

漏洞检索知识图谱系统

介绍资料

北京交通大学电信学院

2020年9月

* **为什么要做漏洞检索知识图谱？**

在信息化社会中，随着互联网技术的快速发展，计算机和网络在军疆、政治、金融、工业、商业、人们的生活和工作等方面的应用越来越广泛，社会对计算机和网络的依赖越来越大，Internet已经成为全球非常重要的一个基础平台。很多企业都将应用架设在该平台上，为用户提供更方便更快捷的服务。如果计算机和网络系统存在安全问题，包括系统本身及其支撑[软件](https://baike.so.com/doc/5366582.html)，网络客户和[服务器软件](https://baike.so.com/doc/5988709.html)，[网络路由器](https://baike.so.com/doc/1473989.html)和安全[防火墙](https://baike.so.com/doc/3110308.html)等，在这些不同的软[硬件](https://baike.so.com/doc/2368908.html)设备中都可能存在不同的[安全漏洞](https://baike.so.com/doc/9807297.html)问题或者计算机和网络系统的信息安全受到威胁，如web攻击等，这将会导致社会的混乱并造成巨大损失。

漏洞的存在，很容易导致[黑客](https://baike.so.com/doc/1571721.html)的侵入及病毒的驻留，会导致数据丢失和篡改、隐私泄露乃至金钱上的损失，如：网站因漏洞被入侵，网站用户数据将会泄露、网站功能可能遭到破坏而中止乃至服务器本身被入侵者控制。所以，人们越来越重视信息安全和网络安全的重要性，提前根据漏洞相关知识防范漏洞问题很有必要。

而漏洞知识又较为分散，虽然可以通过查阅相关安全漏洞库、使用搜索引擎获取知识，但是这些手段并不能精准定位用户所需要的知识，需要用户用自己的专业知识去甄别有效信息，极大地浪费了用户的精力和时间。传统方法不能充分发挥其在漏洞防范工作中的重要作用，难以体现其蕴含的丰富价值。2012 年 5 月 17 日，Google 正式提出知识图谱（Knowledge Graph），用于优化搜索引擎，增强用户的搜索质量和体验。知识图谱，本质上，是一种揭示实体之间关系的语义网络。它把所有不同种类的信息（Heterogeneous Information）连接在一起得到一个关系网络，提供了从“关系”的角度去分析问题的能力。基于知识图谱的漏洞检索系统，拥有大规模的先验背景知识，知识关联全面，能够为查询词赋予丰富的语义信息，从而理解关键字，精确回答，帮助用户更快找到所需要的信息，满足用户需求。

知识图谱是一种基于图的数据结构，由节点(Point)和边(Edge)组成。在知识图谱里，每个节点表示现实世界中存在的“实体”，每条边为实体与实体之间的“关系”。基于信息、知识和智能形成的闭环，从信息中获取知识，基于知识开发智能应用，智能应用产生新的信息，从新的信息中再获取新的知识，不断迭代，就可以不断产生更加丰富的知识图谱，更加智能的应用。构建知识图谱这个过程的本质，就是让机器形成认知能力，理解这个世界。漏洞检索知识图谱中每个漏洞信息都会有详细的各方面分析，包括漏洞名称、对漏洞的定义、漏洞严重性等级说明表、造成漏洞的原因分析、如何修复漏洞、正确的修复方法以及预防建议等。有了知识图谱，漏洞检索系统就可以真正理解用户的意图，为工作人员检索漏洞信息提供方便，充分发挥其在漏洞防范工作中的重要作用。

* **漏洞检索知识图谱的特点和创新点**

知识对于人工智能的价值就在于，让机器具备认知能力和理解能力。构建知识图谱这个过程的本质，就是让机器形成认知能力，理解这个世界。

知识图谱的特点：

1、知识图谱是由一些相互连接的实体和它们的属性构成的，基本单位为(实体1)-(关系)-(实体2)的三元组。

2、知识图谱蕴含着人类大量的先验知识，融合了所有方面，以便于用户搜索时的连贯性。

3、数据关联度-语义理解智能化程度高，数据结构化-检索效率高，能够为用户找出更加准确的信息，作出更全面的总结并提供更有深度的相关信息。

4、把与关键词相关的知识体系系统化地展示给用户。

5、基于知识图谱，漏洞检索系统有简单的推理能力，可以构建知识问答系统。

知识图谱的创新点：

1、知识图谱将搜索引擎从字符串匹配推进到实体层面，可以极大地改进搜索效率和效果。

2、最有效、最直观地表达出实体间的关系，为查询词赋予丰富的语义信息，从而理解关键字，精确回答，帮助用户更快找到所需要的信息，满足用户需求。

* **漏洞检索知识图谱的作用**

漏洞检索系统为了深度挖掘漏洞的信息价值，设计了漏洞信息检索引擎以及关联分析引擎，不仅能够从时间维度、空间维度、漏洞各字段维度进行分析，而且能够从漏洞、事件、产品的关联关系维度进行分析并实现可视化展示。

基于知识图谱的漏洞检索系统能够帮助用户找到最想要的信息，精准回答，提供最全面的漏洞相关信息，帮助用户了解实体之间的关系。而且知识图谱构建了一个与搜索结果相关的完整的知识体系，相关实体之间的关系可以形成若干条通路，如“A漏洞->关键词->B漏洞->B漏洞的补丁信息->相关工程人员……”之间的通路，用户可能会了解到某个新的实体或新的联系，让搜索更有深度和广度。

相比传统的漏洞数据库加搜索引擎，知识图谱有以下优势：

（1）关系的表达能力强

传统的安全漏洞库通常通过表格、字段等方式进行读取，而关系的层级及表达方式多种多样，且基于图论和概率图模型，可以处理复杂多样的关联分析，满足系统的各式检索需要。

（2）像人类思考一样去做分析

基于知识图谱的交互探索式分析，可以模拟人的思考过程去发现、求证、推理，业务人员自己就可以完成全部过程，不需要专业人员使用漏洞的专业知识区分过滤有效信息。

（3）知识学习

利用交互式机器学习技术，支持根据推理、纠错、标注等交互动作的学习功能，不断更新漏洞信息，沉淀知识逻辑和模型，提高系统智能性，将知识沉淀在知识图谱系统内部，降低对专业人员经验的依赖。

（4）高速反馈

图式的数据存储方式，相比传统存储方式，数据调取速度更快，图数据库可计算超过百万潜在的漏洞实体和属性分布，可实现秒级返回结果，真正实现人机互动的实时响应，让用户可以做到即时决策。

知识图谱的价值在于它能够内生地在海量漏洞相关数据中挖掘实体间的深层次关系，让实体之间的关联关系创造出更多的价值。通过知识图谱的图形化展示，可以很直观地呈现给人们错综复杂的实体关系，为工作人员前期搜集漏洞信息提供方便。也可以在知识图谱系统的基础上构建高层级的漏洞知识交互问答系统，让漏洞检索进一步迈向智能化。