

简报

“自然语言多语种文本生成系统”在 上海交通大学研制成功

由上海交通大学计算机科学与工程系盛焕烨教授、姚天 副教授领衔的课题组所承担的“自然语言多语种文本生成理论、技术与应用研究”项目(得到上海市科学技术发展基金和德国大众基金的资助)于1999年5月29日通过了上海市科学技术委员会组织的成果鉴定。鉴定委员会由复旦大学吴立德教授担任主任、东北大学姚天顺教授担任副主任。

鉴定认为:该项成果以包括不同拓扑结构语言(注:即东、西方语言)的多语种和具有实用前景的应用领域为显著特色,在多语种自然语言文本生成技术上达到了国际先进水平,它们的开发和实验成功,在我国自然语言生成研究领域具有重大的现实意义,标志着我国在自然语言多语种文本生成技术方面已达到了一个新的高度。”

从1996年8月该项目立项后,上海交通大学计算机科学与工程系与德国人工智能研究中心进行密切地国际合作,在自然语言多语种文本生成理论和技术方面进行了深入的研究,在研究期间已发表论文19篇(其中包括国际会议论文9篇),待发表2篇。与此同时,课题组又与上海中心气象台和中国人民银行上海分行统计研究处进行合作,选择了两个有实用前景的应用领域——天气预报和银行信息统计进行相应应用系统的开发。这两个应用系统采用内容规划、句子规划和表层生成依次处理的流水线结构做为总体结构。内容规划器采用改进的Schema方法,具有较好的可扩展性和可维护性。句子规划器采用多种语句优化方法和语言资源类映射技术,有效地消除了句间冗余,并就同一谓词提供了多种可供选择的句子结构和短语结构。在语言资源类中采用统一的形式化方法定义不同语种的句子结构,使得不同一机制中可以处理不同语言的句子结构,便于扩展到其它语种。表层生成器采用了以FB-LTAG(基于特征的词汇化树连接方法)为理论基础的汉、英、德句法语义规则和对句法推导树替代、连接、合一操作的组合算法。在构造句法推导树的过程中,提高了生成的有效性,并使得句子的结构组合更具有灵活性。此外,它还具有便于扩展语种和应用领域等特点。系统的试用效果表明,它具有很好的实用性、扩充方便灵活、用户界面友好。它可以生成多城市、海区、雨情、台风、寒潮等天气预报汉、英、德文本以及金融存款、现金统计等汉、英、德文本,基本能满足普通天气预报和银行统计文本的需要。可以根据用户的需求重定义文本结构和修改内容,也可以根据需要增加句型或使句型随机变化。所生成的文本基本符合语言和应用领域的表达习惯。

介绍多语种天气预报文本生成系统(MLWFA)的论文“Multilingual Weather Forecast Generation System”被第九届自然语言生成国际研讨会(加拿大,1998年8月)录用。该研讨会是自然语言生成研究领域最高层语言生成国际研讨会。课题组代表所作的论文报告和系统演示引起与会专家和学者的兴趣和关注。

对自然语言生成理论、技术和应用的研究,将促进我国该领域科学技术的发展。特别是自然语言生成技术的应用将会涉及到社会的各个方面。如果上述两个应用系统一旦投入使用,不但可以大量节省人力、财力、显著提高工作效率,而且还可以改善文本的质量,减少文本的错

误并适应文本实时提供和同时产生多语种文本的需求。从天气预报发布和金融统计报告发表的及时性、文本形式的多样性,以及同时提供多语种文本等方面来看,都具有明显的经济效益和社会效益。

如果针对其它应用领域修改系统的相应部分,可使生成系统用于如民用服务(如保险公司客户信件、股票市场分析报告和医疗文本的生成等)、人机智能化接口(如对话系统、专家系统的解释系统等)、商业系统(如电子商务系统等)、数据库接口(如自然语言数据库查询)和多媒体技术文档等方面。

课题组表示将进一步完善这两个应用系统,向产品化方向发展,并向其它应用领域推广。

上海交通大学计算机科学与工程系 姚天 供稿

联系电话:(021) 62933206 传真:(021) 62933564

电子邮件:yao - tf @cs. sjtu. edu. cn

“机器翻译与计算机语言信息处理国际会议” 圆满成功

中国中文信息学会自然语言处理专业委员会和国家自然科学基金委员会联合主办的《机器翻译与计算机语言信息处理国际会议》(International Conference on Machine Translation & Language Information Processing),于1999年6月26至28日在北京国谊宾馆举行。

会议总结、回顾了当前机器翻译与计算机语言信息处理领域的最新成果、确立了21世纪的发展方向、促进该领域的研究与发展的国际交流与合作,展示了机器翻译与计算机语言信息处理技术和产业的良好前景。

原定会议规模70人,实际与会者145人,这很好地说明了会议受到国内外学者、研究机构 and 企业的广泛关注。

会议论文集收录了76篇(33篇英文论文、43篇中文论文)论文。世界著名机器翻译评论家John Hutchins的论文“The Development and Use of Machine Translation Systems and Computer-based Translation Tools”和我国第一个商品化机器翻译系统——“译星”的主要研制者董振东先生的论文“Bigger Context and Better Understanding - Expectation on Future MT Technology”均使读者受到很大的启发与鼓舞。

会议除了宣读论文外,会议组织者利用晚上的时间召开了专题讨论会,大家对机器翻译的现状和急待解决的关键技术畅所欲言。会议还组织了机器翻译系统演示会,华建集团、日电华、深圳桑夏的精彩演示向与会代表展示了机器翻译及有关信息处理领域的产品和技术的巨大产业华前景。

与会代表在会议上进行了广泛交流和深入研讨,开阔了眼界,探索了合作的途径,我们认为会议达到预期目标,获得了圆满成功。

会议组委会供稿