

# 智能问答中的问句理解

Lin Liu (刘琳)

Yahoo! Senior Engineering Manager



# CNUTCon 2017

## 全球运维技术大会

上海·光大会展中心大酒店 | 2017.9.10-11

智能时代的新运维

大数据运维  
安全  
SRE  
DevOps  
Kubernetes  
Serverless  
游戏运维  
AI Ops  
智能化运维  
基础架构  
监控  
互联网金融



# StuQ

斯达克学院

实践驱动的IT教育



**斯达克学院(StuQ)**，极客邦旗下实践驱动的IT教育平台。通过线下和线上多种形式的综合学习解决方案，帮助IT从业者和研发团队提升技能水平。



10大职业技术领域课程

<http://www.stuq.org>

# SPEAKER INTRODUCE

刘琳      Yahoo! 高级研发经理  
liulinthu@ymail.com

- Yahoo! 雅虎                      2009 至今
  - 负责自然语言理解平台八年
  - 北京研发中心第一个工程师
- 搜狐 / 搜狗                      2006 - 2009
  - 网页信息分析提取模块负责人
- 清华大学计算机系              1999 - 2006



# TABLE OF CONTENTS 大纲

---

- 智能问答应用场景
- 搜索引擎 vs. 智能问答
- 搜索查询理解典型架构
- 从查询理解到问句理解
- 问题与挑战

# TABLE OF CONTENTS 大纲

---

- 智能问答应用场景
- 搜索引擎 vs. 智能问答
- 搜索查询理解典型算法
- 从查询理解到问句理解
- 问题与挑战

# 智能问答应用场景 (1)

- 开放域智能问答
  - 不限定知识领域
  - 混合情感交互问答
- 典型应用
  - 智能个人助理

# 智能问答应用场景 (2)

- 限定域被动式智能问答
  - 机器回答限定知识领域的问题
  - 甚至限制在有限的问题集合中
- 典型应用
  - 用于销售、答疑、咨询的在线智能客服

# 智能问答应用场景 (3)

- 限定域引导式智能问答
  - 机器向用户提问为主
  - 通过分析用户的回答完成全局的逻辑计算
- 典型应用
  - 智能专家系统
  - 问答式表单填写分析

