

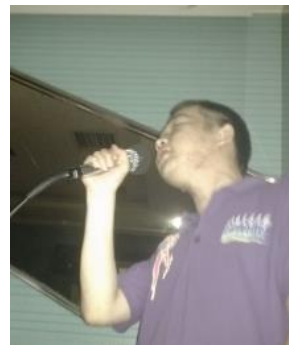


# NLP团队季度总结

语音和语言技术中心 信研院 清华大学  
王晓曦 邢超 骆天一 袁斌 张东旭 马习 曹力 刘荣



# 团队成员

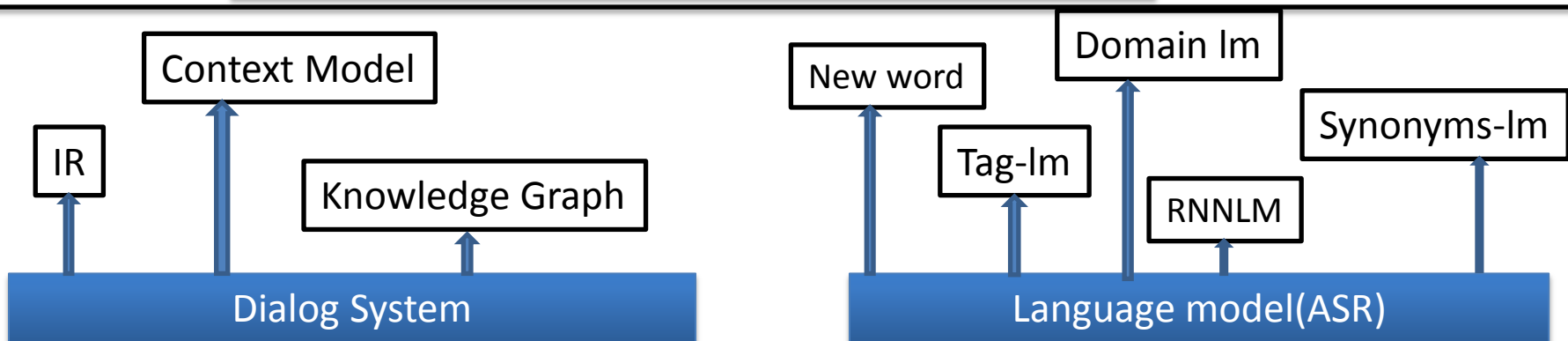


