

Datos abiertos enlazados

¿Qué son los datos abiertos enlazados?

Conjunto de buenas prácticas para exponer, compartir y conectar datos en la Web

Usan **ontologías** para estructurar los datos

Permite interconectar distintos conjuntos de datos

Pueden realizarse consultas muy expresivas sobre los datos

La mejor solución existente para tratar el reto de la heterogeneidad de datos

El esquema de estrellas de los datos abiertos enlazados

¿Cuántas estrellas tienen tus datos?

★ disponible en la Web

★★ disponible en un formato de datos procesable

★★★ lo anterior y en un formato no propietario

★★★★ lo anterior y usando los estándares abiertos del W3C

★★★★★ lo anterior y enlazando tus datos a los de otros

Ontologías

Una ontología es una conceptualización compartida de un dominio descrita en un lenguaje lógico que es interpretable computacionalmente

Las ontologías sirven para modelar la información, típicamente describiendo los aspectos comunes y las diferencias

Una ontología se compone de **clases** (organizadas en taxonomías), **propiedades** y **entidades**

RDFS y OWL son lenguajes del W3C para crear ontologías

Anotando datos

Para anotar datos abiertos enlazados se utiliza el lenguaje RDF del W3C

Puede anotarse cualquier cosa con RDF, diferenciando entre **entidades**, las cosas existentes que se identifican con URIs, y **literales** para datos primitivos

La anotación básica en RDF es una **tripla**, compuesta por sujeto–predicado–objeto

Ejemplo de anotación de datos con RDF extraído de la DBpedia inglesa:

```
dbr:Pinus_nigra rdfs:label "Pinus nigra" .
dbr:Pinus_nigra dbo:kingdom dbr:Plant .
dbr:Pinus_nigra dbo:family dbr:Pinaceae .
dbr:Pinaceae rdfs:label "Pinaceae" .
```

Consultas sobre datos abiertos enlazados

Los datos abiertos enlazados suelen ofrecerse en un **almacén de triplas** –un gestor de bases de datos especializado en RDF

Se llama **punto SPARQL** al procesador de consultas que recupera información de un almacén de triplas a partir de consultas SPARQL

SPARQL es el lenguaje del W3C para realizar consultas sobre datos enlazados; el papel de SPARQL es similar al de SQL en bases de datos relacionales

Elementos de una consulta SPARQL:

- **Declaraciones de prefijos** para abreviar URIs
- **Cláusula SELECT** con las variables a retornar
- **Consulta** compuesta por patrones de triplas en los que una variable puede reemplazar cualquier elemento de una tripla
- **Modificadores** opcionales para ordenación, agrupamientos. . .

Consultas SPARQL de ejemplo

Todos los ejemplos se basan en el punto SPARQL de la DBpedia inglesa: <http://dbpedia.org/sparql>

- Listar las subclases de `dbo:Plant`

```
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?subclass WHERE {
?subclass rdfs:subClassOf dbo:Plant . }
```

- Mostrar 100 instancias de `dbo:Plant` con sus etiquetas en español

```
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT DISTINCT ?indiv ?lab WHERE {
?indiv a dbo:Plant .
?indiv rdfs:label ?lab .
FILTER ((LANG(?lab) = "es")). }
LIMIT 100
```

El Explorador Forestal

El Explorador Forestal es un ejemplo de una aplicación web que consume datos abiertos enlazados

Esta aplicación ofrece un mapa interactivo con el que pueden explorarse de manera sencilla los inventarios y mapas forestales nacionales de España y Portugal

El usuario final no necesita conocer RDF, OWL o SPARQL para poder manejar la aplicación

Puedes probar el Explorador Forestal en este enlace:

<https://forestexplorer.gsic.uva.es/>

Autor

Guillermo Vega Gorgojo



Proyecto de innovación docente

