

KG-E: Um Grafo de Conhecimento Semântico Baseado na Integração de Dados de Empresas e Sancionados

Tulio Vidal Rolim¹, Caio Viktor S. Avila¹, Narciso Moura A. Junior¹,
Francisca Jamires Costa¹, Roberval Gomes Mariano²,
Tainan Calixto², Vânia Maria Ponte Vidal¹

¹Universidade Federal do Ceará (UFC)
Fortaleza, CE, Brasil.

²Secretaria da Fazenda do Maranhão
São Luís, MA, Brasil

{tulio.xcrtf, arlaass, narcisoarruda,
jamirescostaa, vania.pvidal}@gmail.com
{mariano, tainan.calixto}@sefaz.ma.gov.br

Abstract. *The integration of data allows the discovery of information that a priori was not possible from different and isolated sources. For that, the use of Semantic Knowledge Graphs provides a homogeneous view from heterogeneous sources integrated semantically. This work describes the process of construction and validation of KG-E, a Semantic Knowledge Graph containing data from companies registered in the National Network for the Simplification of the Registration and Legalization of Companies and Businesses (REDESIM) and Sanctioned from the Register of Unincorporated Companies and Suspended (CEIS) semantically integrated. KGE-E was validated through a case study based on questions of competence, showing satisfiability regarding issues related to the domain. Finally, the results suggest that KG-E proved to be an integrated and semantically enriched view of the REDESIM and CEIS bases, facilitating the construction of future analysis applications on the integrated data, thus allowing new insights in the fiscal-business scope.*

Resumo. *A integração de dados permite a descoberta de informações que à priori não eram possíveis em fontes distintas e isoladas. Para tanto, a utilização de Grafos de Conhecimento Semântico (Semantic Knowledge Graphs) fornece uma visão homogênea a partir de fontes heterogêneas integradas semanticamente. Este trabalho descreve o processo de construção e validação de KG-E, um Grafo de Conhecimento Semântico contendo dados de empresas cadastradas na Rede Nacional para a Simplificação do Registro e da Legalização de Empresas e Negócios (REDESIM) e de Sancionados a partir do Cadastro de Empresas Inidôneas e Suspensas (CEIS). KGE-E foi validado através de um estudo de caso com base em questões de competência, apresentando satisfatibilidade quanto às questões relativas ao domínio. Por fim, os resultados sugerem que KG-E demonstrou ser uma visão integrada e enriquecida semanticamente das bases do REDESIM e CEIS, facilitando a construção de aplicações para análises futuras sobre os dados integrados, permitindo desta forma novos insights no âmbito fiscal-empresarial.*

1. Introdução

No Brasil, o setor fiscal tem cada vez mais responsabilidades em identificar e fiscalizar irregularidades no âmbito da administração pública, tendo em vista a demanda em correlacionar, integrar e identificar ações envolvendo empresas.

Nessa conjuntura, a Rede Nacional para Simplificação do Registro e Legalização de Empresas e Negócios (REDESIM) foi criada pelas Leis 123/2006 e 11.598/2007 com foco em facilitar o processo de criação de empresas, tornando-o livre de entraves burocráticos, com a adoção de uma tributação sintetizada e garantindo à proteção e apoio do Estado [BRASIL 2021].

A REDESIM dispõe de dados de Empresas e suas Atividades Econômicas Desempenhadas, Naturezas Legais, Estabelecimentos categorizados em Matriz e suas Filiais com dados de Endereço, Dados de Capital Social, Porte e Quadro de Sócios. A REDESIM conta também com informações de eventos realizados por uma empresa, tais como alterações no capital social e / ou quadro societário, bem como eventos e.g., cisão, fusão e incorporação.

Eventualmente, empresas ou pessoas físicas criadas, podem vir a tornar-se sancionadas, e seu administrador vir a realizar aberturas de novas empresas ou a própria empresa sancionada pode chegar a compor o quadro societário de outras e continuar realizando ações não saudáveis na esfera pública, tendo em vista que tais dados não estão presentes na REDESIM.

Para tanto, o Cadastro de Empresas Inidôneas e Suspensas (CEIS) fornece dados de empresas e pessoas físicas que sofreram sanções pela Administração Pública. Contudo, apesar de conterem dados sobre empresas, REDESIM e CEIS são fontes de dados heterogêneas que possuem vocabulário e representação distintos, dificultando assim uma fácil integração dessas fontes.

