

# 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](mailto:baojie@memect.co)

微信 [baojie\\_memect](#)



01

# 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



AI身经百战

## 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



架构大咖

## 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

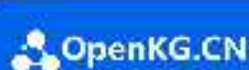
擅长高性能复杂系统架构设计

谷歌Chromecast播放器软件

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



Chromecast



# 文因互联主要荣誉





02

# 人工智能的一些背景



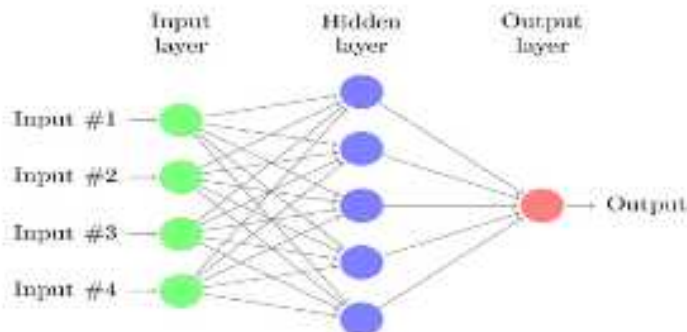
# 人工智能的三大分支

经验主义  
(机器学习)



“学习套路”

连接主义  
(神经网络, 深度学习)



“我认得你”

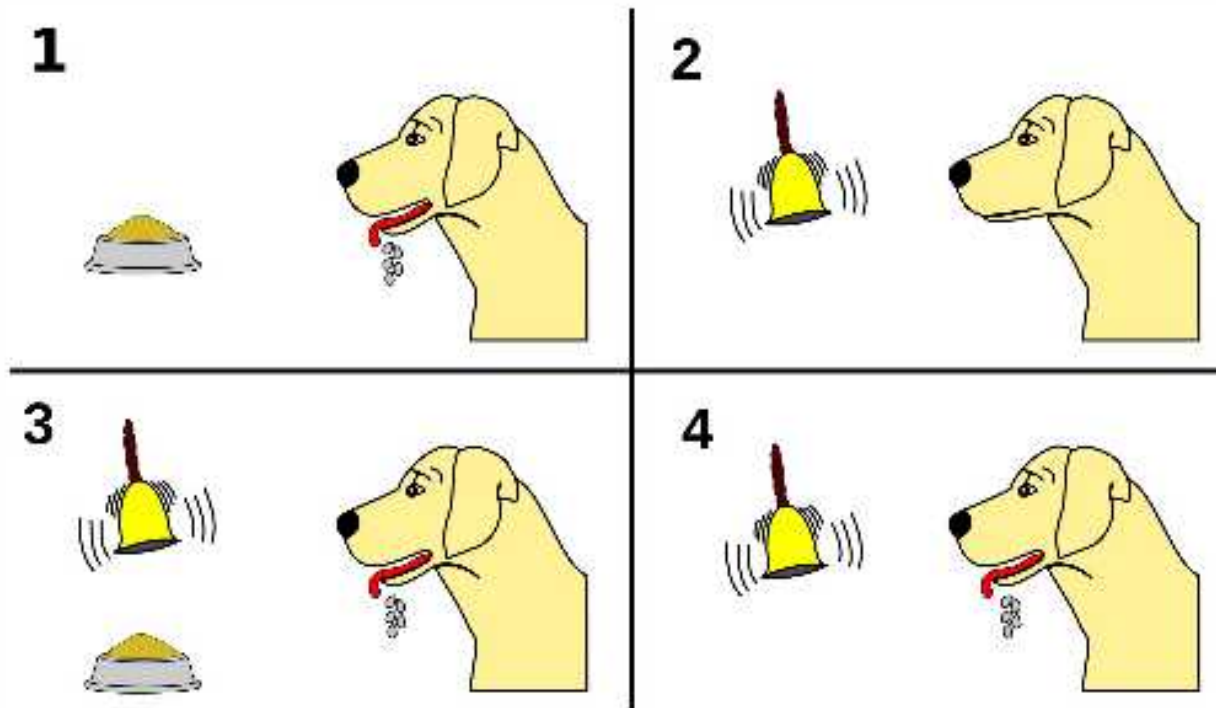
符号主义  
(逻辑, 知识)

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

“明白意思”

