24ºENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação

EXPRESSIVIDADE SEMÂNTICA DO *BIG DATA*: EXTRAINDO INFORMAÇÕES DE DADOS

24 ENANCIB, GT-2, nov. 2024, UFES, Vitória, ES

Grupo de Pesquisa ROCAD* – Representação e Organização do Conhecimento em Ambientes Digitais, + Prof. Linair Campos,
httpo://dgp.cnpq.br/dgp/espelhogrupo/793271

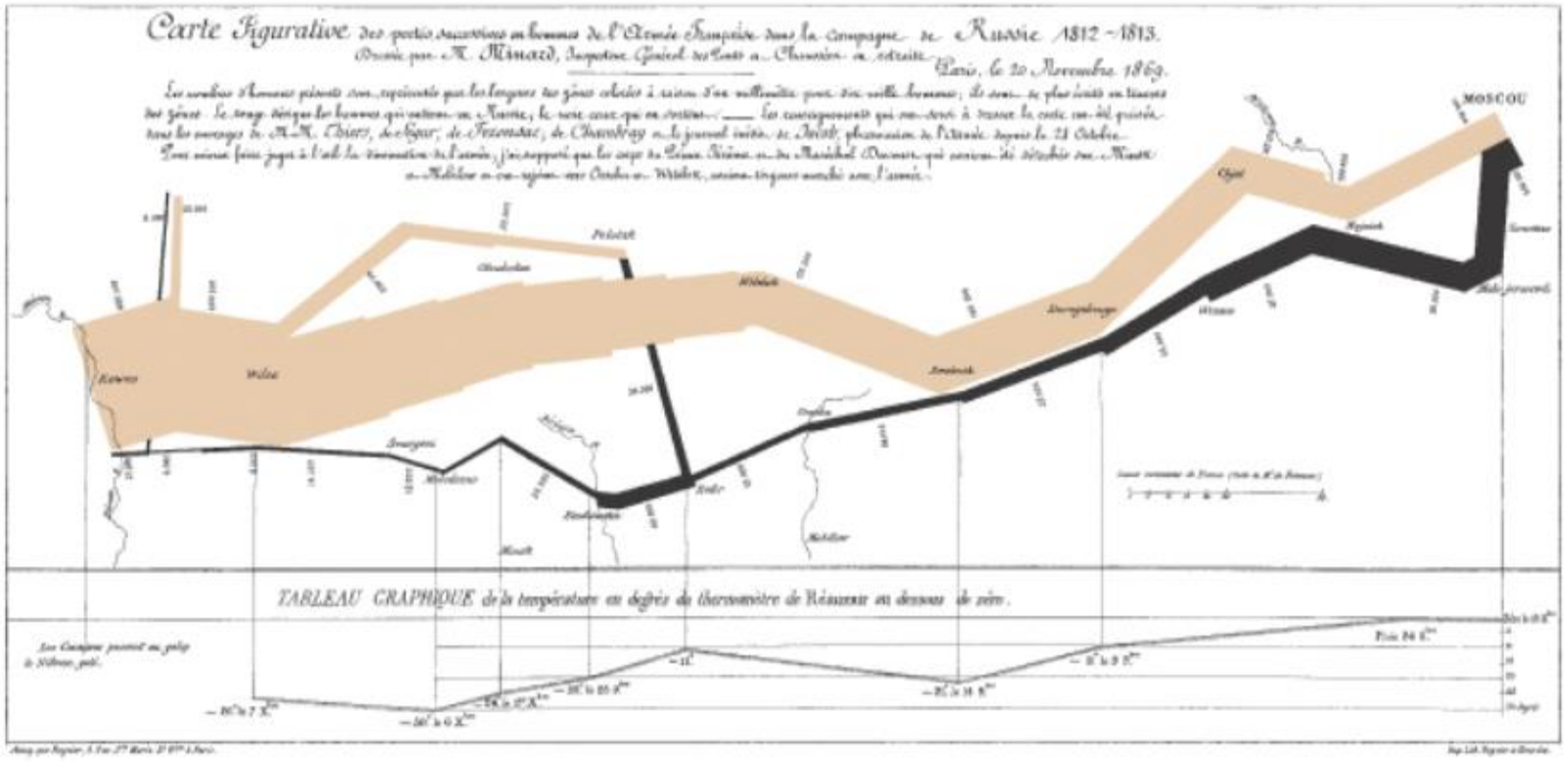
Carlos H. Marcondes, UFMG, UFF
ch_marcondes@id.uff.br
http://profmarcondes.ong.br/

***MARTINS, S; RAMOS JR, M; PEREIRA, D; MARCONDES, C.**

EXPRESSIVIDADE SEMÂNTICA DO *BIG DATA*: EXTRAINDO INFORMAÇÕES DE DADOS

Questões: o que são dados? Qual o potencial semântico dos dados? Como é gerada a semântica a partir de dados digitais?
Hipótese: dados se organizam em diferentes níveis, como campos, registros, tabelas, modelos conceituais, ontologias; à medida que se tornam mais complexos, em maior volume e são processados, estes agregados de dados nos diferentes níveis tornam-se mais expressivos e podem gerar semântica.
Objetivo: discutir teoricamente esta hipótese, propor uma conceituação da questão e mostrar indícios de sua factibilidade, propor uma aproximação CI e OC ao *Big Data* e a CD.

Sumário



Napoleon's March to Moscow The War of 1812

Charles Joseph Minard

The classic of Charles Joseph Minard (1781–1870), the French engineer, shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its heroic eloquence, this combination of data map and time-series, drawn in 1869, portrays the devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Niemen River, the thick band shows the size of the army (422,000 men) as it invaded Russia in June 1812. The width of the band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally struggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn.

