup>2</sup>  
(Institute of Artificial Intelligence System, Chongqing Institute of Technology, Chongqing 400050, China)<sup>3</sup>

**Abstract** Question analysis is the primary task of question answering system. The process of the question analysis in this paper was described as follow: identifies the interrogative phrase class, parses words, identifies the question's sentence type, identifies the question class. A method of Sentence Type Recognition (STR) was put forward as: the sentence type structure of a question results from analyzing the phrase syntactic tree for the sentence, with computing the similarity between the sentence type structure and the sentence type modes, the question's sentence type recognition is achieved. The result of experiments shows that the question analysis implemented by this way increases the precision of question classification.

**Keywords** Question answering system, Question analysis, Sentence type recognition, Similarity

## 1 引言

问答系统(Question Answer System, QA)能更好地满足用户的检索需求,能更快地找出用户所需的答案,目前受到广泛关注<sup>[1-3]</sup>。问答系统一般包括3个主要组成部分:问题理解、信息检索和答案抽取<sup>[1-4]</sup>。问题理解,又叫问句分析,是问答系统中的一个关键步骤,它的准确度对问题的问答有重要影响。问题理解主要分析问题的问句信息,包括:问句类型、问句中心、问句焦点、限制条件以及回答的详细程度,答案是否单一等<sup>[4]</sup>。

目前,问题理解一般是以疑问词短语作为依据对问题进行分类,如询问人物、地点、时间、数量等<sup>[5-7]</sup>,然后再进行关键词提取和扩展。以疑问词短语进行问题分类的方法虽然简便,但对问题的理解比较粗浅,降低了理解的准确度<sup>[2]</sup>。因此目前已经有不少的研究结合句法分析来对问句进行分类。文勋等在文献[8,9]中在汉语依存语法的基础上从句法结构中提取问题分类,张亮等在文献[10]中采用句法树截断形成句法片段的方法对基于短语句法的句法树进行了分析,这些研究都取得了较好的成果。

本文根据上下文无关的短语文法的特点,通过对短语句法树的分析,得到疑问句的关于句型模式的句型结构,再计算

句型结构与句型模式之间的相似度,最后根据相似度来实现问句的句型识别,并进而实现对疑问句的问题理解。测试结果表明,这种方法对疑问句的理解是有效的。

## 2 问句理解的方法和步骤

在自动问答系统中,一般只研究特指问。特指问的显著特征是有疑问词,疑问词也与所问问题的类型以及问题的答案密切相关,也正因为如此,目前多数问题理解的方法就是根据疑问词短语作为依据采用模板匹配的方法来对问题进行分类。但单纯依靠疑问词短语来对问题进行分类是不完善的:(1)对于汉语,同一个问题可以有几种不同的提问方法,疑问词出现的位置很随意。(2)汉语中有些疑问词短语并不能明确代表疑问句的类型,有些有相同疑问词短语的句子,问题类型却并不相同,如:“……的原因是什么?”这个句子从疑问词来划分属于“什么”,即询问事物,但事实上该句涉及的内容却是事物的原因。(3)由于汉语本身的多义,有些句子中的疑问词却是做其他用途,这种句子根本就没有询问任何事物,如:“他不是什么都知道”。

在汉语特指疑问句中,疑问句的结构特点包括由疑问词短语和一些特殊的词语,以及词语在句中出现的词序等成分特征。疑问句的特征词包括疑问句中的疑问词短语和与疑问

刘朝涛 讲师,博士研究生,主要研究方向为智能自动化、文本数据挖掘等;李祖枢 教授,博士生导师,主要研究方向为人工智能、智能自动化、智能机器人、仿人智能控制理论及应用、文本数据挖掘等。

句类型有关的特殊词。人在理解问题时,就是先识别问句中的特征词,然后根据问题中的特征词、句型以及问题可能的回答模式来对问题分类,并进而识别出问题的主题。因此,要准确理解疑问句,应该从疑问句中的特征词、疑问句句型入手,对句子在句法甚至语义的层面进行分析。通过对疑问句的句型识别,就可以得出问题的分类及识别问题的中心等。

因此,本文的问句分析包括疑问词分类、问句分词及问性标注、疑问句句型识别及问题分类等三个步骤。

### 2.1 疑问词分类

特指疑问句都有疑问词短语存在,一组意义相同、用法相同的疑问词短语构成一种疑问词类型。表1列出了部分疑问词的分类情况。

表1 疑问词分类

疑问词短语	疑问词类
谁/什么人/何人/哪个(些)人	什么人
什么原因/什么理由/什么因缘	什么原因
什么时间/什么时候/哪个时候/何时	什么时间
什么地方/什么地点/哪里/哪儿/何处/为什么/为何	什么地方
.....	.....

