

A szemantikus web

Jeszenszky Péter

Debreceni Egyetem, Informatikai Kar

jeszenszky.peter@inf.unideb.hu

Utolsó módosítás: 2022. december 4.

Definíció (1)

- A W3C az alábbi módon definiálja a szemantikus webet:
 - *The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners.*
 - Lásd: <https://www.w3.org/RDF/FAQ>
 - A szemantikus web egy közös keretrendszer biztosít, mely lehetővé teszi az adatok alkalmazások, vállalatok és közösségek közötti megosztását és újrafelhasználását. Egy olyan kollektív munka, melyet a W3C vezet és melyben számos kutató és ipari partner vesz részt.

Definíció (2)

- Egy másik meghatározás ugyancsak a W3C-től:
 - *The term “Semantic Web” refers to W3C’s vision of the Web of linked data.*
 - Lásd: <https://www.w3.org/standards/semanticweb/>
 - A „szemantikus web” kifejezés a W3C az adatwebbel kapcsolatos jövőképét jelenti.

Vízió

- A vízió Tim Berners-Lee-től származik:
 - Tim Berners-Lee, James Hendler, Ora Lassila, *The Semantic Web*, Scientific American, May 2001, p. 29–37.
<https://www.scientificamerican.com/article/the-semantic-web/>

Az alapul szolgáló mechanizmusok szabványok

- Linked Data (RDF)
- Szótárak (OWL, SKOS)
- Lekérdezések (SPARQL)
- Következtetés (RIF, OWL)

Hasznos címek

- *W3C Data Activity – Building the Web of Data*
<https://www.w3.org/2013/data/>
- *Semantic Web Standards*
https://www.w3.org/2001/sw/wiki/Main_Page
- *W3C Semantic Web FAQ*
<https://www.w3.org/RDF/FAQ>

RDF

- Lehetővé teszi kijelentések megfogalmazását erőforrásokról gépi feldolgozásra alkalmas formában.
- Egy kijelentés egy egy tulajdonság értékét adja meg egy erőforráshoz.
 - Példa: <https://www.wikipedia.org/> has a ***creationDate*** whose value is "2001-01-15".
- Így tudásreprezentációt tesz lehetővé.

Specifikációk

- 2014. február 25.-én jelent meg az RDF 1.1 számú verziója.
 - *RDF 1.1 Primer* (W3C munkacsoport feljegyzés) <https://www.w3.org/TR/rdf11-primer/>
 - *RDF 1.1 Concepts and Abstract Syntax* (W3C ajánlás) <https://www.w3.org/TR/rdf11-concepts/>
 - ...

Alapfogalmak

- Erőforrás (*resource*)
- Literál (*literal*)
- Adattípus (*datatype*)
- Üres csomópont (*blank node*)
- RDF term
- RDF hármas (*RDF triple*), RDF kijelentés (*RDF statement*)
- RDF szótár (*RDF vocabulary*)
- RDF gráf (*RDF graph*)
- RDF dokumentum

Erőforrások

- Erőforrás bármi, ami egy IRI segítségével azonosítható.
 - Például a fizikai világ objektumai, dokumentumok, absztrakt fogalmak.
 - Egy IRI által jelölt erőforrást **jelöltnek** (*referent*) nevezünk.
- Erőforrásoknak tekintjük a literálokat is.
 - Egy literál által jelölt erőforrást **literális értéknek** (*literal value*) nevezünk.

Literálok (1)

- Az alábbi két vagy három komponensből állnak:
 - Egy **lexikális alaknak** (*lexical form*) nevezett Unicode karakterlánc.
 - Egy **adattípus IRI**, mely egy olyan adattípust azonosító IRI, mely meghatározza a lexikális alak leképezését egy literális értékre.
 - Egy nem üres **nyelvi címke** kizárólag a <http://www.w3.org/1999/02/22-rdf-syntax-ns#langString> adattípus IRI esetén.

