DefendGraph: Sistema Especialista Semântico para Alertas Wazuh com MITRE D3FEND

Salão de Ferramentas SBSeg 2026 (modalidade codigo aberto)

Eduardo K. M. do Canto[1], Felipe Thomas[1], Patrick d. S. Almeida[1], Jeferson C. Nobre[1], Luciano P. Gaspary[1], Julião Braga[3], Vinicius Tavares Guimarães[2]

[1]Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brasil

[2]Instituto Federal Sul-rio-grandense (IFSul) – Charqueadas – RS – Brasil

[3]Centro de Matemática, Computação e Cognição – Universidade Federal do ABC (UFABC)
Santo André – SP – Brasil

{ekmcanto, felipe.thomas, psalmeida,jcnobre, paschoal}@inf.ufrgs.br,

juliao.braga@ufabc.edu.br, viniciusguimaraes@ifsul.edu.br

Abstract. *The growing complexity of modern networks and cybersecurity ecosystems introduces substantial challenges for the analysis and interpretation of alerts from Security Information and Event Management (SIEM) platforms. High event volumes and heterogeneous security data hinder incident contextualization, thereby increasing the operational burden on analysts and limit timely, high-quality decisions. This article presents a semantic expert-system prototype designed to assist in the analysis of cybersecurity alerts and defensive recommendations. The approach converts native Wazuh alerts into an RDF/OWL knowledge graph enriched with the MITRE D3FEND ontology, aiming to improve explainability, traceability, and formal representation of relationships among evidence, adversarial techniques, and defensive countermeasures.*

Resumo. *O crescente grau de complexidade das redes modernas e dos ambientes de cibersegurança impõe desafios significativos à análise e interpretação de alertas gerados por ferramentas de Security Information and Event Management (SIEM). Nesses ambientes, o volume de eventos e a heterogeneidade dos dados dificultam a compreensão do contexto dos incidentes, aumentando a sobrecarga operacional dos analistas e limitando o processo de tomada de decisão. Este artigo apresenta um protótipo de sistema especialista semântico para apoiar a análise de alertas de cibersegurança e a recomendação de medidas defensivas. A abordagem proposta transforma alertas nativos do Wazuh em um grafo de conhecimento RDF/OWL enriquecido com a ontologia MITRE D3FEND, com o objetivo de ampliar a explicabilidade, a rastreabilidade e a representação formal das relações entre evidências observadas, técnicas adversárias e mecanismos de defesa.*

1. Introdução

As ferramentas *Security Information and Event Management* (SIEM) se consolidaram como componentes centrais para a coleta, correlação e análise de eventos de

cibersegurança. Apesar disso, analistas de cibersegurança ainda enfrentam uma sobrecarga analítica decorrente do elevado volume de eventos, indicadores e alertas produzidos por essas ferramentas [Tailhardat et al. 2025].

Essa sobrecarga não decorre apenas da quantidade de alertas, mas também da dificuldade de interpretar seu significado operacional em relação ao contexto mais amplo do incidente. Em muitos casos, alertas produzidos por ferramentas SIEM permanecem em um nível predominantemente operacional, exigindo que o analista relacione manualmente evidências observadas, técnicas adversárias e possíveis mecanismos de defesa.

Trabalhos como [Huang et al. 2022, Zheng et al. 2025, Mouiche et al. 2025] concentram-se na redução do esforço manual despendido por analistas na compreensão de ameaças e na tomada de decisões quanto à seleção e implementação de medidas defensivas. A proposta deste artigo difere desses trabalhos ao apresentar uma arquitetura baseada em um sistema especialista clássico e ao adotar alertas nativos gerados pela ferramenta SIEM de código-aberto Wazuh como ponto de partida.

Este trabalho recupera princípios fundamentais da tradição de sistemas especialistas, especialmente representação explícita de conhecimento, inferência dedutiva, explicabilidade, recomendação e interface ao usuário, instanciando-os em um contexto contemporâneo baseado em grafos de conhecimento (*Knowledge Graphs*), Web Semântica e ontologias públicas de cibersegurança.

