# 基于知识图谱技术的告警关联性分析可行性报告

知识图谱本质上是语义网络（Semantic Network）的知识库，从实际应用的角度可简单理解为多关系图（Multi-relational Graph）。知识图谱是真实世界的语义表示，其中每一个节点代表实体，连接节点的边则对应实体之间的关系，异构数据通过整合表达为知识，图的表达映射了对世界的认知方式。因此知识图谱非常适合整合非结构化数据，从零散数据中发现知识，从而帮助组织机构实现业务智能化。

可见知识图谱的主要作用在于分析关系，尤其是深度的关系。而在光传输网中，告警间存在着特定的关系，且异常复杂——根源性告警一般会衍生出一些低级别告警，这些告警被称为衍生告警。衍生出的低级别告警对告警定位和处理造成了极大的干扰。告警相关性包括路径上的告警相关性和网元上的告警相关性。因此可以利用知识图谱技术对告警进行相关性分析，有效或者准确地屏蔽衍生告警，压缩告警数量，从而能够快速定位故障。

## 告警涉及的实体及告警间关系

在知识图谱里，通常用“实体（Entity）”来表达图里的节点、用“关系（Relation）”来表达图里的“边”，实体-边-实体则作为知识图谱最基础的单元。对于告警知识图谱，实体可以为告警、故障、网络元件等，关系则用来表达不同实体之间的某种联系，比如断纤故障-“导致”-R\_LOS紧急告警、R\_LOF告警-“某时间段内置信度n%”-B1TCA等。除此之外，实体和关系都拥有各自的属性，比如T\_ALOS告警表示E1或T1接口模拟信号丢失，告警级别为重要告警，告警类型为通信告警等，如图1所示。当一个知识图谱拥有属性时，可以用属性图（Property Graph）来表示。

