

作者：张巧玲 来源：科学时报 发布时间：2009-1-21 2:59:59

小字号

中字号

大字号

全新“网络不良信息检测系统”成功研发

通过语义而非关键词区分网页内容

传统的网络检测系统一般只能基于关键词对网络信息进行机械的识别和过滤，如想屏蔽一些邪教网站，往往那些批判邪教信息的网页也会一并被过滤掉。近日，中国科学院声学研究所成功研发出了一套能通过语义理解信息，进而区分出不良信息和批判不良信息的网页内容的全新的“网络不良信息检测系统”。

“我们不是通过关键词匹配的方式进行处理，而是根据语句的意义来判断哪些网页信息需要过滤。”中科院声学研究所研究员、语言知识处理项目负责人张全向《科学时报》介绍，该系统能模拟人浏览网页的方式对网页进行审查，对于不能作出判断的内容系统还能提出警告，供人工判别。

张全说，摒弃互联网上的不良信息已成为整个社会的共识，在社会范围内已形成对不良信息的“围剿”，不良信息正有如“过街老鼠，人人喊打”。正是面向这一迫切的社会需求，中国科学院声学研究所HNC研发团队集多年从事自然语言理解处理的核心技术，才成功研发出了这套具有语义理解特点的网络不良信息检测系统。

据悉，早在20世纪80年代末，中科院声学所就开始探索模拟人脑语言智能的自然语言理解处理模式，创立了概念层次网络理论（HNC），并发展形成了HNC自然语言理解处理技术。HNC的最大特点是能够进入语义深层处理自然语言的语义内容，而不是仅仅利用语言的表层信息进行处理。因此，尽管涉及同样的关键词，但HNC技术能够区分不良信息和对不良信息进行批判的网页内容。经过多年的科研攻关，HNC团队已取得多项成果，形成了自主知识产权的自然语言理解处理技术，为形成满足信息时代要求的各种特定的信息处理技术奠定了坚实基础。

“网络不良信息检测系统的研发就是将这些自然语言理解处理技术与先进的网络技术结合而形成的。”张全介绍，与传统的检测系统相比，该系统能大幅提高对网页内容判断的准确率，大大降低误判率，从而减少了监管人员的工作强度。

目前，这一系统主要针对网络上出现的色情、反动、低俗等不良信息，根据指定的网站自动进行内容下载、检测并给出检测报告。在面向国家公共信息安全的信息过滤、针对企业应用的信息采集等方面，该系统都大有可为。目前，项目组对该系统进行的语料测试已有3万篇以上，测试成功率达85%以上。

张全告诉记者，通过更换知识库，该系统还可广泛应用在热点信息跟踪、舆情分析等领域。

《科学时报》（2009-1-21 A1 要闻）

发E-mail给：

GO

打印 | 评论 | 论坛 | 博客

读后感言：

发表评论

中国一流大学排行榜出炉 首引网络影响力指标
中国南极长城站卫星网络通讯系统建设成功
科学时报年终专稿：网络之手伸向象牙塔
网络与交换技术国家重点实验室开放课题遴选结果公布
东方科技论坛探讨网络环境下人类的集群行为
高校BBS实名注册再起争议 大学生网络隐私难保护
教育部网络时代科技论文快速共享专项研究课题开始...
我国自主研发的无线网络规范成为IEC国际标准

院士评选**2008**年十大科技进展新闻揭晓
引用次数前**20**国家最高被引单篇论文公布
王华宁研究员：警惕**2012**年太阳风暴
07-08学年度国家奖学金获奖者名单公布
评论：国家科技奖的第一获奖者何以都是领导
美《国家地理》杂志预测**09**年**8**大科学突破
大规模引进“海归”将改变国内学术界生态
“千人计划”从**2008**年开始 **5**到**10**年完成