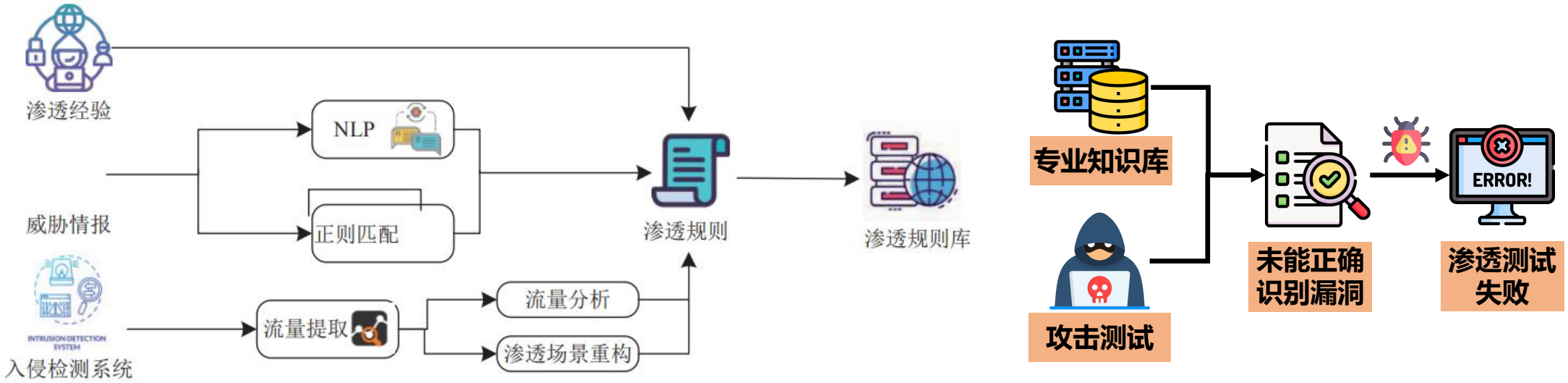
基于大模型智能体的漏洞知识可信检索与匹配方法

研究背景

- 在渗透测试过程中，对扫描获取的目标环境信息及潜在攻击面进行**高效、精准的漏洞关联分析与特征匹配**，是识别系统真实风险、构建有效测试策略的关键

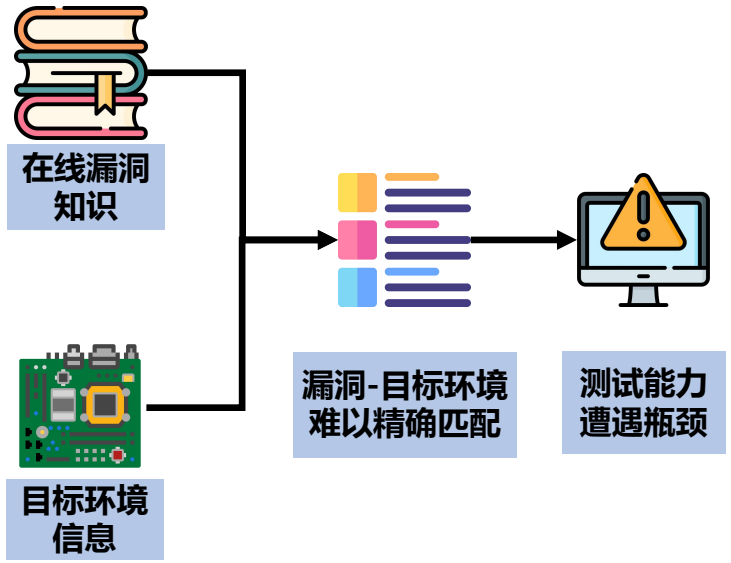


- **关键问题：**已有渗透测试方法在分析阶段受限于**人工分析的片面性**以及**漏洞知识库滞后性**，导致漏洞关联分析效率低下、特征匹配准确率不高

基于大模型智能体的漏洞知识可信检索与匹配方法

研究思路

利用大模型的强大推理能力、智能体的工具调用能力如：数据存储、在线检索、API访问、RAG检索增强等，实现漏洞知识在线可信检索及自动化的漏洞关联分析与匹配

