《现代图书情报技术》版权所有,欢迎下载引用!
请注明引用地址:实体关系抽取的技术方法综述*[J],现代图书情报技术,2008(8):18-23.

实体关系抽取的技术方法综述*

徐　健[1,2,3]　张智雄[1]　吴振新[1]
[1](中国科学院国家科学图书馆　北京 100190)
[2](中国科学院研究生院　北京 100049)
[3](中山大学资讯管理系　广州 510275)

【摘要】对实体关系抽取研究以 MUC 和 ACE 评测为主线的发展进行总结,并指出实体关系抽取任务普遍存在的三个问题是特定领域标引数据集的获取、模式的获取以及共指消解。在对当前关系抽取的相关文献、系统和项目进行分析研究的基础上,将基于非结构化文本的实体关系抽取技术方法归纳为:基于模式匹配的关系抽取、基于词典驱动的关系抽取、基于机器学习的关系抽取、基于 Ontology 的关系抽取以及混合抽取方法,旨在为进一步构建实体关系抽取系统提供良好借鉴。

【关键词】实体关系抽取　信息抽取　关系抽取方法
【分类号】G250.73

Review on Techniques of Entity Relation Extraction

Xu Jian[1,2,3]　Zhang Zhixiong[1]　Wu Zhenxin[1]
[1](National Science Library, Chinese Academy of Sciences, Beijing 100190, China)
[2](Graduate University of the Chinese Academy of Sciences, Beijing 100049, China)
[3](Department of Information Management, Sun Yat-Sen University, Guangzhou 510275, China)

【Abstract】Entity relation extraction is a very important task in text information extraction domain. It first summarizes the development of entity relation extraction related to MUC and ACE, and then points out that main difficulties exist in the process of relation extraction are acquisition of training dataset, acquisition of templates, and co-reference resolution. Based on the analysis of recent related literatures, systems and projects, it concludes the entity relation extraction methods as follows: templates method, lexicon driven method, machine learning method, Ontology driven method, and hybrid method. The analysis of these methods can help to build more efficient entity relation extraction system in further step.

【Keywords】Entity relation extraction　Information extraction　Relation extraction methods

1　引　言

随着数字资源和万维网上的文本信息的飞速增长,从文本中自动抽取知识的技术越来越被关注。具体而言,知识抽取系统从特定领域的文本文档中识别相关知识要素(通常是预先定义的类型),并将它们以结构化形式存储。知识抽取任务在细节和可靠性上有不同的选择,但一般都包括两个普遍存在并且紧密关联的子任务:实体识别和关系抽取。实体识别通过实体抽取技术抽取各个知识要素。抽取出的知识要素以离散的形式存在,只能反映出文本中包含哪些实体,例如人、机构、地点等,却不能反映出知识要素之间的关系,例如机构与人之间的雇用关系、机构与地点之间的位置关系等,而关系抽取则是要解决这一难题。

收稿日期:2008-06-16
*本文系国家社会科学基金项目"从数字信息资源中实现知识抽取的理论和方法研究"(项目编号:05BTQ006)的研究成果之一。

不同的研究者对关系抽取任务的表述不尽相同。Alexander Schutz 等人[1]认为关系抽取是自动识别由一对概念和联系这对概念的关系构成的相关三元组。Sophia Katrenko 等人[2]则从关系抽取的基本过程角度对关系抽取进行了界定。他们认为,关系抽取可以看作是具有两个步骤的过程,即:识别存在关系的证据和检查是否存在关系。维基百科[3]对关系抽取的解释是,关系抽取是在自然语言处理过程中抽取文本中实体间命名关系的任务。抽取的实体间关系能够通过各种形式/语言来表达。其中一种对网络上数据进行表达的语言是 RDF。能用到这些关系的应用领域包括基因 - 疾病关系,蛋白质相互作用等。而作为关系抽取权威评测会议的 ACE(Automatic Content Extraction)[4]将关系抽取任务表述为:探测和识别文档中特定类型的关系,并对这些抽取出的关系进行规范化表示。这些关系中,有些对实体出现顺序敏感,有些对实体出现顺序不敏感。ACE08[5]在以前的关系抽取评测任务基础上新增加了跨文档关系抽取任务。跨文档关系抽取扩展了文档内关系抽取任务,主要用来发现全局实体间的关系。

