

EKGViewExplorer: Uma Ferramenta para Exploração dos Dados e Metadados da Visão Semântica em Enterprise Knowledge Graphs *

José Renato S. Freitas¹, Vânia P. Vidal¹, Tulio V. Rolim¹, Narciso Arruda¹

¹Departamento de Computação – Universidade Federal do Ceará (UFC)
Campus Pici – Bloco 910 – Fortaleza – CE – Brasil

vvidal@lia.ufc.br, {jrenatosfreitas,tulioxcrtf,narcisoarruda}@gmail.com

Abstract. *The semantic view of an Enterprise Knowledge Graph (EKG) offers a semantically coherent, integrated perspective over an organization's heterogeneous data sources. This paper presents EKGViewExplorer, an interactive graphical tool that allows users to navigate both the data and metadata in EKGs built with the DDP_SV, a data design pattern that logically organizes semantic view data and metadata into a four-level hierarchy, enabling scalable maintenance and context-aware exploration. At the heart of this architecture lies the metadata graph, encoded with the VoSV vocabulary, which captures detailed, machine-readable information about structure, provenance, and quality of semantic view. By coupling intuitive exploration capabilities with comprehensive metadata capture, EKGViewExplorer empowers data stewards and domain experts to understand, validate, and evolve enterprise semantic views more effectively.*

1. Introdução

Um *Enterprise Knowledge Graph* (EKG) oferece uma base robusta para a gestão do conhecimento, integração de dados e análises avançadas, ao fornecer uma visão semântica que une fontes de dados heterogêneas presentes no *data lake* de uma organização [Gal-kin et al. 2017]. O objetivo da visão semântica em um EKG é fornecer uma estrutura ontológica unificada, resultante da integração semântica de tais fontes de dados.

Para enfrentar os desafios da construção de uma visão semântica no contexto de *big data*, [Vidal et al. 2024] propuseram um *Data Design Pattern of Semantic View* (DDP_SV), para a organização lógica dos dados e metadados da visão semântica de um EKG. Esse padrão estrutura dados e metadados em quatro camadas hierárquicas que permite uma exploração contextual. No centro dessa estrutura está o grafo de metadados, codificado com o *Vocabulary of Semantic View* (VoSV)¹, que captura informações detalhadas e legíveis por máquina sobre estrutura, proveniência e qualidade da visão semântica.

Este artigo apresenta *EKGViewExplorer*, uma ferramenta gráfica interativa que permite explorar, de forma integrada, dados e metadados em EKGs construídos com o DDP_SV. Evolução da ferramenta *ContextEKG_Explorer*, *EKGViewExplorer* incorpora duas funcionalidades centrais: (i) exploração do grafo de metadados e (ii) visualização

*Link do vídeo de demonstração da ferramenta – <https://tinyurl.com/ysye2ay7>

¹Relatório técnico do vocabulário VoSV – <https://tinyurl.com/3cfxwnbc>

dos metadados das instâncias da visão semântica diretamente a partir do grafo de dados. Além disso, *EKGViewExplorer* é independente de domínio e filtra as informações conforme o idioma selecionado (@pt-BR ou @en).

As demais seções do artigo estão organizadas do seguinte modo. A Seção 2 apresenta a ferramenta de exploração de dados e metadados de uma visão semântica com um estudo de caso. A Seção 3 aborda trabalhos relacionados, fornecendo um breve panorama sobre as soluções encontradas. Por fim, a Seção 4 expõe as considerações finais, analisando resultados e direções futuras.

2. Ferramenta de Exploração da Visão Semântica

Esta seção apresenta *EKGViewExplorer*, uma ferramenta para apoiar a exploração de dados e metadados das visões semânticas em EKGs, construídas com o *DDP_SV*.

