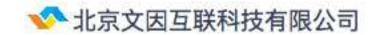
从数据到智慧 知识图谱如何推动金融更加智能

鲍捷

邮箱: baojie@memect.co 微信 baojie_memect





基本介绍

个人简介



- 1998年以来从事人工智能,聚焦在语义网 / 知识图谱
- Iowa State -- 博士学习,逻辑推理
- RPI -- 知识库构造 和 OWL 语言
- MIT -- 证券报表语言语义建模
- SAMSUNG -- 手机语音助手
- 文因互联 -- 智能金融

文因互联简介



作为国内领先的智能金融解决方案提供商,文因互联利用知识图谱技术,用机器阅读和提取文本中的知识和数据,整合金融机构内外部数据,从而高效利用监管机构、银行、券商等持有的海量数据,实现智能监管、智能审计、智能风控等场景的自动化。



提供服务简介:

文因互联提供高效、便捷、专业的自动化监管平台,帮助 客户解决在信息披露、财务复核,财务风险分析等工作中 数据量大、自动化程度低等问题,帮助客户提升工作效率。

文因互联利用领先的知识图谱技术,通过应用网络数据模型,提取行业信息、监管新规,交易数据、交易行为、关联关系等数据,通过对行业的理解、监管的逻辑、风控的原则等规则连接在一起,形成知识网络并且构建证券知识图谱。

创始人介绍



团队现有全职成员40人;硕士及以上占50%,其中拥有博士学位或学历者4人



鲍捷 CEO

创始合伙人

20年全面AI经验 RPI博士后,MIT、BBN研究 Iowa州立大学博士'07 三星问答系统S-Voice架构设计师 W3C知识图谱标准OWL作者之一 XBRL语义建模作者 OpenKG联合发起人



张强 CTO

创始合伙人

15年工程开发经验

美国Marvell资深系统专家

中国科学技术大学硕士 擅长高性能复杂系统架构设计

谷歌Chromecast摄放器软件 架构师(全球超5000万销量)

Google Home智能音箱软件系

统原型架构师

























文因互联主要荣誉



01 2017年获得北京市高新技术企业认证。



03 2017年1月6日,作为用金融+人工智能发起新变革的初创企业,文因互联与科大讯飞等5家企业,共同获得凤凰网-凤凰支持"2016年度科技创新影响力企业"荣誉称号。



CEO鲍捷博士任中国中文信息学会 语言与知识计算专委会金融知识图 谱工作组组长。



受邀与中国信息通信研究院等共同编写 《知识图谱白皮书(1.0版)》,作为业界 第一本介绍知识图谱主要技术内容和应用 场景的白皮书。



CEO鲍捷博士任中国计算机协会会刊 编委。



CEO鲍捷博士为中文开放知识图谱 联盟(OpenKG)发起人之一。





人工智能的一些背景

人工智能的三大分支



经验主义 (机器学习)

Data

proposition in the internal industrial industrial

"学习套路"

连接主义 (神经网络,深度学习) Input Hidden Output layer

Input #1

Input #2

Input #3

Input #4

"我认得你"

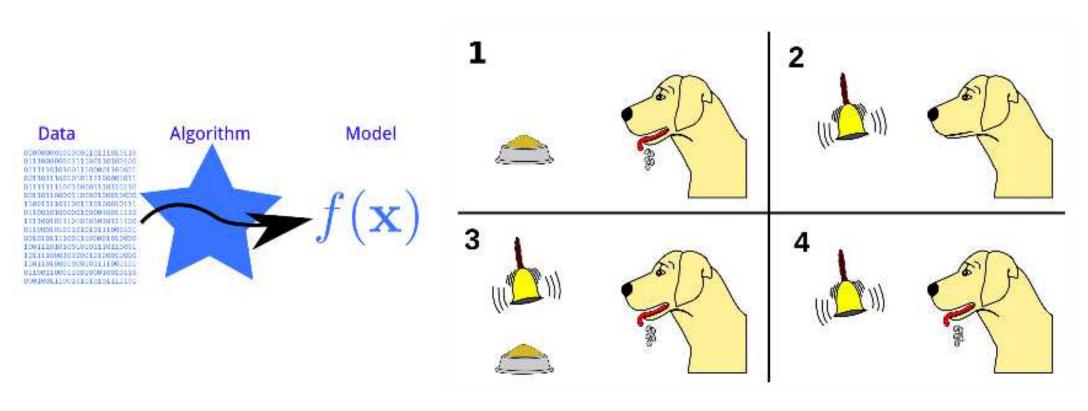
符号主义(逻辑,知识)

if all a are b
and all b are c,
then all a are c

"明白意思"