关系抽取技术在很多领域具有应用价值。在自动问答系统中,关系抽取自动关联相关问题和答案;在检索系统中,关系抽取使类似于“北京有哪些公司?”这样的语义检索功能的实现成为可能;在本体学习过程中,关系抽取能够发现新的实体间关系来丰富本体结构;在语义网标注任务中,关系抽取能够自动关联语义网知识单元。

2　关系抽取研究的发展

关系抽取研究的发展是以 MUC(Message Understanding Conference)评测会议和后来取代 MUC 的 ACE(Automatic Content Extraction)评测会议为主线进行的。在这两个会议上,多种先进的信息抽取方法被提出来,在会议提供的数据平台上进行测试,并组织与会者进行讨论。基本上每年一次的评测会议为关系抽取的发展起到了引导和推动作用。

关系抽取任务最初是由美国国防高级研究计划委员会(Defense Advanced Research Projects Agency,DARPA)资助的 MUC 会议于 1998 年 MUC - 7 上首次正式提出[6]。MUC 的显著特点并不是会议本身,而在于对信息抽取系统的评测。只有参加信息抽取系统评测的单位才被允许参加 MUC 会议。在每次 MUC 会议前,组织者首先向各参加者提供样例消息文本和有关抽取任务的说明,然后各参加者开发能够处理这种消息文本的信息抽取系统。在正式会议前,各参加者运行各自的系统处理给定的测试消息文本集合。由各个系统的输出结果与手工标注的标准结果相对照得到最终的评测结果。最后才是所谓的会议,由参与者交流思想和感受。这种评测驱动的会议模式被证明是行之有效的,对信息抽取的发展起到了推动作用。MUC 主要关注于包括自由文本分析、识别命名实体、识别特定类型的关系等一系列信息抽取任务。在 MUC - 7 之后,MUC 被由 NIST 引导的 ACE(Automatic Content Extraction)[7]评测所取代。

美国国家标准技术研究院(NIST)组织的 ACE 评测从 1999 年开始继续进行信息抽取方面的评测。ACE 评测 1999 年 7 月开始酝酿,2000 年 12 月正式开始启动,迄今已经举办过 8 次评测,最近正在进行第 9 次评测(2008 年 5 月)。ACE08[8]所提出的任务包括:单文档内实体探测和识别以及关系探测和识别;跨文档实体探测和识别以及关系探测和识别。文档语种有英语和阿拉伯语两种。在 MUC 和 ACE 的促进下,关系抽取技术取得了较大进步,研究热点已从最初的语言学单纯模型的应用发展到使用浅解析器或完全解析器的 NLP 技术的应用和复杂机器学习方法的应用,而关系抽取性能也有了大幅提升。

国内信息抽取方面的研究虽然起步较晚,但目前也已经在关系抽取方面做出了一些卓有成效的工作。邓擘等人[9]在使用模式匹配技术的基础上引入了词汇语义匹配技术对汉语实体关系进行抽取,并比较了一般模式匹配技术和词汇语义模式匹配技术在汉语实体关系提取任务中的性能。他们的实验结果表明,一般模式匹配技术在处理中文时效果较差,而词汇语义模式匹配技术更适合于处理汉语实体关系抽取任务。姜吉发等人[10]提出了一种自举的二元关系和二元关系模式获取方法 BRPAM,该方法能够根据用户初始给出的几个种子二元关系从一个大的自由文本集合中抽取出更多的二元关系。顾雪峰[11]针对文本特征粒度对实体关系识别结果影响较大这一问题,应用动态粒度思想,对识别特征进行逐步细化,构建了一个具有偏序

