

ANOTAÇÃO DE DADOS PARA GERAÇÃO DE INDICADORES DE DESEMPENHO EM ORGANIZAÇÕES

Data Annotation for Generating Performance Indicators in Organizations

Marcello Peixoto Bax¹, Evaldo de Oliveira da Silva²

(1) Programa de Pós-Graduação em Ciência da Informação – UFMG, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, bax@eci.ufmg.br

(2) Centro de Ensino Superior de Juiz de Fora (CESJF), Rua Halfeld, 1.179, Centro Campus Academia - MG, 36016-000, evaldosilva@cesjf.br.

Resumo:

Key Performance Indicators (KPIs) são usados por organizações para avaliar o exercício de suas atividades, apoiando a decisão. Com base nesses indicadores, elas revêem seus processos buscando a melhoria contínua das atividades. Modelos de dados dimensionais estruturam os dados agrupados em "fatos" e "dimensões". Os fatos são representados por campos numéricos que alavancam a geração de KPIs. Observe, contudo, a necessidade de boas práticas para nomear e anotar dados com metadados. Assim, diferentes usuários compreendem melhor o conjunto de dados, evitando interpretações divergentes. Descreve-se um processo de anotação semântica usando dicionário de dados, que associa dados a conceitos, permitindo a geração de KPIs. Apresenta-se como se dá a geração de desses indicadores pelo enriquecimento semântico dos dados com ontologias.

Palavras-chave: Modelos Dimensionais, Indicadores de Desempenho, KPI, Dicionário de Dados, Ontologia, Anotação Semântica

Abstract:

Key Performance Indicators (KPIs) are used by organizations to evaluate the performance of their activities and decision support. Based on these indicators, they review their processes seeking continuous improvement of activities. Dimensional data models structure data grouped into "facts" and "dimensions." Facts are represented by numeric fields that leverage the KPIs generation. There is, nevertheless, a need for good practices of naming and annotating data with metadata. Thus, different users better understand the dataset, avoiding divergent interpretations. We describe a semantic annotation process using data dictionary, which associates data with concepts, allowing the generation of KPIs. We present how these indicators are generated by semantic enrichment of data with ontologies.

Keywords: Dimensional Data, Performance Indicators, KPI, Data Dictionary, Ontology, Semantic Annotation

1. Introdução

Um indicador chave de desempenho (KPI, *Key Performance Indicator*) é um valor que pode ser medido e que demonstra a eficácia da organização em alcançar resultados (PARMENTER, 2015). KPIs permitem avaliar o atingimento de metas, avaliar resultados e rever processos capacitando a melhoria contínua das atividades. Valores KPI criam base analítica para tomada de decisões que priorizam ações avaliadas (empiricamente) como as mais relevantes.

KPIs são, p.ex., receitas, lucros, preços e custos, medidas de qualidade ou satisfação. Gestores e executivos interpretam KPIs para decidirem com base científica, empírica. Exemplo comum de mensuração é o percentual de aderência da realização de atividades com o planejamento. KPIs podem

ser vistos também no meio acadêmico. De acordo com o *Central European Research Infrastructure Consortium* (CERIC) KPIs podem avaliar o grau do alcance de objetivos de instituições de ensino ou programas de pesquisa. KPIs são insumos para gerenciar e monitorar o atingimento de objetivos e auxiliar o planejamento estratégico (KOLAR, HARRISON e GLIKSOHN, 2019).

Para Kimball e Ross (2013) a criação de KPIs deve ser disciplinada em suas práticas de nomeação de dados. Assim, caso seja impossível entender o conjunto de dados (*datasets*) a ser utilizado para gerar os cálculos, nomes diferentes serão atribuídos a diferentes interpretações. Com isso, os KPIs acabam resultando de combinações de dados incompatíveis, comprometendo os valores e prejudicando a tomada de decisão.

É necessário garantir a qualidade dos dados (MEDEIROS, 2018) e a Curadoria Digital propõe técnicas de descrição com metadados que favorecem qualidade, preservação e facilitam a descoberta de novos conhecimentos pelo reúso de dados. No entanto, somente a definição dos metadados não basta para extrair e compartilhar *datasets*. Dados usados para geração de KPIs podem vir de estruturas e modelos de dados distintos e requerem informações adicionais para que seus significados sejam explicitados. Bastante aplicados na descrição de *datasets*, os dicionários de dados apoiam atividades de gerenciamento, procedimentos de conversão, validação e critérios para armazenar dados. Ontologias e tecnologias semânticas enriquecem e formalizam o significado dos KPIs, evitando interpretações discrepantes.