# TABLE OF CONTENTS 大纲

---

- 智能问答应用场景
- 搜索引擎 vs. 智能问答
- 搜索查询理解典型算法
- 从查询理解到问句理解
- 问题与挑战

# 搜索引擎 vs. 智能问答 (1) - 输入



- 通常为关键词罗列形式
- 通常无上下文



- 相对完整的自然提问
- 会话上下文的指代和引用
- 非提问形式的上下文表述



# 搜索引擎 vs. 智能问答 (2) - 输出



- 网页、文档列表
- 部分有结构化的图文结果
- 部分有直接的答案
- 要求关键词匹配
- 通常不要求个性化



- 陈述式答案句子
- 直接唯一的答案
- 可能反馈要求问题细化
- 要求语义和意图的匹配
- 通常要考虑用户偏好和上下文

# 搜索引擎 vs. 智能问答 (3) - 反馈



- 点击、停留等交互行为
- 改写查询
- 中止或放弃搜索



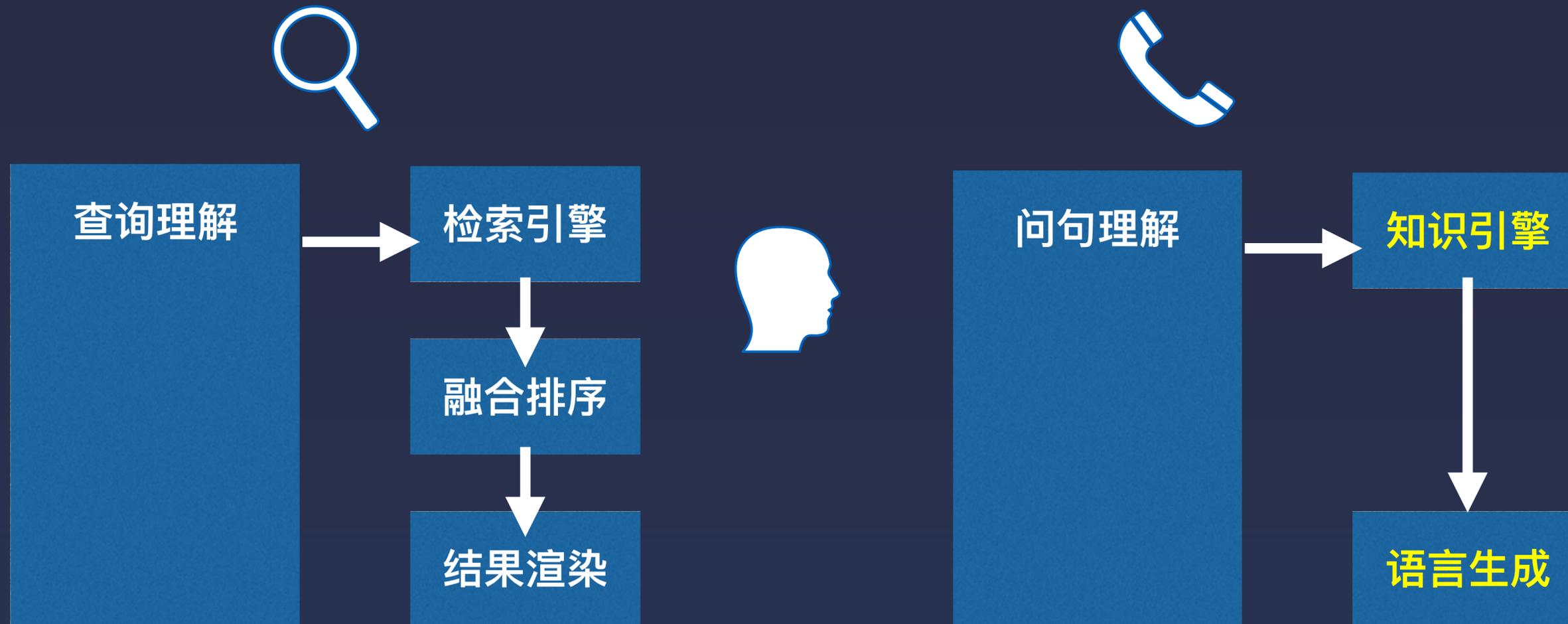
- 对话修正机器理解
- 中止或者放弃对话



# 搜索引擎 vs. 智能问答 (4)



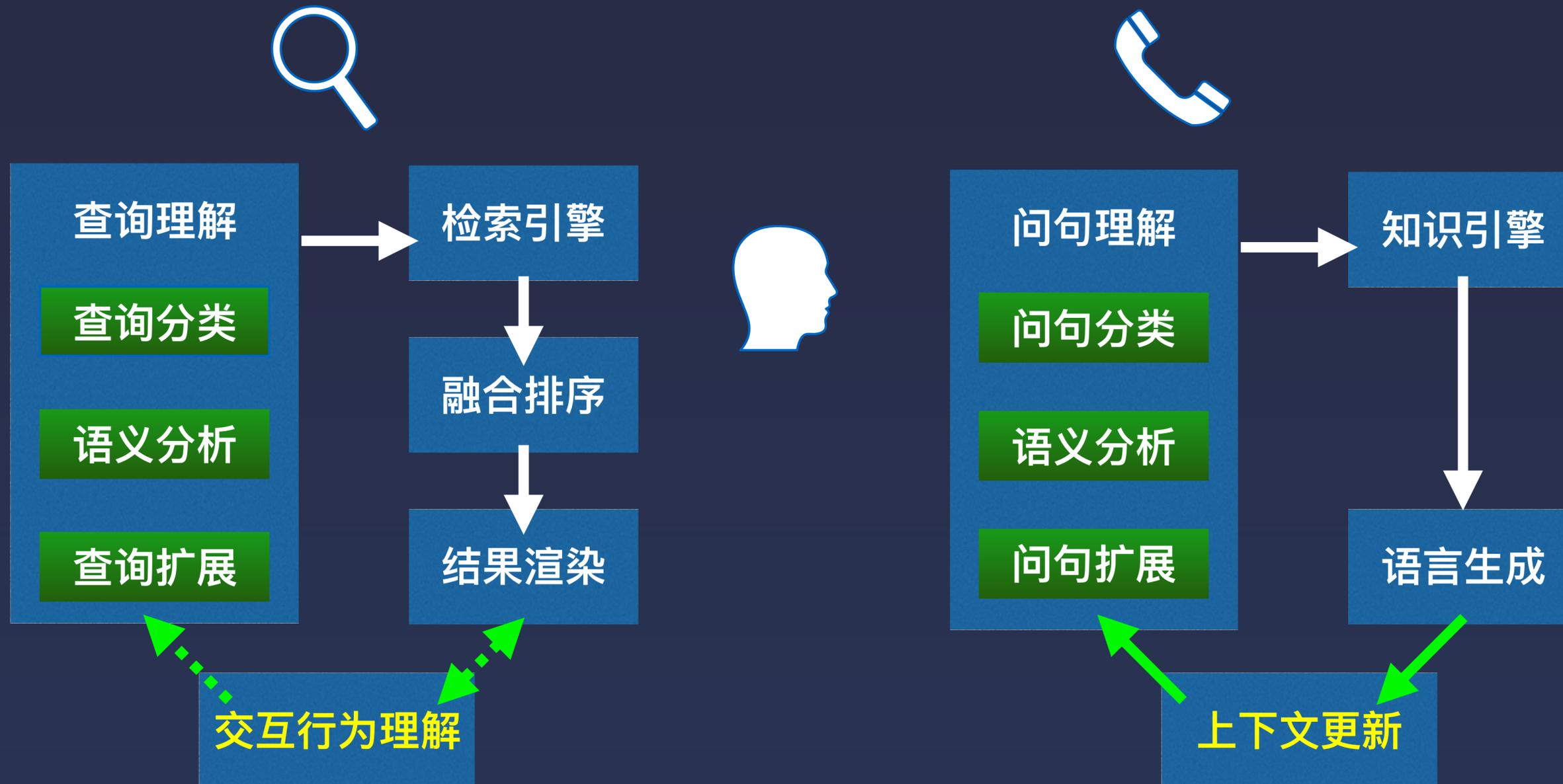
# 搜索引擎 vs. 智能问答 (5)



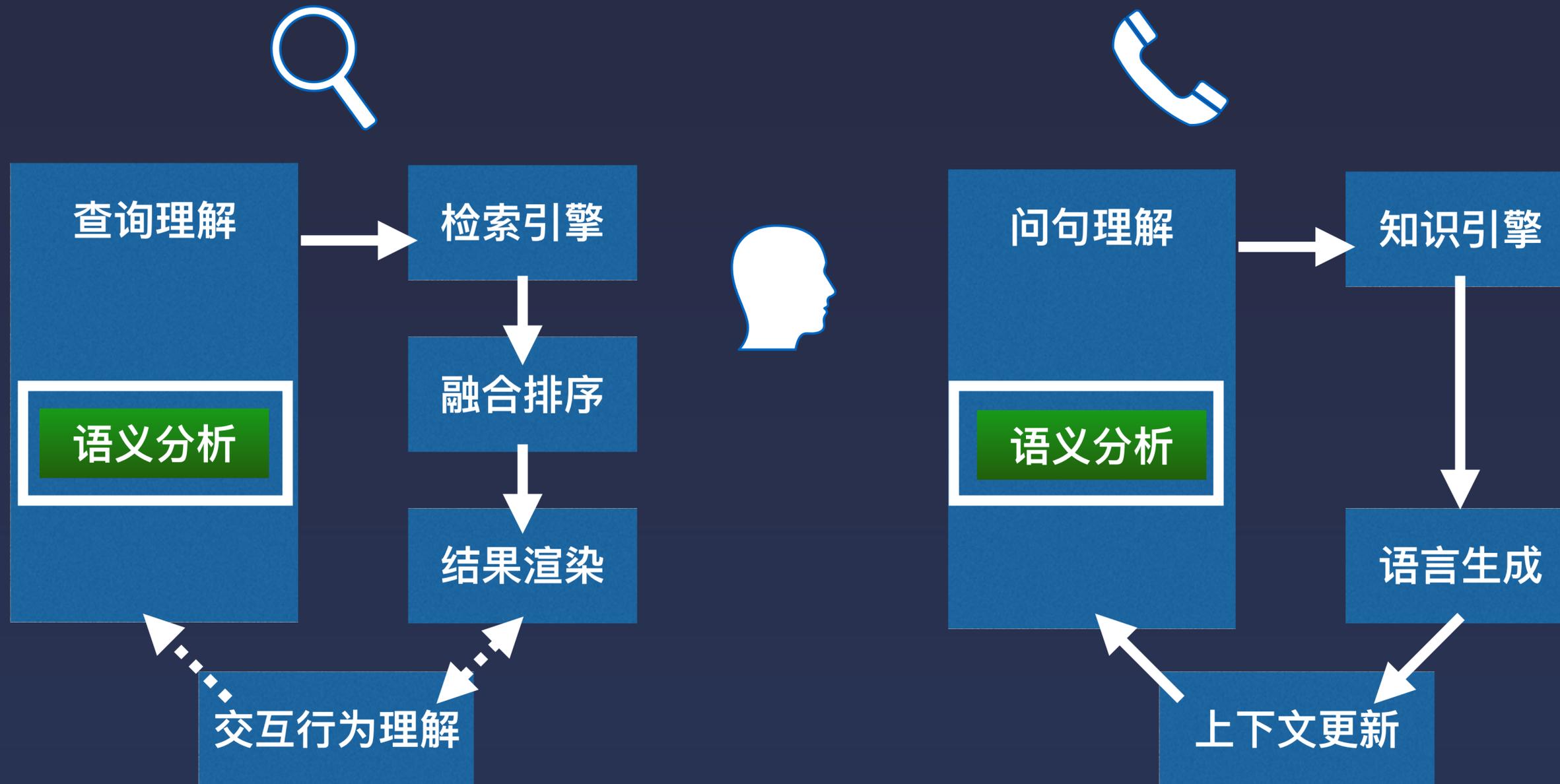
# 搜索引擎 vs. 智能问答 (6)



