

北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2014.005

日语时间表达式识别与日汉翻译研究

赵紫玉 徐金安[†] 张玉洁 刘江鸣

北京交通大学计算机与信息技术学院, 北京 100044; [†] 通信作者, E-mail: xja2010@gmail.com

摘要 基于自定义知识库, 提出一种知识库强化规则集以及与统计模型相结合的日语时间表达式识别方法。旨在不断提高时间表达式的识别精准度按照 Timex2 标准对时间表现的细化分类, 结合日语时间词的特点, 渐进地扩展重构日语时间表达式知识库, 实现基于知识库获取的规则集的优化更新。同时, 融合条件随机场 CRF 统计模型, 提高日语时间表达式识别的泛化能力。通过考察基于短语的翻译模型翻译时间词的精度, 提出统计机器翻译(SMT)结合规则翻译日语时间词的必要性。实验结果显示, 日语时间表达式识别的开放测试 F1 值达到 0.8987, 基于日汉时间词平行字典与规则的翻译精度和召回率都略高于基于统计机器翻译模型。

关键词 知识库; 规则; 统计模型; 统计机器翻译; 时间词平行字典

中图分类号 TP391

Japanese Time Expression Recognition and Translation

ZHAO Ziyu, XU Jin'an[†], ZHANG Yujie, LIU Jiangming

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044;

[†] Corresponding author, E-mail: xja2010@gmail.com

Abstract Based on the defined knowledge base, the authors presented a Japanese time expression recognition method through combining rules set strengthened by knowledge base with statistical model. In order to increase recognition accuracy, according to the Timex2 standards' granular classification on time, the knowledge base was progressively expanded and reconstructed given the Japanese time characteristic to achieve rules set optimization and update. Simultaneously, CRF model was fused to enhance the generalization ability of Japanese time expression recognition. The authors studied the time translation accuracy of phrase-based translation model and proved the necessity of combing rules with statistical machine translation (SMT). Experiment results show that the F1 value of Japanese time expression recognition reaches 0.8987 on open test, and both the precision and recall by the method based on rules and parallel dictionary of Japanese to Chinese time are a bit higher than those by the method based on statistical translation model.

Key words knowledge base; rule; statistical model; statistical machine translation; time parallel dictionary

时间表达式是句子中的重要成分, 是关键信息的载体。时间表达式的抽取和处理是当前自然语言处理中的一个重要研究方向。正确识别和翻译时间表达式具有重要的意义。

近年来, 时间表达式的识别、规范化以及翻译在事件跟踪、时间关系推理、时序定位等方面的应用愈加广泛, 不仅可以提高分词、句法分析的精度,

还可改善机器翻译、信息抽取、文本摘要、对话系统的性能。比如, 在机器翻译中, 可以使译文更加流畅^[1]; 在多文档自动摘要中, 可以对文档信息进行时序排序^[2]; 在自动问答系统中, 可以用于回答“多久, 何时”等与时间相关的问题。

信息理解会议(Message Understanding Conference, MUC)在 1995 年首次将时间表达式的识别

国家自然科学基金(61370130)、科技部国际科技合作计划(K11F100010)、中央高校基本科研业务费专项资金(2010JBZ2007)、中国科学院计算技术研究所智能信息处理重点实验室开放课题(IIP2010-4)和北京交通大学人才基金(2011RC034)资助

收稿日期: 2013-06-15; 修回日期: 2013-08-15; 网络出版时间: 2013-11-11 10:25

作为命名实体识别的一个子任务。美国国家技术标准局 (NIST)、ACE2005(Automatic Content Extraction)和 SemEval2007(Semantic Evaluations)都有举办时间表达式识别与归一化(Time Expression Recognition and Normalization, TERN)的评测。目前, TERN 评测涉及阿拉伯语、英语和汉语等, 对于韩语、法语和西班牙等语言也有人进行初步的探索性研究, 但是针对日语时间表达式识别的评测工作比较少。

