



自然语言理解和知识图谱技术应用探索

(第三方bot/skill开放平台)

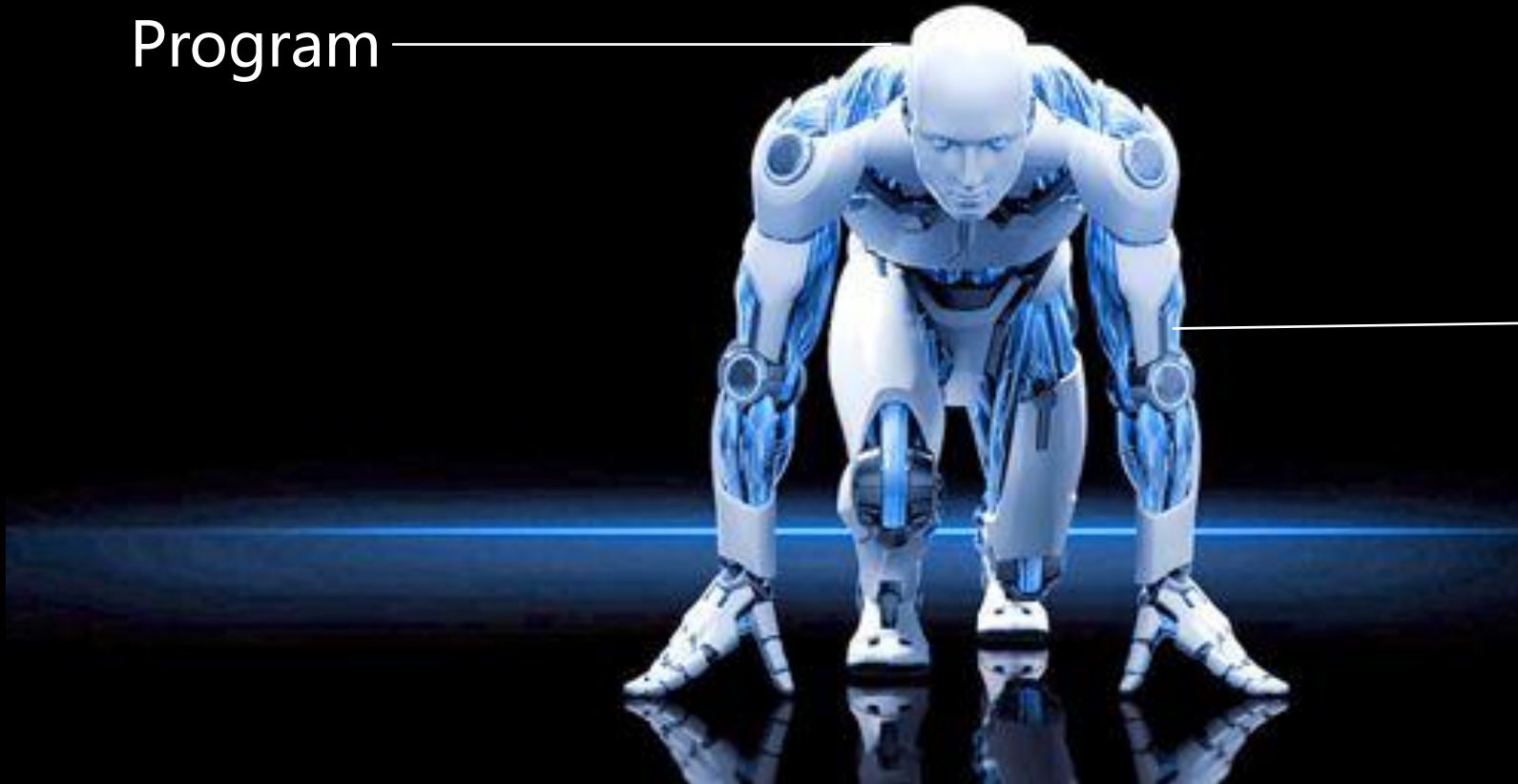
2018.1@wang yu

新华社见习记者：“i思” 机器人



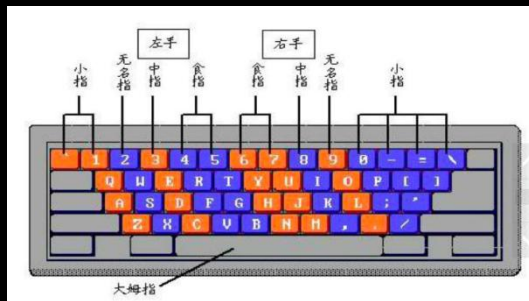
机器人：bot & robot

Program



Machine

三十年小机遇：人机交互升级



PC
键盘鼠标
学习周期：周

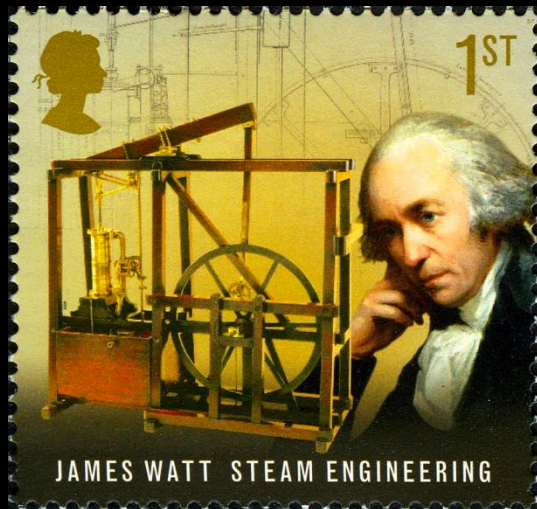


Mobile
触摸屏
学习周期：日



iOT&Robot
说话
学习周期：分钟

两百年来大变革：智能化



蒸汽机 (1785)

工业革命
解放体力劳动者
从手工作坊到自动化工厂



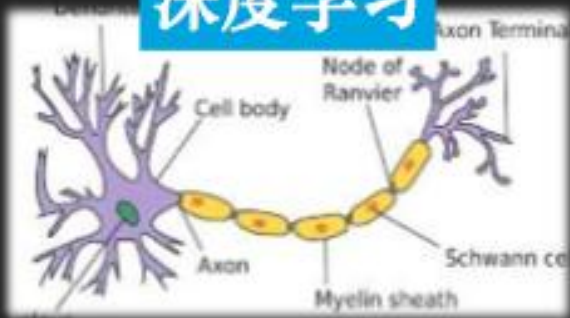
智能机器人 (2017)

人工智能
解放脑力劳动者
知识工作自动化

bot : 更聪明 or 更有知识



深度学习



知识图谱



知识工作者：解放，not代替

?