# 目录

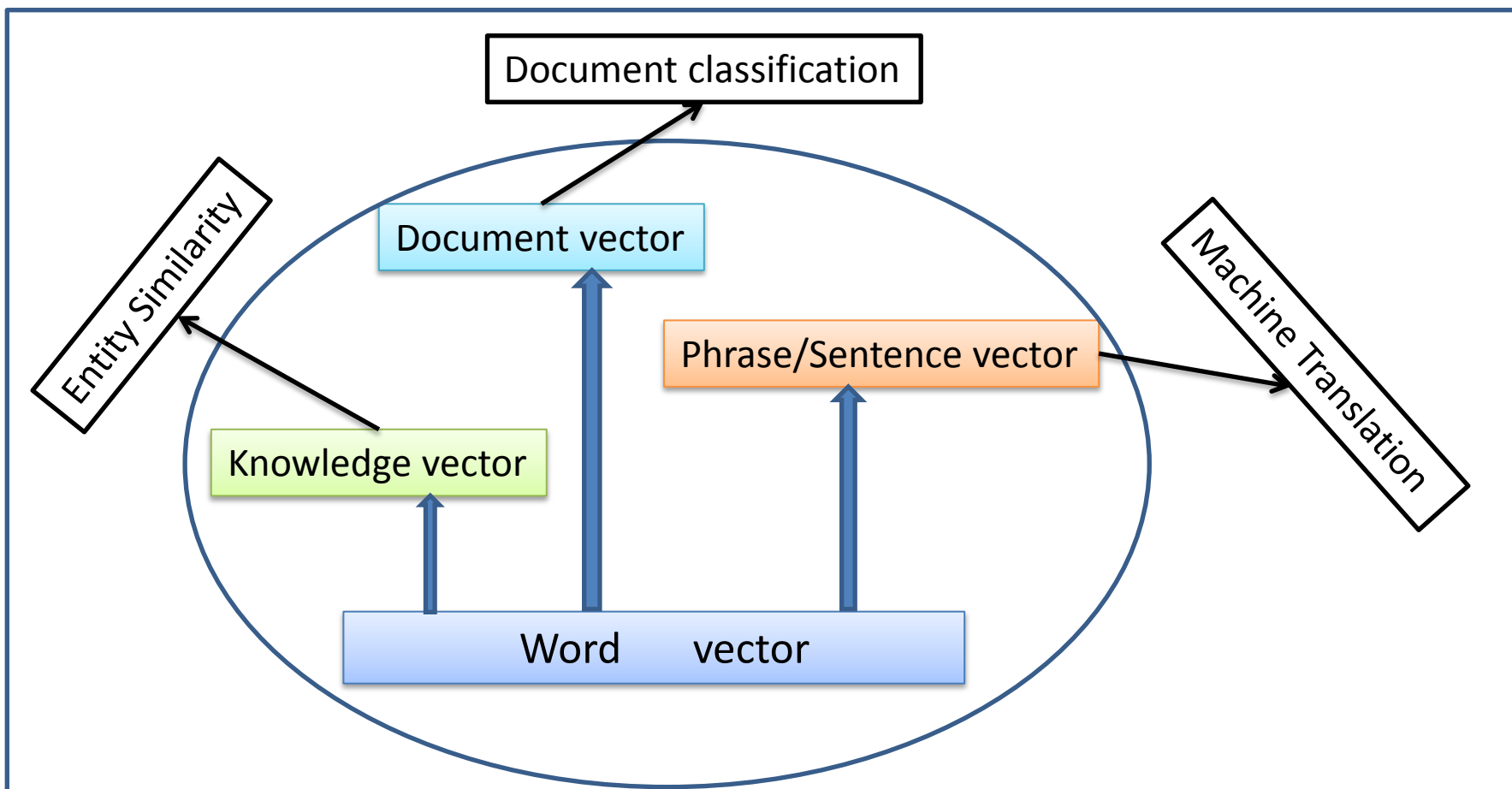
- 目前工作
  - 整体框架
  - 成果贡献
  - 学术氛围
- 目标

# 领域应用

## 系统构建领域

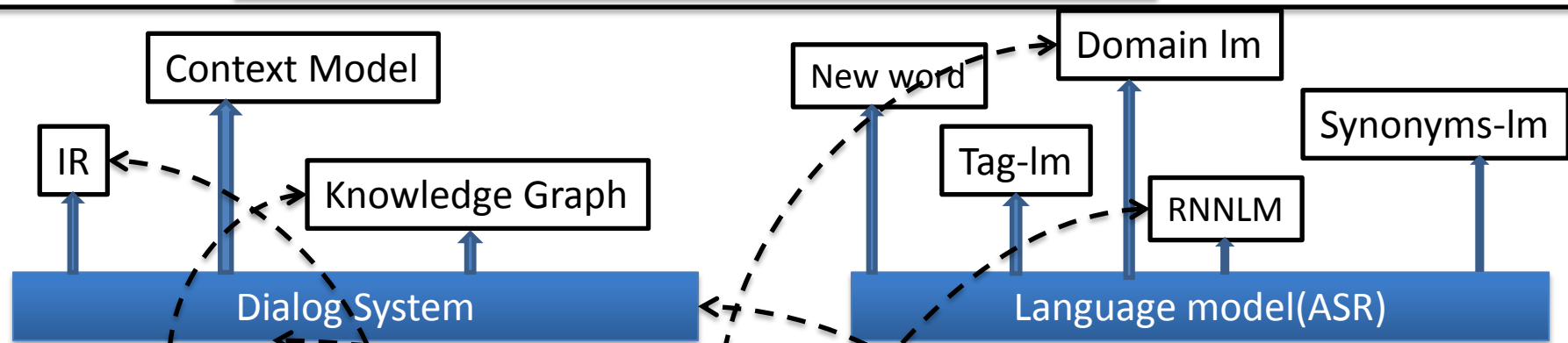


## 核心研究领域

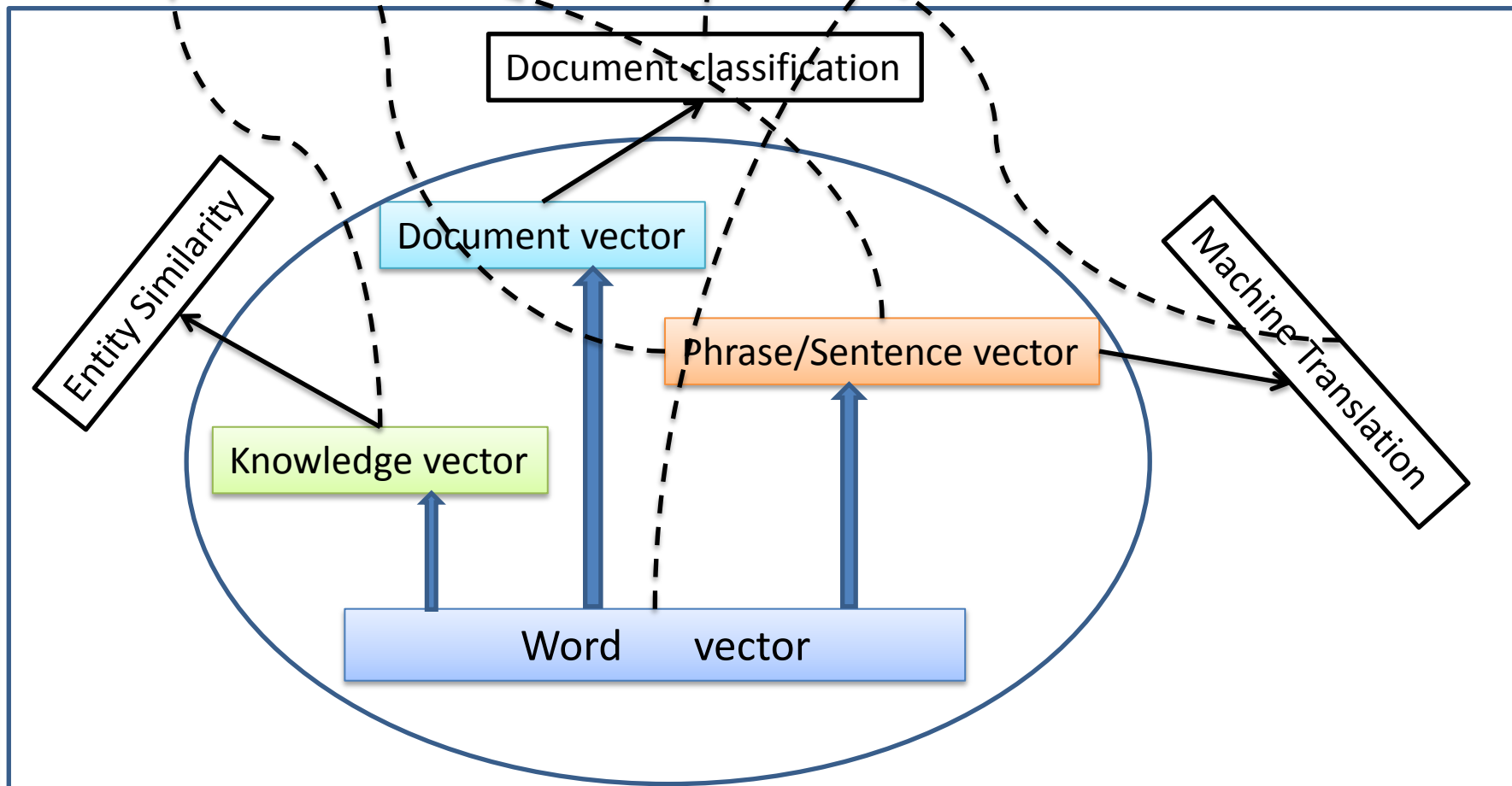


# 领域应用

## 系统构建领域



## 核心研究领域



# 成果贡献

- 论文

Chao Xing, Dong Wang, Xuwei Zhang, Chao Liu, document classification based on i-vector distributions, AP

