

Vocabulários controlados e web semântica

Terminologia e Organização do Conhecimento

27/04/26

Bruno Almeida

brunoalmeida@fcsb.unl.pt

Fundamentos da web semântica e dos *linked data*

- [illegible]

2.

Web semântica: a web de dados

- “*Semantic Web*” é o termo cunhado por Tim Berners-Lee em 1999 para designar a evolução da **web de documentos** para uma **web de dados estruturados**, tornando possível a sua interpretação pelas máquinas:

*Machines become capable of analyzing all the data in the Web – the content, links, and transactions between people and computers. A “**Semantic Web**”, which would make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy, and our daily lives **will be handled by machines talking to machines**, leaving humans to provide the inspiration and intuition.*

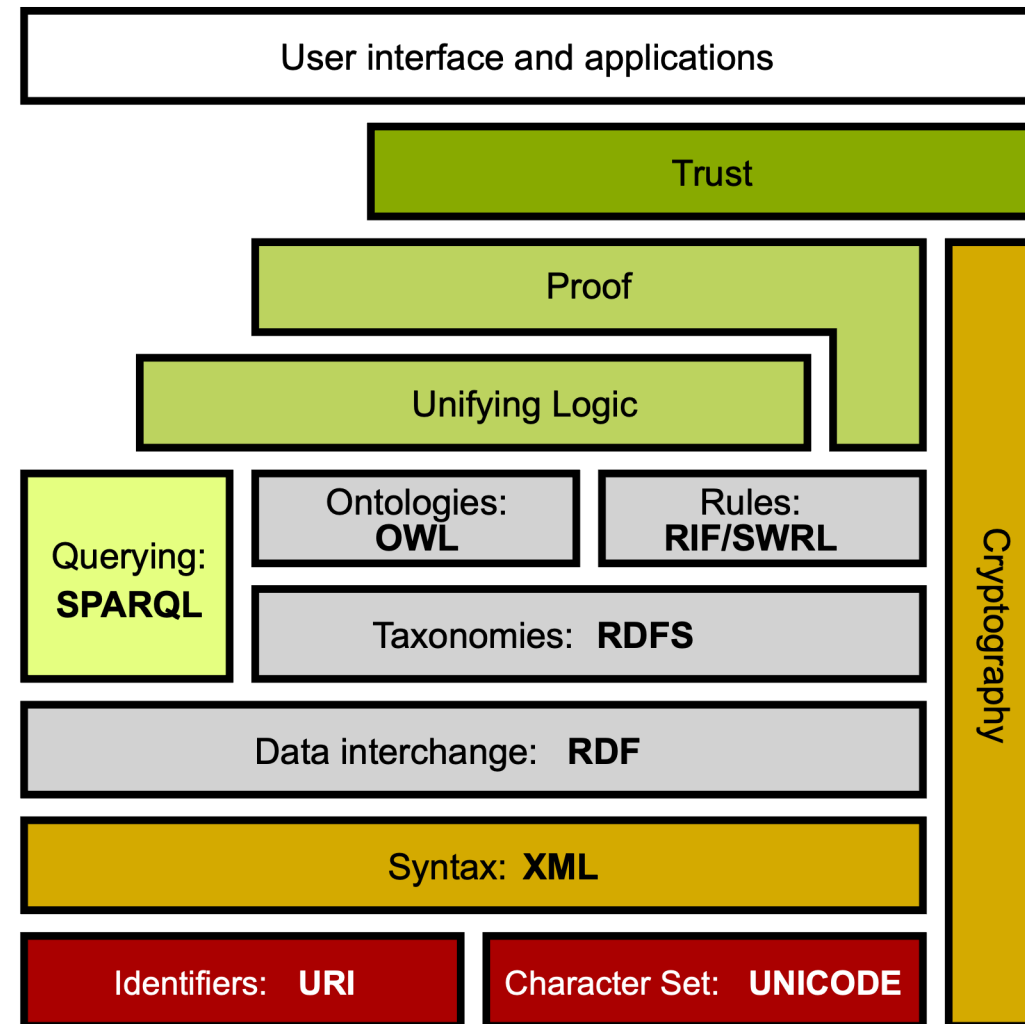
Berners-Lee, *Weaving the Web*, 1999, pp. 157-158

Web semântica: a web de dados

- No âmbito do W3C, tem vindo a ser desenvolvido ao longo dos últimos 25 anos um conjunto de tecnologias para a construção da web semântica.
- As tecnologias da web semântica têm vindo a ser aplicadas com sucesso em diversas comunidades especializadas, tais como a **biomedicina**, as **ciências da informação**, e as **humanidades digitais**.