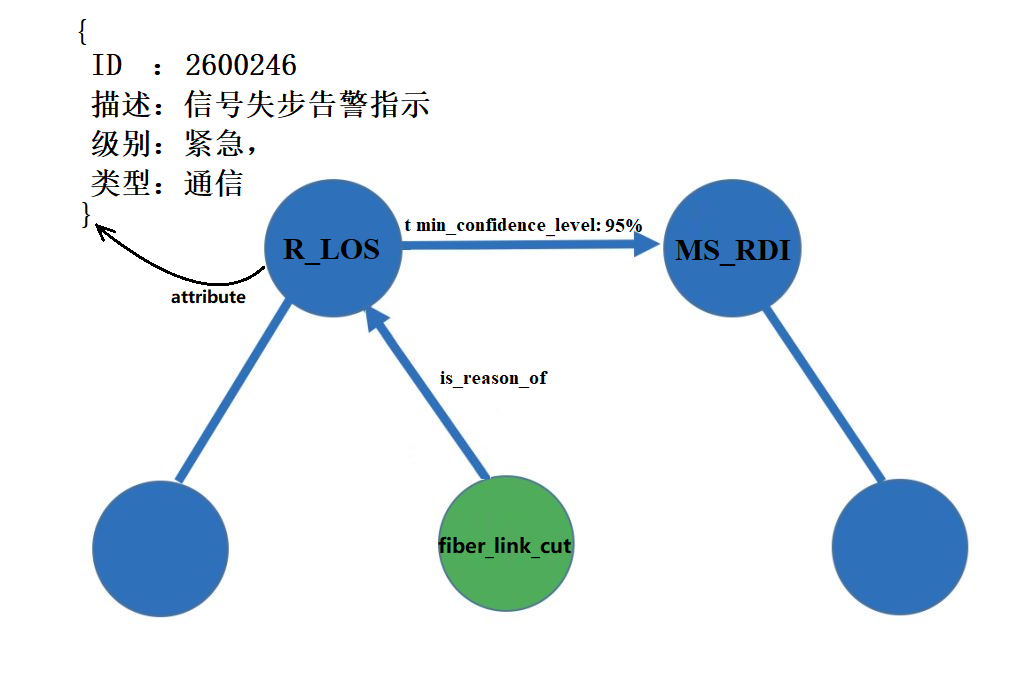


图 1 告警图谱基础单元示例

## 告警知识图谱初步构建思路

构建告警知识图谱是一个迭代更新的过程，根据知识获取的逻辑，每一轮迭代包含三个阶段，如图2所示：

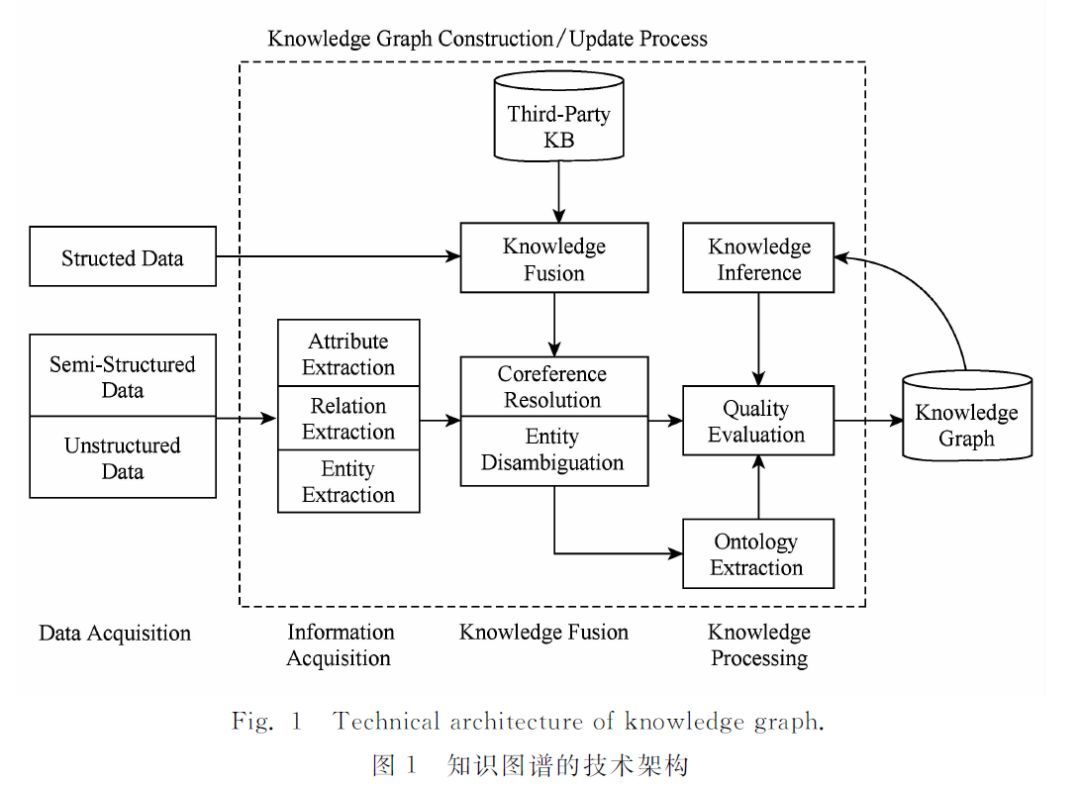


图 2 知识图谱构建技术框架

**信息抽取（Information Acquisition）**

告警知识图谱构建的前提是确定数据来源并从中抽取实体及其关系，涉及三大技术：实体抽取技术、关系抽取技术和属性抽取技术。

首先，数据源主要来自两种渠道：一种是业务本身的数据，这部分数据通常包含在各机构或者公司内的数据库表并以结构化的方式存储，比如通信运营公司的历史告警数据等；另一种通常是以网页、文档的形式存在，实体和关系藏于大片文字中，是非结构化的数据，比如网络设备资料等白皮书。前者一般只需要简单预处理即可以完成“知识抽取”，R\_LOF和MS\_RDI的置信度就可以从结构化的历史告警数据中简单计算出来；但后者一般需要借助于自然语言处理等技术来提取出实体及其关系，如图1例子里，R\_LOF和fiber\_link\_cut的关系、R\_LOF的属性信息等需从非结构化的设备资料中提炼出来。

**知识融合（Knowledge Fusion）**

此环节包括2部分内容：知识合并技术和实体链接技术。通过从各种类型的数据源抽取的各种候选实体(概念)及其属性关联，形成了一个个孤立的抽取图谱(Extraction Graphs)。为了形成一个完整的告警知识图谱，需要将这些信息孤岛集成在一起。在此环节中还需消除矛盾和歧义，比如实体光纤可能表达成光链路、光物理通道等；某个特定称谓也许对应于多个不同的实体等。

**知识加工（Knowledge Processing）**

对于经过融合的光通信网告警知识，需要经过质量评估之后（部分需要人工参与甄别），才能将合格的部分加入到告警知识库中，以确保知识库质量。

### 通过告警原因间接关联告警实体及应用思路



图 3 告警间接关联图谱示例

### 通过告警间支持度及其置信度直接关联告警实体

### 关于告警原因实体化用于故障定位的分析

### 告警实体属性初步压缩告警数量

## 基于自然语言处理技术的告警图谱的构建

信息自动抽取是知识图谱构建的前提，也是关键环节之一。信息抽取的难点在于处理非结构化数据，主要涉及以下几个方面的自然语言处理技术：

(1)实体命名识别（Name Entity Recognition）

(2)关系抽取（Relation Extraction）

(3)实体统一（Entity Resolution）

(4)指代消解（Coreference Resolution）

## 告警知识图谱上的数据挖掘

通过各种信息抽取和数据集成技术已经可以构建初具规模的告警知识图谱。为了进一步提高告警图谱的覆盖率，需要进一步在知识图谱上进行挖掘。目前已有几种基于知识图谱的挖掘技术：

(1)推理(Reasoning或Inference)

(2)实体重要性排序

(3)相关实体挖掘

(4)基于图形神经网络(GNN)的信息挖掘

用关系数据库来存储，尤其是存储简单的知识图谱，从技术上来说是完全没问题的。但若知识图谱变复杂，图数据库在关联查询的效率上会比传统的关系数据存储方式有显著的提高。据测试当涉及到2,3度的关联查询，基于知识图谱的查询效率会高出几千倍甚至几百万倍。除此之外，基于图的存储在设计上会非常灵活，一般只需要局部的改动即可。