Literálok (2)

- A konkrét szintaxisok támogathatják csupán egy lexikális formából álló, adattípus IRI vagy nyelvi címke nélküli **egyszerű literálok** használatát.
 - Ezek <http://www.w3.org/2001/XMLSchema#string> típusú literálok.
 - Példa: "stdout"
- A legtöbb konkrét szintaxis a nyelvi címkével ellátott karakterláncokat az adattípus IRI nélkül ábrázolja, mivel az mindig <http://www.w3.org/1999/02/22-rdf-syntax-ns#langString>
 - Példa: "Hello, World!"@en, "Helló, világ!"@hu

Adattípusok (1)

- Az RDF által használt adattípus fogalom kompatibilis az XML Schema 1.1 szabvánnyal.
 - Egy adattípust egy olyan rendezett hármas határoz meg, melyet az alábbi komponensek alkotnak:
 - **Lexikális tér (*lexical space*)**: az értékeket ábrázoló Unicode karakterláncok halmaza.
 - **Értéktér (*value space*)**: az értékek halmaza.
 - **A lexikális tér leképezése az értéktérre (*lexical-to-value mapping*)**: egy olyan függvény, mely a lexikális tér minden egyes eleméhez az értéktér egy elemét rendeli hozzá.
- Egy adattípust egy vagy több IRI jelöl.

Adattípusok (2)

- Példa: az <http://www.w3.org/2001/XMLSchema#boolean> adattípus
 - **Lexikális tér:** {"true", "false", "1", "0"}
 - **Értéktér:** {*true*, *false*}
 - **A lexikális tér leképezése az értéktérre:**
{("true", *true*), ("false", *false*), ("1", *true*), ("0", *false*)}

Adattípusok (3)

- Az RDF átveszi az XML Schema 1.1 sok beépített adattípusát.
 - Az alábbiak kivételével valamennyi beépített adattípus rendelkezésre áll:
 - `xsd:ENTITY`, `xsd:ENTITIES`, `xsd:ID`, `xsd:IDREF`,
`xsd:IDREFS`, `xsd:NOTATION`, `xsd:NMTOKENS`,
`xsd:QName`
 - Két további adattípus bevezetése:
 - <http://www.w3.org/1999/02/22-rdf-syntax-ns#HTML>
 - <http://www.w3.org/1999/02/22-rdf-syntax-ns#XMLLiteral>

Üres csomópontok

- Olyan erőforrásokat jelölnek, melyeket nem nevezünk meg explicit módon egy IRI-vel.
- Bizonyos konkrét szintaxisok és RDF tárolók **üres csomópont azonosítónak** (***blank node identifiers***) nevezett lokális hatáskörű azonosítókat használnak az üres csomópontok ábrázolásához.
 - Ezek mindig lokális hatáskörűek egy állományra vagy RDF tárolóra nézve és nem tartós vagy hordozható azonosítók üres csomópontokhoz.

RDF termek

- Az IRI-ket, literálokat és üres csomópontokat összefoglaló néven **RDF termek**nek nevezzük.
 - Az IRI-k, literálok és üres csomópontok különbözőek és megkülönböztethetőek.
 - Azaz az IRI-k, literálok és üres csomópontok halmaza páronként diszjunkt.

RDF hármas

- Egy **RDF hármas** (***RDF triple***) az alábbi komponensekből álló rendezett elemhármas:
 - **Alany** (***subject***): egy IRI vagy egy üres csomópont.
 - **Állítmány** (***predicate***): egy IRI, mely egy tulajdonságot jelöl.
 - **Tárgy** (***object***): egy IRI, literál vagy üres csomópont.
- Egy RDF hármas azt fejezi ki, hogy az alany és a tárgy által jelölt erőforrások az állítmány által jelölt kapcsolatban vannak.
 - A hármasok tehát binér relációkat képesek kifejezni.
- Egy RDF hármasnak megfelelő kijelentést **RDF kijelentésnek** (***RDF statement***) nevezzük.