关系的特征族来进行关系抽取,取得了较好的效果。刘克彬等人[12]实现了基于核函数的中文实体关系自动抽取系统,应用改进的语义序列核函数,结合 KNN 机器学习算法构造分类器来分类并标注关系的类型。通过对 ACE 评测定义的 3 大类 6 子类实体关系的抽取,关系抽取的平均精度达到了 88%。车万翔等人[13]以 2004 年 ACE 评测训练数据作为实验数据,使用两种基于特征向量的机器学习算法 Winnow 和 SVM 进行实体关系抽取,并指出在关系抽取时,应当集中尽力寻找好的特征。Intel 中国研究中心的 ZHANG Yi – Min 和 ZHOU Joe F 等人[14]在 ACL – 2000 上演示了他们开发的一个抽取中文命名实体以及这些实体间相互关系的信息抽取系统,该系统利用基于记忆的学习(Memory – Based Learning,MBL)算法获取规则用以抽取命名实体及它们之间的关系。

3　关系抽取面临的困难

成功的关系抽取取决于正确探测实体,正确判断实体类型,以及正确判断实体间关系的类型。目前对于命名实体的探测以及实体类型的判断技术已经相对成熟,其准确率和召回率一般都能达到 90% 以上。因此能否正确判断实体间关系的类型成为影响关系抽取最终性能的决定性因素。一个比较完整的关系抽取系统应包括依次相连的 5 个模块:NLP 处理和实体抽取、模式匹配或分类、共指消解、新关系处理以及规范化输出。在这个抽取过程中面临的困难基本上可以归纳为以下 3 个方面:

(1)特定领域标引数据集的获取。关系抽取核心部分多采用基于模式匹配和基于机器学习的算法来判断关系是否存在以及关系的类型,而使用这些算法的先决条件是需要预先通过对一个特定领域手工标引的数据集进行学习以获取领域内关系类型的各项特征。此外,这些用于学习的数据集的大小和标引质量都会影响到关系抽取的效果。已标引数据集通常是通过手工标引获得的,这使得特定领域标引数据集的获取比较困难。针对这个问题,一些学者已经提出在关系抽取过程中尽量引入领域 Ontology、领域词表以及 WordNet 等资源来减少关系抽取对已标引数据集的依赖性。此外,学习算法的不断改进也能够减少学习过程中所需的已标引数据量。

(2)模式的获取。基于模式匹配原理的关系抽取方法在很多关系抽取系统中得到了应用。然而,定制特定领域的恰当的关系模式存在较大困难。在以手工方式编制模式过程中,用户必须首先确定给定文集中所有的目标信息表达方式,然后考虑所有的那些表达方式中的变量,最后写出恰当的规则模式。为了使模式编制过程能够更加方便,有学者提出了基于宏的模式编写方式,即给定若干具有特定变量集的宏,当编写领域相关模式时,只要按需要设置宏中的某些变量,就可以自动生成大量相关模式。另外,借助于先进算法的自动模式获取方法也被应用到很多系统中。

(3)共指消解。一个命名实体在文本中可能出现多次,其表现形式也可能不同(例如代名词、反身代词、名词性时间表述等),因此实体间的关系经常被重复探测到。这些指向相同实体间关系的关系实例需要进行合并。在目前的关系抽取系统中,一般使用首语重复法(Anaphors)来解决共指消解问题。通过首语重复法,相关联的实体被合并,并在候选短语表中选择一个最恰当的表达形式。

4　关系抽取的几种技术方法

针对关系抽取过程中的难题,信息抽取领域的学者们进行了长期探索和不懈努力。到目前为止,已经有许多关系抽取方法被应用在各种实验系统当中。这些方法所遵循的技术方法基本可以归纳为:基于模式匹配的关系抽取、基于词典驱动的关系抽取、基于机器学习的关系抽取、基于 Ontology 的关系抽取以及混合抽取方法。

