

AISWare Onta KG

亚信科技知识图谱产品白皮书

AISWare Onta KG是亚信科技全力打造的知识图谱一站式知识计算平台，提供源数据管理、本体管理、图谱管理等功能，帮助用户从0到1快速构建知识图谱，管理知识图谱流水线任务。

声明

任何情况下，与本软件产品及其衍生产品、以及与之相关的全部文件（包括本文件及其任何附件中的全部信息）相关的全部知识产权（包括但不限于著作权、商标和专利）以及技术秘密皆属于亚信科技（中国）有限公司（“亚信”）。

本文件中的信息是保密的，且仅供用户指定的接收人内部使用。未经亚信事先书面同意本文件的任何用户不得对本软件产品和本文件中的信息向任何第三方（包括但不限于用户指定接收人以外的管理人员、员工和关联公司）进行开发、升级、编译、反向编译、集成、销售、披露、出借、许可、转让、出售分发、传播或进行与本软件产品和本文件相关的任何其他处置，也不得使该等第三方以任何形式使用本软件产品和本文件中的信息。

未经亚信事先书面允许，不得为任何目的、以任何形式或任何方式对本文件进行复制、修改或分发。本文件的任何用户不得更改、移除或损害本文件所使用的任何商标。

本文件按“原样”提供，就本文件的正确性、准确性、可靠性或其他方面，亚信并不保证本文件的使用或使用后果。本文件中的全部信息皆可能在没有任何通知的情形下被进一步修改，亚信对本文件中可能出现的任何错误或不准确之处不承担任何责任。

在任何情况下，亚信均不对任何因使用本软件产品和本文件中的信息而引起的任何直接损失、间接损失、附带损失、特别损失或惩罚性损害赔偿（包括但不限于获得替代商品或服务、丧失使用权、数据或利润、业务中断），责任或侵权（包括过失或其他侵权）承担任何责任，即使亚信事先获知上述损失可能发生。

亚信产品可能加载第三方软件。详情请见第三方软件文件中的版权声明。

亚信科技控股有限公司（股票代码：01675.HK）

亚信科技创立于1993年，依托产品、服务、运营和集成能力，为电信运营商及其它大型企业客户提供业务转型及数字化的软件产品及相关服务，致力于成为大型企业数字化转型的使能者。

根据弗若斯特沙利文的资料，我们是中国电信行业最大的电信软件产品及相关服务供应商，按2017年收益计，我们的市场份额为25.3%。根据同一资料来源，我们也是中国电信行业最大的BSS软件产品及相关服务供应商，按2017年收益计，我们的市场份额为50.0%。我们是中国第一代电信软件的供应商，从20世纪90年代开始与中国移动、中国联通和中国电信长期合作，支撑全国超过十亿用户。与电信运营商的长期合作关系让我们对电信运营商的IT及网络环境以及业务运营需求有了深度理解，使我们能够开发出拥有500多种任务关键型电信级软件的丰富的产品组合（软件产品主要面向电信运营商，对其业务运营至关重要），包括客户关系管理、计费账务、大数据、物联网及网络智能化产品。我们有214家电信运营商客户，包括中国移动、中国联通和中国电信的总部、省级公司、地市级公司、专业化公司和合营企业。

我们也正在积极拓展在中国非电信企业软件产品及相关服务市场的市场份额。凭借我们在电信软件产品及相关服务市场丰富的行业知识及专长及稳固的领导地位以及全方位、高度专业化的电信级产品图谱，我们相信我们也已经就解决各类企业，尤其是大型企业在业务转型与数字化方面与电信运营商相类似的、最为根本的需求占据了有利地位。我们有38家广电、邮政及金融、电网、汽车等行业的大型企业客户。通过资源、管理、专业知识及技术专长的共享，我们能够同时服务电信和非电信企业市场，凭借协同效应赢取新业务并保持竞争优势。

部分企业荣誉资质

ISO 9001质量管理体系认证	国家规划布局内重点软件企业
ISO 20000IT服务管理体系认证	2018年中国软件业务收入前百家企业前20强
信息系统集成及服务资质（一级）	2018年中国电子信息行业社会贡献500强
CMMI 5级（能力成熟度模型集成5级）认证	2018年中国电子信息研发创新能力50强企业

