

小爱：语音识别和NLP在语音交互中的实践

王刚

小米

小爱团队负责人



极客时间

重拾极客精神·提升技术认知

下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新



扫一扫下载极客时间App

人工智能基础课

“通俗易懂的人工智能入门课”

王天一
博士 副教授



扫一扫，免费试读

AI技术内参

你的360度人工智能信息助理

洪亮劫
Etsy 数据科学主管



扫一扫，免费试读



关注落地技术，探寻AI应用场景

- 14万AI领域垂直用户
- 8000+社群技术交流人员，不乏行业内顶级技术专家
- 每周一节干货技术分享课
- AI一线领军人物的访谈
- AI大会的专家干货演讲整理
- 《AI前线》月刊
- AI技能图谱
- 线下沙龙



扫码关注带你涨姿势

QCon

全球软件开发大会

成为软件技术专家 的必经之路

[北京站] 2018

会议：2018年4月20-22日 / 培训：2018年4月18-19日

北京·国际会议中心

8折

购票中, 每张立减1360元

团购享受更多优惠



识别二维码了解更多

ArchSummit

全球架构师峰会

2018 · 深圳站

从2012年开始算起，InfoQ已经举办了9场ArchSummit全球架构师峰会，有来自Microsoft、Google、Facebook、Twitter、LinkedIn、阿里巴巴、腾讯、百度等技术专家分享过他们的实践经验，至今累计已经为中国技术人奉上了近千场精彩演讲。

限时**7折**报名中，名额有限，速速报名吧！

● 2012.08.10-12 深圳站



2018.07.06-09 深圳站

会议：07.06-07.07

培训：07.08-07.09



TABLE OF CONTENTES

背景介绍

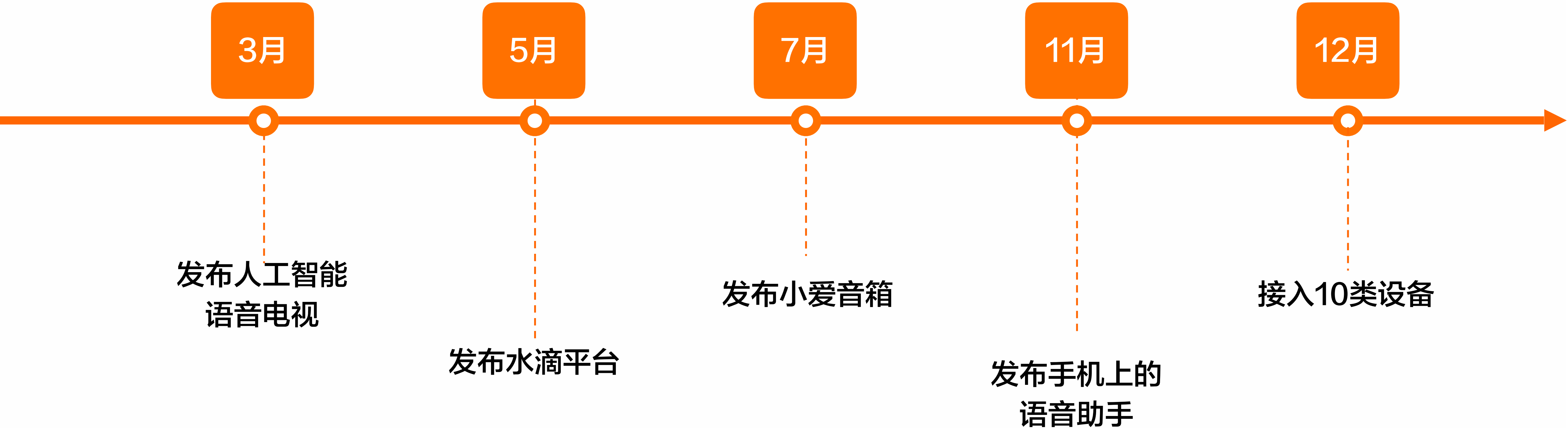
语音交互和搜索的区别

语音技术

自然语言理解

用户行为反馈

小爱语音交互系统



视频

视频

多场景应用



TABLE OF CONTENTES

背景介绍

语音交互和搜索的区别

语音技术

自然语言理解

用户行为反馈

1: 要完成更广泛的任务

信息查找
内容消费
设备控制
生活服务
...

语音交互

信息查找

搜索

2: 唯一一次的交互机会

更精简和精准

语音交互

列表多选

搜索

3: 有记忆的对话

个人信息
历史行为

语音交互

独立的

搜索

4: 用户行为反馈

多模反馈
(如何建模)

语音交互

直接反馈
(Click Model)

