

知乎

首发于  
文因互联

# 金融知识图谱论坛

## Financial Knowledge Graph Workshop (FinKG)

2017.03.29 杭州



组织方：中文信息学会



承办方：浙江大学



承办方：文因互联

## 金融知识图谱的现状和展望



鲍捷

文因互联 CEO

关注他

82 人赞了该文章

本文整理自文因互联CEO鲍捷在3月29日举办的金融知识图谱论坛上所做的开场报告。



▲ 赞同 82 ▼

● 4 条评论

➤ 分享

★ 收藏

...

知乎

首发于  
文因互联

融从业公司。除了今天的主讲嘉宾之外，还有其他的好多家也过来了。BAT 也悉数到场，其他的加在一起有超过 10 家大型的互联网公司，投资机构来了至少有 30 家。可以看到整个智能金融知识图谱领域受到了极大关注，跟去年对比非常明显。我们去年也是在 3 月份的时候，在北京召开了第一次语义对话金融沙龙。那次我们请了一部分嘉宾，大概是来了 30 多个人，那次会场可以看比今天这个会场要小得多。

## 回顾2016 “语义对话金融沙龙”



这一年的时间，发生了很大的变化。我们今天讨论的很多内容，都是在这过去一年里面新增出来的。一年之前大家还把这个东西叫做语义，现在基本上都已经转叫知识图谱了。虽然名字发生了变化，但是这个技术的本质并没有变化，其应用却发生了巨大的变化。像自动化报告，很多跟智能投研相关的技术，去年都还只处于雏形构思阶段，到今天已经得到了落实，在市场上已经可以看到这样的产品出现。

如果我们把金融知识图谱放在一个更大的图景里来看的话，可以看到整个 FinTech 行业的发展。过去三四年大家听到最多的一个词就是互联网金融，最近一两年有智能金融这个词冒出来。在我的理解里面，它代表了 FinTech 的上半场和下半场。任何技术在去改变，或者革新一个已有的产业，通常就会有这样一个上半场和下半场。

▲ 赞同 82 ▼

● 4 条评论

➤ 分享

★ 收藏

...

知乎

首发于  
文因互联

# 触达 → 重构

(互联网金融) (智能金融)

上半场重点在提高效率，在互联网金融这个领域，主要是解决**触达**的问题。不管是交易也好还是平台也好，P2P 也好，其实它是在解决如何把原有的资源更好地组织在一起的问题。它把线下的搬到线上了，把原来效率低的提高效率，主要解决了一个触达的问题。但是当技术真正深入到一个领域，它将行业效率提升到了一个很高的程度，之后会发生怎样的情况？通常会有新的商业模式应运而生，产生一个**重构**。以前它不能够形成的商业模式，不能够做的很多应用，因为新技术的支持可以做了。我觉得这块恰恰就是我们今后两三年，甚至三四年的时间内的一个重心。这一块的重心就不仅仅是由于互联网而更多的是由人工智能来支持的。

在这个从**触达**到**重构**的变迁里，其实很多经验不完全适用。在更早的互联网应用里面，无非就是加钱、加人、加机器解决问题，因为互联网的应用，它是一种复杂的系统。但是对于智能系统，它也是一种复杂系统，但它不是 complex system，它是一个 complicated system，它里面有非常精细的小结构在里头，不仅是通过加钱加人就能够解决这个问题，所以两者的发展路径上会有点不一样的地方。

▲ 赞同 82 ▼

● 4 条评论

➤ 分享

★ 收藏

...





