



2012年中文信息处理战略研讨会

内外兼修 百花齐放 促进中文信息处理事业发展

黄萱菁

xjhuang@fudan.edu.cn

Watson和Siri

➤ 问答系统 - Watson

- 2011年2月，由IBM和美国德克萨斯大学联合研制的超级电脑Watson在美国最受欢迎的智力竞猜电视节目《危险边缘》中击败该节目历史上两位最成功的选手，成为《危险边缘》节目新的王者。



➤ 个人助理 - SIRI

- 2011年10月，苹果公司发布了iPhone 4S，推出个人智能助理系统SIRI。SIRI使用自然语言处理技术，用户可以使用自然的对话与手机进行交互，完成搜索数据、查询天气、设置手机日历、设置闹铃等服务。

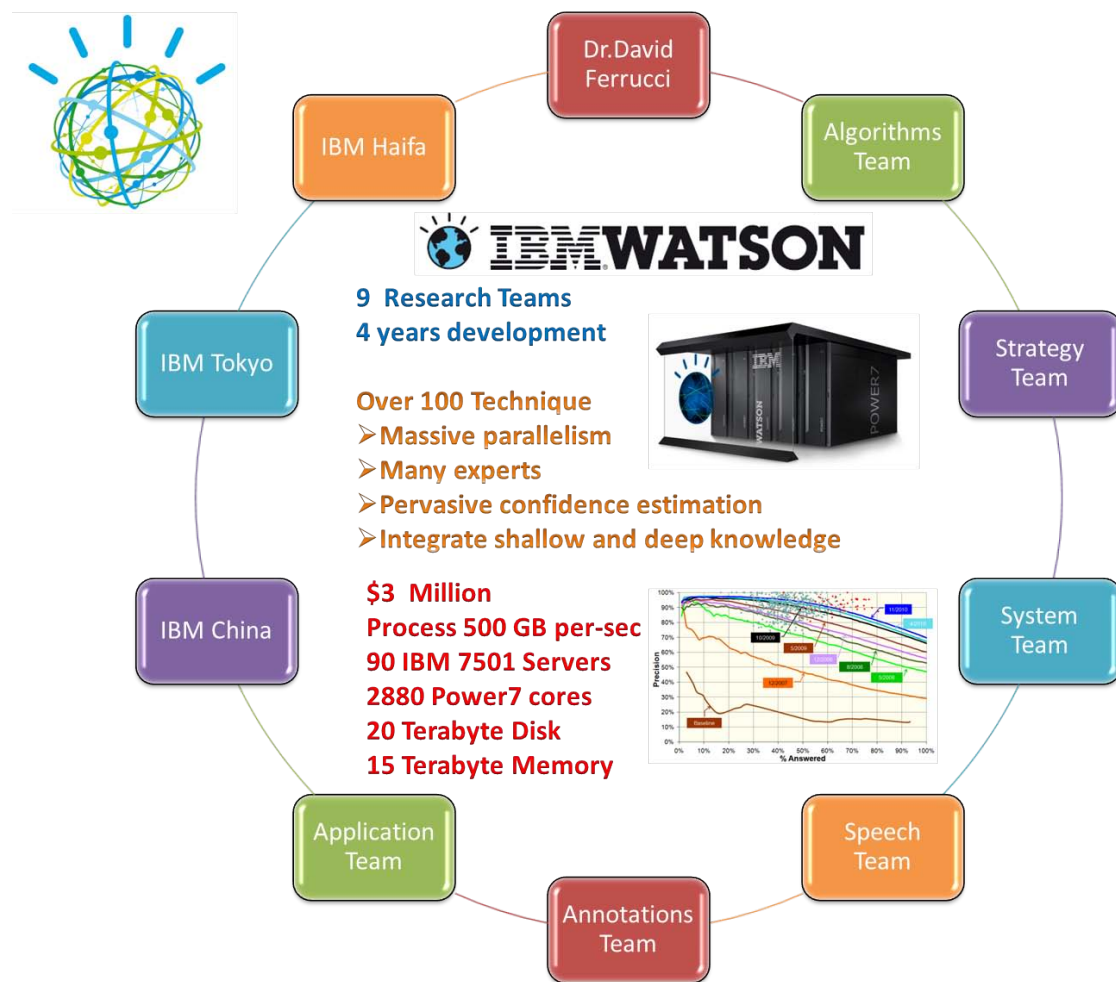


学术影响力：Invited Talks (2011)