Como proposta, um grafo de conhecimento semântico ou (*Semantic Knowledge Graphs*) pode ser utilizado para realizar uma integração semântica entre as fontes de dados do REDESIM e CEIS de modo a propor um processo que faz uso de uma representação conceitual dos dados e seus relacionamentos para eliminar possíveis heterogeneidades.

Grafos de Conhecimento Semânticos vem sendo utilizados como um mecanismo para consolidar e integrar semanticamente um grande número de fontes de dados heterogêneas em um espaço de dados abrangente. [Soylu et al. 2018]

Neste trabalho é proposto KG-E, um grafo de conhecimento enriquecido semânticamente com dados da REDESIM e CEIS, com ênfase em fornecer informações a partir de dados de empresas cadastradas e relações com pessoas jurídicas e físicas sancionadas na esfera pública de forma integrada.

O restante do artigo está organizado como se segue. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve o processo de construção do KG. A Seção 4 expõe o estudo de caso utilizado para validação do KG construído. Já na Seção 5 são expostos os resultados. E por fim, a Seção 6 apresenta as considerações finais do trabalho.

2. Trabalhos Relacionados

[Tosin 2015] desenvolve uma solução para interligar bases de dados governamentais isoladas, disponibilizando os dados em formato aberto e de forma interligada com base em ontologias de domínio. O autor também aborda o Sistema de Informações Administrativas do Serviço Geral (SIASG) como uma proposta de integração com outras bases.

Tabela 1. Trabalhos Relacionados.

	Integração Semântica	Ontologia	REDESIM	CEIS
[Tosin 2015]	X	X	-	-
[Marques 2013]	-	-	-	X
[da Cruz et al. 2019]	X	X	-	-
[Bernardi et al. 2017]	-	X	-	-
[da Fonseca et al. 2014]	X	X	-	-
[Fonseca et al. 2016]	X	X	-	-
[Rolim et al. 2019]	X	X	-	X
Este estudo	X	X	X	X

Fonte: Elaborado pelos autores, 2021.

Em seu trabalho [Marques 2013] realiza um *case* de integração com uma aplicação utilizando o Portal Brasileiro de Dados Abertos e a API de Dados Abertos do TCM-CE, tendo como base a Lista de Empresas Inidôneas em conjunto com a lista de Notas de Pagamentos efetuados pelos órgãos do Estado do Ceará.

[Bernardi et al. 2017] propõe uma ontologia de domínio sobre licitações como base de conhecimento primário, visando facilitar a elicitação de requisitos para novos portais de transparência municipal.

O trabalho desenvolvido por [da Fonseca et al. 2014] relata uma abordagem de integração de dados governamentais sobre organizações e estruturas organizacionais em conformidade com a ontologia (*ORG Ontology*).

[Fonseca et al. 2016] utilizam uma ontologia de referência para abranger os conceitos do processo de gerenciamento e autorização orçamentária do Sistema Integrado de Orçamento e Planejamento (SIOP), bem como aos relacionados ao processo da despesa pública do SIAFI. Assim, os autores sugerem melhorias à compreensão dos dados governamentais publicados viabilizando seu uso e sua integração ao nível semântico.

Em [Rolim et al. 2020] e [da Cruz et al. 2019] os autores fazem a integração semântica das fontes de dados SIM e SINASC, extraídas do Sistema Único de Saúde (SUS). Os autores criam um grafo de conhecimento KG SUS dessas duas fontes de dados. Já em [Rolim et al. 2019] o trabalho apresenta uma proposta de portal semântico baseado em ontologias para as fontes de gastos públicos, SIASG, CEIS e CNEP.

Apesar de tratarem problemáticas relacionadas, os trabalhos [Tosin 2015], [Marques 2013], não realizam uma integração semântica, dificultando o entendimento dos dados e também a adição de novas fontes de dados. Em [da Cruz et al. 2019] e [Rolim et al. 2019] os autores realizam uma integração semântica, no entanto, essas integrações não fazem uso dados públicos de Empresas do REDESIM.

A Tabela 1 expõe os aspectos diferenciais apresentados aqui neste trabalho frente aos demais relacionados.

3. Processo de Construção

KGs Semânticos devem possuir um grafo semântico associado, e neste trabalho KG-E adota a utilização de um grafo semântico representado em uma Camada Semântica.

A Camada Semântica de KG-E é compreendida formalmente como uma tripla $\lambda = (O_D, E, E_L)$.