今天我们的主题是知识图谱，但是知识图谱其实是整个智能化的一部分。我们金融的智能化也不能脱离过去20年的整个工作。按照我自己的理解把它分为四个阶段：

第一个阶段就是**信息化**，包括银行之前做的就是这件事情，比如我们用纸质的文档，现在变成一种机器的文档，从线下到线上解决这个问题。

第二个阶段是**大数据化**。过去其实十年的时间，一开始不叫大数据，但其实做的东西就是大数据化，就是把割裂的分散的数据把它集中在一起。一开始它叫数据仓库，后来叫大数据，最近又叫云，每家银行都在做，券商也在做，交易所也在做。实际上是要把这个数据从割裂状态，变成一个融合的状态，变成一个大数据，解决这个问题。

这两年应该说从 2015 年开始，我们又进入了下面一个阶段，就是一个**自动化**的阶段。在我们有了这么多数据之后，我们发现很多东西是非常繁琐的，用人工来转移这些数据，不如用机器来做这种重复性的工作。我们用更智能的程序，来帮我们自动化这个过程，把繁琐变为简单。这也是最近两三年的时间，我们在很多分支里都看到了的尝试。

**智能化**可以说是紧跟着自动化开始，我把它称为从画龙到点睛，也就是整个系统里面，我们以前是人做所有的工作。但是人应该和机器在一起形成一个协作系统，机器做复杂的事情，而人是真正的价值创造者。只有人能够点睛，智能化就是要把这个过程变成让机器来画龙——人来点睛的过程。

▲ 赞同 82 ▼

● 4 条评论

➤ 分享

★ 收藏

...

知乎



首发于  
文因互联



人工智能现在有很多不同的分支，我这里列了三个主要的分支，就是**经验主义**，或者称为机器学习方法，还有**联结主义**，最近最火就是深度学习，在这之前是神经网络方法，还有今天我们要谈到的知识的方法，就是**符号主义**，更早的时候叫逻辑，后来在90年代末的时候，开始有一个领域叫语义网，后来演化成了知识图谱。在金融领域里面，早期大家对于机器学习比较熟悉，包括像信用卡导购，营销，用户画像技术，我们都会用到很多机器学习的方法；最近一两年时间，深度学习开始广泛的应用，知识图谱进入还是算比较晚的吧。

大家对知识图谱可能不是很熟悉，我这里就来简单讲一下我个人理解的知识图谱。在我看来知识其实就是结构，我们最简单的知识就是字典，我们用一些词去定义另外一些词，所以这就是形成了词和词之间的结构。

知乎



首发于  
文因互联



谷歌知识图谱

这副图是谷歌的知识图谱。它放出来的一个广告图，代表了每一个节点，就是一个实体，这里有一个蒙娜丽莎是一幅画，然后达芬奇是一个人，这代表了一个关系，就是蒙娜丽莎是达芬奇画的这样一个关系。回头陈华钧老师会更详细地介绍。

知乎

首发于  
文因互联

另外一个结构，就是行业上下游。每一种行业，我们都可以画出这个行业的上游产品、下游产品是什么；在每一种产品上有什么样的公司在提供这样一种服务。以前我们都是用人工来做这样的事情，它可不可以自动化呢？以前我们对于比如A股上市公司，可能只有那么几百个行业，我们可以人工来做。但现在我们面临的是上万家上市和挂牌公司，上千个细分行业，我们能不能用机器来提高效率，来做出这样的知识图谱呢？应该是可以的。

这是用了 JSON 的格式从公告里面提取出来的一个高管变动，里面提到了其中某一个人，他因为什么原因，他在什么时间辞去了什么职位，这样一个知识提取，也是机器可以帮我们做出来的。

所以整个知识图谱技术的核心，就是如何把数据**结构化**。在传统的数据库研究里面，我们已经看到了数据是结构化的。知识图谱技术，从某种程度上来说，是数据库技术往前的一步，把传统的表格结构所不能处理的复杂的关系，用更新型的数据库来进行存储、表达。这个技术虽然是从2012年开始才为大家所知，但从渊源上讲，已经有差不多15年到20年的发展时间了。

▲ 赞同 82 ▼

● 4 条评论

➦ 分享

★ 收藏

...