Edward R. Tufte, The Visual Display of Quantitative Information Graphics Press, Box 430, Cheshire, Connecticut 06410

Nesta imagem (Fig.7), que Tufte considera um dos diagramas científicos mais "eficazes", entendemos a origem desta aposta dupla que faz o cientista ganhar sempre que parece ter pedido contacto direto com o mundo. Marey, o grande fisiologista, pode sobrepor o mapa da Rússia, a medição das temperaturas, a rota da Grande Armée, a data das suas implantações e, mais tragicamente, o número de soldados sobreviventes de cada bivaque! (LATOUR 2000, p. 30)

1. Introdução, Problema

"*While data volume proliferates, the knowledge it creates has not kept pace*", Cognizant Newsletter (2011),

As atividades humanas são cada vez mais mediadas pelas tecnologias de informação, o chamado processo de datificação (Mejias; Couldry, 2019) da sociedade contemporânea.

Grande quantidade de dados digitais, chamados de Big Data, que se tornaram fundamentais nas organizações por terem alto potencial semântico.

Como recurso semântico, o *Big Data* só atinge seu potencial máximo quando processado por tecnologias da informação.

1. Introdução, Questões de Pesquisa

- **O que são dados? Como a semântica (para humanos e máquinas) emerge de dados digitais?**

“The “strength” of the semantic, in these cases, is linked to ”semantic expressivity,” associated with the tractability of the KOS for different kinds of formalism” (SOUZA et al. 2012, p. 183).

“Representational power, Semantic Expressiveness, Intelligibility” (SOUZA et al. 2012, p. 188).

“Expressivity or semantic expressivity, is a notion used to nominate how accurate a knowledge representation express a phenomena of reality”. (MARTINS et al 2024, p. 6)

- **Como a EXPRESSIVIDADE SEMÂNTICA aumenta no contexto do Big Data?**

✓ **Dados são organizados em diferentes níveis, desde os dados mais simples até os agregados de dados mais complexos (Foskett, 1961; Barreto, 2008), formando conjuntos de dados ou sistemas de dados, como campos, registros, tabelas, modelos conceituais, ontologias. À medida que se tornam mais complexos e volumosos, esses agregados de dados potencialmente se tornam mais expressivos e podem gerar semântica, informações, insights para humanos e máquinas.**

1. Introdução, definições, pressupostos

- **EXPRESSIVIDADE SEMÂNTICA (ainda uma noção) – Precisão na representação da realidade, resultando em informação potencial, potencial informativo da SAÍDA de um processamento de conjunto de dados, imediatamente oferecida a um usuário final, permitindo-lhe ação ou tomada de decisão imediata.**

- **SAÍDA do processamento de um conjunto de dados (um agregado de dados) – "saída" de um sistema de informação imediatamente oferecida a um usuário final.**

1. Introdução ..., Referências, Bases Conceituais

- Web Semântica (BERNERS-LEE, HENDLER, LASSILA 2001)
- Teoria dos Níveis Integrativos (HARTMANN 1952), (FEIBLEMAN 1954), (GNOLI, 2018)
- Modos de existir SNAP e SPAN (GRENON, SMITH 2004)
- Big Data and Knowledge Organization (IBEKWE-SANJUAN, BOWKER 2017)
- Data, Big Data and semantics (HJØRLAND , 2018)
- The semantic of the Semantic Web (SHETH, RAMAKRISHNAN, THOMAS 2005)
- A taxonomy of KOS according to their semantic expressivity (SOUZA, TUDHOPE, Douglas, ALMEIDA 2012)

O que são dados?

Várias áreas científicas também têm feito esforços para entender, conceituar e instrumentar o Big Data, como Ciência da Computação, Ciências da Saúde e Organização do Conhecimento (Shet, 2020) (Huang et al. 2015).

Big Data hoje X explosão de informações, o fenômeno que deu origem à CI e OC na década de 1960

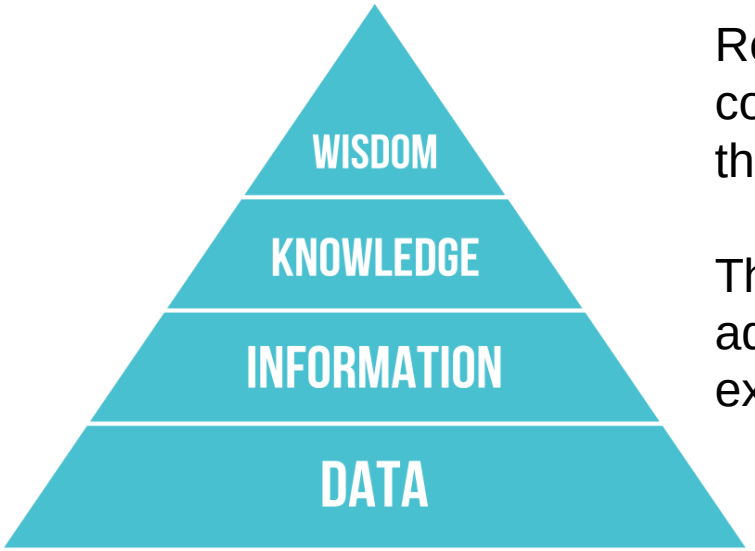
“Big Data is the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value” Mauro, Greco e Grimaldi (2016, p. 126).