时间表达式识别方法一般可以分为以下两类。一类是基于规则的方法^[3], 该类方法通过分析短语内部的构成规律和短语外部的约束信息来识别时间表达式^[4], 利用了时间表现形式规范的特点, 但是规则的撰写耗时耗力, 可移植性较差, 而且构建的规则往往会有粒度过粗的缺点。另一类是基于机器学习的时间序列标注方法^[5-7], 这类方法最大的特点是可以充分利用已标注上下文信息, 使得识别召回率较高, 而且无需消耗太多的人力。但是受限于日语时间表达式语料规模的局限性和质量, 训练语料容易存在数据稀疏问题, 也无法充分利用时间表现形式相对规范的特点, 使得基于机器学习的序列标注方法难以充分发挥它的优势。

本文结合以上两类方法的优点, 提出规则与统计相结合的日语时间表达式识别方法, 旨在提高识别的精准度、召回率以及识别的泛化能力。另外, 本文基于规则以及日汉时间词平行字典, 探索性地考察基于短语的统计翻译模型翻译时间词的性能, 以期提高时间词翻译精度, 为机器翻译等工作提供技术基础。

1 基本概念及问题分析

1.1 时间表达式

日语和汉语在时间的使用上有很多相同和相似之处。参照 Timex2^[8]中关于中文时间表达式的描述, 本文将日语时间表达式定义为由一个或多个时间基类组成的时间短语, 即时间表达式为时间基类的序列, 如“每年夏休み”此时间表达式由 2 个时间基类集合而成: “每年”和“夏休み”。“时间基类”, 即基本时间类型, 本文提出 7 种基本时间类型, 是构成时间表达式的最小组成类型。由此可定义日语时间表达式为

$$T_e = (t_1, t_2, \dots, t_i, \dots, t_m), m \geq 1, \quad (1)$$

其中, T_e 为时间表达式, 是 m 元组; 其中 $t_1, t_2, \dots, t_i, \dots, t_m$ 是 m 个独立的时间基类。

1.2 时间基类

本文提出的 7 种时间基类的具体描述说明如表 1 所示。

1.3 日语时间表达式知识库

知识库是关于某一项领域的陈述性知识、过程性知识和策略性知识的集合^[9]。在该集合中各类知识通过一定的表示方法表示, 并建立相互之间的联系。从存储知识的角度来看, 知识库以描述型方法来存储和管理知识。我们初始创建的知识库如表 2~5 所示。

1.4 问题分析

本文通过对大量日语语料分析, 发现时间表达式的多样性问题及若干歧义现象:

- 1) 一般日语时间表达式中会包含触发词^[8], 但是也存在不包含触发词的时间表达式; 而且并非所有包含触发词的表达式都是时间, 如“日中経済協

表 1 日语时间表达式时间基类
Table 1 Time basic class unit of Japanese time expression

时间基类名称	时间基类描述
绝对时间	指一个具体固定且与时间的推移无关的时间词, 如“2013 年 5 月 27 日”
相对时间	指随着时间的推移, 所指的时间产生变化的时间词, 如“一昨年、夕方”
段时间	指一个持续的时间段, 如“一週間”
集合时间	指某一类时间的集合, 如“梅雨期間、毎日”
事件触发时间	指跟特定事件相关的时间词, 如“天和元年、平成 11 年”
文化相关时间	指日本文化纪念日的时间词, 如“建国記念の日、公休日”
不特定时间	指日本二十四节气、七十二候、季节、六曜历法、天干地支等时间词, 如“立春、雨水”

表 2 日语时间触发词知识库
Table 2 Japanese time trigger knowledge base

触发词种类	触发词个数	触发词示例
绝对时间触发词	7	時、分、秒
相对时间触发词	122	夜明け、朝午前
段时间触发词	19	秒間、分間、時間
集合时间触发词	2	期間、毎
事件触发时间触发词	557	世紀、紀元、王朝、時代
文化相关时间触发词	115	の日、纪念日、誕生日
不特定时间触发词	148	曆、盛夏、夏、立春

