

金融知识图谱的发展近况 和应用回顾

Jie Bao
Memect CEO
baojie@memect.co
2018-08-17, CCKS



March 2016, First workshop (Beijing)



attendees: 30+

March 2017, FinKG Hangzhou



金融知识图谱论坛

Financial Knowledge Graph Forum (FinKG)

2017.03.29 杭州



组织方：中文信息学会
语言与知识计算专委会



承办方：浙江大学



承办方：文因互联

FinKG 2017 统计

270+ 报名者

150+ 参与单位

2 家交易所

7 家券商

10+ 银行

10+ 家大学

10+ 智能金融从业公司

10+ BAT等大型互联网公司

30+ 投资机构

FinKG 2018

计划中（第四季度，地点待定）

触达 → 重构

（互联网金融）

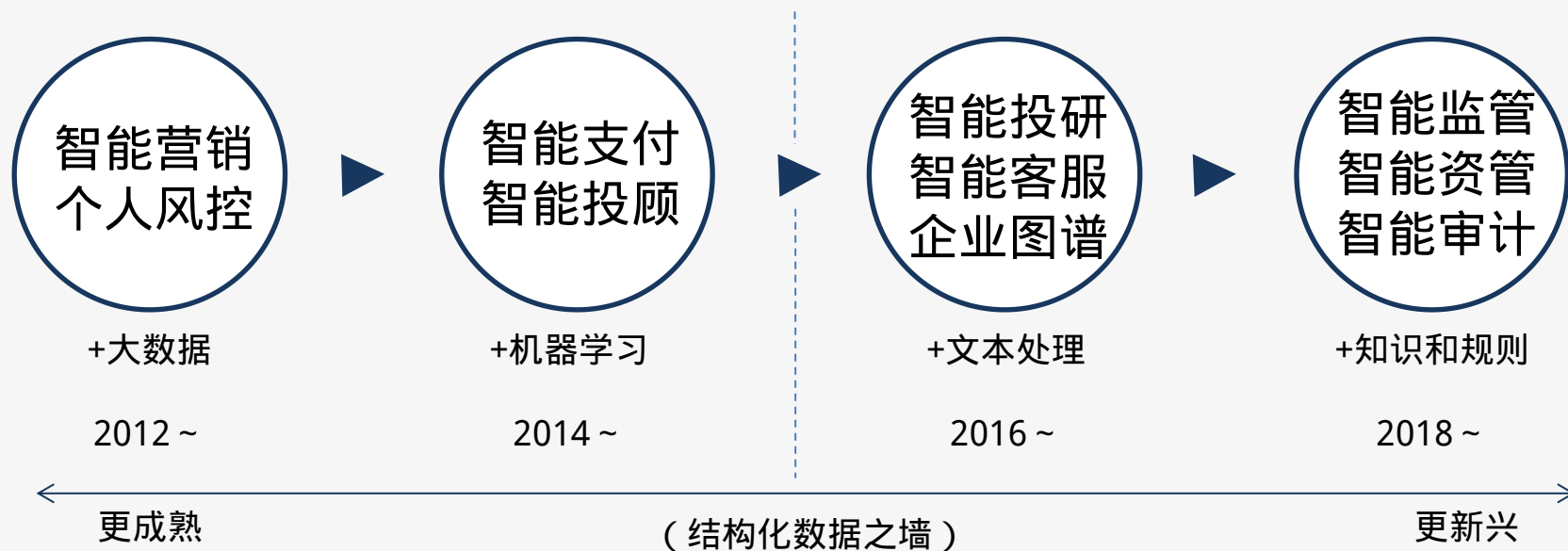
（智能金融）



智能金融行业现状和应用领域

智能金融行业于2015年开始在中国起步，仍然处于发展早期。按成熟度可分为：

- 1) 智能营销、面向个人的智能风控，主要依托大数据；
- 2) 以算法驱动的智能支付（利用图像处理的身份识别）、智能投顾（资产配置）；
- 3) 利用多种AI技术（特别是文本处理）的智能投研、智能客服、企业图谱的银行应用（获客、风控）；
- 4) 新兴场景，如智能监管、智能资管、自动化审计等，还处于市场启蒙期。