Desta forma, descreve-se aqui um processo de anotação baseado em Dicionários Semânticos de Dados (SDDs) que contribui com a curadoria, dentre outros elementos, por estar alinhado com princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) de Wilkinson et. al, 2016). O "caso de uso" apresentado neste artigo anota os dados de um modelo dimensional para cálculos de KPIs.

A Seção 2 traz o conceito de modelagem dimensional de dados para KPIs e trabalhos correlatos. A Seção 3 descreve o processo de anotação proposto por Rashid et. al (2017). A Seção 4 relata a anotação para criação de um KPI para monitoramento de projetos de pesquisa. A Seção 5 faz considerações finais e sugere trabalhos futuros.

2. Modelagem de KPIs

Um modelo de dados dimensional é construído, agrupando dimensões ao redor de dados numéricos. Os fatos são estruturados relacionando dados e suas dimensões. A análise dos fatos usa as dimensões (facetas) combinando filtros que atendem as necessidades do usuário, na tomada de decisão (KIMBAL e ROSS, 2013). O modelo dimensional da Figura 1 (Apêndice A) usa esquema estrela e permite o cálculo dos montantes de publicações pelas dimensões: fator de impacto, centro de pesquisa, mês e ano.

2.1. A Ontologia KPIOnto

A anotação pelo SDD, exige compreender o domínio por sua modelagem conceitual prévia. Deve-se selecionar os dados e encontrar os termos/vocabulários existentes que referenciam os conceitos do domínio, explicitando e formalizando sua semântica com o uso de ontologias. A anotação ontológica, permite a geração de fragmentos (declarações em formato de triplas) do conhecimento do domínio em RDF (*Resource Description Framework*).

Diamantini, Potena e Storti (2016) propõem a KPIOnto que usamos para anotação e alinhamento conceitual de diferentes profissionais sobre os KPIs. A KPIOnto constitui-se de classes como: Indicador, Dimensão e Fórmula; sendo "Indicador" a classe principal. Ela especifica um indicador pelas propriedades: *hasDimension*, *hasFormula* e *hasAggrFunction* (para uso de funções de agregação).

2.2. Dicionário Semântico de Dados

Rashid et. al (2017) utiliza padrões de metadados para configurar a anotação semântica por um SDD. Recomenda ainda a utilização da ontologia SIO (*Semanticscience Integrated Ontology*) que fornece propriedades para descrever os relacionamentos entre objetos e atributos como modelo de representação do conhecimento. A anotação semântica proposta por Rashid et. al (2017) utiliza os seguintes documentos:

- *InfoSheet*: referências para descrição dos SDDs;
- *Dictionary Mapping*: anotação semântica das colunas das coleções de dados;
- *CodeBook*: códigos correspondentes a conceitos de ontologia;
- *Code Mapping*: mapeamento de termos dos *datasets* que correspondem a conceitos existentes na ontologia;
- *TimeLine*: anotação de intervalos temporais;
- *Properties Table*: para fins de customizar a descrição por outras ontologias de topo.

A ferramenta *sdd2rdf* (SEMANTIC DATA DICTIONARY, 2019) interpreta o SDD e "ingere" os dados, formando um grafo RDF. Para acessar os dados anotados, o *sdd2rdf*

cria consultas no formato SPARQL¹. São geradas também regras SWRL² que auxiliam em novas inferências. O grafo RDF gerado pelo script *sdd2rdf* utiliza o vocabulário formal ontológico, e possibilita a interoperabilidade dos dados.

2.3. Trabalhos Correlatos

Kritikos (2017) descreve que os dados vinculados (*Linked Data*) representam um grande mecanismo para a integração de informações entre fontes distintas, permitindo a realização de inferências para derivar conhecimento. Utiliza esta ideia no contexto do processo de negócios como serviço (BPaaS) a fim de coletar e vincular informações originadas de diferentes sistemas. Propõe o uso de ontologias principais que visa melhorar a comparação de KPIs gerados dos dados integrados entre os sistemas.

Wetzstein, Ma e Leymann (2008) propõe que KPIs sejam modelados por analistas de negócios que exploram anotações semânticas de processos de negócios. Os modelos de KPI são automaticamente calculados para serem geridos por meio de um painel de monitoramento em tempo real.

