

# 基于词语形式标记的句际语义关系自动识别方法

钟茂生<sup>1,2</sup>

1. 华东交通大学信息工程学院, 南昌 330013; 2. 上海交通大学计算机科学与工程系, 上海 200240

E-mail: [zhongmaosheng@sjtu.edu.cn](mailto:zhongmaosheng@sjtu.edu.cn)

**摘 要:** 句际语义关系是指语篇上下文中相邻句子之间存在的语义关系。准确的识别句际语义关系对于文本理解、文本推理和文本结构分析具有重要的意义。然而, 由于受到语篇上下文环境、指代消解、句法结构等多个因素的影响, 使得句际语义关系自动分析研究具有很大的困难。该文在语言学界总结的句际语义关系和句际语义关系对应的词语形式标记的基础上, 提出了一种机器自动识别上下文句际语义关系的方法, 包括词语形式模板的获取、模板冲突消解的方法以及句际语义关系识别算法。实验结果表明, 该文提出的方法具有较好的识别效果。

**关键词:** 句际语义关系; 词语形式模板; 模板冲突消解; 句际语义关系识别

## Method on Automatic Recognizing the Inter-Sentence Semantic Relation Based on the Words Sign Template

Zhong Maosheng<sup>1,2</sup>

2. School of Information Engineering, East China Jiaotong University, NanChang, 330013; 2. Department of Compute Science and Engineering, Shanghai Jiaotong University, ShangHai 200240

E-mail: [zhongmaosheng@sjtu.edu.cn](mailto:zhongmaosheng@sjtu.edu.cn)

**Abstract:** The inter-sentence semantic relation is a semantic relation where exists between adjacent sentences in the context of the discourse. Accurate recognizing of inter-sentence semantic relation is of great significance to text understanding, text reasoning and text structure analysis. However, because of the impact of a number of factors, such as discourse context environment, anaphora resolution and syntactic structure, these make that the research on automatic recognizing the inter-sentence semantic relation is very difficult. On the base of the classification of inter-sentence semantic relation and the words sign template corresponding to a inter-sentence semantic relation summarized by the language researcher, this paper presented a method on automatic recognizing the inter-sentence semantic relation in the context of the discourse, included the way of obtaining the words sign template, the way of template conflict resolution and the algorithm of recognizing the inter-sentence semantic relation. The experiment results show that the proposed method has high performance for recognizing the inter-sentence semantic relation.

**Keywords:** Inter-sentence semantic relation, Words sign template, Template conflict resolution, Inter-sentence semantic relation recognizing.

## 1 引言

人类自然语言中的语义关系错综复杂, 语义关系存在于词汇、短语、句子、语篇等不

\*基金资助: 国家自然科学基金面上项目 (No. 60873135)

作者简介: 钟茂生 (1974-), 男, 江西兴国人, 讲师, 博士研究生; 研究方向: 自然语言处理、智能信息检索、信息抽取;

同的层次，层次越高，语义关系越抽象。句际语义关系是指存在于语篇上下文句子（简单句或复杂句）之间的语义关系，如因果关系、结果关系、条件关系、转折关系等，它属于句子层次上的语义关系。考察文本篇章中上下文句际之间是否具有某种语义关系，是判断上下文相邻句子是否能够衔接和连贯的一个依据，也是分析上下文在语义上是否有话题转换的趋势和存在话题转换的可能。因此，句际语义关系分析对于文本理解、文本推理和文本结构分析具有重要的意义。

现有面向机器的语义关系分析研究工作，大部分是词语或短语层次上的分析，如词语内部的概念语义分析[1]、复合词内部语义关系的识别[2,3]、文本中命名实体之间的语义关系识别[4~12]。而句子层次上的语义关系分析，由于受到语篇上下文环境、指代消解、句法结构等多个因素的影响，导致句际语义关系的识别任务更加困难，句际语义关系识别研究工作也相对较少。

本文根据语言学研究关于句际之间存在的各种语义关系分类，以及这些语义关系对应的连接词形式标记特征，来探讨机器自动识别语篇内部句际之间语义关系的可能性、方法和识别效果。据此，下面首先介绍语言学界关于句际语义关系分类体系；然后介绍本文提出的句际语义关系自动识别方法的基本思想和具体步骤；最后介绍应用本文方法进行句际语义关系识别的一些实验及结果分析。

