

# ANOTAÇÃO DE DADOS PARA GERAÇÃO DE INDICADORES DE DESEMPENHO EM ORGANIZAÇÕES

*III Workshop de Informação, Dados e Tecnologia*

Marcello P. Bax e Evaldo O. Silva, PPGGOC/UFMG.



## Conteúdo

1 | 23

- Dicionário Semântico de Dados
- Indicadores de Desempenho
- Ontologia KPIOnto
- Trabalhos correlatos
- Processo de Anotação Semântica
- Considerações Finais

- É a **descrição** de uma **coleção de dados**:
  - Tabelas com informações sobre o conteúdo, descrição e formato de cada variável de dados.
- Útil para a **legibilidade humana**
- **Dificuldade** de processar os **significados** dos dados por meio de aplicações de software:
  - Combinação de dados de várias fontes ou mesclagem de esquema, não são facilmente automatizados.
- Caso seja impossível entender o conjunto de dados, nomes diferentes serão atribuídos a **diferentes interpretações** (KIMBALL e ROSS, 2013)

- Um **SDD** (*Semantic Data Dictionary*) é uma abordagem que utiliza **ontologias** para fornecer **propriedades** e descrição de **relacionamentos** entre objetos a fim representar o **conhecimento** em um domínio de aplicação (RASHID et. al, 2017).
- Um SDD é criado a partir de **documentos** que auxiliam na sua **descrição**, anotação das **colunas de dados** e os **conceitos** correspondentes na ontologia.
- Este trabalho utiliza SDD como padrão para **anotação semântica de metadados** para **indicadores de desempenho** (PARMENTER, 2015).

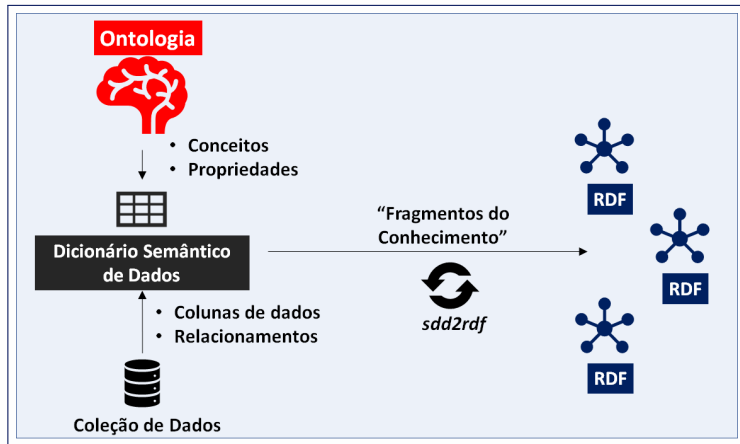


Figure: *Uso do dicionário semântico de dados (RASHID et. al, 2017)...*



Figure: *Geração de KPIs...*

## Indicadores de Desempenho

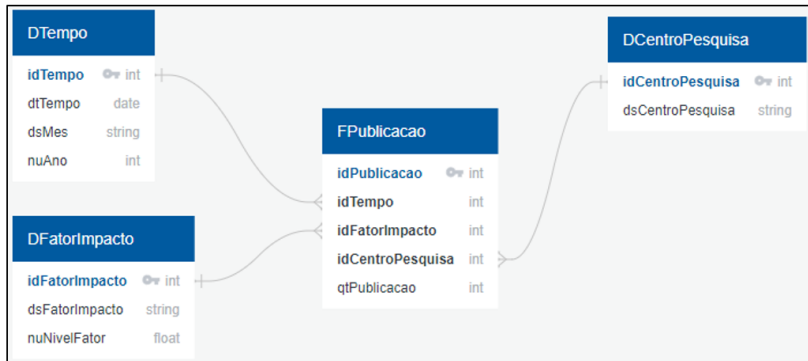


Figure: *Modelo Dimensional...*

## Ontologia KPIOnto

- A **anotação pelo SDD** exige compreender o **domínio** por sua modelagem **conceitual** prévia.
- A **KPIOnto** é a **ontologia** utilizada neste trabalho para anotação e alinhamento conceitual de diferentes profissionais sobre os **KPIs**.
- Com base na KPIOnto, os dados a serem anotados devem ser selecionados para que sejam relacionados aos conceitos existentes na ontologia.