Tecnologias da web semântica

- A arquitetura da web semântica assenta no **Unicode**, para codificação de texto na generalidade das línguas, e no uso de **URI/IRI** (*Uniform/Internationalized Resource Identifier*) para identificação unívoca de recursos na Web.
- O **XML** (Extensible Markup Language) funciona como “língua franca” para a codificação dos dados na Web.
- O **RDF** (*Resource Description Framework*) é o modelo de base para troca de dados na Web.



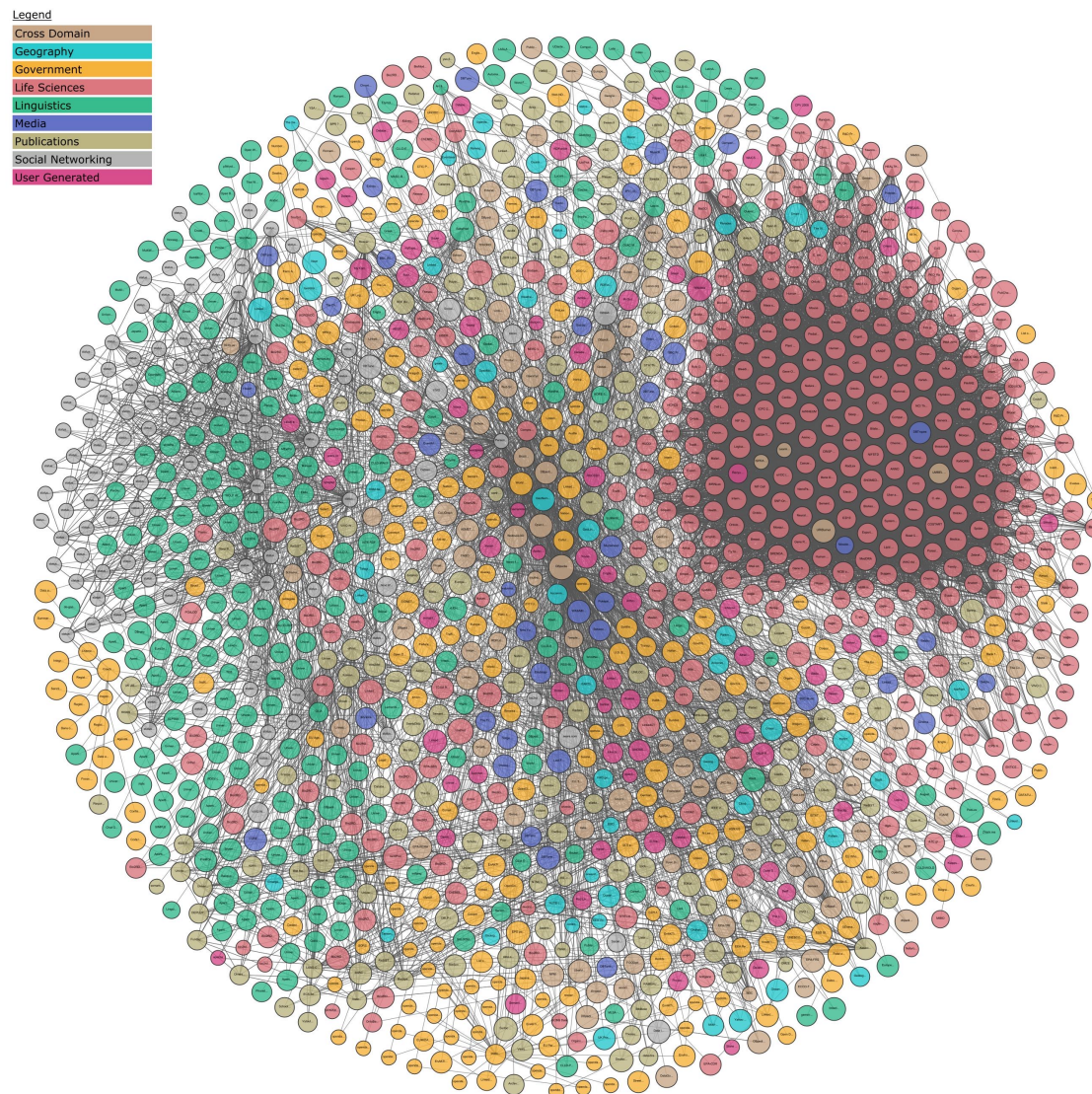
Linked Data: dados interligados

- Os *linked data* são um dos principais pilares da web semântica.
- Trata-se de um conjunto de princípios para publicar dados estruturados na Web e ligá-los a outras fontes de dados (Berners-Lee, *Linked Data*, 2006.
<https://www.w3.org/DesignIssues/LinkedData.html>):
 - Utilizar **URI/IRI** para identificar recursos;
 - Utilizar **URI/IRI HTTP**, de forma a que os recursos possam ser recuperados na Web;
 - Fornecer informação útil sobre os recursos através de **normas abertas**, como o **RDF**;
 - Incluir **ligações** para outros URI/IRI, de forma a que possam ser descobertos.

Linked Open Data: dados abertos e interligados

- Dados interligados publicados através de uma **licença aberta** (Berners-Lee, 2006).
- Esquema de 5 estrelas para avaliar a qualidade dos dados abertos e interligados:
 - ★ Disponíveis com **licença aberta** na Web, em qualquer formato.
 - ★★ Além de (1), utilização de um **formato estruturado** (p. ex., XLS).
 - ★★★ Além de (2), utilização de um **formato estruturado não-proprietário** (p. ex., CSV).
 - ★★★★ Além de (3), utilização de **normas abertas do W3C**, como RDF.
 - ★★★★★ Além de (4), **ligar os dados** a outras fontes de dados abertos e interligados.

Linked Open Data Cloud



Um exemplo: Getty AAT

- **Getty Art & Architecture Thesaurus:**

- <https://www.getty.edu/research/tools/vocabularies/aat>
- Tesauro do Instituto Getty para a **descrição documental e recuperação de informação** em **arte, arquitetura e património cultural**.