表 3 日语时间表达式边界知识库
Table 3 Boundary knowledge base of Japanese time expression

边界种类	边界词个数	边界词示例
助词	23	格助词、提示助词
日语标点符号	47	、。『』「」
其他	2	約

表 4 日语时间表达式规则关键词知识库
Table 4 Rules' key words knowledge base of Japanese time expression

关键词所属类别	关键词个数	关键词示例
触发词	970	7类时间基类时间触发词
边界	71	助词、日语标点符号、其他
阿拉伯数字	10	0、1、2
中文简体数字	10	零、一、二
中文繁体数字	10	零、壹、貳

表 5 基于日语维基百科的知识库
Table 5 Knowledge base based on Japanese Wikipedia

列表名称	词条举例
世界性节日	國際労働者の日、母の日
日本国节日	国民の祝日、建国記念日
基督教节日	クリスマス(圣诞节)、キリストのはりつけ(基督受难日)
二十四節気	立春、雨水
七十二候	東風解凍、黄鶯睨院

会理事長”中的“日中”是相对时间的触发词，结合上下文可以看出此处的“日中”并不是时间表达式。因此单纯的基于触发词的规则方法不能准确地识别时间表达式。

2) 某些完整时间表达式中包含非时间概念词的时间单元，如“九時三分前”中的方位词“前”，由于其与时间概念词结合起来可表达完整的时间意义，因此这类非时间概念词也需要准确识别。

综上所述，不能仅通过时间触发词等词形信息来制定规则，还应该结合知识库来强化规则集并结合统计模型，提高识别准确率。因此，本文针对时间表达式的内部组成结构，提出并构建基于知识库强化规则集和统计模型相结合的时间表达式识别系统。

2 规则与统计相结合的日语时间表达式识别

2.1 识别系统结构

本文提出基于知识库强化获取规则集以及规则集与统计模型相结合的识别方法。首先通过初始构建的知识库强化获取规则集，训练统计模型；其次分别基于规则和统计两种方法进行识别，并整合二者的识别结果；然后基于错误驱动学习^[10]的思想，根据整合后的识别结果，比较初始构建的知识库与识别结果的差异来校正知识库及规则，体现重构知识库的触发环境。识别系统流程如图 1 所示。

2.2 基于条件随机场的日语时间表达式识别

条件随机场(conditional random field, CRF)^[11]在观测序列的基础上对目标序列进行建模，重点解决序列化标注的问题。条件随机场模型既具有判别式模型的优点，又具有产生式模型考虑到上下文标记间的转移概率，以序列化形式进行全局参数优化和解码的特点，解决了其他判别式模型(如最大熵马尔科夫模型)难以避免的标记偏置问题。因此，本文基于 CRF 模型，将日语时间表达式识别看作一个序列标注过程，观察值为所有日语字符集合，并基于由字构词的理念，利用词位信息来标记时间词。本文采用 BIEO 的组块表达方法来标识时间表达式，一个组块即视为一个时间表达式，而对于不同种类

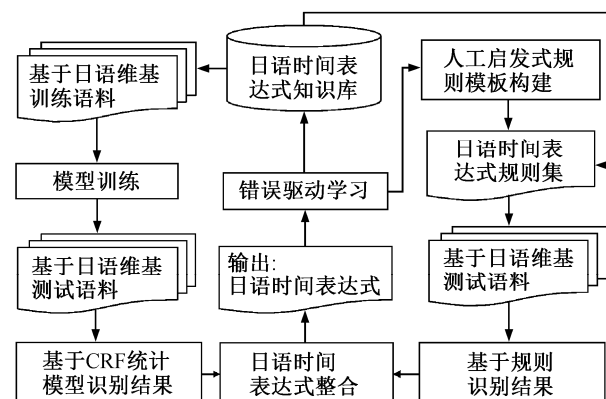


图 1 日语时间表达式识别流程图

Fig.1 Flow chart of Japanese time expression recognition

的时间表达式,即绝对时间、相对时间、段时间、集合时间、事件触发时间、文化相关时间以及不特定时间,分别用 Absolute, Relative, Duration, Set, EventAnchored, CultureRelated 和 Fuzzy 标识。根据 BIEO 分类标记,本系统中的分类标注集说明如表 6 所示。