秘书

客服

陪伴

分析员

财务
咨询

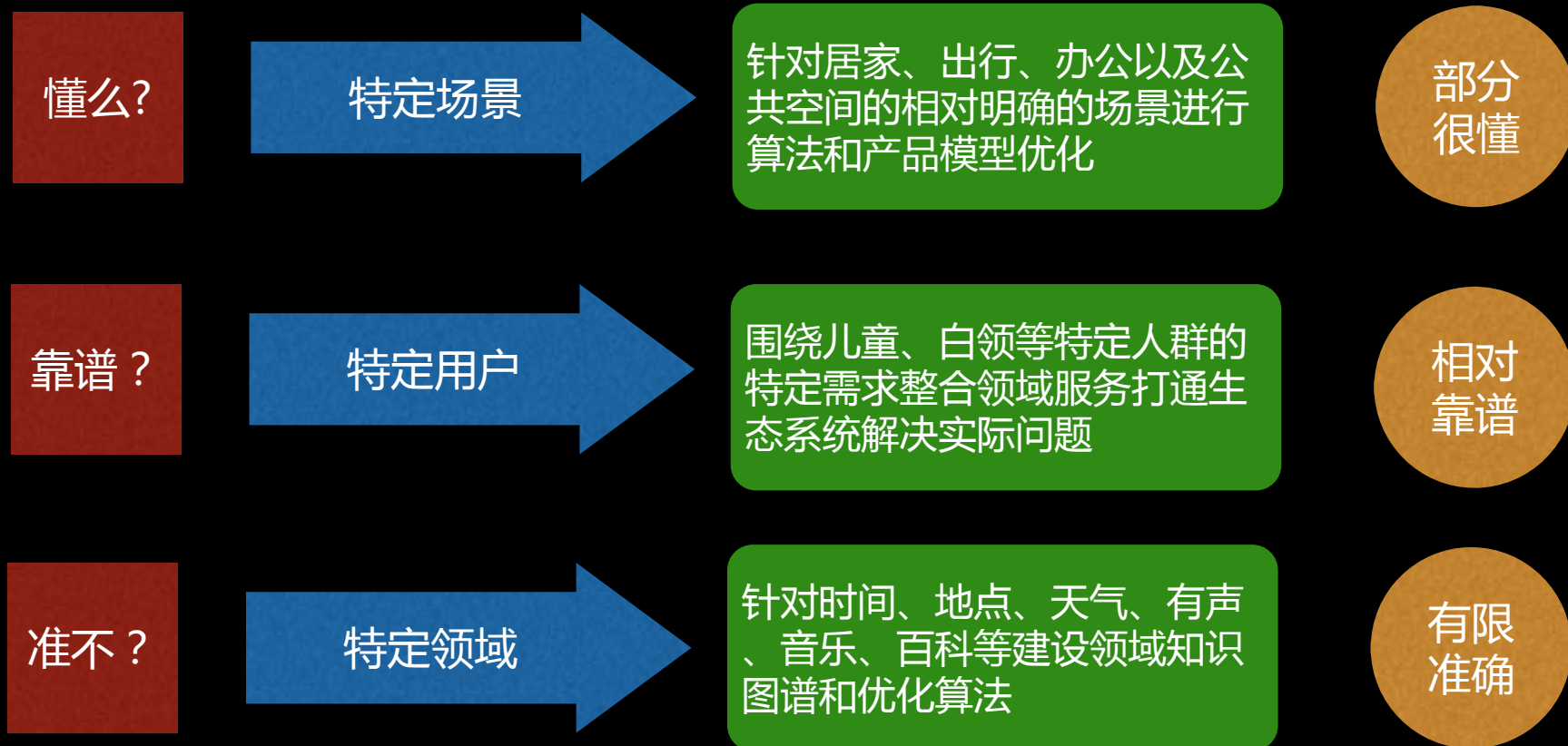
记者
编辑

教师
培训师

项目
管理

律师
医生

用户痛点：演示系统 to 落地产品



开发者痛点：用电=建电厂？

自行 开发

- 自然语言理解和知识图谱技术人才难招且成本高
- 技术需要积累且实现有一定门槛和难度
- 数据和知识图谱的积累
- 领域服务对接

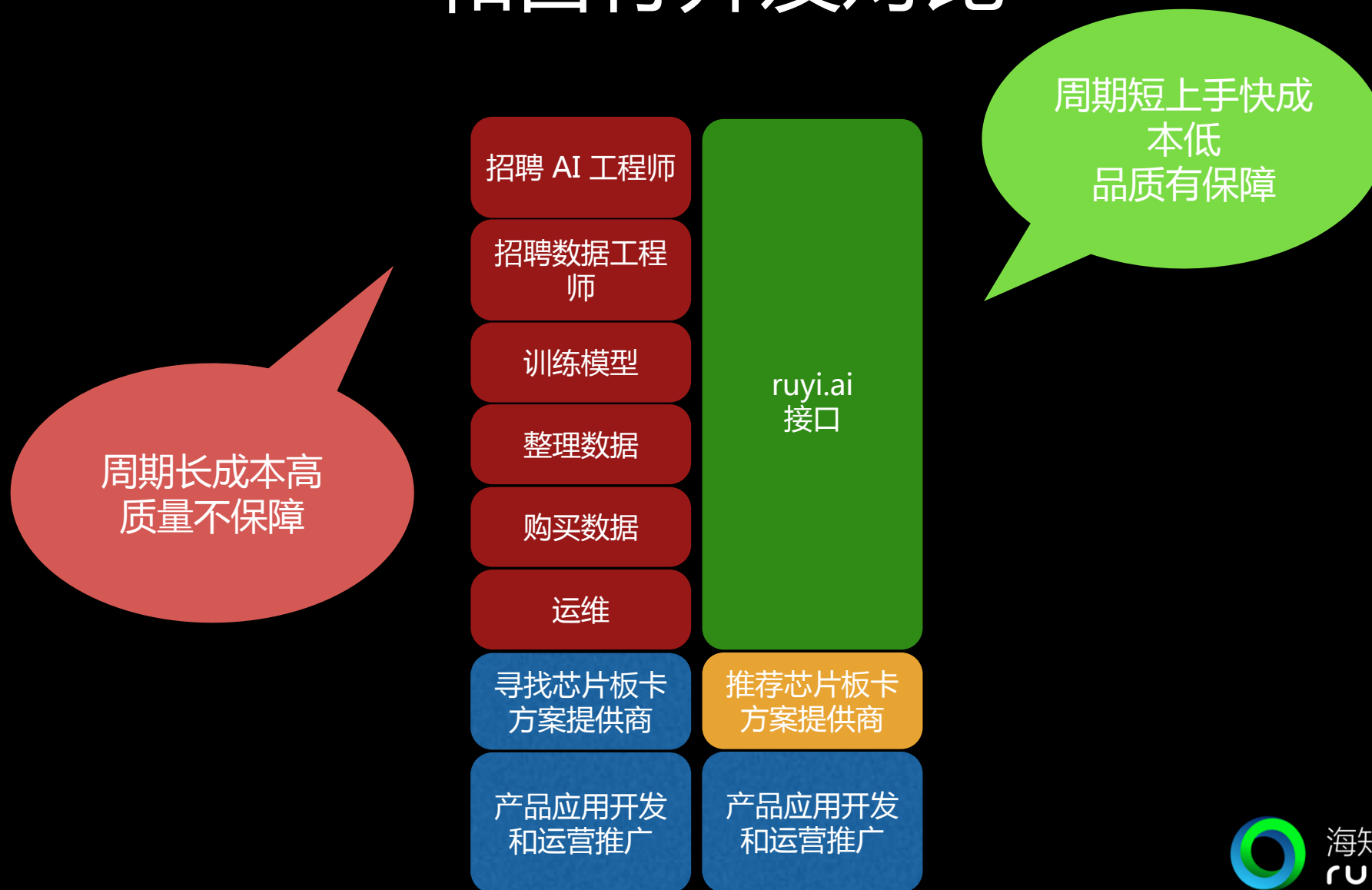
第三方 API

- 当前开放 API 的语义能力在不同领域参差不齐
- 知识库规模和完整性准确性以及版权风险
- 个性化弱开放程度不高
- 国外的技术中文理解以及本地化支持不够

bot/skill开发者痛点(不含硬件)

产品经理	Android 工程师	NLP工程师	系统架构师	系统管理员	语言学家
客户服务	iOS 工程师	知识图谱工程师	爬虫工程师	运维工程师	心理学家
行业专家	前端 工程师	机器学习工程师	数据工程师	安全工程师	音乐专家
文案	java 工程师	数据标注员	数据工程师	DBA	营养专家

和自行开发对比



ruyi.ai产品线

魔戒

技能插件工具
botOfRings
V 1.0

- ①提供语义理解 API: 接收文本, 输出 Json
- ②面向模板进行机器学习的知识库问答系统
- ③开放对接第三方 bot 平台

水晶球

智能分析员
botAnalyst
V 0.9 beta

- ① 结构化、半结构化数据, 非结构文本采集
- ② 自然语言理解形成领域知识图谱
- ③ 结合量化模型形成预测、判断、风险预警和线索发现等辅助决策支持

聚宝盆

知识图谱生产线
KnowledgePot
V 0.1 alpha

- ① 人工&半自动化 构建领域知识图谱
- ② 覆盖相对全面 质量相对可靠 更新比较及时
- ③ 医药行业, 金融领域, 法律服务

ruyi.ai业务线

智能
家居

智能电视、冰箱、音箱等硬件，智能服务机器人

新智能
客服

大型企业售后客服、售前营销导购以及业务对接

企业
智能化

营销分析员、招标采购分析员等机器人大脑平台

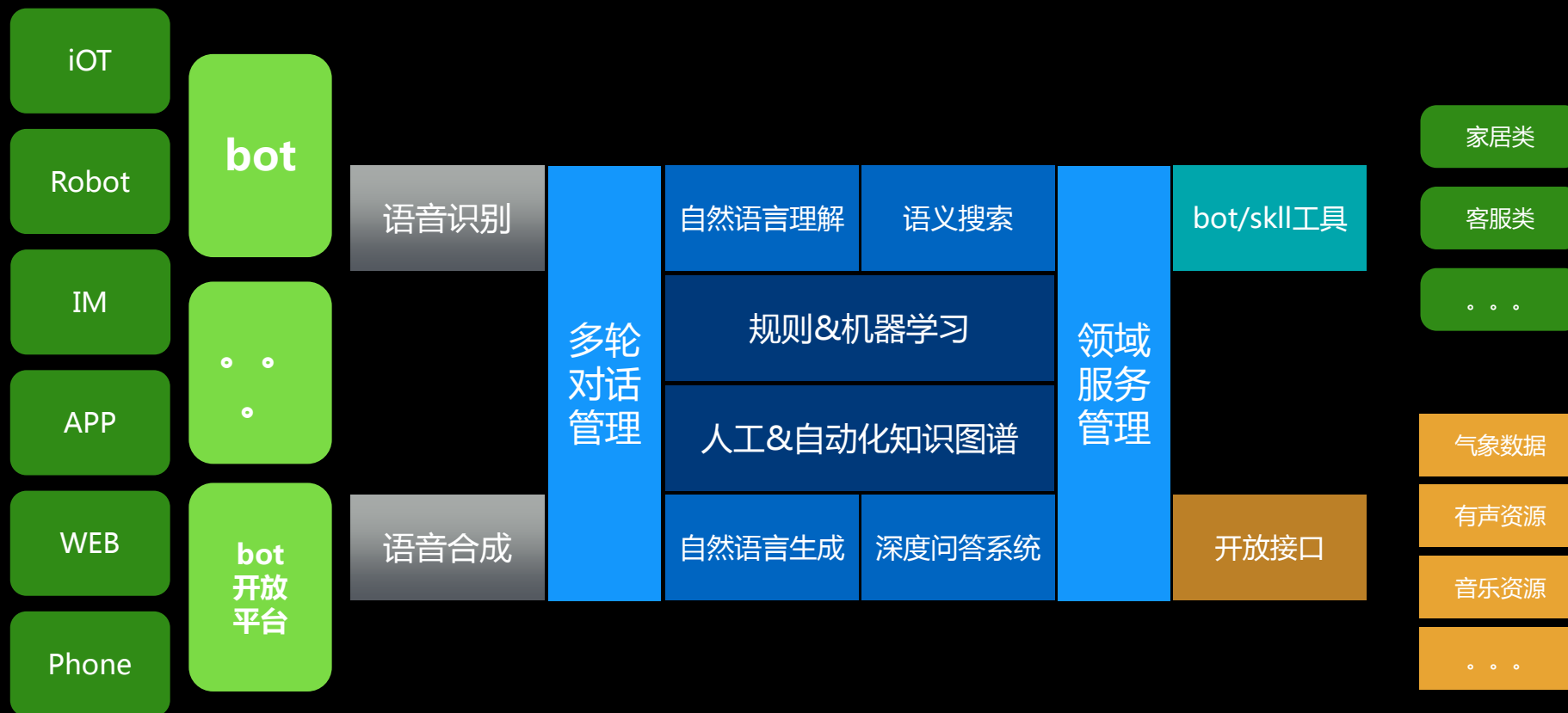
ruyi.ai产品线-1

魔戒

技能插件工具
botOfRings
V 1.0

- ①提供语义理解 API: 接收文本, 输出 Json
- ②面向模板进行机器学习的知识库问答系统
- ③开放对接第三方 bot 平台

bot 引擎：人机结合



"给我放一首周杰伦的稻香"

```
{
  code: 0,
  msg: "ok",
  - result: {
    _text: "给我放一首周杰伦的稻香",
    msg_id: "fff57183-c89e-474d-a350-71cf2c43c9ca",
    - intents: [
      - {
        - parameters: {
          musicGroupId1: "e202ad75-fb8b-43d7-a0d5-5f56de97851e",
          musicRecordingId: "156f0b73-9abf-4651-88fd-10ad08c688d6",
          artistName1: "周杰伦",
          songName: "稻香",
          artist1: "{\"score\":1498943,\"type\":\"musicgroup\",\"id\":\"e202ad75-fb8b-43d7-a0d5-5f56de97851e\",\"name\":\"周杰伦\"}",
          song: "{\"score\":3862598,\"type\":\"musicrecording\",\"id\":\"156f0b73-9abf-4651-88fd-10ad08c688d6\",\"name\":\"稻香\"}",
          service: "music"
        },
        action: "sys.action.kg/music/artist_song",
        name: "-1-按歌手歌名组合点歌",
      },
      - result: {
        - music_list: [
          - {
            name: "稻香",
            id: "156f0b73-9abf-4651-88fd-10ad08c688d6",
            pic: "http://img.xiami.net/images/album/img60/1260/3046831471875446_1.jpg",
            music163: "5235486",
            xiami: "1770153970",
            artist: "周杰伦"
          },
          - {
            name: "黑色幽默(live版)",
            id: "20eb0a66-5496-4da2-a141-09084e3ffa46",
            music163: "26632221",
            artist: "周杰伦"
          },
          - {
            name: "双截棍 (Live)",
            id: "a6e5111f-ade1-4e7f-86bb-a2e092d10227",
            music163: "433681297",
            artist: "周杰伦, TAHITI The Gaugin Years: Songs and Dances"
          },
        ],
      },
    ],
  },
}
```