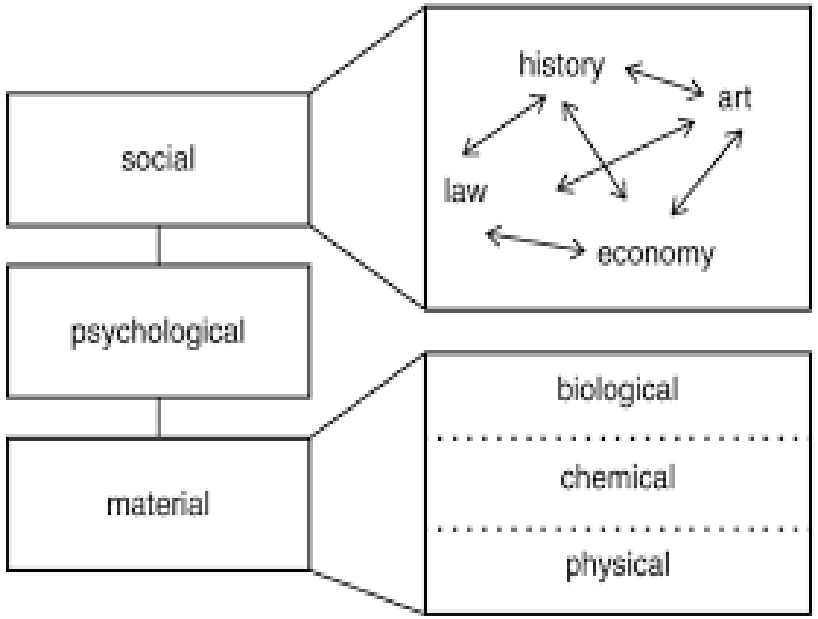
“Data are social artefacts” (Ibekwe-Sanjuan, Bowker 2017, p. 195)

“Dados são instanciações concretas de representações simbólicas de proposições descritivas, informadas por observação empírica, sobre as propriedades quantitativas e qualitativas de fenômenos do mundo real”; “Dados são sempre produzidos para alguns propósitos e perspectivas”. (Hjørland, 2018, s.p.).
“Um dado ou item de dados como um triplo <e, a, v>, onde e é uma entidade em um modelo conceitual, a é um atributo da entidade e, e v é um valor do domínio do atributo a. Um dado afirma que a entidade e tem valor v para o atributo a” (Hjørland, 2018, s.p.).

Dados são representações de entidades ou fenômenos, entidades semióticas

Fundamentos ... -> Hierarquias, Teoria dos Níveis Integrativos



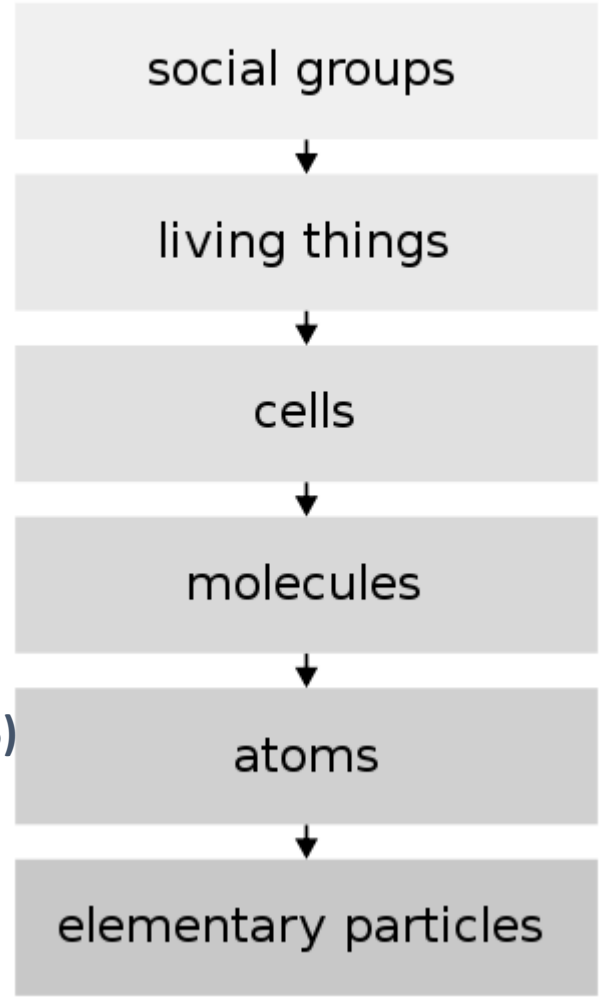


Reality is organized as crescente levels of complexity, as Phisycal level, the Biological level, the Psycologicallevel, the Cultural level

The levels overlap each other. Each higher level adds an emergent property (or quality) that did not exist at the lower level

1. Structures and Form
2. Mater and Energy
3. Cosmos and Earth
4. Life (biological systems)
5. Human beings
6. Societies
7. Material artefacts
8. Intellectual artefacts
9. Spiritual artefacts

(DAHLBERG, 1995)



KANISTO, Tony (Quora, 2018)
httpo://philosophicallp.quora.com/Emergence-vs-Supervenience

Fundamentos ... -> Teoria dos Níveis Integrativos

Our list of the strata can be compared with the terminology of philosophers dealing with levels:

	Lloyd Morgan	RW Sellars	Hartmann	Poli	Popper
form			ideal being		
matter	matter	inanimate	material	material	world 1
life	life	animate	organic		
mind	mind	mind	psychic	psychological	world 2
society heritage		society	personal spirit objective spirit objectivated spirit	social	world 3

DADOS

The material and living strata can be decomposed quite easily into their layers: eg for Lloyd Morgan matter can be either physical or chemical, while mind can be conscious or reflective; for Hartmann, the spiritual stratum includes personal, objective (social), and objectivated (cultural) spirit. Modern science often acknowledges matter as including subatomic particles, atoms, molecules, celestial objects; and life as including cells, organisms, and biological populations. Layers within higher strata are less immediately identified, but levelled structures are often cited, eg families, clans, cities, nations, and the global community can be listed in the social stratum.

(GNOLI 2008, http://www.iskoi.org/ilc/book/strata.php)

Fundamentos ...

