

我的思考和回答

宗成庆

中国科学院自动化研究所
模式识别国家重点实验室

北京市海淀区中关村东路95号
邮编：100190



主页：<http://www.nlpr.ia.ac.cn/cip/cqzong.htm>
邮件：cqzong@nlpr.ia.ac.cn

1. 定义和理解

扼要阐述新时期下中文信息处理的定义和理解：

研究中国语言（包括中国少数民族语言）文字和语音的特点及规律，建设相关的资源和制定相关的标准，提出有效的方法和模型，面向各类应用，尤其是大规模网络和移动通讯环境下的用户实际需求，开发高性能实用系统，并建立适用于中文语音语言信息处理的理论体系。

- ◆3个层面：资源和标准、模型和方法、系统和技术
- ◆2个特色：中文（语言）、网络与通讯（时代）
- ◆1个体系：中文信息处理的理论体系

2. 着眼点与着力点

下一个三十年中文信息处理的主要着眼点和着力点应该在哪里？

◆ 汉语语义理解的理论和方法

— 从自然语言处理的角度如何建立汉语的语法体系、语义表示方法和计算模型？

◆ 少数民族语言的基础资源建设和关键技术研究

— 无论从语言文化保护的角度，还是从自然语言处理的角度，太多的问题需要进一步加大研究和开发的力度。

3. 可能的创新点

中文信息处理领域的重要原始创新可能在哪里？

◆ 汉语语义理解的理论和方法

— 从自然语言处理的角度建立汉语的语法体系、语义表示和计算方法

◆ 多语言、多媒体信息理解的融合技术

— 文本、音视频、图像等多模态信息并存，英语、汉语、阿拉伯语等多语言信息并存……

方法：大规模数据计算+机器学习

4. 产业化方向

中文信息处理相关产业发展的重要方向或方面，
及“产学研”战略模式

◆ 重要方向：

— 面向网络和移动通讯用户的个性化信息服务，
包括翻译、检索、抽取、问答、推荐等

◆ “产学研”模式：

— ?

5. “核高基”

如何有效推动中文信息处理进入“核高基”之类的国家重大科研计划？

在2006年国务院发布的《国家中长期科学和技术发展规划纲要（2006-2020年）》中，“核高基”是“核心电子器件、高端通用芯片及基础软件产品”的简称。

操作系统、数据库 和 中间件的统称

红旗Linux

神舟通用、达梦、金仓

可以媲美国际品牌。



5. “核高基”

◆ 锁定任务和目标

- 语言的问题 (仅有汉语是不够的)
- 技术梳理和目标定位
- 与相关学科专家的联合

◆ 组织起步

- 组织预研
- 与相关学科专家的联合

◆ 宣传和推介

- 让相关专家和主管部门了解我们的工作

6. 几点具体建议

- **有计划、有组织地申办国际学术活动，提升CIPS的国际影响力**
 - 组织申办 ACL、EMNLP、IJCNLP 等高水平国际学术会议
 - 举办 Summer school
 - 组织相关评测，如 SIGHAN 等
- **在保持学会尊严的前提下，加强与兄弟学会的合作**
 - 通过各种活动，提高学会在国内的影响
- **提升《中文信息学报》的地位**
 - 争取 EI 索引
 - 改进审稿方式



6. 几点具体建议

我自己的经验：

ScholarOne Manuscripts

IEEE Intelligent Systems, Associate Editor

ACM Trans. on ALIP, Associate Editor

Inter. J. of Comp. Process of Lang. (IJCPOL), Associate Editor

Machine Translation, Editor

J. of Comp. Sci. & Tech. (JCST), Editor

《自动化学报》编委

《智能系统学报》创刊仅6年，被收录为中文核心期刊。



6. 几点具体建议

■ 关于学科竞争力

同行的合作与“宽容”是学科发展的基本要求。