2.2.1 特征模板与特征

本文分析并研究日语时间表达式内部结构和上下文环境对其的影响程度,使用词形与日语形态素信息作为特征(表 7)。具体特征模板的设置如表 8 所示,其中 C 代表词形, S 代表词性。

2.2.2 基于 CRF 识别方法

基于条件随机场理论(CRFs)统计模型的识别流程图如图 2 所示。

首先预处理维基语料;其次将对预处理后的观

表 6 四词位分类标注集
Table 6 Label set of four words

标注符号	符号说明
B-[时间基类]	时间表达式的开始位置
I-[时间基类]	时间表达式的中间位置
E-[时间基类]	时间表达式的结尾位置
O	非/其他位置

说明:时间基类是指 Absolute, Relative, Duration, Set, EventAnchored, CultureRelated 和 Fuzzy。

表 7 特征
Table 7 Features

特征描述	说明
词形	词本身
形态素信息	包括词性,日语活用型等

表 8 特征模板
Table 8 Feature template

特征类型	特征模板
Unigram(一元)	$C_n, S_n, n=-2, -1, 0, 1, 2$
Sigram(二元)	$C_n C_{n+1}, S_n S_{n+1}, n=-2, -1, 0, 1$

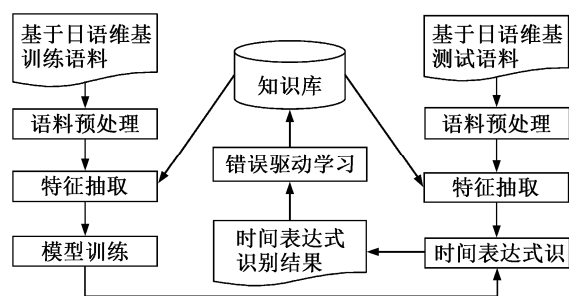


图 2 基于条件随机场的日语时间表达式识别

Fig. 2 Japanese time expression recognition based on CRF

察序列进行特征抽取,使用 CRF^[12]完成参数训练过程;识别过程即解码过程,在测试语料上使用已经训练好的 CRF 模型识别日语时间表达式;参考基于错误驱动算法的思想,动态扩展重构知识库。

2.3 基于知识库强化获取规则集的日语时间表达式识别

本文采用的规则集由自定义的人工启发式规则模板结合日语时间表达式知识库自动生成。

2.3.1 人工启发式规则模板

由于目前较少人在做日语时间词识别的研究工作,因此很难进行对比试验,我们根据自己的研究经验分别对 7 种日语时间基类构建人工启发式规则模板,如表 9 所示。

2.3.2 基于知识库强化获取规则集识别方法

本文在人工规则模版上利用知识库信息,生成强化的规则集。利用正则表达式匹配方式,以句子为单位识别时间表达式。基于错误驱动学习的思想,根据识别结果与标准数据的差异,校正知识库。

2.4 规则和统计的融合策略

一方面,基于规则的方法可以很好地表达语言的不确定性现象,从而克服统计模型在此方面的缺点;另一方面,统计模型的泛化能力可以弥补基于规则方法的移植性差等缺点。二者的结合可以达到很好的互补效果。因此,融合模型的建立已成为时间表达式识别过程中的关键问题。本文提出的融合策略如下。