- ACL: Building **Watson** - An Overview of the DeepQA Project
- AAAI: Building **Watson**: An Overview of DeepQA for the Jeopardy! Challenge
- ICML: Building **Watson** - An Overview of the DeepQA Project
- NTCIR: Building **Watson**: A Grand Challenge in Automatic NL Question Answering



成功的背后 -- Watson



成功的背后 -- Siri



Personal Assistant

Understand you

- Speech to Text
- Text To Speech
- Dialog Flow

Do things for you

- Book table
- Play Music
- Read Message
- Traffic Info on Map

Search

- Local search
- Answer question

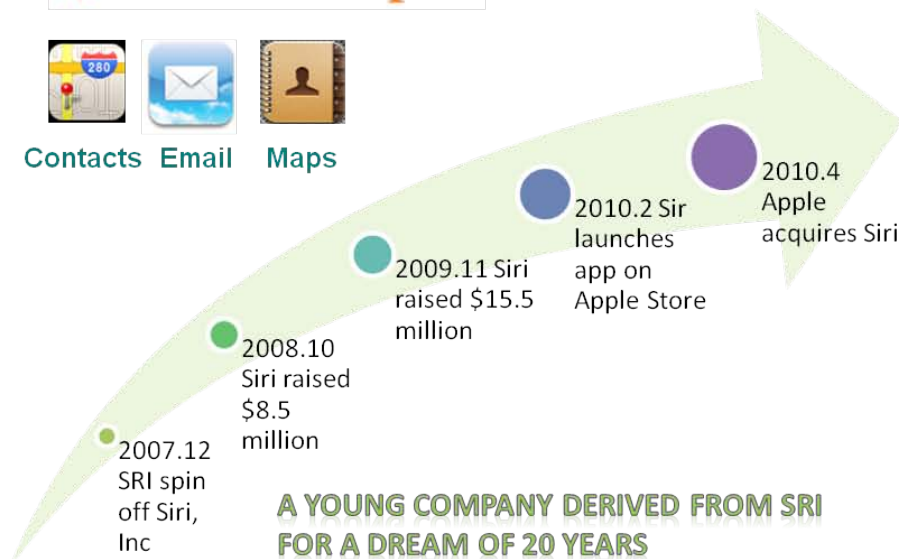
NUMEROUS PARTNERS
Google bing YAHOO!

yelp OpenTable®

WolframAlpha



Contacts Email Maps



NLP的未来方向--1989年

- **Knowledge acquisition** from natural language text and from interaction with human beings
- **Interaction with multiple underlying systems** to give NLP the utility and flexibility
- **Partial understanding** from multi-sentences language, or fragments of language
- Future Directions in Natural Language Processing: The Bolt Beranek and Newman Natural Language Symposium—BBN, Cambridge, Massachusetts, from 29 November to 1 December 1989



NLP的未来方向--2004年

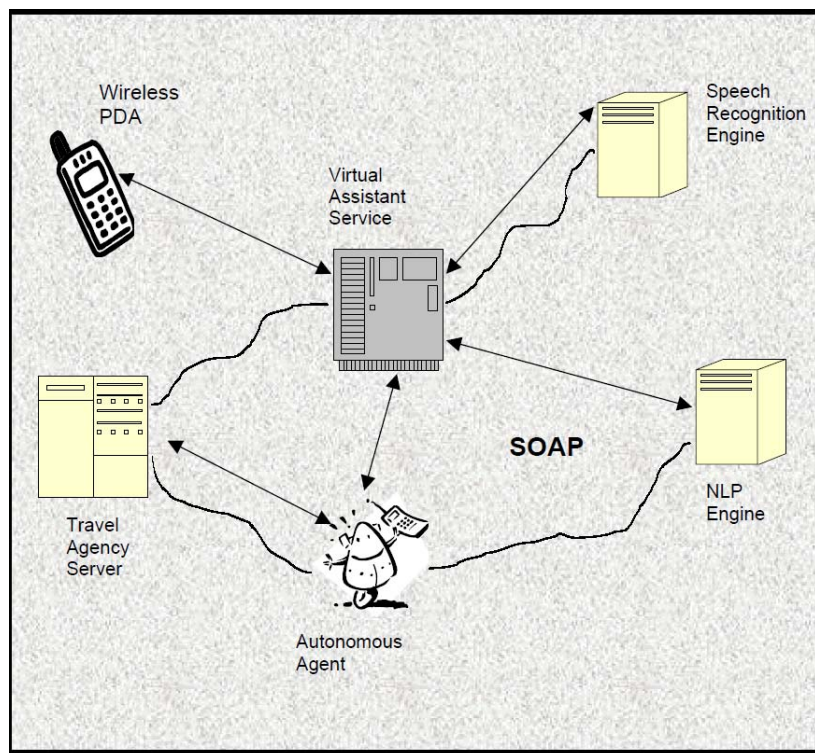


Figure 3. Consumer use of NLP via wireless PDA

- Consumer use of NLP via wireless PDA
- Conversational systems
- Systems where a computer would be able to read a book
- Microsoft MindNet – an extensive database and algorithms
- Chatterbots