Kourtesis e Alvarez-Rodrigues (2014) sugerem uma estrutura semântica para gerenciamento de QoS (*Quality of Service*). Utilizam abordagens para o gerenciamento de QoS baseado em semântica, bem como os principais métodos, técnicas para explorar diversos dados.

Silva et. al. (2018) propõe um conjunto de funções para compor a estrutura semântica para definição de dicionário de dados. Apresenta ainda como estrutura semântica está relacionada à configuração sintática dos dicionários de dados, a fim de identificar padrões que possam ser usados no desenvolvimento de procedimentos para extração de informações e modelos semânticos.

3. Processo de Anotação Semântica

A anotação baseia-se em Rashid et. al (2017), que segue princípios FAIR e permite gerar o grafo RDF (*script sdd2rdf*) persistido no *triplestore*. A ontologia formaliza o

vocabulário e abre caminho para interoperabilidade de dados. Após escolher que dados do *dataset* anotar, segue-se para a criação dos artefatos abaixo, em cada etapa do processo:

1. Ontologia de domínio. Criação/ajuste de ontologia de domínio para formalização dos conceitos tratados no problema de pesquisa. Buscar reutilizar ontologias consolidadas no domínio do problema.
2. Dictionary Mapping. Cada linha do DM mapeia uma coluna do *dataset*, formalizando-a conceitualmente e também suas relações e proveniência.
3. CodeBook. Permite a criação dos seguintes campos: Coluna (entidade a ser anotada), Código, Descrição e a Classe da Ontologia.
4. Infosheet. Metadados de um SDD que organiza e descreve a coleção de arquivos de metadados (planilhas do Excel) usados pelo SDD em questão.
5. Grafo RDF. Interpretação da dupla: "SDD + Dados" pelo script *sdd2rdf*, gerando o RDF e armazenando-o em *triplestore* para consulta posterior.

Os dados dos objetos mapeados pelo SDD são as colunas do próprio *dataset*. Porém, Rashid et al. (2017) afirmam que os objetos descritos no *dataset* podem encontrar-se ali explícita ou implicitamente. Ou seja, no mesmo *dataset* podem aparecer também atributos de outros objetos implicitamente representados ali. Estes objetos serão explicitados no SDD e formalizados no grafo final gerado (pelo script *sdd2rdf*), favorecendo a sua integração nos níveis conceituais (ou intencionais) mais abstratos do projeto.

4. Anotação de Dados e Geração de KPIs

Descreve-se exemplo de anotação de dados para geração de KPIs a partir da necessidade do acompanhamento de índices de publicação em centros de pesquisa. O modelo da Figura 1 foi utilizado como fonte de dados. Abaixo segue a descrição da execução do processo de anotação:

¹ SPARQL Protocol and RDF Query Language

² Semantic Web Rule Language

Coleta de dados. O *dataset* a ser anotado foi obtido por uma *view* criada a partir do modelo da Figura 1.

Dictionary Mapping (DM). O DM (Tabelas 2 e 3) mapeia para ontologias (Sio e KPIOnto) as seguintes características dos KPIs: ResearchField, ImpactFactor e PubQuantity. A Tabela 1 traz o Codebook, que descreve os dados categoriais do *dataset*: DTempo, DFatorImpacto e DCentroPesquisa.

Criação da tabela de Infosheet. A *Infosheet* (Tabela 4) possui as propriedades:

- dct:creator: Responsável pelo preenchimento.
- dct:contributor: Contribuidores na criação do *Infosheet* e execução do processo.
- dct:created: Data de criação.
- dct:description: Propósito do *Infosheet*.
- owl:imports: Endereço da Ontologia.
- schema:keywords: Palavras-chave.
- dct:publisher: Responsável por publicar.
- dct:title: Título do *Infosheet*.

Grafo RDF. Geração dos RDFs para representar os fragmentos de conhecimento a respeito do KPI (do exemplo apresentado). O RDF foi persistido no Virtuoso (ERLING e MIKHAILOV, 2009).

Visualização dos dados. Um *dashboard* genérico conecta-se ao Virtuoso, via ODBC³, e executa consultas SPARQL para ilustrar como os dados, extraídos do grafo, aparecem no *dashboard* (Figura 2).