Rong Liu, Dong Wang, Chao Xing, Text classification based on word vectors, 2014, Singapore (in Chinese)

Semi-Dynamic Graph Embedding for Large Scale Language Model Adaptation(yb,wxx)

- 专利

刘荣,王东,郑方,"基于分布式神经网络的语言模型训练方法及其系统",公布日期: 2014.05.21

王东, 王晓曦, 赵芳, 刘荣, 游世学, 一种信息查询方法和系统, 公布日期: 2014.12.10

王东, 刘荣, 游世学, 杜新凯, 基于问答系统的信息匹配的方法和装置 (提交)

王东, 刘超, 刘荣, 郑方, 一种基于稀疏矩阵的快速语音识别方法 (提交)

王东, 游世学, 刘荣, 基于字的神经网络语言模型训练方法及其系统 (提交)

基于多层贝叶斯的在线用户自学习的问答系统 (汇联)

# 成果贡献

- 问答系统（汇联公司）

1. 检索
2. 多场景识别
3. 拼写检查
4. 上下文处理
5. 在线学习

- 语言模型（捷通华声）

1. 工具（tag-lm, 新词爬取与解析，领域语料爬取）
2. 语言模型（v1.0, v2.0, v2.5的发布）：
  - 分布式训练LM的方法
  - 对于pruned语料做kndiscount的方法
  - 还有各种优化LM的实验
3. RNNLM的探索