### 2.2 问题分类

疑问词模式与问题类型存在对应关系。一个疑问词模式可能对应一个或多个问题类型,而一个问题类型又可能对应多个疑问词模式。问题分类需要识别出一个问句所属的问题类型。

表2 问题分类

疑问词类	句型模式	问题类型
什么人	(vp +)np + 是 + ~	who
	np + 是 + ~ (+vp)	
	~ + vp + np	
	~ + (说) + s	
为什么	s + 是(,) + ~	why
	np + ~ + vp	
什么原因	s + 是 + ~	why
	~ + s	
什么	s + 的 + 原因/理由/原由 + 是 + ~	
	.....	

## 3 句型识别

在问题理解中,重要的步骤是问句句型的识别。设待识别的疑问句为  $S$ , 属于  $S$  的疑问词类的句型模式集合为  $M = \{M_1, \dots, M_n\}$ , 则这个过程就是从  $M$  中找出与  $S$  最适合的句型模式  $M_s$ 。实现句型识别的方法可以有两种:(1)根据句型模式生成句子  $SM$ , 再同  $S$  进行比较。(2)对  $S$  进行句法分析, 然后再同句型模式进行比较。本文采用方法 2, 因此句型的识别过程包括 3 个步骤: 对  $S$  进行句法分析, 得到  $S$  的句法树  $T$ ; 以句型模式  $M_i$  为模板, 通过对  $T$  的计算得到关于  $S$  的基于句型模板  $M_i$  的句型结构  $SM_i$ ; 计算  $M_i$  与  $SM_i$  之间的相似度。最后选择最大相似度所对应的句型模式作为  $S$  的句型, 若最大相似度小于一个设定的阈值, 则匹配失败, 句子  $S$  不是问句。

### 3.1 句型结构分析

句法分析是应用句法规则和其它知识, 将句子中的单词, 表示成合乎语法的句法结构, 最常见的句法结构是短语句法结构树<sup>[12]</sup>。短语句法结构树表示了一个句子或短语的生成

过程, 在句法树中, 结点与子结点之间表示着一种生成关系, 即任意一个非叶子结点是由其子结点推导生成。

设疑问句为  $S$ ,  $S$  的句法树为  $T$ ,  $S$  的疑问词类所对应的疑问句句型模式集合  $\{M_1, \dots, M_i, \dots, M_n\}$ 。则句法树  $T$  基于句型结构模式  $M_i$  的句型结构分析的过程如下:

step 1 对树  $T$  中为特征词的叶子节点, 从该节点向上逐个删去该节点的度为 1 的祖先节点, 直到第一个度不为 1 的祖先节点, 并把该叶子节点直接与这个度不为 1 的祖先节点相连。

step 2 对  $T$  中的其他节点, 若其子孙结点中没有特征词, 则删除所有的子树。

step 3 删除树  $T$  中所有度为 1 且没有兄弟节点的非叶子节点。

step 4 将最后得到的树的叶子节点按从左到右写成字符串形式, 这就是句子  $S$  的以句型模式  $M_i$  为模板的句型结构  $SM_i$ 。

图 1 示出了“他受大家喜爱的原因是什么”在句型模式“dj + 的 + 原因 + 是 + 什么”下的句型结构的分析过程。其中图 a 为句子的句法结构树  $T$ , 图 b 为  $T$  经过 step 1 之后的结果, 图 c 为 step 2 的结果, 本例中无须 step 3。把图 3 中树的叶子节点写成字符串即“DJ + 的 + 原因 + 是 + 什么”, 这就是“他受大家喜爱的原因是什么”在句型模式“dj + 的 + 原因 + 是 + 什么”下的句型结构。