今天的论坛是由中文信息协会来主持的，中文信息协会以前是侧重在自然语言处理这个技术上面的。自然语言处理里有一个分支叫**知识提取**，从非结构化的数据里面，把结构化数据提取出来。后来到了2012年左右的时候，这一块就融入了知识图谱技术，白硕老师待会给大家详细地讲这一块技术。**知识表现**是另外一块，更古老的时候叫做逻辑，或者再早的时候叫专家系统。漆桂林老师是这块的大拿。在语义网研究的十几年的时间中起起伏伏，当然也有一些非常辉煌的成功，包括像 Siri、IBM Watson 这样的项目，在2010年、2012年的时候，证明了知识表现的价值。到2012年的时候，也被并入了知识图谱这个领域。**知识存储**的渊源，就是刚才提到的数据库技术。现在大家听到得很多，像 Neo4j 这种图数据库，RDF 数据库，都是代表了新的知识存储引擎。最后一块就是**知识检索**，今天在座的也有很多从搜索引擎公司出来的朋友。谷歌有一句话是说我们现在搜索的东西，不再是字符串-string，而是实体-entity。

这四个不同的领域，发展到了某一个关键节点，都发现要想解决各自领域里的问题，必须用结构化数据，进而就形成了知识图谱这个技术。



知乎

首发于  
文因互联

我们具体再来看金融知识图谱，这是我们今天在中国市场上看到主要的各种类型的金融知识图谱。这是一个很粗的分类，每一个分类下面，都会有很多更细分的分类。比如像前面的创投类数据库，滕放腾总，文飞翔文总，还有郭颖哲郭总，他们接下来会分别向大家介绍自己的经验。另外在很多其他类型的市场上，包括一级市场、二级市场，我们都看到各种不同的知识图谱、数据库陆陆续续出现。比如做A股、新三板的公司，其实还有做港股、美股的公司，各种基本面的数据，行情的数据，现在都在陆陆续续地知识图谱化。以前大家只看F10，现在一种智能的“F10”出现了，还有公告数据、研报数据的提取、公告研报的检索等等。泛舆情数据、泛企业数据和各种工商数据，都是我们今天看到的金融知识图谱的不同分支。

从应用上来讲，我能想到的有十几种。但实际上会远远不止这十几种。让人兴奋的是，这里面的大多数应用都是在过去一年里头出现的。一年之前我列这个列表可能只有四五种，而现在我们可以列出十几种。所以我们很难想象，明年我们会列出多少种。在各种投资的分支上面，我们可以看到知识图谱有非常广阔的应用。今天在座的也有很多嘉宾，他们公别都在做各自的工作，这个论坛的目的就是把大家凑在一起，互通有无，

▲ 赞同 82 ▼ ● 4 条评论 ➤ 分享 ★ 收藏 ...

知乎



首发于  
文因互联

经验，最后是陈华钧教授和丁力老师，来跟大家讲一下开放知识图谱的愿景。

(更加精彩的论坛主题报告将于下周陆续发布，敬请关注)

编辑于 2017-04-06

知识图谱 金融 金融科技

文章被以下专栏收录



文因互联

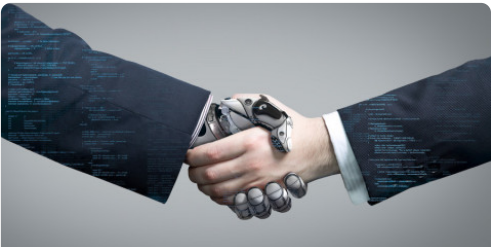
关注专栏

推荐阅读



中科院自动化所陈玉博：事件抽取与金融事件图谱构建

鲍捷



文因互联入驻京东AI加速器：以知识图谱赋能金融机构

鲍捷



行业研究  
FinTech

星起

4 条评论

切换为时间排序

写下你的评论...



技无止境

写的很好啊

赞同 82



4 条评论

分享

收藏





herbtert拒绝吹b

1 年前

期待后文

👍 赞



见舟

1 年前

坐等后文

👍 赞



gudx

1 年前

说好的后文呢

👍 赞