- É um **vocabulário controlado e estruturado**.
- Informação sobretudo em inglês, com alguns equivalentes em chinês, alemão, neerlandês, espanhol e português.
- Ex. de conceito (<caixas de música>):
- <http://vocab.getty.edu/aat/300042659> (**URI do conceito**).
- <https://vocab.getty.edu/page/aat/300042659> (página Web com informação sobre o conceito).

Um exemplo: Getty AAT

Click the  icon to view the hierarchy.


Ligações e identificadores

[Semantic View](#) ([JSON](#), [RDF](#), [N3/Turtle](#), [N-Triples](#))

ID: 300042659

Page Link: <http://vocab.getty.edu/page/aat/300042659>

Record Type: [concept](#)

 **music boxes** (mechanical instruments, musical instruments, ... Furnishings and Equipment (hierarchy name))

Nota de âmbito do conceito

Note: Mechanical musical instruments in which tuned steel prongs are made to vibrate by contact with moving parts driven by a clockwork mechanism, the whole contained in a box.

Termos preferenciais e alternativos









Terms:

music boxes ([preferred](#), [C,U,English-P,D,U,PN](#))
music box ([C,U,LC,English,AD,U,SN](#))
boxes, music ([C,U,English,UF,U,N](#))
musical box ([C,U,English,UF,U,N](#))
musical boxes ([C,U,English,UF,U,N](#))
muziekdozen ([C,U,Dutch-P,D,U,U](#))
muziekdoos ([C,U,Dutch,AD,U,U](#))
boîtes à musique ([C,U,French-P,D,U,PN](#))
boîte à musique ([C,U,French,AD,U,SN](#))
Spieldose ([C,V,German-P,D,SN](#))
Spieldosen ([C,V,German,AD,PN](#))
caixas de música ([C,U,Portuguese-P,D,U,PN](#))
caixa de música ([C,U,Portuguese,AD,U,SN](#))
cajas de música ([C,U,Spanish-P,D,U,PN](#))
caja de música ([C,U,Spanish,AD,U,SN](#))

Facet/Hierarchy Code: [V.TT](#)

Posição do conceito no sistema conceptual

Hierarchical Position:

 [Objects Facet](#)
 [Furnishings and Equipment \(hierarchy name\)](#) (G)
 [Sound Devices \(hierarchy name\)](#) (G)
 [sound devices \(equipment\)](#) (G)
 [<sound devices by function>](#) (G)
 [musical instruments](#) (G)
 [mechanical instruments](#) (G)
 [music boxes](#) (G)

