

Relato de eventos assistido por ontologias para geração de triplas semânticas

Javan O. Cruz¹, Paulo C. Prandel¹, Marcio C. Victorino¹, Edison Ishiwaka¹

¹Departamento de Ciência da Computação
Universidade de Brasília (UnB) – Brasília – DF – Brasil

{javan.cruz,paulo.prandel}@aluno.unb.br

{ishikawa,mcvictorino}@unb.br

Abstract. *In intelligence operations, due to the dynamics involved, the agility in data collection and the structuring of the information to be made available are factors of considerable relevance. The purpose of this work is to provide a system that allows obtaining data through field observers, in an assisted manner, making available the information collected. For this, ontology resources were used for the representation of knowledge, as well as concepts aimed at organizing and structuring information. As a result, semantics triple are generated that structure the report of the observed event, allowing better conditions for the decision-making process by authorities.*

Resumo. *Em operações de inteligência, pela dinâmica envolvida, a agilidade na coleta de dados e a estruturação das informações a serem disponibilizadas são fatores de considerável relevância. A proposta deste trabalho é fornecer um sistema que permita a obtenção de dados por meio de observadores de campo, de forma assistida, bem como disponibilizar as informações coletadas. Para isso, foram empregados recursos de ontologias para a representação do conhecimento, bem como conceitos voltados à organização e estruturação da informação. Como resultado, foram geradas triplas semânticas que estruturam o relatório do evento observado, permitindo melhores condições para o processo de tomada de decisão por autoridades.*

1. Introdução

O relato de eventos tem importância especial nas áreas militar e de segurança pública, sendo um dos principais meios de obtenção de informações de inteligência. Agentes em campo encontram-se quase sempre conectados a uma base de operações, a qual registra, processa e distribui informações aos demais agentes e às autoridades. Esses eventos são considerados qualquer incidente não planejado que possa comprometer a segurança de uma atividade, requerendo tratamento adequado.

A despeito dos constantes avanços tecnológicos, que permitem a obtenção de informações de forma ágil e precisa, a coleta de informações por observadores humanos continua sendo essencial. Porém, a simples obtenção de dados, de forma não estruturada, não é julgada suficiente para a formação de conhecimentos específicos sobre uma determinada situação conflituosa. A estruturação das informações é de destacada relevância para prover a consciência situacional de uma operação em andamento com eficiência.

Nesse contexto, este trabalho apresenta uma proposta de implementação de um sistema de construção de relatórios estruturados, por meio do preenchimento de um formulário assistido por ontologias, buscando entregar uma interface ágil e intuitiva para os agentes de campo. Como resultado apresentado, o sistema gera triplas semânticas que permitem a representação do conhecimento, estruturando o relatório do evento registrado.

No trabalho apresentado por [Christian Mårtenson and Kabir 2011] é realizada a descrição conceitual de uma ferramenta de criação de relatórios estruturados, orientados por ontologia, sem ser efetivada a implementação do que foi teorizado. Na presente proposta foi buscado cobrir lacunas deixadas no trabalho citado, contemplando a aplicação prática de um sistema de estruturação de relatórios com base em ontologias.

Como contribuição, o relatório estruturado provido pelo sistema facilita a análise das informações obtidas, sendo pertinente para proporcionar melhores condições para o processo de tomada de decisão, a ser efetivado pelas autoridades competentes.

2. Referencial Teórico

2.1. Ontologias

Uma das formas de estruturar dados, fazendo a sua representação de forma eficiente, é por meio do uso de ontologias. Aplicado à ciência da computação, o termo ontologia pode ser definido como um conjunto de conceitos fundamentais e suas relações [Mizoguchi 2004]. [Gruber 1993] afirma que ontologia é a especificação de uma conceituação.

Na área computacional, [Guarino et al. 2009] afirma que ontologias são um meio de modelar formalmente a estrutura de sistemas, ou seja, as entidades e relacionamentos relevantes, sendo úteis para os seus propósitos.

Ontologias propiciam um vocabulário controlado de uso geral, possibilitando o entendimento de termos que descrevem dados, explicitando entidades e seus relacionamentos, atribuindo sentido e significado a eles. De acordo com [Noy 2004] as ontologias ganharam grande popularidade na comunidade, como uma forma de estabelecer um vocabulário formal para compartilhamento entre aplicações.

2.2. Triplas RDF

O *RDF (Resource Description Framework)* é reconhecido como uma linguagem que descreve recursos de informação na Web, permitindo a sua descrição formal bem como o seu acesso por máquinas.