目录

一. 摘要	7
二. 缩略语与术语解释	8
三. 产品概述	9
3.1 趋势与挑战	9
3.2 产品定位	10
四. 技术介绍	11
4.1 亚信科技知识图谱产品整体架构	11
4.2 亚信科技知识图谱产品功能架构	11
4.3 关键技术能力	12
五. 功能介绍	17
5.1 基础功能	17
5.2 特色功能	20
5.2.1 图谱知识推理	20
5.2.2 图谱浏览	20
六. 场景应用方案	21
6.1 疫情防控	21
6.2 AIOps智能知识问答	22
6.3 识别潜在社群支撑魔百盒点播推荐	23
6.4 数据资产快速溯源，发现资产异常变化	24

七. 带给客户的价值	25
八. 产品优势	26
九. 联系我们	27

AsialInfo PRD Confidential

一. 摘要

知识图谱 (Knowledge Graph) 的概念是由Google公司在2012年提出的[1], 指代其用于提升搜索引擎性能的知识库。同近年来学者的习惯, 本文借用此概念泛指早期至于如今的知识库项目, 而非特指Google的知识图谱项目。知识图谱的出现是人工智能对知识需求所导致的必然结果, 但其发展又得益于很多其他的研究领域, 涉及专家系统、语言学、语义网、数据库, 以及信息抽取等众多领域, 是交叉融合的产物而非一脉相承。

AsialInfo PRD Confidential



二. 缩略语与术语解释

缩略语或术语	英文全称	解释
CSV	Comma-Separated Values)	CSV(Comma-Separated Values), 逗号分隔值, 有时也称为字符分隔值, 因为分隔字符也可以不是逗号, 其文件以纯文本形式存储表格数据 (数字和文本)。
JSON	<u>JavaScript</u> Object Notation	JSON(JavaScript Object Notation) 是一种轻量级的数据交换格式。易于人阅读和编写。同时也易于机器解析和生成。它基于JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999的一个子集。
JDBC	Java DataBase Connectivity standard	Java语言中用来访问数据库应用程序接口规范, 提供了诸如查询和更新数据库中数据的方法。通过标准化的访问接口, 可以访问任何符合JDBC规范的数据库。
RDF	Resource Description Framework	RDF是一种资源描述语言, 它受到元数据标准、框架系统、面向对象语言等多方面的影响, 被用来描述各种网络资源, 其出现为人们在Web上发布结构化数据提供一个标准的数据描述框架

三. 产品概述

3.1. 趋势与挑战

在人工智能领域，知识建模必须在知识库和两个子系统之间建立联系：agent行为（问题求解技能）和环境（问题存在的领域）。而长期以来，AI的研究者较为注重前一个子系统，而领域知识的表达依赖于特定的任务。这样做的好处是只需要考虑相关的领域知识。但是，大规模的模型共享、系统集成、知识获取和重用依赖于领域的知识结构分析。

因此，进入九十年代以来，与任务独立（task-independent）的知识库（本体，英文Ontology）的价值被发现，并受到广泛关注。

Google于2012年5月16日发布了Knowledge Graph，特指其使用的极大增强搜索结果的知识库。随后Knowledge Graph广泛被互联网和IT行业认知。Ontology和相关的技术工程实践被统一称作知识图谱。

AsialInfo PRD Confidential

3.2. 产品定位

这个时代，信息极大丰富，人每时每刻都被各种各样的知识、信息轰炸着。如何有效的选择对自己有价值的知识，如何构建一个独属于自己的知识体系并让它为自己创造价值，变得越来越重要。只有知识经过了你的选择和应用，内化为自己的隐性经验，纳入到你的知识体系中，才能真正地为你创造财富。

AsialInfo PRD Confidential

[返回目录](#)