CONTINUANDOS E OCORENTES

GRENON, Pierre; SMITH, Barry. SNAP and SPAN: Towards dynamic spatial ontology. *Spatial cognition and computation*, v. 4, n. 1, p. 69-104, 2004

Temporal Modes of Being

The central dichotomy among the perspectives represented in BFO concerns the modes of existence in time of the entities populating the world. BFO endorses first of all a view according to which there are entities that have continuous existence and a capacity to endure (persist self-identically) through time even while undergoing different sorts of changes. We will henceforth use the terms 'continuant' and 'endurant' interchangeably for such entities. These entities come in several kinds. Examples are: you, the planet Earth, a piece of rock; but

also: your suntan, a rabbit-hole, Leeds. All of these entities exist in full in any instant of time at which they exist at all and they preserve their identity over time through a variety of different sorts of changes. You are the same person today as you were yesterday.

In addition, however, BFO endorses a view according to which the world contains *occurrents*, more familiarly referred to as processes, events, activities, changes. Occurrents include: your smiling, her walking, the landing of an aircraft, the passage of a rainstorm over a forest, the rotting of fallen leaves. These entities are four-dimensional. They occur in time and they unfold themselves through a period of time.

UM EXEMPLO...

World Health Organization

Global COVID-19 Clinical Platform
RAPID CORE CASE REPORT FORM (CRF)

INTRODUCTION

In response to the COVID-19 pandemic, the World Health Organization (WHO) has launched a global COVID-19 anonymized clinical data platform (the "COVID-19 Data Platform") to enable State Parties to the International Health Regulations (IHR) (2005) to share with WHO anonymized clinical data related to patients with suspected or confirmed infections with SARS-CoV-2 (collectively "anonymized COVID-19 data"). The anonymized COVID-19 data received by WHO will remain the property of the contributing Entity and will be used by WHO for purposes of verification, assessment and assistance pursuant to the IHR (2005), including to inform the public health and clinical operation response in connection with the COVID-19 outbreak. To help achieve these objectives, WHO has established an independent Clinical Advisory Group to advise WHO on global reporting and analysis of the anonymized clinical COVID-19 data. State Parties and other entities are invited to contact WHO to obtain more information about how to contribute anonymized clinical COVID-19 data to the WHO Data Platform. To preserve the security and confidentiality of the anonymized COVID-19 data, State Parties and other entities are respectfully requested to take all necessary measures to protect their respective log-in credentials and passwords to the COVID-19 Data Platform.

The anonymized COVID-19 data will be stored in the WHO COVID-19 Data Platform, which is a secured, access-limited, password protected electronic platform. WHO will (i) protect the confidentiality and prevent the unauthorized disclosure of the anonymized COVID-19 data; (ii) implement and maintain appropriate technical and organizational security measures to protect the security of the anonymized COVID-19 data and the COVID-19 Data Platform. In accordance with Article 11(4) of the IHR (2005), WHO will not make the anonymized COVID-19 data generally available to other State Parties or entities until such time as any of the conditions set forth in paragraph 2 of Article 11 are first met, and following consultation with affected countries/entities. Pursuant to that same Article 11, WHO will not make the anonymized COVID-19 data available to the public, unless and until the anonymized COVID-19 data have already been made available to State Parties, and provided that other information about the COVID-19 epidemic has already become publicly available and there is a need for the dissemination of authoritative and independent information. To contribute data to the WHO COVID-19 Data Platform or to receive more information, please contact: COVID_ClinPlatform@who.int

DESIGN OF THIS CASE REPORT FORM (CRF)

The Rapid Core CRF is designed to collect data obtained through examination, interview and review of hospital notes. Data may be collected prospectively or retrospectively. The data collection period is defined as the period from hospital admission to discharge, transfer, death, or continued hospitalization without possibility of continued data collection.

This CRF has 3 modules:

Module 1: to be completed on the first day of admission to the health centre.

Module 2: to be completed daily during hospital stay for as many days as resources allow. Continue to follow-up patients who transfer between wards.

Module 3: to be completed at discharge or death.

GENERAL GUIDANCE

- Participant identification numbers consist of a site code and a participant number. You can register on the data management system by contacting COVID_ClinPlatform@who.int, and our data management team will contact you with instructions for data entry and will assign you a 5-digit site code and participant number.
- Please contact us at COVID_ClinPlatform@who.int for any information.

PARTICIPANT ID [__][__][__][__][__][__]--[__][__][__][__] Module 1 – page 1

Complete on hospital admission (within 24 hrs from hospital admission)

Country _____

Date of enrolment [D][D]/[M][M]/[2][0][Y][Y]

1a. CLINICAL INCLUSION CRITERIA

One or more of these during this illness	A history of self-reported feverishness or measured fever of ≥38°C	☐Yes	☐No
	Cough	☐Yes	☐No
	Dyspnoea (shortness of breath) OR Tachypnoea*	☐Yes	☐No
	Clinical suspicion despite not meeting criteria above	☐Yes	☐No

** Respiratory rate ≥ 50 breaths/min for < 1 year; ≥ 40 for 1–4 years; ≥ 30 for 5–12 years; ≥ 20 for ≥ 13 years*

1b. DEMOGRAPHICS

Sex at birth ☐Male ☐Female ☐Not specified **Date of birth** [D][D]/[M][M]/[Y][Y][Y][Y]
If date of birth is unknown, record: **Age** [___][___] **years** OR [_][_] **months** OR [_][_] **days**
Health care worker? ☐Yes ☐No ☐Unknown **Laboratory worker?** ☐Yes ☐No ☐Unknown
Pregnant?* ☐Yes ☐No ☐Unknown ☐N/A **If yes: Gestational weeks assessment** [___][_] weeks
If currently pregnant or recently pregnant (delivery within 21 days of symptom onset), complete Pregnancy CRF

1c. DATE OF ONSET AND ADMISSION VITAL SIGNS *(first available data at presentation/admission)*