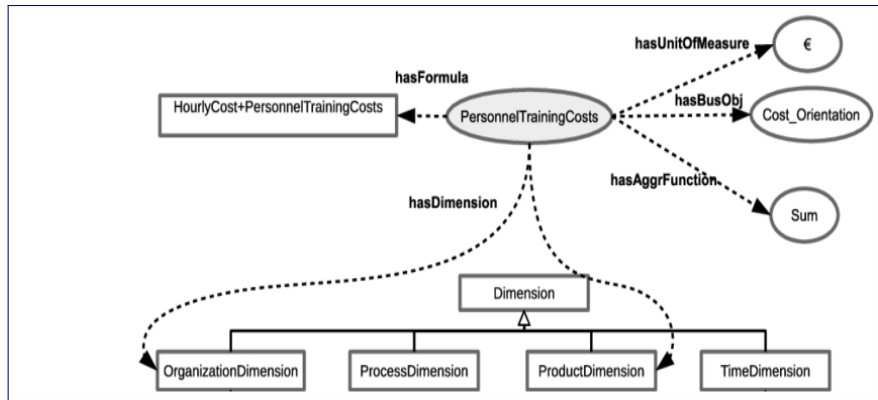


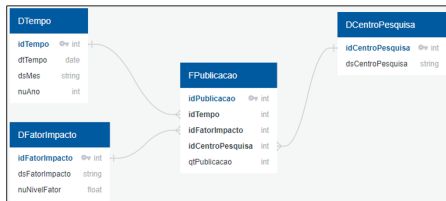
Figure: Ontologia KPIOnto...

- Uso de **ontologias** no contexto de **BPaaS** (*Business Process as a Service*) para **descrever** os dados de **KPIs** gerados pela execução dos processos e serviços de integração (KRITIKOS, 2017).
- **KPIs** descritos por analistas de negócios a partir de **anotações semânticas**, visando a recuperação de dados para criação de painéis de **monitoramento em tempo real** (WETZSTEIN, MA e LEYMAN, 2008).
- **Estrutura semântica** para gerenciamento de **QoS** (*Quality of Service*) (KOURTESIS e ALVAREZ-RODRIGUES, 2014).
- Estrutura semântica para definição de um **dicionário de dados**. Os autores utilizam procedimentos para **extração de informações e modelos semânticos** (SILVA, et. al. 2018).



### 1. Coleta de Dados

O conjunto de dados (*dataset*) a ser anotado deve ser obtido através do modelo de dados (ou modelo dimensional), por exemplo, a partir da elaboração de uma *view* de dados.



ID	Data	Mes	Ano	Centro de Pesquisa	Descrição do fator de impacto
1	01/01/2000	JANEIRO	2000	Ciência da Informação	Impacto entre 2 e 4
2	01/01/2000	JANEIRO	2000	Ciência da Computação	Impacto entre 5 e 7
3	01/01/2000	JANEIRO	2000	Linguística	Impacto entre 5 e 7
4	01/01/2000	JANEIRO	2000	Ciência da Informação	Impacto entre 5 e 7
5	01/01/2000	JANEIRO	2000	Ciência da Computação	Impacto entre 7 e 10
6	01/01/2000	JANEIRO	2000	Ciência da Informação	Impacto entre 7 e 10
7	01/01/2000	JANEIRO	2000	Linguística	Impacto entre 7 e 10
8	01/01/2000	JANEIRO	2000	Linguística	Impacto entre 2 e 4
9	01/01/2000	JANEIRO	2000	Ciência da Informação	Impacto entre 2 e 4
10	02/01/2000	JANEIRO	2000	Ciência da Computação	Impacto entre 5 e 7
11	02/01/2000	JANEIRO	2000	Linguística	Impacto entre 5 e 7
12	02/01/2000	JANEIRO	2000	Ciência da Informação	Impacto entre 5 e 7
13	02/01/2000	JANEIRO	2000	Ciência da Computação	Impacto entre 7 e 10
14	02/01/2000	JANEIRO	2000	Ciência da Informação	Impacto entre 7 e 10

### 2. Criação do Dictionary Mapping (DM)

O DM mapeia para a ontologia KPIOnto as seguintes características dos KPIs: ResearchField e ImpactFactor. A tabela abaixo traz o *Codebook*, para descrever os dados categoriais do *dataset*: DFatorImpacto e DCentroPesquisa.

Column	Code Label	Class
DCentroPesquisa	1 CIÊNCIA DA INFORMAÇÃO	kpionto:researchField
DCentroPesquisa	2 LINGUÍSTICA	kpionto:researchField
...		
DFatorImpacto	40 IMPACTO ENTRE 2 E 4	kpionto:ImpactFactor
DFatorImpacto	42 IMPACTO ENTRE 7 E 10	kpionto:ImpactFactor

Column	Attribute Label	AttributeOf
Id.Kpi	sio:Identifier Identificador do KPI	??KpiPublication
ResearchField	kpiOnto:hasDimension Centro de Pesquisa	??KpiPublication
ImpactFactor	kpiOnto:hasDimension Nível do Fator de Impacto	??KpiPublication
PubQuantity	kpiOnto:hasAggFunction Quantidade de Publicação	??KpiPublication

Table: *Especificação do DM para dados explícitos com o uso da Ontologia SIO (Semanticscience Integrated Ontology)*

Column	Entity Role	InRelationTo
??KpiPublication	kpiOnto:Indicator -	??researchInstitute
??researchInstitute	sio:institute -	-

Table: *Especificação do DM para dados implícitos*

### 3.Criação da Tabela de Infosheet

A Infosheet possui as propriedades definidas abaixo:

- 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.



