
Models RDF

PID_00272093

Blas Torregrosa García

Temps mínim de dedicació recomanat: 1 hora



**Blas Torregrosa García**

Enginyer en Informàtica i màster universitari en Seguretat de les Tecnologies de la Informació i de les Comunicacions (MISTIC) per la Universitat Oberta de Catalunya (UOC). Especialitzat en ciberseguretat. Professor col·laborador del màster de Ciència de Dades de la UOC i professor associat a la Universitat de Valladolid (UVA).

L'encàrrec i la creació d'aquest recurs d'aprenentatge UOC han estat coordinats pel professor: Ferran Prados Carrasco (2020)

Primera edició: febrer 2020
© Blas Torregrosa García
Tots els drets reservats
© d'aquesta edició, FUOC, 2020
Av. Tibidabo, 39-43, 08035 Barcelona
Realització editorial: FUOC

Cap part d'aquesta publicació, incloent-hi el disseny general i la coberta, no pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, tant si és elèctric com químic, mecànic, òptic, de gravació, de fotocòpia o per altres mètodes, sense l'autorització prèvia per escrit dels titulars dels drets.

Índex

Introducció	5
1. Model de dades RDF	7
1.1. Contenedors	8
1.2. Col·leccions	9
2. Representacions RDF	10
2.1. Notació N3	10
2.2. RDF/XML	10
2.3. Turtle	11
2.4. RDFa	12
2.5. JSON-LD	13
Bibliografia	15

Introducció

Aquest mòdul tracta del model de dades RDF (*Resource Description Framework*) com a pedra angular de tota la web semàntica i com un estàndard flexible per a l'estructuració de la informació. Veurem en què consisteix el model i les diferents representacions (serialitzacions) que admet.

«El Marc de Descripció de Recursos (RDF o, en anglès, *Resource Description Framework*) és un llenguatge dissenyat per recolzar la web semàntica, de la mateixa manera que HTML va ajudar a iniciar la web original. RDF és un marc per a la descripció de recursos o metadades per a la web. RDF proporciona estructures comunes que es poden usar per a l'intercanvi de dades interoperables amb XML.»

Declaració d'activitat de web semàntica del W3C (World Wide Web Consortium).

RDF es va crear el 1997 i llavors XML era un format que començava a ser reconegut i que va tenir una gran influència en la sintaxi d'RDF. El format RDF/XML resulta molt prolix i van aparèixer diverses modificacions per fer RDF més llegible.

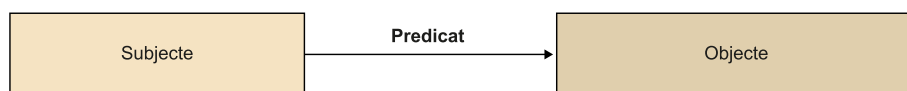
El 2014 apareix RDF 1.1 com una versió revisada que segueix mantenint la major part del model de dades, però afegeix altres formats de representació com Turtle o JSON-LD. En aquest mòdul veurem un breu resum del model de dades RDF seguint la versió RDF 1.1.

1. Model de dades RDF

El **model de dades RDF** es basa a definir sentències (*statements*) sobre recursos utilitzant tripletes. Cada tripleta consta d'un subjecte, un predicat i un objecte. Les **tripletes RDF** es representen com un graf dirigit que connecta dos nodes (subjecte i objecte) mitjançant un arc (predicat).

El coneixement en RDF s'expressa mitjançant una llista de sentències que segueixen el mateix esquema (tripletes RDF).

Figura 1. Exemple de tripleta RDF



Per tant, una tripleta RDF expressa una relació binària entre dos recursos: el **subjecte** i l'**objecte**. El **predicat** representa la naturalesa d'aquesta relació formulada de manera direccional (del subjecte a l'objecte) i es denomina *propietat*. Hi pot haver tres tipus de nodes:

1) Un **URI/IRI** que denota un recurs. Un URI pot ser un subjecte, un objecte o un predicat. La majoria de formats RDF permet definir els prefixos (com un àlies) per simplificar l'escriptura d'URI extensos. Hi ha alguns àlies freqüentment usats com `rdf`, `xsd`, `rdfs`, `owl`, etc.