行业现状和应用领域

根据亿欧《2017中国智能金融产业研究报告》，截止2017年中国智能金融公司有160余家，累计融资**250亿人民币**，其中智能营销占37%，智能风控（主要是面向个人数据）占26%，智能支付占19%，智能投顾占13%，以上（**即前段1、2两类**）**合计占95%**，竞争已很激烈。3、4两类还处于起步阶段，市场处于蓝海。

结构化数据的缺失是当前智能金融深化和应用扩展的**核心瓶颈**，特别是3、4类应用发展壮大的先决条件。

金融业务自动化的前提是有丰富的结构化数据，进一步加强智能风控、营销和投研也不能仅依靠现有的结构化数据，必须利用隐藏在非结构化文本数据中的大量数据。**知识图谱**是数据结构化主要的技术实现形式，

金融知识图谱 (2016)

- 创投类数据库；

金融知识图谱（2017）

- 创投类数据库；
- 公众公司（A股和新三板等）基本面数据、行情数据；
- 公告数据提取；
- 研报数据提取；
- 泛舆情数据（如新闻、诉讼）
- 泛企业数据（如工商）

金融知识图谱 (2018)

- 创投类数据库；
- 公众公司（A股和新三板等）基本面数据、行情数据；
 - 机器财报提取已达人类准确度（>99.5%）
 - 监管企业画像
 - 高管画像
- 公告数据提取；上交所已上线，深交所、股转（新三板）执行中
- 研报数据提取；
 - 产业链图谱
- 泛舆情数据（如新闻、诉讼）
 - 价格图谱，竞争情报图谱，基于的KG的负面新闻
 - 债券风险预警
- 泛企业数据（如工商）
 - 中小企业风控图谱
 - 营销获客图谱

应用（2016）

- 信用监控（Garlik）
- 数据集成（摩根大通）
- 风控（宜信，普惠）
- 财务建模（富国银行）
- 财务报表自动化（XBRL技术相关）
- 语义搜索（FIBO工作组）

应用（2017）

- 传统数据终端++
- 金融搜索；
- 金融问答；
- 公告、研报摘要；
- 个人信贷反欺诈；
- 信贷准备自动化；
- 信用评级自动化；
- 自动化报告；
- 自动化新闻；
- 自动化监管和预警；
- 自动化审计；
- 法规和案例搜索；
- 自动化合规检查；
- 产业链自动化分析；
- 跨市场对标；
- 营销和客户推荐

应用（2018）

- 传统数据终端++;
- 金融搜索;
- 金融问答;
- 公告、研报摘要;
- 个人信贷反欺诈;
- 信贷准备自动化;
- 信用评级自动化;
- 资管投研
- 资管流动性管理;
- 金融法规和案例搜索;
- 银行业数据治理;
- 证券业数据治理;
- 研报智能搜索;
- 报告图表搜索;
- 银行业务流程管控;
- 自动化报告;
- 自动化新闻;
- 自动化监管和预警;
- 自动化审计;
- 法规和案例搜索;
- 自动化合规检查;
- 产业链自动化分析;
- 跨市场对标;
- 营销和客户推荐;
- 债券风控;
- 财务自动化复核和诊断;
- 机器人流程自动化;
- 保险客服机器人;
- 基金客服机器人;
- ICO评级;
-