搜索

TABLE OF CONTENTES

背景介绍

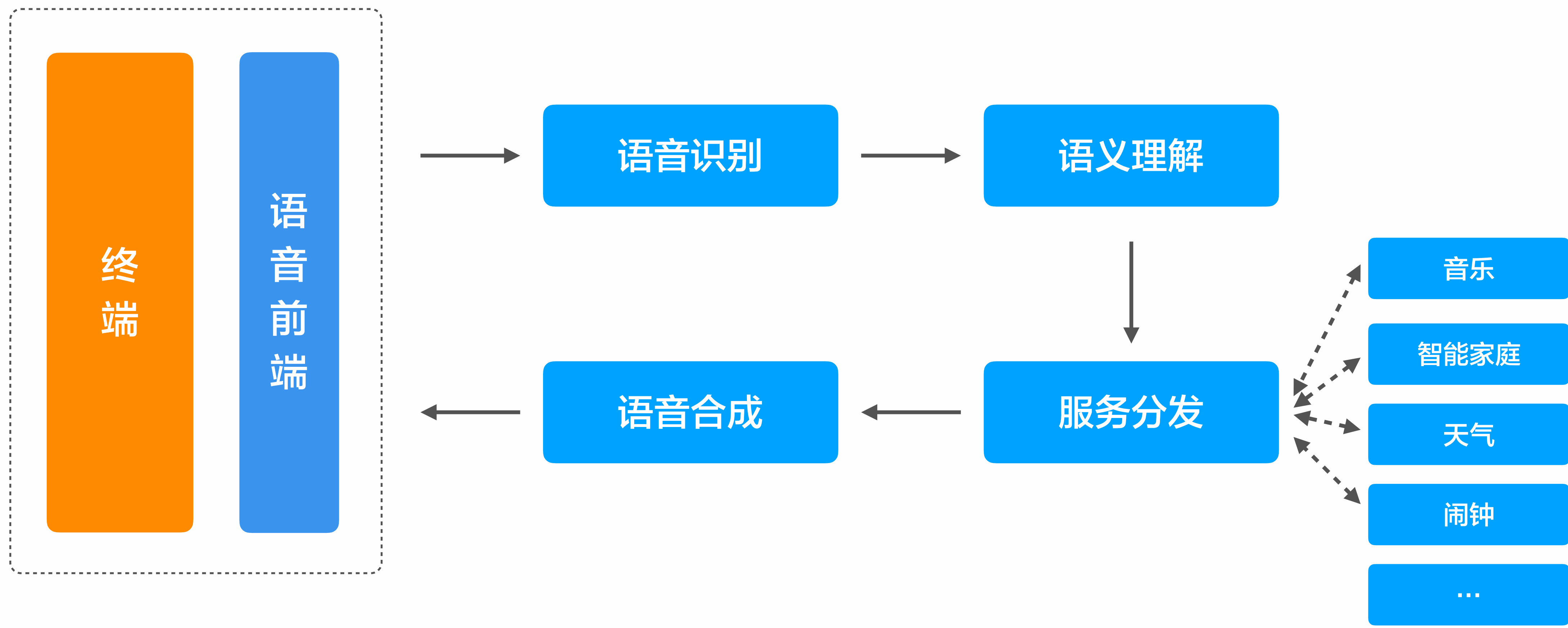
语音交互和搜索的区别

语音技术

自然语言理解

用户行为反馈

系统架构



语音技术的难点

唤醒
回声消除
声源定位

前端

噪音环境下的识别
中英文混杂
不同人声(老人/小孩)
方言

识别

实体词
中英文混杂

合成

语音识别

开放性策略



	电视（近场）	音箱（远场）
1	98%	87%
2	96%	97%
3	98%	98%

字准确率

TABLE OF CONTENTES

背景介绍

语音交互和搜索的区别

语音技术

自然语言理解

用户行为反馈

三个基础

产品的
清晰定义

垂域的
深度优化

全局的
统一决策

四种方式

内置垂域

内部

设备定制

设备开发者

开放平台

技能开发者

AI训练计划

用户

垂域的语义理解

翻译一下 [北京明天晴转多云]
[translate]

[北京] [明天] 的 [空气质量] 是多少?
[city] [date] [quality]

[1] [加] [2] [加] [3] 等于几
[val] [op] [val] [op] [val]

韩磊 南山南
[artist=张磊] [song]

天猫精灵怎么样

文法

词表

知识库

问答对

垂域的语义理解

	属性	举例
1	影片名	星球大战、活着、非诚勿扰
2	演员名	刘德华、周星驰、舒淇
3	导演名	张艺谋、周星驰、陆川
4	角色	伏地魔、六小龄童演的孙悟空
5	年份	90年代、今年、去年
6	分类	电影、电视剧、综艺、纪录片、体育、电子竞技
7	类型	科幻片、动作片、爱情片、警匪片
8	地域	中国、美国、法国、香港
9	发行商	福克斯、漫威
10	获奖名	奥斯卡、金棕榈
11	奖项	最佳男主角、最佳女主角
12	播放量	按播放量排名，最火的、最热的

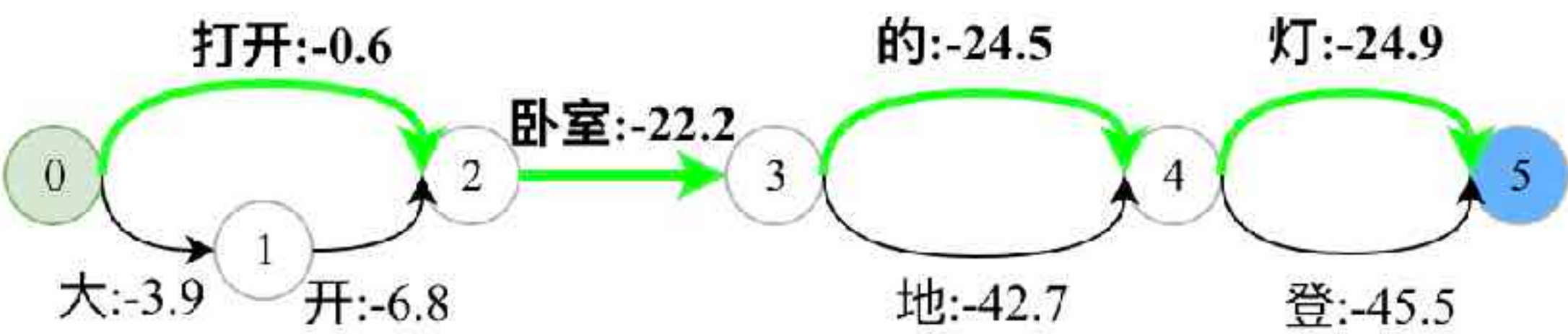
	属性	举例
13	评分	按评分排名，评分大于八分
14	清晰度	高清的，4K
15	付费	免费的
16	年龄	幼儿、小孩、3岁、5到7岁
17	语言	英语、粤语、华语
18	情节	地震、结婚、旅行、历险
19	风格	暴力、残忍、烧脑、荒诞
20	主题题材	励志、毒品、友谊、间谍
21	情感	母子情、师徒情、可爱
22	集	第三集、大结局
23	季	第五季
24	音效	杜比、杜比全景声
25	预告片	预告片