经验主义:"学习套路"

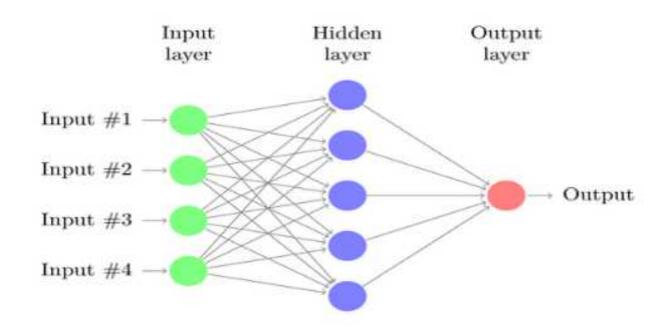


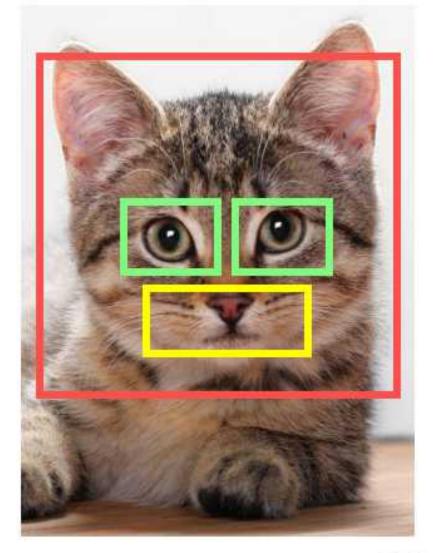


巴甫洛夫的"套路"

连接主义: "我认得你"



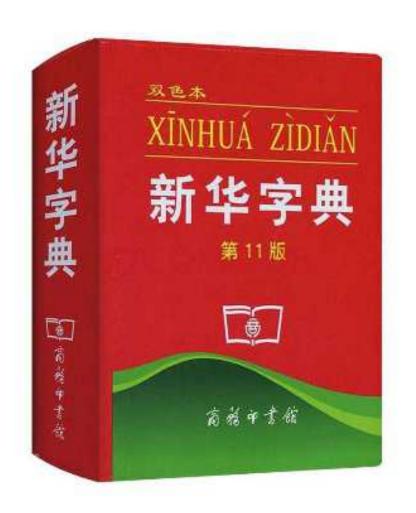




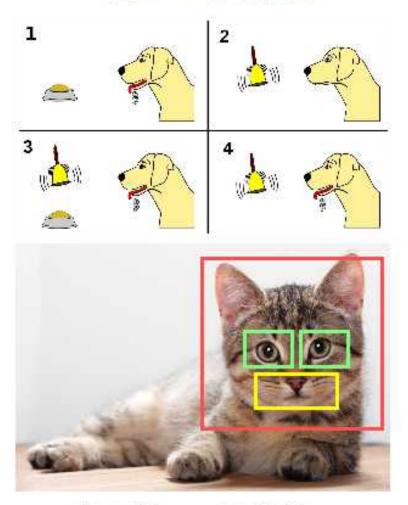
符号主义:"明白意思"