2 句际语义关系分类

句子是在词汇的基础上按照一定的语法规则构成的，是信息交流的基本语言单位，也是表达语义相对完整的语言单位。一句话不论长短，如果有相对完整的意义，用一定的语调，末尾有较长的停顿，则它就是句子[13]。语篇中的句子之间总是采用某种手段使其上下文相互衔接，语义连贯。语篇上下文中相邻句子之间存在的语义关系称为句际语义关系。

语言学界关于句际语义关系有很多分类方法。韩礼德[14]早期将篇章中的句子语义关系划分为“加合”(additive)、“转折”(adversative)、“因果”(causal)和“时间”(temporal)四大类，到了 80 年代，韩礼德采用了“详述”(elaboration)、“延伸”(extension)和“增强”(enhancement)的新分类方法。国内廖秋忠“根据连接成分的语义和用法近似原则”推出了著名的“廖氏分类法”[15]。沈开木[16]将句段内语义联系分为联合关系和偏正关系两大类。吴为章等[17]列举了 12 种常见的句段结构分类。马楠[18]在上述研究成果基础上，进行梳理和补充，提出了对句际语义关系的时间、空间和逻辑三分法。本文第 3 节中关于句际语义关系自动识别主要基于马楠的三分法方法而进行研究，因此这里主要介绍这种分类方法，其分类结果如表 1 所示。从表 1 中可以看出，句际语义关系分为三个层次，最顶层为粗分类，即分为“时间联系”、“空间联系”、“逻辑联系”三个类，然后每个类再进行第二层分类，即亚类，有 11 个，而每个亚类还可以进行再分为第三层，即具体各种句际语义关系。

表 1 句际语义关系分类表  
Tab.1 Classification of inter-sentence semantic relation

类	亚类	语义关系名称
时间联系	时序关系:	包括“过程”、“先后”、“持续”、“重复”、“中断”、“接续”等语义关系
	时点关系:	包括“即时”、“特指”、“推定”等语义关系
空间联系	方位关系:	包括“描写”、“引发”等语义关系
	序列关系:	包括“承转”、“插入”等语义关系

逻辑联系	因果关系:	包括“原因”、“结果”等语义关系
	条件关系:	包括“正向”、“逆向”、“毋论”等语义关系
	转折关系:	包括“急转”、“婉曲”等语义关系
	联合关系:	包括“并列”、“选择”等语义关系
	比较关系:	包括“对比”、“类比”、“对照”等语义关系
	诠释关系:	包括“解说”、“总分”、“总结”、“转换”、“例证”、“校正”、“确认”、“强调”、“重申”等语义关系
	加合关系:	包括“递进”、“推论”、“转移”、“补充”等语义关系

3 基于词语形式标记模板的句际语义关系自动识别

3.1 基本思想

本文认为,语篇上下文句子之间具有的某种语义关系本质上是人的分析和认知的结果,这种分析是建立在两个句子所使用的一些特殊词语,本文称之为词语形式标记(或者称为词语形式特征、词语形式标记模板)基础上进行的。也就是说,语篇句子之间具有的某种语义关系通常会对应着两个句子中所使用的一些词语形式标记,通过识别和匹配这些词语形式特征,就可以识别句际之间的语义关系。例如:

“浪漫主义者大概比任何人都更加偏爱流浪。因为流浪最能实现浪漫情调,满足浪漫主义者的心理上的与美学上的需要。”

该例子中,前后两句之间的语义关系因为有连接词“因为”做形式标记,而且,语义侧重点(中心)在原因,因此两个句子之间的语义关系可以认为是“因果关系”。

据此,本文关于句际语义关系自动识别的基本思想就是,假定句子中词语形式特征的使用与句子间语义关系具有对应关系,通过分析和匹配词语形式特征,即可将文本中上下文句子之间的语义关系识别出来。

3.2 相关定义

定义 1 定义  $D=\{d<S_1,S_2>\}$  为待分析的文本语篇数据集,其中  $d<S_1,S_2>$  为数据集中的一个数据元素,  $S_1$  和  $S_2$  是语篇上下文中的两个句子;

定义 2 定义  $R=\{r_1,r_2,\cdots,r_n\}$  为句子之间语义关系集合。对于任意语义关系  $r_i\in R$ ,它可能在  $D$  中出现一次或若干次;