视频领域

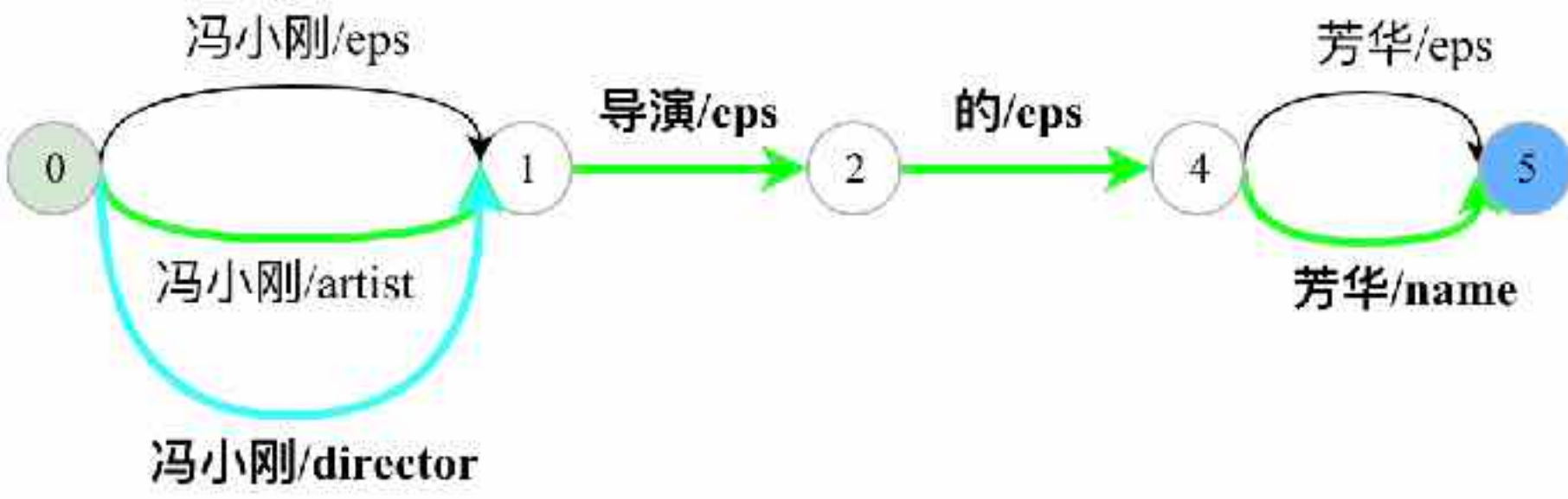
词网格

语音识别:

打开卧室地灯
大开卧室的灯
打开卧室地登
打开卧室的登



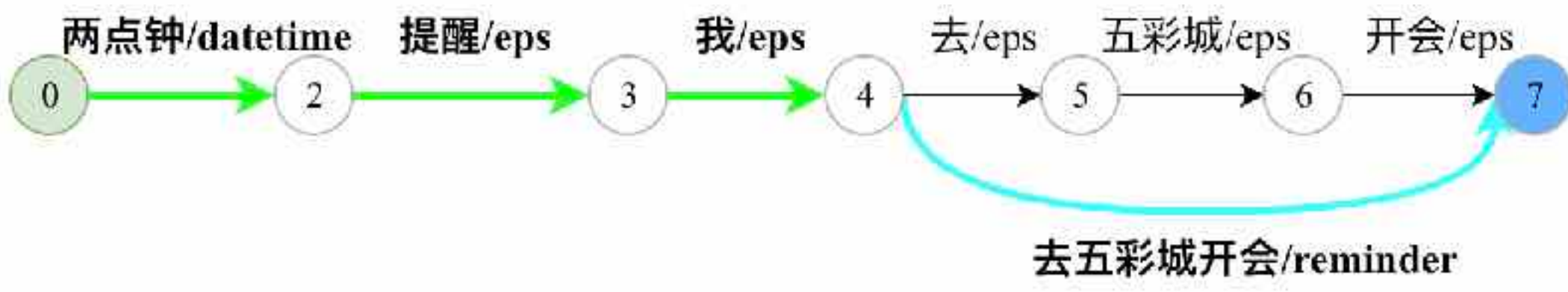
Query: 冯小刚导演的芳华



冯小刚为artist, 当其上下文出现“导演”时, 将冯小刚变换为director

变换文法

Query: 两点钟提醒我去五彩城开会