Grafos de conhecimento (*Knowledge Graphs*) são relevantes neste contexto porque permitem representar entidades, relações e axiomas de domínio em estruturas consultáveis e passíveis de inferência. No domínio da cibersegurança, essa característica é reforçada pela disponibilidade de bases de conhecimento públicas, mantidas por organizações internacionais de referência, como a MITRE Corporation, que favorecem a reutilização sistemática de conhecimento especializado em cibersegurança.

O problema central investigado neste trabalho não reside na melhoria da ingestão contínua, escalabilidade ou latência dos painéis de ferramentas SIEM, mas na lacuna de explicabilidade, inferência lógica e rastreabilidade na interpretação de alertas de cibersegurança. Busca-se investigar a transformação de um alerta nativo produzido pela ferramenta SIEM de código-aberto Wazuh em um grafo de conhecimento passível de inferência, processado com tecnologias de Web Semântica (RDF, OWL, SPARQL e raciocinadores). Argumenta-se que essa abordagem, além de prestar suporte analítico ao analista, fornece uma base formal e processável por máquina para futuras extensões voltadas à automação de processos de resposta a incidentes, pesquisa alinhada com discussões recentes sobre operações, automação, inteligência e gerenciamento de rede [IAB 2025, Martinez-Casanueva et al. 2025, Tailhardat and Troncy 2025, Mackey et al. 2025, Pang et al. 2025].

A abordagem proposta consiste em um protótipo de sistema especialista semântico que recebe como entrada um alerta nativo do Wazuh e produz como saída recomendações de mecanismos defensivos. Essas recomendações são apresentadas ao usuário em dois formatos complementares: texto estruturado e um grafo esquemático que representa o encadeamento semântico da recomendação.

As principais contribuições deste trabalho são:

(i) um protótipo para transformar alertas nativos da ferramenta SIEM Wazuh em um grafo de conhecimento RDF/OWL alinhado à ontologia MITRE D3FEND;
(ii) um mecanismo de explicação baseado em questões de competência, consultas SPARQL e reconstrução dos encadeamentos semânticos que conectam o alerta analisado à recomendação defensiva;
(iii) uma avaliação experimental em cenário controlado, destinada a verificar a viabilidade da abordagem proposta.

2. Arquitetura do Sistema Especialista Semântico

A arquitetura proposta foi concebida de modo a refletir a estrutura clássica de um sistema especialista, conforme caracterizada em [Tan 2017]. A interface gráfica do protótipo foi projetada para espelhar essa base conceitual, organizando o fluxo operacional de maneira estritamente sequencial em quatro estágios (*Stages*), os quais são subdivididos em passos específicos (*Steps*).

2.1. Arcabouço Tecnológico e Componentes do Sistema

O arcabouço tecnológico adotado foi concebido e organizado de modo a viabilizar a implementação dos componentes clássicos de um sistema especialista.

Segundo [Tan 2017], a estrutura básica de um sistema especialista compreende: Base de Conhecimento, Memória de Trabalho, Máquina de Raciocínio (Mecanismo de Inferência), Interpretrador (Módulo de Explicação) e Interface Humano-Computador (Interface ao Usuário). A Base de Conhecimento armazena fatos e regras, a Memória de Trabalho registra os fatos de entrada, a Máquina de Raciocínio combina esses fatos com o conhecimento disponível para produzir novas informações e o Interpretador explica os resultados inferidos e as razões da conclusão.

Na Tabela 1, apresenta-se o mapeamento do arcabouço tecnológico em relação aos blocos funcionais teóricos que lhe servem de fundamento.