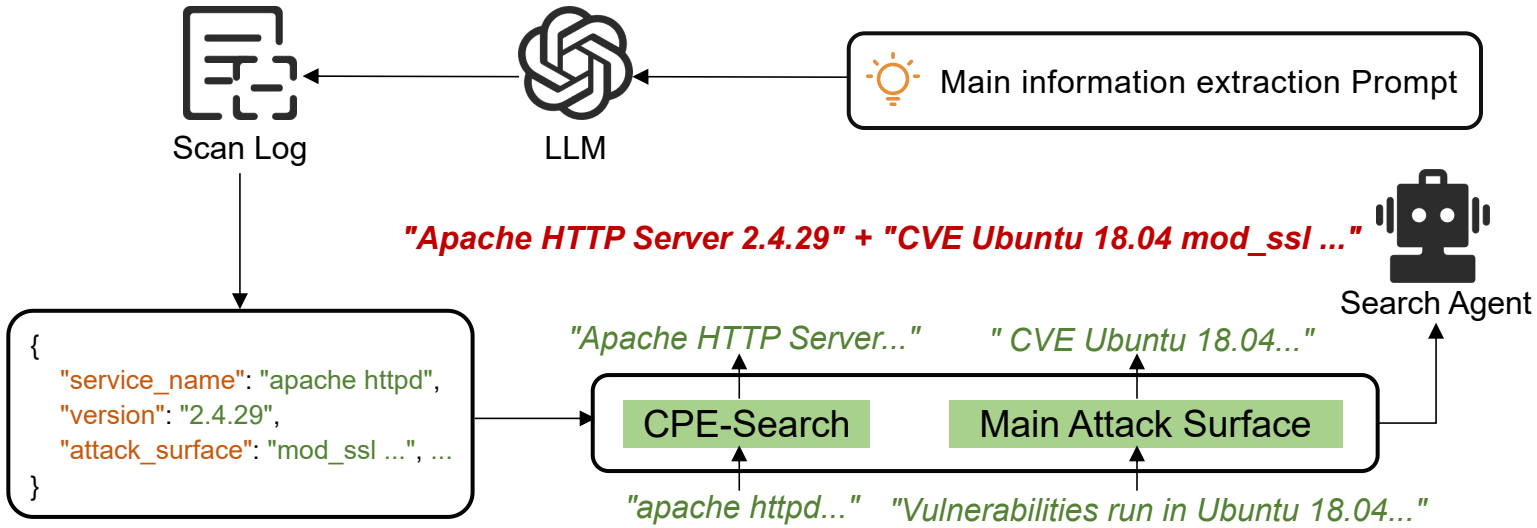


- **面临挑战：** 目标环境多样异构，大模型**缺乏特征统一提取标准**；智能体对漏洞知识的在线检索来源**缺乏可信评估**；检索内容**难以关联目标特征**，影响漏洞匹配准确性
- **解决方案：** 目标环境特征提取与检索词构建方法；基于多维可信评估的漏洞知识源过滤机制；基于语义对齐的漏洞-环境关联匹配策略

基于大模型智能体的漏洞知识可信检索与匹配方法

方案设计

针对**目标环境多样异构，大模型缺乏特征统一提取标准**等问题，提出一种**目标环境特征提取与检索词构建方法**，引导大模型提取目标环境的关键信息，并输出标准化的检索关键词



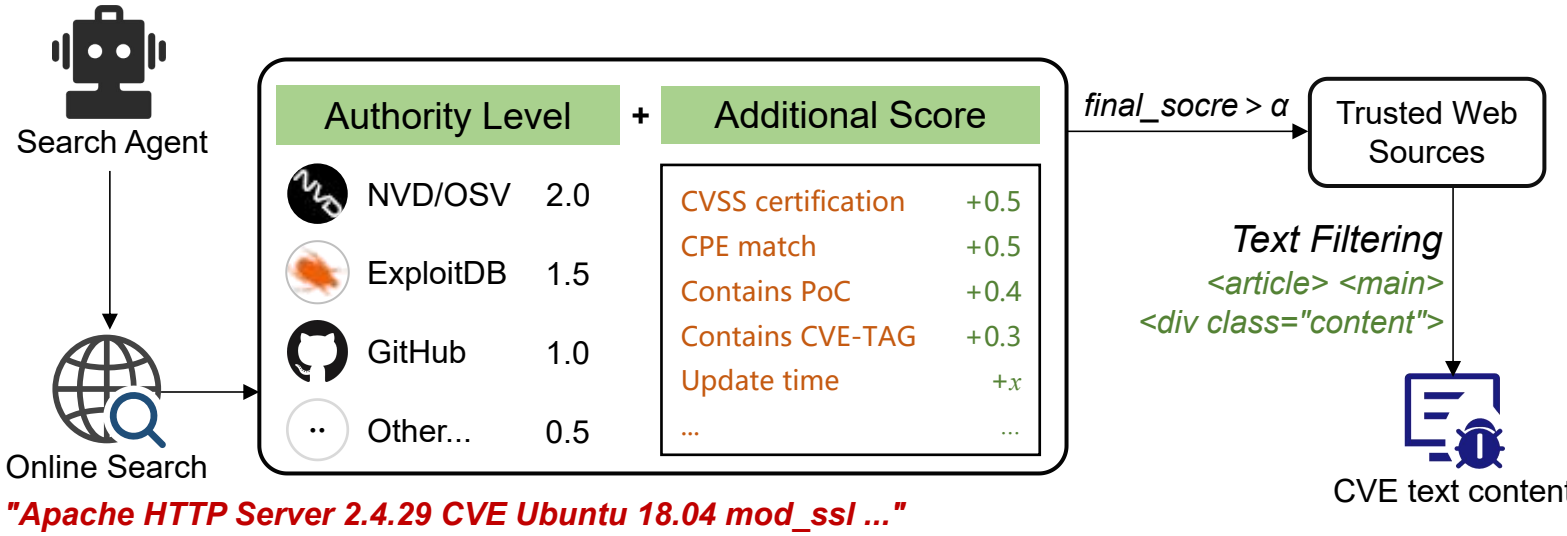
设计适配Prompt模板，引导**大模型结构化提取关键字段**，如：服务名、版本、潜在攻击面等