- O_D representa a ontologia de domínio. O_D é responsável por estabelecer um vocabulário comum a ser compartilhado a fim de descrever os dados integrados advindos das fontes de dados;
- E : É um conjunto de Visões Exportadas definida sobre uma fonte de dados S , utilizando uma ontologia exportada O_E e mapeamentos M_E , formando uma Especificação da Visão Exportada E_S ;
- E_L é uma visão de *linkset*, as quais especificam como identificar objetos em diferentes fontes de dados que representam o mesmo objeto no mundo real. Cada E_L possui uma especificação da visão de *linksets* E_{LS} .

O processo de construção do KG-E foi realizado com base na construção da camada semântica e integração de dados baseados em *linked data* propondo fornecer dados relevantes para o domínio fiscal e de empresas.

3.1. Construção da Camada Semântica

3.1.1. Descrição das Bases de Dados

As Bases de Dados utilizadas foram a REDESIM que é uma fonte de dados em formato relacional e o CEIS que se encontra em formato original CSV, ambas bases com *schemas* e proveniências distintas.

O REDESIM contém dados de empresas, estabelecimentos e quadro societário, já o CEIS possui dados referentes às sanções aplicadas em pessoas físicas ou jurídicas que participaram de alguma modalidade de compra pública com a administração pública.

A integração das bases de dados foi feita utilizando uma abordagem *pay-as-you-go* [Madhavan et al. 2007]. Dessa forma, a integração das bases de dados pode ser realizada de forma contínua e incremental. Isso permite que a integração já especificada possa ser utilizada, mesmo que ainda esteja sendo construída e modificada.

A adoção deste tipo de abordagem faz com que a adição de novas fontes de dados seja bastante simples e possa ser feita sem que seja necessário alterar o resultado da integração semântica já realizada, logo:

1. Caso a ontologia de domínio não contenha os conceitos e relacionamentos presentes na fonte de dados a ser adicionada, a ontologia deve ser modificada com a finalidade de incluir esses conceitos;
2. Mapeamentos e regras de links semânticos devem ser especificados e adicionados à camada de integração semântica. Nenhuma das especificações já existentes, precisa ser alterada.

3.1.2. Extração e anotação de dados:

Os dados da REDESIM foram obtidos especificamente a partir de convênio com a Secretaria de Fazenda do Maranhão (SEFAZ-MA). Já os dados do CEIS foram obtidos no Portal da Transparência¹. Ambas as fontes de dados foram tratadas em formato relacional, tendo em vista um maior conjunto de ferramentas e recursos consolidados disponíveis.

A frequência de atualização dos dados da REDESIM é feita a cada 30 minutos, e a do CEIS é feita a cada quatro horas, aspecto que demanda um processo de extração contínuo e de difícil manutenção.

Para cada fonte, foram fornecidos um conjunto de metadados baseados no uso do RDFS² e Dublin Core³ para descrever informações que vão desde o formato da fonte até comentários sobre os significados de elementos do esquema utilizado pela fonte.

3.1.3. Construção da Ontologia de Domínio

A ontologia atua como uma camada de abstração. Ela provê uma visão integrada das fontes de dados da REDESIM e CEIS. Desta forma, usuários e aplicações podem ter acesso a uma abstração semântica sobre os dados que lhe façam sentido, dispensando detalhes técnicos subjacentes às fontes originais de forma transparente. Além disso, a ontologia de domínio também é responsável pela adição de semântica aos dados.

A ontologia de domínio O_{DE} consiste na união das ontologias exportadas $O_{REDESIM}$ e O_{CEIS} , abrangendo todos os termos e conceitos presentes em ambas as fontes.

3.1.4. Especificação das Visões Exportadas

Cada Visão Exportada E possui uma Especificação de Visão Exportada E_S contendo uma fonte de dados, ontologia exportada e mapeamentos. Uma E_S é uma tripla $\{S, O_E, M_E\}$, onde:

- S : é uma fonte de dados;
- O_E : é a ontologia exportada, sendo um recorte de O_D , o qual contém os termos de O_D que descrevem o *schema* de uma fonte S ;
- M_E : é um conjunto de mapeamentos que relacionam termos do vocabulário de O_E com termos do *schema* de uma fonte S .

A Figura 1 apresenta uma visão da ontologia de domínio O_{DE} composta pelas ontologias exportadas O_{CEIS} e $O_{REDESIM}$ contendo os termos e conceitos das fontes do REDESIM e CEIS.