自然语言理解

意图识别

触发器执行

skill : bot平台上的APP



小视

上线**9款**语音交互
7款来**海智**

第三方专区（需通过语音遥控器操作，点击上图购买）

 旅行家 ★★★★★ “我想去看星空”	 谜语大作战 ★★★★★ “打开猜谜语”	 词语接龙 ★★★★★ “打开词语接龙”
 词汇乐园 ★★★★★ “词汇乐园”	 成语接龙 ★★★★★ “刚才这个成语啥意思”	 翻译官 ★★★★★ “你今天很好看用英语怎么说呢”
 古诗词 ★★★★★ “苏轼的代表作是什么”	 饮食百科 ★★★★★ “吃蓝莓对什么好”	 菜谱 ★★★★★ “你会不会做红烧肉？”



skill-饮食百科

猪肉和什么相克

如意饮食百科



杏仁
会引起腹痛



羊肝
共烹炒易产生怪味



甲鱼
会引起腹痛



牛排
性味有所抵触



香菜
伤身

ruyi.ai
海知智能

你可以说

“菠菜和什么不能一起吃” “蜂蜜和豆腐可以一起吃吗” “哈密瓜和香蕉相克吗” “萝卜和橘子适合一起吃吗” “鱼腥草的功效”

按住  说话

skill-如意菜谱

您还可以这样说

说菜名

“农家小炒肉怎么做”

说类别

“川菜怎么做”

说场景

“早餐菜谱”

翻页

“换一批”

进入菜谱详情

“打开第一个”

“第一步怎么做”

“下一步”

今日推荐



黄金玉米烙

1



飘香酸菜鱼

2



可乐鸡翅[简单到没下过厨也会做]

3



skill-古诗词

少小离家老大回的下一句

如意古诗词

回乡偶书

贺知章

少小离家老大回，乡音无改鬓毛衰
儿童相见不相识，笑问客从何处来

ruyi.ai
海知智能

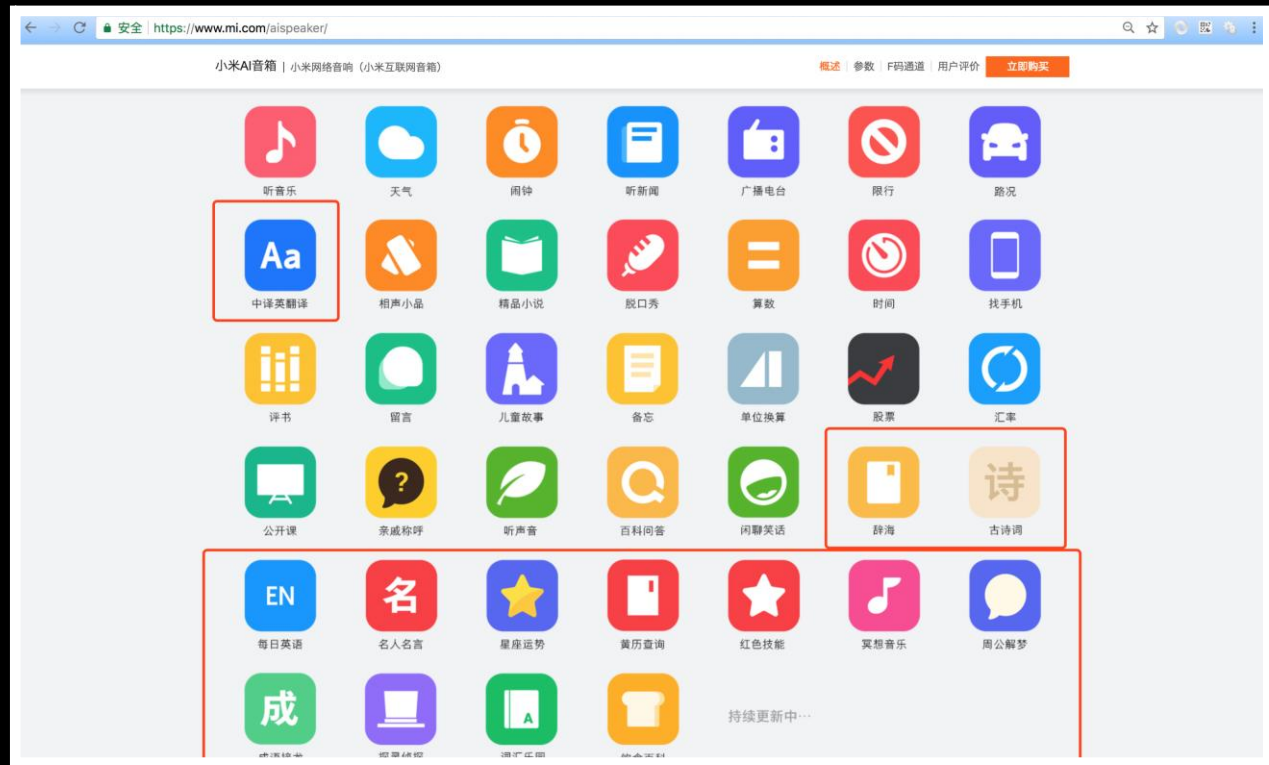
你可以说

“静夜思” “李白的古诗” “苏轼”

按住  说话

小音

17款语音交互功能前3名畅销来自海智



bot/skill-其他应用

Haier

馨厨冰箱

智享珍馐 尚品人生

海尔干湿分储互联网冰箱



平台合作伙伴

水滴计划

小米电视人工智能开放平台



DUER OS



www.dui.ai

思必驰DUI开放平台

AI Speech Dialogue User Interface

一站式智能对话定制平台

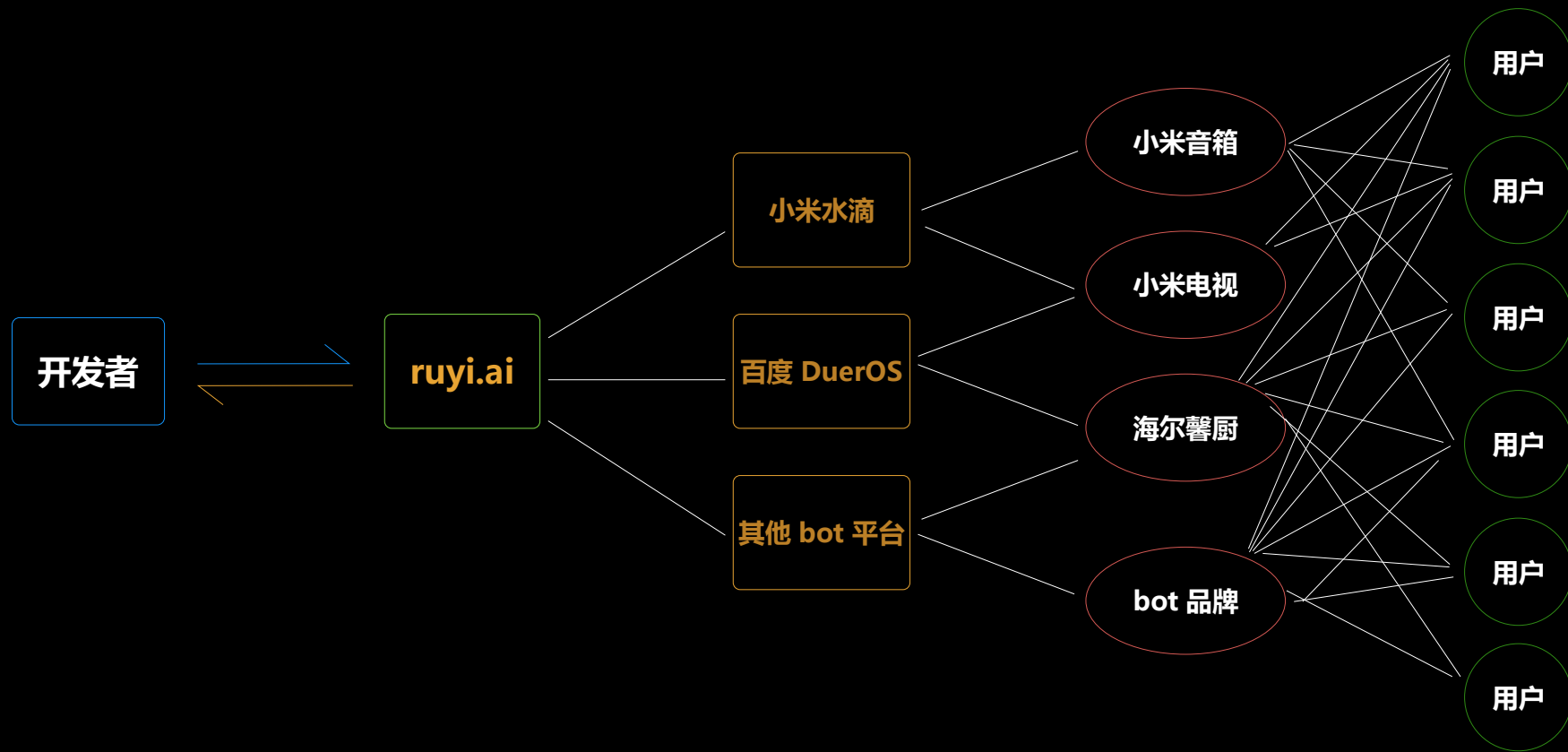
国内最专业的AI技能商店

了解更多

智能车载 | 智能家居 | 智能机器人 | 智能手机



第三方bot/skill开放平台-生态



三步做个”机器人”