文因互联服务客户



某证券交易所



某大型会计事务所



金融知识图谱技术架构

知识提取

使用自然语言处理、机器学习、模式识别、信息论等诸多人工智能技术，从大量复杂、异构、多媒体原始数据中进行表格识别、实体提取、关系发现等知识提取的任务。

知识表现

处理结构化和非结构化数据混合的知识表示与存储，拥有自主开发的推理机和基于百科的常识知识库和行业知识库，采用优化的表示方法以降低整个工程周期的成本。



知识融合

基于实体链接，实体消歧，精益本体构建等技术手段，将来自不同数据源知识快速融合，及时更新，发现存在冲突的节点，提升图谱的一致率。

知识检索（应用方向）

由关键词、词联想、词本体、短程关系、长程关系、自由问答构成的六级知识检索技术体系，可以有效过滤信息噪声，为用户提供信息的友好交互。

金融知识图谱产品典型技术架构

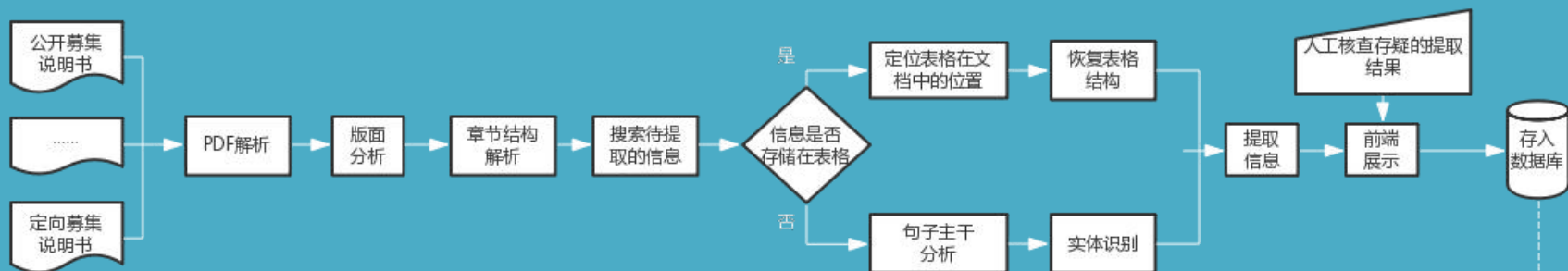


金融图谱数据（文因云）

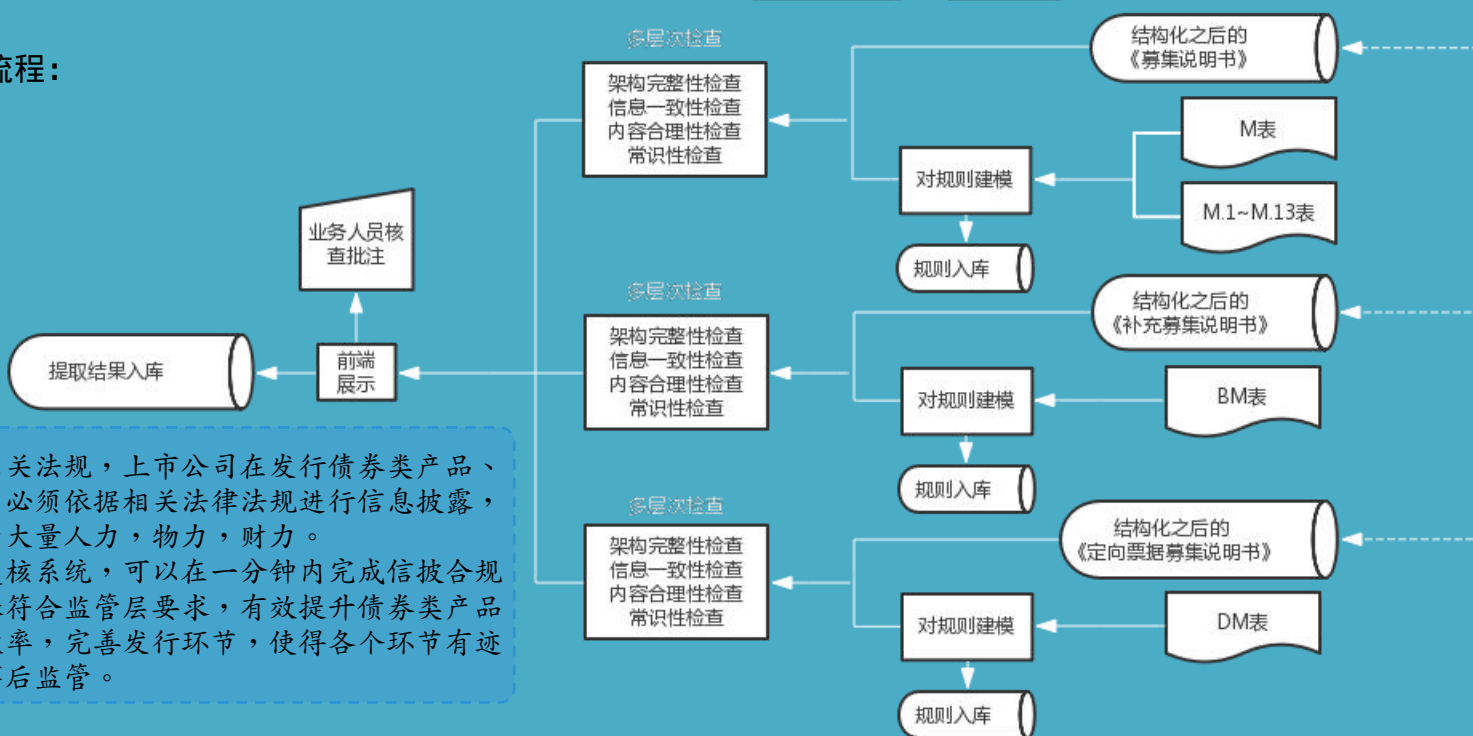
模块分类	模块详情概览			
公司基本信息	基本信息概览	团队与核心成员	股东持股信息	股权变更历史
	参控股公司关系	资本市场情况	财务与估值指标	销售生产资质
	产品结构分析	知识产权一览	技术与研发能力	销售模式
	同业企业对比	客户与市场	公司新闻	高管最近讲话
公司财务分析	关键审计事项	前五欠款人	五大供应商	利润表分析
	现金流量表分析	资产负债表分析	损益表分析	