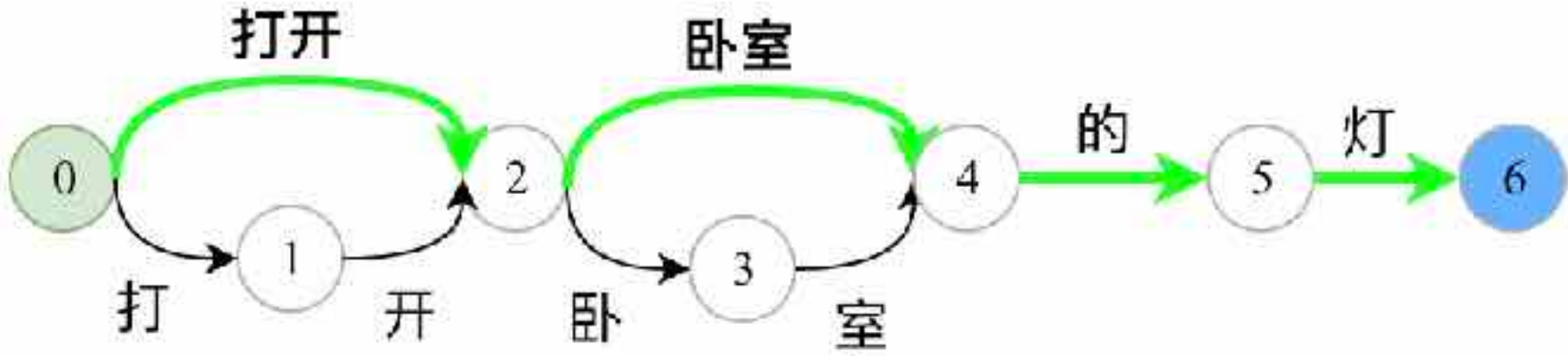
去五彩城开会为需要抽取的提醒消息, 需要识别为一个整体

限定文法

词网格的其他应用

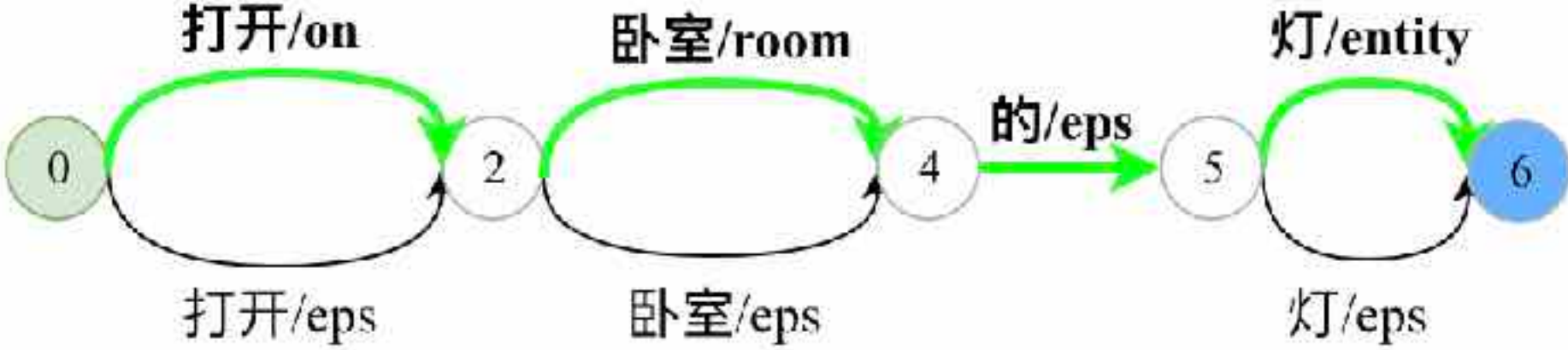
●中文分词

打开卧室的灯



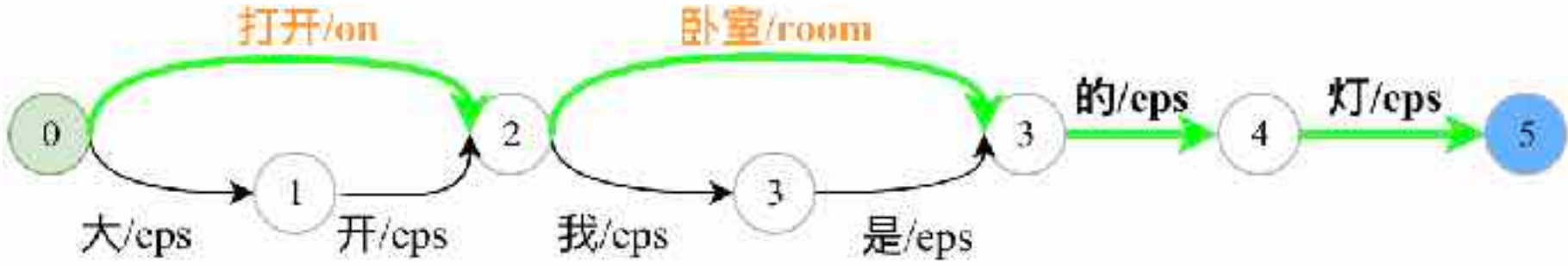
●Slot抽取