四. 技术介绍

4.1. 亚信科技知识图谱产品整体架构



4.2. 亚信科技知识图谱产品功能架构

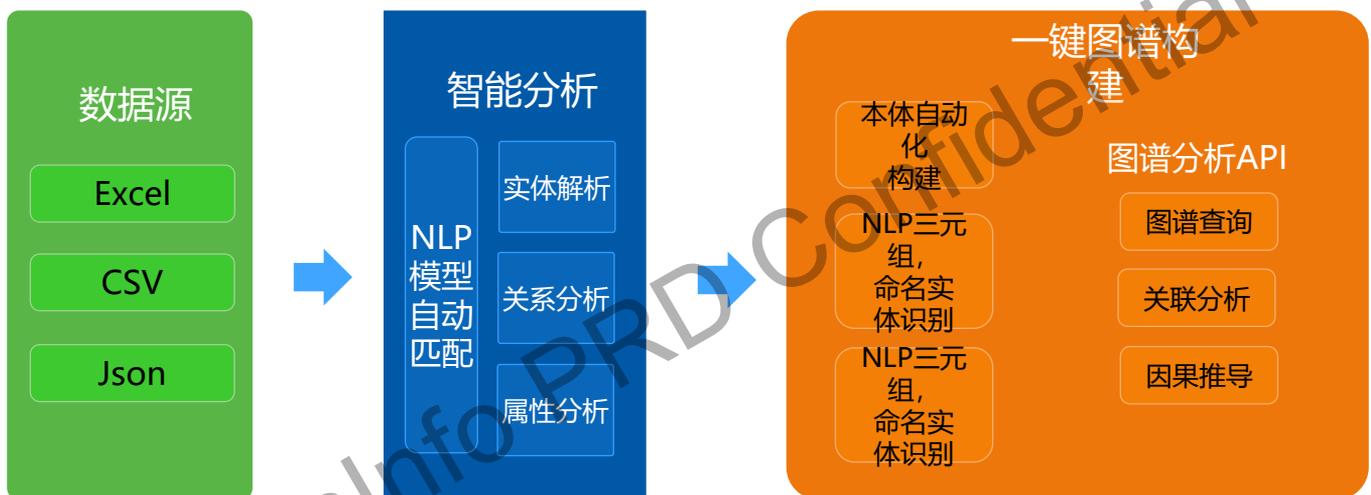


4.3. 关键技术能力

4.3.1 一键智能构建图谱

智能构建图谱功能只需提供源数据，无需其他配置，即可快速构建图谱。具体任务包括：

- 定义数据源及数据格式。
- 自动根据数据定义及格式生成本体定义及图谱数据导入。



本体自动构建，图谱数据自动导入，
图谱数据稽核及人工校对提醒，图谱