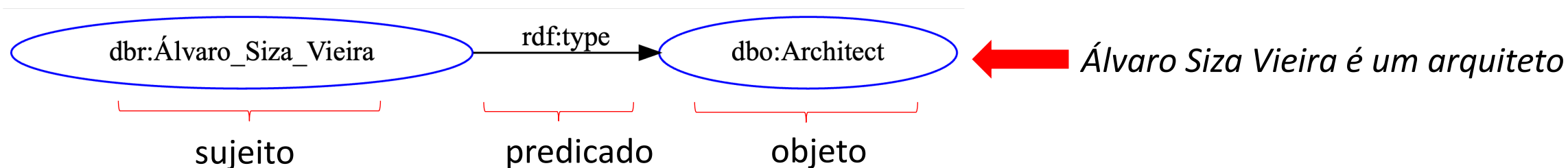
Related concepts:

created by [music box makers](#)
 (<people in crafts and trades by product>, <people in crafts and trades>, ... People (hierarchy name)) [300386288]

Representar informação em RDF

Representar informação em RDF

- A informação em RDF assume a forma de **triplos**, i.e., declarações com **sujeito**, **predicado** e **objeto**.
- Representação gráfica de um triplo (dois **nós** e um **arco**):



Sintaxes RDF: Turtle

- Existem diversas sintaxes para representar informação em RDF: RDF/XML, Turtle, JSON-LD, etc.
- O mais legível para humanos é o **Turtle** (*Terse RDF Triple Language*) (ficheiros *.ttl):

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dbo: <http://dbpedia.org/ontology/> .

dbp:Álvaro_Siza_Vieira rdf:type dbo:Architect .
```

Declaração do **espaço de nomes** (*namespace*) com **prefixos**

Declaração de um **triplo**

sujeito

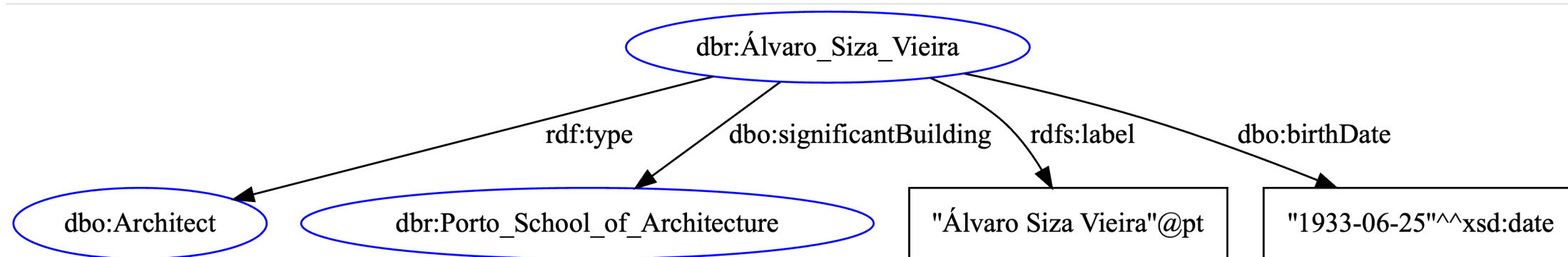
predicado

objeto

Nota: a propriedade `rdf:type` por ser substituída por `a`: `dbp:Álvaro_Siza_Vieira a dbo:Architect`

Grafos RDF

- Um **grafo RDF** corresponde a um conjunto com 1 ou mais triplos.
- Exemplo de um **grafo com 4 triplos**:



Álvaro Siza Vieira é um arquiteto. O seu nome é escrito como “Álvaro Siza Vieira” em português. Nasceu em 25/06/1933. Tem como edifício significativo a Faculdade de Arquitetura da Universidade do Porto.

- Grafo RDF em Turtle:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dbr: <http://dbpedia.org/resource/> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

dbr:Álvaro_Siza_Vieira rdf:type dbo:Architect;
    dbo:significantBuilding dbr:Porto_School_of_Architecture;
    rdfs:label "Álvaro Siza Vieira"@pt;
    dbo:birthDate "1933-06-25"^^xsd:date .
```


Principais tipos de nós em RDF

- **URI/IRI (Uniform/Internationalized Resource Identifier):**
 - Identificadores unívocos para identificar **recursos** na Web (documentos, pessoas, conceitos, etc.);
 - Os **IRI** incluem mais caracteres (ex. acentos gráficos, caracteres chineses);
 - Alguns URI/IRI são também **URL** (<http://...>), i.e. permitem também **localizar** recursos na Web;
 - Podem aparecer na sua forma completa (ex. http://dbpedia.org/resource/Álvaro_Siza_Vieira) ou abreviada através de prefixos (ex. **dbr:Álvaro_Siza_Vieira**);
 - Os URI/IRI **podem aparecer em todas as posições de um triplo** (sujeito, predicado ou objeto).