4.1　基于模式匹配的关系抽取

在关系抽取研究领域,普遍使用基于模式匹配的关系抽取方法。这种抽取方法通过运用语言学知识,在执行抽取任务之前,构造出若干基于语词、基于词性或基于语义的模式集合并存储起来。当进行关系抽取时,将经过预处理的语句片段与模式集合中的模式进行匹配。一旦匹配成功,就可以认为该语句片段具有对应模式的关系属性。

在应用基于模式匹配的关系抽取方法时,最困难的步骤是关系模式的建立。最初关系模式的建立需要依靠语言学家对抽取任务涉及的领域语料进行深入分析,借鉴已有语言学成果,穷举各种可能的关系表达,

手工编制关系模式。这样的方法一方面使编制模式的周期太长,应用成本很高;另一方面,当抽取系统被用来进行新领域的关系抽取时,就需要语言学家根据新的领域抽取特点重新编制关系模式,这在现实应用中实现起来非常困难。针对这一问题,一些学者提出了不同的解决思路。

Douglas E. Appelt 等人[15]在 MUC－6 上提出的 FASTUS 抽取系统中,通过引入“宏”的概念将各种领域依赖规则以一种具有扩展性的、通用方式表达。用户只需要修改相应“宏”中的参数设置,就可以快速配置好特定领域任务的关系模式规则。FASTUS 系统中的所有模式规则被分成领域依赖和领域独立两部分。领域独立部分可以看作确定参数的宏。这些模式规则在一个相对粗的粒度层次上覆盖各种句法结果,目标是要对于符合模式的动词构造恰当的谓词－参数(Predicate－argument)关系。领域依赖的规则包含一些参数,这些参数必须通过“宏”的实例化来产生实际模式规则。这些领域依赖规则会指定哪一个动词载有领域相关信息,以及这些参数的领域依赖限制以及规则的语义。FASTUS 系统采用的编译时转换的方式实现了使用 12 个宏规则和 15 个领域依赖的规则就可以实现大概 100 个明确表达的模式的效果,这为系统在处理领域关系抽取任务时的配置工作节约了大量时间。

Roman Yangarber 等人[16]在 MUC－7 上提出的 Proteus 抽取系统采用了基于样本泛化的关系抽取模式构建方法。用户通过 Proteus 系统提供的模式构建界面,对含有某种关系的例句进行分析,识别出所含关系的要素,并将这些要素泛化,最后经用户确认存储经泛化表达的模式。系统还会应用集成的 Meta－rules,从用户生成的简单的主动句模式或独立的名词短语产生一组句法转换器,例如某词的被动词、关系词,以及被动关系、减少关系模式等。Proteus 也能将可选修饰部分插入到产生的变量(例如:临近句子等),来扩展模式的覆盖范围。

4.2　基于词典驱动的关系抽取

与基于模式匹配的关系抽取方法相比,基于词典驱动的关系抽取方法显得非常灵活。新的关系类型能够仅仅通过向词典添加对应的动词入口而被抽取。用户不需要具备复杂的模式语言知识就可以轻松配置抽取系统。

Chinatsu Aone 等人[17]在 MUC－7 上提出了一个快速、灵巧的大规模事件和关系抽取系统(Large－Scale Relation and Event Extraction System,REES)。该系统采用的基于词典驱动的关系抽取方法旨在能够抽取尽可能多类型的关系和事件,但耗费的努力最小,准确率较高。在 REES 系统中,当输入语料经过名称标识和名词短语标识阶段的处理,形成基于 XML 的输出。接着关系识别模块应用词典驱动模型,通过基于句法的一般模式来识别关系和事件。REES 的词典驱动方法需要对于每一个事件指示词设置一个词典入口,而这个词通常是动词。词典入口具体化了该动词参数的句法和语义限制。

基于词典驱动的关系抽取方法的缺点也非常明显。它只能识别以动词为中心词的关系,而对于名词同位语之类的关系抽取就很难实现了。另外,使用这种方法无法对系统中没有对应词汇入口的新关系进行探测。