2.1. Arquitetura da *EKGViewExplorer*

A Figura 1 mostra a arquitetura cliente-servidor da *EKGViewExplorer* onde o usuário interage com uma interface gráfica no *front-end*, desenvolvida como uma aplicação web utilizando as bibliotecas *React*, *Material-IU* e *Axios*. O *back-end* é composto por uma API REST, implementada com o *framework FastAPI*. Além disso, a ferramenta conta com o *triplestore GraphDB*, responsável por armazenar os grafos da visão semântica. O código-fonte da ferramenta está disponível neste repositório: <https://github.com/renato-freitas/ekg-view-explorer>.

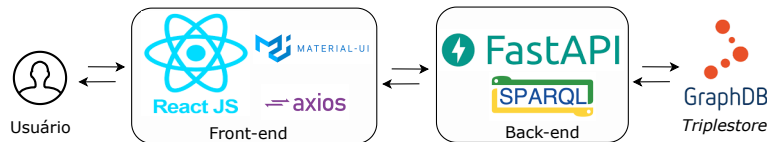


Figura 1. Arquitetura da *EKGViewExplorer*

2.2. Estudo de Caso: A Visão Semântica ‘SV_Music’

No *DDP_SV*, as quatro camadas das visões semânticas são: visões exportadas (mapeamento das fontes de dados para um vocabulário compartilhado); visões de ligação (estabelecimento de conexões semânticas entre entidades equivalentes); visões de unificação (integração de entidades semanticamente equivalentes em uma única representação) e visões de fusão (resolução de conflitos para produzir uma visão consistente).

A Figura 2 ilustra um fragmento representativo da visão semântica *SV_Music* (prefixo “*svm:*”), que serve como exemplo prático ao longo das seções seguintes. A *SV_Music* adota o padrão *DDP_SV* e resulta da integração de dados de música das fontes *Music-Brainz* e *DBpedia*. A Figura 2(a) mostra o grafo de metadados, anotado com o *VoSV*, exibindo as seguintes visões:

- *svm:EV_DBpedia* (visão RDF exportada pela *DBpedia*), *svm:EV_MusicBrainz* (visão RDF exportada pela *MusicBrainz*);
- *svm:LV_Music_Artist* (visão de ligação entre artistas das visões exportadas);
- *svm:UV_Music_Artist* (visão de unificação artistas equivalentes);
- *svm:FV_Music_Artist* (visão de fusão com resolução de conflitos em propriedades na visão de unificação).