➤ Natural Language Processing: Current state and future directions, Ranjit Bose, International Journal of the Computer, the Internet and Management, 12-1

NLP的未来方向--2006年

- Machine Reading
 - natural language understanding in an open-ended domain
- Socially-Aware Language Understanding
 - the recognition of communicative goals, of how a conversation proceeds and varies in interesting ways depending on the conversants and the situation, and of how time and place affect interpretation.
- Annotation Science



何谓“好”的长期研究

- **Understandable**. The goal should be simple to state
- **Challenging**. It should not be obvious how to achieve the goal
- **Useful**. If the goal is achieved, the results should be clearly useful to many people
- **Testable**. Solutions to the goal should have a simple test so that one can measure progress and one can tell when the goal is achieved
- **Incremental**. It is very desirable that the goal has intermediate milestones so that progress can be measured along the way
- What makes a good long-range research goal or a grand challenge?
----- Jim Gray

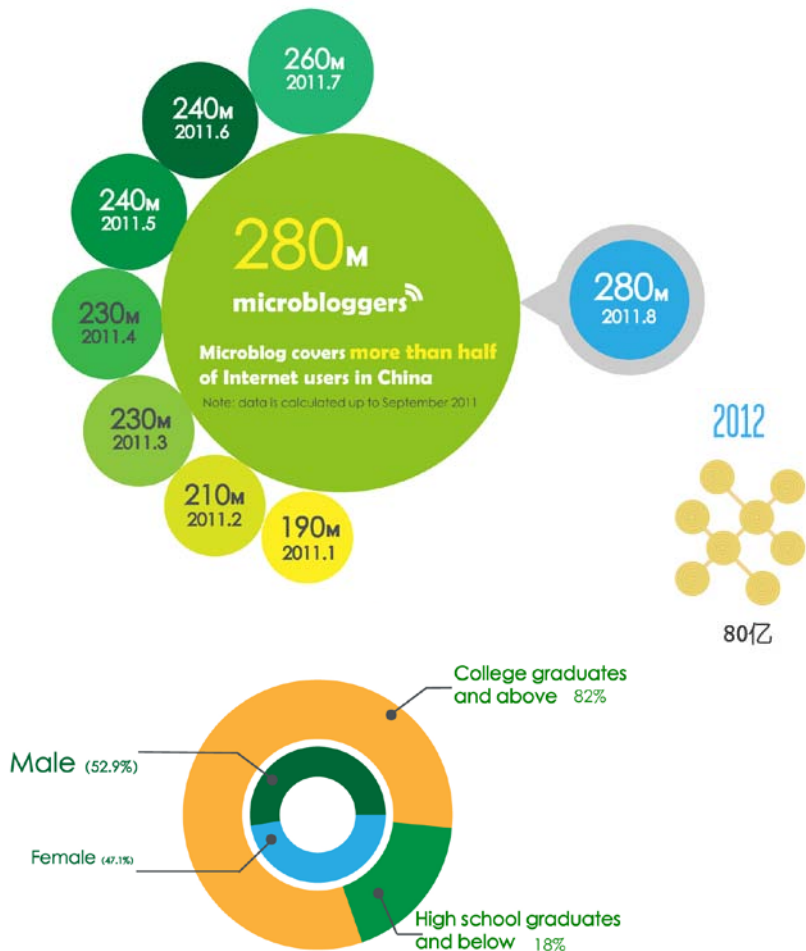


内功 – 解决核心问题

- Understanding
 - Discourse, semantic, pragmatic
- Generation
 - Translation, summarization,...
- Cognition + linguistic
- Multi-lingual, multi-modal, interactive
- Knowledge acquisition
- Efficiency: learning + decoding