2) **Literals** denoten recursos que tenen associat un valor com un enter o una cadena. Els literals solament poden aparèixer en l'objecte de les tripletes. Tots els literals RDF tenen associat un tipus de dades. Per exemple, `"33"^^xsd:integer` representa l'enter 33 i `"1969-07-21"^^xsd:date` representa el 21 de juliol de 1969. En cas que no tinguin declarat el tipus, se suposa per defecte que és una cadena `xsd:string`.

Un tipus especial de literals són les cadenes amb marques d'idioma que indiquen l'idioma en què està escrita aquesta cadena. Per exemple, `"hola"@es` representa el literal amb valor «hola» escrit en espanyol (es).

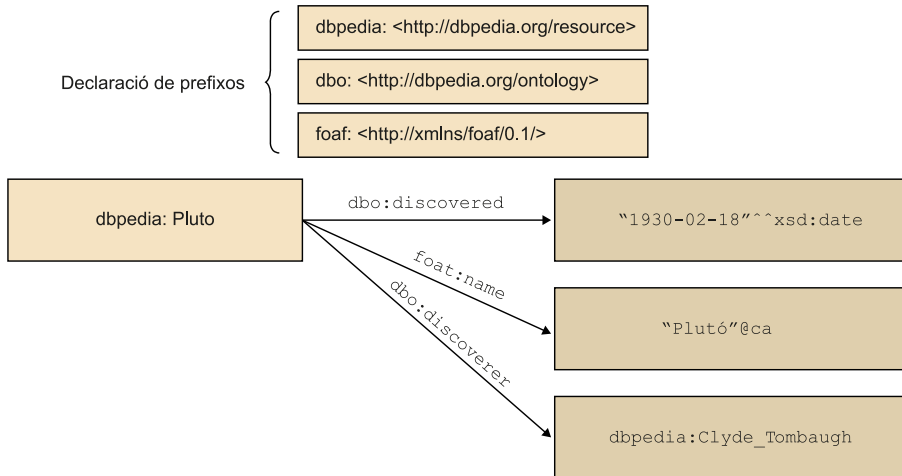
3) Nodes en **blanc** són identificadors locals que no referencien recursos. Els nodes en blanc es poden usar com a subjectes o objectes en les tripletes, especificant que hi ha la relació donada, sense nomenar-la explícitament. Els nodes en blanc es poden usar per fer afirmacions sobre elements els URI dels quals no són coneguts.

Enllaç d'interès

El servei `prefix.cc` es pot usar per buscar URI associats amb àlies coneguts.

Un **graf RDF** és un conjunt de tripletes RDF. Els subjectes solament poden ser URI o nodes en blanc, mentre que els objectes poden ser URI, nodes en blanc o literals.

Figura 2. Exemple de tripletes que defineixen el fet del descobriment de Plutó



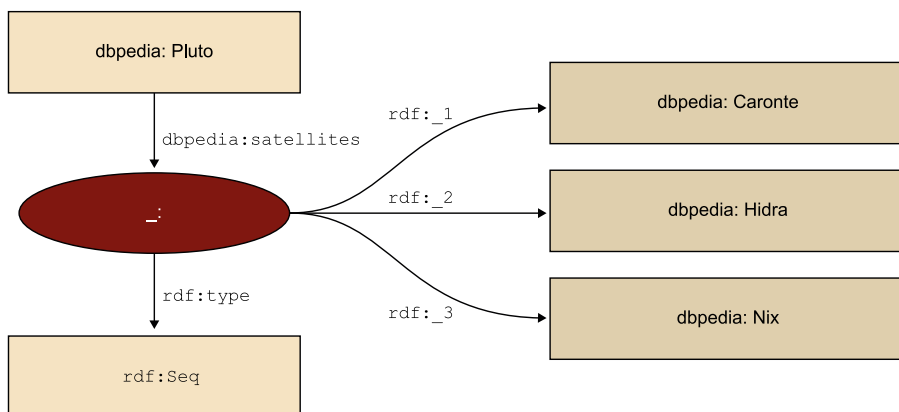
Una característica important dels grafs RDF és que dos grafs independents es poden fusionar per obtenir un altre graf format per la unió dels seus conjunts de tripletes. Atesa la naturalesa global dels URI, els nodes amb el mateix URI es poden unir automàticament.