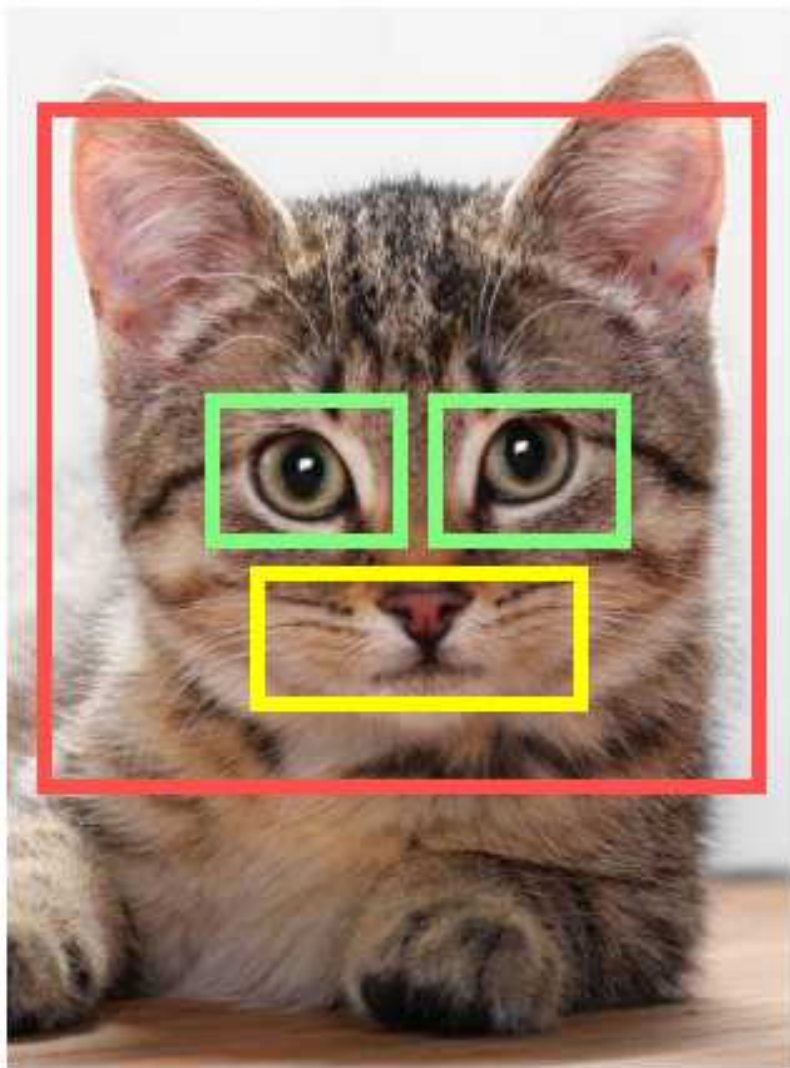
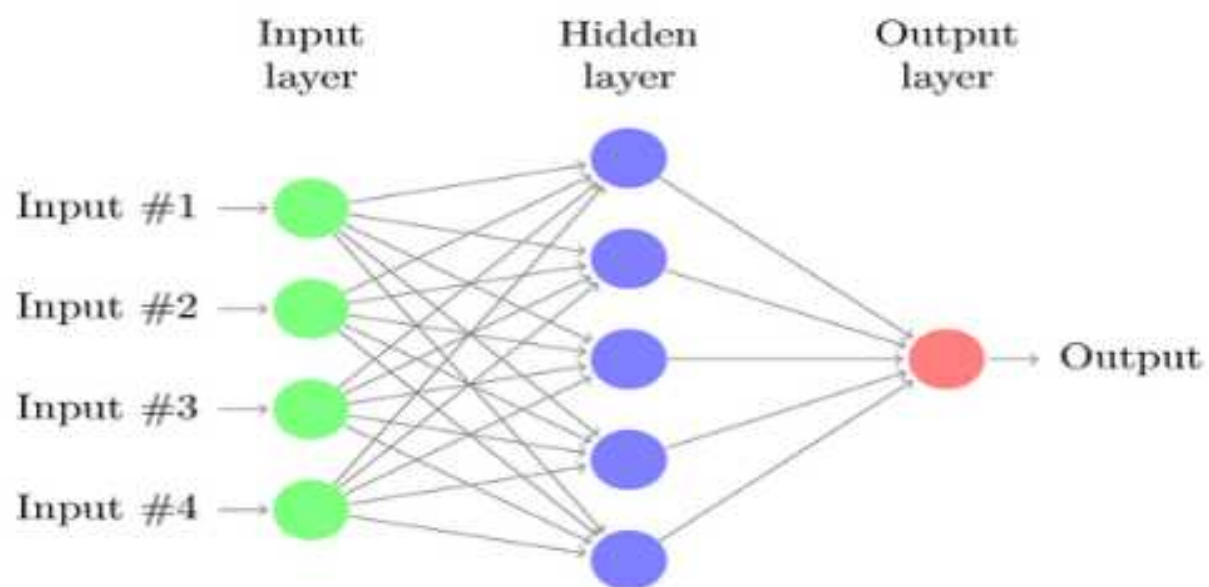


# 经验主义：“学习套路”



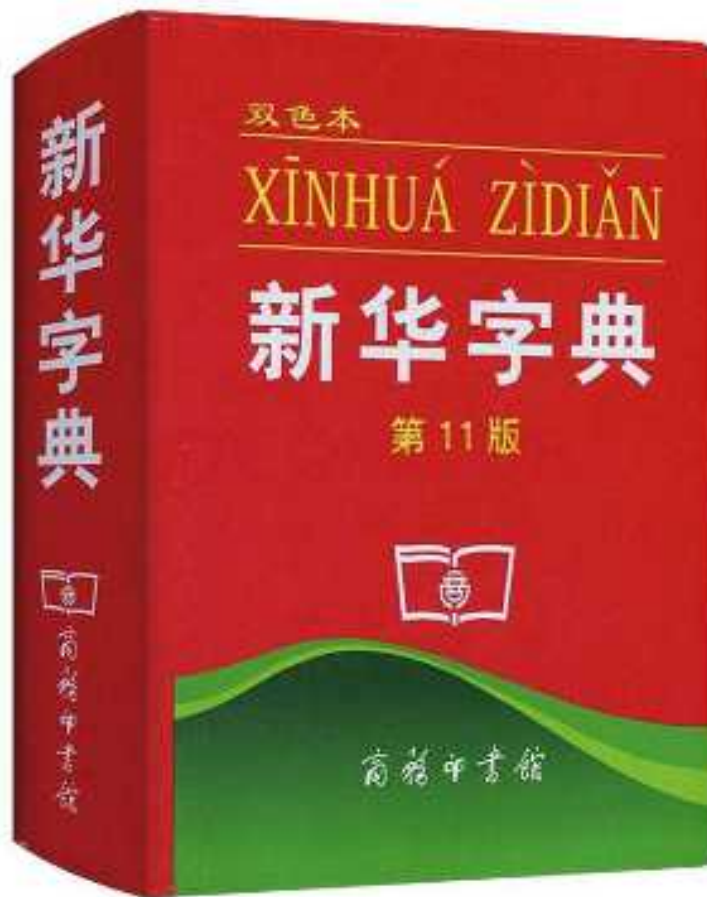
巴甫洛夫的“套路”

