

# ***ContextEKG Explorer: Uma Ferramenta Interativa para Exploração Contextual da Visão Semântica em Sistemas de Grafo de Conhecimento Empresarial***

**Renato Freitas<sup>1</sup>, Vânia Ponte Vidal<sup>1</sup>**

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

renato.freitas@alu.ufc.br, vvidal@lia.ufc.br

**Resumo.** *Um sistema de Grafo de Conhecimento Empresarial (Enterprise Knowledge Graph ou EKG) é um paradigma baseado em tecnologias da Web Semântica e Grafos de Conhecimento para integrar fontes de dados heterogêneas. Um EKG fornece uma visão ontológica e unificada, para que as aplicações tenham acesso integrado aos dados através da visão semântica. A exploração de dados na visão semântica requer ferramentas que apresentem grafos de forma compreensível, facilitando a interpretação e a tomada de decisão. Este artigo apresenta a ContextEKG Explorer<sup>1</sup>, uma ferramenta gráfica interativa para a exploração de dados na visão semântica de um EKG. A visão semântica é organizada em uma hierarquia de três níveis, simplificando a exploração das entidades em múltiplos contextos. A ferramenta oferece visualizações intuitivas das entidades e navegação fluida entre esses contextos.*

## **1. Introdução**

Um sistema de *Enterprise Knowledge Graph* é um paradigma que estabelece uma base robusta de conhecimento, integração de dados em larga escala e análises avançadas para as organizações [Ehrlinger and Wöß 2016].

A arquitetura de sistemas de EKG, proposta por [Galkin et al. 2017], consiste em três camadas distintas: (i) a Camada de Dados, onde um *Data Lake* funciona como um sistema de armazenamento centralizado, projetado para acomodar dados brutos de vários tipos, incluindo dados não-estruturados e (semi)estruturados; (ii) a Camada Semântica é responsável por integrar semanticamente dados das diversas fontes no *Data Lake*, visando construir uma apresentação unificada e coerente, conhecida como “visão semântica”. Nessa camada, o acesso integrado aos dados é realizado através de um *endpoint* SPARQL, usando o vocabulário da ontologia da visão semântica. (iii) a Camada de Aplicação implementa aplicações ou serviços específicos que utilizam o grafo da visão semântica e seu modelo de dados unificado para extrair *insights* e facilitar a tomada de decisões.

Este trabalho apresenta *ContextEKG Explorer*, uma ferramenta para a exploração contextual da visão semântica de um EKG. Suas principais características incluem: (i) a exploração de recursos da visão semântica em diversos contextos, permitindo rastrear a linhagem de dados (*data lineage*); (ii) a aplicabilidade em qualquer domínio do conhecimento; (iii) a realização de consultas SPARQL de forma otimizada, adotando paginação

<sup>1</sup>URL da demonstração da ferramenta: <https://youtu.be/czEKmBoBPzI>

de dados e grafos nomeados, evitando examinar toda a visão semântica desnecessariamente, ideais para explorar grandes grafos de conhecimento; e (iv) a alternância de idioma entre português e inglês, exibindo os rótulos da ferramenta e os resultados das consultas SPARQL conforme o idioma selecionado.

O restante do artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 detalha a ferramenta e mostra um exemplo de exploração de uma visão semântica, enquanto a Seção 4 expõe as considerações finais.

## 2. Trabalhos Relacionados

Aplicações de visualização como as apresentadas em [Bikakis and Sellis 2016] podem facilitar a compreensão e interpretação dos dados no grafo da visão semântica. No entanto, o estudo de [Li et al. 2023] aponta desafios quando essas aplicações exploram grafos grandes, com inúmeras triplas e relacionamentos complexos. Um exemplo disso é quando a visualização se torna confusa, dificultando a descoberta de padrões.

Estudos sobre a exploração de grafos de conhecimento (GC) incluem o trabalho de [Haase et al. 2019] que introduziu a *metaphactory*, uma plataforma para construção de aplicações de GC. Sua interface de usuário é baseada em componentes HTML, renderizados a partir da URI dos recursos no GC. O estudo de [de Souza et al. 2022] apresentou a ferramenta *TKGEvolViewer* para visualizar GCs ao selecionar “questões-guia” pré-definidas. [Sellami and Zarour 2022] introduziram o *KeyFSI*, uma interface de navegação facetada, que visa facilitar a exploração de dados em um GC. Por fim, [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. Embora promissoras, essas ferramentas não permitem uma exploração de recursos em 3 níveis de visões distintas nem exibem a proveniência dos recursos e propriedades na visão semântica, como a *ContextEKG Explorer* possibilita.