基于**标准化软件名**与**漏洞库高频词**生成复合查询，最终输出标准化的检索关键词至**检索智能体**

基于大模型智能体的漏洞知识可信检索与匹配方法

方案设计

- 针对**智能体对漏洞知识在线检索来源缺乏可信度评估标准**等问题，提出一种**基于多维可信评估的漏洞知识源过滤机制**，通过构建多维度评分方法，高效选择可信在线漏洞信息来源

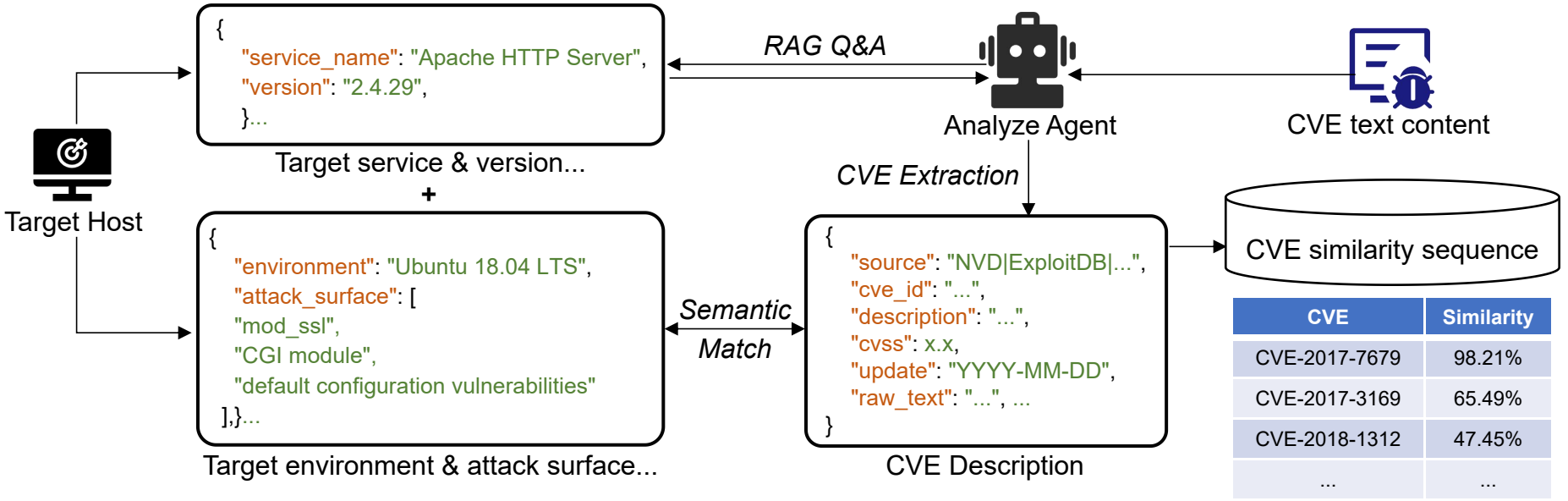


- 通过**数据源权威分级、内容完整性、时效性**等多维指标进行可信评分，保留高于阈值的信息来源
- 对可信来源中的摘要进行**CVE编号正则快速抽取**，若存在则提取该网页信息，过滤输出正文内容

