

# 自然语言处理技术在下一代搜索引擎中的关键作用

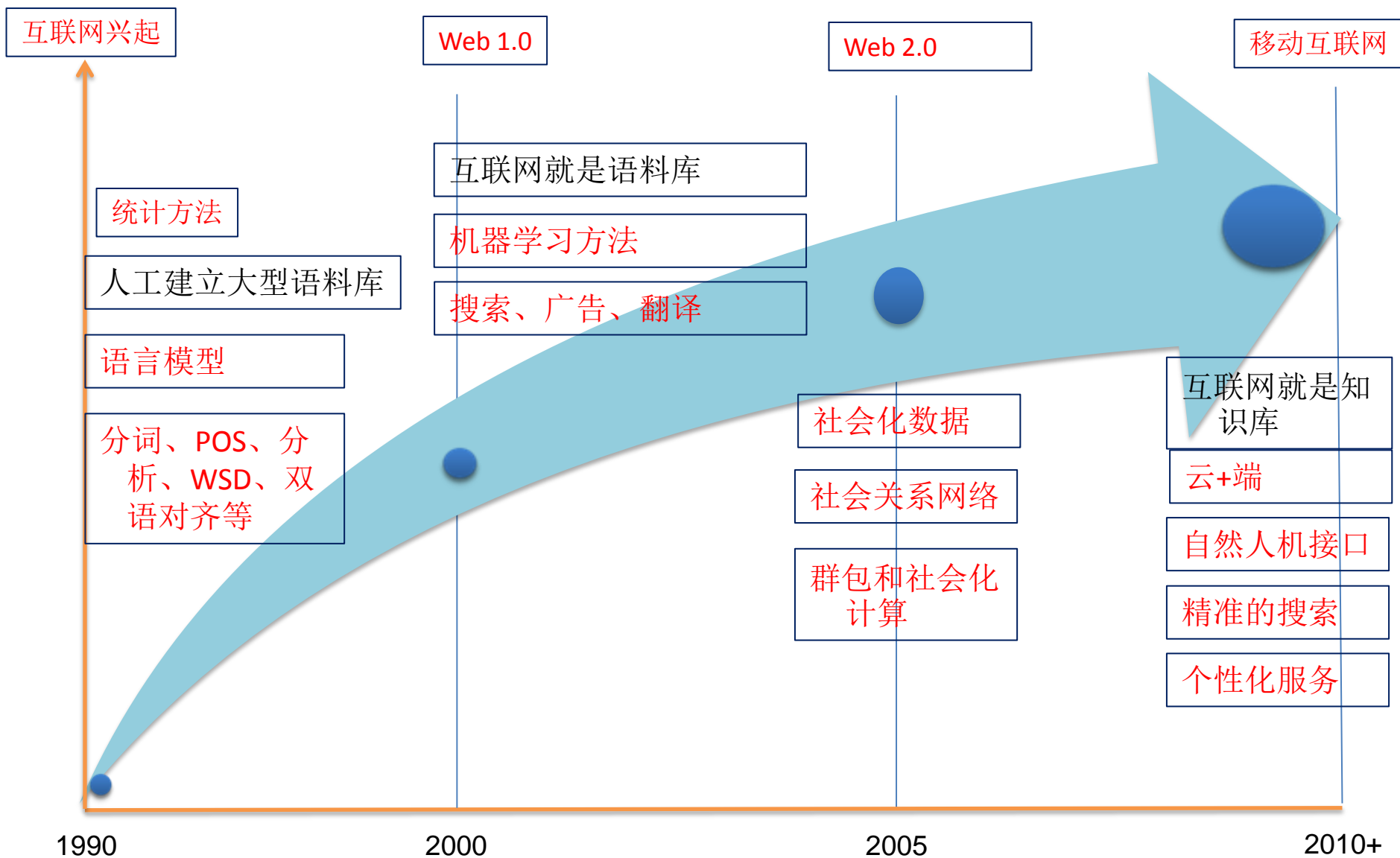
周明

微软亚洲研究院

2012年4月13日婺源

中文信息处理学会战略研讨会

# 自然语言处理技术的二十年的演变和未来热点



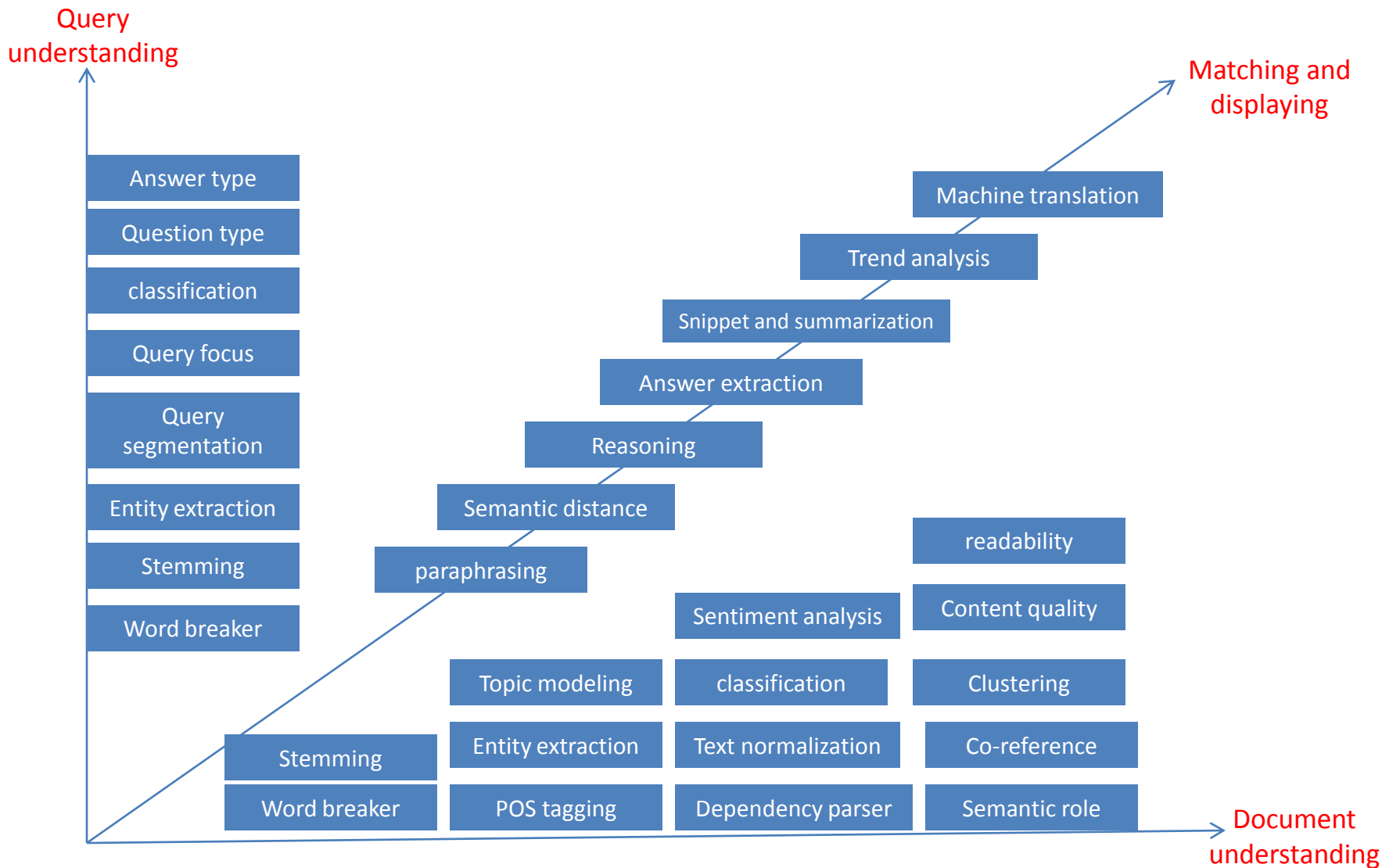
# 搜索引擎的八大趋势

- 强大的信息的整合能力
- 实时搜索能力
- 社会关系网络和搜索的互相影响
- 搜索结果需要更加精准
- 搜索的移动化
- 由搜到用的快速迁移
- 灵活的人机接口
- 个性化

# 重要的研究方向

- 用户意愿的理解（利用上下文搜索意图和社会关系网络作为背景）和个性化的搜索和推荐
- 实体和知识的获取、推理在搜索中的应用
- 自然语言和语音的紧密结合的伟大创新
  - 手机人机接口和语音翻译系统
- 社会关系网络的文本挖掘和新型的搜索
- 新一代的问答系统和精准的搜索

# 自然语言对搜索的渐次展开的作用



# 自然语言处理的研究策略（NLP2.0）

- 把互联网当成一个研究平台，在其上进行数据获取、模型训练、实施、与用户互动。
- 把自然语言研究和用户需求和市场因素适度地联络起来，争取互动。
- 呼吁快速实施。系统在雏形阶段就需考虑尽快投放使用以便根据用户反馈改善功能并且完备界面设计。