	主营构成	不良贷款率		
行业分析	行业竞争	市场规模	政策环境	产业链分析
	行业增长速度	行业上下游	行业重要会计政策	
研报摘要	公司研报摘要	宏观研报摘要	龙头企业	投资观点
	目标价格一致性预期			
公告摘要	股东大会/董事会决议	召开股东大会通知	利润分配实施公告	停、复牌公告
	新股/限售股上市	名称及简称变更	暂停/恢复上市	业绩预测
	融资融券	子公司担保	其他担保	重大行政处罚
	提供反担保	仲裁事件	政府补助	股东增持计划
	新建项目	重大损失	人事变动	高新技术企业
	股权激励计划	抵押借款	签订战略框架协议	诉讼事件
	股东减持	出售资产	签订日常经营合同	计提大额资产减值准备
关系数据	定向增发	限售股解禁	股权激励回购注销	出售资产
	产量销量	股份被质押	股份质押解除	股权质押展期
	特殊交易事项	重大处罚	股东增持计划完成	
	高管关系数据	产业链关系	公司产品关系	公司和行业关系
基础技术服务	研报搜索	公告搜索	图表搜索	财报搜索
	新闻标签服务	自动化报告引擎	问答引擎	

金融报告自动化复核

文本结构化处理：



合规性检查流程：



依据证监会相关法规，上市公司在发行债券类产品、定向增发时，必须依据相关法律法规进行信息披露，人工审查耗费大量人力，物力，财力。通过自动化复核系统，可以在一分钟内完成信披合规性检查，确保符合监管层要求，有效提升债券类产品的注册发行效率，完善发行环节，使得各个环节有迹可循，方便事后监管。