if all a are b and all b are c, then all a are c

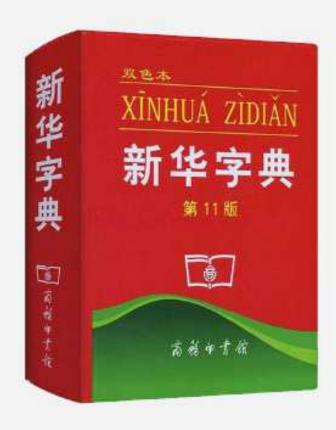


感知智能



机器学习、深度学习

认知智能



知识图谱、机器阅读

知识图谱



知识图谱(Knowledge Graph)是一种应用于诸多人工智能相关领域的关键技术,主要应用在数据结构化处理、解析、关联以及后续的分析与推理。



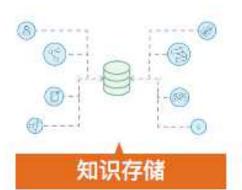
知识提取

利用自然语言处理、机 器学习、模式识别等解 决结构化数据生成问题。



知识表现和推理

组织结构化数据,使 得机器能够自动化处 理。



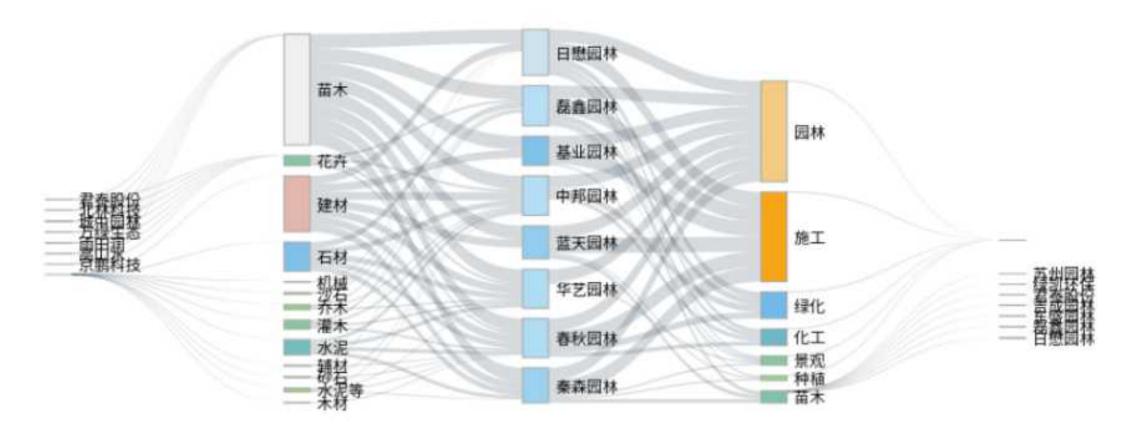
进行大量的结构化数据 管理,同时混合管理结 构化和非结构化数据。



用语义技术提高搜索与 查询的精准度,为用户 展现最合适的信息。

知识 = 结构





行业上下游



知识 = 结构



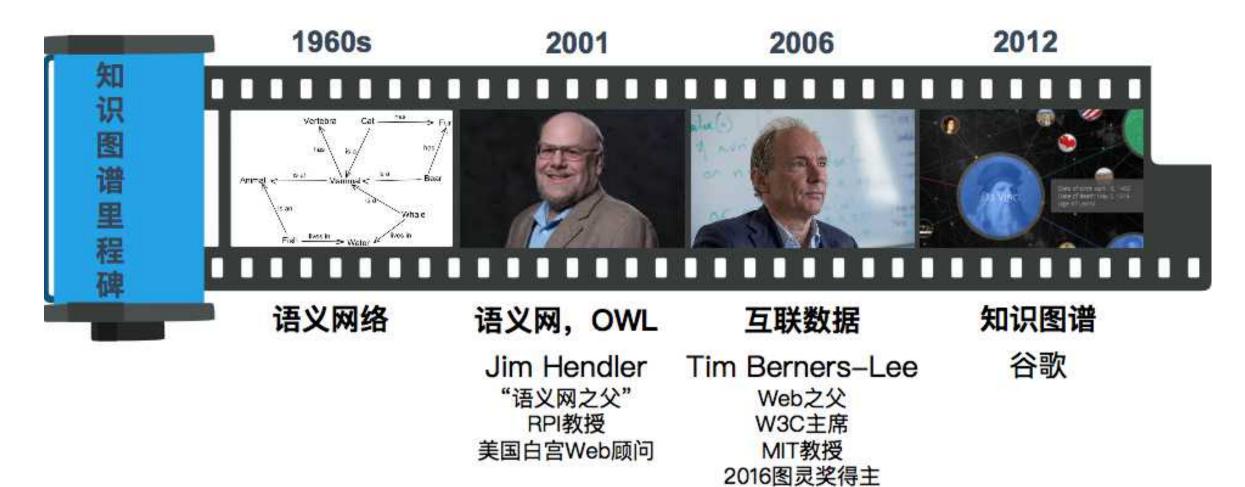
```
"日期": "2015-09-11",
 "摘要": "上海宝钢包装股份有限公司(以下简称"宝钢包装"或"公司")于
2015年 9 月 10 日收到公司独立董事孙业利先生提交的书面辞职报告,孙业利先
生因个人原因申请辞去公司独立董事及董事会专门委员会委员职务。",
 "职位":[
 "变动": "辞职",
 "名字": "孙业利",
 "原因": "个人原因"
```