¹<http://www.portaltransparencia.gov.br/download-de-dados/ceis>

²<https://www.w3.org/TR/rdf-schema/>

³<https://www.dublincore.org/resources/glossary/ontology/>

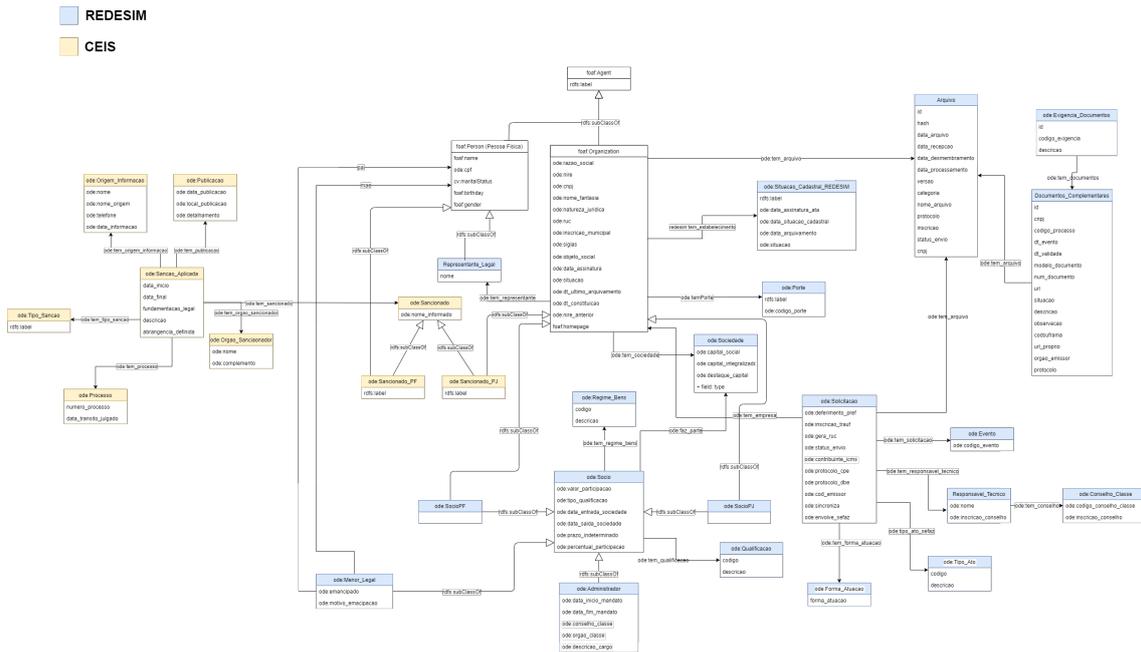


Figura 1. Recorte da Ontologia de Domínio O_{DE} .

Por conseguinte, as visões exportadas foram especificadas através de mapeamentos entre O_{DE} e as bases do REDESIM e CEIS, no qual os dados estavam armazenados. Os mapeamentos foram especificados utilizando a linguagem padrão R2RML [W3C 2012].

3.1.5. Especificação das Visões de Linksets

A compreensão da Especificação da Visão de *Linksets* do *KG*, E_{LS} descrita neste trabalho é vista através de uma adaptação de [Vidal et al. 2016] como uma quintupla = $\{E_{LS} = P, O_E, F, G, \mu\}$, onde:

- P : é uma *link type*, este conceito é um predicado *owl:sameAs*;
- O_E : é o vocabulário de correspondência de E_L . Uma O_E está contida em uma E_S de uma E ;
- F e G são definições de visões exportadas. Então, O_E é a ontologia comum para as visões exportadas F e G ;
- μ é uma relação $2n$, uma (*Linkage Rule*) de E_{LS} é definida por uma tripla = $\{M_C, M_P, L_F\}$, onde:
 - M_C : é a *Match Class*, classe utilizada para estabelecer a relação de link *owl:sameAs* entre duas instâncias;
 - M_P : é a *Match Property*, propriedade de uma O_D utilizada para calcular uma relação de link entre duas instâncias;
 - L_F : é uma *Link Function*, função de link, podendo ser uma Função de Agregação (*Aggregation Function*), Comparação (*Comparator Function*) ou Transformação (*Transformation Function*);

As visões de *linksets*, foram construídas utilizando o *Silk Link Discovery Framework* [Volz et al. 2009] de modo a identificar os links *owl:sameAs* entre as bases. Através do *Silk* são definidas heurísticas para identificação de dois registros similares. Na visão desenvolvida neste trabalho, o intuito foi identificar se duas empresas, ou sócios (pessoas físicas), representam um mesmo indivíduo no mundo real.