---

Atributo	Valor
dct:creator	Marcello P. Bax e Evaldo de Oliveira da Silva
dct:contributor	Marcello P. Bax
dct:created	19/09/2019
dct:description	Anotação semântica do dicionário de dados de KPI
owl:imports	<a href="http://semanticscience.org/ontology/sio-subset-labels.owl">http://semanticscience.org/ontology/sio-subset-labels.owl</a>
schema:keywords	KPI, Publicação
dct:publisher	Evaldo de Oliveira da Silva
dct:title	Geração de KPIs e anotação semântica de modelos dimensionais

---

Table: *Especificação da tabela de Infosheet*

### 4. Geração dos RDFs

Geração dos RDFs para representar os fragmentos de conhecimento a respeito do KPI (do exemplo apresentado). O RDF pode ser persistido em banco de dados semânticos.

```
...  
<#ID_KPI:1> <kpionto/hasdimension/mes> "JANEIRO" .  
<#ID_KPI:1> <kpionto/hasdimension/ano>2000 .  
<#ID_KPI:1> <kpionto/hasdimension/ ResearchField>"CIENCIA INF."  
<#ID_KPI:1> <kpionto/hasdimension/fatorimpacto>"IMPACTO ENTRE 2  
<#ID_KPI:1> <kpionto/hasdimension/nivelfatorimpacto>4 .  
<#ID_KPI:1> <kpionto/hasggrfunction/contagempublic> 6 .  
...
```

### 5. Visualização dos Dados

Como estudo de caso neste trabalho, um *dashboard* genérico conecta-se ao Virtuoso, via ODBC , e executa consultas SPARQL para ilustrar como os dados, extraídos do grafo, aparecem no *dashboard*.

```
SELECT DISTINCT
  kpi_public.s as id_kpi,
  kpi_public.p as dimension,
  kpi_public.o as contagempublic
FROM
  (SPARQL SELECT ?s ?p ?o
  FROM <http://kpi_publication>
  WHERE { ?s ?p ?o FILTER (?o IN (2000, "JANEIRO", "FEVEREIRO"
    , "MARÇO")) }
  ) as kpi_publication
```

# ANOTAÇÃO DE DADOS PARA GERAÇÃO DE INDICADORES DE DESEMPENHO EM ORGANIZAÇÕES

## Processo de Anotação Semântica

19 | 23

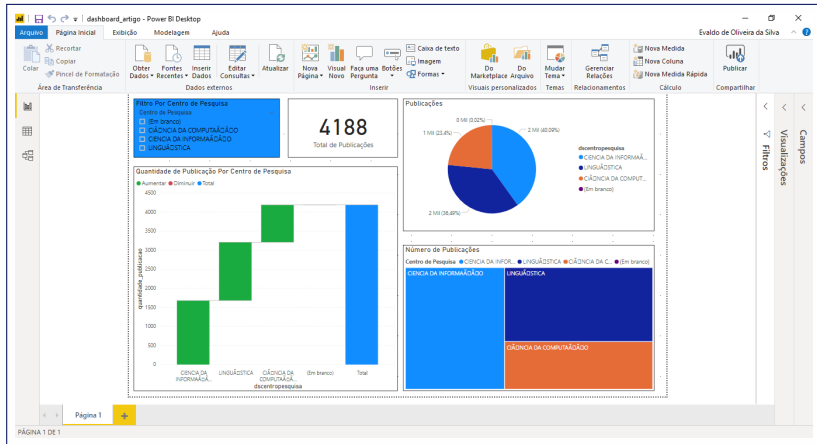


Figure: Ontologia KPIOnto...

## Considerações Finais

20 | 23

- **Processo** apresentado visa organizar etapas para **anotação** com **dicionários semânticos de dados**.
- **SDDs** contribuem para **organização conceitual** dos dados, visando estruturar o conhecimento sobre KPIs.
- Também os SDDs podem permitir alinhamentos dos KPIs a partir de uma abordagem de **modelagem de dados** ampla, do tipo *top down*, e não apenas *bottom up*.
- Como trabalhos futuros pretende-se **investigar alternativas de modelagem** de *data mart* ou *data warehouse* com SDD.
- Espera-se poder alavancar a flexibilidade de modelos conceituais "**livres de esquemas**" (*schema free*) para **facilitar a geração** de KPIs.

- KIMBALL, Ralph; ROSS, Margy. The data warehouse toolkit: The definitive guide to dimensional modeling. John Wiley Sons, 2013.
- 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.
- 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.

## Referências

22 | 23

- 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.

Obrigado!

Evaldo de Oliveira da Silva

PPGGOC/UFMG