创意设计

特定用户 特定场景 特定服务

训练学习

问答对儿 知识库 训练集

接入成“人”

智能硬件、机器人、微信公众号等

ruyi.ai-后台操作界面

测试专用

数据统计

粉丝管理

对话场景

测试自定义触发器

JS模板语言

词典实体

导入知识库

机器人设置

技能插件

素材管理

JS模板语言

暗号-雨天统计

暗号-降温

暗号-计算温差

JS-温度

JS-计算温差

JS-降温

JS-雨天统计

JS-温度

保存并应用

添加认证号多条回复

对话能力拓展

接受暗号

添加接受暗号

发出暗号

天气

添加发出暗号

意图处理

是否必须	名称	值	类型	提示语	默认值
☐	date	\$date	@sys.date	-	默认值
☐	city	\$城市	@城市	-	默认值
☐	key	\$key	@sys.any	-	9e746fa0e9t
☐	date_raw	\$date.raw	@sys.date	-	默认值
☐	参数名称	值	参数类型	-	默认值

触发器

对话场景-意图模版

用户说
机器人答
暗号
意图处理
触发器
优先级

用日期查询天气 • 1. 意图名称

• 用户说 ? 2. 机器学习 机器学习 ?

今天的天气	
明天的天气	✕ 已 回 答
周一的天气	
周二的天气 • 3. 列举用户说	
.....	

事件 ? <

机器人答 ?

硬件 微信

微信回复

暂未开放天气查询功能哦~ • 4. 自定义回复

系统词典 - 系统词典+知识图谱词典 - 累计上亿词典条目 (属性体)

词典名称	词典释义
sys.number	整数, 不包负数
sys.duration	时间段, 如 1 小时 分钟
sys.timedelta	时间差, 如提前半小时 延迟五分钟
sys.time	时间实体可识别, 如 日期星期、今明、农历、节日 别格式为时间戳
sys.date	日期识别格式为 YYYY-MM-DD
sys.year	年份识别格式为 YYYY
sys.city	国内主要城市, 如上海
sys.location	地点, 如盛夏路
sys.verb	常用动词
sys.flightno	航班号识别, 例如东航 234, 识别为 MU 234
sys.email	邮箱域名
sys.any	任意字符串, 可为空
sys.any.not-blank	不为空的任意字符串

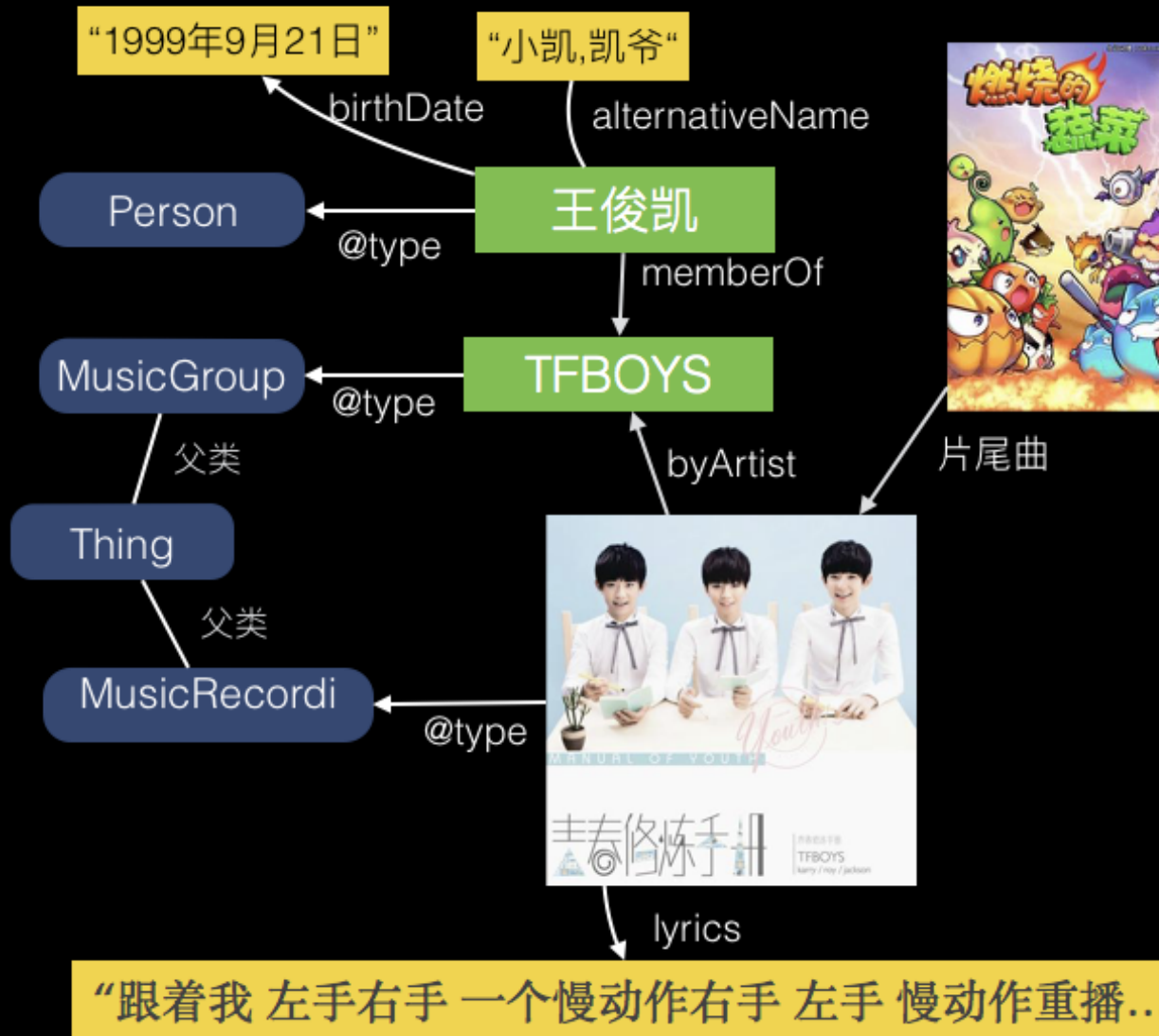
图谱	实体	属性
POI	5000万	10
音乐	2000万	10
机构	3000万	10
百科	2000万	5
有声	500万	20
饮食	1000	30
医药	7000医院, 9万科室, 24万医生 10万资质证书	50
家谱	5000	10
.....		

音乐知识图谱应用

- 实体
 - @id 标识
 - name 名字 / 别名
 - property 属性
 - relation 实体间二元关系
- 本体
 - @type 实体分类
 - taxonomy 分类树
 - mapping 映射关系
- 复杂数据结构: 时间, 数字

应用场景

- 理想状态, 用户这么说
 - 我要听“tfboys”的“青春修炼手册”
- 但用户也可能这么说
 - 我要听“王俊凯”的“青春修炼手册”
 - 我要听“左手右手”那首歌
 - 我要听“燃烧的蔬菜”的片尾曲
- 用户还会问
 - “青春修炼手册”是谁唱的?
 - “小凯”多大了?
 - “王俊凯”是“tfboys”里最大的吗?

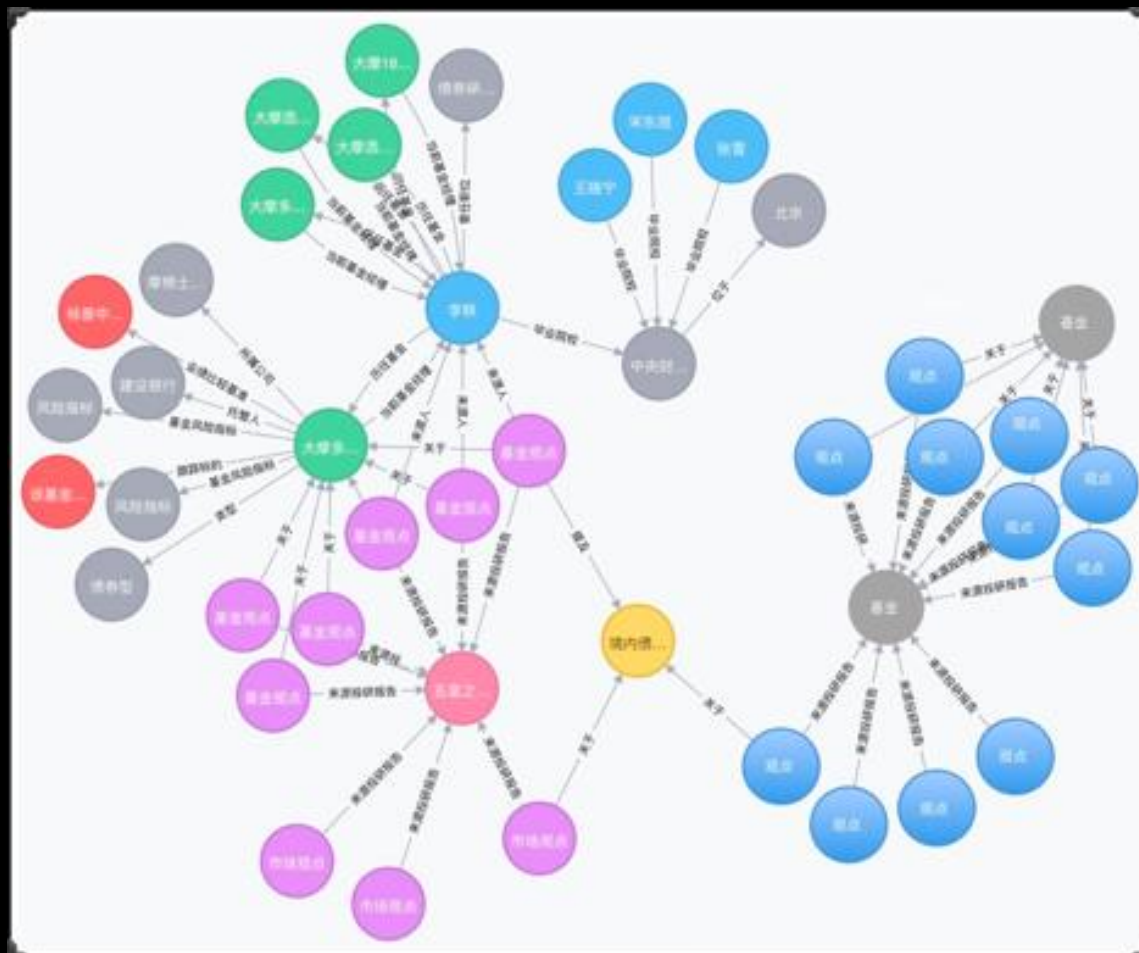


“跟着我 左手右手 一个慢动作右手 左手 慢动作重播...”