Em relação às empresas, para realizar o *match*, foi utilizado apenas o CNPJ raiz (*root_cnpj*), já que todas as bases envolvidas publicam o CNPJ dos indivíduos que são empresas. No entanto, para os sócios pessoas físicas, foi necessário utilizar o (*cpf*) a ser comparado.

3.2. Integração dos dados

Neste passo, é feito o processo de resolução da heterogeneidade dos dados, incluindo a heterogeneidade semântica. Aqui foi feito o aproveitamento do resultado da camada semântica para integração dos dados.

O resultado da camada semântica é representado por um conjunto de especificações. Cada uma dessas especificações está associada a um passo necessário para a construção da integração dos dados contidos no KG.

Após especificadas, as visões exportadas, os mapeamentos de cada visão foram utilizados, onde os mapeamentos da REDESIM foram utilizados para realizar a triplificação dos dados REDESIM e fornecer acesso através de um grafo RDF disponibilizado em um Endpoint SPARQL. Já os mapeamentos do CEIS, foram utilizados para construção de um Endpoint Virtual. Em ambos foi utilizado o Ontop [Calvanese et al. 2017].

Para estabelecer as ligações de links *owl:sameAs*, o *Silk* foi utilizado para calcular as ligações entre instâncias de empresas (pessoas jurídicas) e sócios (pessoas físicas e jurídicas) idênticas.

3.2.1. Acesso aos Dados através da Camada Semântica

A Camada Semântica provê um único ponto de acesso aos dados, e permite que consultas sejam formuladas com base nos termos da ontologia de domínio, de forma que o usuário não precisa entender das fontes de dados, nem das relações entre elas.

Para tanto, a integração dos dados demanda de uma forma de acesso ou publicação, e neste trabalho o acesso aos dados integrados através da camada semântica contida em KG-E é realizado com base na abordagem Híbrida, onde mediante utilização de um Federador, cada fonte de Dados Ligada é vista sob um *Endpoint*, ou acessada com *wrappers* em casos de fontes relacionais ou CSV, o *wrapper* é responsável por transformar essas fontes em Dados Ligados. A Figura 2 apresenta uma visão desse acesso.

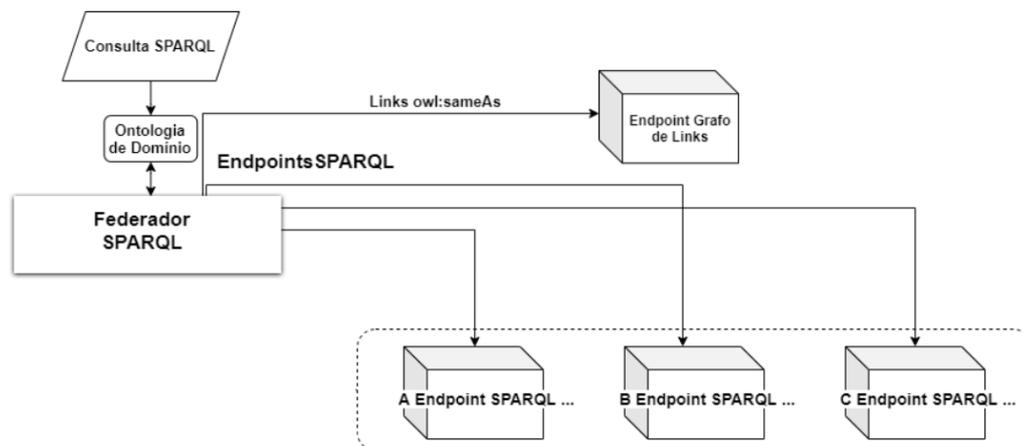


Figura 2. Acesso ao KG-E.

4. Estudo de Caso

O estudo de caso baseou-se em validar a integração semântica dos dados resultante em KG-E através de questões de competência. Essas questões visam identificar a cobertura dos dados e conceitos do KG gerado sob o domínio de empresas e fiscal. Para tanto, foram utilizadas as seguintes questões:

1. (QC1) Existem empresas baixadas suspensas que efetuaram algum processo de cisão, fusão ou incorporação?
2. (QC2) Existem sócios que constituem o quadro societário de empresas distintas?