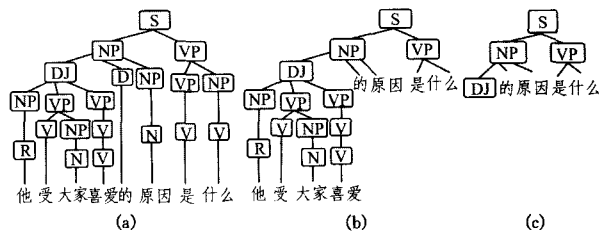


图1 句型结构生成示例

### 3.2 相似度计算

得到句型结构  $SM_i$  后, 需要识别该结构是否与对应的句型模式  $M_i$  相匹配。最简便的办法当然是逐词进行模式匹配, 但由于自然语言的变化多端, 而且汉语本身是弱语法语言, 这种模式匹配的方法误差很大, 且严重受句子成分的变化变化的影响, 实际上根本不具备实用性。但句型结构  $SM_i$  与句型模式  $M_i$  都是由短语标记和特征词组成的字符串, 可以认为是两个句子, 因此可以采用向量空间模型(SVM), 把这两个“句子”表示成向量的形式, 然后通过计算二者之间的相似度来实现句型结构与句型模式  $M_i$  的匹配识别。

在信息检索中, 常用相似度来表示待检索文档与检索需求之间的相关性程度。句子相似度则在自然语言处理的各个领域如信息检索、信息过滤等都有着非常重要的作用<sup>[8,12]</sup>。而在问句句型识别中, 问句  $S$  以句型模式  $M_i$  为模板生成了句型结构  $SM_i$ ,  $M_i, SM_i$  都是由短语和表示词性的标记所组成的符号集合。

为了判断  $M_i$  与  $SM_i$  之间的相关程度, 作者采用了句型相似度来分析和判断。计算句子与句型的相似度时, 只以句型为参考, 计算句子与句型之间的相似性, 而不考虑句型与句子之间的相似特征。

句型模式和基于句型模式的句型结构都是由短语标记和特征词组成的字符串, 如“s + 的 + 原因 + 是 + 什么”。将句型

模式  $M_i$  中的字符串项按照在  $M_i$  中的出现次序组成向量  $Term = \{t_1, \dots, t_j, \dots, t_n\}$ 。在此设  $M_i$  中的项数为  $n$ ,  $SM_i$  中的项数为  $m$ 。则  $M_i$  可以表示成以  $M_i = W_{M_i} * Term^T$ , 其中权向量  $W_{M_i} = \{w_1, \dots, w_j, \dots, w_n\}$  中的权重  $w_j$  表示项  $t_j$  对句型模式识别的重要程度, 由人工在描述句型模式时给出, 与识别该句型有较大影响的成分的权值较大; 而  $SM_i$  可以表示成以  $Term$  为基的向量  $SM_i = W_{SM_i} * Term^T$ ,  $W_{SM_i} = \{ws_1, \dots, ws_j, \dots, ws_n\}$ , 其中  $ws_j$  表示项  $t_j$  在  $SM_i$  中是否出现, 用 0, 1 表示。句型模式  $M_i$  与句子  $S$  的相似度由  $M_i$  与句型结构  $SM_i$  的相似度表示, 主要包括 3 个部分: 句型成分相似度、序列相似度、句型长度相似度。

(1) 句型成分相似度。句型成分相似度通过权向量  $W_{M_i}$ ,  $W_{SM_i}$  来计算  $M_i$  与  $SM_i$  之间的成分的相似度, 可以内积的方法来计算。为了突出特征词对句型识别的作用, 引入了系数  $\kappa = (1 - e^{-\alpha \sum_j w_j})$ :

$$Sim_{Term} = \kappa * W_{M_i} \cdot W_{SM_i} / |W_{M_i}| = \frac{\sum_j w_j * t_j}{\sum_j w_j} (1 - e^{-\alpha \sum_j w_j}) \quad (1)$$