公告中抽取的实体的描述和关系



知识图谱里程碑







金融科技发展趋势

Fintech上下半场

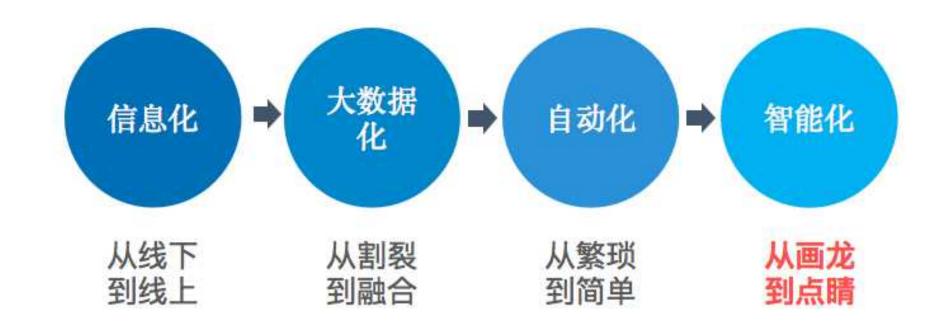
触达重构

(互联网金融)

(智能金融)

金融IT四阶段





智能金融现状









怎么往猫的脖子上面挂铃铛?





智能金融的本质是

金融信息处理过程的 (部分)标准件化







自动化,标准化,工业化





由标准化而工业化

现状: 金融农业, 金融手工业

农业、手工业、大工业





农业:从金融"自然资源"里刨数据,没有规模化扩大再生产的能力



手工业: 依赖于个人经验和人脉的金融决策 过程



大工业: 依靠可复制的技术和系统, 建立协作, 制造出千千万万种有质量保证的产品

金融大工业



大工业之所以能超越手工业

- 不在于制造出更精细的产品
- 在于可以大批量的标准化生产
- 在于通过协作和分工提高整体的生产效率。

在金融,就是不是超出市场收益率的投资决策, 而是建立**金融信息处理的协作系统**

趋势的大胆展望



- 机构知识资产化(5年内)
- 金融结构化数据高速公路(10年内)
- 可重复过程标准化,智力劳动大工业化(10-20年)



金融科技场景需求

金融数据治理的共性要求



全覆盖, 无死角

集成海量数据

元数据管理

非结构化数据结 构化 支持业务流程自 动化 支持智能分析

为什么需要KG: 以监管为例





阶段一 人工处理

人工获取数据



Microsoft Excel



阶段二 工作流自动化 合规软件促成自动化工 作流

Metric Stream





阶段三 持续监控 应用大数据推动后台业 务自动化







阶段四 智能分析 用人工智能主动识别和 预测风险





例: 从XBRL到KG





Squart SSC gav Q

PERSONALISM EDUCATION FLUNCS AHOUR DIVERSIONS **ENFORCEMENT** Structured Disclosure RSS Feeds 8 F 9 0 + O'ber of Whenless Decoure. Circles De-What it Szurzood Duty? Hittory and Rollersking. HAVE ABOVE TOWN LIBERT. Conserva-The SEC provides three Really Simple Syntication (RSS) leads for EDGAR structured disclanure automissions as a coursely for those invested it viewing, analyzing, and munipulating the submissions. These leeds are updated every ten minutes Monday. through Priday, 6am - 16pm EST. The content and format of these files may change without prior notice. Techencial These leads are take of the 100 most recent fillings for: 1. Corporate societ through minimum transpal steements buyger using the US GAAP XIVII. Issummy 2. Mutual fund flings containing the riskneturn summary tagged using the XBRL US Mutual Fund Riskneturn Summary 3. Himgs in both of the above 2 categories. The above RSS links may be opened in a web browser, but the resulting page will tack functionality. Rather, they are designed to be accessed with specialized RSS software products, which are widely available for free Each entity provides metadata (company name, accession number, CIK, filling date, form type, and the number), a link to the filling EDGAR, and a link to the interactive structured data files in a compressed (.zip) format. For a testigated auctive of all things with 2000, exhibits authorited to the 2010 beginning with the unseption of the violating program. in 2005, go in http://www.sec.goal/actives/enjormunity/. Each mainthly and the contains a complete set of all simulated data Hings received chang that month arrays all programs.