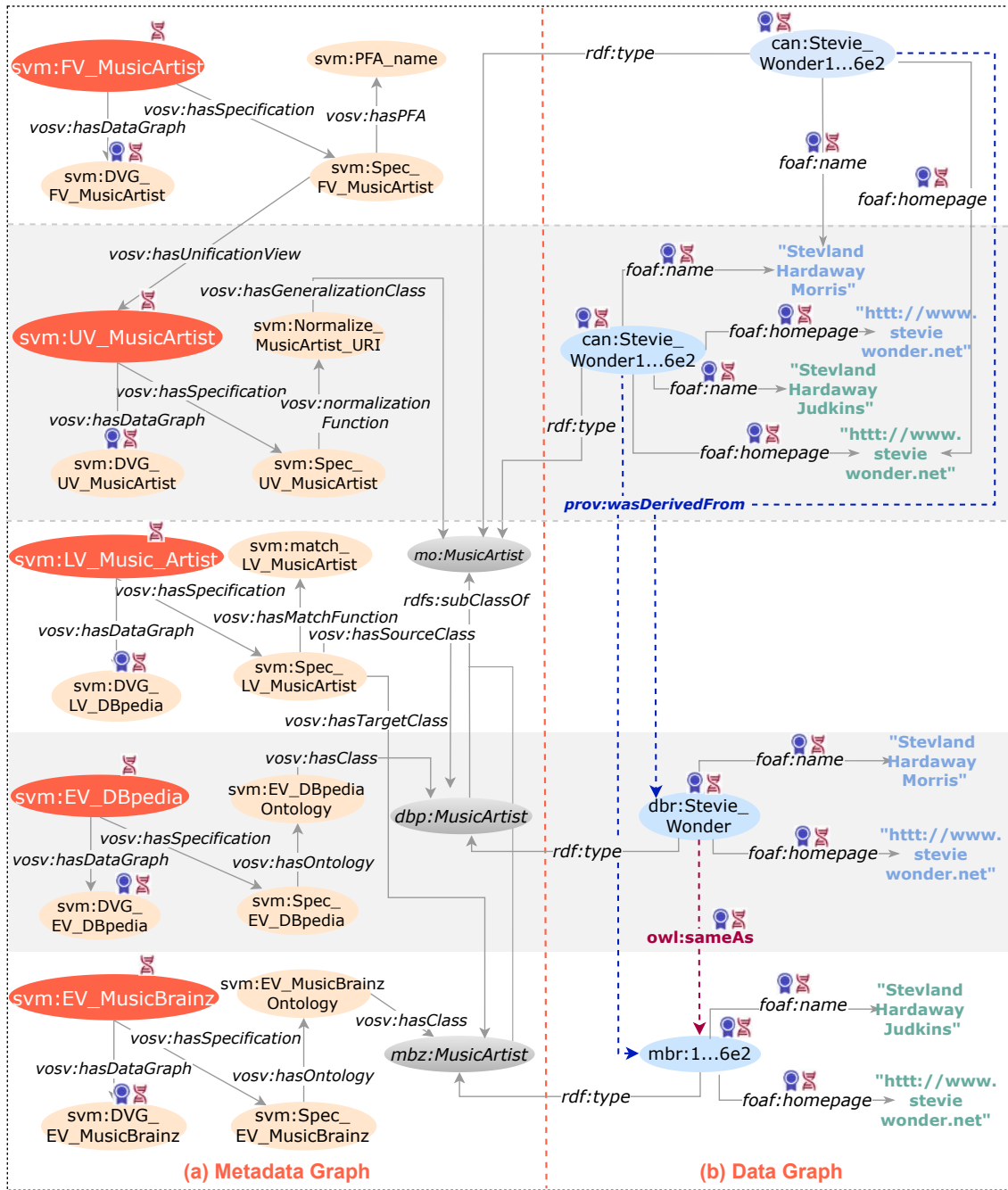


Figura 2. Fragmento da *SV_Music*

A Figura 2(b) apresenta o grafo de dados correspondente. Esta figura destaca a capacidade do *DDP_SV* de vincular explicitamente elementos de metadados aos dados que eles descrevem, permitindo exploração integrada e compreensão contextual.

O *DDP_SV* permite a representação estruturada e em conformidade com os padrões de metadados essenciais—como qualidade e proveniência dos dados—em vários níveis de granularidade. Na Figura 2, os metadados de qualidade e proveniência são indicados visualmente pelos ícones  (qualidade) e  (proveniência), respectivamente. Esses ícones aparecem ao lado de recursos específicos e triplas RDF dentro da visão semântica,

indicando a presença de metadados associados e facilitando a rápida identificação e exploração de indicadores-chave de confiabilidade.

2.3. Explorando Dados e Metadados da ‘SV_Music’ com *EKGViewExplorer*

A *EKGViewExplorer* permite que os usuários inspecionem recursos selecionados do grafo de dados em detalhes, incluindo seus atributos e metadados associados. Por exemplo, a Figura 3 exibe a tela com a visualização do recurso *dbr:Stevie_Wonder* dentro do grafo de dados mostrado na Figura 2. Os ícones exibidos ao lado do rótulo do recurso e das propriedades fornecem acesso direto aos metadados de proveniência e qualidade associados, permitindo que os usuários avaliem rapidamente a confiabilidade e a origem dos dados.

Os metadados de qualidade são recuperados consultando as medições de qualidade (*dqv:QualityMeasurement*) dos recursos e triplas, pré-computadas, armazenadas no grafo de metadados da *SV_Music*. As informações de qualidade, tanto para o recurso quanto para triplas individuais, são visualizadas em um gráfico de radar com os valores das dimensões de qualidade avaliadas. Esse tipo de visualização facilita uma análise da qualidade observada. Por exemplo, os indicadores de qualidade do gráfico de radar da Figura 3(b) nos dizem que a tripla (*dbr:Stevie_Wonder*, *foaf:name*, “Stevland Hardaway Judkins”) é consistente, com um alto grau de completude, mas os dados não são recentes.