# 连接主义：“我认得你”

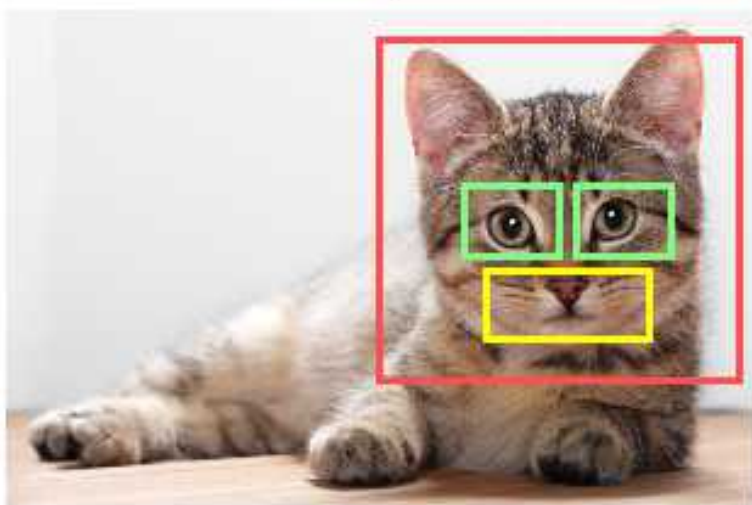
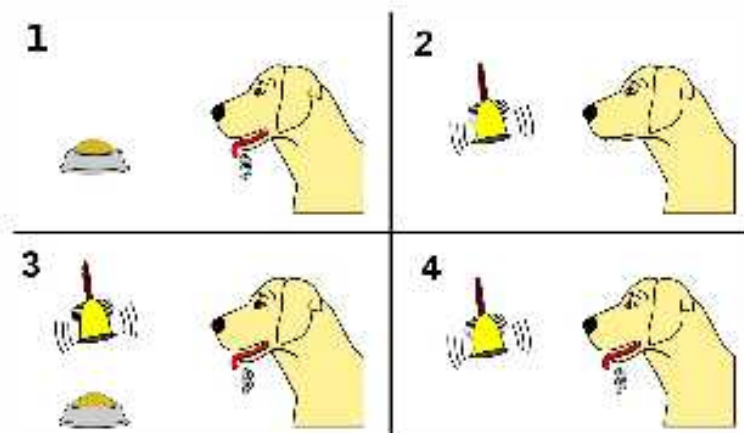