1

XBRL的技术局限



- 规范了数据的结构而不是数据的语义
 - · 概念 (concept) 难以继承和扩展
 - 关系 (link, relation) 无法推理
 - · 基于linkbase的数据引用非常受限
 - 难以表达稍微复杂的概念约束和算数约束
 - 扩展困难
- XBRL非常啰嗦,数据一致性保证很弱
- 繁琐, 枯燥



XBRL的实施局限

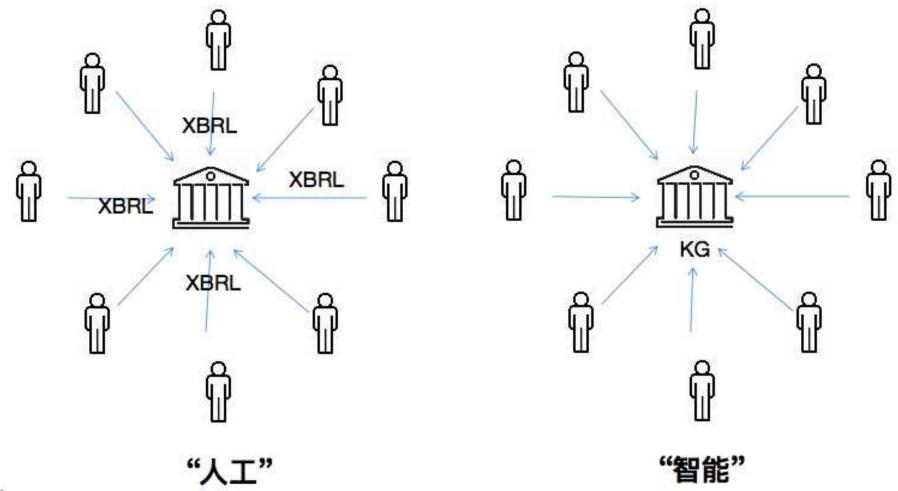


需要改变 成千上万人 的 现有工作流!

从XBRL到KG



00



Al



从数据到智慧 应用场景举例

解决问题:金融流程自动化和业务智能化



利用自然语言处理和知识图谱等人工智能技术,实现金融领域海量文本信息的结构化和图谱化。助力监管、银行、会计师事务所、券商等金融机构进行业务流程的自动化,赋能金融机构提升效率。



公告自动摘要



证券代码: (0247)

证券摘除: 全正大

公告编号: 2017-041

金正大生态工程集团股份有限公司第四届董事会第一次会议决议公告

本公司及董事会全体或其保证信息兼實力容的真实。推确和完整,沒有雇債 证載、提展性防地或重义者需。

一、会议召开情况

2017年6月24日9时,金王大生志工整集制股份可限公司《以下资格》公司") 在由原稿深提供大西街19号公司会议案召开第四届董事第一次会议。会议通知及 会议资料于第17年6月23日以电子邮件或直接透过方式透过全体董事、董事及高 级管理人员。会议应到董事9名,实到董事9名。全体董事一公同承额免本次会议 通知展院。董事万差步,张继义、高义武、陈宏冲、董师既希出居了本次会议。 独立董事王等、王学峰、杨祖集、吕晓峰以逾高的方式参加了本次会议。公司董 事以及高级管理人员到第了会议。经全体与会董事一致推举,会议由董事万进步 失生主持。会议的通知、召集、召开和表达程序符合《中学人民共和国公司法》 (以下资格》《公司法》》) 作《全正大生卷工程集团股份有限公司章程》(以下 简称"《公司章程》") 的规定。