Példa RDF hármaskra

- `<http://dbpedia.org/resource/Tim_Berners-Lee>`
`<http://dbpedia.org/ontology/birthDate>`
`"1955-06-08"^^`
`http://www.w3.org/2001/XMLSchema#date>`
- `<http://dbpedia.org/resource/Tim_Berners-Lee>`
`<http://dbpedia.org/property/birthName>`
`"Timothy John Berners-Lee"@en`
- `<http://dbpedia.org/resource/Tim_Berners-Lee>`
`<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>`
`<http://xmlns.com/foaf/0.1/Person>`

RDF szótárak (1)

- Egy RDF szótár RDF gráfokban való felhasználásra szánt IRI-k egy összessége.
- Egy RDF szótárat alkotó IRI-k gyakran ugyanazzal a karakterlánccal kezdődnek, melyet **névtér IRI**-nek nevezünk.
 - Bizonyos sorosítási formátumok lehetővé teszik az IRI-k **névtér előtag**okkal való rövidítését.
 - Például a <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> IRI `rdf:type` módon rövidíthető.

RDF szótárak (2)

- Néhány elterjedten használt RDF szótár:

RDF szótár	Névtér IRI	Névtér előtag
RDF	http://www.w3.org/1999/02/22-rdf-syntax-ns#	rdf
RDF Schema	http://www.w3.org/2000/01/rdf-schema#	rdfs
XML Schema	http://www.w3.org/2001/XMLSchema#	xsd
Dublin Core	http://purl.org/dc/elements/1.1/ http://purl.org/dc/terms/	dc dcterms
FOAF	http://xmlns.com/foaf/0.1/	foaf

RDF gráfok

- Egy **RDF gráf** RDF hármások egy halmaza.
 - Mivel halmaz, lényegtelen a hármások sorrendje.
- Egy RDF gráf jelentését az őt alkotó RDF hármások jelentésének összessége adja.
 - A hármások között konjunkció van (negáció és diszjunkció nem fejezhető ki).

RDF dokumentumok

- Egy RDF gráfot az RDF egy konkrét szintaxisával kódoló dokumentumok.
- Lehetővé teszik RDF gráfok rendszerek közötti cseréjét.

Az RDF konkrét szintaxisai

- **RDF/XML:**

- *RDF 1.1 XML Syntax* (W3C ajánlás, 2014. február 25.) <https://www.w3.org/TR/rdf-syntax-grammar/>

- **N-Triples:**

- *RDF 1.1 N-Triples – A line-based syntax for an RDF graph* (W3C ajánlás, 2014. február 25.) <https://www.w3.org/TR/n-triples/>

- **N-Quads:**

- *RDF 1.1 N-Quads – A line-based syntax for RDF datasets* (W3C ajánlás, 2014. február 25.) <https://www.w3.org/TR/n-quads/>

- **Turtle:**

- *RDF 1.1 Turtle – Terse RDF Triple Language* (W3C ajánlás, 2014. február 25.) <https://www.w3.org/TR/turtle/>

- **TriG:**

- *RDF 1.1 TriG – RDF Dataset Language* (W3C ajánlás, 2014. február 25.) <https://www.w3.org/TR/trig/>