Symptom onset (date of first/earliest symptom) [D][D]/[M][M]/[2][0][Y][Y]
Admission date at this facility [D][D]/[M][M]/[2][0][Y][Y]
Temperature [_][_].[_] °C **Heart rate** [___][___][___] beats/min
Respiratory rate [_][___] breaths/min
BP [_][_][_] (systolic) [___][___] (diastolic) mmHg **Severe dehydration** ☐Yes ☐No ☐Unknown
Sternal capillary refill time > 2 seconds ☐Yes ☐No ☐Unknown
Oxygen saturation: [__][__]% on ☐Room air ☐Oxygen therapy ☐Unknown **A V P U** (circle one)
Glasgow Coma Score (GCS/15) [_][_] **Malnutrition** ☐Yes ☐No ☐Unknown
Mid-upper arm circumference [_][_][___] mm **Height** [___][___][___] cm **Weight** [_][_][___] kg

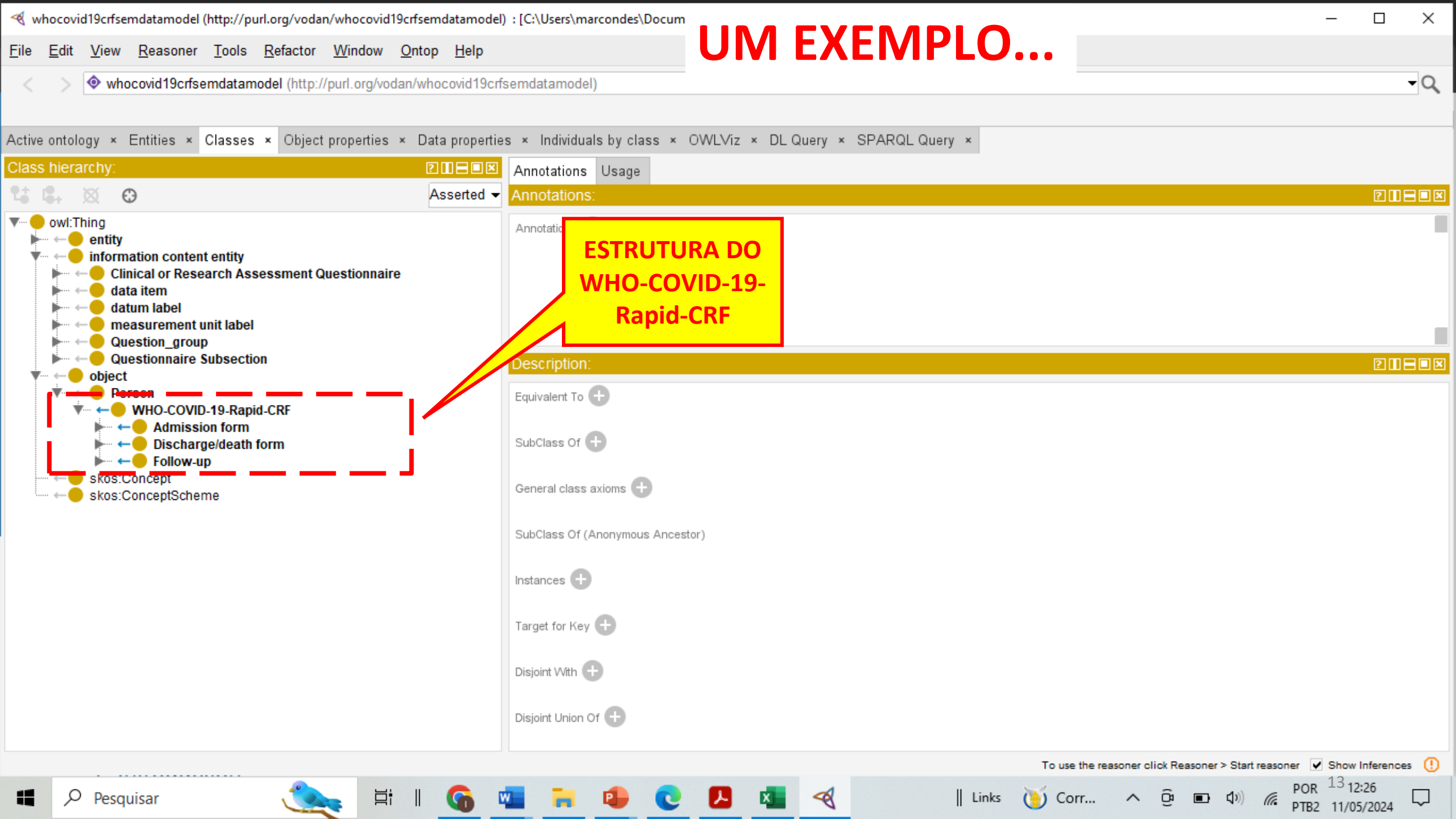
1d. CO-MORBIDITIES *(existing at admission) (Unk = Unknown)*

Chronic cardiac disease *(not hypertension)*	☐Yes	☐No	☐Unk	Diabetes	☐Yes	☐No	☐Unk
Hypertension	☐Yes	☐No	☐Unk	Current smoking	☐Yes	☐No	☐Unk
Chronic pulmonary disease	☐Yes	☐No	☐Unk	Tuberculosis (*active*)	☐Yes	☐No	☐Unk
Asthma	☐Yes	☐No	☐Unk	Tuberculosis (*previous*)	☐Yes	☐No	☐Unk
Chronic kidney disease	☐Yes	☐No	☐Unk	Asplenia	☐Yes	☐No	☐Unk
Chronic liver disease	☐Yes	☐No	☐Unk	Malignant neoplasm	☐Yes	☐No	☐Unk
Chronic neurological disorder	☐Yes	☐No	☐Unk	Other	☐Yes	☐No	☐Unk
				If yes, specify: _____			
HIV	☐Yes (on ART)	☐Yes (not on ART)	☐No ☐Unknown	ART regimen_____			