# 符号主义：“明白意思”

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

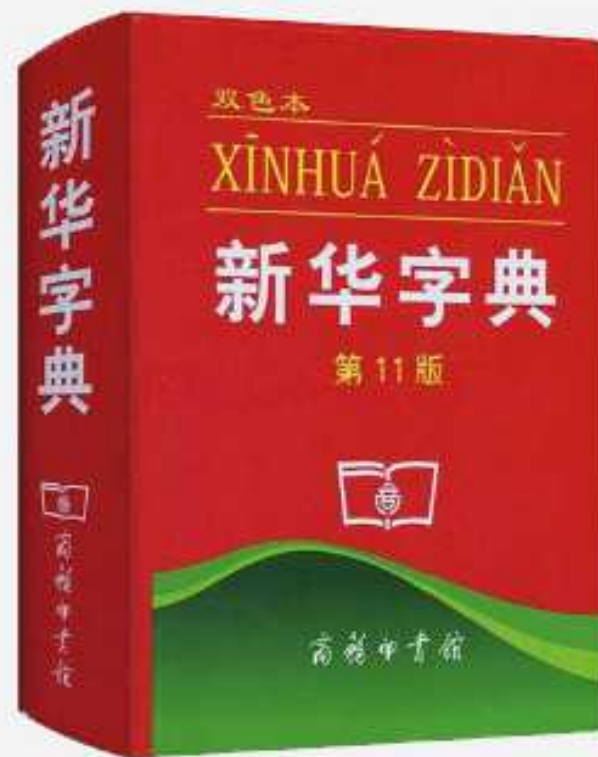


# 感知智能



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**。



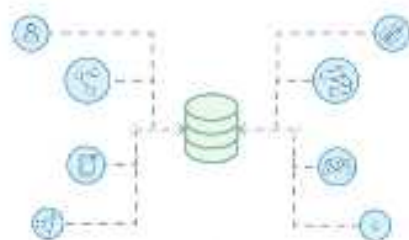
知识提取

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。



知识存储

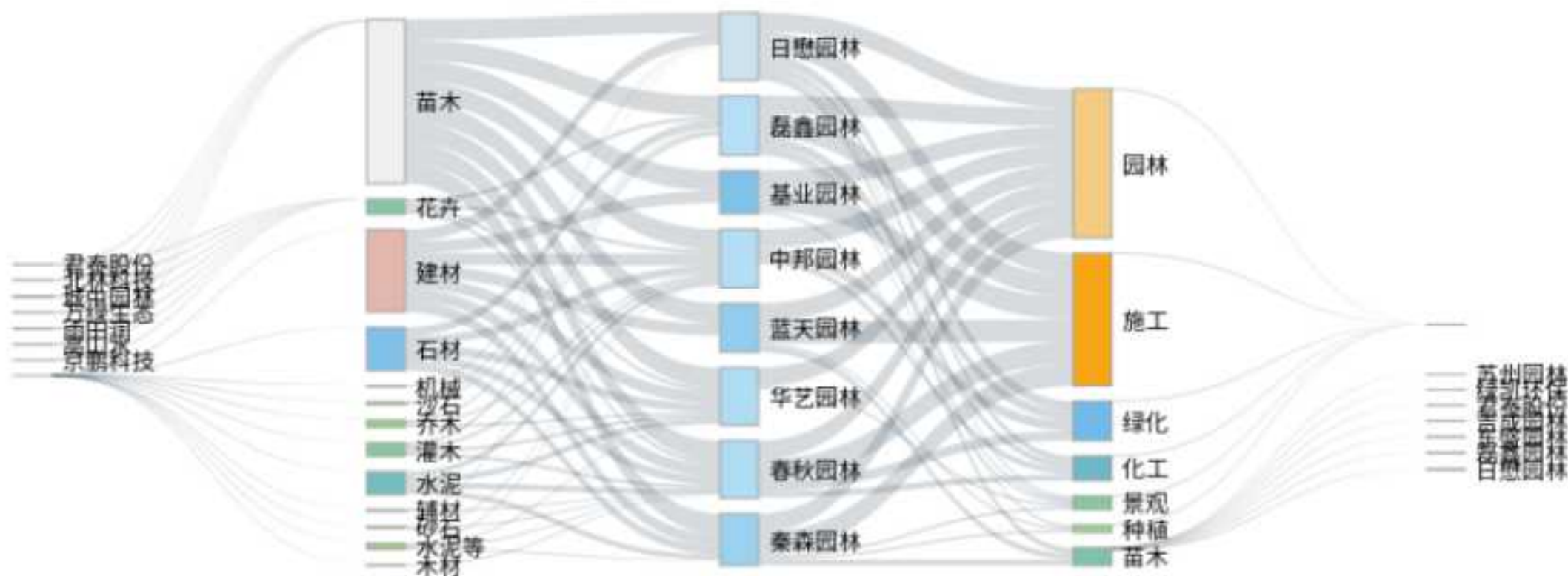
进行大量的结构化数据管理，同时混合管理结构化和非结构化数据。



知识检索

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": [  
    "独立董事"  
  ],  
  "公司": "宝钢包装",  
  "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

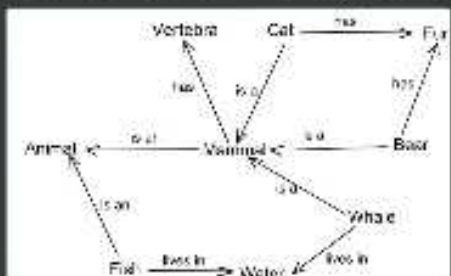
1960s

2001

2006

2012

知识图谱里程碑



语义网络

语义网, OWL

互联数据

知识图谱

Jim Hendler

“语义网之父”

RPI教授

美国白宫Web顾问

Tim Berners-Lee

Web之父

W3C主席

MIT教授

2016图灵奖得主

谷歌





03

# 金融科技发展趋势



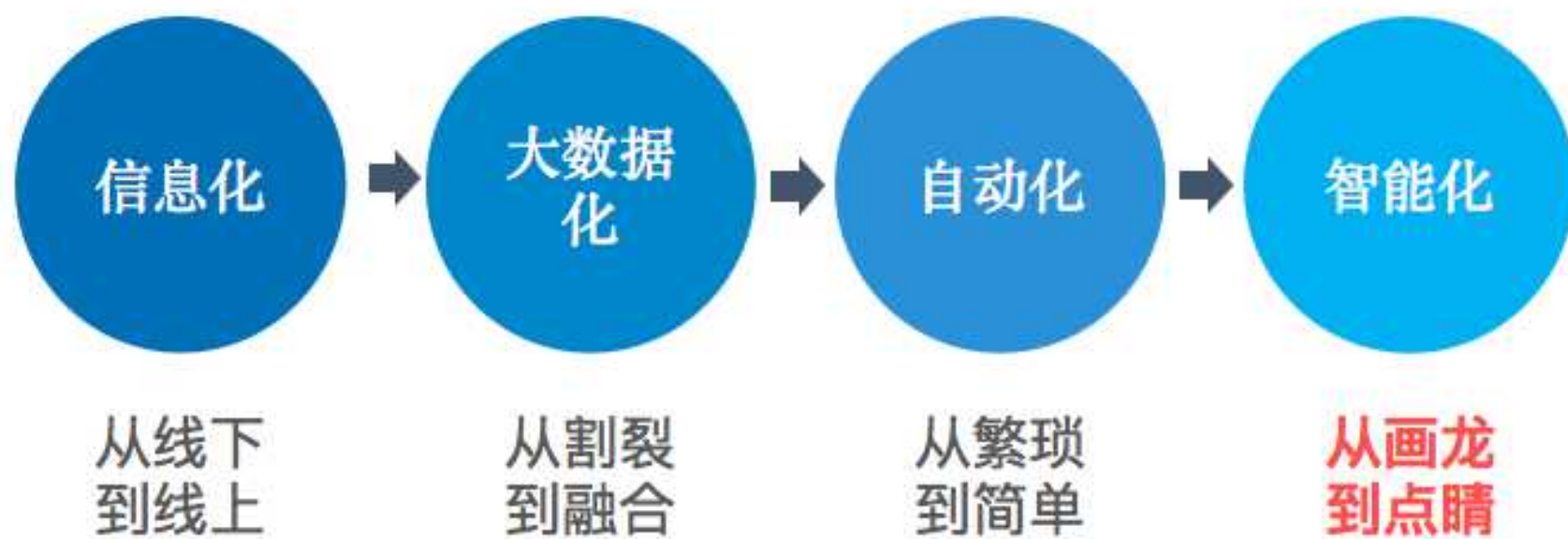
Fintech上下半场

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) 标准化

# 目标



## 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）



04

## 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



# 例：从XBRL到KG



U.S. SECURITIES AND  
EXCHANGE COMMISSION

Search-SEC.gov

EDGAR FILING | MORE SEARCH OPTIONS

ABOUT | DIVISIONS | ENFORCEMENT | REGULATION | EDUCATION | FILINGS | NEWS

Office of Investor  
Education

Contact Us

What is Structured  
Data?

