

博士论坛

## 非结构化信息抽取关键技术研究探讨

周法国<sup>1</sup>, 王映龙<sup>2</sup>, 杨炳儒<sup>3</sup>, 宋泽锋<sup>3</sup>

1.中国矿业大学(北京) 机电与信息工程学院, 北京 100083

2.江西农业大学 软件学院, 南昌 330045

3.北京科技大学 信息工程学院, 北京 100083

收稿日期 2008-10-7 修回日期 2009-1-15 网络版发布日期 2009-5-8 接受日期

**摘要** 以基于内在认知机理的知识发现理论为指导, 针对汉语命名实体识别的难点, 充分考虑专家知识在命名实体识别中的作用; 根据不同的实体类型, 采用灵活变化的统计与规则相结合的方式; 采用各种技术来研究信息抽取的任务, 如: 机器学习技术、篇章分析与理解技术、句法分析技术、图算法与图挖掘技术、词计算技术、快速全文检索技术等; 该文探讨的是不仅要从文本中获取简单子句中的关系, 还要获得跨句子、段落中的实体关系。

**关键词** [信息抽取](#) [内在认知机理](#) [命名实体识别](#) [共指消解](#) [机器学习](#)

分类号

## Research on key technologies of unstructured information extraction

ZHOU Fa-guo<sup>1</sup>, WANG Ying-long<sup>2</sup>, YANG Bing-ru<sup>3</sup>, SONG Ze-feng<sup>3</sup>

1.School of Mechanical Electronic & Information Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China

2.School of Software, Jiangxi Agriculture University, Nanchang 330045, China

3.School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China

### Abstract

Under the guidance of the Knowledge Discovery Theory based on Inner Cognitive Mechanism (KDTICM), this paper focuses on the difficult points of the Chinese named entity recognition. It takes into full consideration the role that expert knowledge plays in the named entity recognition. According to the different types of entities, this paper flexibly combines the method based on statistics and rules. A series of techniques are adopted to deal with the tasks of information extraction, such as machine learning, document understanding and analysis, parsing technique, graph algorithm and graph mining, computing with words, rapid full-text retrieval, etc. The object of exploration of this paper is to obtain from text not only the relations in simple sentences but also the entity relations across sentences and paragraphs.

**Key words** [information extraction](#) [inner cognitive mechanism](#) [name entity recognition](#) [anaphora resolution](#) [machine learning](#)

DOI: 10.3778/j.issn.1002-8331.2009.14.001

通讯作者 周法国

### 扩展功能

#### 本文信息

▶ [Supporting info](#)

▶ [PDF\(1339KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [参考文献](#)

#### 服务与反馈

▶ [把本文推荐给朋友](#)

▶ [加入我的书架](#)

▶ [加入引用管理器](#)

▶ [复制索引](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

#### 相关信息

▶ [本刊中包含“信息抽取”的相关文章](#)

▶ [本文作者相关文章](#)

- [周法国](#)
- [王映龙](#)
- [杨炳儒](#)
- [宋泽锋](#)