O modelo *RDF* é uma coleção de triplas que podem ser visualizadas como um multigrafo dirigido, sendo adequado para a representação de informações [Leal 2013]. Uma instância *RDF* representa a relação entre dois recursos, o sujeito e o objeto, a partir de um predicado que os conecta [Isotani and Bittencourt 2015], consolidando uma tripla.

Dentre os principais objetivos do *RDF*, destaca-se a construção de uma rede de informações formada por nós ligados semanticamente, com informações de diversas fontes. Paralelamente, as ontologias são recursos empregados para efetivar uma relação organizada entre termos pertencentes a um domínio contextualizado, permitindo a interpretação e a recuperação de dados de forma eficiente [Victorino et al. 2018]. Nesse contexto, o *RDF* apresenta um formato robusto para representação do conhecimento por meio de ontologias.

A interação entre o sistema implementado e a ontologia seguiu padrões do *World Wide Web Consortium* (W3C), empregando recursos do *RDF*, de sua extensão que amplia o vocabulário existente, o *RDFS* (*Resource Description Framework Schema*), da *OWL* (*Web Ontology Language*), para definir e instanciar ontologias, além da *SPARQL* (*Protocol and RDF Query Language*), para realização de consultas em grafos *RDF*.

2.3. Ontologias de alto nível

Existem milhares de ontologias na *World Wide Web*, compostas por diversos conceitos e relacionamentos entre eles. Algumas dessas ontologias possuem um domínio específico de aplicação, enquanto outras são mais primitivas, conhecidas como ontologias de alto nível (*upper ontologies*), possuindo especificações para diversas áreas do conhecimento.

A utilização de ontologias de alto nível permite descrever classes gerais, presentes no contexto de vários domínios, bem como suas relações e atributos básicos [Kiryakov et al. 2004]. De acordo com [Mascardi et al. 2007], essas ontologias são essenciais para a integração de conhecimentos heterogêneos oriundos de fontes diversas.

Neste trabalho foi utilizada a *PROTON*, uma ontologia de alto nível leve. Ela fornece cobertura de conceitos necessários para uma ampla gama de tarefas [Mascardi et al. 2007]. Sua última atualização foi em junho de 2020 [DBPedia 2021], sendo julgada pertinente para a utilização neste trabalho. Possui mais de 500 classes, com propriedades específicas, cobrindo conceitos que contemplam operações de inteligência.

Cabe destacar que o sistema desenvolvido não se restringe à utilização da *PROTON*, sendo possível utilizar outra ontologia que possua termos adequados à temática buscada. Isso proporciona flexibilidade, uma vez que a entrada de dados estará condicionada à ontologia carregada no sistema. Assim, a proposta apresenta vantagem em relação a um cadastro de eventos que possui um formulário com termos previamente definidos.

3. Metodologia

3.1. Visão geral do sistema

A Fig. 1 apresenta a visão geral do sistema, no qual um observador, após presenciar um evento, preenche um relatório por meio de uma aplicação. Esse sistema é baseado em formulários e o seu preenchimento é assistido pelo emprego de uma ontologia. Ao final do preenchimento, o sistema realiza a indexação dos termos, gerando triplas semânticas as quais irão compor o relatório estruturado, a ser enviado para um repositório central.

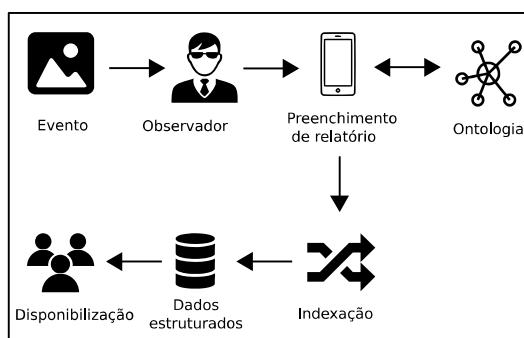


Figura 1. Visão Geral do Processo

O sistema foi escrito em *Python*, utilizando-se o *framework Django*. A interface do usuário é baseada no preenchimento e no envio de formulários. Na lógica da exibição dos formulários, utiliza-se o pacote *RDFLib* [RDFLib Team 2020] para as funções relacionadas ao processamento de ontologias.