ruyi.ai: 面向模板基于机器学习的KBQA



基金KG应用（正在进行）



- 查找并对比基金的属性
- 查找基金概念的定义，并对比之
- 研报报告关联到“基金”，“基金经理”
大摩多元收益债券A怎么样？
它的基金经理采用什么投资策略？
- 发现基金相关实体之间的关系
哪些基金被中财毕业生管理？
债券型基金的基金经理大多从哪里毕业？

业务知识

知识建模-架构层

结构化知识图谱

意图构建-服务层

问答模版

满足需求

知识图谱结构辅助意图理解

基金.属性

* 用户说 ?

@kg.基金:entity1 和 @kg.基金:entity2 的 @基金属性:property 相比怎么样

基金知识图谱->基金实体

基金知识图谱->基金属性

实体数



实体数



属性数



问答模版数



问答对

范例: 6000基金实体

6000基金实体

100基金属性

10模版

三百六十亿



海知智能
ruri.ai

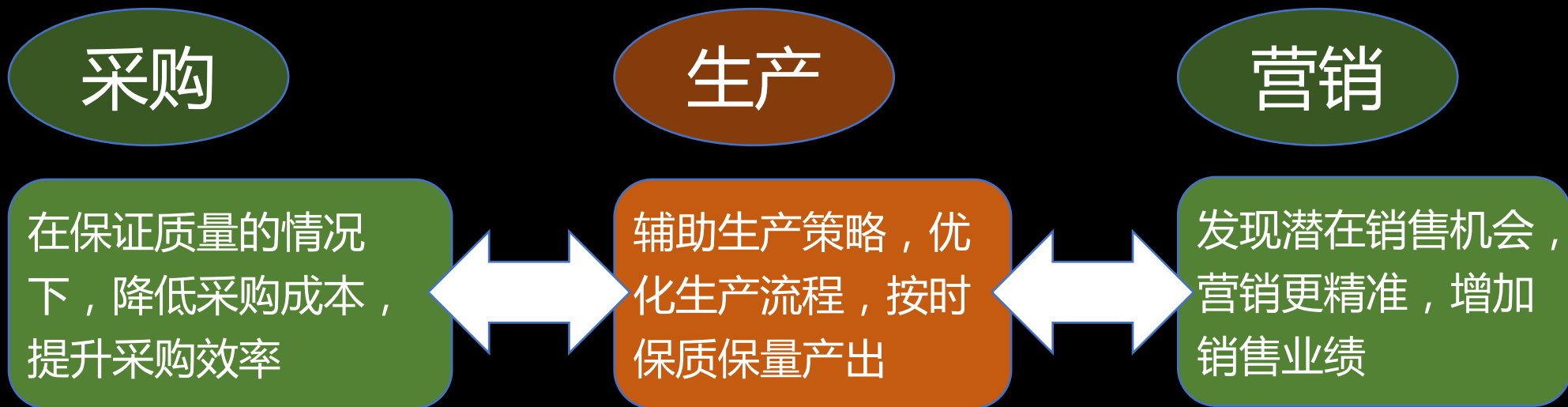
ruyi.ai产品线-2

水晶球

智能分析员
botAnalyst
V 0.9 beta

- ① 结构化、半结构化数据，非结构文本采集
- ② 自然语言理解形成领域知识图谱
- ③ 结合量化模型形成预测、判断、风险预警和线索发现等辅助决策支持

制造业智能分析员



招采智能分析员

数据采集

- 数据爬取和清洗原始数据
- 广覆盖面
- 高更新频率

信息处理

- 关联多源数据进行数据融合
- 多重渠道交叉验证数据可信度
- 建立并形成数据信息模型

知识组织

- 实体关系识别和可信度计算
- 领域本体设计建模
- 形成领域知识图谱

决策支持

- 价格判断：判断现有选择（供应商）价格的合理性，给出适当的价格采购区间
- 价格预测：根据各方面各层级影响因素，预测可能的价格走势以及采购策略建议
- 供应商发现：筛选推荐适合的潜在供应商
- 风险预警：在全景分析基础上，对于价格及相关影响因素的变化，进行风险预警