分析API自动化生成

4.3.2 基于图谱的数据治理

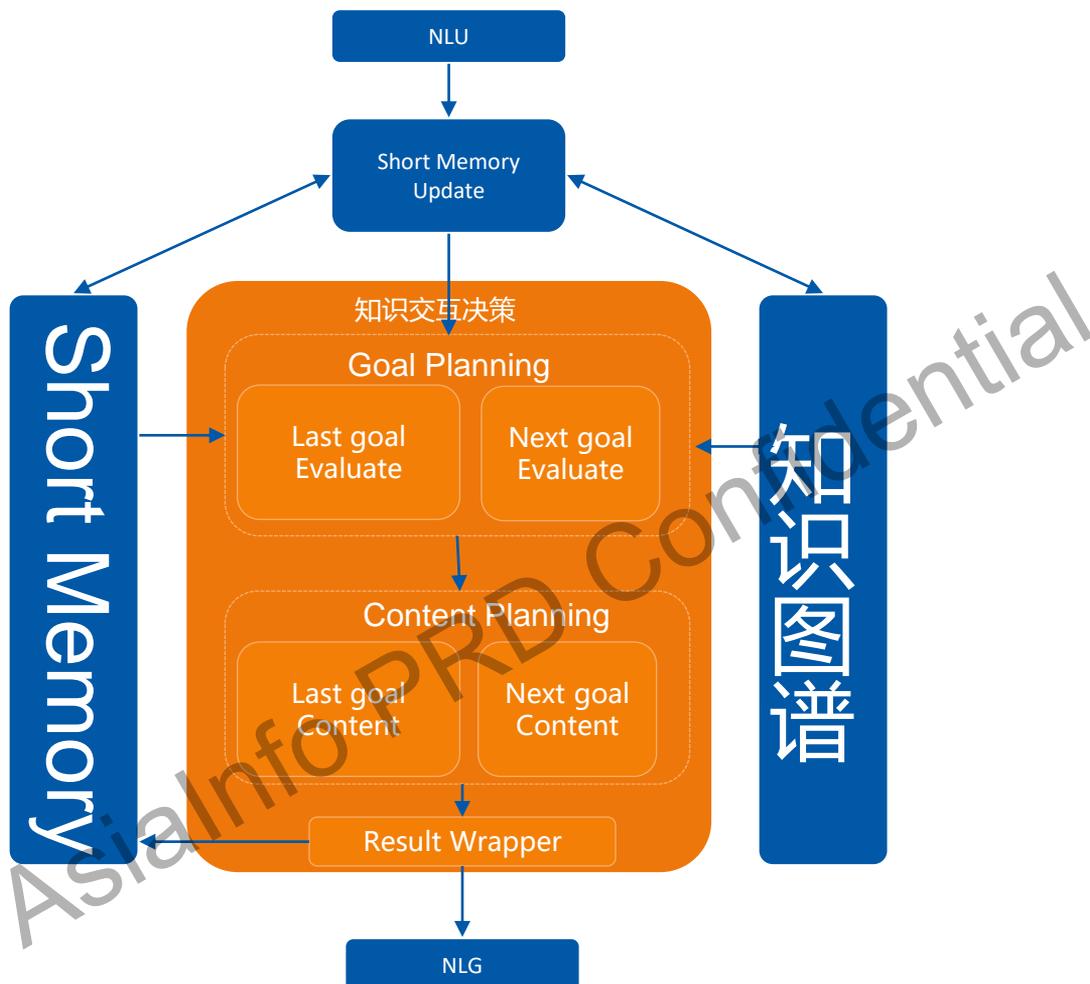
构建数据治理图谱，实现治理能力的赋能应用。



4.3.3 知识驱动的多轮对话

[返回目录](#)

借助于知识驱动的多轮对话技术，可以实现更为流畅的人机对话，并为用户提供智能化的知识协助。常用于智能客服机器人，导购机器人等诸多领域。



- High-Level Policy(Goal Planning)