A Consulta 1 apresenta uma possível consulta SPARQL para QC1, onde a Figura 3 apresenta o seu retorno, nesta consulta são filtradas as empresas cuja situação cadastral esteja baixada e cuja motivação esteja dentro da faixa de valores que representam os eventos de cisão, fusão ou incorporação. Por sua vez, QC2 é respondida pela Consulta 2, onde a Figura 4 apresenta o retorno da consulta, neste caso a consulta realiza a contagem de participações de um sócio em empresas distintas, filtrando aqueles que participam em mais de uma empresa.

Consulta 1. Existem empresas baixadas suspensas que efetuaram algum processo de cisão ou fusão ou incorporação?

```

1 PREFIX ode: <http://www.redesim.ceis.gov.br/ontology/>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
5 SELECT ?cnpj ?empresa WHERE {
6   ?emp a foaf:Organization;
7     ode:nome_fantasia ?empresa;
8     ode:cnpj_raiz ?cnpj;
9     ode:tem_situacao_cadastral ?situacao.
10  ?situacao ode:codigo_baixa ?codigo_baixa.
11  FILTER (?codigo_baixa >= 2 || ?codigo_baixa <= 4)
12 }

```

	cnpj	empresa
1	"0727940"	"INTERVILAS"
2	"1919210"	"SELECTA SERVICES"
3	"0807236"	"IOBV"
4	"1696558"	"MAXIMO ENGENHARIA"
5	"1357152"	"MARKETUP"
6	"0221526"	"LOC MAXX"
7	"0858776"	"WT LOCACOES E EVENTOS"
8	"1078862"	"DELTA CONSTRUCAO"
9	"1526224"	"J.M. LOTERIA"
10	"0253412"	"V I P A C"

Figura 3. Resultado da QC1.
Consulta 2. Existem sócios que constituem o quadro societário de empresas distintas?

```

1 PREFIX ode: <http://www.redesim.ceis.gov.br/ontology/>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
5 SELECT ?socio ?cpf (COUNT(DISTINCT ?x) AS ?quantidade)
WHERE {
6     ?x a foaf:Organization;
7     ode:tem_sociedade ?sociedade.
8     ?sociedade ode:tem_socio ?socio_empresa.
9     ?socio_empresa rdfs:label ?socio;
10    sefazma:cpf ?cpf.
11 }
12 GROUP BY ?socio ?cpf
13 HAVING (?quantidade > 1)
14 ORDER BY desc (?quantidade)

```

	socio	cpf	quantidade
1	"ALTEMIR BRAZ DANTAS"	****569928***	"15"^^xsd:integer
2	"ARMANDO JOSE RINALDI BALBI"	****440807***	"8"^^xsd:integer
3	"WALTER VOGEL"	****513929***	"7"^^xsd:integer
4	"ANDRE MARQUES RECACHO"	****226058***	"6"^^xsd:integer
5	"MARCIA GALLUCCI TOMASELLI"	****302718***	"6"^^xsd:integer
6	"MIRIAM DE OLIVEIRA MARQUES RECACHO"	****819218***	"6"^^xsd:integer
7	"GABRIEL ALFIO TOMASELLI"	****992558***	"6"^^xsd:integer
8	"CLAUDIO HENRIQUE FONTENELLE SANTOS"	****284983***	"5"^^xsd:integer
9	"MARIA AMALIA DELFIM DE MELO COUTRIM"	****298507***	"5"^^xsd:integer
10	"ALEXIS STAHIS PANAGIDES"	****437423***	"5"^^xsd:integer
11	"ARTHUR SCHAPER NETO"	****756576***	"5"^^xsd:integer
12	"VERONICA VALENTE DANTAS"	****853205***	"5"^^xsd:integer
13	"ADRIANO LINO SANTOS"	****106965***	"4"^^xsd:integer

Figura 4. Resultado da QC2.