打开卧室的灯



●纠错

大开我是的灯



●依存分析

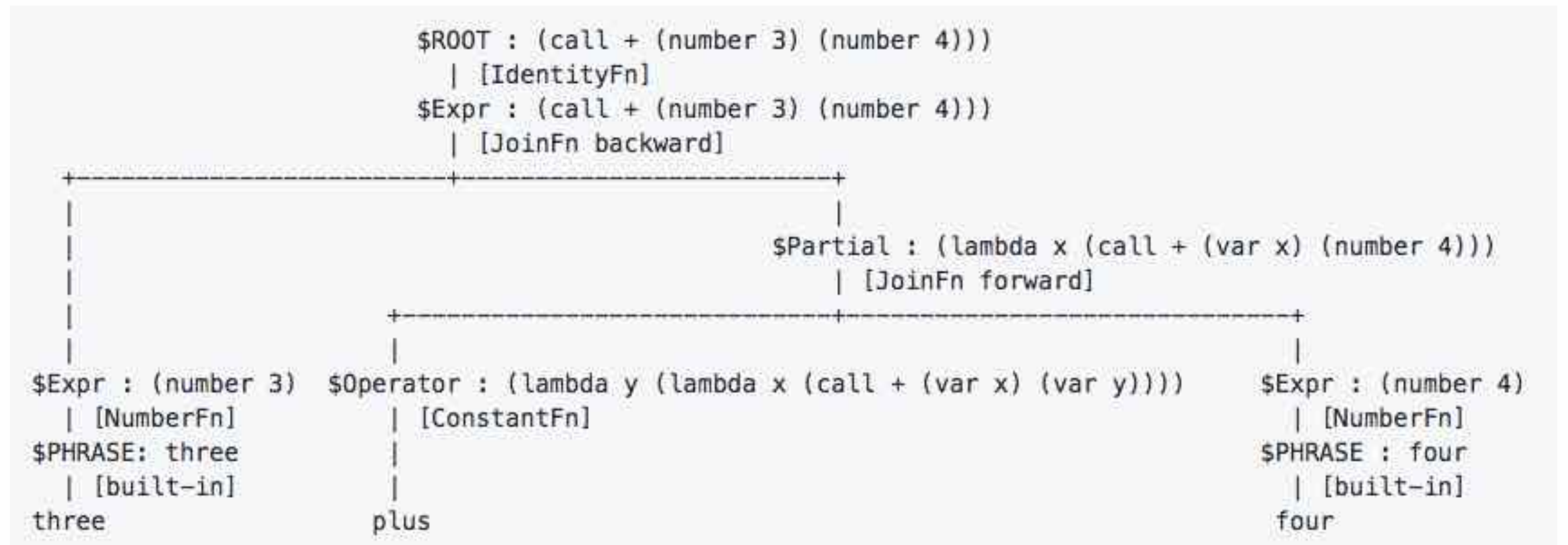
打开卧室的灯



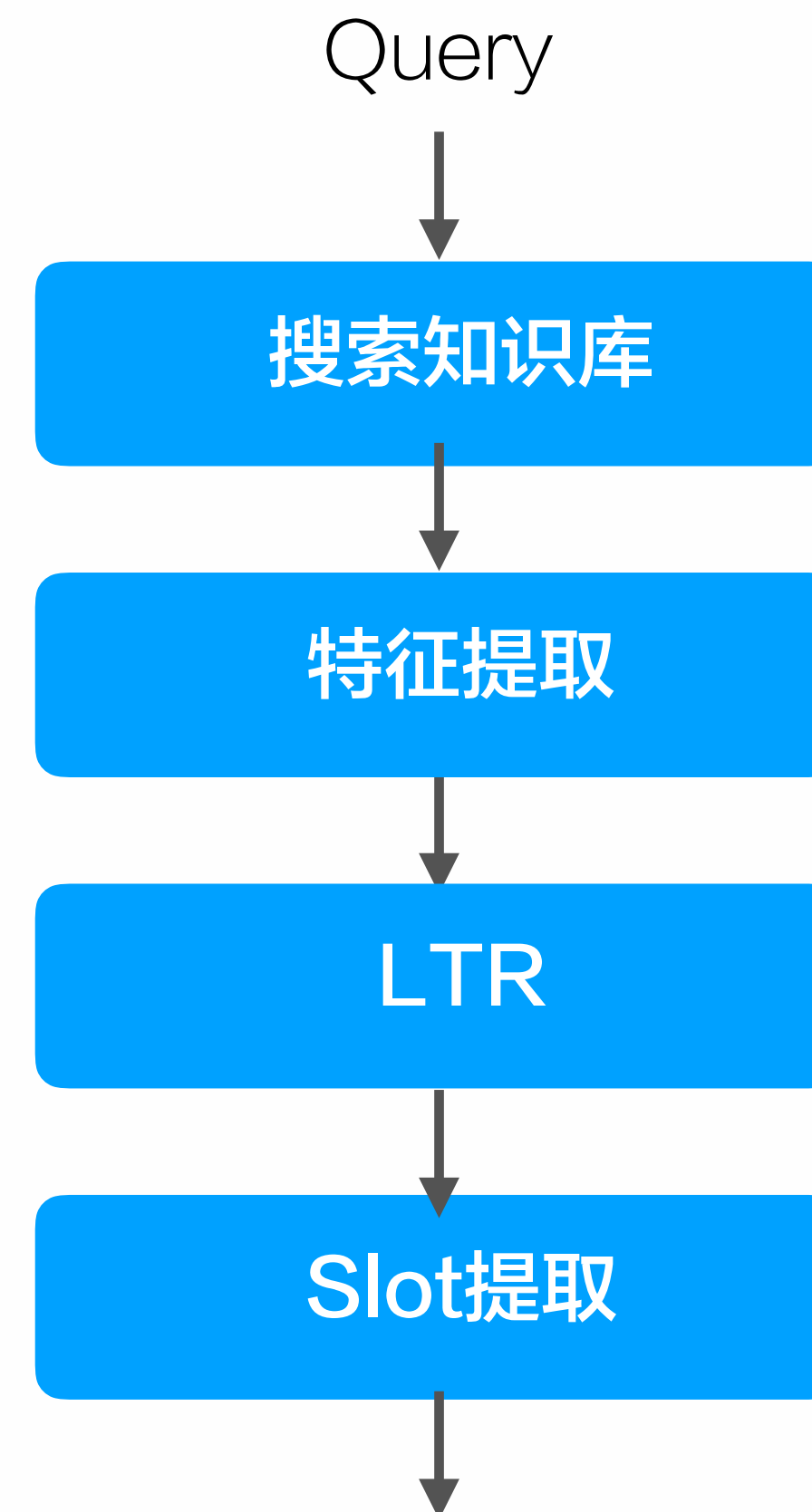
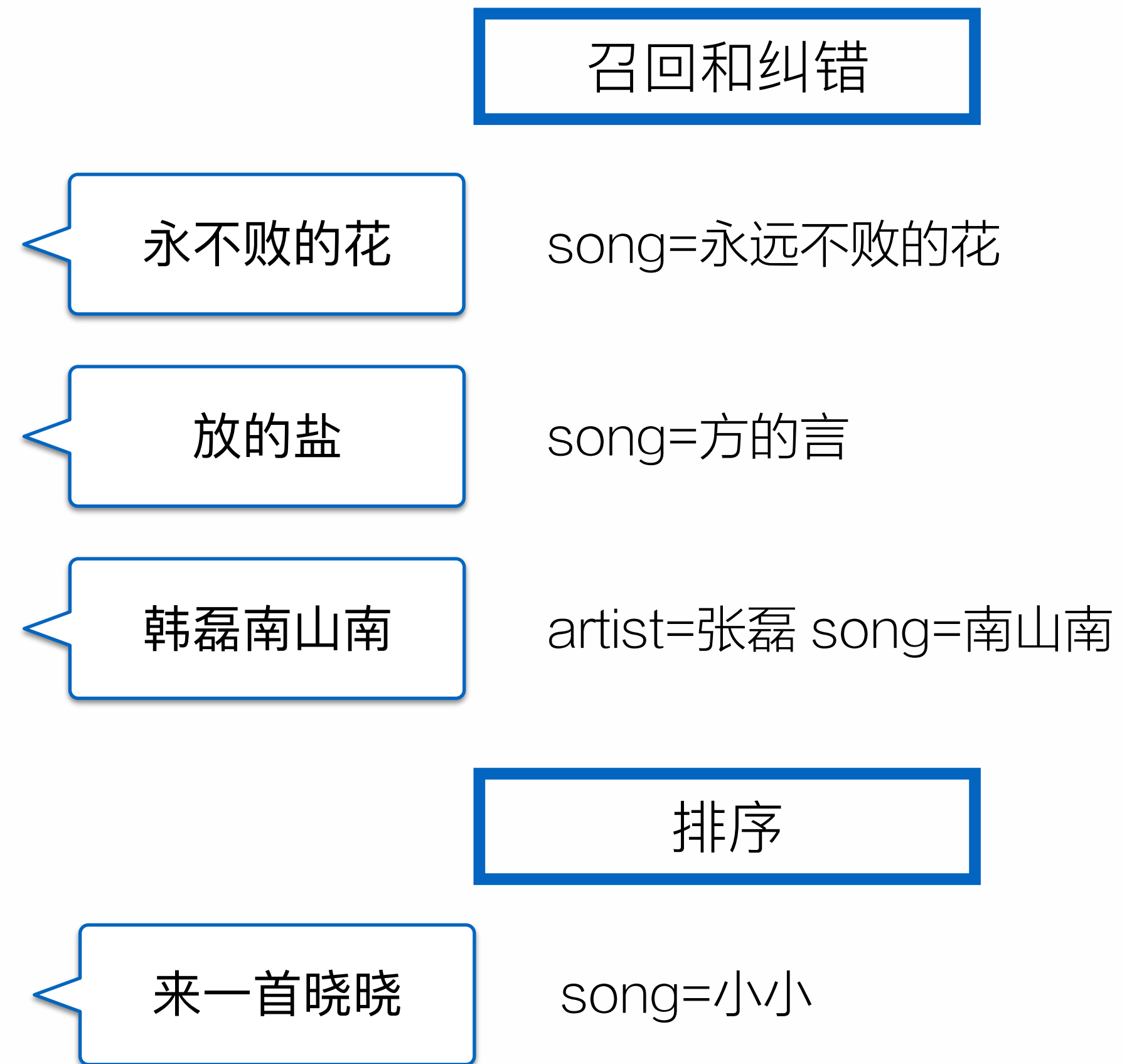
上下文无关文法