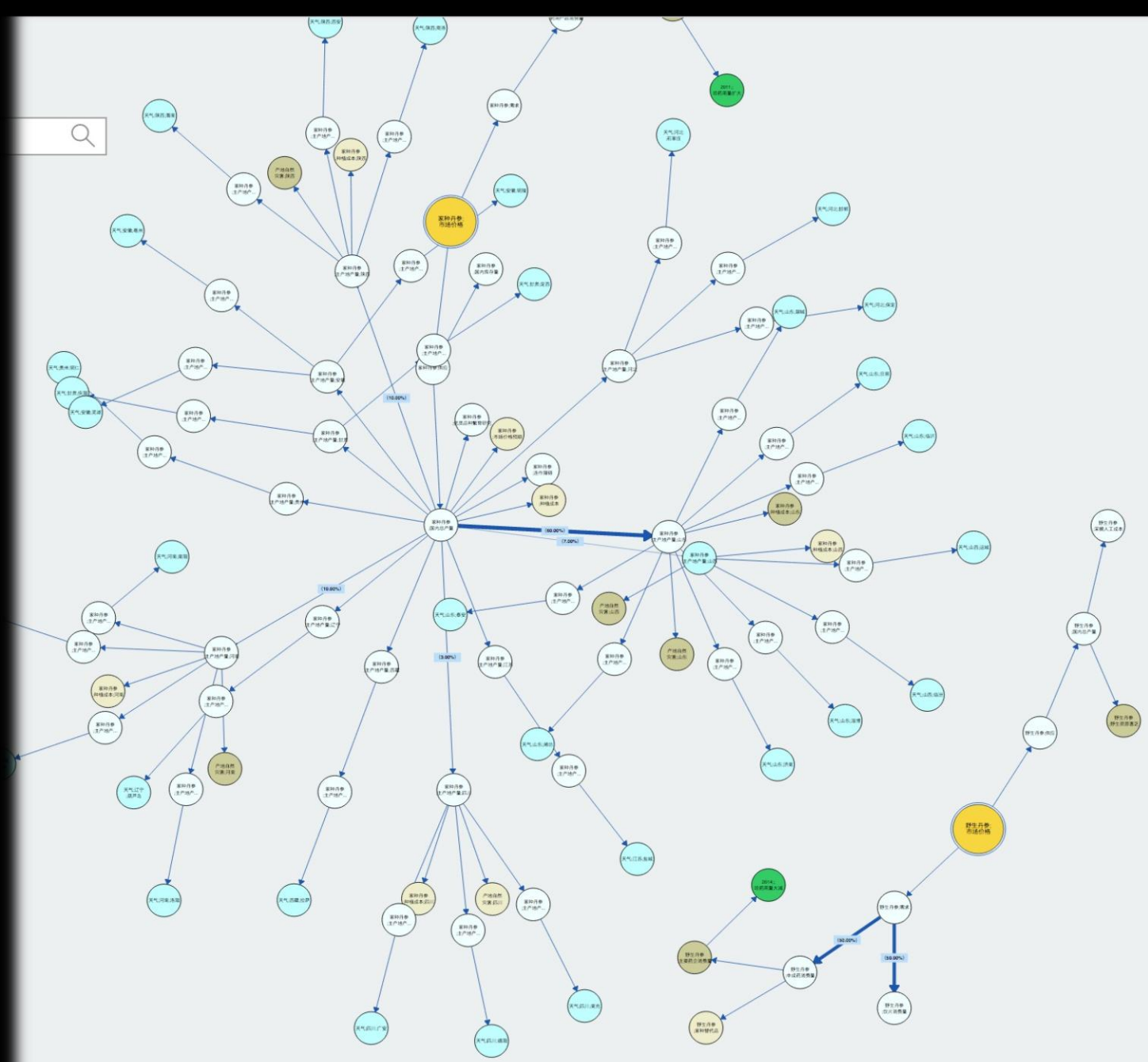
智能招采分析员能力举例



HDGC 招采智能化系统

新闻



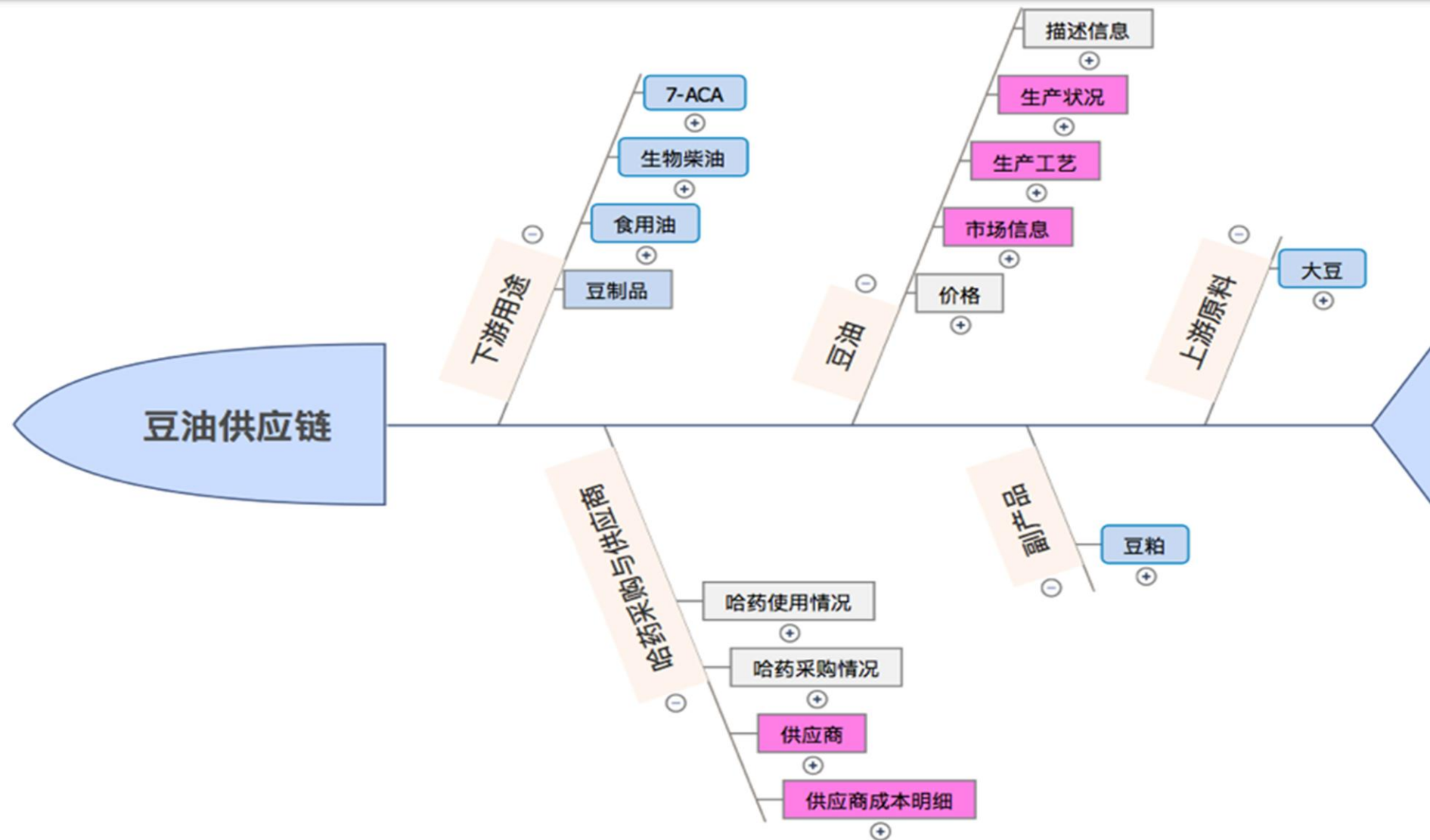


图例

- 物料
- 市场信息（有详情）
- 市场信息（暂无详情）
- 历史预警信号（有详情）
- 历史预警信号（暂无详情）
- 需关注的预警信号（涨价趋势）
- 需关注的预警信号（降价趋势）
- 需关注的预警信号（未知趋势）