5. Considerações Finais

O processo especificado neste trabalho visa organizar etapas para anotação com dicionários semânticos de dados para gerar fragmentos de conhecimento em RDF, i.e., conjunto de fatos originados da combinação de dados de diferentes fontes. Um exemplo usando um modelo dimensional para geração de KPI na área de publicação de pesquisa ilustrou o processo constituindo uma validação preliminar do método ("prova de conceito").

No contexto organizacional, a modelagem conceitual adequada dos dados envolve questões complexas de interpretação conceitual e negociação de significados sobre

entidades, relacionamentos e regras de negócios, todas envolvidas no processo de comunicação entre as "partes interessadas". Argumentou-se neste texto sobre como o processo, fundamentado em SDDs, contribui para organização e integração conceitual dos dados oriundos de diferentes nichos da organização, gerando informações que fundamentam a estruturação de conhecimentos sobre diversos indicadores empresariais (KPIs). Isso facilita os alinhamentos dos KPIs a partir de uma abordagem de modelagem de dados ampla, do tipo *top down*, e não apenas *bottom up*.

Futuras pesquisas investigarão as possibilidades da modelagem com SDD, tal como foi apresentada, constituir alternativa superior à modelagem dimensional do tipo "*data mart*" ou "*data warehouse*". Espera-se poder alavancar a flexibilidade de modelos conceituais "livres de esquemas" (*schema free*) para facilitar a geração de KPIs. Isso tornaria a evolução do conhecimento sobre os indicadores de desempenho das organizações mais flexível, incremental e conceitualmente enriquecido, agregando ainda a explicitação da semântica formal, advinda do uso de ontologias representadas em Lógica de Descrições (*Description Logic*).

Referências

- DIAMANTINI, C., POTENA, D. and STORTI, E. SemPI: A Semantic Framework for the Collaborative Construction and Maintenance of a Shared Dictionary of Performance Indicators. *Future Generation Computer Systems (FGCS)*, vol. 54, pages 352-365, Elsevier, 2016.
- ERLING, Orri; MIKHAILOV, Ivan. RDF Support in the Virtuoso DBMS. In: *Networked Knowledge-Networked Media*. Springer, Berlin, Heidelberg, 2009. p. 7-24.
- KIMBALL, Ralph; ROSS, Margy. *The data warehouse toolkit: The definitive guide to dimensional modeling*. John Wiley & Sons, 2013.
- KOLAR, Jana; HARRISON, Andrew e GLIKSOHN, Florian. Key performance indicators of Research Infrastructures. Disponível em: <https://www.cericro.eu/2018/08/30/key-performance->

³Open Database Connectivity

- indicators-of-research-infrastructures/. 30 de Ago de 2018.
- KOURTESIS, Dimitrios; ALVAREZ-RODRÍGUEZ, Jose María; PARASKAKIS, Iraklis. Semantic-based QoS management in cloud systems: Current status and future challenges. *Future Generation Computer Systems*, v. 32, p. 307-323, 2014.
- KRITIKOS, Kyriakos; PLEXOUSAKIS, Dimitris; WOITSCH, Robert. Towards Semantic KPI Measurement. In: CLOSER. 2017. p. 63-74.
- MEDEIROS, Claudia B. Gestão de Dados Científicos – da coleta à preservação. Disponível em <https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/#.XXZ82ChKjIV>. Acesso em 04 de Set de 2019.
- PARMENTER, David. Key performance indicators: developing, implementing, and using winning KPIs. John Wiley & Sons, 2015.
- RASHID, Sabbir M. et al. The Semantic Data Dictionary Approach to Data Annotation & Integration. In: SemSci@ ISWC. 2017. p. 47-54.
- SEMANTIC DATA DICTIONARY. SDD Specification. Disponível em: <https://github.com/tetherless-world/SemanticDataDictionary>. Acesso em 22 de set de 2019.
- SILVA, Vivian S.; HANDSCHUH, Siegfried; FREITAS, André. Categorization of semantic roles for dictionary definitions. arXiv preprint arXiv:1806.07711, 2018.
- WETZSTEIN, Branimir; MA, Zhilei; LEYMANN, Frank. Towards measuring key performance indicators of semantic business processes. In: International Conference on Business Information Systems. Springer, Berlin, Heidelberg, 2008. p. 227-238.
- WILKINSON, M. D., DUMONTIER, M., AALBERSBERG, I. J., APPLETON, G., Axton, M., BAAK, A., and BOUWMAN, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3.