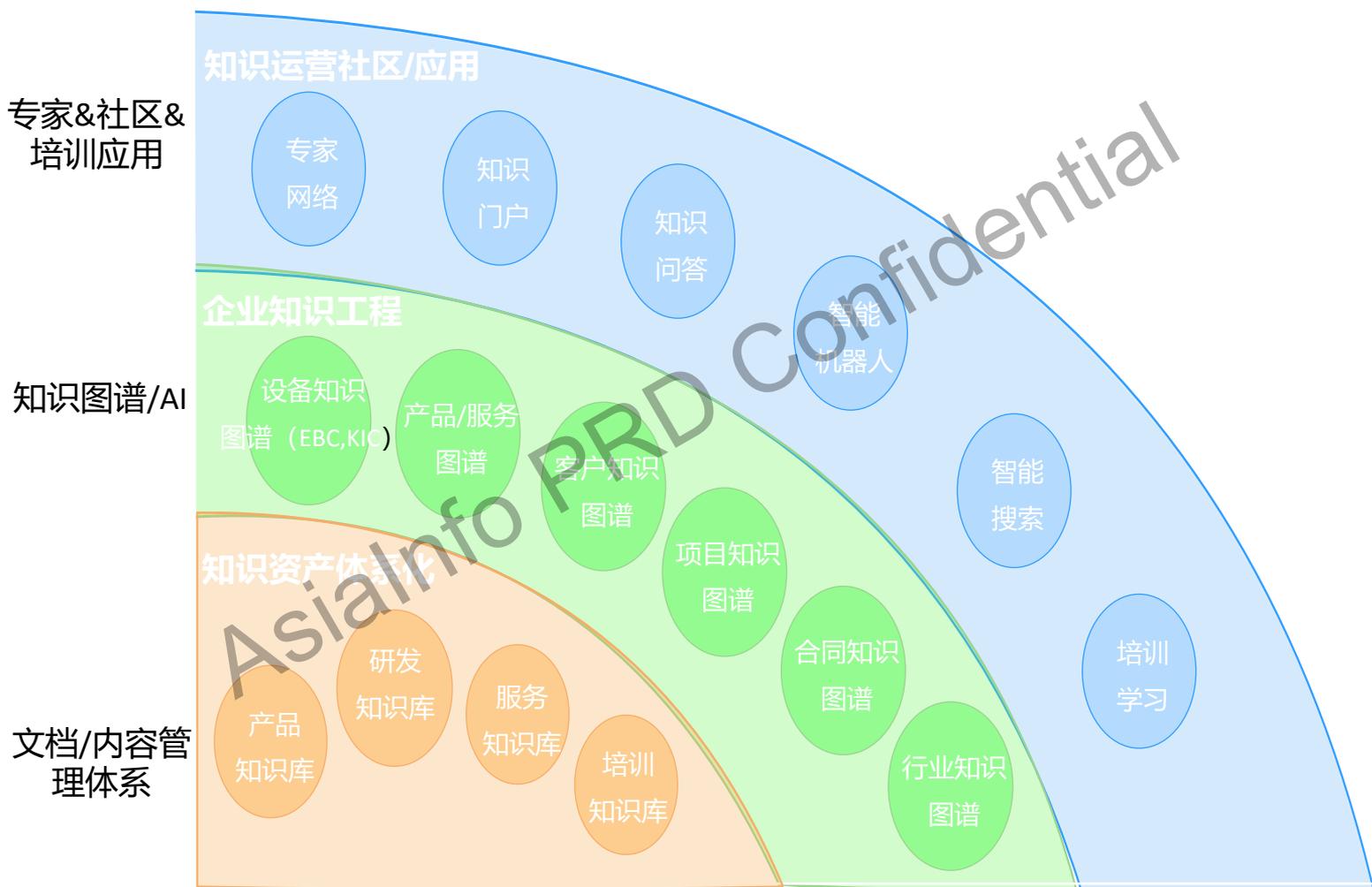
- **机器主动触发知识**：根据上下文内容主动触发知识
- **机器主动推荐知识**：根据聊天内容触发知识推荐
- **机器主动切换知识**：主动切换到关联知识上

- Low-Level Policy(Goal Planning)

- **决定回复方式**
比如：决定回复act方式，是询问，还是提供信息等
- **领域相关的先验规划**
比如：推荐一部电影时，优先给出剧情简介
- **基于知识图谱提供答案**
比如调用问题答案结果

4.3.4 多模态企业知识库

- 以非结构化的文档和内容作为为知识库的数据源，构建结构化的知识图谱
- 以构建好的知识图谱来服务于各类企业知识应用场景



4.3.5 AI集成支持知识抽取融合推理

通过集成大量数据挖掘算法和能力商城提供的优质算法，降低算法使用门槛，实现知识抽取、融合、推理计算。基于AI平台提供的词向量学习、强化学习，图神经网络等AI算法，支撑更高级的智能推理能力。



五. 功能介绍

5.1. 基础功能

功能点	功能点描述
知识获取	<p>知识获取分为数据集管理和知识采集，功能包括：</p> <ol style="list-style-type: none"> 1、数据集新增； 2、数据集目录树管理； 3、数据集业务口径描述； 4、语句中变量支持； 5、文件名规则通配符支持； 6、关系数据库采集； 7、本地文件手动采集； 8、FTP/SFTP远程文本采集； 9、数据实时采集； 10、网络爬虫网络公开数据采集； 11.图片、语音、非结构化文本数据采集；
知识建模	<p>知识建模首先需要对知识的逻辑组织结构，即本体 (Schema)进行设计。功能包括：</p> <ol style="list-style-type: none"> 1、图形化边界本体； 2、本体关系图形化展示； 3、本体版本管理； 4、本体审批上线管理； 5、公有本体库同步； 6、百科类本体库扩充； 7、本体构建规范定义；

[返回目录](#)