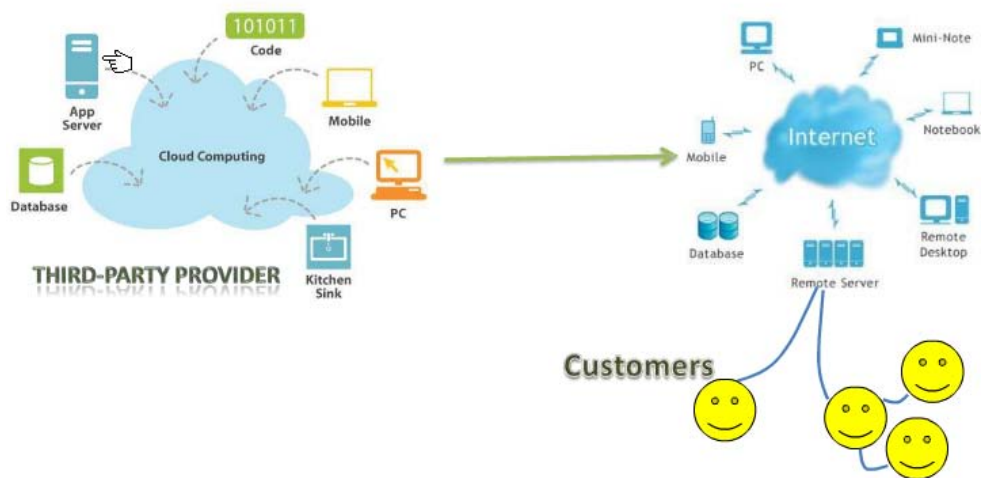


外功 – 关注用户需求

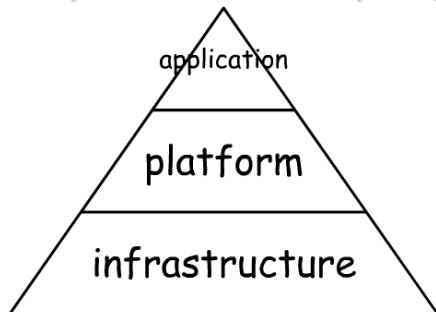


- UGC (User Generated content), TU/ SIGIR
- Web Search Queries as a Corpus TU/ ACL
- Web Retrieval: The Role of Users, TU/SIGIR
 - User behavior (Usage data analysis and metrics)
 - Assist user interaction
- Practical Online Retrieval Evaluation, TU/ SIGIR
- Crowdsourcing for IR, WS/SIGIR

外功 -- 云+端



Three layers to cloud computing...



➤ Computing power consumption ❌

➤ Storage Limitation ❌

➤ Application

➤ ASR

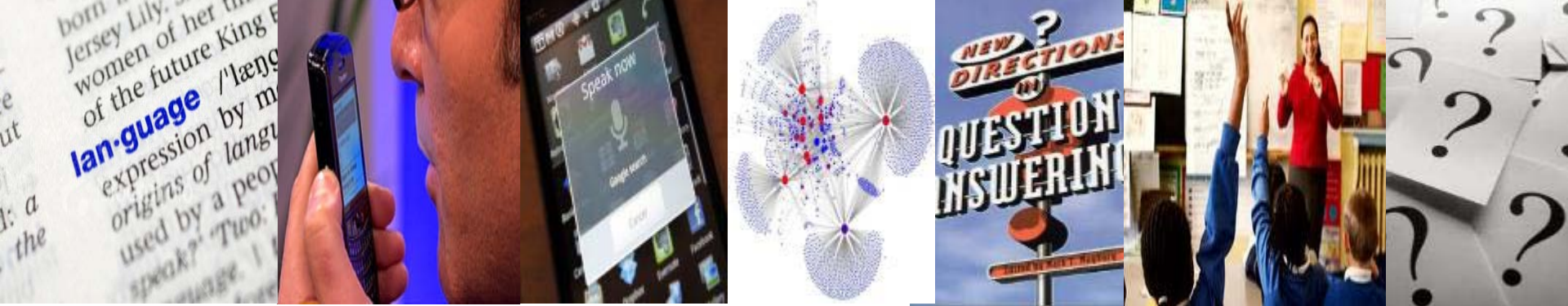
➤ MT

➤ ?

结语

- 中文信息处理
 - 实现人与计算机之间用中文进行有效通信的各种理论和方法
 - 主要内容包括中文语音和语言的编码、识别、理解、翻译和生成
- 新的历史条件下的中文信息处理
 - 结合计算机科学、人工智能、语言学、认知科学、统计学等相关学科
 - 解决中文信息处理的核心科学问题
 - 服务国家重大需求，促进产业发展
 - 结合互联网发展，促进信息社会进步和全球化进程
- 未来突破方向
 - 机器翻译、人机对话、机器阅读、……





谢谢

Q&A