(2) 序列相似度。与句型模式  $M_i$  相比照, 句型结构  $SM_i$  中字符串项为逆序的个数为  $Reorder$ , 则序列相似度

$$Sim_{order} = 1 - \frac{Reorder}{n} \quad (2)$$

(3) 语句长度相似度:

$$Sim_{Len} = 1 - \frac{|m-n|}{n} \quad (3)$$

(4) 句型相似度。总的句型相似度由句型成分相似度、序列相似度、句型长度相似度线性组合而成:

$$Sim = \lambda_1 * Sim_{Term} + \lambda_2 * Sim_{order} + \lambda_3 * Sim_{Len} \quad (4)$$

### 3.3 句型匹配的结果及评判

采用以上方法, 计算出句子与句型模式集中各个句型之间的相似度, 若最大相似度大于预先设定的阈值  $\sigma$ , 则选择具有最大相似度的句型为该句子的疑问句句型, 同时可以根据疑问句句型实现该问句的问题理解; 若最大相似度小于  $\sigma$ , 则句子  $S$  不是疑问句。

例如句子“他没来的原因是什么”, 该句的疑问词短语为“什么”, 因此只需要计算句子同“什么”一类问题句型的相似度, 通过对句子分词、句法分析后, 按照第 4 节中的步骤分析和计算, 得到如下的相似度, 从中可以得出, 该句子最接近的疑问句句型为“s+的原因+是+~”, 并进一步得出该句子的疑问句类型为“WHY”, 问题的问句中心为“他没来”。

表 3 句型相似度计算举例

句型模式 $M_i$	$SM_i$	相似度	排序
np+是+~	np+是+~	0.997781	2
ap+是+~	np+是+~	0.943438	3
vp+~+np	np+vp+~	0.570713	7
np+vp+~	np+vp+~	0.570713	7
~+np+vp	np+vp+~	0.670713	4
~+np+ap	np+vp+~	0.632040	6
s+是+~	np+是+~	0.943438	3
vp+~+vp	np+vp+~	0.570713	7
np+vp+~+np	np+vp+~	0.655403	5
s+的原因+是+~	s+的原因是+~	0.999988	1

## 4 实验结果与分析

在实验中, 作者利用 BAIDU 以各种疑问词短语搜索了

共 8000 个搜索结果, 通过对这些结果进行人工处理, 剔除掉其中一些重复的和广告性质的句子, 然后选择了其中 1000 个句子作为测试集, 并对测试问句采用人工方式标注出句子的疑问句句型、问题类型。通过本文中前述的方法进行分析, 得到如表 4 中所示的测试结果。其中测试 1 是在对这些测试问句进行全集测试的结果, 由于测试中所用的句法分析程序精度不高, 对测试结果影响较大, 因此在剔除句法分析不正确的问句后, 得到了测试 2 中的测试结果。在测试 2 中分析错误的问句中, 主要的原因仍然与句法分析有关, 例如把  $d_j$  分析成了  $vp+np$ ,  $np$  分析成  $ap$ ,  $np$  分析成  $tp$ 。测试 3 中的测试问句的句法分析都不正确, 可见, 本文中的句型识别方法对相当一部分句法分析错误的问句也能正确识别其问句类型。

表 4 问句句型识别实验结果

	句子数量	正确识别数量	正确率(%)
测试 1	1000	622	62.2000
测试 2	421	411	97.6247
测试 3	579	211	36.44214

对于问题理解的实验见表 5。由于实验所用的测试问题集来自真实的句子, 其中有疑问词的非疑问用法, 因此测试的精确度并不特别高, 但通过与基于疑问词为特征的问题分类方法进行了对比, 可以看出, 该方法比以疑问词为特征的问题分类方法有所提高。本文中的问题理解不是简单地分类, 而是要识别出问题的类型、句型, 并找出问题的问句中心等, 问题理解不正确的主要原因仍然是句法分析错误造成的, 但就是这样所得到的结果也是很让人满意的了。另外, 通过对分类错误的句子进行分析, 发现理解错误的问句中疑问词的非疑问用法<sup>[1]</sup>的数量为 79 个, 占总测试问句的 7.9%, 而在分类错误的测试问句中占 75.96%。