Em síntese, o arcabouço tecnológico baseia-se na linguagem de programação Python, utilizando as bibliotecas *Owlready2*, *rdflib* e *OWL-RL* para manipulação de ontologias, inferência lógica e consultas SPARQL, e as bibliotecas *Streamlit*, *Mermaid* e *Cytoscape* para o desenvolvimento da interface do usuário e visualização das recomendações em formato textual, de gráficos e de grafos.

2.2. Descrição dos Estágios (*Stages*) e Passos (*Steps*)

Com o objetivo de conciliar flexibilidade com rigor metodológico, a execução das etapas do processo é parametrizada, sempre que viável, por meio de arquivos de configuração independentes (*templates*). Destaca-se, adicionalmente, que cada passo da arquitetura opera com entradas e saídas materializadas em formatos de arquivo padronizados. A geração desses arquivos a cada transição de estado não apenas modulariza o sistema, como também viabiliza a auditoria completa de todo o ciclo de inferência.

A Figura 1 apresenta a arquitetura e o fluxo de execução dos estágios (*Stages*) e de seus respectivos passos (*Steps*).

Tabela 1. Mapeamento dos componentes de um sistema especialista (SE) no protótipo.

Componente SE	Implementação no protótipo	Tecnologias
Base de Conhecimento	A ontologia MITRE D3FEND é processada via as bibliotecas Python *rdflib*, *Owlready2* e *OWL-RL*	d3fend.owl (MITRE D3FEND), *rdflib*, *Owlready2* e *OWL-RL*
Memória de Trabalho	O fato primário é um arquivo JSON que representa um alerta nativo do Wazuh. Resultados intermediários persistidos por etapa.	Alerta Wazuh (JSON) e arquivos de estado (`.json`, `.rdf`, `.owl`, `.html`, `.mmd` e `.md`)
Mecanismo de Inferência	A consistência lógica é verificada pelo *reasoner* Pellet, via *Owlready2*. As relações implícitas são materializadas pelo OWL 2 RL, via *OWL-RL*. As consultas ao grafo seguem *Competency Questions* (CQs) definidas em SPARQL e executadas com *rdflib*.	Pellet (*Owlready2*), OWL 2 RL *OWL-RL*, SPARQL (*rdflib*)
Módulo de Explicação	Converte as deduções em narrativas estruturadas ao usuário, conforme *templates* predefinidos: texto em Markdown (`.md`) e diagramas em Mermaid (`.mmd`).	Markdown (`.md`) e diagramas Mermaid (`.mmd`).
Interface ao Usuário	Gerencia entradas, resultados, explicações e notificações ao usuário.	*Streamlit*, *Cytoscape*, *Mermaid*