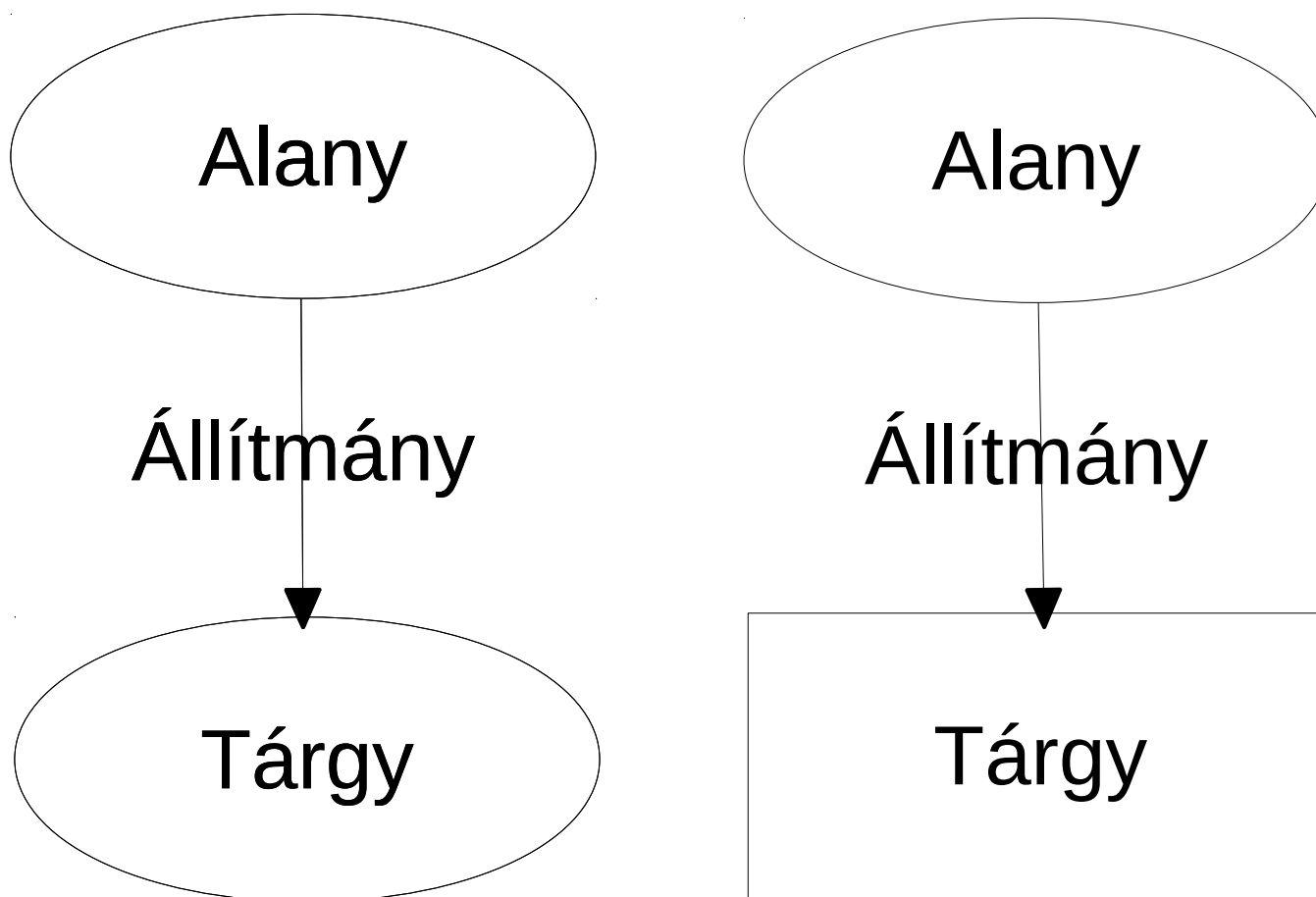
- **JSON-LD:**

- *JSON-LD 1.1 – A JSON-based Serialization for Linked Data* (W3C ajánlás, 2020. július 16.) <https://www.w3.org/TR/json-ld/>