两个节点之间关系的强度体现为连线粗细。关系越强，数字越大，线越粗。

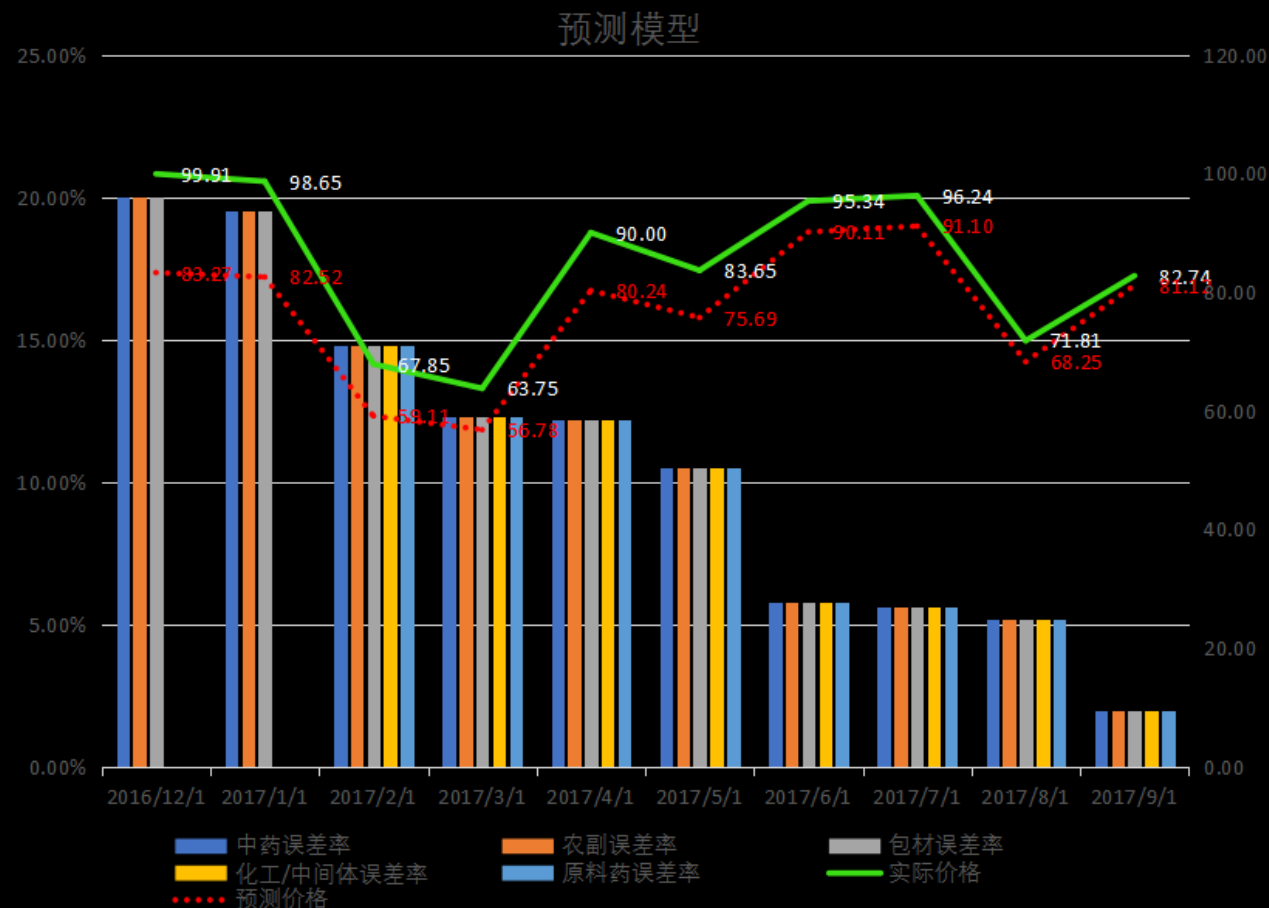
招采智能分析员- 7ACA与大豆



招采智能分析员——价格预测

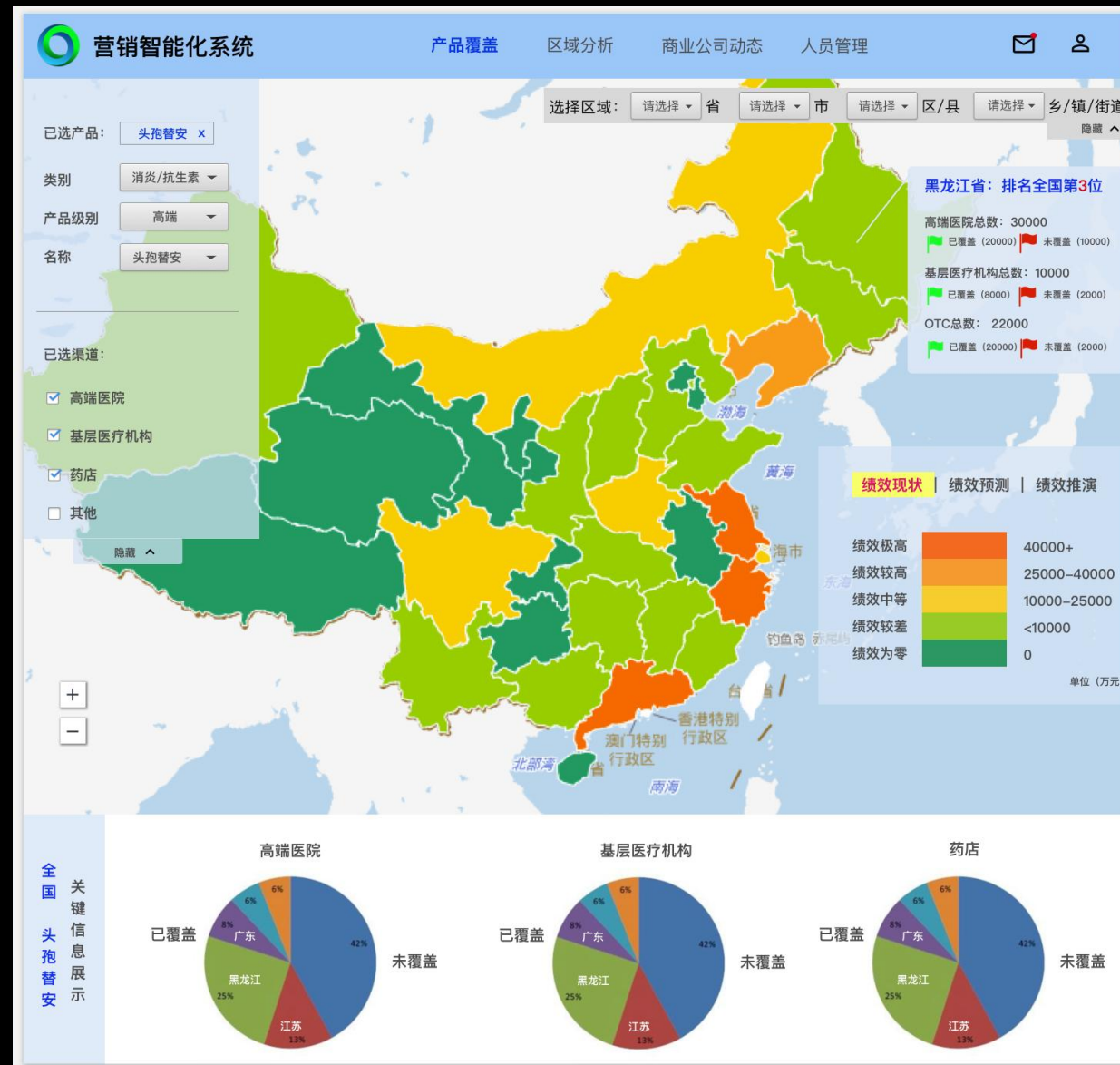
•支持128种物料市场价格的预测

•80%以上的物料预测误差
≤5%



营销智能分析员-基于地理信息智能分析

1. 选择关注的产品
2. 选择关注的渠道
3. 区域及数据鸟瞰
4. 一览无余的绩效现状
5. 关键信息聚合及排名



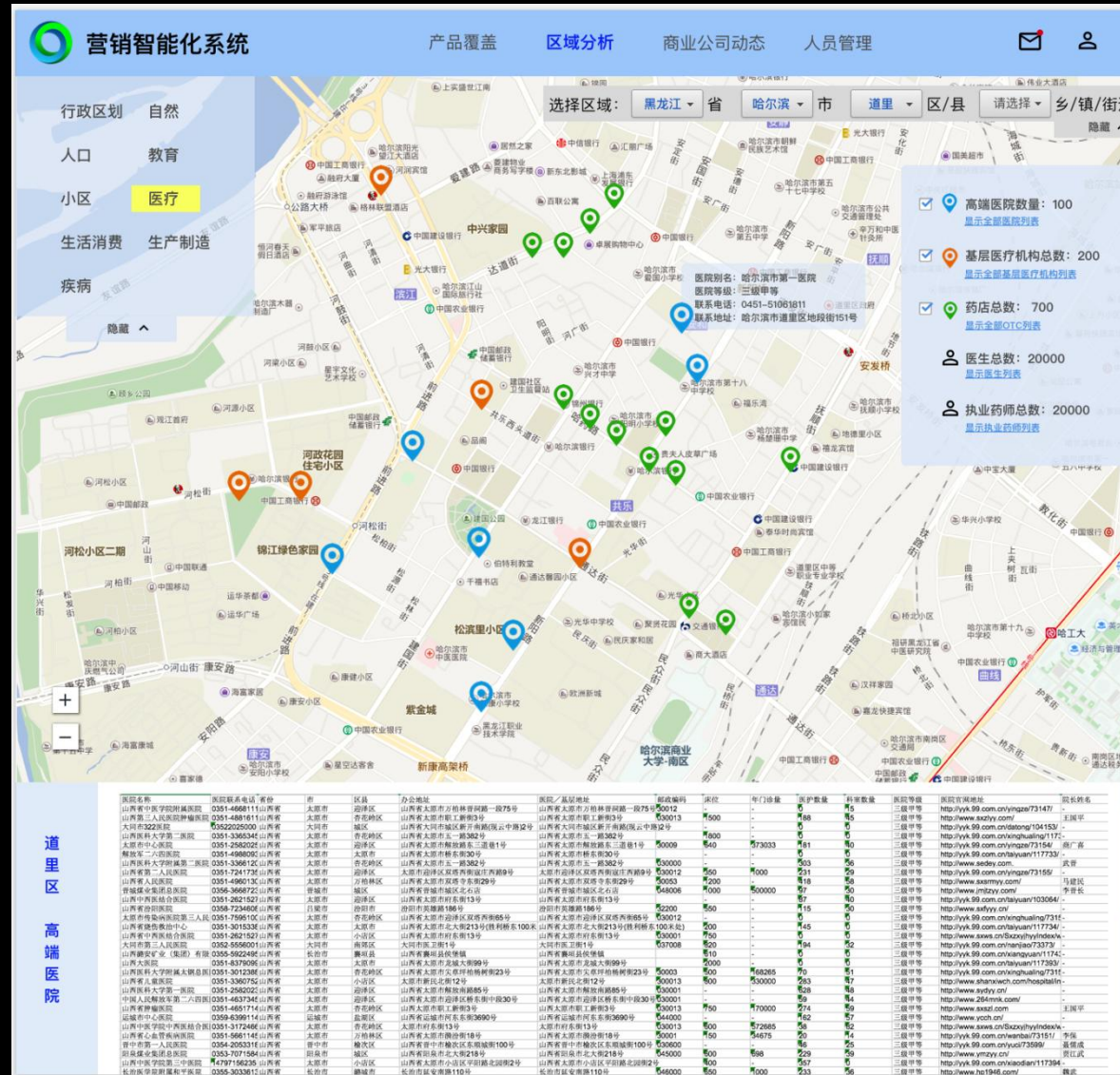
营销智能分析员-基于地理信息智能分析



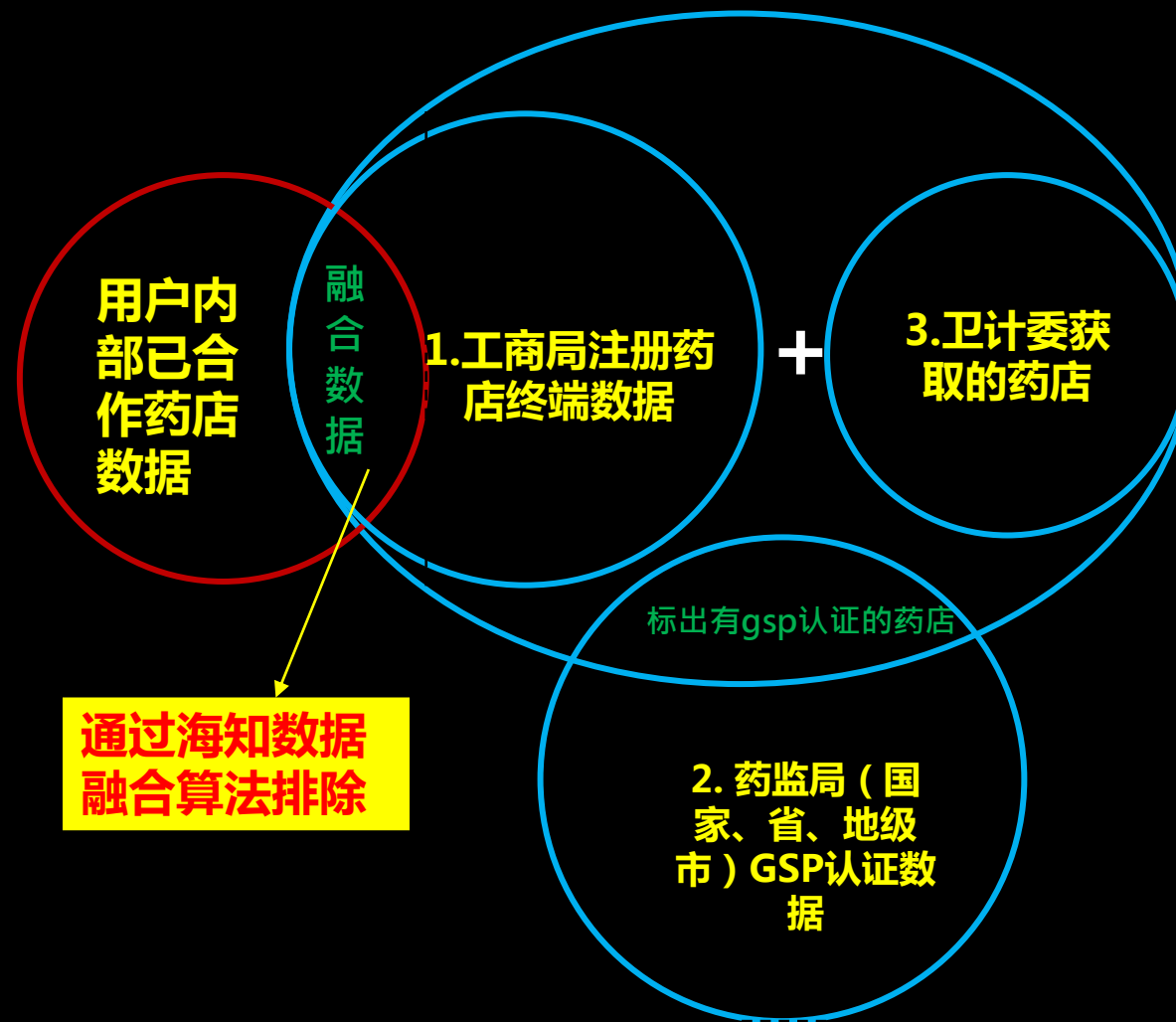
1. 详细到区县级别的数据总览
2. 根据现有数据得出未来周期内的绩效预测结果
3. 根据目标绩效作出推演并得出合理建议