二、会议审议情况

经与会董事认真审信。会议以举于表决和遵循表决相结合的方式形成如下决

ŕ

1、申议经过《关于这举公司董事圣的仪案》。

选择赛事万选步先生为公司赛事长,任期与本程董事任规相民, 在本次赛事 会通过之日忠计。

表央學院: 建成 9 票, 反对0 票, 弃权0 票。

2. 市议建设《关于选举公司原董事长的议案》

选择董事陈宏绰先生为公司副董事长,任期与本届董事任职租司,自本次董 事会通过之目出计。

表决等次。要成 9 票, 反对0 票, 弃权0 票。

3、申议通过《关于坟举公司董事会各专门委员会委员的议案》。

本次董事会选举公司董事会下被各专门委员会组成如下。

(1) 高華王拳鳴《和文童事》、張和舜(和文董事)、万建参为公司提名 委而会委而、王拳雖担任王洛(召集人)。

表决情况、赞成 9 茶。反对0 茶。弃权0 类。

(2) 选择方定步、陈太卓、王睿《和立董事》。王孝總《和立董事》。后 修理《独立董事》为公司战略委员会委员、万姓尹担任王淳(召集人)。

表決情况, 赞成 9 辰, 反对0 亮, 弃权0 系。

(3) 查移模相類(對立重導)、月轉峰(對立重導)、万進以为公司市計 委員会委員,把組織批任主席(召集人)。

表达情况, 赞成 9 素, 反对0 素, 弃权0 素。

(4) 查科王者(程文董事)、共務地(独立董事)、方连方为公司辦理与 老特委员会委员、王莽担任主席(召集人)。

表征情况, 赞成 9 医, 反对0 医, 弃权0 医。

上述各专门委员会委员的任弊与本届董事任弊相同,自本次董事会通过之日

Sec.

中省通过《关于参与认购集合资金相托计划并对控制子公司进行增货管设金》。

阿蒙公司出售?亿元认购北京包托·德昭亚财 024 号集合资金信托计划的贫 后级份额,并同意北京信托·德昭亚财 024 号集合资金信托计划时公司校股子公司合本农业服务有限公司编资。认证公本农业服务有限公司编资。 213,000 万元。

模拟公司法定代表人为连步先生代表公司签署相次的信托文件及与履行上 域义务而置要签署的其他文件,并代表公司处理与上述审查有关的一切事务。陈 时在需要时于有关文件上加盟公司的公章。

表识情况, 赞成 9 赛, 反对 0 厚, 弃权 0 聚。

具体内容详见整数于巨视验机网(www.cninfo.com,cn)的公司《关于参与 认购集全接金值托计划并对构股子公司进行增货的公告》。

三、各省文件

- 1、公司等四周董事会等一次会议决议
- 7、独立宣事关于郑周周宣享会第一次会议相关事项的独立意为

Case1 股东大会决议公告类

案例名称:《金正大生态工程集团股份有限公司 第四届董事会第四次会议 决议公告》

提取结果:

公告提取:人物-亲属关系发现

19,774,442

14,693,820

13,191,869

10.395,060



⊕N-9	24,937,438								
£8			39,446,	196 人民市	并地段		39,446,496		
	18.21 所作的人用的名词形以包裹			100	9件类	- 3	W.R.		
	报告期末将有天制者条件助价数量			股份种类					
	前1	8 老无關售条	件数字持款情况						
	公司未知其他股东之间是否存在关联关系成一致行动关系。								
散行动的说	锋系权性关系。上述因非公开发行认购股份成为公司前 10 名股东之一的股系与公司投 股股东湖上洋丰集组股份有限公司、实际控制人榜才学不存在关联关系。								
乗)(計画)5	日、相关股	价解除报告后	股份報告期为 12 个 己于 2016 年 5 月	19日上市曲	进-	N. C.	608-35-04		
四配告新双	∰ 10 ¢	5股东中。南方	· 债本管理有關公	可到 2015年	认购公司非	5开发行	散份成为公		
境内自然 人	0.68%	8,884,856	-128,000	8,692,856	192,000	MH	8,212,800		
境内自然 人	0,71%	9,323,846	-162,000	9,205,846	118,000				

19,774,442 人民币普递股

14,693,820 人民币普通提

13,191,869 人托币普通股

10.795.060 人民币普通股

杨才斌: 男,1963年5月出生,经济师,现任湖北洋丰 这居直业股份有限公司董事长,湖北洋丰股份有限公 司董事,许昌洋丰逸居置业有限公司董事长,重庆洋 丰逸居房地产有限公司董事长,湖北新洋丰肥业股份 有限公司副总经理。