History and Policy Making

XBRL

Data Users

Products

Technical

## Structured Disclosure RSS Feeds



The SEC provides three Really Simple Syndication (RSS) feeds for EDGAR structured disclosure submissions as a courtesy for those interested in viewing, analyzing, and manipulating the submissions. These feeds are updated every two minutes Monday through Friday, 6am – 10pm EST. The content and format of these files may change without prior notice.

These feeds are lists of the 100 most recent filings for:

1. [Corporate event filings](#) containing financial statements tagged using the US GAAP XBRL taxonomy.
2. [Mutual fund filings](#) containing the risk/return summary tagged using the XBRL US Mutual Fund Risk/Return Summary Taxonomy.
3. Filings in both of the above 2 categories.

The above RSS links may be opened in a web browser, but the resulting page will lack functionality. Rather, they are designed to be accessed with specialized RSS software products, which are widely available for free.

Each entry provides metadata (company name, accession number, CIK, filing date, form type, and the number), a link to the filing EDGAR, and a link to the interactive structured data files in a compressed (.zip) format.

For a historical archive of all filings with XBRL exhibits submitted to the SEC beginning with the inception of the voluntary program in 2005, go to <http://www.xbrl.gov/ArchiveWebCommunity/>. Each monthly archive contains a complete set of all structured data filings received during that month across all programs.



Home | Latest Filings | Previous Page

U.S. Securities and Exchange Commission

Search the Next-Generation  
EDGAR System

## Filing Detail

SEC History > Search the Next-Generation EDGAR System > Company Search > Crows Page

Form 10-K Annual Report (Section 13 and 15(d)) SEC 10-K (Form 10-K)

SEC Accession No. 0001103175 15-00008

Filing Date: 2015-07-09  
Accepted: 2015-07-09 17:31:06  
Documents: 99

Period of Report: 2015-12-31  
Filing Date Changed: 2015-07-09

### Document Format File

Seq	Description	Document	Type	Size
1	FORM 10-K	<a href="#">e122835d10k.htm</a>	10-K	872,703
2	EX-10.1	<a href="#">e122835dex101.htm</a>	EX-10.1	92295
3	EX-10.2	<a href="#">e122835dex102.htm</a>	EX-10.2	164263
4	EX-21	<a href="#">e122835dex21.htm</a>	EX-21	1207
5	FX-31.1	<a href="#">e122835dxc311.htm</a>	FX-31.1	9387
6	FX-31.2	<a href="#">e122835dxc312.htm</a>	FX-31.2	9179
7	FX-32.1	<a href="#">e122835dxc321.htm</a>	FX-32.1	3538
8	FX-32.2	<a href="#">e122835dxc322.htm</a>	FX-32.2	3551
9	FX-99.1	<a href="#">e122835dxc991.htm</a>	FX-99.1	47043
16	GRAPHIC	<a href="#">e122835g77n79.jpg</a>	GRAPHIC	7777
17	GRAPHIC	<a href="#">e122835g74n90.jpg</a>	GRAPHIC	4795
18	GRAPHIC	<a href="#">e122835g71n81.jpg</a>	GRAPHIC	8739
Complete Submission Size:				1,001,135,125 (6,544,603 kb)

### Data Files

Seq	Description	Document	Type	Size
10	XBRL INSTANCE DOCUMENT	<a href="#">xmg-20151231.xml</a>	EX-101 IAS	390,107
11	XBRL TAXONOMY EXTENSION SCHEMA	<a href="#">xmg-20151231.xsd</a>	EX-101 SCH	30329
12	XBRL TAXONOMY EXTENSION CALCULATION LINKBASE	<a href="#">xmg-20151231_cal.xml</a>	EX-101 CAL	100,069
13	XBRL TAXONOMY EXTENSION DEFINITION LINKBASE	<a href="#">xmg-20151231_def.xml</a>	EX-101 DEF	275,108
14	XBRL TAXONOMY EXTENSION IASB LINKBASE	<a href="#">xmg-20151231_iasb.xml</a>	EX-101 IAS	549,177
15	XBRL TAXONOMY EXTENSION PRESENTATION LINKBASE	<a href="#">xmg-20151231_pre.xml</a>	EX-101 PRE	490,009

TRANS ENERGY INC (Filer) CIK: 0000918721 (see all company filings)