1e. PRE-ADMISSION AND CHRONIC MEDICATION Were any of the following taken within 14 days of admission

Angiotensin converting enzyme inhibitors (ACE inhibitors)? ☐Yes ☐No ☐Unknown
Angiotensin II receptor blockers (ARBs)? ☐Yes ☐No ☐Unknown
Non-steroidal anti-inflammatory (NSAID)? ☐Yes ☐No ☐Unknown
Antivirals? ☐Chloroquine/hydroxychloroquine ☐Azithromycin ☐Lopinavir/Ritonavir ☐Other: _____

https://iris.who.int/rest/bitstreams/1287200/retrieve


whocovid19crfsemdatamodel (http://purl.org/vodan/whocovid19crfsemdatamodel) : [C:\Users\marcondes\Docum
UM EXEMPLO...
File Edit View Reasoner Tools Refactor Window Ontop Help
whocovid19crfsemdatamodel (http://purl.org/vodan/whocovid19crfsemdatamodel)
Active ontology × Entities × Classes × Object properties × Data properties × Individuals by class × OWLViz × DL Query × SPARQL Query ×
Class hierarchy:
Asserted
owl:Thing
entity
information content entity
Clinical or Research Assessment Questionnaire
data item
datum label
measurement unit label
Question_group
Questionnaire Subsection
object
Person
WHO-COVID-19-Rapid-CRF
Admission form
Discharge/death form
Follow-up
skos:Concept
skos:ConceptScheme
ESTRUTURA DO WHO-COVID-19-Rapid-CRF
Annotations Usage
Annotations:
Description:
Equivalent To
SubClass Of
General class axioms
SubClass Of (Anonymous Ancestor)
Instances
Target for Key
Disjoint With
Disjoint Union Of
To use the reasoner click Reasoner > Start reasoner
Show Inferences
Pesquisar
Links
Corr...
POR
PTB2
12:26
11/05/2024

UM EXEMPLO...

World Health Organization PARTICIPANT ID [__][__][__][__][__][__] -- [__][__][__][__] Module 1 – page

MODULE 1. Complete on hospital admission (within 24 hrs from hospital admission)

Facility name _____ **Country** _____

Date of enrolment [_D_][_D_]/[_M_][_M_]/[_2_][_0_][_Y_][_Y_]

1a. CLINICAL INCLUSION CRITERIA

One or more of these during this illness	A history of self-reported feverishness or measured fever of ≥38°C	☐Yes	☐No
	Cough	☐Yes	☐No
	Dyspnoea (shortness of breath) OR Tachypnoea*	☐Yes	☐No
	Clinical suspicion despite not meeting criteria above	☐Yes	☐No

** Respiratory rate ≥ 50 breaths/min for < 1 year; ≥ 40 for 1–4 years; ≥ 30 for 5–12 years; ≥ 20 for ≥ 13 years*

1b. DEMOGRAPHICS

Sex at birth ☐Male ☐Female ☐Not specified **Date of birth** [_D_][_D_]/[_M_][_M_]/[_Y_][_Y_][_Y_][_Y_]

If date of birth is unknown, record: **Age** [___][___][___] **years** OR [_][_] **months** OR [_][_] **days**

Health care worker? ☐Yes ☐No ☐Unknown **Laboratory worker?** ☐Yes ☐No ☐Unknown

Pregnant?* ☐Yes ☐No ☐Unknown ☐N/A **If yes: Gestational weeks assessment** [___][_]weeks

If currently pregnant or recently pregnant (delivery within 21 days of symptom onset), complete Pregnancy CRF

1c. DATE OF ONSET AND ADMISSION VITAL SIGNS *(first available data at presentation/admission)*

Symptom onset (date of first/earliest symptom) [_D_][_D_]/[_M_][_M_]/[_2_][_0_][_Y_][_Y_]

Admission date at this facility [_D_][_D_]/[_M_][_M_]/[_2_][_0_][_Y_][_Y_]

Temperature [_][___].[_]°C **Heart rate** [___][___][_]beats/min

Respiratory rate [_][___]breaths/min

BP [_] [_] [_](systolic) [___][_][_](diastolic)mmHg **Severe dehydration** ☐Yes ☐No ☐Unknown

Sternal capillary refill time > 2

Oxygen saturation: [__][ ][__]

Glasgow Coma Score (GCS/15

Mid-upper arm circumference

1d. CO-MORBIDITIES *(existing*

Chronic cardiac disease *(not hypertension)*

Hypertension

EXPRESSIVIDADE SEMANTICA

ID_PACIENTE	ID_aTENDIMENTO	DT_INTERNACAO	UNID_ATENDIMENT	DE_EXAME	DT_OBITO
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP A	CULTURA AERÃ“BIA	21/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP B	URÃ‰IA	07/06/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP C	URINA TIPO 1	25/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP D	URINA TIPO 1	21/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP A	URINA TIPO 1	20/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP B	URINA TIPO 1	23/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP C	URINA TIPO 1	17/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP D	URINA TIPO 1	19/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP A	URINA TIPO 1	20/05/2020
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP B	URINA TIPO 1	27/05/2020

NOSSO EXEMPLO:

EXAMES

ID_PACIENTE	ID_ATENDIMENTO	DT_COLET	DE_ORI	DE_EXAME	DE_ANALITO	DE_RESULTADO
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	CULTURA AERÃ“BIA	CULTURA AERÃ“BIA	Negativa
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	URÃ‰IA	URÃ‰IA	10
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	URINA TIPO 1	pH :	7.0
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	URINA TIPO 1	Urobilinogenio :	0.2
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf					
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf					
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf					
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf					
004056afeb7b5441855846349123f686	0ae989c74676b6b5b8bf					