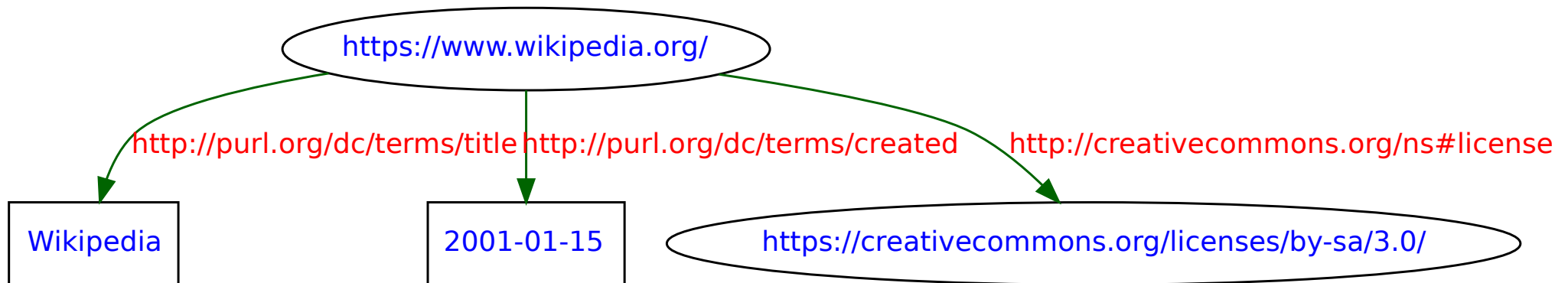
RDF gráfok vizuális ábrázolása (1)

- Egy RDF gráf egy olyan irányított gráffal ábrázolható, melyben:
 - A csúcsok halmazát a hármasok alanyai és tárgyai alkotják.
 - Az élek a hármasok állítmányainak felelnek meg, egy él az alanyról a tárgyra mutat.
- Ellipszisek ábrázolják az IRI-vel azonosított és az üres csomópontokat, téglalapok a literálokat.

RDF gráfok vizuális ábrázolása (2)



Példa (1)



Példa (2)

- Az RDF gráf az N-Triples szintaxissal ábrázolva:

```
<https://www.wikipedia.org/> <http://purl.org/dc/terms/title>  
"Wikipedia"@en .  
<https://www.wikipedia.org/> <http://purl.org/dc/terms/created>  
"2001-01-15"^^<http://www.w3.org/2001/XMLSchema#date> .  
<https://www.wikipedia.org/> <http://creativecommons.org/ns#license>  
<https://creativecommons.org/licenses/by-sa/3.0/> .
```

Példa (3)

- Az RDF gráf az RDF/XML szintaxissal ábrázolva:

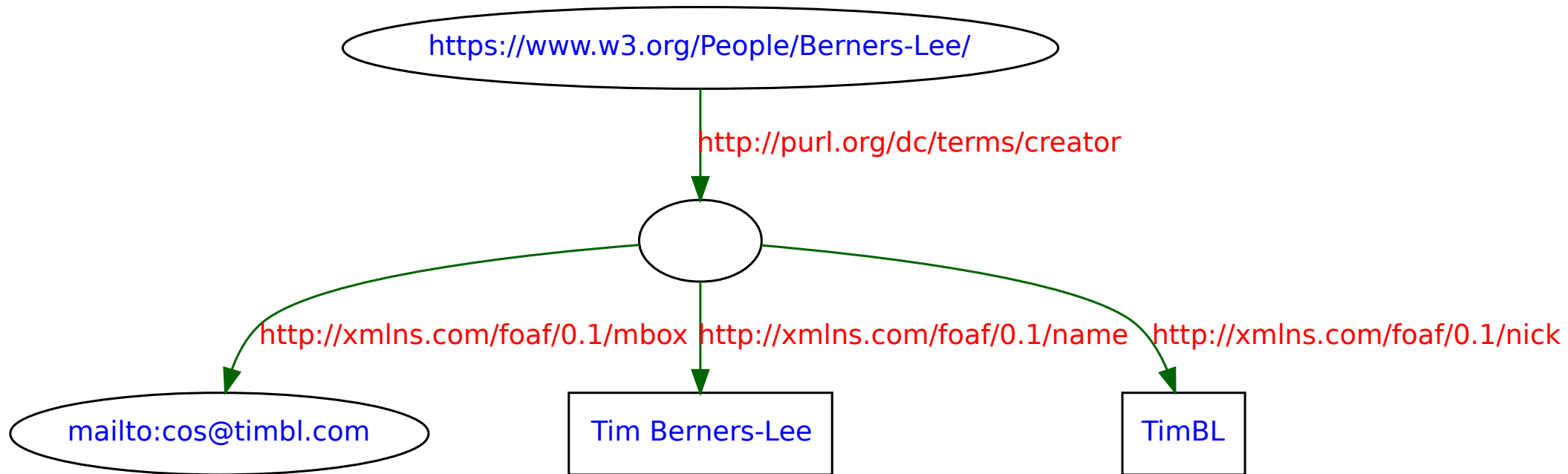
```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cc="http://creativecommons.org/ns#">
  <rdf:Description rdf:about="https://www.wikipedia.org/">
    <dcterms:title xml:lang="en">Wikipedia</dcterms:title>
    <dcterms:created
rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2001-01-15
    </dcterms:created>
    <cc:license
rdf:resource="http://creativecommons.org/licenses/by-sa/3.0/" />
  </rdf:Description>
</rdf:RDF>
```