# 学术氛围

- 周三讨论会

- 1 ACL 会议论文
- 2 NIPS 会议论文



# 目标

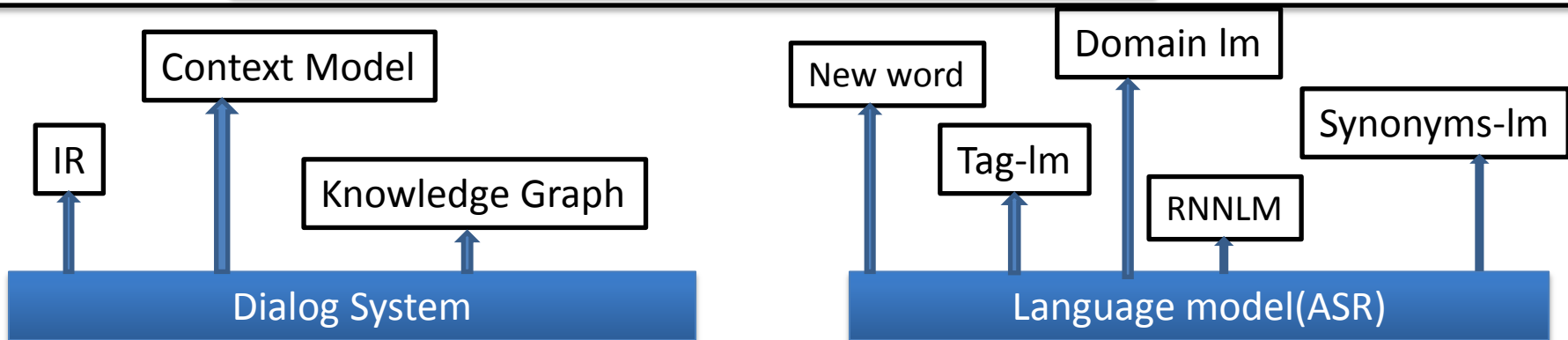
学术 + 氛围 = 成果

高效 + 自由 = 快乐

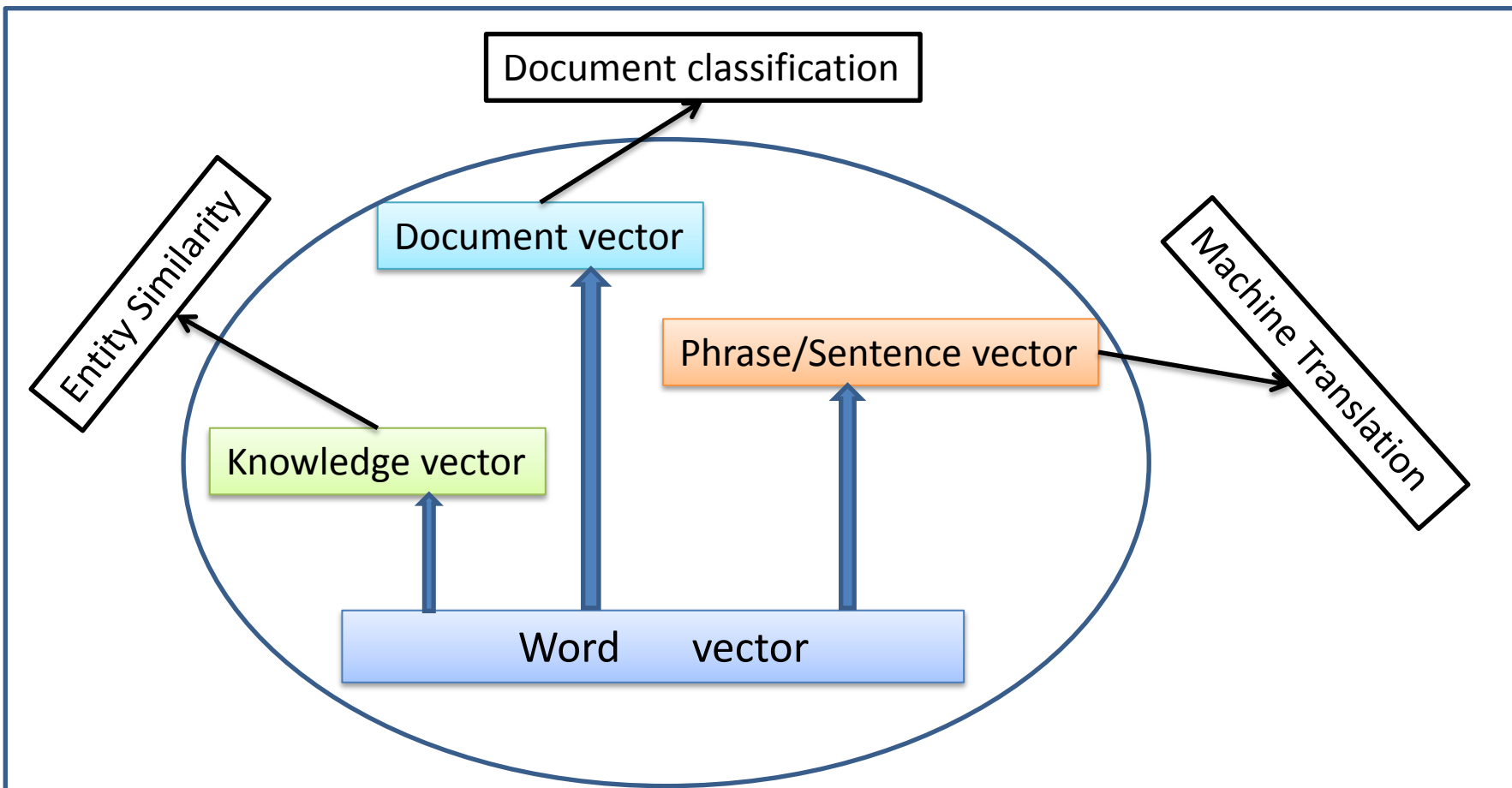
10篇论文+3篇专利

# 领域应用

## 系统构建领域



## 核心研究领域



Thanks!