基于大模型智能体的漏洞知识可信检索与匹配方法

方案设计

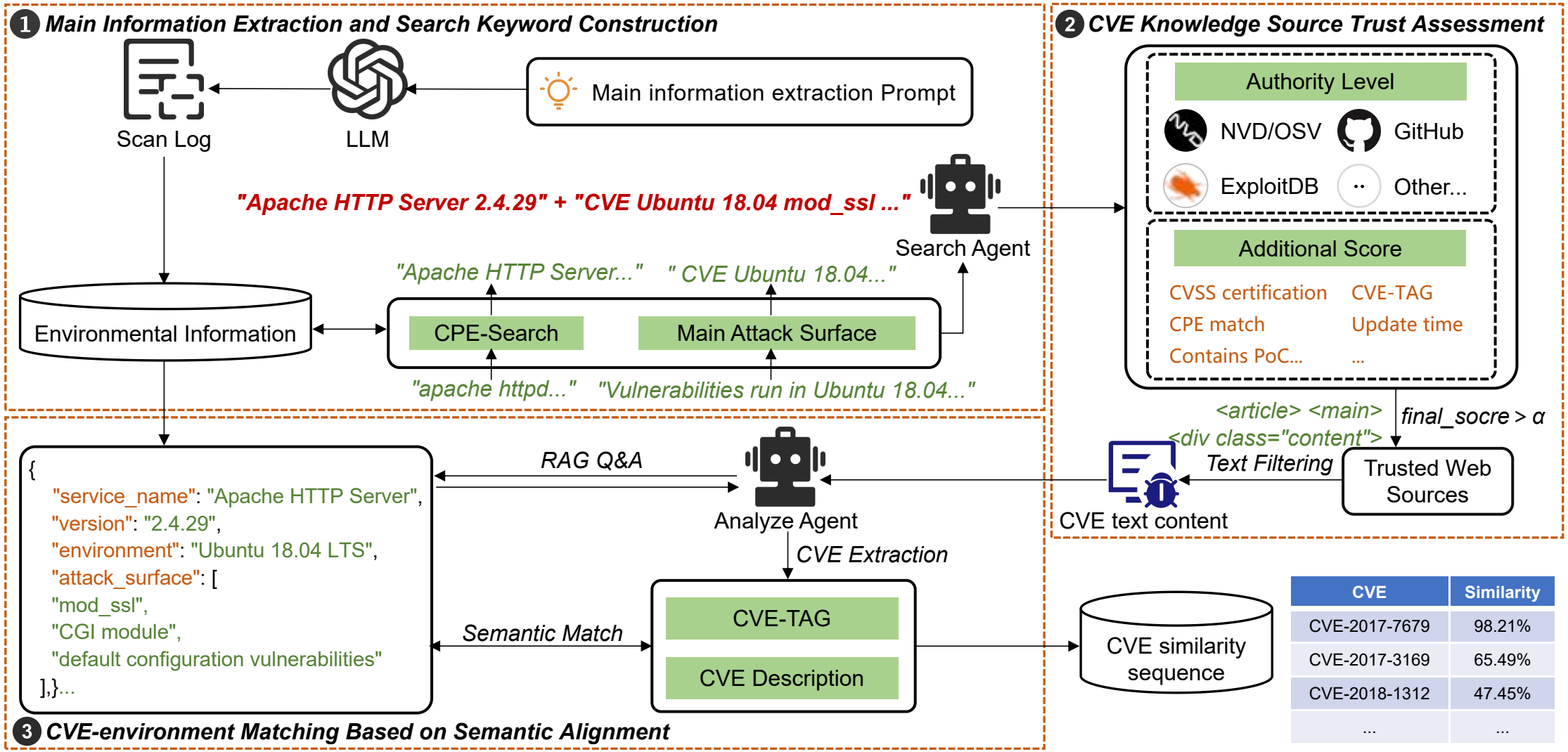
针对**大模型难以关联检索内容与目标特征，降低漏洞匹配准确性**等问题，提出一种**基于语义对齐的漏洞-环境关联匹配策略**，深度关联漏洞描述与环境特征，提高漏洞匹配准确性



设计**分析智能体调用RAG组件**，根据目标服务及版本初步挖掘CVE，并输出为**[CVE-ID+描述]**

根据目标环境特征、潜在攻击面等信息，对每个CVE描述进行**语义匹配**，输出相似度排序结果

基于大模型智能体的漏洞知识可信检索与匹配框架图



CVE	Similarity
CVE-2017-7679	98.21%
CVE-2017-3169	65.49%
CVE-2018-1312	47.45%
...	...

一种基于大模型智能体的漏洞知识可信检索与匹配方法

VulMatchAgent: Reliable Vulnerability Search and Match via LLM Agents