# 搜索引擎 vs. 智能问答 (7)



# 搜索引擎 vs. 智能问答 (8)



# TABLE OF CONTENTS 大纲

---

- 智能问答应用场景
- 搜索引擎 vs. 智能问答
- 搜索查询理解典型算法
- 从查询理解到问句理解
- 问题与挑战

# 搜索查询理解典型架构 (1)

- 白名单精确匹配完整或者部分查询串



# 搜索查询理解典型架构 (2)

- 词典加正则表达式匹配多个关键词的组合



# 搜索查询理解典型架构 (3)

- 词标注、语义标注的分离和强化



# TABLE OF CONTENTS 大纲

---

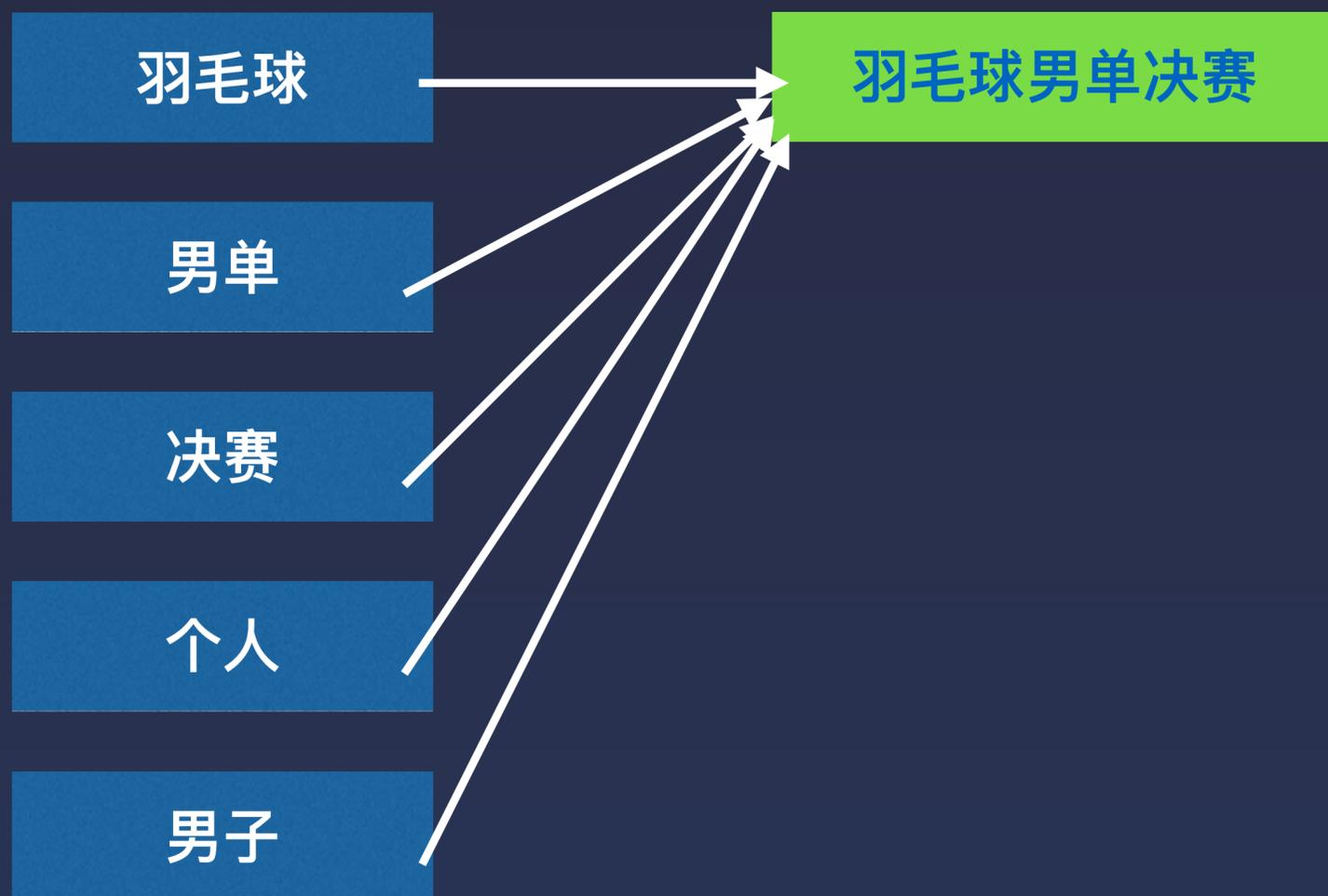
- 智能问答应用场景
- 搜索引擎 vs. 智能问答
- 搜索查询理解典型算法
- 从查询理解到问句理解
- 问题与挑战

# 从查询理解到问句理解 ( 1 )

- 词标注到语义标注的瓶颈
- e.g. 去年羽毛球奥运会比赛单打冠军是谁, 男子的?
- 知识库词典分类空间爆炸问题
- 语法规则的空间爆炸问题
- 语义标注模型的扩展性问题
- 不连续单元的语义合并问题