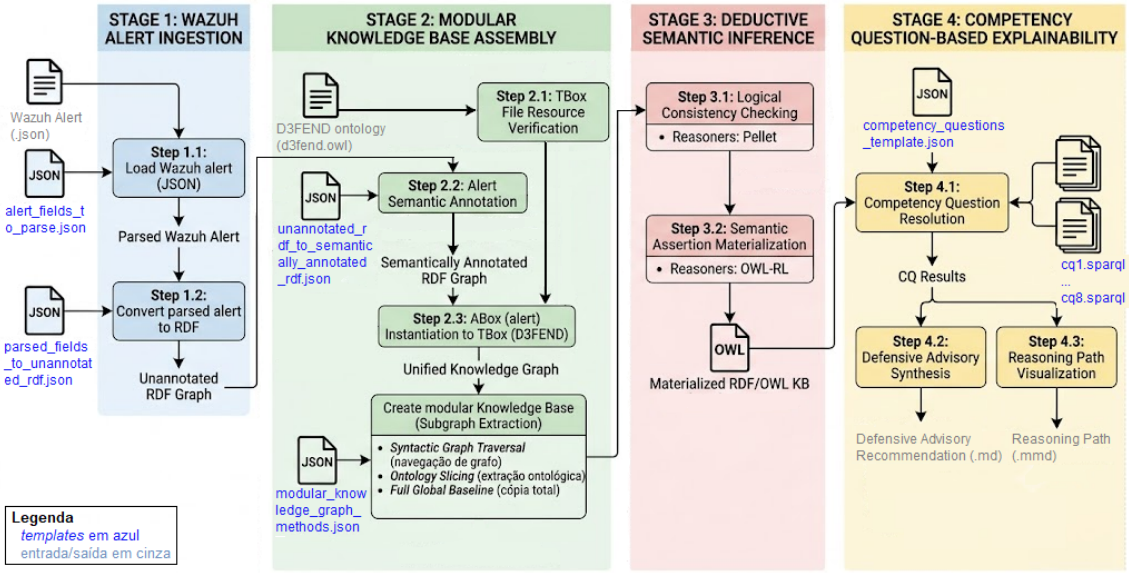

Figura 1. Arquitetura e fluxo de execução dos estágios (*Stages*) e passos (*Steps*)

3. Wazuh em Ambiente Experimental para Geração de Alertas

O ambiente experimental foi concebido para avaliar a abordagem proposta em um cenário controlado de geração e enriquecimento semântico de alertas de cibersegurança. Esse

ambiente é composto por um Wazuh Manager e Wazuh Agents. A Figura 2 apresenta o ambiente experimental.

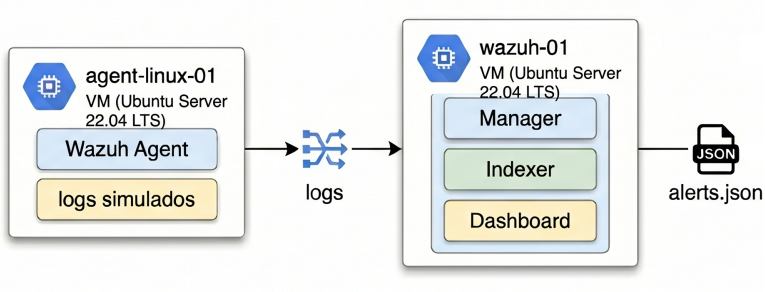

Figura 2. Visão geral do ambiente experimental

A configuração da plataforma Wazuh e a simulação dos ataques seguiram as diretrizes descritas nos casos de uso documentados oficialmente pela Wazuh, em especial no guia *Proof of Concept guide*, garantindo que os alertas produzidos sejam representativos de cenários reais de operação. Nesta versão do estudo, a avaliação empírica concentrou-se em dois alertas: uma tentativa de *password guessing* e uma tentativa de *SQL injection*.

4. Resultados e Discussão

Os resultados são apresentados em três partes. Inicialmente, descrevem-se os resultados obtidos a partir dos estágios funcionais do protótipo. Em seguida, discute-se a avaliação da recomendação defensiva produzida pela abordagem. Por fim, apresentam-se a discussão metodológica e as principais limitações do estudo.

4.1. Resultados por Estágio

4.1.1. Stage 1: Wazuh Alert Ingestion

A separação do *Stage 1* mostrou-se adequada para delimitar o escopo dessa etapa, evidenciando que, nesse momento, ocorre apenas uma conversão estrutural de formatos, do JSON para RDF. O RDF gerado ainda não corresponde a um vocabulário formal, a uma ontologia OWL ou a uma representação semanticamente anotada com MITRE D3FEND.

4.1.2. Stage 2: Modular Knowledge Base Assembly

O alerta convertido em RDF no *Stage 1* é anotado semanticamente e incorporado à base de conhecimento. Nesta implementação, foram investigadas duas abordagens principais: (i) o uso do identificador MITRE ATT&CK do alerta nativo do Wazuh, que permite vincular o alerta e uma técnica ofensiva ATT&CK já representada na ontologia D3FEND; e (ii) o estabelecimento de um vínculo com o artefato digital efetivamente monitorado pelo agente Wazuh, segundo o conceito de Artefatos Digitais (*Digital Artifact Ontology* - DAO) definido na D3FEND.