法规智能检索系统

知识检索系统

关键字检索

语义标签检索

分面浏览器筛选

文本分类系统

规则分类引擎

有监督学习的分类引擎

无监督学习的热词生成

知识提取系统

篇章识别

正文提取

关键词抽取

关系抽取

内容脱敏

数据

数据

数据

大规模分布式爬虫

法规智能检索系统



法规宝

输入关键词



☒ 查案例

☐ 查法规

☐ 讨论区

提问

- ☐ 全部标签 精神损害 **合同** 利息 **买卖合同** **传票** 债权 给付 交付 保险合同 贷款 处分 违约责任 驳回 精神损害 清偿
交通事故损害赔偿 保证 缺席判决 **传唤** 清偿 合同约定 民间借贷 债权人 违约金 债务人 残疾赔偿金 交通事故 借款合同
- ☐ 发布部门 中央 四川 北京 贵州 上海 天津 甘肃 山东 安徽 江西 新疆 江苏 广东 河南 广西 浙江 吉林 湖南
云南 辽宁 山西 西藏 重庆 河北 宁夏 青海 海南 福建 陕西 黑龙江 内蒙古
- ☐ 全部来源 证监会 银监会 保监会 微信公众号
- ☐ 全部时间 2018 2017 2016 2015 2014 **2018.01.12—2018.06.12**

收起 ▲

中国证监会行政处罚决定书（文高永权、王交英、宋翼湘等4名责任人员）

众益传媒

民间借贷

合同

2018-06-15 13:21 银监会

关联法规：《证券法》第二百零三条

关联法规：《证券法》第七十七条第一款第（一）项

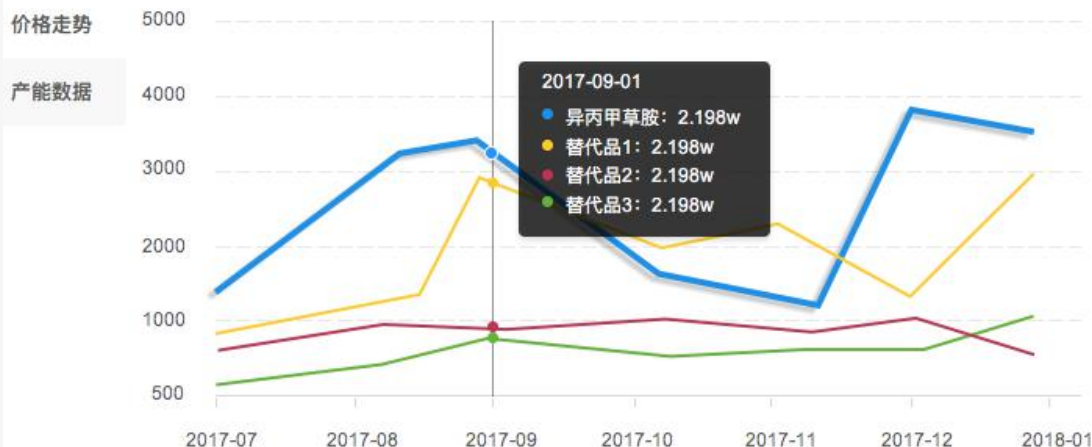
研报智能检索与管理



行业智能决策情报

- **产业链研究：**从采购的角度来看，上游产业波动，会传导影响企业利润率波动。该系统能够帮助企业实现智慧采购。
- **利润率关系研究：**从生产的角度来看，产量、成本、利润率三者都会影响到产品的价格以及市场前景。该系统分析同行业其他企业的量、本、利状况，有利于未来对产品布局提供支撑。
- **供需关系研究：**从销售的角度来看：市场容量、供应量、存量都会影响到销售策略，该系统自动计算其中的关系及波动规律、各因素之间的相关性，可以减少公司定价政策的盲目性。
- **行业数据的提取：**数据来源较为繁杂，仅靠手动导入汇总，效率较低，该情报系统实现了信息数据联动和实时信息查询。

市场数据



变动解读

供给（环保影响）下降厉害，造成恐慌；抢货

相关登记证

[查看全部](#)