1.1. Contenedors

Els **contenedors** (*containers*) i les **col·leccions** (*collections*) s'utilitzen per descriure conjunts de recursos. A RDF hi ha tres tipus de contenidors:

- `rdf:Bag`. Contenedor en què es permeten duplicats i l'ordre dels recursos no és significatiu.
- `rdf:Seq`. Contenedor seqüència en què es permeten duplicats i l'ordre sí que és significatiu.
- `rdf:Alt`. Contenedor que guarda una llista de recursos o literals alternatius (se seleccionarà un dels membres, per defecte el que tingui la propietat `rdf:_1`).

Figura 3. Exemple de contenidors



1.2. Col·leccions

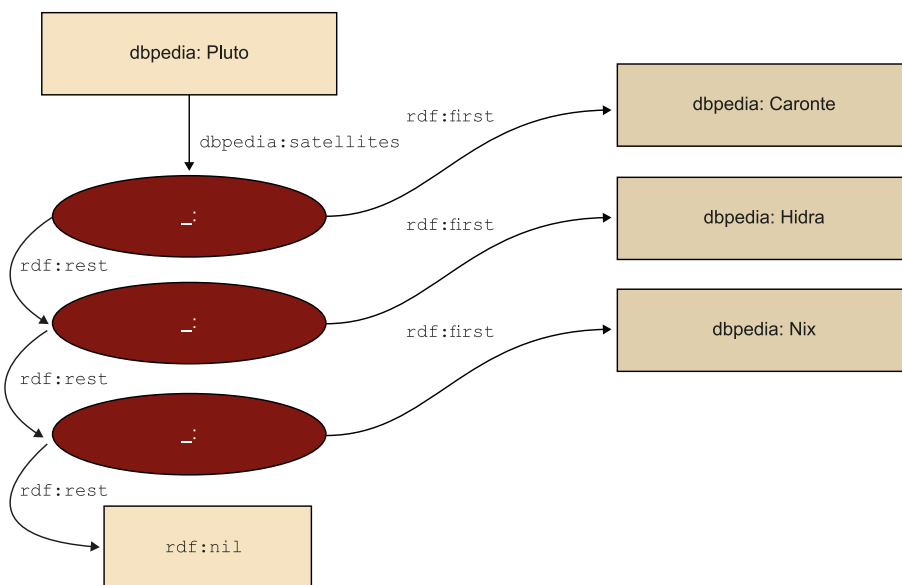
Una característica dels contenidors és que són llistes obertes, és a dir, no hi ha cap limitació en el nombre de membres que poden contenir. Per aconseguir llistes tancades, RDF disposa de les **col·leccions** el tipus bàsic de les quals és `rdf:List` i defineix una llista ordenada de recursos.

S'utilitzen unes propietats especials per estructurar aquest tipus de llistes:

- `rdf:first` per al primer element de la llista.
- `rdf:rest` per a la resta de la llista.
- `rdf:nil` per al final de la llista.

La llista s'estructura encadenant el conjunt de recursos (es poden usar nodes en blanc) en què cada node té una relació amb un recurs i una altra amb la resta.

Figura 4. Exemple de col·lecció



2. Representacions RDF

RDF és un model de dades abstracte i no un format en si mateix, de manera que no està vinculat a una sintaxi concreta. Hi ha diverses representacions sintàctiques o serialitzacions per al model RDF, algunes de les quals més adequades per al processament per ordinador i altres més comprensibles per a les persones.