功能点	功能点描述
图谱生产	<p>图谱生产线从数据抽取清洗到图谱生成，主要功能包括：</p> <ol style="list-style-type: none"> 1、图谱项目层级管理； 2、图谱项目结构展示； 3、图谱版本管理； 4、多类型数据清洗实现数据结构化； 5、映射任务定义； 6、映射任务批量导入； 7、映射任务手工执行和监控； 8、融合任务规则定义； 9、融合任务手动运行； 10、融合任务算法集成； 11、融合任务监控；
图谱发布	<p>开源图数据库和商业图数据库的不同特性，知识图谱平台支持发布到多类图数据库主要功能：</p> <ol style="list-style-type: none"> 1、neo4j数据装载； 2、JanusGraph数据装载； 3、ElasticSearch数据装载；
调度	<p>图谱生产的整个流程中，涉及多个处理阶段，需要将这些阶段打包组装，形成一个可整体运行的任务，主要功能：</p> <ol style="list-style-type: none"> 1、任务调度配置； 2、调度任务监控；

[返回目录](#)

功能点	功能点描述
系统管理	系统管理主要能力包括： 1、用户账户管理； 2、用户权限管理； 3、用户角色管理； 4.审核管理； 5.知识图谱权限管理； 6.平台功能权限；
基础数据中心	基础数据中心涉及两个部分公共属性管理和数据源管理，功能包含： 1、公共属性管理； 2、资源信息管理； 3、集群信息管理； 4、爬虫基础信息管理；

5.2. 特色功能

5.2.1 图谱知识推理

知识推理的核心是算法，主要功能：

- 1、预置推理算法集成；
- 2、能力商城算法介入；

5.2.2 图谱浏览

通过图谱浏览器访问图谱
首页主要功能包括：

- 1、图谱总览；
- 2、知识图谱详细列表；
- 3、图谱全景图展示；
- 4、图谱能力开放；
- 5、本体结构展示；
- 6、实体排行榜；
- 7、实体分类展示；
- 8、图谱探索；

六. 场景应用方案

6.1. 疫情防控

运用知识图谱构建技术，结合运营商数据、政府公开数据，构建融合人物知识图谱、地点知识图谱的图谱联邦，搭建一个可以快速查询分析预警敏感、高危、隔离人员的应用系统，协助政府做疫情防控。

场景一

敏感人员排查

基于当前疫情形势，疫情防控进入关键阶段，春节后返工人群为疫情防控带来非常大的压力。通过应用系统识别人员身份后，可对到过疫情地区或高风险地区的返工人员快速排查，及时预警，提升防控一线现场核查的工作精度和效率

敏感人员图谱

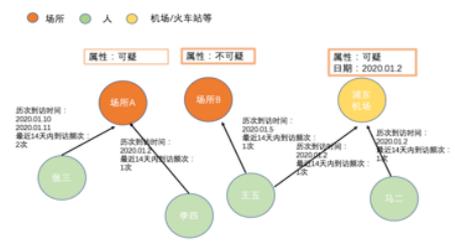


场景二

高危人群分析预警

基于卫健委发布的已确诊患者乘坐公共交通、到访过的公共场所信息，和电信用户行为轨迹大数据分析，定位出和已确诊患者在同一时间空间有交集的人群，主动推送预警信息，提示居家自我隔离，有症状及时就医，降低疾病扩散风险。

高危人群图谱



场景三

隔离人员风险管控预警

对已登记或系统识别的需居家隔离人员进行监控，基于位置数据，判断是否离开隔离区域，如不遵守隔离要求，依据设定的规则推送信息劝返本人，乃至防控部门。

隔离人员管控

姓名	性别	身份证号	手机号	住址	隔离开始时间	隔离结束时间	当前状态	备注
张三	男	310101199001010001	13912345678	上海市浦东新区张江路100号	2020-01-10	2020-01-20	正常	
李四	女	310101199001010002	13912345679	上海市浦东新区张江路100号	2020-01-10	2020-01-20	异常	离开隔离区域
王五	男	310101199001010003	13912345680	上海市浦东新区张江路100号	2020-01-10	2020-01-20	正常	

6.2. AIOps智能知识问答

引入知识图构建技术，基于投诉工单、问题单、事件单等工单样本数据，对历史数据进行知识抽取知识挖掘，进而推理总结出问题知识库，供后续类似问题自动化处理或者人工处理时提供参考，支持运维知识应答能力，提升故障处理效率。



