



中国中文信息学会2012战略研讨会  
2012年4月13日-15日 江西婺源



# 关于中文信息处理事业未来发展的 几点想法

赵铁军  
哈尔滨工业大学

# 提纲

Web为平台

理解为核心

应用为目标

# Web为平台

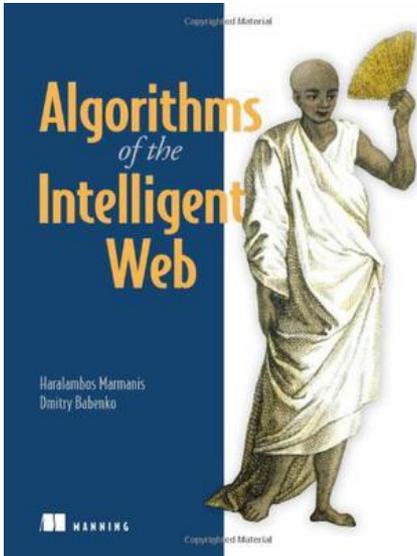
**Web=**当今世界最大的信息平台、交流平台、知识平台

**Web**平台支撑中文信息处理技术的研究和应用; 终端: 固定设备+移动设备(移动互联网)

**Intelligent Web**具有广阔的空间, 一款有用的、显示智能特点的程序模块均可获得成功

目前正在应用和正在研发的技术的目标: 朝着更加智能的服务努力

# Intelligent Web: Google搜索结果



powerful techniques  
to process information  
intelligently and offer  
features based on  
patterns and relationships  
in data

Web 3.0: The Intelligent  
Web

**Tailor made search**  
**Contextual Search**  
**Personalized Search**  
**Interoperability**  
**Transforming the Web  
into One Big Database**  
**Evolution of 3D Web**



开源的内容管理系统  
(CMS) 平台  
Drupal有一个优秀的模  
块化结构，提供了许多模  
块，包括短消息、个性化  
书签、网站管理、Blog、  
日记、电子商务、电子  
出版、留言簿、Job、  
网上电影院、论坛、投票  
等模块

- Google搜索前3结果

# 理解为核心

一定意义上，智能=自然语言理解  
理解有多种层面：从浅层到深层

语义是理解自然语言的基石之一

中文语义理论体系探索 + 应用需求牵引：好的形式满足应用需要

以往的词语、句法等方面的中文信息处理技术成果应充分利用和结合

# 语义计算: 从具体任务出发

中文尚语义缺乏体系，存在很多模糊的内容  
定义明确的任務，为具体任务的完成建立一套可行的中文语义计算表示框架和有效算法  
实例: 理解式中文文摘 or 关于复述的某个展现形式 (理解的4大表现: 翻译、问答、复述、文摘)

语言语义只是一种形式; 探索多媒体语义、数据语义.....

# 应用为目标

一切自然语言理解技术成果最终都要为应用服务，否则经不起历史考验

Web平台(及传感器网络)的一大特色:海量数据;要求云计算处理平台、并行处理等

自然语言理解算法的现实要求:海量数据下有效和高效计算

应用驱动研究 + 核心科学问题驱动研究

# 应用的多模式

## 应用多种多样

政府投资：战略性、大规模、大团队、大成果...

企业投资：产品与应用解决方案、赚钱或开拓市场、公益事业...

团队投入：前瞻研究、扩大影响、技术储备、未来项目...

吸引眼球之妙想 + 耐得住寂寞之深入研究

# 学科的交叉

## 与机器学习研究交叉

不仅停留在应用层面，探索自然语言理解应用中学习算法的特性

## 与智能人机接口技术交叉

多媒体、虚拟现实、数据库、知识工程...在具体任务或系统中融合

## 与脑科学交叉

但其成果如何在自然语言理解中应用必须探索

# 当务之急

急于扩大影响？有影响力的成果展现  
用户喜闻乐见 or 政府战略需求？

成功离不开多年的潜心研究：如IBM“沃森”  
系统参加Jeopardy节目(100人年)

立足于现有成果和技术，寻求一种突破式  
的结果：怎样协调？