Példa (4)

- Az RDF gráf Turtle szintaxissal ábrázolva:

```
@prefix cc: <http://creativecommons.org/ns#> .  
@prefix dcterms: <http://purl.org/dc/terms/> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
  
<https://www.wikipedia.org/>  
  dcterms:title "Wikipedia"@en ;  
  dcterms:created "2001-01-15"^^xsd:date ;  
  cc:license <http://creativecommons.org/licenses/by-sa/3.0/> .
```

Példa: üres csomópont (1)



Példa: üres csomópont (2)

- Az RDF gráf az N-Triples szintaxissal ábrázolva:
 - Az üres csomópontok jelölése `_:xyz` formájú üres csomópont címkékkel.

```
<https://www.w3.org/People/Berners-Lee/> <http://purl.org/dc/terms/creator> _:tbl .  
_:tbl <http://xmlns.com/foaf/0.1/mbox> <mailto:cos@timbl.com> .  
_:tbl <http://xmlns.com/foaf/0.1/name> "Tim Berners-Lee" .  
_:tbl <http://xmlns.com/foaf/0.1/nick> "TimBL" .
```


Példa: üres csomópont (3)

- Az RDF gráf az RDF/XML szintaxissal ábrázolva:

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

  <rdf:Description rdf:about="https://www.w3.org/People/Berners-Lee/">
    <dcterms:creator rdf:parseType="Resource">
      <foaf:mbox rdf:resource="mailto:cos@timbl.com"/>
      <foaf:name>Tim Berners-Lee</foaf:name>
      <foaf:nick>TimBL</foaf:nick>
    </dcterms:creator>
  </rdf:Description>

</rdf:RDF>
```

Példa: üres csomópont (4)

- Az RDF gráf a Turtle szintaxissal ábrázolva:

```
@prefix dcterms: <http://purl.org/dc/terms/> .  
@prefix foaf: <http://xmlns.com/foaf/0.1/> .  
  