定义 3 假定  $P$  是各种文本语篇中所有出现的词语形式标记(词语形式标记模板)集合,定义函数  $f_{RC}:P\rightarrow R$  为在约束条件  $RC$  的约束下,从词语形式标记集合  $P$  到句际语义关系集合  $R$  的映射。因此,对于特定的语义关系  $r_i\in R$ ,它在各种文本语篇中出现时所对应的词语形式标记集合,记为  $P_{r_i}=\{p|p\in P\wedge r_i\in R\wedge(r_i=f_{RC}(p))\}$ 。

词语形式标记(词语形式标记模板)  $P$  可以用正则表达式来进行形式表示。例如表示“选择”语义关系的词语形式标记“要么...要么...”可以用下列正则表达式来表示:  
 $p(\text{“要么}\cdots\text{要么}\cdots\text{”})\equiv\text{“要么}\backslash\backslash S+\text{要么”}$ 。因此,下文中对“词语形式标记”和“词语形式标记模板”的描述将不加区分。

定义 4 假定  $p$  为词语形式标记模板,定义  $M_D(p)$  为模板  $p$  在  $D$  中所有能匹配到的元组集合,  $M_D(p)$  中元素个数记为  $|M_D(p)|$ ;

定义 5 假定  $p$  为词语形式标记模板, 其对应的语义关系为  $r_p = f_{RC}(p)$ , 定义  $M_D^{f_{RC}(p)}$  为  $D$

中所有语义关系为  $f_{RC}(p)$  的元组集合,  $M_D^{f_{RC}(p)}$  中元组个数记为  $|M_D^{f_{RC}(p)}|$ 。

定义 6 假定  $p$  为词语形式标记模板,  $p$  的覆盖能力定义为  $C_D(p) = \frac{|M_D(p) \cap M_D^{f_{RC}(p)}|}{|M_D^{f_{RC}(p)}|}$ ,

$p$  的错误率 (error rate) 定义为  $E_D(p) = \frac{|M_D(p) - M_D^{f_{RC}(p)}|}{|M_D(p)|}$ 。

模板  $p$  的覆盖能力  $C_D(p)$  反映了其在语义关系识别时的概括能力, 而其错误率  $E_D(p)$  反映了模板  $p$  在语义关系识别时的出错情况, 因此, 对于词语形式标记模板  $p$  的要求是尽可能有较强的覆盖能力和较低的出错率。

### 3.3 词语形式标记模板的获取

如何获取具有较强覆盖能力、较低出错率的词语形式标记模板, 对于句际语义关系识别的正确率有着重要影响。本文为了获取理想的词语形式标记模板, 采用如下的方法:

(1) 根据语言学专业人员在句际语义关系分类时归纳总结得到的各种词语形式标记, 经过形式转换后, 生成候选的形式标记模板。例如“对照”语义关系中的“与……相比”, 可以转换成正则表达式形式的模板为“与\\S+相比”作为候选模板。

(2) 人工收集一定规模的文本语篇, 手工标记好这些语篇内部的句子之间的语义关系, 作为试验测试语料, 用候选模板对标记好的语料进行匹配和句子语义关系识别, 通过分析识别结果, 用覆盖能力  $C_D(p)$  和错误率  $E_D(p)$  两个指标对每个候选模板  $p$  进行评价, 选择

具有较强覆盖能力和错误率低的模板作为最终模板, 存于模板库, 当覆盖能力  $C_D(p)$  和错

误率  $E_D(p)$  两个指标矛盾时, 以较低错误率指标作为模板的优选条件。

事实上, 上述方法是以手工总结、机器验证的方式来获取理想的词语形式标记模板, 因此模板的形式固定, 数量也是确定的。然而, 由于语言的开放性和表达形式的多样性, 试图以形式固定、数量确定的模板, 来覆盖各种语言现象, 显然是不合理的。据此, 很多研究人员研究如何通过标记语料自动扩展和生成新模板[19,20], 作者所在的课题组也提出了一种在 Bootstrapping 框架下, 用生物序列比对方法来抽取和生成特定语义关系的语义模板, 并定义了新的模板评价机制来估计模板的置信度[21]。本文关于这个工作没有进一步扩展, 后面的试验主要是使用语言学专业人员总结的各种词语形式标记, 作为候选, 并通过机器验证来最终获取可用的模板。

### 3.4 模板冲突消解方法

模板冲突是指对于一些语篇上下文, 可以使用模板库中一个以上的模板予以匹配, 最终分析得到语篇上下文句子之间有多组语义关系的现象。模板冲突发生后必须进行冲突消解, 由于语篇内句子之间是否有多种语义关系, 是个理解和认知问题, 因此, 模板冲突消解本身是个难题。