Embora ambas as abordagens de anotação semântica dependam de conhecimento especializado humano prévio, essa dependência ocorre em pontos distintos. A vinculação

via MITRE ATT&CK ID baseia-se no mapeamento das regras XML do Wazuh, responsáveis por associar determinados alertas a uma ou mais técnicas ofensivas. Já a vinculação via DAO requer a construção de um dicionário interno no protótipo, capaz de associar cada tipo de alerta Wazuh ao artefato digital correspondente da D3FEND. Assim, a primeira abordagem reutiliza a classificação ofensiva já incorporada ao SIEM, enquanto a segunda transfere a curadoria para a própria estrutura do protótipo.

No *Step 2.4: Create Modular Knowledge Base (Subgraph Extraction)*, investigou-se a modularização da base antes da inferência. Embora não tenha sido realizada uma comparação exaustiva entre os métodos, a etapa explicita uma decisão metodológica relevante: preservar a completude semântica ou restringir previamente o escopo inferencial, com impacto direto no custo computacional do *Stage 3*.

4.1.3. Stage 3: Deductive Semantic Inference

No primeiro passo do *Stage 3*, a verificação de consistência lógica constitui uma boa prática no tratamento de ontologias OWL. No caso concreto, essa verificação apoiou, sobretudo, a investigação aberta no *Step 2.4* sobre enviar a base ontológica completa ou apenas um subgrafo modularizado ao estágio de inferência.

No *Step 3.2*, realizou-se a materialização semântica com o *reasoner* OWL-RL da biblioteca `owlrl` para explicitar relações inferidas a partir do grafo de conhecimento da etapa anterior. Essa etapa mostrou-se necessária porque consultas SPARQL com `rdflib` consideram apenas triplas explícitas, ignorando relações implícitas da semântica OWL.

Durante o desenvolvimento do protótipo, buscou-se aproximar os resultados em Python daqueles do GraphDB, usado como referência para verificar códigos e consultas. Nessa perspectiva, o perfil OWL-RL mostrou-se mais adequado à metodologia deste trabalho do que o *reasoner* Pellet via `Owlready2`.

Mesmo em um cenário controlado, com apenas um alerta incorporado à base construída a partir da ontologia D3FEND, a materialização do *Step 3.2* apresentou o principal gargalo temporal do protótipo. Em um computador com processador Intel Core i5-4200M, 8 GB de RAM e armazenamento SSD, cada execução levou cerca de cinco minutos. Esse resultado sugere que o custo computacional não decorre do volume de dados acrescentado pelo alerta, mas sim do processamento inferencial sobre toda a base ontológica, dada sua estrutura, expressividade e quantidade de relações a serem materializadas. Portanto, o gargalo observado reforça a importância do debate no *Step 2.4*.

4.1.4. Stage 4: Competency Question-Based Explainability

Nesta versão, o relatório de recomendação defensiva baseou-se em oito *Competency Questions*. Duas delas recuperam informações diretamente vinculadas ao alerta processado. Embora essas informações de alerta pudessem ser obtidas em etapas anteriores do fluxo, diretamente na camada de aplicação, sua recuperação pelo grafo de conhecimento integrado mostrou-se arquiteturalmente mais adequada.

O uso de Markdown demonstrou ser adequado para estruturar um relatório textual

dinâmico, mantendo legibilidade e simplicidade na geração. O protótipo também produziu uma visualização esquemática do caminho de raciocínio em Mermaid, que sintetiza a sequência entre alerta, evidência, artefato digital e recomendação defensiva. Por decisão de escopo, a visualização em Mermaid enfatiza que o grafo de conhecimento atua como base para processamento e consulta, enquanto o diagrama visual desempenha uma função explicativa e didática sobre o processo dedutivo.