4.3　基于机器学习的关系抽取

基于机器学习的关系抽取方法是目前应用比较广泛的方法。该方法实质是将关系抽取看作是一个分类问题。通过具体的学习算法,在人工标引语料的基础上构造分类器,然后将其应用在领域语料关系的类别判断过程中。目前使用比较多的学习算法有 MBL 算法和 SVM 算法。

Intel 中国研究中心的 ZHANG Yi－Min 和 ZHOU Joe F 等人[18]在 ACL－2000 上演示了他们开发的一个抽取中文命名实体以及这些实体间相互关系的信息抽取系统,该系统就是利用 MBL 算法获取规则用以抽取命名实体及它们之间的关系。ZHANG Yi－Min 等人将中文实体名和关系识别看作一系列分类问题。整个过程能够被分成两个阶段:第一阶段是学习过程,若干分类器从训练数据构建起来;第二阶段是抽取过程,通过使用学习得到的分类器抽取中文实体名和它们的关系。之所以选择 MBL 作为学习算法,是因为它非常适合处理从大量不同来源获取的特征,并且能记住例外案例和低频案例,而这对于后续的推断阶段非常有用。该系统已经能够抽取的关系类型包括 Employee－of,Location－of,Product－of,和 No－relation。通过提供更多的训练数据,能够轻易扩展关系抽取类型。

Zhu Zhang[19]提出的基于SVM的弱监督关系分类系统应用SVM算法进行关系抽取。Zhu Zhang提出的弱监督学习过程包括两个组件:一个底层监督学习器和一个在其上的Bootstrapping算法。底层监督学习器是一个支持向量分类器,它使用从当前可获得的已标注数据训练而来的模型,对未标记的数据进行分类。Bootstrapping算法则负责选择最有可能被正确标记的实例,并通过使用它们来增强已标记数据的训练效果。该系统在进行分类任务时用到了词语特征、浅句法特征、深层句法特征以及序列标志、实体类型等特征。

Michele Banko等人[20]提出了一个新颖的开放信息抽取方法(Open IE,OIE),实现了对网络上海量异构信息中可能存在的关系的抽取。该方法既不需要手工标注训练集作为训练语料,也不局限于特定领域,而是通过自动学习和统计来实现关系抽取。开放信息抽取方法的实现分为3个阶段:通过对一个相对较小的语料集进行深层解析,自动抽取并标注可信的和不可信的关系三元组。这些三元组的特征向量被作为训练样例进行幼稚贝叶斯分类器的训练;在训练好的分类器上进行大量网络文献的关系抽取。为了确保较高的处理效率,抽取器并不使用解析器对文献进行深层解析,而是将较容易获得的词性标注、序列等特征作为分类器的输入。这一阶段的输出是去除了不必要的修饰词后的候选关系三元组集合;对这些候选三元组进行合并,通过统计的方法计算各个关系三元组的可信度,并建立索引。

4.4　基于Ontology的关系抽取

知识管理过程中,利用信息抽取技术抽取的实体以及实体间的关系来构建和丰富本体,是一种行之有效的方法。另一方面,借助已有的本体层次结构和其所描述的概念之间的关系来协助进行关系的抽取,也不失为一种行之有效的关系抽取方法。

José Iria等人[21,22]提出了一个基于本体的关系抽取通用软件框架—可训练关系抽取框架(Trainable Relation Extraction Framework,T-Rex)。设计该框架的目的是要提供语义网自动化语义标注任务需要的灵活度。由于T-Rex采用了参数化的插件结构,因此可以对多种基于不同抽取算法的插件进行集成和测试。T-Rex最具特色的地方是它采用了规范的基于图的数据模型。该数据模型借助本体实现等级层次的表达结构,并允许以一致的方式任意链接子图,例如共指关系链接,语法关系链接,与HTML格式相关的链接等。T-Rex数据模型的表示是等级化的,能够将语料模型化到字符级、语词级、短语级、语句级和文档级层次。通过对本体的定义和扩充,可以实现使用该多层次数据模型对于语料的多种特征集表达的一致性。