DICIONÁRIO

Nome	Descrição	Formato	Conteúdo, domínio e restrições
ID_PACIENTE	Identificação única do paciente (correlaciona com o ID_PACIENTE de todos os arquivos onde aparece (por exemplo, EXAMES e DESFECHOS)	32 caracteres alfanuméricos	string, anonimizado conforme critério da instituição
ID_ATENDIMENTO	Identificação única do atendimento. Correlaciona com o ID_ATENDIMENTO de todas as tabelas onde aparece (por exemplo, DESFECHO)	Alfanumérico	string, anonimizado conforme critério da instituição
DT_COLETA	Data em que o material foi coletado do paciente	Data (DD/MM/AAAA)	DD = Dia / MM = Mês / AAAA = Ano
DE_ORIGEM	Local de coleta daquele exame.	4 caracteres alfanuméricos	LAB – Exame realizado por paciente em uma unidade de atendimento laboratorial HOSP – Exame realizado por paciente dentro de uma Unidade Hospitalar
DE_EXAME	Descrição do exame realizado	alfanumérico	nome do exame + nome do material
DE_ANALITO	Descrição do analito	2 caracteres alfanuméricos	string com nome do analito
DE_RESULTADO	Resultado do exame, associado ao DE_ANALITO	alfanumérico	Se DE_ANALITO exige valor numérico, NNNN se inteiro ou NNNN,NNN se casas decimais Se DE_ANALITO exige qualitativo, String com domínio restrito Se DE_ANALITO por observação microscópica, String conteúdo livre
CD_UNIDADE	Unidade de Medida utilizada na Metodologia do laboratório específico para analisar o exame	Alfanumérico	string restrita
DE_VALOR_REFERENCIA	Faixa de valores de referência	Alfanumérico	String - Resultado ou faixa de resultados em que é considerado normal para este analito, na população

PACIENTES

	A	B	C	D	E	F	G
1	Column1	Column2	Column3	Column4	Column5	Column6	Column7
2	ID_PACIENTE	IC_SEXO	AA_NASCIMENTO	CD_PAIS	CD_UF	CD_MUNICIPIO	CD_CEPREDUZIDO
3	9698838a8fa8a01ff[illegible]5ed5c71e8e17a3	F	1962	BR	SP	SAO PAULO	CCCC
4	d9fec23b3820f93a9[illegible]1841d569db8cb5	F	1974	BR	SP	MMMM	CCCC
5	ee507ba3a9959fdf31bca52852fd5715	F	1962	BR	SP	MMMM	CCCC
6	51590e8c53f4e8e332c05d7e6cee35c7	F	1960	BR	SP	SAO PAULO	CCCC
7	13699f0f7714fdaba277c5e360c6869c	M	1967	BR	SP	MMMM	CCCC
8	9bd839fe149857321685b1e1d8a55cbd	F	1948	BR	SP	SAO PAULO	CCCC
9	b723165a04c8e28fe2b4dcff8dc8dab3	F	1963	BR	SP	MMMM	CCCC

Ministério da Saúde - COVID-19 NO BRASIL

A DASHBOARD

Online, real time data

Informações | Exportar dados

Filtros

Região | UF | Município | Reg.Metropolitana/Interior | Ano | Semana Epidemiológica

Ano: 2024 X

Limpar filtros

Atualização do painel em **16/05/2024** às **15:31:58**, com dados contidos nas Secretarias Estaduais de Saúde.

BRASIL
01/01/2024 a 11/05/2024

População
210.147.125

CASOS

Casos novos notificados na semana epidemiológica
6.849

Casos Acumulados
38.802.815

Incidência covid-19 (100 mil hab)
281,68

ÓBITOS

Óbitos novos notificados na semana epidemiológica
52

Óbitos Acumulados
712.090

Taxa mortalidade (100 mil hab)
1,64

Fundamentos ...

Modelo Computacional de Processamento de Dados

Free Pascal 3.2

File Edit Search Run Compile Debug Tools

SOL4.PAS

```
var F,L:real;
    i,j,n:integer;
    x:array[1..10] of real;
    y:array[1..10] of real;

begin
    write('n=');readln(n);
    FOR i:=1 TO n DO
    begin
        write('x[',i,']=');readln(x[i]);
        write('y[',i,']=');readln(y[i]);
    end;
    begin
        write('x[ n+1 ]=');readln(x[n+1]);
    end;
```

PROCESSAMENTO

DADOS

TABELA DE NÍVEIS DE EXPRESSIVIDADE SEMÂNTICA

Concepção dos dados	Nivel 0	Modelagem conceitual	Continuantes	Ocorrentes
			Processamento, consulta	Modelagem estatística
Dados digitais, arquivo digital	**Nivel 1**	Dados textuais	PLN, REN	
Dados não estruturados	**Nivell 2**	Tokens discretos, quase-signos (NÖTH, 2002), “data point” (SHAH 2020, p. 16), delimitados mas decontextualizados	MIN. TEXTOS	
Dados estruturados	**Nivel 3**	Dados contextualizados, um estado de coisas (JANSEN, 2008, 188), um “datum” Hjørland (2018): triplas entidade, atributo, valor, a representação de um fato ou fenômeno isolado	SPARQL	Ciên. Dados
	Nivel 4	Agregação de triplas referenciando uma única entidade ou fenômeno	SQL	Dados em tempo real, um Painel
	Nivel 5	2 ou more agregações de triplas referenciando 2 ou mais entidades ou fenômenos interrelacionados; esquema implícito	SQL	
	Nivel 6	Dados incluem o modelo conceitual/esquema; esquema explícito	SPARQL	

EXEMPLOS

Nível 1 – “O paciente 0ae989c74676b6b5b8bf1a5f57be45f7 fez o exame URINA 1” (Um texto)

Nível 2 – “paciente”, “0ae989c74676b6b5b8bf1a5f57be45f7”, “URINA 1”, “exame” (Tokens discretos)

Nível 3 - <paciente_0ae989c74676b6b5b8bf1a5f57be45f7> <exame> <URINA 1>. (Triplas RDF)