表 5 问题理解实验结果

	以词和词性标注为特征	以句型为特征
分类正确率	0.885	0.896
问题理解正确率	—	0.622

**结束语** 本文的问题理解方法结合了疑问句的语法分析, 问题理解的准确度得到了提高。同时, 在问题理解过程中, 本文提出了对问句句型的识别方法, 即通过对短语句法树的分析, 得到疑问句的关于句型模式的句型结构, 再计算疑问句的句型结构与句型模式之间的相似度, 最后根据相似度来识别句子的句型。但在本文的分析中, 还没有涉及语义的分析, 同时, 对疑问词的非疑问用法也没有做进一步的分析。因此, 进一步的工作包括: 结合疑问词的非疑问用法, 进一步提高问题理解的精确度; 结合问句的语义分析, 从更高更深层上来完善对疑问句的分析和理解。

**致谢** 本文中的词法分析采用了中科院计算所提供的开放资源的词法分析系统 ICTCLAS(Copyright: 2002-2005), 而句法分析则采用了中科院计算所自然语言处理课题组提供的概率句法分析器, 在此一并表示感谢。

## 参考文献

- [1] Voorhees E M. Overview of the TREC 2003 Question Answering Track[C]//Proc. The Twelfth Text Retrieval Conference. Gaithersburg, Maryland, 2003

Mstream DDoS可分解为5个阶段:1) 探测主机;2) 扫描端口;3) 入侵系统;4) 安装木马;5) 跳板攻击。其基于OWA-PFCM的攻击效能评估模型如图2所示。

基于OWA-PFCM的攻击效能评估模型为一个五元组: $\{A, V, E, P, \phi\}$ 。其中, $A$ 为攻击警报信息论域, $V$ 为脆弱性信息论域, $E$ 为攻击后果论域, $P$ 为条件概率权重, $\phi$ 为OWA算子集合。概念 $E_0$ 是最终的后果节点,表示复合攻击总的攻击后果。 $A_i, V_i$ 与 $E_i (i \geq 1)$ 分别表示对应于第 $i$ 个攻击阶段的入侵警报信息、脆弱性信息和攻击后果。 $A_i, V_i$ 均为原因节点,其状态值由各自的隶属函数定义。 $E_i$ 既是当前攻击阶段的后果节点,又是后续阶段攻击后果 $E_{i+1}$ 的原因节点。OWA-PFCM根据原因节点的当前状态值、重要性 $p_{ij}(t)$ 及与或组合关系,利用不同的OWA算子 $\phi_i$ 推理并融合结果。由于后继阶段的节点值 $V_i(t), A_i(t)$ 和 $E_i(t)$ 还会受到其前驱阶段攻击效果 $\{E_1(t), E_2(t), \dots, E_{i-1}(t)\}$ 的影响( $i > 1$ ),因此 $p_{ij}(t)$ 可随攻击进程动态变化。该OWA-PFCM攻击效能评估模型的推理过程如下:

1) 计算 $\{E_1(t), E_2(t), \dots, E_5(t)\}$ ,得到各攻击阶段的效能评估:

$$E_1(t+1) = \phi_1[A_1(t), V_1(t)];$$

$$E_i(t+1) = \phi_i[A_i(t), V_i(t), E_1(t), E_2(t), \dots, E_{i-1}(t)],$$

$$i > 1;$$

2) 计算 $E_0(t); E_0(t+1) = \phi_0[E_1(t), E_2(t), \dots, E_5(t)]$ ,得到总的攻击效能评估结果。

**结束语** 传统FCM能够有效表示节点状态的不确定性,但仅限于表示单调的或对称的因果关系。PFCM利用条件概率测度表示因果关系的不确定性,扩展了FCM的性能。但FCM与PFCM均不能表示节点间与或组合关系的不确定性。OWA算子是泛化的析取/合取运算,能够利用不同的权向量模拟任意的与或组合关系。基于有序加权平均算子的概率模糊认知图OWA-PFCM模型在继承PFCM优点的前提下,兼顾各原因节点的自主性、重要性及与或组合关系,能同时表示因果节点状态的不确定性、因果联系强度的不确定性、与或组合关系的不确定性,因此具有更强的认知与模拟能力。