Principais tipos de nós em RDF

- **Literais:**

- Utilizados para **valores de dados** (texto, números, datas, etc.);
- Existem ***typed literals*** (onde se declara um tipo) e ***plain literals***;
- Ex. "1933-06-25"^^xsd:date (*typed literal*);
- Ex. "Álvaro Siza Vieira"@pt (*plain literal*, com etiqueta de língua);
- Os literais **apenas podem aparecer na posição de objeto** nos triplos.

RDFS: Resource Description Schema

- O **RDF Schema** estende o RDF com elementos para representar **relações género-espécie** entre classes, **etiquetas linguísticas de conceitos**, entre outros aspetos.

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
  
foaf:Person rdfs:subClassOf foaf:Agent;  
           rdfs:label "Pessoa"@pt, "Person"@en.
```

Pessoa é uma subclasse de Agente na ontologia FOAF (Friend of a Friend), tendo como designação “Pessoa” em português e “Person” em inglês.

SKOS: Simple Knowledge Organization System

SKOS: Simple Knowledge Organization System

- Modelo do W3C para representar em RDF vocabulários controlados.
- O SKOS é definido formalmente como uma **ontologia** para modelizar vocabulários controlados como grafos RDF.
- Surgiu da necessidade de um modelo comum para a representação e troca de vocabulários na web, permitindo a sua reutilização e interligação com diversos conjuntos de dados.
- Facilitou a transição de vocabulários controlados para a web semântica, contribuindo para a *Linked Open Data Cloud*:
 - LC Authorities (LCSH, etc.)
 - Getty Vocabularies (AAT, TGN, etc.)
 - GEMET

SKOS: Simple Knowledge Organization System

- O SKOS tem por base os **conceitos**, definidos de forma análoga à **terminologia** e à **organização do conhecimento**.
- Um vocabulário controlado é modelizado em SKOS como um **esquema de conceitos**.

Concepts are the units of thought—ideas, meanings, or (categories of) objects and events—which underlie many knowledge organization systems. As such, concepts exist in the mind as abstract entities which are independent of the terms used to label them.

<https://www.w3.org/TR/skos-primer/>

Declaração de um conceito em SKOS

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix ex: <http://www.example.com/vocab/> .  
  
ex:concept_001 rdf:type skos:Concept .
```

- Este triplo declara concept_001 como um **tipo** (i.e. uma instância) da **classe** skos:Concept.
- “concept_001” é um conceito no modelo SKOS.

Designações em línguas naturais

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.com/vocab/> .

ex:concept_001 rdf:type skos:Concept;
               skos:prefLabel "arquiteto"@pt, "architect"@en;
               skos:altLabel "architectural designer"@en;
               skos:hiddenLabel "arquitecto"@pt .
```

- O SKOS inclui propriedades para indicar **formas preferenciais** (skos:prefLabel), **alternativas** (skos:altLabel) e **ocultadas** (skos:hiddenLabel) para as designações.
- As formas ocultas permitem pesquisar e recuperar o conceito, mas **não são visíveis** para os utilizadores do vocabulário controlado.

Relações conceptuais

- O SKOS permite representar dois tipos de relações entre conceitos:
 - **Hierárquicas** (genéricas ou partitivas): através das propriedades `skos:broader` e `skos:narrower`.
 - **Associativas**: através da propriedade `skos:related`.
- A propriedade `skos:broader` relaciona um **conceito subordinado** com um seu **conceito superordenado** (**générico** ou **compreensivo**).
- A propriedade `skos:narrower` é a sua inversa, i.e. relaciona um **conceito superordenado** com um seu **conceito subordinado** (**específico** ou **partitivo**).

Relações hierárquicas

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.com/vocab/> .

ex:concept_001 rdf:type skos:Concept;
               skos:prefLabel "arquiteto"@pt;
               skos:narrower ex:concept_002 .

ex:concept_002 rdf:type skos:Concept;
               skos:prefLabel "arquiteto paisagista"@pt;
               skos:broader ex:concept_001 .
```