区域的详情信息展示可以提供用户全面的区域情况，来帮助用户做出合理的产品营销策略。



营销智能分析员-更多药店的数据源拆解



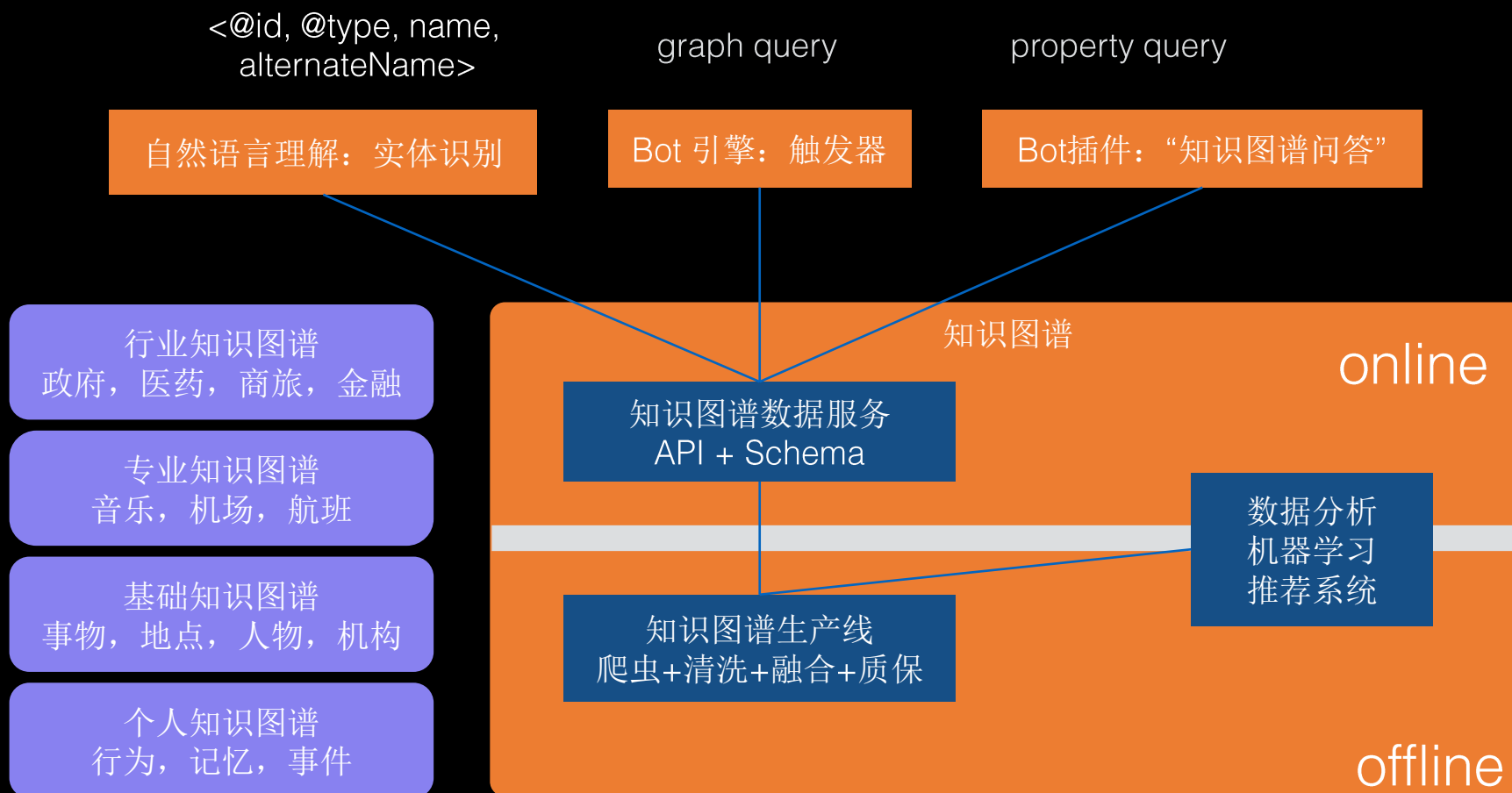
ruyi.ai产品线-3

聚宝盆

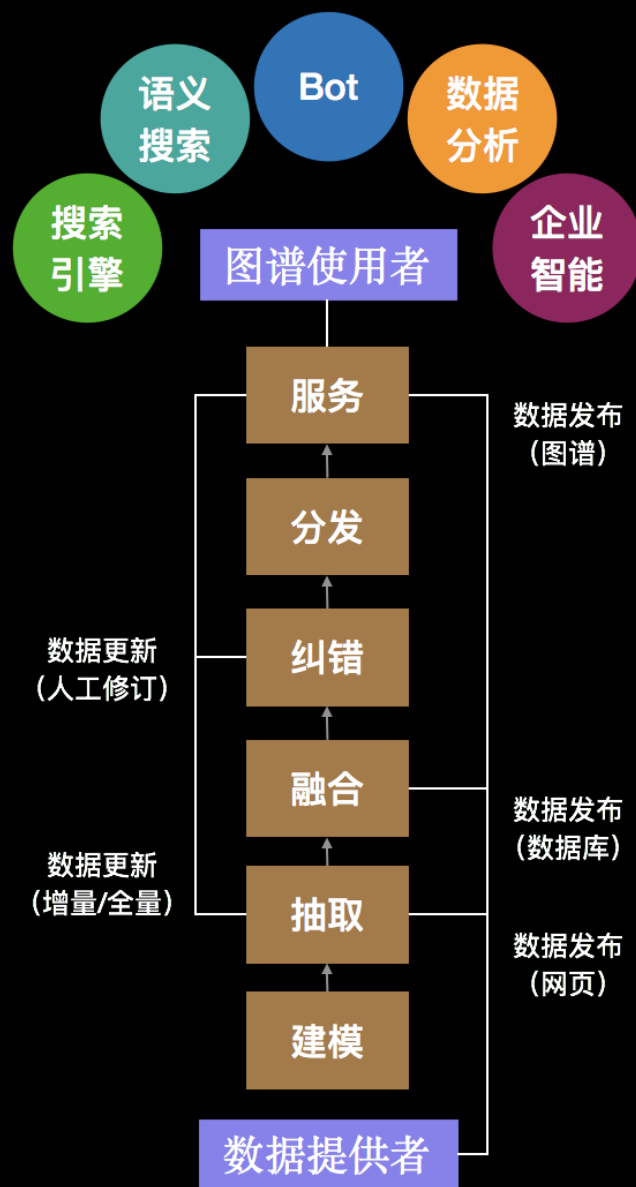
知识图谱生产线
KnowledgePot
V 0.1 alpha

- ① 人工&半自动化 构建领域知识图谱
- ② 覆盖相对全面 质量相对可靠 更新比较及时
- ③ 医药行业，金融领域，法律服务

kg 体系



KG生产线



- **coverage** , 全面覆盖领域知识 , 多数据源融合互补 , 关联领域相关实体
- **fresh** , 数据内容的定期更新 , 新词新知识的发现
- **quality** , 多来源数据交叉验证正确性 , 利用人机协同纠错保障图谱质量

OpenKG简介

中文开放知识图谱（简称OpenKG.CN）旨在促进中文知识图谱数据的开放与互联，促进知识图谱和语义技术普及和广泛应用

开展活动和业务

1. 培育开放社区
2. 组织社区活动
3. 构建开放域连接知识图谱
4. 与OMAHA开放医疗与健康联盟合作
5. 起草面向 bot 的中文知识图谱标准 cnSchema

部分发起机构

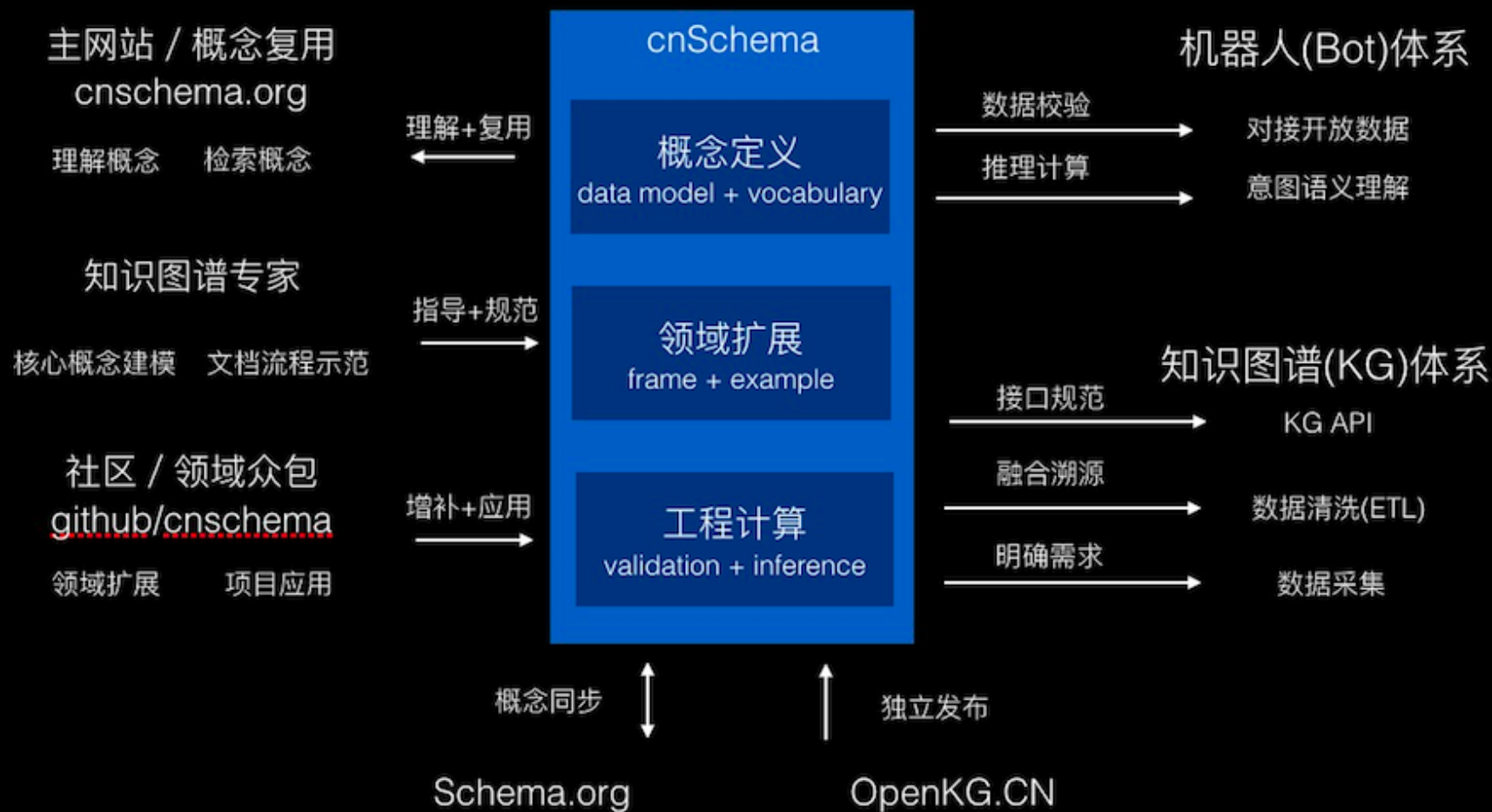


已经聚集的部分知识库

- ★ Zhishi.me : 近1000万实体，10亿三元组
- ★ CN-DBPedia : 近1000万实体，8000多万三元组
- ★ XLORE双语百科知识图谱：1000万实体
- ★ Belief Engine : 5000万三元组
- ★ PKU-PIE知识库：70万个实体，900万条三元组
- ★ ConceptNet中文部分：2800万三元组
- ★ WikiData中文部分：60万Items



cnSchema生态体系



bot：超越商业



NPI恩派



海内知己，因 AI如意

使命：人人会做bot
愿景：人人能有bot