## 3. ContextEKG Explorer

### 3.1. Arquitetura da Visão Semântica

A ferramenta *ContextEKG Explorer* foi desenvolvida sobre uma arquitetura que estrutura, logicamente, os dados e metadados da visão semântica. O grafo da visão semântica compreende dois subgrafos distintos: o GC de Dados e o GC de Metadados, conforme

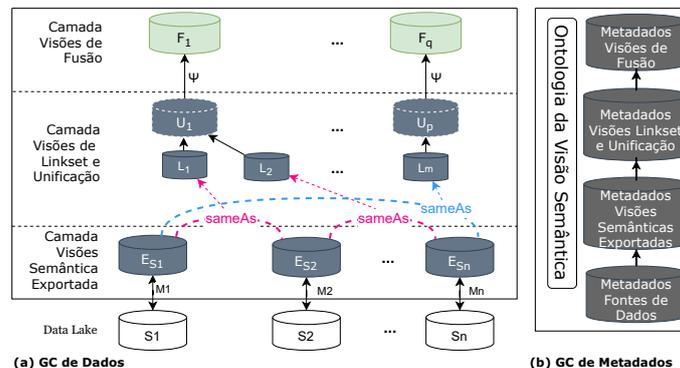


Figura 1. Arquitetura da Visão Semântica.

ilustrado na Figura 1. O GC de dados é formalmente estruturado por meio de três níveis hierárquicos de visões, enquanto o GC de Metadados fornece detalhes cruciais sobre as especificações das visões em vários níveis no GC de Dados.

Cada fonte de dados no *Data Lake* exporta uma visão RDF (**visão semântica exportada**) definida por mapeamentos entre a fonte de dados e a ontologia da visão semântica, que resolve a heterogeneidade de tipo de dados e de esquema. Para resolver a questão de “resolução de identidade”, as **visões de ligação** definem *links* “*sameAs*” entre instâncias em diferentes visões semânticas exportadas. Esses *links* desempenham um papel crucial na definição de **visões de unificação** que ajudam a formar uma representação unificada de entidades ao reunir atributos e relacionamentos de múltiplas visões semânticas exportadas ligadas pela relação *owl:sameAs* [Vidal et al. 2021]. Isso facilita a visualização e consulta de recursos em um contexto unificado, permitindo uma compreensão e análise mais abrangentes dos dados interconectados. O objetivo de uma **visão de fusão** é resolver conflitos quando diferentes fontes têm informações divergentes sobre a mesma entidade ou relacionamento, visando melhorar a qualidade, precisão e confiabilidade das informações na visão semântica. A partir da visão de unificação, é possível detectar informações conflitantes sobre a mesma entidade ou relacionamento. A solução para esse tipo de inconsistência requer a implementação de mecanismos de resolução de conflitos para identificar e resolver discrepâncias.

### 3.2. Explorando a Visão Semântica com *ContextEKG Explorer*

Em um EKG, a visualização de recursos é geralmente centrada na entidade onde o foco se dá, por exemplo, em indivíduos do tipo “Artista Musical” e “Álbum”. Com a *ContextEKG Explorer* um recurso pode ser explorado em três contextos diferentes: o contexto da visão semântica exportada; da visão de unificação e da visão de fusão. Para demonstrar a exploração nesses contextos, criamos um EKG com uma visão semântica sobre músicas, álbuns e artistas musicais brasileiros, denominada *SV\_MUSIC\_BR*, que integra três fontes de dados heterogêneas: *MusicBrainz*<sup>2</sup>(relacional), *Spotify*<sup>3</sup>(JSON) e *Wikidata*<sup>4</sup>(RDF).



(a) Tela de Seleção de Classe exibindo a VSE MusicBrainz e suas classes. (b) Tela de Exploração de Recurso na VSE MusicBrainz.