黄镇先生:汉族,1966年9月出生,硕士研究生。曾担任中国国际钢铁投资公司业务经理,中宏集团综合部部门总助,中信证券投资银行部经理,海通证券北京投行二部部门副总经理,北京恒银投资管理有限公司副总经理,和丰资本(中国)有限公司执行合伙人,和丰财富(北京)投资有限公司总经理,现任湖北新洋丰肥业股份有限公司董事、执行总裁。

杨小红;女,1977年1月出生,硕士学位,经济师,曾 担任湖北洋丰股份有限公司督办审计室副主任、人事 处处长、财务部副部长、财务部部长、副总经理(分 管财务)。现任湖北新洋丰肥业股份有限公司董事、 财务总监。

杨华锋: 汉,1974年10月出生,高级工程师,历任湖 北洋平股份有限公司物资供应部部长、销售公司河南 分公司经理,湖北新洋丰肥业股份有限公司董事、常 务副总经理,现任湖北洋丰集团股份有限公司董事, 湖北新洋丰肥业股份有限公司董事,

杨才学: 男,1964年6月出生,高级经济师,曾担任荆 门市第二磷肥厂厂长、党支部书记,湖北洋丰股份有

会司一传

LCT001 BE

战策定向

会司一長

計划 公司一位





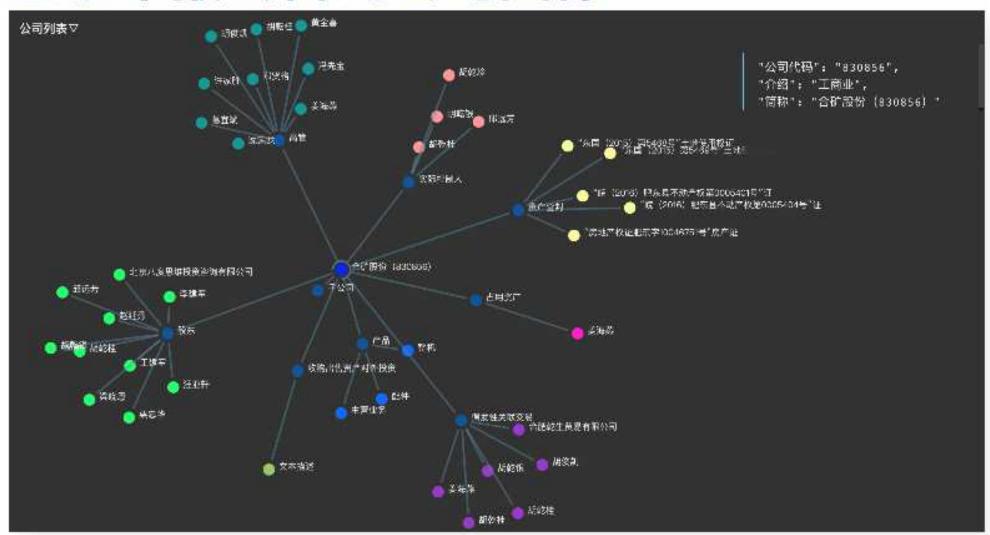
模块分类 公司基本信息	模块详情概览						
	基本信息概范	团队与核心成员	股东持股信息	股权变更历史			
	参控股公司关系	资本市场情况	财务与估值指标	销售生产资质			
	产品结构分析	知识产权一拡	技术与研发能力	销售模式			
	同业企业对比	客户与市场	公司新闻	高管最近讲话			
公司财务分析	关键审计事项	前五欠款人	五大供应商	利润表分析			
	现金流量表分析	资产负债表分析	损益表分析	45 244 400 25 401			
	主蓋构成	不良贷款率					
行业分析	行业竞争	市场规模	政策环境	产业链分析			
	行业增长速度	行业上下游	行业重要会计政策				
7T +C + 0 TE	公司研报掩要	Control of the Contro	龙头企业	投资观点			
研报摘要	目标价格一致性预期	宏观研报摘要					
	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			
公告摘要	提供反担保	仲裁事件	政府补助	股东墙持计划			
ACION	新建项目	重大报失	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			
	股东减护	出售资产	签订日常经营合同	计提大额资产减值准6			
	定向增发	限售股解號	般权激励回购注销	出售资产			
关系数据	产量销量	股份被原押	級份质押解除	股权质押展期			
	特殊交易事項	重大处罚	股东增持计划完成				
	高韶关系数据	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	財报搜索			
圣 斯技术服务	新闻标签服务	自动化报告引擎	问答引擎				