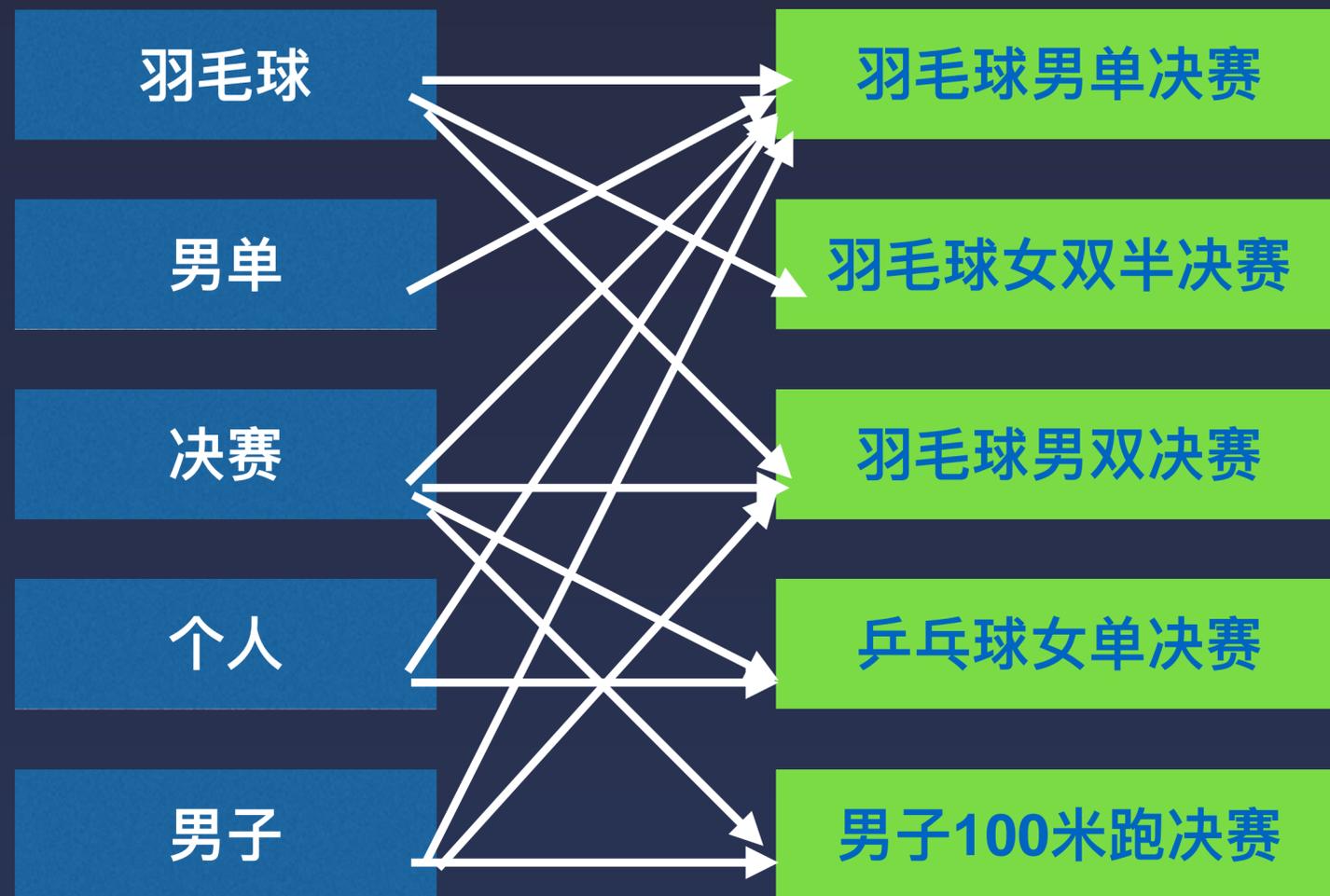
# 从查询理解到问句理解 (2)

- 建立语义索引



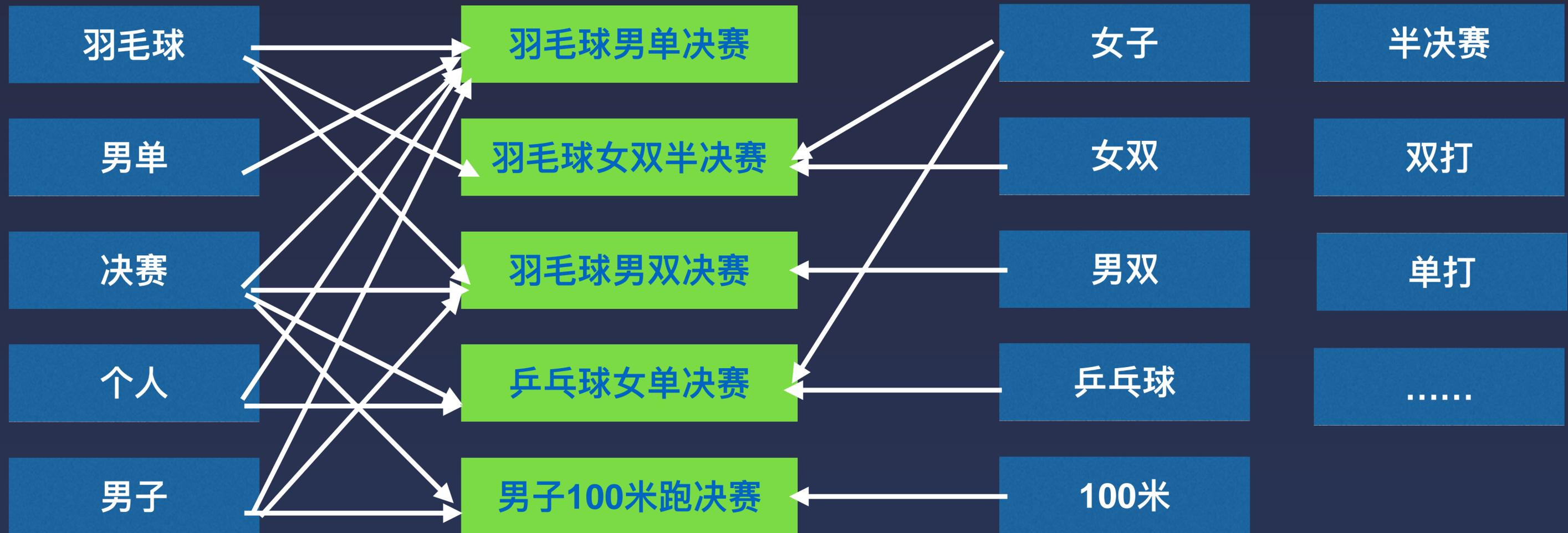
# 从查询理解到问句理解 ( 3 )