Business Address:  
210 SECOND ST  
FOURTH FLOOR

Mailing Address:  
210 SECOND ST  
FOURTH FLOOR



# 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

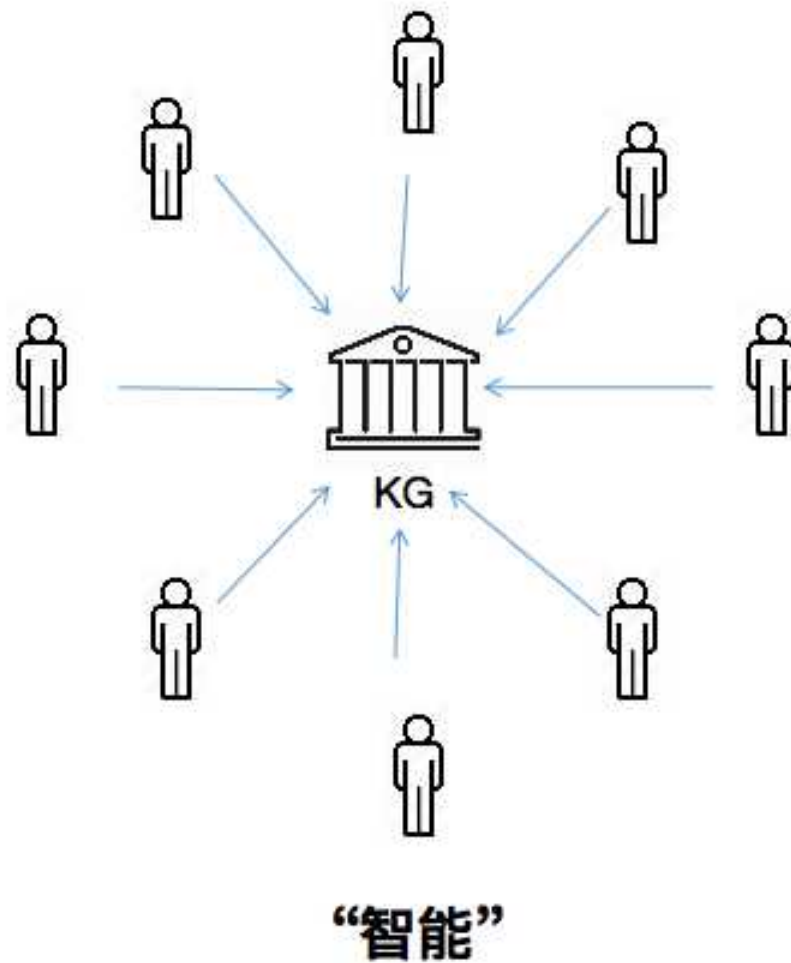
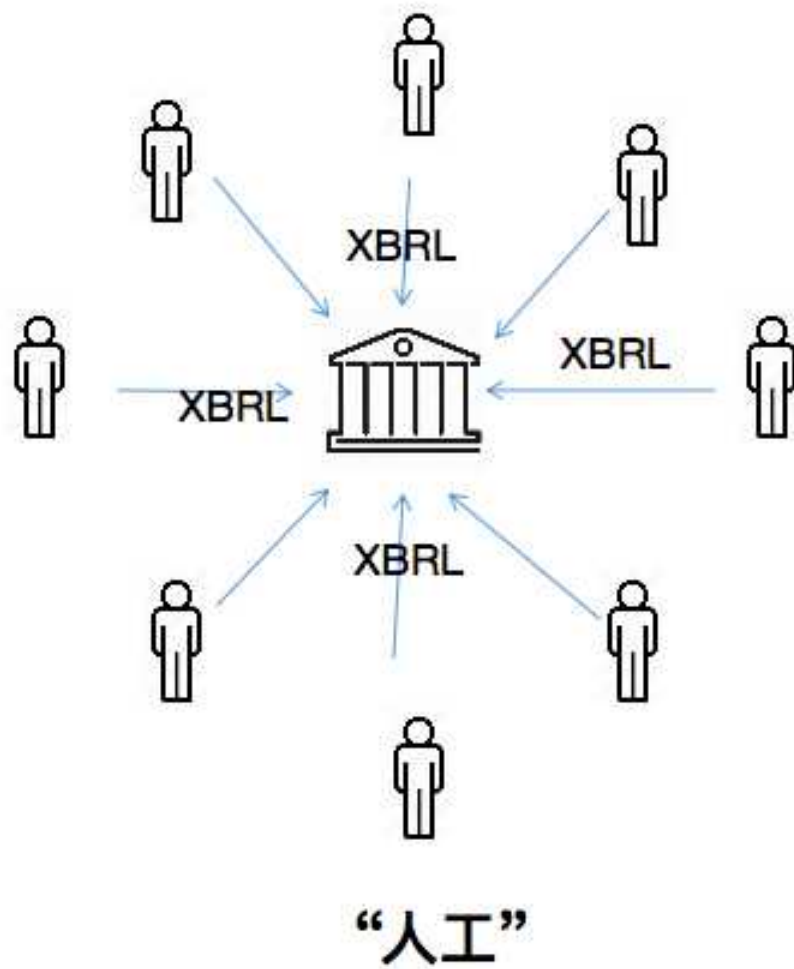




# XBRL的实施局限

需要改变  
成千上万人的  
现有 workflow!