Alexander Schutz等人[1]将DOLCE、SUMO、SportEventOntology等本体有机结合在一起来描述足球领域的相关概念及概念之间的关系,建立了能够自动识别高相关三元组(概念对和概念之间的关系)的RelExt系统。该系统通过从文本集合抽取相关词项和动词,借助语言学和统计学处理过程计算词项之间的相关关系。该系统目前能够处理1 570个足球领域相关概念(类)和487个直接关系。

Marta Sabou和Mathieu d'Aquin等人[23]提出的SCARLET系统通过自动选择和查询本体的方法来发现概念实体之间的关系。例如,当要确定两个概念实体Researcher和AcademicStaff之间的关系时,SCARLET先识别网络上能够提供上述概念实体相关信息的本体,然后综合这些信息来推断概念实体之间的关系。当上述两个概念实体已经在某个本体中被定义,这两个实体间的关系就可以通过本体获取到。如果上述的概念实体在不同的本体中被描述,例如在一个本体中描述了Researcher属于ResearchStaff的关系,在另一个本体中描述了ResearchStaff属于AcademicStaff的关系,则通过关系逻辑推导可以获取Researcher和AcademicStaff之间的关系。

4.5　混合抽取方法

在关系抽取研究的初期阶段,无论是基于词典的抽取方法还是基于模式的抽取方法,都仅将一种抽取方法作为整个关系抽取过程的核心。随着关系抽取研究的不断深入,研究者逐渐意识到,单纯的抽取方法在识别特征和识别模式方面难以避免地会具有局限性。为了将更多的已有关系识别特征加入到关系抽取过程中来,一些将多种现有关系抽取方法相结合的混合抽取方法被提出来。其中具有代表性的是Lucia Specia和Enrico Motta[24]提出的一个抽取语义关系的混合方法。

该方法通过管道(Pipeline)方式引入了解析器(Parser),词性标注器(Part-of-speech Tagger),命名实体识别系统,基于模式的分类器以及词义辨析模块,

并用到了领域本体,知识库以及词语数据库等资源。

该方法的核心策略是匹配一个语言学三元组和他们对应的语义组件。这不仅包括匹配关系,还包括匹配这些关系相关联的项。语言学三元组的探测包括一系列的语言学处理步骤。词项和概念的匹配通过一个领域本体和一个命名实体识别系统引导。关系识别依赖于在领域本体和词库中的知识,以及基于模式的分类和词义辨析模块。除了抽取已经在领域本体中存在的关系,该框架还可以通过模式匹配策略来发现词项类型之间的新关系。

5　结　语

经过 20 多年的发展,关系抽取理论和方法愈加完善。从最初的手工编写模式和词典进行关系抽取,发展到目前借助 Ontology 和知识库等多种知识资源的综合关系抽取,关系抽取的正确率和召回率在不断提高,对不同领域的适应性也在不断加强。目前仍然存在一些比较实际的问题阻碍了关系抽取在实际中的应用,这包括已标引数据集的获取、关系模式的构建、共指消解等问题。随着这些问题的进一步解决,关系抽取技术必然会在增强检索系统功能、语义网标注、本体学习等领域得到广泛应用。

参考文献:

[1] Schutz A, Buitelaar P. RelExt:A Tool for Relation Extraction from Text in Ontology Extension[C]. 4*th International Semantic Web Conference*, Galway, Ireland, November 6 - 10, 2005:593 - 606.

[2] Katrenko S, Adriaans P. Learning Relations from Biomedical Corpora Using Dependency Tree Levels[C]. In:*Proc. BENELEARN conference*(2006), 2006.

[3] Relationship Extraction[EB/OL]. [2008 - 05 - 30]. http://en.wikipedia.org/wiki/Relationship_extraction.

[4] The ACE 2004 Evaluation Plan[EB/OL]. [2008 - 05 - 30]. http://www.nist.gov/speech/tests/ace/2004/doc/ace04 - evalplan - v7.pdf.