Relações associativas

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix ex: <http://www.example.com/vocab/> .
```

```
ex:concept_001 rdf:type skos:Concept;  
               skos:prefLabel "arquiteto"@pt;  
               skos:related ex:concept_003 .  
  
ex:concept_003 rdf:type skos:Concept;  
               skos:prefLabel "arquitetura"@pt;  
               skos:related ex:concept_001 .
```

Documentação de conceitos

- O SKOS inclui diversas propriedades para modelizar diversos tipos de **anotações** associadas a conceitos num vocabulário controlado:
 - **skos:note** – propriedade para uma nota em geral.
 - **skos:scopeNote** – explicação do uso do conceito na descrição documental, etc.
 - **skos:definition** – descrição do significado do conceito.
 - **skos:example** – exemplos do conceito, ou conceitos subordinados.
 - **skos:historyNote** – descrição da evolução histórica do conceito.
 - **skos:editorialNote** – informação administrativa, mudanças a realizar, etc.
 - **skos:changeNote** – documentação de alterações realizadas num conceito.

Documentação de conceitos

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
```

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
@prefix ex: <http://www.example.com/vocab/> .
```

```
ex:concept_002 rdf:type skos:Concept;
```

```
    skos:prefLabel "arquiteto paisagista"@pt;
```

```
    skos:definition "Arquiteto especializado na ordenação do espaço  
    exterior em relação ao ser humano"@pt;
```

```
    skos:scopeNote "Usar para classificar os arquitetos paisagistas na  
    base de dados. Para as obras sobre a área do conhecimento, usar  
    "arquitetura"."@pt.
```

Coleções de conceitos

- Uma **coleção de conceitos** corresponde a um agrupamento temático de conceitos de um vocabulário controlado.
- A classe `skos:Collection` permite representar coleções de conceitos como suas instâncias.
- A propriedade `skos:member` permite indicar quais os conceitos que são membros de uma coleção.
- As coleções de conceitos podem ser utilizadas para modelizar os chamados “termos-guia”, i.e. **grupos de conceitos** com base num mesmo critério de divisão.

Coleções de conceitos

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.com/vocab/> .

ex:collection_001 rdf:type skos:Collection;
    skos:prefLabel "pessoas por atividade"@pt;
    skos:member ex:concept_001, ex:concept_005 .
ex:concept_001 rdf:type skos:Concept;
    skos:prefLabel "arquiteto"@pt;
    skos:inScheme ex:my_vocab .
ex:concept_005 rdf:type skos:Concept;
    skos:prefLabel "designer"@pt;
    skos:inScheme ex:my_vocab .
```

Esquemas de conceitos

- Um **esquema de conceitos**, instância de `skos:ConceptScheme`, corresponde à **representação do vocabulário controlado em si**.
- A propriedade `skos:inScheme` é usada para agregar todos os conceitos que integram o respetivo esquema / vocabulário controlado.
- Os conceitos de topo de um vocabulário controlado são modelizados através das propriedades `skos:hasTopConcept` e `skos:topConceptOf`.

Esquemas de conceitos

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.com/vocab/> .

ex:my_vocab rdf:type skos:ConceptScheme;
    skos:prefLabel "0 meu primeiro vocabulário controlado"@pt;
    skos:hasTopConcept ex:concept_004, ex:concept_006 .