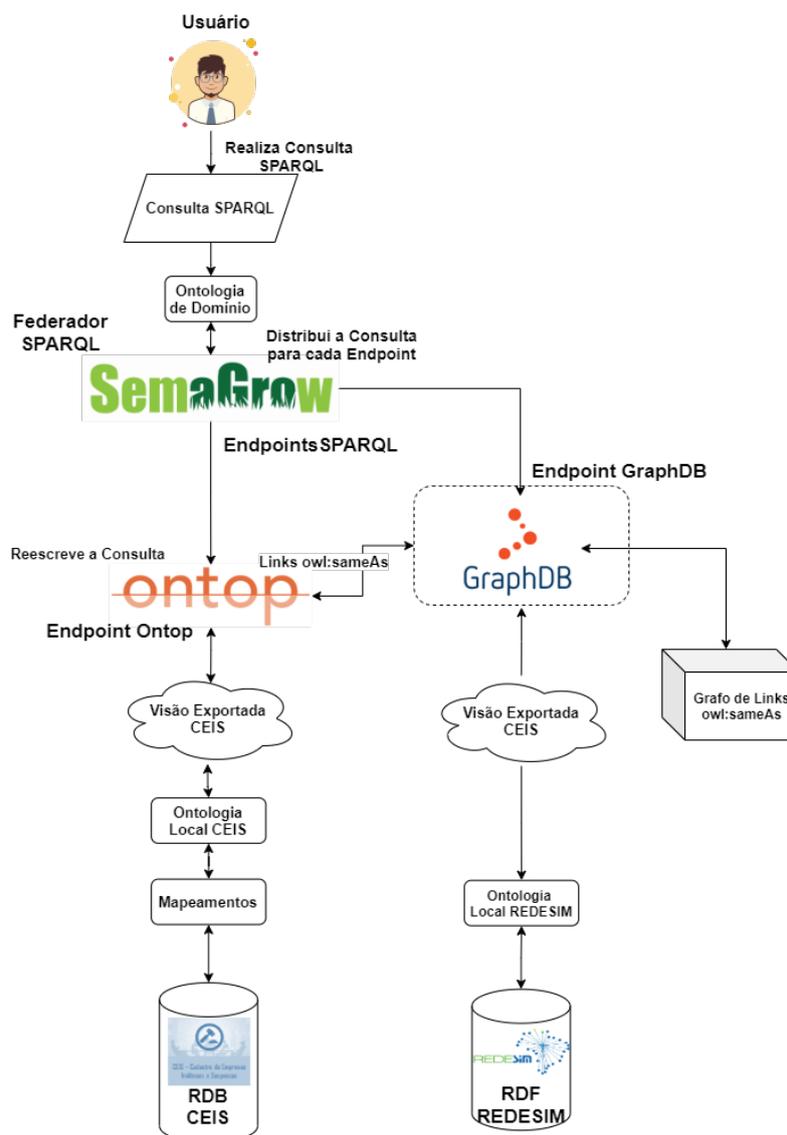


Figura 5. Processo de consulta ao KG-E

Cada questão de competência descrita abaixo foi mapeada e reescrita em SPARQL seguindo o processo apresentado na Figura 5, sendo realizados os seguintes passos:

1. Uma consulta SPARQL Q é realizada por um usuário com base nos termos de uma ontologia de domínio;
2. A Consulta Q é processada através do Semagrow que é um Federador SPARQL F' responsável por fazer a seleção das fontes $S_1 \dots S_n$ aptas a responderem um *triple pattern* t ;
3. Cada fonte é vista pelo Federador como um Endpoint SPARQL $E_1 \dots E_n$. Posteriormente a consulta Q é dividida em sub-consultas $Q_K \dots Q_{K_n}$ e executadas em E_1 e E_2 ;
4. Cada Endpoint é tratado individualmente, onde para o armazenamento relacional é utilizado o Ontop para realização da virtualização através de um *wrapper* que obtém uma sub-consulta Q_K e faz a reescrita em uma consulta SQL Q_Q sob S_1 .

Para S2 no formato RDF os dados são armazenados no Triplestore (GraphDB) e acessados diretamente através de E2.

5. As ligações *owl:sameAs* disponíveis em um Endpoint são consultadas pelo Fedeador em tempo de consulta;
6. Após processar cada sub-consulta *QK..QKn* o resultado é retornado ao Semagrow que faz a junção dos resultados e retorna diretamente ao usuário;

5. Conclusões

Este trabalho apresentou a construção de um KG Semântico com base em um processo de integração semântica com base em *Linked Data* sob as bases heterogêneas do REDESIM e CEIS.

