
基于概念的自然语言处理

马颖华, 苏贵洋

(上海交通大学计算机系 3029 室, 200030);

E-mail: ma-yh@cs.sjtu.edu.cn

摘要: 基于“概念”进行自然语言处理是近期的研究热点。本文在分析近几年概念的研究成果和经验基础上, 从统计模型和概念词典两个方面对基于“概念”的自然语言处理进行了总结, 并分析了现存的问题, 以及未来研究的方向。作者提出对于“概念”的研究应重视其主观性的一面, 并从“概念”自身的规定性出发, 逐步规范“概念”在语言学中的定义, 从而理清应用“概念”作为语言处理基本单元的处理思路。“概念”作为语言表义(完整意义上的语义)的基本单元, 它的内涵和外延都表现为知识。而知识的表达与描述正是人工智能中的难点。随着“概念”在语言处理中越来越广泛的应用, 解决这个难点将越发迫切, 应得到更多的关注和重视。

关键词: 概念; 自然语言处理

作者简介: 马颖华(1973—), 女, 山东济南人, 博士研究生, 主要研究领域为自动标引及算法设计; 苏贵洋(1974—), 男, 山东济宁人, 博士研究生, 主要研究领域为信息检索、网络信息智能处理。

Natural Language Processing Based on Concepts

MA Ying hua, SU Gui yang

(Shanghai Jiaotong University, Shanghai, 200030, China)

E-mail: ma-yh@cs.sjtu.edu.cn

Abstract: Natural Language Processing based on concepts is a hotspot of NLP. Experience are summarized from two aspects: statistics model and concept dictionary. Problems now existing and the direction of researching in the future are discussed. In author's opinion, researches on concepts should pay more attention on subjectivity of concepts. The definition of concept in linguistics should be standardized, in order to make the thought of using concepts as basic unit of language processing clear. "Concept" is the basic ideographic unit of language, and its intension and extension are all knowledge. How to describe and express knowledge is difficult. Along with the researches on concept become more and more popular, the problem will be inescapable.

Key words: Concept; Natural Language Processing;