6.3. 识别潜在社群支撑魔百盒点播推荐

社交网络精准营销以客户交往群划分为营销推荐的基础，同时对客群的特性进行分析发掘，不同拓扑结构的客群具有不同的特性，处于一个客群不同位置的用户对该客群也有不同的影响。

引入知识图谱技术打造全视角“社交点播类图谱”，针对视屏节目魔百盒，结合“社群影响因子”和“节目热度”，分析其“推荐指数”，分析潜在社群进行扩展营销

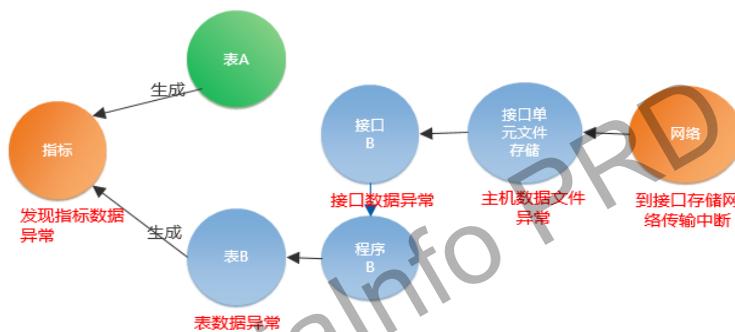


6.4.数据资产快速溯源，发现资产异常变化

数据资产图谱是以各类元模型为核心实体，通过采集、转换和扩展现有系统的资产元数据，构建资产实体和资产关系为主体的关系网络，并扩展资产信息、资产事件信息，形成一个初步的数据资产图谱。并通过数据资产图谱提供资产可视化查询、图关系全景展现、隐含关系发现、事件影响等应用，帮助数据中心建设人员迅速了解数据资产信息。

● 辅助数据质量问题进行快速溯源

利用数据资产知识图谱完整的实体间关系，基于图路径遍历算法跨实体域发现数据资产异常情况。



如左图所示：系统发现某一指标存在异常，自动检索与其相关的表并进行检查，由此发现表B存在异常并继而发现接口及其接口机存储存在异常，通过检索与接口机存储相关的实体，发现存在网络中断的问题。最终通过溯源，跨数据实体域、资源实体域发现了指标存在异常的最终原因。

七. 带给客户的价值

- 规范化图谱开发流程，降低开发难度，提升开发效率，减轻客户成本投入。
- 提供统一管理模块，增强图谱管控，提高客户领域本体、图谱项目的有效性、可用性。
- 通过图谱浏览器，加强了图谱对外开放能力，提升图谱应用价值。
- 为医疗行业客户打造出可预测、可预防、精准医疗的健康解决方案。
- 为金融行业客户针对风险评估、预测、反欺诈、精准营销、智能搜索等提供技术支撑。

AsialInfo PRD Confidential

八. 产品优势

亚信AISWare Onta KG产品的优势集中体现在：

多元化场景抽取：从知识的数据结构和抽取的手段出发，评估用户能够出现的场景，实现抽取全场景覆盖。

多维建模能力：通过自顶向下和自底向上的两种途径相结合构建本体模型。知识建模过程是构建知识图谱的基础。集成多开放域高质量本体模型，避免不必要、重复性的知识获取，提高知识图谱构建效率降低领域数据融合成本。

多元融合算法：提供多种融合算法：主键法、相似度法，并结合算法能力商城支撑融合算法的扩展。

商城化数据推理：通过已集成大量数据挖掘算法和能力商城提供的优质算法，降低算法使用门槛，让图数据挖掘变得简单易懂，从而降低人员投入。

AsialInfo PRD Confidential

[返回目录](#)

九. 联系我们

亚信科技（中国）有限公司

地址：北京市海淀区中关村软件园二期西北旺东路10号院东区亚信大厦

邮编：100193

传真：010-82166699

电话：010-82166688

Email：5G@asiainfo.com

网址：www.asiainfo.com



Thank you



亚信科技依托产品、服务、运营、集成能力助力企业数字化，持续创造新价值。

亚信科技（中国）有限公司保留所有权利