- 建立语义索引



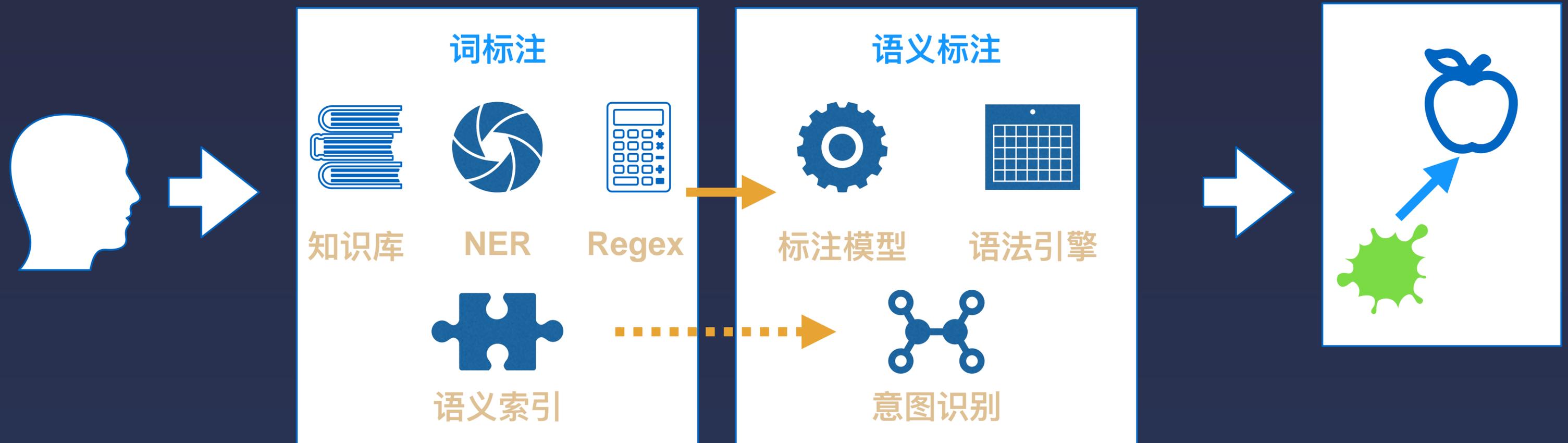
# 从查询理解到问句理解 (4)

- 建立语义索引



# 从查询理解到问句理解 (5)

- 应用语义索引改进词标注



# TABLE OF CONTENTS 大纲

---

- 智能问答应用场景
- 搜索引擎 vs. 智能问答
- 搜索查询理解典型算法
- 从查询理解到问句理解
- 问题与挑战

# 问题与挑战

- 语义知识库的设计与自动构建
- 用于上下文理解的个性化知识图谱
- 语音等主观情感信息的分析引入
- 模糊问题的理解和引导细化

# Q&A

## 智能问答中的问句理解

刘琳 [liulinthu@ymail.com](mailto:liulinthu@ymail.com)

# THANKS