为了进行模板冲突消解, 本文在模板进行匹配语篇上下文之前, 首先进行“模板适应度分析”, 即判断一个模板在当前上下文  $C_X$  中的适应性。为此, 定义函数  $\partial$  来进行模板适应度分析。

$$\partial(p, C_X) = \arg \max_{c' \in \{c, \bar{c}\}} P(c' | C_X) \quad (1)$$

式(1)中,  $c$  表示当前模板  $p$  在上下文环境  $C_X$  中是适用的, 而  $\bar{c}$  则表示不适用。因此模板的适应度分析转变成为一个二元分类问题。本文用于分类的特征主要源于句际语义关系定义时所限定的对象, 如人物、时间、地点、事件等之间的关系, 即考察语篇前后句子在“人物 (Agent)”特征上是否有“同指”、“时间 (Time)”特征上是否有先后、“地点 (Place)”特征上是否有改变、“事件 (Event)”特征上是否相同, 来确定模板  $p$  的适应度。

由于模板  $p$  的适应度分析转变成了一个二元分类问题, 而且使用的分类特征也已经确定, 因此很容易使用现有的经典分类方法, 如 Bayesian 模型、决策树等方法进行分类。本文使用最基本的朴素贝叶斯分类模型进行分类, 即, 对以某个特定的模板  $p$ , 如果语篇上下文用特征表示, 则模板  $p$  的适应度分类公式如(2)式所示。

$$P(c' | C_X) = \frac{P((Agent, Time, Place, Event) | c') \cdot P(c')}{P(Agent, Time, Place, Event)} \quad (2)$$

由于是二元分类, 所以只需比较概率的大小(如1式所示), 而(2)式中的  $P(Agent, Time, Place, Event)$  与最终比较结果无关, 因此只需考虑(2)式中的分子部分。其中, 分子部分  $P(c')$  为模板  $p$  关于适应度的先验概率, 可从标记好的训练语料中估计。对于  $P((Agent, Time, Place, Event) | c')$ , 如果假定特征是独立的, 则可以用式(3)进行估计, 即

$$\begin{aligned} & P((Agent, Time, Place, Event) | c') \\ &= P(Agent | c') \cdot P(Time | c') \cdot P(Place | c') \cdot P(Event | c') \end{aligned} \quad (3)$$

式(3)中等式右边每个概率都可以从标记好的训练语料中进行估计。

根据模板的“适应度分析”结果, 就可以确定当前模板能否适应待分析的上下文, 以便尽可能消解模板使用的冲突。

### 3.5 句际语义关系识别算法

根据获取的模板和模板冲突消解策略, 即可设计句际语义关系识别算法, 来识别语篇上下文句子之间的语义关系。算法如图1所示。

---

Algorithm 1 (RSRIS, Recognize\_Semantic\_Relation\_of\_Inter-Sentences)

输入：待分析语篇的两个句子  $S_1$ 、 $S_2$ ，模板库  $DB_M$ ，模板与语义关系的映射  $f_{RC}$ ：

输出：句子的语义关系  $R(S_1, S_2)$ ；

Initialization:  $C_x = S_1 \& S_2$ ,  $DB_M$ ,  $f_{RC}$ ,  $R(S_1, S_2) = \emptyset$

For each  $p_i \in DB_M$  do

If  $\partial(p_i, C_x) = c \ \&\& \ Match(p_i, C_x) = 1$  (其中  $\partial(p_i, C_x)$  为模板  $p_i$  在当

前上下文  $C_x$  的适应度分析,  $Match(p_i, C_x)$  为模板  $p_i$  与  $C_x$  的匹配)

$r \leftarrow f_{RC}(p_i)$

$R(S_1, S_2) \leftarrow R(S_1, S_2) \cup \{r\}$

Endif

Endfor

return  $R(S_1, S_2)$

图 1 句际语义关系识别算法

Fig.1 Recognizing algorithm of inter-sentence semantic relation

## 4 实验及结果分析

为了验证本文提出的句际语义关系识别方法，本文设计了两个实验：一个实验的数据来自于语言学研究文献中标记好的语篇上下文例句，总共有 136 对；另外一个实验的数据是从网页文本中随机选择的上下文句子对，数量为 1150 对。实验使用准确率和召回率两个指标进行评价，其定义如 (4)、(5) 式所示。

$$\text{准确率 (P)} = \frac{\text{机器识别的语义关系中正确结果数}}{\text{机器识别的语义关系总数}} \times 100\% \quad (4)$$

$$\text{召回率 (C)} = \frac{\text{机器识别的语义关系中正确结果数}}{\text{人工标记的语义关系总数}} \times 100\% \quad (5)$$