[5] Automatic Content Extraction 2008 Evaluation Plan (ACE08)[EB/OL]. [2008 - 05 - 30]. http://www.nist.gov/speech/tests/ace/2008/doc/ace08 - evalplan.v1.2.pdf.

[6] MUC[EB/OL]. [2008 - 05 - 30]. http://www.itl.nist.gov/iaui/894.02/related_projects/muc/.

[7] ACE[EB/OL]. [2008 - 05 - 30]. http://www.nist.gov/speech/tests/ace/.

[8] ACE08 Annotation Tasks[EB/OL]. [2008 - 05 - 30]. http://projects.ldc.upenn.edu/ace/annotation/.

[9] 邓擘, 樊孝忠, 杨立公. 用语义模式提取实体关系的方法[J]. *计算机工程*, 2007,33(10):212 - 214.

[10] 姜吉发, 王树西. 一种自举的二元关系和二元关系模式获取方法[J]. *中文信息学报*, 2005,19(2):71 - 77.

[11] 顾雪峰. 基于动态粒度思想的实体关系识别方法研究[EB/OL]. [2008 - 05 - 30]. http://www.cnki.com.cn/grid20/Detail.aspx.

[12] 刘克彬, 李芳, 刘磊,等. 基于核函数中文关系自动抽取系统的实现[J]. *计算机研究与发展* 2007,44(8):1406 - 1411.

[13] 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.

[14] Zhang Y M, Zhou J F. A Trainable Method for Extracting Chinese Entity Names and Their Relations[C]. In: *Proceedings of the Second Chinese Language Processing Workshop*, Hong Kong, 2000:66 - 72.

[15] Appelt D E, Hobbs J R, Bear J, et al. SRI International FASTUS System:MUC - 6 Test Results and Analysis[C]. In: *Proceedings of the 6th Message Understanding Conference* (*MUC* - 6), 1995:237 - 248.

[16] Roman Y, Grishman R. NYU:Description of the Proteus/PET System as Used for MUC - 7 ST[C]. In: *Proceedings of the 6th Message Understanding Conference* (*MUC* - 7), 1998.

[17] Aone C, Ramos2Santacruz M. Rees:A large - scale relation and event extraction system[C]. In:*Proc of the 6th Applied Natural Language Processing Conference*, New York, 2000:76 - 83.

[18] Zhang Y, Zhou J F. A Trainable Method for Extracting Chinese Entity Names and Their Relations[C]. In: *Proceedings of the second Chinese Language Processing Workshop*, ACL, 2000:66 - 72.

[19] Zhu Z. Weakly - supervised Relation Classification for Information Extraction[C]. In:*Proceedings of the Thirteenth ACM conference on Information and Knowledge Management*, Washington D. C., 2004:581 - 588.

[20] Banko M, Cafarella M J, Soderland S, et al. Open Information Extraction from the Web[C]. In: *Proceeding of the International Joint Conferences on Artificial Intelligence*, 2007.

[21] Iria J. T - Rex:A Flexible Relation Extraction Framework[C]. In: *Proceeding of the 8th Annual Colloquium for the UK Special Interest Group for Computational Linguistics* (*CLUK'05*), Manchester, January 2005.

[22] Iria, Mr. José, Ciravegna, Fabio. Relation Extraction for Mining the Semantic Web[C]. In: *Proceedings Machine Learning for the Semantic Web Dagstuhl Seminar* 05071, Dagstuhl, 2005.

[23] Sabou M, Mathieu d'Aquin, Motta E. SCARLET:SemantiC relAtion DiscoveRy by Harvesting onLinE onTologies[C]. In: *Proceedings of the 5th European Semantic Web Conference*, June, 2008.

[24] Specia L, Motta E. A Hybrid Approach for Extracting Semantic Relations from Texts[EB/OL]. [2008 - 05 - 30]. http://www.dcs.shef.ac.uk/~lucia/publications/SpeciaMotta_OLP2 - 2006.pdf.

(作者 E - mail:xujian@mail.las.ac.cn)