O *RDFLib* emprega uma estrutura de grafo para representar as triplas semânticas da ontologia. Dessa maneira, além das consultas SPARQL, o pacote oferece uma série de métodos para a extração de informações, como por exemplo funções que iteram sobre o grafo na busca de recursos da ontologia.

3.2. Arquitetura

Para a organização do relatório de evento, foi concebida a construção de uma estrutura inicial. Assim, um formulário de entrada de dados a ser preenchido pelo observador do evento foi construído, possuindo três divisões fixas, baseadas nos atributos “atores”, “lugar” e “ação”.

Essa subdivisão padronizada tem por finalidade facilitar a anotação da ontologia, permitindo a subdivisão de um evento observado em três triplas semânticas principais, as quais possuem o evento como sujeito, além dos três atributos citados reconhecidos como objetos, conforme a Fig. 2.

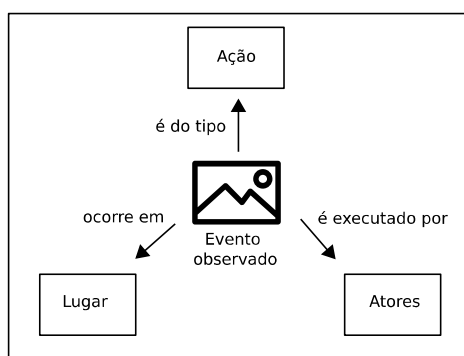


Figura 2. Representação do evento por meio de triplas *RDF*

Para cada um desses atributos, é buscada uma classe na ontologia que melhor represente o atributo em questão. A busca por essa classe é iniciada por uma palavra-chave inserida pelo usuário, que resultará em uma *query SPARQL* na ontologia. A palavra deverá ser inserida em inglês, uma vez que este é o idioma da ontologia empregada (PROTON). Essa palavra-chave deverá sintetizar, em um único termo, a definição do atributo do evento. A busca é realizada sobre os *labels* e os *comments* das classes da ontologia, utilizando-se para isso métodos de busca com expressões regulares (*regular expressions - regex*). Os resultados da busca da palavra-chave retornarão por meio de uma lista em ordem decrescente de relevância, conforme o que segue (da mais relevante para a menos relevante):

- palavra-chave encontrada no *rdfs:label* (*match* exato)
- palavra-chave encontrada no *rdfs:label* (*match* parcial)
- palavra-chave encontrada no *rdfs:comment* (*match* exato)
- palavra-chave encontrada no *rdfs:comment* (*match* parcial)

Após o resultado da busca, caso o usuário identifique a classe de interesse, realizará a seleção. Se nenhuma classe for retornada, o usuário deverá inserir uma nova palavra-chave que represente o atributo. Após a seleção da classe inicial, o usuário poderá refinar a sua busca. Para isso, o sistema buscará as subclasses da classe selecionada (*rdfs:subClassOf*), de forma que sejam disponibilizadas ao usuário, para seleção.

Com a classe/subclasse encontrada, será instanciado o atributo utilizando o recurso *rdf:type* para efetivar o relacionamento. Com o atributo instanciado, o formulário exibirá então as *data properties* da classe selecionada, permitindo ao usuário o preenchimento das propriedades julgadas pertinentes. Essas propriedades dos atributos do evento são as informações que descrevem em detalhes a classe selecionada e são dinâmicas, de acordo com as classes da ontologia utilizada no sistema.

Ao final do processo, o sistema instancia o evento. A partir das três triplas *RDF* iniciais representadas por seus atributos (“atores”, “lugar” e “ação”), serão geradas as demais triplas, detalhando classes/subclasses as quais os atributos pertencem, bem como as suas propriedades. Com isso, serão geradas as triplas *RDF* que descrevem o evento observado, proporcionando a construção de um relatório estruturado.

3.3. Interface

Uma parte do formulário utilizado para a descrição do evento e criação das triplas pode ser visualizado na Fig 3-A, a qual apresenta o registro de um ator no sistema. Uma classe pode ser refinada após ser selecionada, conforme mostrado na Fig 3-B. O sistema busca por subclasses dessa classe selecionada (*rdfs:subClassOf*).

Após a seleção da classe desejada, as propriedades sob seu domínio serão exibidas em um novo formulário. As Fig. 3-C e 3-D, apresentam o detalhamento de propriedades a serem preenchidas pelo usuário, sendo exemplificadas as propriedades do Lugar e da Ação do Evento. Cabe destacar que os termos apresentados para descrição das propriedades da classe a serem preenchidas são os *labels* das *data properties* das classes da ontologia.

