



内海 彰
Akira UTSUMI

研究课题 开发 Web 信息检索支援系统、探明语言理解的认知

关键词

语言信息处理, 语言认知科学, 网站信息检索辅助系统, 信息组织化, 认知科学, 认知心理学, 修辞现象, 语义空间

所属专业	研究生院信息理工学研究科 综合信息学专业
研究成员	内海 彰 副教授
所属学会	日本认知科学会, 人工智能学会, 信息处理学会, 语言处理学会, 日本语用论学会
研究设备	服务器, 工作台, 台式机等各种计算机 30 台, 其它, AV 设备

研究概要

人类传达信息的中枢是「语言」, 该研究的目标是探明语言的认知机制

世间大部分的信息是通过语言来传达的。不论通信技术多么高级也好, 社会网络多么发达也好, 这个事实都是不会改变的。

在通过互联网就能简单获取庞大信息的当今, 索性让计算机来处理语言表达的信息或者探明计算机语言处理背后的人类的语言认知机制变得日益重要。

该研究室以语言的工学处理(语言信息处理)和语言认知的科学研究(语言认知科学)为两大支柱, 从事语言及其背后潜藏的与人的智能相关的学术性研究。

语言信息处理

在这个领域, 该研究室以内容、风格、外观等各种角度为基础, 以网页自动分类为中心, 主要从事网站检索辅助系统的研究开发。

平时用户在互联网上检索某些信息时, 看着罗列了庞大信息的检索结果的画面, 有很多人会产生厌烦的情绪吧。如果能将检索结果再进一步分类, 或者附带某些比较容易看明白的副标题来表示的话, 网络调查就会变得格外容易且能缩短调查的时间。

该研究室正在为此开发网站信息检索辅助系统。

具体来讲就是他们正在设计信息提示系统, 该系统不仅能自动将网页分类, 还能让计算机自动推断网页显示信息的「新旧度」及「可靠性」, 再以这些信息为基础对庞大的信息进行过滤, 显示出用户所需要的信息。此外他们还在研究自动归纳电子化文书的技术。

语言认知科学

在这个领域, 该研究室通过计算机模拟和心理实验来研究如何理解、鉴赏人类使用的比喻、讽刺这类高级修辞法, 尝试着去探明修辞法的生成和认知机制。

并且他们用统计法来处理小说、新闻报道等大量电子化文本中的各种语言出现的方式、频率、多个近似词出现(共现关系)的频率等, 构建将语言意义表现为多维矢量的「语义空间」, 再用构建的这种结构去尝试在计算机上再现语言的认知过程。

接着再详细研讨引起这种语言之间的相互关联性、多个「语言群」的语义空间图是一种什么样的东西, 这样就有望将其结果反馈到语言信息处理研究所代表的计算机语言信息处理技术上, 并推动它的发展。

优势

该研究室是同时从事语言的工学处理和科学研究的学术性研究室。在修辞现象的结构研究中已经获得了世界性的评价

该研究室是同时从事检索辅助系统所代表的计算机语言工学处理和探知人类实际心理活动的认知科学与心理学的比较少见的研究室。

通常情况下, 语言的工学处理属于信息工学领域, 语言认知则属于心理学、语言学这类领域, 它被分为理科和文科, 分别进行完全不同的研究。在一个研究室, 同时研究、实践这两个项目, 在日本还是非常罕见的案例, 因此该研究室是有其独特性的。从这两大方向性来思考事物, 就能扩大研究的视野, 实现更有实效性的研究。

此外在人类认知系统的研究中导入工学的视点, 也是一个新的亮点。再者该研究室在修辞法认知结构的研究中, 已经具有世界级的水平, 受到了高度的评价。

未来展望

研究的目标是实现实用化应用。网页自动分类和检索辅助系统正处在试制品阶段

网页的自动分类和检索辅助系统中, 实现自动分类的程序本身要已经相当完善。该研究室目前的课题是为了在画面上向用户简洁、清晰地显示检索结果, 正在摸索其表现形式及方法。如果能做成这个系统, 他们会将其作为软件安装在用户的系统上, 计划在一两年后完成, 并开始制作试制品。期待感兴趣的企业前来咨询。

除了网站检索之外, 该研究室今后还会将②语言认知科学的研究成果向实用化推进。非常欢迎对此感兴趣的企业发出共同研究的申请。从实用性观点来看, 必须通过成本、普遍的方便性这种比较严格的评价, 因此非常希望能够收到普通企业的各位专业人士提出的有意义的启发。

另外在更具研究性的一面, 该研究室还在广泛地分析、研究修辞技巧、幽默、玩笑等修辞法的生成以及「微笑」这种风格。

将来制造出「知道微笑的计算机」、「写短篇推理的计算机」的日子不久也会来临吧。该研究室怀着这一大梦想, 并承担着这个梦想中的一部分研究内容。