- 充分利用社区问答、维基百科、用户标签、众包等加强训练和测试。
- 要加强多学科的交叉，数据挖掘，机器学习,搜索、广告、人机交互、互联网经济学和社会学和自然语言处理充分交融。
- 建立大规模的数据处理平台以支持网络挖掘、信息抽取、模型训练和系统实施。
- **Case study:**
  - 英库词典和翻译[www.engkoo.com](http://www.engkoo.com)
  - 微软对联<http://duilian.msra.cn>

# 详细观点请参考如下文章

周明，新一代搜索引擎的研发战略，中文信息学报第25卷第六期。

摘要：

理想的搜索引擎不仅应该快捷地帮助找到所需要的信息，还应该像银行的窗口一样提供个性化的服务。更进一步，它还应该象专家一样辅助用户做出决策并快速完成各种任务。当前，互联网搜索领域酿着前所未有的新机遇。庞大的网民数量为搜索的发展提供了广阔的空间。社会关系网络和移动互联网推动搜索引擎的技术将会发生重大的改变。自然语言处理对用户意图的理解和对文本的理解能力的提高，将使得搜索的准确性也将持续改善。然而搜索引擎的成功是多方面因素共同作用的结果，需要技术和战略的巧妙配合。本文将深入分析目前搜索引擎技术发展的趋势，指出应该关注的若干重要领域，并且探讨了搜索引擎的研发战略。