4.2. Avaliação da Recomendação Defensiva

Como o problema investigado não envolve escalabilidade nem correlação de grandes volumes de eventos, a análise concentrou-se na coerência do encadeamento produzido pelo sistema. Nesse sentido, a recomendação foi considerada adequada por manter uma cadeia semântica verificável entre o alerta original, a evidência observada, o artefato digital associado e a técnica defensiva da D3FEND. Essa estrutura permitiu confirmar que a medida sugerida não resultava de associação arbitrária, mas de um percurso consultável no grafo de conhecimento.

Adicionalmente, a recomendação foi comparada com os recursos disponibilizados pelo MITRE D3FEND, como a planilha em formato textual e os pontos de acesso (*endpoints*) da interface de programação de aplicações (API) relacionados a artefatos, técnicas ofensivas, táticas e técnicas defensivas.

Como decisão de escopo, o relatório de recomendação defensiva foi direcionado principalmente às táticas *Harden* e *Detect* da D3FEND. Essa delimitação conferiu maior objetividade e concisão ao resultado apresentado ao usuário. Contudo, essa escolha também restringe deliberadamente o alcance da recomendação, deixando em segundo plano táticas como *Isolate*, *Deceive*, *Evict* e *Restore*. Portanto, a ênfase em *Harden* e *Detect* deve ser entendida como uma decisão metodológica adequada ao escopo do protótipo, e não como uma limitação conceitual da D3FEND.

4.3. Discussão Metodológica e Limitações

A execução dos estágios do protótipo mostra a importância de distinguir conversão de formatos, anotação semântica, inferência ontológica e consultas ao grafo de conhecimento. Essa distinção é metodologicamente relevante porque explicita o papel de cada etapa na arquitetura.

A incorporação de conhecimento adicional ao modelo informacional não deve ser tratada apenas como expansão quantitativa do grafo. Ela requer uma definição explícita da intenção analítica do sistema, isto é, se a recomendação deve ser mais abrangente, mais restritiva, mais operacional ou mais explicativa para o usuário. Quanto maior o volume de relações incorporadas, maior também a necessidade de critérios metodológicos para seleção, filtragem, priorização e validação das recomendações.

Embora o protótipo esteja ancorado em uma arquitetura definida e tenha sido testado com o objetivo específico de produzir recomendações defensivas a partir de alertas nativos do Wazuh, sua avaliação permanece limitada. Ainda assim, a abordagem de vincular regras nativas de geração de alertas da plataforma Wazuh a *Digital Artifacts Objects* da MITRE D3FEND mostrou-se promissora, pois se alinha à forma como a própria D3FEND estrutura e mantém relações entre artefatos digitais, técnicas ofensivas e mecanismos defensivos.

5. Planejamento para Demonstração

Para a demonstração no Salão de Ferramentas, serão necessários dois monitores e acesso à Internet. Os autores disponibilizarão dois laptops, organizados em estações complementares.

Na primeira estação, um laptop será conectado ao ambiente experimental Wazuh hospedado no Google Cloud. Além da apresentação dos alertas nativos em formato JSON, será possível demonstrar sua geração no ambiente descrito na Seção 3. Na segunda estação, será executado o protótipo do sistema especialista semântico, permitindo acompanhar o processamento dos alertas, a construção do grafo de conhecimento e a geração das recomendações defensivas.

A demonstração terá como foco os conceitos de grafos de conhecimento, Web Semântica, ontologia MITRE D3FEND e alertas Wazuh (SIEM). Essa organização permitirá demonstrar, de forma integrada, a motivação, a arquitetura e os principais resultados da linha de pesquisa.

O código-fonte e a documentação encontram-se disponibilizados em repositório[1], enquanto o vídeo de demonstração está publicado em uma plataforma[2] de compartilhamento de vídeos.

6. Considerações Finais

A principal contribuição deste trabalho consiste em estruturar a recomendação defensiva como um processo rastreável. A arquitetura proposta, inspirada em sistemas especialistas, aproxima a abordagem dos princípios de inteligência artificial explicável, na medida em que a justificativa das recomendações candidatas decorre do percurso semântico explicitamente representado, em vez de depender de mecanismos posteriores de interpretação para justificar uma saída pouco transparente.