Após construção e validação através do estudo de caso, os resultados sugerem que KG-E permite a descoberta de informações que à ‘piori’ não eram possíveis em fontes isoladas, podendo vir a ser utilizado em aplicações e análises, possibilitando as seguintes contribuições:

- Busca pela descoberta de relações significativas de empresas e sancionados através da semântica;
- Obter *insights* a partir de grandes quantidades de informações, independentemente da variedade de formato ou fonte;
- Maximizar o uso e reutilização dos dados de empresas, estabelecendo as bases para construção de aplicações semânticas e de Inteligência Artificial (IA) a partir dos dados da REDESIM e CEIS com ênfase em identificar padrões de empresas e sócios sancionados.

Este trabalho possui limitações quanto ao conjunto de experimentos realizados, tendo em vista a não presença de especialistas de domínio durante o processo. Ainda, outra limitação é expressa na não atualização frequente dos dados, demandando da implementação de um ”Scrapper” para monitoramento e atualização.

Como trabalhos futuros, pretende-se obter novas questões bem como permitir a descoberta de novas informações de interesse ao domínio.

Agradecimentos

Este trabalho é uma produção do Projeto ”Uso das Tecnologias da Web Semântica e Big Data para Integrar, Enriquecer e Explorar Dados da SEFAZ”, sob acordo de parceria nº 0001/2019-SEFAZ/PROFISCO II realizado entre a Secretaria da Fazenda do Maranhão e a Universidade Federal do Ceará.

Referências

- Bernardi, T. L. et al. (2017). *Uma ontologia sobre licitações aplicada na elicitação de requisitos de portais de transferência municipal*. PhD thesis, Universidade de Passo Fundo.
- BRASIL (2021). Redesim. disponível em: <<http://https://www.gov.br/empresas-e-negocios/pt-br/redesim/sobre-a-redesim>>. Acessado em: 21-03-21.

- Calvanese, D. et al. (2017). Ontop: Answering sparql queries over relational databases. *Semantic Web*, 8(3):471–487.
- da Cruz, M. M. L., Avila, C. V. S., Vidal, V. M. P., and Junior, N. M. A. (2019). Semanticus: Um portal semântico baseado em ontologias e dados interligados para acesso, integração e visualização de dados do sus. In *Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 13–18. SBC.
- da Fonseca, L. B., Azevedo, C. L., and Almeida, J. P. A. (2014). Mapeando dados governamentais com uma ontologia de organizações.
- Fonseca, L. B. R., Detoni, A. A., Almeida, J. P. A., and de Almeida Falbo, R. (2016). Uma proposta de ontologia de referência para autorização orçamentária e execução da despesa pública.
- Madhavan, J., Jeffery, S. R., Cohen, S., Dong, X. L., Ko, D., Yu, C., and Halevy, A. (2007). Web-scale data integration: You can only afford to pay as you go.
- Marques, C. G. N. (2013). Estudo de Caso na Aplicação de Web Semântica no Desenvolvimento de Sistemas Voltados para Dados Abertos. Monografia (Bacharel em Ciência da Computação), Faculdade Farias Brito, Fortaleza, Brasil.
- Rolim, T., Avila, C., Arruda, N., Silva, J., Maia, J., Oliveira, M., Andrade, L., and Vidal, V. (2020). Um enfoque incremental para construção do grafo de conhecimento do sus. In *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 72–83. SBC.
- Rolim, T. V., Vidal, V. M. P., Avila, C. V. S., Cruz, M. M. L. d., Barrio, M., and Queiroz, D. (2019). Semanticsefaz: an ontology-based semantic portal for the government spending. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, pages 493–496.
- Soylu, A., Corcho, O., Simperl, E., Roman, D., Francisco, Y. M., Taggart, C., Makgill, I., Elvesæter, B., Symonds, B., McNally, H., et al. (2018). Towards integrating public procurement data into a semantic knowledge graph.
- Tosin, T. d. M. (2015). *Um modelo para integração de informações de bases de dados abertos, com uso de ontologias*. PhD thesis, Universidade do Vale do Rio dos Sinos.
- Vidal, V. M., Casanova, M. A., Menendez, E. S., Arruda, N., Pequeno, V. M., and Leme, L. A. P. (2016). Using changesets for incremental maintenance of linkset views. In *International Conference on Web Information Systems Engineering*, pages 196–204. Springer.
- W3C (2012). R2RML: RDB to RDF Mapping Language. disponível em: <<https://www.w3.org/TR/r2rml/>>. Acessado em: 14-04-2021.