1) 以基于规则的识别结果为基础,观察未被识别到的时间表达式,通过错误驱动更新日语时间表达式知识库,并使用人工启发式的方法,对规则模版进行修正。最后以更新的知识库和修正的模版为基础重构规则集。错误主要表现在两个方面:一方面,规则识别错误,例如“二日市”等专有地名,利用形态素分析信息(名词-固有名词-地域-一般),加强规则的限制;另一方面,规则识别召回率低,例如“卒業後しばらくして(毕业后一段时间)”,“から……(从……到)”等,总结时间表达式,提高规则的表现能力。

2) 使用更新的知识库和规则模板处理语料,提高训练语料的质量,然后利用统计模型的泛化能力进行时间表达式的识别。

3) 在规则与统计结果整合的过程中,采取贪心策略。最终的时间表达式为规则与统计同时覆盖到的片段的最长序列,使用如下公式:

表 9 人工启发式规则模板和规则集示例
Table 9 Artificial heuristic rule templates and rules set samples

时间表达式类型	规则模板示例	规则示例(正则表达式)
绝对时间	[数字]+[绝对时间触发词]	\d+年\d+月\d+日
相对时间	[边界词][相对时间触发词]+[边界词]	を.*(夜明け 朝午前).*も
段时间	[数 何 半][段时间触发词]	数(秒間 時間)
集合时间	每[绝对时间出发词 集合时间触发词]	每(年 月 日)
事件触发时间	[事件触发时间触发词][数字]+年	平成\d+年
文化相关时间	[文化相关时间触发词][节日]	國際労働者の日
不特定时间	[不特定时间触发词]	立春 雨水 啓蟄 春分

$$\text{TimeExpression} = \text{sequence}(\text{maxleft}(R, S), \text{maxright}(R, S)),$$

当 $R \cap S \neq \emptyset$, (2)

其中, $\text{maxleft}(R, S)$ 表示规则结果与统计结果最左边字符, $\text{maxright}(R, S)$ 表示规则结果与统计结果最右边字符, $\text{sequence}(i, j)$ 函数表示从字符 i 到字符 j 的连续字符串。

综上所述, 本文提出的融合方法包括: 一方面, 系统采用错误驱动人工启发式的方法, 利用知识库强化规则集, 并融合统计模型泛化能力, 识别时间表达式; 另一方面, 系统采用贪心策略, 整合规则和统计的识别结果。因此, 本文提出的规则和统计结合的日语时间表达式识别框架及方法是有效的。

2.5 日语时间词的识别与翻译

图 3 是日语时间词翻译成汉语的实验系统框图。具体翻译算法如下。

1) 翻译模型/语言模型训练。用 28 万句子级日语语料训练 Moses 机器翻译模型。

2) 识别时间词。用本文提出的日语时间词识别模型识别并抽取语料中的日语时间词。

3) 翻译过程。一方面, Moses 解码过程中使用训练好的翻译模型翻译步骤 2 识别出的日语时间词; 另一方面, 使用基于词典与规则翻译步骤 2 识别出

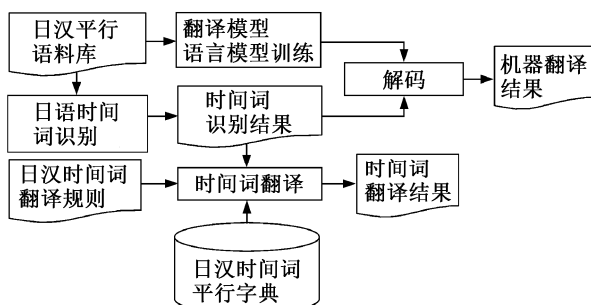


图 3 日语时间词识别与翻译

Fig. 3 Japanese time recognition and translation

的日语时间词。

3 实验及分析

3.1 识别实验语料与评测方法

本文识别实验语料采用具有实时性的日语维基资源库, 将该语料经过去标签、篇章分割等预处理后, 随机分为测试语料以及训练语料两部分, 具体语料信息如表 10 所示。

本文使用 PRF 评测指标^[13]评测识别结果, 并提出使用覆盖度(Coverage)指标评测系统正确识别的时间表达式覆盖 7 类时间单元的程度, 计算公式如下:

$$\text{覆盖度(Coverage)} = \frac{\text{正确覆盖的时间单元个数}}{\text{时间单元总数}}, \quad (3)$$

3.2 识别实验结果及分析

本文对时间表达式识别系统进行多次实验, 比较基于规则方法和基于统计方法的识别结果的差异, 采用融合策略, 使本文提出的方法取得突出的识别效果。实验结果如表 11 和 12。