Figura 3. Tela de exploração de recurso com metadados de proveniência e qualidade.

A exploração dos metadados de proveniência depende do contexto do recurso, se ele pertence a uma visualização exportada, visualização de unificação ou visualização de fusão, conforme discutido a seguir.

No contexto de uma visão exportada, a proveniência de um recurso é recuperada consultando os metadados da especificação da visão exportada correspondente. Os principais metadados recuperados incluem: (i) a fonte de dados da qual o recurso foi derivado e (ii) as regras de mapeamento que geraram o recurso. A proveniência de triplas individuais também é derivada dos metadados da especificação da visão exportada, particularmente, por meio das regras de mapeamento definidas para o predicado da tripla (Figura 3(a)).

Os metadados de proveniência de um recurso em uma visão de unificação incluem: a própria especificação da visão de unificação e os recursos provenientes de distintas visões exportadas que originaram o recurso unificado. A Figura 4 mostra a proveniência do recurso *dbr:Stevie_Wonder*, no contexto de uma visão de unificação. Nesta visualização, um usuário pode explorar cada recurso original separadamente.

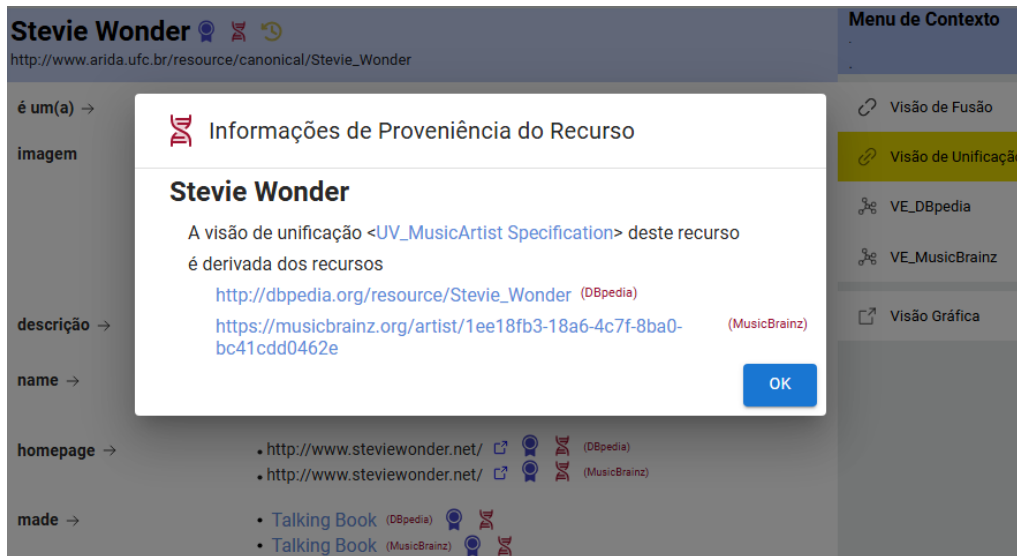


Figura 4. Tela de exploração de recurso com a proveniência no contexto de visão de unificação.

No contexto da visão de fusão, a proveniência do recurso canônico inclui a classe de generalização e os recursos utilizados na construção da fusão. A proveniência de uma tripla, neste contexto, inclui a assertiva de fusão de propriedade, possibilitando saber qual função ou método de resolução do conflito foi utilizado.

Por fim, a ferramenta também permite a exploração diretamente dos metadados dos componentes da visão semântica: *smv:EV_DBpedia*, *smv:EV_Musibrainz*, *smv:LV_Music_Artist*, *smv:UV_Music_Artist* e *smv:FV_Music_Artist*. Além disso, é possível explorar a ontologia da *SV_Music* e visualizar graficamente os recursos e suas relações por meio do item “Visão Gráfica”.