登记名称	类别	生产企业
精异丙甲草胺	水剂	浙江禾本科技有限公司
异丙甲草胺	可湿性粉剂	山东潍坊润丰化工股份...
精异丙甲草胺	水剂	浙江禾本科技有限公司
异丙甲草胺	可湿性粉剂	山东潍坊润丰化工股份...
精异丙甲草胺	水剂	浙江禾本科技有限公司
异丙甲草胺	可湿性粉剂	山东潍坊润丰化工股份...
精异丙甲草胺	水剂	浙江禾本科技有限公司
异丙甲草胺	可湿性粉剂	山东潍坊润丰化工股份...
精异丙甲草胺	可湿性粉剂	浙江禾本科技有限公司
异丙甲草胺	水剂	山东潍坊润丰化工股份...
精异丙甲草胺	可湿性粉剂	浙江禾本科技有限公司
异丙甲草胺	水剂	山东潍坊润丰化工股份...

企业综合评价系统

从财报中自动抓取财务指标，然后根据评分体系，对每个指标进行自动评分。分析维度包括偿债能力、经营效率、盈利能力、成长能力、现金流水平等。

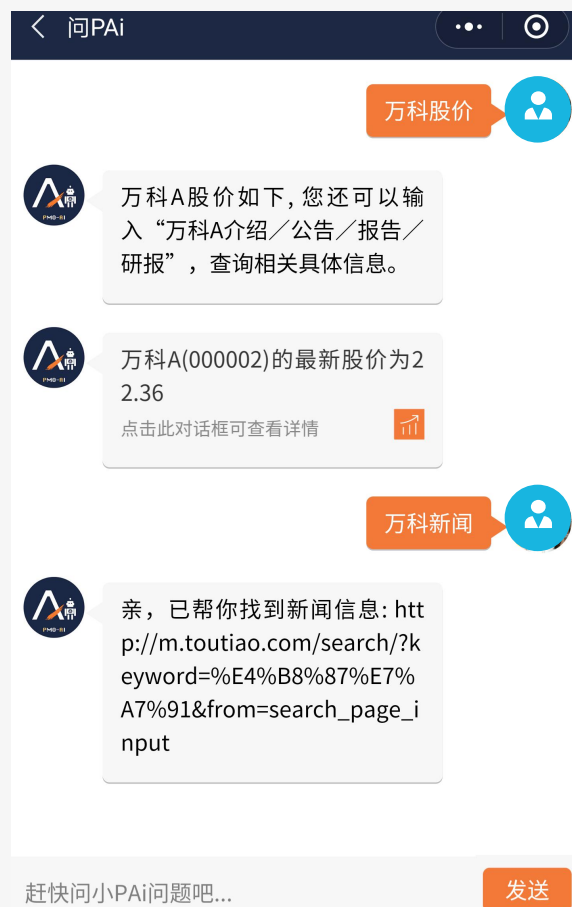
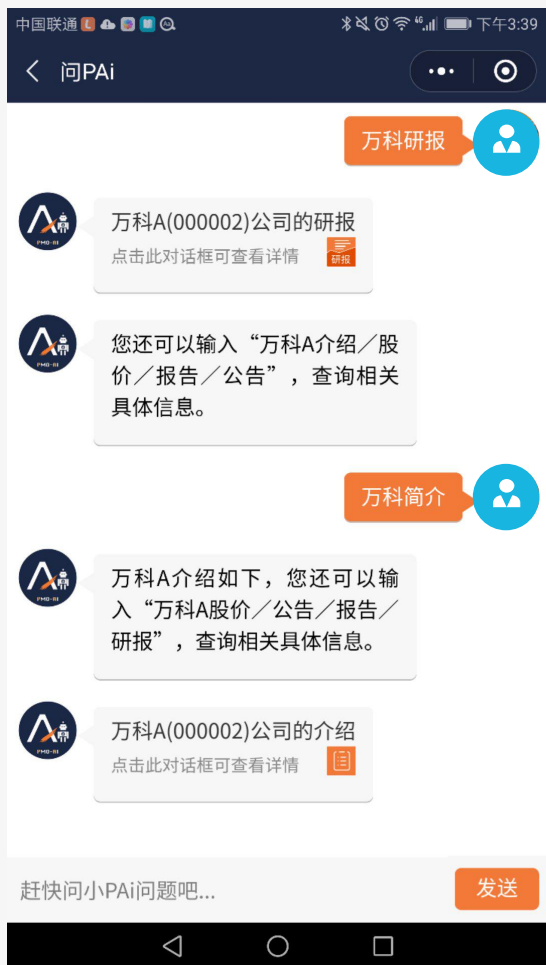
效果示例：