Serialització

El concepte **serialització** es refereix a la representació de dades en un format per emmagatzemar-les, transportar-les o qualsevol altre motiu.

2.1. Notació N3

Notation3 (N3) és una forma de representar RDF amb la finalitat de facilitar la lectura per a les persones i la compactació.

La notació N3 té les característiques següents:

- Els URI/IRI van entre angles (< >).
- Els literals van entre cometes (" ").
- Cada tripleta acaba amb un punt (.

Figura 5. RDF amb N3

```
@prefix dbpedia: <http://dbpedia.org/resource/>.
@prefix dbo: <http://dbpedia.org/ontology/>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
```

```
dbpedia:Pluto dbo:discovered "1930-02-18".
dbpedia:Pluto foaf:name "Plutó".
dbpedia:Pluto dbo:discovered dbpedia:Clyde_Tombaugh
```

2.2. RDF/XML

La primera notació utilitzada va ser **RDF/XML**, estandarditzada pel W3C. El seu principal avantatge radicava en el fet que els llenguatges de programació disposen de llibreries per tractar amb XML. A més, es poden utilitzar els espais de noms (*Namespaces*) d'XML, evitant URI completes i fent-ho menys extens. No obstant això, RDF/XML resulta poc llegible per a les persones.

Com tot document XML ha de tenir una única arrel i cal indicar que és una arrel RDF juntament amb l'espai de noms de W3C anomenat *rdf*. Aquest espai de noms informa que el document és RDF i que l'etiqueta `rdf:RDF` forma part d'aquest.

Figura 6. RDF amb XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

<!-- Cos del document -->

</rdf:RDF>
```

RDF/XML pot tenir diversos enunciats (tripletes), per a cadascun dels quals cal afegir una etiqueta `rdf:Description`. L'atribut `rdf:about` declara el subjecte de l'enunciat. Es poden afegir diversos enunciats sobre el mateix subjecte.

Les propietats s'afegeixen com a nodes que pengen d'`rdf:Description` i contindran un identificador de la propietat (etiqueta XML), i també l'objecte de l'enunciat, que pot ser un literal o un URI/IRI. Els nodes en blanc s'identifiquen amb l'atribut `rdf:nodeID`, relatiu al document en qüestió.

Figura 7. Exemple d'RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dbo="http://dbpedia.org/ontology/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" >

  <rdf:Description rdf:about="http://dbpedia.org/resource/Pluto">
    <dbo:discovered>1930-02-18</dbo:discovered>
    <foaf:name xml:lang="ca">Plutó</foaf:name>
    <dbo:discoverer rdf:resource="http://dbpedia.org/resource/Clyde_Tombaugh"/>
  </rdf:Description>
</rdf:RDF>
```

2.3. Turtle

Turtle (*Terse RDF Triple Language*) és una extensió de Notation3.

Les característiques d'aquesta notació són les següents:

- El punt i coma (;) indica que les tripletes següents tenen el mateix subjecte.
- La coma (,) indica que les tripletes següents comparteixen el mateix subjecte i propietat (llista d'objectes).
- En els literals és possible indicar el tipus de dades (data, sencer, punt flotant, etc.).
- Els nodes en blanc s'indiquen amb `_:`. Si n'hi hagués diversos poden portar un identificador intern per al document, és a dir, `_:ID1`, `_:ID2`, etc.

Figura 8. Exemple d'RDF en Turtle

```
@base <http://dbpedia.org/resource/>.
@prefix dbo: <http://dbpedia.org/ontology/>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.

<Pluto> dbo:discovered "1930-02-18"^^xsd:date;
        foaf:name "Plutó"@ca;
        dbo:discovered <Clyde_Tombaugh>.
```

2.4. RDFa

RDFa (*Resource Description Framework in Attributes*) és un mecanisme que permet definir tripletes RDF dins de documents HTML de forma integrada amb la pàgina web.