# 从XBRL到KG





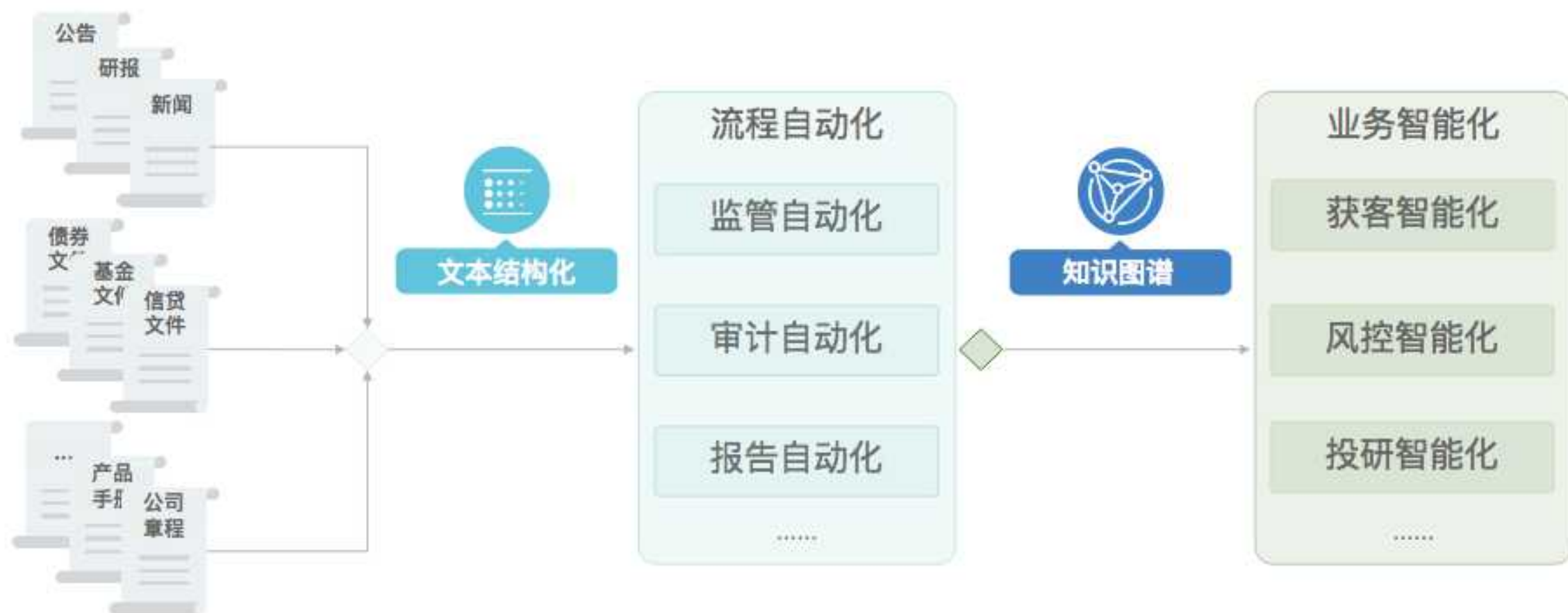
05

# 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: 603470 证券简称: 金正大 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成员保证信息披露内容的真实、准确和完整,没有虚假记载、误导性陈述或重大遗漏。

### 一、会议召开情况

2017年6月24日9时,金正大生态工程集团股份有限公司(以下简称“公司”)在山东临沂经济开发区19号公司会议室召开第四届董事会第一次会议。会议通知及会议资料于2017年6月23日以电子邮件或直接送达方式送达全体董事、监事及高级管理人员。会议应到董事9名,实到董事9名。全体董事一致同意豁免本次会议通知期限。董事万连步、张继义、高义武、陈宏坤、祝祖强出席了本次会议,独立董事王晖、王孝峰、祝祖强、吕晓峰以通讯的方式参加了本次会议,公司监事以及高级管理人员列席了会议。经全体与会董事一致推举,会议由董事万连步先生主持。会议的通知、召集、召开和表决程序符合《中华人民共和国公司法》(以下简称“《公司法》”)和《金正大生态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

经与会董事认真审议,会议以举手表决和投票表决相结合的方式形成如下决议:

#### 1、审议通过《关于选举公司董事长的议案》。

选举董事万连步先生为公司董事长,任期与本屆董事任期相同,自本次董事会通过之日起计。

表决情况:赞成9票,反对0票,弃权0票。

#### 2、审议通过《关于选举公司副董事长的议案》。

选举董事陈宏坤先生为公司副董事长,任期与本屆董事任期相同,自本次董事会通过之日起计。

表决情况:赞成9票,反对0票,弃权0票。

#### 3、审议通过《关于选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设各专门委员会组成如下:

(1) 选举王孝峰(独立董事)、祝祖强(独立董事)、万连步为公司提名委员会委员,王孝峰担任主席(召集人)。

表决情况:赞成9票,反对0票,弃权0票。

(2) 选举万连步、陈宏坤、王晖(独立董事)、王孝峰(独立董事)、吕晓峰(独立董事)为公司战略委员会委员,万连步担任主席(召集人)。

表决情况:赞成9票,反对0票,弃权0票。

(3) 选举祝祖强(独立董事)、吕晓峰(独立董事)、万连步为公司审计委员会委员,祝祖强担任主席(召集人)。

表决情况:赞成9票,反对0票,弃权0票。

(4) 选举王晖(独立董事)、吕晓峰(独立董事)、万连步为公司薪酬与考核委员会委员,王晖担任主席(召集人)。

表决情况:赞成9票,反对0票,弃权0票。

上述各专门委员会委员的任期与本屆董事任期相同,自本次董事会通过之日起计。

#### 4、审议通过《关于参与认购集合资金信托计划并对控股子公司进行增资的议案》。

同意公司出资7亿元认购北京信托·鹏程理财024号集合资金信托计划的优先级份额,并同意北京信托·鹏程理财024号集合资金信托计划对公司控股子公司金丰农业服务有限公司增资。以金丰农业服务有限公司新增注册资本213,000万元。

授权公司法定代表人万连步先生代表公司签署相应的信托文件及与履行上述义务所需签署的其他文件,并代表公司处理与上述事宜有关的一切事务,同时在需要时于有关文件上加盖公司公章。

表决情况:赞成9票,反对0票,弃权0票。

具体内容详见刊登于巨潮资讯网(www.cninfo.com.cn)的公司《关于参与认购集合资金信托计划并对控股子公司进行增资的公告》。

### 三、备查文件

1、公司第四届董事会第一次会议决议

2、独立董事关于第四届董事会第一次会议相关事项的独立意见

## 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:

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于近日召开,会议审议通过《关于选举公司董事长的议案》、《关于选举公司副董事长的议案》、《关于选举公司董事会各专门委员会委员的议案》、《关于参与认购集合资金信托计划并对控股子公司进行增资的议案》等事项。

# 公告提取：人物-亲属关系发现

(限公司)