3.4. Geração de Relatório

Após o preenchimento dos três formulários, o sistema gera as triplas semânticas que representam o evento, com base nas classes selecionadas e nos valores preenchidos como propriedades. Esses eventos poderão ser salvos em um repositório que permitirá a recuperação dos eventos registrados, para posterior análise por autoridade competente. Um exemplo de evento relatado pelo sistema pode ser visto na Tab. 1:

Tabela 1. Exemplo de Evento Relatado

evento01 <i>rdf:type</i> Evento	ator01 <i>hasApproximateAge</i> '25'
ator01 <i>rdf:type</i> Man	ator01 <i>hasRace</i> 'Branco'
acao01 <i>rdf:type</i> SuspiciousActivity	ator01 <i>isWearing</i> 'Camiseta azul'
lugar01 <i>rdf:type</i> Street	lugar01 <i>hasName</i> 'Pistão Sul'
evento01 <i>hasActor</i> ator01	lugar01 <i>hasCity</i> 'Brasília'
evento01 <i>hasAcao</i> acao01	lugar01 <i>hasNeighborhood</i> 'Taguatinga'
evento01 <i>hasLocation</i> lugar01	acao01 <i>hasDetails</i> 'Provável traficante'



Figura 3. Interface do sistema de geração de triplas semânticas

4. Conclusão e Trabalhos Futuros

Neste trabalho foi apresentada a implementação de um sistema capaz de registrar eventos observados por agentes de campo, tomando por base o fornecimento de formulários assistidos por ontologias. Com isso, foi proporcionada a construção de relatórios estruturados, nos quais os conhecimentos são representados por meio de triplas semânticas.

A partir da recuperação dos relatórios gerados pelo sistema, são proporcionadas melhores condições para a análise das informações obtidas. Isso contribuirá para um processo decisório mais assertivo, principalmente no que diz respeito às ações a serem desencadeadas para tratar incidentes relatados.

Uma oportunidade para trabalhos futuros seria a construção de uma ontologia específica para o contexto dos eventos observados em operações de inteligência, de preferência na língua portuguesa. Pelo fato de ser muito abrangente, a ontologia de alto nível deixa lacunas durante a busca por classes dentro do contexto específico apresentado.

Referências

- Christian Mårtenson, A. H. and Kabir, Z. (2011). An ontology-based adaptive reporting tool. In *6th International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS) 2011*, pages 85–88. CEUR Workshop Proceedings.
- DBPedia (2021). Dbpedia arquivo. Disponível em <https://databus.dbpedia.org/ontologies/>. Acessado em: dezembro. 08, 2020.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5:199–220.
- Guarino, N., Oberle, D., and Staab, S. (2009). What is an ontology? *Handbook on Ontologies - International Handbooks on Information Systems*, pages 1–17.
- Isotani, S. and Bittencourt, I. I. (2015). *Dados Abertos Conectados*. Nnovatec Editora, São Paulo.
- Kiryakov, A., Popov, B., Terziev, I., Manov, D., and Ognyanoff, D. (2004). Semantic annotation, indexing, and retrieval. *Journal Of Web Semantics*, 2(1):49–79.
- Leal, J. P. (2013). Using proximity to compute semantic relatedness in rdf graphs. *Computer Science And Information Systems*, 10(4):1727–1746.
- Mascardi, V., Cordì, V., and Rosso, P. (2007). A comparison of upper ontologies. In *WOA 2007 - 8th AI*IA/TABOO Joint Workshop "from Objects to Agents": Agents and Industry: Technological Applications of Software Agents*, pages 55–64. CEUR Workshop Proceedings.
- Mizoguchi, R. (2004). Tutorial on ontological engineering - part 3: Advanced course of ontological engineering. *New Generation Computing*, 22(2):198–220.
- Noy, N. F. (2004). Semantic integration: A survey of ontology-based approaches. *ACM SIGMOD Record*, 33:65–70.
- RDFLib Team (2020). Rdfliib. Disponível em: <https://rdflib.readthedocs.io/en/stable/>. Acesso em: 05 novembro 2020.
- Victorino, M., de Holanda, M. T., Ishikawa, E., Oliveira, E. C., and Chhetri, S. (2018). Transforming open data to linked open data using ontologies for information organization in big data environments of the brazilian government: the brazilian database government open linked data dbgoldbr. *Knowledge Organization*, 45(6):443–466.