小微企业信贷审批流程自动化

分析场景		内部数据	难点	结合外部数据分析
企业经营分析	产品情况	<ul style="list-style-type: none"> • 采购与销售情况 • 产能与成本情况 	<ul style="list-style-type: none"> • 市场占有率如何判断 • 同业毛利能力对比 • 政策保护情况如何 • 技术先进性如何 	<ul style="list-style-type: none"> • 对比同业公司情况 • 通过招投标分析现金流能力 • 企业研发与招聘、专利等是否匹配 • 相关政策的关联
	供应链情况	<ul style="list-style-type: none"> • 供应商与客户名称 • 供应与销售业绩 	<ul style="list-style-type: none"> • 供应链上企业是否能持续发展 • 供应链上企业是否是存在强依赖 	<ul style="list-style-type: none"> • 通过诉讼信息统计欠款情况 • 类似的供应链上企业的风险判断
	核心团队分析	<ul style="list-style-type: none"> • 高管基本信息 • 不良信用记录 • 借款人关系 	<ul style="list-style-type: none"> • 高管对外投资的情况 • 高管隐性行为的跟踪 	<ul style="list-style-type: none"> • 投资企业的情况如何 • 通过运营商数据等推测风险行为
行业趋势分析		同类企业同比营收增长	<ul style="list-style-type: none"> • 细分行业如何切割与描述 • 财务信息容易粉饰 • 上下游关联分析 	<ul style="list-style-type: none"> • 关联上下游企业情况 • 招聘需求趋势 • 同类新企业增加速度 • 相关专利、商标、著作权情况 • 行业在新闻中曝光程度
同业授信分析		通过客户的跨行贷款信息	<ul style="list-style-type: none"> • 关联公司的贷款信息聚合 • 跨行信息的整合统一分析 • 区域性分析 	<ul style="list-style-type: none"> • 整合关联公司的同业授信数据分析 • 统计识别同业中的不良信贷模式

问答引擎为某大型银行提供微信智能系统



金融工作组计划

目标1 金融知识图谱标准化

- 配合cnSchema中关于金融的schema的标准化
- 组织专委会与监管机构间的沟通，积极推进知识图谱在银行业数据治理、证券业数据治理、监管3.0中的应用。
- 推动和XBRL、DAMA中国等的合作
- 推动加强和W3C的合作，特别是FIBO工作组、金融支付工作组等的合作。

金融工作组计划

目标2 活动组织

- 与OpenKG合作，每年一次 FinKG（金融知识图谱论坛），规模目标300-500人。
- 每个季度一次轮流在北京、上海、杭州、深圳等主要的金融业中心举办金融知识图谱沙龙，规模目标100人左右。
- 不定期针对不同的目标人群，如金融机构、投资机构、学生、政府、一般科普对象等，参与或举办小型的金融知识图谱相关的活动。

金融工作组计划

目标3 普及和工具服务

- 在成员单位现有工作基础上，推出证券业基础数据的浏览器，如年报数据、公告数据、事件数据等。
- 提供一些基础的金融实体提取展示性工具（文因互联参与人为首席科学家郑锦光）、金融事件提取展示性工具。
- 在openbot框架下提供金融基础问答的展示工具（与OpenKG合作）。



个人微信 baojie_memect

欢迎参与 金融KG工作



主页: memect.cn

微博: weibo.com/wenyinhulian

邮箱: contact@memect.co