## 参 考 文 献

[1] Kosko B. Fuzzy Engineering[M]. New Jersey: Prentice-Hall,

1997

- [2] Aguilar J. A Survey about Fuzzy Cognitive Maps Papers[J]. International Journal of Computational Cognition, 2005, 3(2): 27-33
- [3] Hagiwara M. Extended Fuzzy Cognitive Maps[C]// Proc. of the 1st IEEE International Conference on Fuzzy Systems. New York: IEEE, 1992: 795-801
- [4] 骆祥峰, 高隼. 概率模糊认知图[J]. 中国科技大学学报, 2003, 33(1): 26-33
- [5] Lv Zhenbang, Zhou Lihua. A Hybrid Fuzzy Cognitive Model Based on Weighted OWA Operators and Single-antecedent Rules[J]. International Journal of Intelligent Systems, 2007, 22(11): 1189-1196
- [6] Yager R R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1988, 18(1): 183-190
- [7] Calvo T, Mesiar R, Yager R R. Quantitative Weights and Aggregation[J]. IEEE Transactions on Fuzzy Systems, 2004, 12(1): 62-69
- [8] 骆祥峰, 高隼, 张旭东. 基于信任知识库的概率模糊认知图[J]. 计算机研究与发展, 2003, 40(7): 925-933
- [9] Fullér R, Majlender P. An Analytic Approach for Obtaining Maximal Entropy OWA Operator Weights[J]. Fuzzy Sets and Systems, 2001, 124(1): 53-57
- [10] Fullér R, Majlender P. On Obtaining Minimal Variability OWA Operator Weights[J]. Fuzzy Sets and Systems, 2003, 136(2): 203-215
- [11] Carver C A, Hill J M D, Pooch U W. Limiting Uncertainty in Intrusion Response[A]// Proceedings of the 2nd IEEE Workshop on Information Assurance and Security[C]. New York, USA: IEEE Computer Society, 2001: 142-147
- [12] MIT Lincoln Laboratory. 2000 DARPA Intrusion Detection Scenario Specific Data Sets[EB/OL]. [http://www.ll.mit.edu/IST/ideval/data/2000/2000\\_data\\_index.html](http://www.ll.mit.edu/IST/ideval/data/2000/2000_data_index.html)

(上接第153页)

- [2] 刘朝涛, 李祖枢. 一种汉语问题理解方法及其实现[C]// Proc. The 7th World Congress on Intelligent Control and Automation. Chongqing, China, 2008
- [3] 曹志娟, 李祖枢, 刘朝涛. 自动问答系统中的问题理解研究[J]. 计算机科学, 2005, 32(11): 158-161
- [4] 卢志坚, 张冬莱. 中文问答系统中的问句理解[J]. 计算机工程, 2004, 30(18)
- [5] Li X, Roth D. Learning Question Classifiers[C]// Proc. the 19th International Conference on Computational Linguistics (COLING). Taipei, Taiwan, 2002
- [6] Zhang D, Lee W S. A Language Modeling Approach to Passage Question Answering[C]// Proc. The Twelfth Text Retrieval Conference. Gaithersburg, Maryland, 2003
- [7] Ravichandran D, Hovy E. Learning Surface Text Patterns for a Question Answering System[C]// Proc. the 40th Annual Meeting of the Association for Computational Linguistics (ACL). Philadelphia, July 2002
- [8] 文勤, 张宇, 刘挺, 等. 基于句法结构分析的中文问题分类[J]. 中文信息学报, 2006, 120(12): 33-39
- [9] 李鑫, 杜永萍, 等. 基于句法信息和语义信息的问题分类[C]// 第一届全国信息检索与内容安全学术会议. 复旦大学, 2004: 243-251
- [10] 张亮, 王树梅, 黄河燕, 等. 面向中文问答系统的问句句法分析[J]. 山东大学学报: 理学版, 2006, 41(3): 30-33
- [11] 伍雅清. 汉语特殊疑问词的非疑问用法研究[J]. 语言教学与研究, 2002(2): 41-49
- [12] 孙建军, 成颖, 等. 信息检索技术[M]. 北京: 科学出版社, 2004