2016 年年度报告全文

境内自然人	0.71%	9,323,846	-162,000	9,205,846	118,000		
境内自然人	0.68%	8,884,856	-128,000	8,692,856	192,000	质押	8,212,800
因配售新股 元(如有)(参	前 10 名股东中,南方资本管理有限公司因 2015 年认购公司非公开发行股份成为公司前 10 名股东之一,该股份限售期为 12 个月,自 2015 年 5 月 19 日至 2016 年 5 月 19 日,相关股份解除限售后已于 2016 年 5 月 19 日上市流通。						
致行动的说	上述股东中,湖北洋丰集团股份有限公司与杨才斌之存在关联关系,属于《上市公司收购管理办法》中规定的一致行动人。杨才学与杨才斌系兄弟关系;杨才学与杨华锋系叔侄关系,上述因非公开发行认购股份成为公司前 10 名股东之一的股东与公司控股股东湖北洋丰集团股份有限公司,实际控制人杨才学不存在关联关系。 公司未知其他股东之间是否存在关联关系或一致行动关系。						
前 10 名无限售条件股东持股情况							
	报告期末持有无限售条件股份数量	股份种类					
		股份种类		数量			
公司	39,446,496	人民币普通股		39,446,496			
公司一分 302 深	24,937,438	人民币普通股		24,937,438			
公司一传 LCT001 深	19,774,442	人民币普通股		19,774,442			
路策定向 计划	14,693,820	人民币普通股		14,693,820			
公司一保	13,191,869	人民币普通股		13,191,869			
公司一长	10,395,060	人民币普通股		10,395,060			

**杨才斌**：男，1963年5月出生，经济师，现任湖北洋丰逸居置业有限公司董事长，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，历任湖北洋丰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、总裁。

**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



# 财报信息披露自动提取



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	业绩预告
	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	

99.8%

财报提取正确率  
达到人工质量

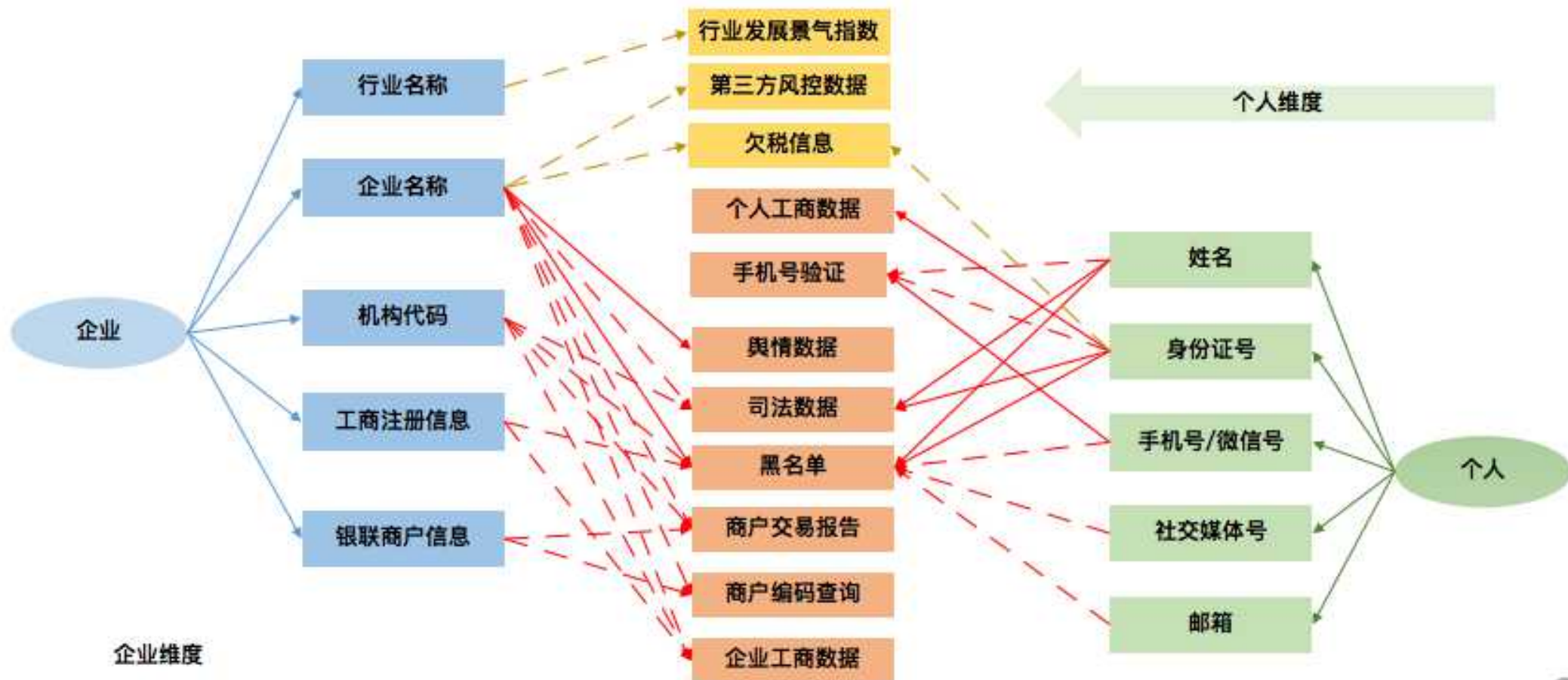






# 银行跨部门数据治理

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



# 2016应用列表

---

- 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)



# 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;
- .....



# QCon

## 全球软件开发大会

### 北京·2019

更多技术干货分享，北京站精彩继续  
提前参与，还能享受更多优惠

识别二维码  
查看了解更多

[2019.qconbeijing.com](http://2019.qconbeijing.com)



# 人工智能基础课

“通俗易懂的人工智能入门课”



扫一扫, 回复「AI」  
领取 AI 学习图谱



王天一  
工学博士，  
副教授

# 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+**  
企业研发团队  
的选择

**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



扫码了解更多官方咨询



# TGO 鲲鹏会

## 汇聚全球科技领导者的高端社群

 全球9大城市

 700+ 高端科技领导者

使命  
Mission

为社会输送更多  
优秀的科技领导者

愿景  
Vision

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