El fonament d'RDFa és utilitzar un conjunt d'atributs per afegir informació RDF als llenguatges d'etiquetatge. Per a això es marca cada entitat amb un tipus d'element RDF. Aquest element tindrà el paper de subjecte. A continuació, es defineixen les propietats (equivalents als predicats) i els valors de les propietats (equivalents als objectes).

Els atributs bàsics per a l'etiquetatge de dades RDF són:

- `vocab`: indica el vocabulari que s'utilitzarà o de manera alternativa també es poden indicar espais de noms.
- `typeof`: indica el tipus de subjecte que es declara.
- `about`: indica l'URI del subjecte.
- `property`, `rel` o `rev`: indiquen les propietats del subjecte, és a dir, els predicats.
- `href`, `resource`, `src`: indiquen l'URI dels recursos relacionats, és a dir, els objectes. En el cas dels literals, estaran emmarcats per l'etiqueta.

En l'exemple següent s'utilitza RDFa per estendre una pàgina web amb contingut semàntic. La pàgina original és la següent:

Figura 9. Pàgina web convencional

```

<div>
  <span>Plutó</span>
  <div>
    Descobert el <span> 18 de febrer de 1930 </span>.
  </div>
  <div>
    Descobert per <a>Clyde Tombaugh</a>.
  </div>
</div>

```

A continuació tenim la pàgina amb etiquetatge RDFa.

Figura 10. Exemple de pàgina web amb RDFa

```

<html version="XHTML+RDFa 1.1"
  xmlns="http://www.w3.org/1999/xhtml" xml:lang="ca" lang="ca" >
  <div xmlns:dbpedia="http://dbpedia.org/resource/"
    xmlns:dbo = "http://dbpedia.org/ontology/"
    xmlns:foaf= "http://xmlns.com/foaf/0.1/"
    about="dbpedia:Pluto">
    <span property="foaf:name">Plutó</span>
  </div>
  Descobert el <span property="dbo:discovered"
    content="1930-02-18">18 de febrer de 1930</span>.
  </div>
  <div>
  Descobert per <a rel="dbo:discoverer"
    href="dbpedia:Clyde_Tombaugh">Clyde Tombaugh</a>.
  </div>
</div>
</html>

```

2.5. JSON-LD

JSON-LD (*JavaScript Object Notation for Linked Data*) és un format de serialització de tripletes RDF que té una representació més llegible per a les persones.

Mitjançant JSON, les tripletes es representen com a conjunts de parells atribut-valor.

Un dels elements principals de JSON-LD és la idea de **context**. Un context en JSON-LD permet que dues aplicacions utilitzin termes abreujats per comunicar-se entre si amb una major eficiència i sense perdre precisió.

Figura 11. Exemple d'RDF amb JSON-LD

```
{
"@context": { "dbpedia": "http://dbpedia.org/resource/",
              "dbo": "http://dbpedia.org/ontology/",
              "foaf": "http://xmlns.com/foaf/0.1/",
              "xsd": "http://www.w3.org/2001/XMLSchema#"
            },

"@id": "dbpedia:Pluto",

"dbo:discovered": { "@value": "1930-02-18" , "@type" : "xsd:date" },
"foaf:name": { "@value": "Plutó", "@language": "ca" },
"dbo:discoverer": { "@id": "dbpedia:Clyde_Tombaugh" }
}
```

Bibliografia

DuCharme, R. (2013). *Learning SPARQL* (2a. ed.). O'Reilly Media.

Guarino, N.; Oberle, D.; Staab, S. (2009). *What Is an Ontology?* [en línia]. [Data de consulta: gener 2020]. Disponible a: <https://iaoa.org/isc2012/docs/Guarino2009_What_is_an_Ontology.pdf>

Kumar, A. (2018). *Architecting Data-Intensive Applications*. Packt Pub.

Noy, N. F.; McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report (Informe SMI-2001-0880).

Powers, S. (2003). *Practical RDF*. O'Reilly Media.