Nível 4 –

ID_aTENDIMENTO	DT_COLETA	DE_ORIGE	DE_EXAME	DE_ANALITO	DE_RESULTADO	CD_UNIDAD	DE_VALOR_REFERENCI
0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	CULTURA AERÃ“BIA	CULTURA AERÃ“BIA	Negativa		Negativa
0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	URÃ‰IA	URÃ‰IA	10	mg/dL	10 a 50 mg/dL
0ae989c74676b6b5b8bf1a5f57be45f7	17/05/2020	HOSP	URINA TIPO 1	pH :	7.0		5.0 a 6.0

(Uma tabela)

Nível 5 –

Column1	Column2	Column3	Column4	Column5	Column6	Column7
ID_PACIENTE	IC_SEXO	AA_NASC	CD_PAIS	CD_UF	CD_MUNICIPIO	CD_CEPREDUZIDO
9698838a8fa8a01ffd5ed5c71e8e17a3	F	1962	BR	SP	SAO PAULO	CCCC
d9fec23b3820f93a961841d569db8cb5	F	1974	BR	SP	MMMM	CCCC

(2 ou mais tabelas interrelacionadas)

Nível 6 -

(Uma ontologia)

Tipos de Saídas

- **Consulta SQL** sobre 1 ou mais tabelas interrelacionadas

  Saída - o **subconjunto** de linhas e células de uma tabela que atendem determinada condição

- **Consulta SPARQL** em um grafo

  Saída - o **subconjunto** das triplas que atendem determinada condição

ESTÁTICO, Modelagem Conceitual, um campo de dados, uma célula

DINÂMICO, Modelagem Estatística, AO LONGO DO TEMPO, variáveis

- **Análise descritiva** (SHAH 2020, p. 67) sobre 1 ou mias tabelas interrelacionadas

  Saída – informações gerais sobre um conjunto de dados

- **Correlações** (SHAH 2020, p. 82) sobre 1 ou mais tabelas interrelacionadas

  Saída – uma visão (possivelmente uma visão gráfica) sobre como as variáveis correlacionadas variam uma em relação à outra ao longo do tempo

- **Painel** (em tempo real) sobre 1 ou mais fontes de informação

  Saída – uma visão em tempo real sobre como variáveis correlacionadas variam uma em relação a outra ao longo do tempo

6. Considerações finais

A semântica “emerge” dos dados em níveis crescentes de acordo com 2 eixos:

– Eixo da complexidade das representações - À medida que dados se organizam em conjuntos ou sistemas mais complexos, eles podem se tornar mais expressivos e representar coisas em um domínio com mais precisão;

- Eixo do processamento - ao processar as coisas representadas pelos dados

MODELAGEM CONCEITUAL X MODELAGEM ESTATÍSTICA

Aproximar OC, RC, MC do *Big Data*, da Ciência de Dados

References

LATOUR, Bruno. Redes que a razão desconheco: laboratórios, bibliotecas, coleções. In: Baratin, M; Jacob, C. O poder das bibliotecas. Rio de Janeiro : Ed. UFRJ, 2000.

COGNIZANT. Making Sense of Big Data in the Petabyte Age. Cognizant, 20- 20 insights, jun. 2011. Disponível em: httpo://www.cognizant.com/whitepapers/Making-Sense-of-Big-Data-in-thePetabyte-Age.pdf. Acesso em: 02 abr. 2021.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. *The semantic web*. Scientific American, May, 2001.

FEIBLEMAN, JAMES K. Theory Of Integrative Levels. The British Journal for the Philosophy of Science, v5, n.17, p. 59-66, 1954.

GRENON, Pierre; SMITH, Barry. SNAP and SPAN: Towards dynamic spatial ontology. *Spatial cognition and computation*, v. 4, n. 1, p. 69-104, 2004.

IBEKWE-SANJUAN, F.; BOWKER, G. C. Implications of big data for Knowledge Organization. Knowledge Organization, Baden-Baden, v. 44, n. 3, p. 187-198, 2017.

HJØRLAND, Birger. Data with big data and database semantics. In. IEKO, ISKO Encyclopedia of Knowledge Organization. ISKO: CastanhoClaro. Disponível em: httpo://www.isko.org/cyclo/data. Acesso em: 02 dez. 2020.

HARTMANN, NICOLAI. New ways of ontology. Henry Regnery Company, Chicago, Illinois, 1952.

GNOLI, Claudio. Mentefacts as a missing level in theory of information science. *Journal of Documentation*, v. 74, n. 6, p. 1226-1242, 2018. Disponível em: httpo://www.gnoli.eu/mentefacts.docx. Acesso em: 07 ago. 2023.

SHETH, Amit; RAMAKRISHNAN, Cartic; THOMAS, Christopher. Semantics for the semantic web: The implicit, the formal and the powerful. International Journal on Semantic Web and Information Systems (IJSWIS), v. 1, n. 1, p. 1-18, 2005. Disponível em: httpo://www.academia.edu/download/90817267/JSWIS.pdf#page=19. Acesso em: 12 set. 2019.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. Handbook on Constructing Composite Indicatoro: Methodology and User Guide, 2008. Disponível em: httpo://compositeindicators.jrc.ec.europa.eu/. Acesso em: 22 mai. 2024.

SOUZA, Renato Rocha; TUDHOPE, Douglas; ALMEIDA, Maurício Barcellos. Towards a taxonomy of KOo: Dimensions for classifying Knowledge Organization Systems. KO KNOWLEDGE ORGANIZATION, v. 39, n. 3, p. 179-192, 2012. Disponível em: .

MAC Niterói

Museu de Arte Contemporânea, Niterói, Rio de Janeiro

Comentários são bem vindos

Obrigado !

ch_marcondes@id.uff.br

http://profmarcondes.ong.br