从表 11 可以看出基于规则与统计融合的识别方法优于单独使用规则或统计的方法, 主要体现在

表 10 规则与统计结合的日语时间表达式识别实验语料信息
Table 10 Experiment corpus of Japanese time expression recognition based on rules and statics

实验数据	句子数	平均句长(字数)	时间表达式数
训练语料	236206	60.8	452498
测试语料	1691	57.8	3766

表 11 时间表达式识别结果
Table 11 Time expression recognition result

方法	P	R	F1
基于规则方法	0.8110	0.7534	0.7811
基于统计方法	0.9432	0.8383	0.8877
规则统计融合	0.9364	0.8639	0.8987

表 12 时间基类识别结果覆盖度
Table 12 Time basic class unit recognition result by coverage

方法	绝对时间	相对时间	段时间	集合时间	事件触发时间	文化相关时间	不特定时间
基于规则方法	0.8964	0.4005	0.8874	0.8902	0.9383	0.9478	0.4500
基于统计方法	0.9678	0.8816	0.9665	0.8140	0.9515	0.9395	0.1214
规则统计融合	0.9698	0.8816	0.9686	0.8963	0.9515	0.9560	0.5142

召回率上;然而准确率有细微的降低,主要原因是融合算法目前按照时间表达式的字表面特征进行融合,还有很大的优化空间,此外,规则的不完备性可带来的噪声和语料自身的噪音也很难避免。表 12 说明基于规则的方法在有明显格式的时间基类(集合时间和文化相关时间)上表现较好,例如“期間”,“平成年間”和“國際労働者の日”。但是,基于规则的方法在格式相对不明显的时间识别上效果较差,特别是相对时间。一方面,相对时间识别规则中的边界词难以确定,本文暂使用日语的格助词及标点符号作为边界词;另一方面,统计模型识别的方法能够通过统计特征识别边界。因此基于统计的方法在相对时间基上的识别更为突出。

分析实验结果发现,无论是基于规则还是统计的方法在不特定时间上的识别效果尤其差,其中原因包括:1) 相对训练数据稀少且难以保证其准确性;2) 存在严重的词义歧义问题,例如“雨水”、“小雪”和“大雪”等,不仅作为二十四节气,且更为普遍地作为天气的自然现象词汇。

综上所述,利用统计与规则的融合策略,弥补各自方法的不足。实验结果显示基于规则的方法和基于统计模型的方法,能够有效地识别日语时间表达式,并且基于规则与统计相结合的方法能够提高日语时间表达式的识别效果。

3.3 翻译实验语料与评测方法

本实验语料如表 13 所示。

3.4 翻译实验结果及分析

实验结果如表 14 所示。可以看出,本文提出的基于规则与统计的日语时间词识别方法效果明显; Moses 模型对时间词的翻译效果仍然不是特别理想,原因是统计机器翻译模型尤其是基于短语的翻译模型即使用大数据语料训练,其对未登录词的翻译效果也不佳^[14],而具有多样性特点的时间词在未登陆词中分布很典型也很广泛,因此时间词的翻译需要结合规则。

表 13 时间词翻译实验结果

Table 13 Experiment corpus of Japanese time translation

实验数据	日汉平行语料	含日语时间词的句子数	汉语时间词数
实验语料	28 万	18780	20658

表 14 时间词翻译结果

Table 14 Japanese time translation result

方法	P	R	F1
规则与统计结合的时间词识别模块	0.9517	0.8652	0.9064
Moses 翻译模型	0.4891	0.3762	0.4253
基于字典与规则的时间词翻译模块	0.5060	0.4745	0.4897

