

计算语言学专业设置的几个问题 及其课程名目内容提要

Some Problems of Instituting the Course
of Computational Linguistics

孙介铭

根据权威人士预测,智能化自然语言产品在下一个十年的国际市场中将会增加100倍!为了满足这方面的需要,有必要建立一个以计算语言学(CL)理论为基础以及自然语言计算机处理为技术目标的交叉学科,以便系统正规地培养这一领域高层次复合型人才。

CL是由(1).自然语言形式化描述(2).自然语言计算机处理以及(3).两者有机结合的方法学所组成。

计算语言学家应具备下列基础知识:(1).现代语言学最新知识(2).数学基础(3).计算机科学基础(4).形式语法与形式理论(5).信息系统和数据库系统(6).解决问题的方法(7).测试理论。

计算语言学家还应具备下列专业知识:(1).人-机通讯基础(2).知识表达形式系统及其加工处理过程(3).句法分析方法和策略,规范化语言和语言生成(4).自然语言系统结构和组成(5).自然语言系统类型、接口、对话系统、文本分析、理解系统、机器翻译系统(6).系统的评估方法、工艺过程优化程度的评估。

对CL专家能力的要求是:(1).能判断某一语言借助计算机来解决的可能性(2).鉴别未解决问题的性质及寻求解决该问题的新方法(3).对用计算机来解决语言学问题时提出独立见解和实际方法(4).对软件开发工程中的中间成果能做出达到最终目标的综合处理(5).对用CL来解决问题的方法作出有科学根据的评价(6).能从用户角度提出解决问题的方案。

CL专业本科生的学制五年,相当于精选压缩的双学位课程,其课程名目与内容如下:

(1).计算语言学基础(上,下).绪论,CL的性质、任务与应用范围,语言处理的基本算法与规则系统,自然语言句法分析基础,信息交换原理,人际通讯,人-机通讯,信息系统基础,自然语言处理的目标(索引生成,自动分类,自动文摘,事实抽取,机器翻译,文本生成,用户接口等),信息系统互作用概论(问答系统,对话系统,多模式系统)。

(2).现代语言学基础.作为经验科学的语言学,描写,形式,比较语言学,语言学子系统(形态学,词汇学,句法学,语义学,语用学以及语法理论)。

(3).数学与信息科学基础(上,下).代数,数学分析,离散数学,图论,形式逻辑(包括定理证明法),模型论,统计学,形式语言理论,自动机原理,算法语言,数据结构。

(4).程序设计语言和程序设计实践.C, LISP, SMALLTALK, PROLOG, 以及DOS和UNIX操作系统,程序设计环境,工程模型设计,软件工程。

(5).经验研究工作的基本方法.观察,描述,说明,数据的指出,语料库建设,分类方法,排序方法,假设的建立,抽象方法,理论的建立与论证。

(6).选定语言现象的理论描述(上,下).话语行为典型化,配价理论的语法/语义描述,转换语法或格语法的语法/语义描述,标记法,参照语义学,句法学,构词学。

(7).信息/数据库系统.信息系统的类型,数据库系统的基本设计方法,数据的整体结构,数据库系统的实际通讯。

(8).解决问题的方法.generate-and-test法,means-end-analysis法,直接探索法Heuristic Search,模式匹配法,形式逻辑判断法,演绎,归纳,外延法,学习法,规划法。

(9).人-机通讯模型设计.协作原理,自适应系统,用户模型设计,翻译系统,话语行为识别与生成,行动计划生成,论证模型,对话模型设计,答疑系统。

(10).知识表达方法.知识表达形式系统,知识表达语言(FRL, UL-ONE, KRYPTON, OPS5等),自然语言描述形式体系(依存语法,转换语法,合一语法(UNIFICATION GRAMMER),语法规范语言。

(11).句法分析方法和语言生成.编译程序原理,限定关系句法分析,概率关系句法分析,自顶向下分析,自底向上分析,岛分析,图分析,合一语法分析,句法分析生成器,语言生成方法。