**Figura 2. Tela de Seleção de Classe e Tela de Exploração de Recurso.**

<sup>2</sup><https://musicbrainz.org/>

<sup>3</sup><https://developer.spotify.com/documentation/web-api>

<sup>4</sup><https://query.wikidata.org/>

### 3.2.1. Visualização no Contexto da Visão Semântica Exportada

Considerando a *SV\_MUSIC\_BR*, a Figura 2(a) exibe a tela de seleção de classe com a visão semântica exportada (VSE) “MusicBrainz” escolhida e suas classes. Ao selecionar, por exemplo, o artista “Gilberto Gil” da classe “Músico | Artista MusicBrainz”, o usuário é direcionado para a tela de exploração de recurso no contexto da visão semântica exportada “MusicBrainz”.

No painel esquerdo da tela de exploração de recurso na Figura 2(b) é exibido o atual conjunto de propriedades do recurso “Gilberto Gil” recuperadas do grafo da VSE “MusicBrainz”. O *rdfs:label* do recurso é evidenciado na parte superior do painel e os *rdfs:type* são destacados em azul. A *ContextEKG-Explorer* também oferece a funcionalidade de navegar para páginas externas, tais como as conectadas pela propriedade “site”. No painel direito o **menu de contexto** destaca em amarelo o contexto atual e lista as opções de contexto “Visão de Fusão”, “Visão de Unificação”, “VSE MusicBrainz”, “VSE Spotify” e “VSE Wikidata”, nos quais “Gilberto Gil” pode ser explorado.

### 3.2.2. Visualização no Contexto da Visão de Unificação

A Figura 3(a) mostra a tela de exploração de recurso exibindo o recurso “Gilberto Gil” no contexto da visão de unificação cujas propriedades (*é um(a)*, *imagem*, *descrição*, *nome*, *site*, *gravou | fez*, *data de nascimento*) são recuperadas das visões semânticas exportadas *MusicBrainz*, *Wikidata* e *Spotify*.

Uma visão de unificação é um grafo RDF virtual, computado em tempo de exploração. Durante a computação, a ferramenta agrega todas as propriedades do recurso selecionado com as propriedades dos recursos que têm *link owl:sameAs* com ele, e também com as propriedades dos recursos inferidos, detectando discrepâncias entre os valores das propriedades. Além disso, a ferramenta captura a proveniência de cada valor. A discrepância e a proveniência são exibidas ao lado dos valores correspondentes, como mostra a Figura 3(a). Para as propriedades de relacionamento, como “gravou | fez”, a ferramenta permite a navegação interativa através da URI do objeto. Por exemplo, o usuário pode navegar para o álbum “Louvação” gravado pelo “Gilberto Gil”.



(a) Propriedades do Artista “Gilberto Gil” no Contexto da Visão de Unificação. (b) Propriedades do Artista “Gilberto Gil” no Contexto da Visão de Fusão.

**Figura 3. Tela de Exploração de Recurso em Diversos Contextos.**

### 3.2.3. Visualização no Contexto da Visão de Fusão

A partir da tela de exploração de recurso, um usuário pode visualizar um recurso no contexto da visão de fusão ao selecionar a opção “Visão de Fusão” no menu de contexto.

Uma visão de fusão é um grafo virtual gerado em tempo de exploração. Durante a fusão, conflitos de valores de propriedades detectados no processo de agregação das propriedades, explicado na Subseção 3.2.2, são resolvidos com o auxílio de assertivas de fusão de propriedade (AFP). Considerando o recurso “Gilberto Gil”, a discrepância nos valores da propriedade “nome” nas fontes *MusicBrainz* e *Spotify* é resolvida por uma *função de resolução* definida na AFP. A Figura 3(b) mostra, como resultado da *função de resolução*, o único valor para “nome” que vem da fonte de dados com melhor qualidade.

A *ContextEKG Explorer* permite, ainda, visualizar o recurso graficamente selecionando a opção “Visão Gráfica”. Também é possível visualizar o histórico de eventos dos recursos na tela de *timeline* e consultar questões de competência pré-salvas.

### 3.2.4. Visualização das *Timelines* dos Recursos

A opção “Timeline” é disponibilizada na tela de exploração de recurso quando o objeto de *timeline* do recurso selecionado apresenta algum *tl:Instant* no grafo nomeado de *timeline*. Isso significa que o recurso sofreu algum evento em um determinado momento.