- 词网格：有限状态自动机
- 上下文无关文法：下推自动机
 - 通过一个“无限长度”的栈，下一个状态可能跟之前出现过的状态有关

<number> <- 1|2|3|4|5|6|7|8|9|0
<op> <- + | - | * | /
<expr> <- (<expr>) | <number>
 | <expr> <op> <expr>



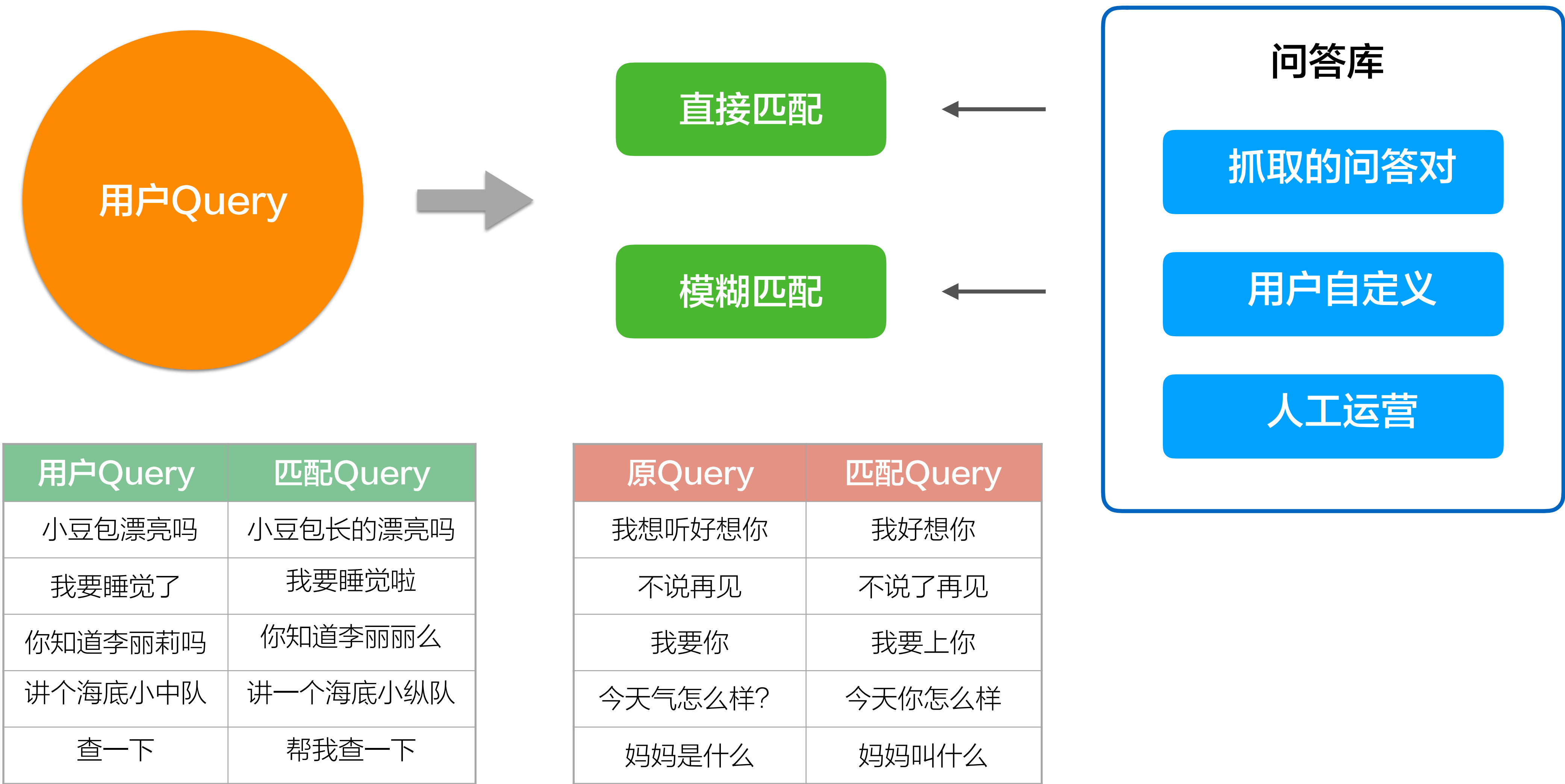
知识库搜索



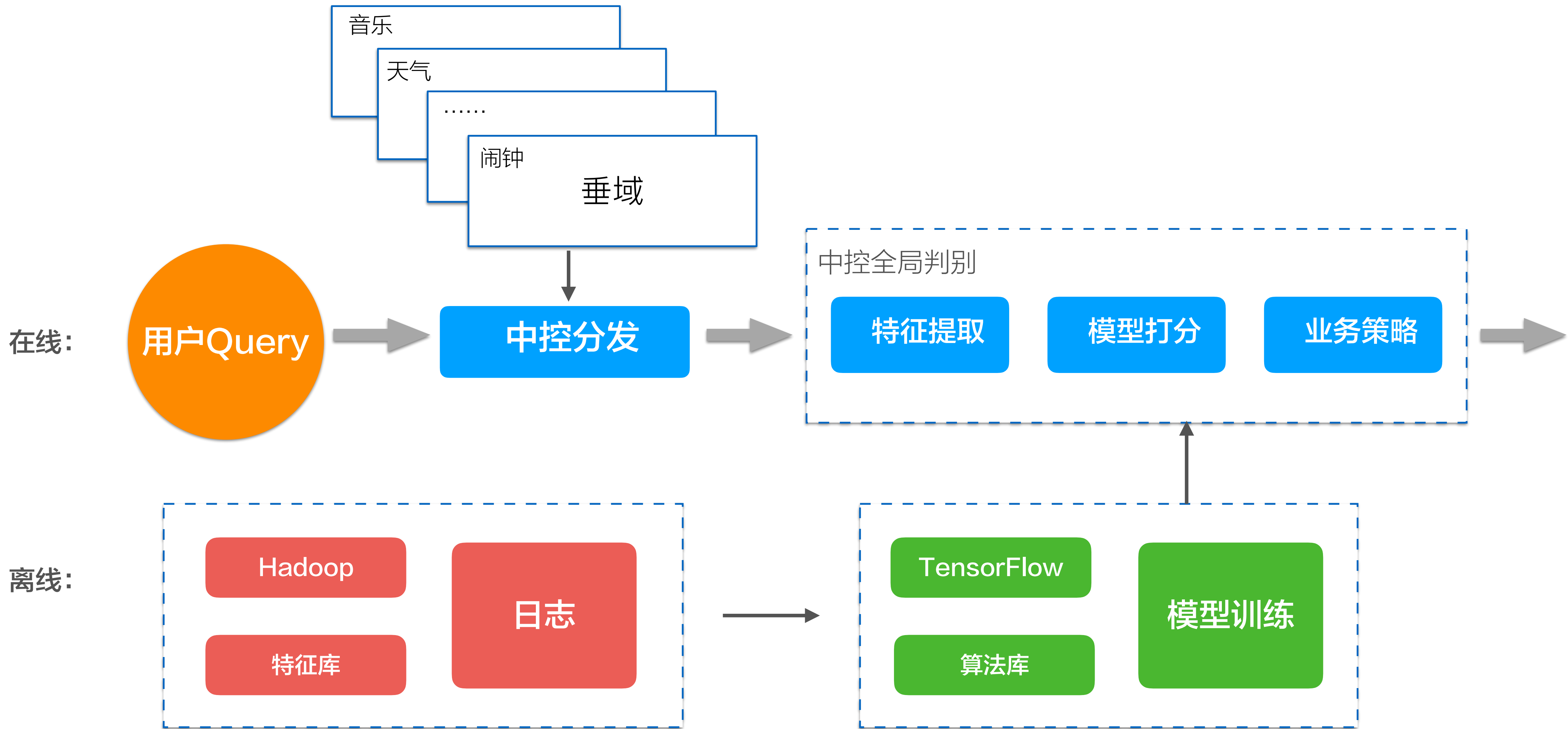
使用50个特征，如：

- query倾向类型
- 拼音匹配
- 歌手热度
- 歌曲热度
- ...

问答对匹配

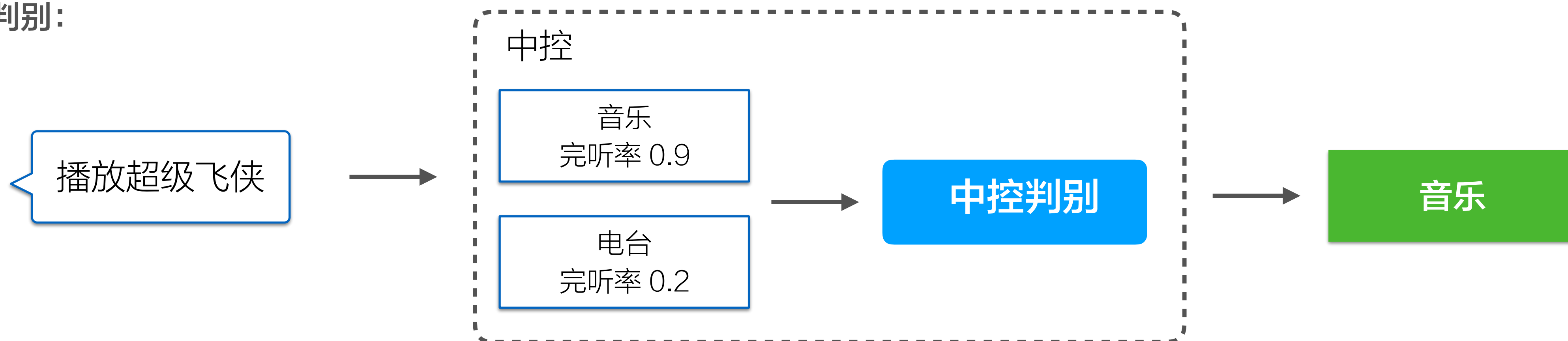


中控全局判别



中控全局判别

歧义判别:



多轮对话管理:

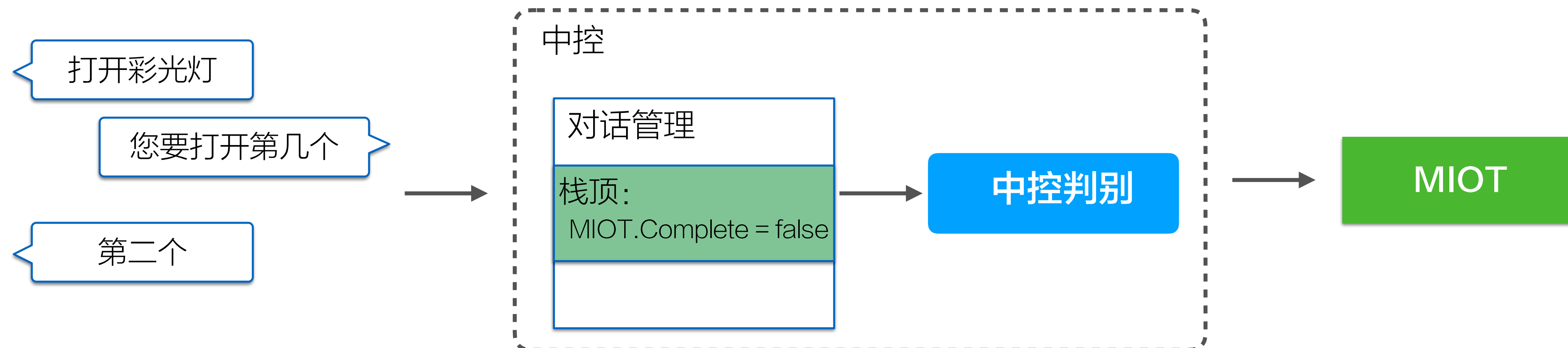


TABLE OF CONTENTES

背景介绍

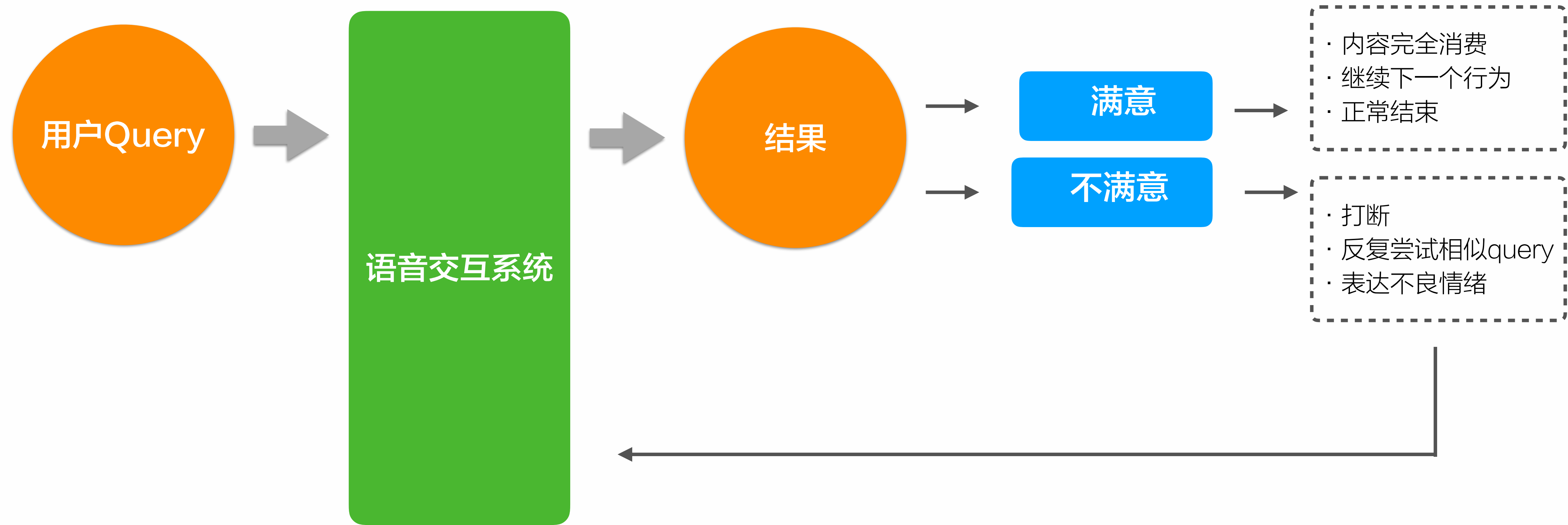
语音交互和搜索的区别

语音识别和合成

自然语言处理

用户行为反馈

用户行为反馈



用户行为反馈

Query	首条完听率
说散就散	62.8%
小苹果	68.8%
体面	74.1%
播放体面	70.2%
播放说散就散	60.9%
远走高飞	66.1%
小跳蛙	80.2%
我们不一样	61.4%
成都	57.7%
播放小苹果	65.3%

满意

Query	首条完听率
YOU	6.0%
LETGO	6.4%
SHAPE OF YOU	2.3%
TFBOYS唱的朵妈妈	10.0%
你帮我换一首广州爱情故事	9.5%
一生中所爱	0.7%
修炼爱情	1.3%
BABY上	2.8%
一个什么	11.0%
你放错了要放歌沧海一声笑	6.3%

不满意

用户行为反馈

构建具有自主学习能力的对话系统

Thanks!