ex:concept_004 rdf:type skos:Concept;
    skos:prefLabel "pessoa"@pt;
    skos:inScheme ex:my_vocab;
    skos:topConceptOf ex:my_vocab .
```

Mapeamento entre esquemas de conceitos

- O SKOS inclui diversas propriedades para mapear conceitos, **através dos respectivos URI/IRI**, entre esquemas de conceitos diferentes.
- A existência destes mapeamentos promove a **interoperabilidade** entre vocabulários controlados diferentes publicados como *linked open data*.

Mapeamento entre esquemas de conceitos

- **skos:exactMatch** – o conceito do esquema de origem corresponde **exatamente** ao conceito do esquema externo.
- **skos:closeMatch** – o conceito do esquema de origem corresponde **parcialmente** ao conceito do esquema externo.
- **skos:broadMatch** – o conceito do esquema de origem tem um **conceito mais amplo, i.e. superordenado** (genérico ou partitivo) no esquema externo.
- **skos:narrowMatch** – o conceito do esquema de origem tem um **conceito mais restrito, i.e. subordinado** (genérico ou partitivo) no esquema externo.
- **skos:relatedMatch** – o conceito do esquema de origem tem um **conceito relacionado** (associativo) no esquema externo.

Mapeamento entre esquemas de conceitos

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix ex: <http://www.example.com/vocab/> .
```

```
ex:concept_001 rdf:type skos:Concept;
```

```
    skos:prefLabel "arquiteto"@pt;
```

```
    skos:inScheme ex:my_vocab;
```

```
    skos:exactMatch <https://vocab.getty.edu/aat/300024987>;
```

```
    skos:narrowMatch <https://vocab.getty.edu/aat/300024993> .
```

<architects>
(Getty AAT)

<landscape architects>
(Getty AAT)

- Mapeamento do conceito <arquiteto> no nosso vocab com um conceito equivalente e com outro mais restrito do Getty AAT.

Exemplo: tipos de dados arqueológicos (em SKOS)

Tipos de dados arqueológicos

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix ex: <http://www.example.com/vocab/> .  
  
ex:scheme_01 rdf:type skos:ConceptScheme;  
             skos:prefLabel "Vocabulário de conceitos arqueológicos"@pt,  
                           "Vocabulary of archaeological concepts"@en;  
             skos:hasTopConcept ex:concept_001 .
```

Tipos de dados arqueológicos

```
ex:concept_001 rdf:type skos:Concept;  
    skos:prefLabel "archaeological data"@en;  
    skos:definition "cultural item, cultural feature, ecofact or  
        site."@en;  
    skos:inScheme ex:scheme_01;  
    skos:topConceptOf ex:scheme_01;  
    skos:narrower ex:concept_002, ex:concept_003, ex:concept_004,  
        ex:concept_005 .
```

Tipos de dados arqueológicos

```
ex:concept_002 rdf:type skos:Concept;  
    skos:prefLabel "cultural items"@en;  
    skos:definition "archaeological data that are discrete and  
        transportable entities partially resulting from cultural  
        activity, and that can be formally analyzed without  
        information about their provenience."@en;  
    skos:inScheme ex:scheme_01;  
    skos:broader ex:concept_001 .
```


Tipos de dados arqueológicos

```
ex:concept_003 rdf:type skos:Concept;  
    skos:prefLabel "cultural features"@en;  
    skos:definition "archaeological data that partially result from  
        cultural activity, which cannot be observed or analyzed if  
        removed from the field."@en;  
    skos:inScheme ex:scheme_01;  
    skos:broader ex:concept_001 .
```

Tipos de dados arqueológicos

```
ex:concept_004 rdf:type skos:Concept;  
    skos:prefLabel "ecofacts"@en;  
    skos:definition "archaeological data that result from natural  
        processes and that represent or inform about articulations  
        between a culture and its surrounding natural systems."@en;  
    skos:inScheme ex:scheme_01;  
    skos:broader ex:concept_001 .
```

Tipos de dados arqueológicos

```
ex:concept_005 rdf:type skos:Concept;  
    skos:prefLabel "sites"@en;  
    skos:definition "archaeological data that are spatial clusters of  
        cultural features and/or cultural items."@en;  
    skos:inScheme ex:scheme_01;  
    skos:broader ex:concept_001 .
```

Exercício prático: tipos de coleções de documentos em SKOS

Exercício prático

- Transferir e editar o seguinte ficheiro para um ficheiro de texto simples:
<https://tinyurl.com/yzrt7cy7>
- Incluir os conceitos de **arquivo** e **coleção factícia**
- Incluir as seguintes definições dos conceitos em português:
 - coleção de documentos: arquivo ou coleção factícia
 - arquivo: coleção de documentos reunida organicamente
 - coleção factícia: coleção de documentos reunida artificialmente
- Mapear o conceito de arquivo para um conceito equivalente no Getty AAT:
 - **<https://www.getty.edu/research/tools/vocabularies/aat>**
- Consultar o SKOS Primer para resolver dúvidas (Secções 2-3):
 - **<https://www.w3.org/TR/skos-primer>**
- No final, visualizar o vocabulário no RDF Grapher:
 - **<https://www.ldf.fi/service/rdf-grapher>**