A Figura 4 exibe um fragmento da *timeline* do artista “Gilberto Gil” na visão de unificação, mostrando eventos ocorridos tanto na *MusicBrainz* quanto na *Spotify*. Os instantes são ordenados por data/hora, em cores diferentes para cada fonte, listando as propriedades com os valores antigos e os atualizados. Os instantes podem ser filtrados por propriedade. Essa funcionalidade da ferramenta pode auxiliar nas investigações.

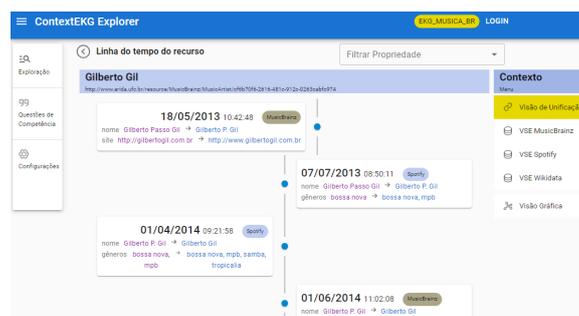


Figura 4. Fragmento da *Timeline* do Artista “Gilberto Gil” no Contexto Unificado.

## 4. Considerações Finais

Este artigo apresentou a *ContextEKG Explorer*, uma ferramenta para explorar entidades na visão semântica em sistema de EKG em vários contextos e perspectivas. Foram abordadas as visualizações no contexto da visão semântica exportada, da visão de unificação e da visão de fusão; além da visualização de *timeline* dos recursos.

A *ContextEKG Explorer* é composta por uma interface gráfica e uma API REST. A interface gráfica é uma aplicação web construída com a biblioteca *React*<sup>5</sup> e a API REST,

<sup>5</sup><https://react.dev/>

criada com o *framework FastAPI*<sup>6</sup>, gerencia as requisições da interface, além manter uma conexão com um banco de triplas RDF (*triplestore*) que armazena os grafos e processa as consultas SPARQL. A *ContextEKG\_Explorer* utiliza a versão gratuita do *GraphDB*<sup>7</sup> como seu *triplestore*. O código-fonte da *ContextEKG\_Explorer* é encontrado em [https://github.com/renato-freitas/ContextEKG\\_Explorer](https://github.com/renato-freitas/ContextEKG_Explorer).

Para trabalhos futuros, pretendemos expandir esta ferramenta e transformá-la em uma plataforma do tipo *Software as a Service* (SaaS) para, além da exploração, auxiliar na construção e manutenção de EKGs e *Data Mashups* para aplicações inteligentes.

Este trabalho foi realizado com apoio da Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (Funcap) - 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.
- Bikakis, N. and Sellis, T. (2016). Exploration and visualization in the web of big linked data: A survey of the state of the art. *6th International Workshop on Linked Web Data Management (LWDM 2016)*.
- de Souza, E. M. F., Rossanez, A., dos Reis, J. C., and da Silva Torres, R. (2022). Visualização interativa da evolução de grafos de conhecimento. In *Anais do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 343–354. SBC.
- Ehrlinger, L. and Wöß, W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48(1-4):2.
- 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.
- Haase, P., Herzig, D. M., Kozlov, A., Nikolov, A., and Trame, J. (2019). metaphactory: A platform for knowledge graph management. *Semantic Web*, 10(6):1109–1125.
- Li, H., Appleby, G., Brumar, C. D., Chang, R., and Suh, A. (2023). Knowledge graphs in practice: characterizing their users, challenges, and visualization opportunities. *IEEE Transactions on Visualization and Computer Graphics*.
- Sellami, S. and Zarour, N. E. (2022). Keyword-based faceted search interface for knowledge graph construction and exploration. *International Journal of Web Information Systems*, 18(5/6):453–486.
- Vidal, T., Viktor, C., Avila, S., Mariano, R., Calixto, T., Ivo, P., Filho, J., Brayner, A., and Vidal, M. (2021). Uso das tecnologias da web semântica na construção de grafos de conhecimento semântico baseado no enfoque híbrido. In *Proceedings of the XIV Seminar on Ontology Research in Brazil (ONTOBRAS 2021)*.

<sup>6</sup><https://fastapi.tiangolo.com/>

<sup>7</sup><https://graphdb.ontotext.com/>