3. Trabalhos Relacionados

Diversos estudos têm proposto ferramentas para exploração de grafos de conhecimento (GCs) [Lissandrini et al. 2022]. Por exemplo, [Hirsch et al. 2009] apresentam as ferramentas *Thinkbase* e *Thinkpedia*, que utilizam visualização de GCs para facilitar o acesso e a exploração de grandes repositórios de conhecimento na Web. [Liebig et al. 2023] introduziram a ferramenta *SemSpect*, com uma abordagem que utiliza agregação visual para resolver o problema da visualização “bola-de-pelo” de GCs. Já [Avila and Vidal 2023] apresentaram *LiRB*, uma interface baseada nas tradicionais páginas da Web que permite uma exploração de GCs. *LiRB* atua sobre endpoints SPARQL para acessar os dados.

Alguns bancos de dados de grafos como o *GraphDB*² e o *Neo4J*³, oferecem fun-

²<https://www.ontotext.com/products/graphdb/>

³<https://neo4j.com/>

cionalidades de visualização de grafos baseadas em representações do tipo nó-aresta. Contudo, tais abordagens não atendem plenamente às demandas dos atuais casos de uso relacionados à exploração de GCs. Ademais, essas ferramentas apresentam limitações na exibição de informações de proveniência e outros metadados associados aos recursos. Nesse sentido, *EKGViewExplorer* emerge como uma abordagem inovadora para a exploração e visualização de dados e metadados dos GCs da visão semântica, visando facilitar a compreensão do conteúdo e obter *insights* relevantes.

4. Conclusão

Este artigo apresentou *EKGViewExplorer*, uma ferramenta gráfica interativa, projetada para explorar e visualizar dados e metadados da visão semântica de EKGs construídas sobre o padrão de *design* de dados *DDP_SV*. Demonstramos a ferramenta explorando recursos no grafo de dados de uma visão semântica e visualizando metadados de proveniência e qualidade associados a esses recursos.

Nossa expectativa é transformar esta ferramenta em uma plataforma do tipo *Software as a Service* (SaaS), tornando-a um produto comercial. Para além da exploração, ela auxiliará na construção e manutenção de EKGs e *Data Mashups* para aplicações inteligentes. Como trabalhos futuros, planejamos incluir uma interface para perguntas e respostas baseada em *Large Language Model* (LLM), onde a visão semântica do EKG servirá como uma fonte confiável de informação.

Por fim, ressaltamos que esse trabalho foi apoiado pela Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap), órgão vinculado a Secretaria de Ciência, Tecnologia e Educação Superior do Ceará (Secitece) - Código de Financiamento #BMD-0008-00739.01.10/24

Referências

- Avila, C. and Vidal, V. (2023). Lirb: Um navegador leve baseado em texto para knowledge graphs rdf. In *Anais Estendidos do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 102–107, Porto Alegre, RS, Brasil. SBC.
- Galkin, M., Auer, S., Vidal, M.-E., and Scerri, S. (2017). Enterprise knowledge graphs: A semantic approach for knowledge management in the next generation of enterprise information systems. In *International Conference on Enterprise Information Systems*, volume 2, pages 88–98. SCITEPRESS.
- Hirsch, C., Hosking, J., and Grundy, J. (2009). Interactive visualization tools for exploring the semantic graph of large knowledge spaces.
- Liebig, T., Opitz, M., Vialard, V., and Wenzel, M. (2023). Scalable no-code knowledge graph exploration and querying with semspect. In *SEMANTiCS (Posters & Demos)*.
- Lissandrini, M., Mottin, D., Hose, K., and Pedersen, T. B. (2022). Knowledge graph exploration systems: Are we lost? In *Annual Conference on Innovative Data Systems Research*.
- Vidal, V., Freitas, R., Arruda, N., Casanova, M. A., and Renso, C. (2024). A data design pattern for building and exploring semantic views of enterprise knowledge graphs. In *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 1–13, Porto Alegre, RS, Brasil. SBC.