Apesar dos resultados obtidos, a avaliação permanece limitada a um cenário controlado, com número reduzido de alertas e sem validação sistemática por especialistas humanos. Assim, o protótipo deve ser compreendido como uma demonstração arquitetural da rastreabilidade e da explicabilidade do processo de recomendação defensiva, e não como uma validação operacional da redução da carga analítica ou da superioridade das recomendações em relação a métodos alternativos.

Agradecimentos

Os autores agradecem ao Instituto de Informática da Universidade Federal do Rio Grande do Sul (UFRGS) pelo Programa de Pós-Graduação em Ciência da Computação. Os autores declaram, adicionalmente, o uso de ferramentas de inteligência artificial como suporte complementar ao processo de pesquisa e redação. O modelo ChatGPT foi empregado para revisão ortográfica e gramatical e no aprimoramento da fluidez, da coesão textual e da clareza da redação em língua portuguesa. O modelo Gemini Banana Pro foi utilizado na geração de imagens. O ChatGPT também foi empregado como apoio ao desenvolvimento do código-fonte do protótipo em linguagem Python.

[1] `https://github.com/felipethomas82/DefendGraph`
[2] `https://www.youtube.com/watch?v=mpG0RzIeado`

Referências

[Huang et al. 2022] Huang, C.-C., Huang, P.-Y., Kuo, Y.-R., Wong, G.-W., Huang, Y.-T., Sun, Y. S., and Chang Chen, M. (2022). Building Cybersecurity Ontology for Understanding and Reasoning Adversary Tactics and Techniques. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 4266–4274.

[IAB 2025] IAB (2025). Report from the iab workshop on the next era of network management operations (nemops). Internet-Draft draft-iab-nemops-workshop-report-04, IAB. Work in progress. Expires: March 2026.

[Mackey et al. 2025] Mackey, M., Claise, B., Voyer, D., Graf, T., and Keller, H. (2025). Knowledge graph framework for network operations. Internet-Draft draft-mackey-nmop-kg-framework-02, IETF. Work in progress.

[Martinez-Casanueva et al. 2025] Martinez-Casanueva, I. D., Claise, B., and Mackey, M. (2025). Knowledge graph construction from network data sources. Internet-Draft draft-marcas-nmop-knowledge-graph-yang-04, IETF. Work in progress.

[Mouiche et al. 2025] Mouiche, I., Merbouh, H., and Saad, S. (2025). Context-aware entity-relation extraction pipeline for threat intelligence knowledge graphs. *Available at SSRN 5152241*.

[Pang et al. 2025] Pang, R., Zhao, J., and Zhang, S. (2025). Knowledge graph for network traffic monitoring and analysis. Internet-Draft draft-pang-idr-kg-traffic-monitoring-00, IETF. Work in progress.

[Tailhardat et al. 2025] Tailhardat, L., Chabot, Y., and Troncy, R. (2025). Anomaly Detection using Knowledge Graphs: A Survey for Network Management and Cybersecurity Application. Preprint, HAL. HAL Id: hal-04930539; revised version, 2026.

[Tailhardat and Troncy 2025] Tailhardat, L. and Troncy, R. (2025). Knowledge graphs for enhanced cross-operator incident management and network design. Internet-Draft draft-tailhardat-nmop-kg-cross-operator-incident-01, IETF. Work in progress.

[Tan 2017] Tan, H. (2017). A brief history and technical review of the expert system research. *IOP Conference Series: Materials Science and Engineering*, 242:012111.

[Zheng et al. 2025] Zheng, M., He, T., Wang, N., Wang, X., and Zuo, J. (2025). Network Security Threat Detection System Based on Knowledge Graph. In *2025 40th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pages 998–1004.