由于句子间的语义关系可能有或没有，即语义关系数  $\geq 0$ ，因此将句子间语义关系看成是一个集合，当机器识别到一个句子对的语义关系集合  $R_{machine}$  与人工标记的句子对的语义关系集合  $R_{human}$  满足  $(R_{human} \subseteq R_{machine}) \vee (R_{machine} \subseteq R_{human})$  时，就认为机器标记的结果是正确的结果。下面是两个实验的结果汇总。

### 4.1 实验一

该实验使用的数据选自于语言学研究文献中标记好的语篇上下文例句，本文

假定这些标记的结果都是正确的结果，机器识别的结果只需和其进行对比即可。实验分“模板冲突消解”和“模板冲突不消解”两种情况，实验结果如表 2 所示。

表 2 句子语义关系自动识别结果

Table 2 Result of automatic recognizing the inter-sentence semantic relation

	准确率	召回率
模板冲突不消解	72.7%	64.7%
模板冲突消解	87.3%	75.7%

表 3 句子语义关系自动识别结果实例

Tab. 3 Some results of automatic recognizing the inter-sentence semantic relation

编号	语篇句子	$R_{human}$	$R_{machine}$	词形标记特征
1	老虎和鸡都是生命，老虎是吃鸡的；鸡和小虫子都是鸡吃小虫子的。都是为了生存。	原因	原因	“为了...”
2	像那群脸上生出痘痘的年轻人，去山上打鸟，或者干些什么。反正，需要找点什么事做。	毋论	毋论	“反正...”
3	估计，它要么是饿极了，打算冒险。要么是曾受到过人类的攻击，此刻决心报复。	选择	选择	“要么... 要么...”
4	于是先把东城放着，搞城南。城南相对容易，搞起来也不轻松。	对照	对照	“相对...”
5	同时，我们不应该忽略，云南现代化的进展带来了省内一元化支配的强化和均质化。以民族问题为例，它表现在对少数民族的改土归流和汉化的推进。	例证	例证	“以... 为例.....”
...	...	...	...	...

从表 2 中可以看出，对模板冲突采取本文中的消解策略，可以在很大程度上提高识别的精度，总体准确率达到 87%左右。表 3 是机器自动识别句子间语义关系的一些结果实例。

4.2 实验二

该实验的数据是从网页文本中随机选择的上下文句子对，实验过程中采用本文的模板冲突消解策略，对实验结果分三次随机选择部分结果，进行人工分析和判定，每次的句子对数和最后分析结果如表 4 所示。

表 4 句子语义关系自动识别结果

Tab. 4 Result of automatic recognizing the inter-sentence semantic relation

编号	句子对数量	准确率	召回率
1	56	75.0%	64.3%
2	52	78.6%	63.5%
3	60	72.5%	61.7%

从表 4 中可以看出，其准确率平均为 75.4%，比实验一中的结果低。可能的原因是，实验一中的实验数据是语言学专业人员的选定并标记好的例句，具有典型性和语言代表性，而实验二中的数据是网页文本中的句子对，这些句子是普通的上下文，因此句子中使用的词形标记特征不一定都很明显，所以识别的正确率要低。

总体来看，上面实验一和实验二的结果还是比较理想，它表明本文提出的基于词语形

式标记模板的句际语义关系自动识别方法是可行的,但也存在一些问题,主要是本文所用到的模板全部由语言学专业人员人工总结,这些模板的覆盖能力有限,因此后面的工作是如何从现有的语言事实中自动抽取到更多的可用模板,以提高句际语义关系识别的正确率。

## 5 结束语

考察文本篇章中上下文句际之间是否具有某种语义关系,是判断上下文相邻句子是否能够衔接和连贯的一个依据,也是分析上下文在语义上是否有话题转换的趋势和存在话题转换的可能,它对于文本理解、文本推理和文本结构分析具有重要的意义。