99.8%

财报提取正确率 达到人工质量

基于年报公告的公司画像

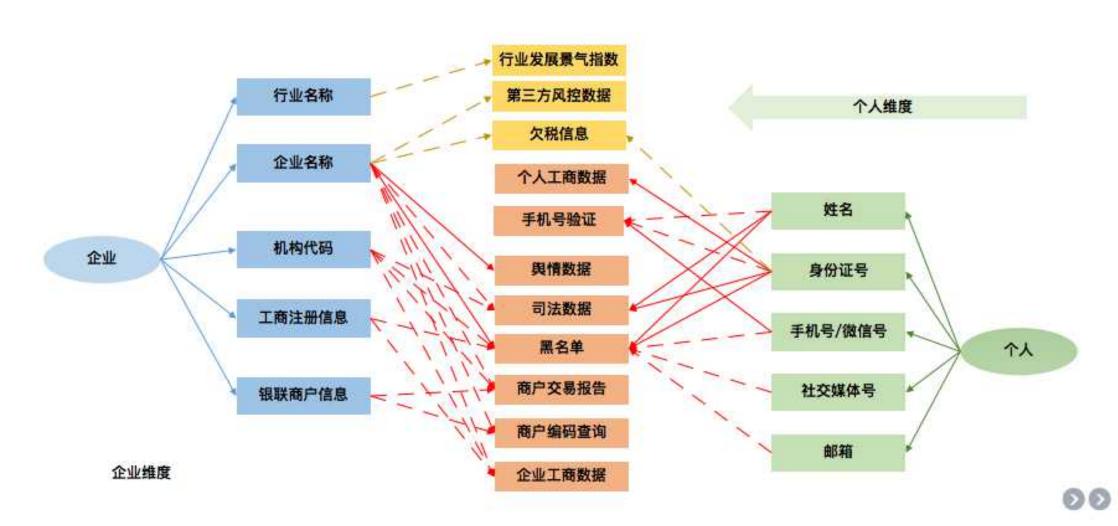




银行跨部门数据治理



帮助银行治理跨部门数据以及外连接部数据以支持银行对个人用户+企业用户画像和风险分析



银行跨部门数据治理



智能营销推荐系统	潜在客户识别	潜在客户识别 沉睡客户的唤醒		流失客户预量	Non-America.		
特征监测 个性化服务	客户风险 识别系统	客户质量 自动评分系统	市场推广辅助 决策系统	标签智能 检索系统	应用层		
		结合	业务逻辑				
AI服务	知识存储	知识表现	知识融合	知识检索			
技术应用层	知识图谱数据库						
文本分词 新词发现 关键词识别 词义消歧 半监督文本学习 人机交互 数据清洗 技术领域 深度学习 自然语言生成 知识表现 知识获取 基于知识的推荐算法 基于内容的 推荐算法					核心技术层		
算法层面 数据挖掘与聚类算法 机器学习算法 有限自动机 关系网络分析 LSTM RDF推理 SPARQL规则 人名识别 语义消歧 推荐算法 词嵌入与词联想							
数据存储 文档-	割关系数据库 图	数据库 RDF数	据库 大规模知识	2缓存和索引	.,,.		
客户基本信息	客户资产数据	客户交易数据	消费行为信	息 其他	基础数据层		
生活关系信息	社交网络关系 信息	营销活动信息	兴趣爱好信	近期無 涼息 求信息			

2016应用列表

- 信用监控 (Garlik)
- 数据集成 (摩根大通)
- 风控(宜信,普慧)
- 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)

2017应用列表

- 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

2018应用列表

- 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;
-



全球软件开发大会

北京·2019



极客邦企业培训与咨询

「帮助企业和技术人成长」

10 余年 经验技术专家

200+ 国内外一线技术 专家团队 800+

(10000+ 学员参与学习 交流

助力企业提升技术竞争壁垒,让技术驱动业务发展



人工智能基础课

"通俗易懂的人工智能入门课,,





TGO鲲鹏会

汇聚全球科技领导者的高端社群

■ 全球9大城市





为社会输送更多 优秀的科技领导者



构建全球顶尖的有技术背景的 优秀人才成长平台



扫码了解更多内容