近年来, BLEU^[15]等一些常用的自动评测方法能够很好地评价句子级和篇章级文本的翻译效果,但是对短语级文本的翻译效果评测不是特别理想。因此本文提出使用 PRF 值评测时间词的翻译结果。

4 结语

本文提出的方法在训练语料规模匮乏的条件下,减少了人工参与,使得系统在拥有较好的模型学习能力的同时高质量识别日语时间表达式,这是本方法的优点之一,但也会引起一些识别歧义问题,识别只包含一个词或知识库特征不明显的时间表达式带来的识别歧义问题更为突出。本文另一优点是用知识库增强规则集解决歧义问题,实现知识库的优化扩展与重构。今后,我们将尝试更多有效的特征,提高统计模型的识别精度,特别针对不特定时间,使用深层语义特征提高识别效果,以及尝试运用错误驱动思想的规则筛选策略,达到自动学习规则,提高识别性能和效率。另外,本文探索性地研究了日语时间的翻译,发现基于短语的统计翻译模型对时间词的翻译效果不佳,在未来研究工作中,我们将着重基于规则与统计的融合,进一步提高时间词翻译精度,并将其应用于篇章级文本翻译,使译文更加流畅。

参考文献

[1] 郭桐,周雅倩,黄萱菁,等. 自动构建时间基元规

- 则库的中文时间表达式识别. 中文信息学报, 2010, 24(4): 3-10
- [2] 贺瑞芳, 秦兵, 刘挺, 等. 基于依存分析和错误驱动的中文时间表达式识别. 中文信息学报, 2007, 21(5): 36-40
- [3] Maqur P, Dale R. A rule based approach to temporal expression tagging // Proceeding of the International Multi Conference on Computer Science and Information Technology. 2007, 293-03
- [4] Wu Mingli, Li Wenjie, Lu Qin, et al. A Chinese temporal parser for extracting and normalizing temporal information // International Joint Conference on Natural Language Processing(IJCNLP), 2005. Volume 3651: 694-706
- [5] Ahn D, Adafre S F, de Rijke M. Recognizing and interpreting temporal expressions in open domain texts. Digital Information Management, 2005, 3(1): 14-20
- [6] Ahn D D, Adafre S F, de Rijke M. Towards task-based temporal extraction and recognition. 2005.
- [7] Hacioglu K, Chen Y. Benjamin douglas automatic time expression labeling for English and Chinese text// Computational Linguistics and Intelligent Text Processing, 2005, 3406: 548-559
- [8] ACE (Automatic Content Extraction) Chinese Annotation Guidelines for TIMEX2(Summary). Version 1.2 20050610
- [9] 刘成亮, 韩海伟. 知识库系统的原理及其在智能搜索引擎中的应用. 电脑知识与技术, 2008, 8: 1512-1514
- [10] Brill, Eric. Transformation-based error-driven learning and natural language processing: a case study in part of speech tagging. Computational Linguistics, 1995, 21(4): 543-565
- [11] Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. The Journal of Machine Learning Research, 2001, ICML01: 282-289
- [12] CRF++: Yet another CRF toolkit[CP/OL]. (2012-05-30)[2012-08-21]. <http://crfpp.googlecode.com/svn/trunk/doc/index.html>
- [13] NIST. The ACE 2007(ACE07)Evaluation Plan: Evaluation of the Detection and Recognition of ACE Entities, Values, Temporal Expressions, Relations and Events[EB/OL], 2007
- [14] Zong, Mei Tu Yu Zhou Chengqing. A universal approach to translating numerical and time expressions // Proceedings IWSLT, 2012(2012)
- [15] Papineni K, Roukos S, Ward T, et al. Bleu: a method for automatic evaluation of machine translation. 2001