本文在语言学界总结的句际语义关系和句际语义关系对应的词语形式标记的基础上,提出了一种机器自动识别上下文句际语义关系的方法,包括词语形式模板的获取、模板冲突消解的方法以及句际语义关系识别算法,并用实验验证了该方法的有效性和识别效果。

未来的主要工作是进一步研究影响句际语义关系识别性能的命名实体识别、命名实体语义关系分析、指代消解等自然语言处理任务,此外,利用本文提出的方法,进行语篇层次上的语义分析、文本结构分析,也是重要的后续工作。

## 参 考 文 献

- [1] 陆汝占. 汉语内涵逻辑及其应用[J]. 中国中文信息学会二十周年学术会议论文集[C]. 北京: 清华大学出版社, 2001.11
- [2] Kim S. N. and Baldwin T. Interpreting Semantic Relations in Noun Compounds via Verb Semantics[C]. Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, 2006, pp:491-498
- [3] Costello F.J., Veale T., and Dunne S.. Using WordNet to Automatically Deduce Relations between words in Noun-Noun Compounds[C]. Proceedings of the COLING/ACL on Main Conference Poster Sessions, 2006, pp:160-167
- [4] 赵军. 命名实体识别、排歧和跨语言关联[J]. 中文信息学报, 2009, 23(2):3-17
- [5] ACE08 Evaluation Plan v1.2d. <http://www.itl.nist.gov/iad/mig/tests/ace/2008/doc/ace08-evalplan.v1.2d.pdf>
- [6] Hobbs J R, Bear J, Israel D, et al. SRI International FASTUS System: Muc-6 Test Results and Analysis[C]. In Proceedings of the 6<sup>th</sup> Message Understanding Conference(MUC-6). 1995. pp:237-248
- [7] Roman Y, Grishman R. NYU:Description of the Proteus/PET System as Used for MUC-7 ST[C]. In Proceedings of the 7<sup>th</sup> Message Understanding Conference(MUC-7). 1998
- [8] Aone C. and Ramos-Santacruz M.. REES:A Large-Scale Relation and Event Extraction System[C]. In Proceedings of the 6th Applied Natural Language Processing Conference, 2000. pp:76-83.
- [9] Suzuki J. Isozaki H. and Maeda E. Convolution Kernels with Feature Selection for Natural Language Processing Tasks[C]. In Proceedings of the 42<sup>nd</sup> Meeting of Association for Computational Linguistics. 2004, pp:119-126
- [10] 车万翔, 刘挺, 李生. 实体关系自动抽取[J]. 中文信息学报, 2005, 19(2):1~6.
- [11] Zelenko D, Aone C, Richardella A. Kernel Methods for Relation Extraction[J]. Journal of Machine Learning Research, 2003(2): 1083-1106
- [12] Zhao Shubin, Ralph Grishman. Extracting Relations with Integrated Information Using Kernel Methods[C]. In Proceedings of the 43<sup>rd</sup> Annual Meeting of Association for Computational Linguistics. 2005. pp: 419-426
- [13] 黄国文. 语篇分析概要[M]. 长沙: 湖南教育出版社出版, 1988.4
- [14] Halliday M.A.K., Hasan R., Cohesion in English[M]. London: Longman, 1976.
- [15] 廖秋忠, 篇章与语用和句法研究[M], 廖秋忠文集. 北京: 北京语言学院出版社, 1992.10



- [16] 沈开木,田树生. 句段分析(超句体的探索)[M]. 语文出版社. 1987.9
- [17] 吴为章,田小琳. 汉语句群[M]. 商务印书馆. 2002.4
- [18] 马楠. 现代汉语句段内的语义联系[D]. 黑龙江大学硕士学位论文. 2008年5月
- [19] Agichtein, E. and Gravano, S.. Snowball: Extracting relations from large plain-text collections[C]. In Proceedings of the 5<sup>th</sup> ACM International Conference on Digital Libraries. San Antonio, Texas, United States, 2000, pp:85-94
- [20] Brin, S. Extracting Patterns and Relations from the World Wide Web[C]. In Proceedings of the International Workshop on the Web and Databases. Valencia, Spain, 1998. 172-183.
- [21] Hu Yi, Lu Ruzhan, Chen Yuquan and Pei Bingzhen. Text Retrieval Oriented Auto-construction of Conceptual Relationship[C]. In Proceedings of ICCS 2007, Lecture Notes in Computer Science, Springer-Verlag, 2007, Vol.4488, pp:1214-1217