<https://www.w3.org/People/Berners-Lee/> dcterms:creator [  
  foaf:mbox <mailto:cos@timbl.com> ;  
  foaf:name "Tim Berners-Lee" ;  
  foaf:nick "TimBL" ] .
```

RDF programkönyvtárak

- Szabad és nyílt forrású RDF programkönyvtárak:
 - **Java:**
 - *Apache Jena* (licenc: *Apache License 2.0*) <https://jena.apache.org/>
 - **.NET:**
 - *dotNetRDF* (licenc: *MIT License*) <https://www.dotnetrdf.org/>
<https://github.com/dotnetrdf/dotnetrdf>
 - **PHP:**
 - *EasyRdf* (licenc: *New BSD License*) <http://www.easyrdf.org/>
<https://github.com/easyrdf/easyrdf>
 - **Python:**
 - *RDFLib* (licenc: *New BSD License*) <https://rdflib.readthedocs.io>
<https://github.com/RDFLib/rdflib>

RDF tárolók (1)

- Egy RDF (*RDF store*, *triple store*) tároló egy olyan speciális célú adatbázis, mely nagy mennyiségű RDF hármas tárolására szolgál.
- RDF tárolók népszerűsége:
 - Lásd: *DB-Engines Ranking of RDF Stores*
<https://db-engines.com/en/ranking/rdf+store>

RDF tárolók (2)

- Nem szabad szoftverek:
 - *GraphDB* <https://graphdb.ontotext.com/>
 - *MarkLogic* <https://www.marklogic.com/>
 - *Stardog* <https://www.stardog.com/>
 - *Virtuoso Universal Server*
<https://virtuoso.openlinksw.com/>

RDF tárolók (3)

- Szabad és nyílt forrású szoftverek:
 - *Apache Rya* (programozási nyelv: Java; licenc: *Apache License 2.0*)
<https://rya.apache.org/>
 - *BrightstarDB* (programozási nyelv: C#; licenc: *MIT License*)
<https://brightstardb.com/> <https://github.com/BrightstarDB/BrightstarDB>
 - *Cayley* (programozási nyelv: Go; licenc: *Apache License 2.0*) <https://cayley.io/>
<https://github.com/cayleygraph/cayley>
 - *TDB* (programozási nyelv: Java; licenc: *Apache License 2.0*)
<https://jena.apache.org/documentation/tdb/>
 - Az Apache Jena komponense.
 - *Eclipse RDF4J* (programozási nyelv: Java; licenc: *New BSD License*)
<https://rdf4j.org/> <https://github.com/eclipse/rdf4j>
 - *Virtuoso Open-Source Edition* (programozási nyelv: C; licenc: *GPLv2*)
<http://vos.openlinksw.com/> <https://github.com/openlink/virtuoso-opensource>

SPARQL

- RDF gráfok lekérdezésére és manipulálására szolgál nyelv és protokoll.
- Jelenleg aktuális verziója az 1.1 számú.
- Specifikációk: összesen 11 W3C ajánlás (2013. március 21.)
 - Lásd:
https://www.w3.org/2009/sparql/wiki/Main_Page

SPARQL lekérdezések

- Egy SPARQL lekérdezés végrehajtása egy részgráfillesztési feladat megoldását jelenti egy RDF gráfban.
- A részgráfillesztési feladatok megadása gráfminták formájában történik.
- A részgráfillesztési feladatok megoldása a gráfmintákban előforduló változók helyettesítési értékeinek meghatározását jelenti.

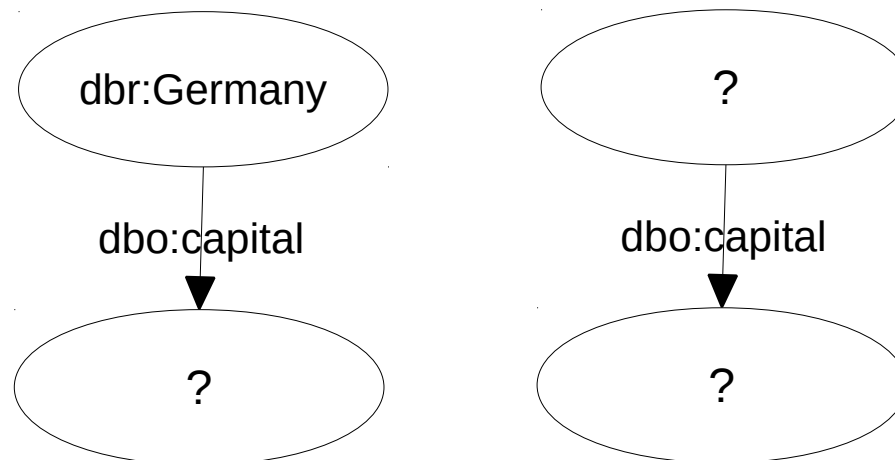
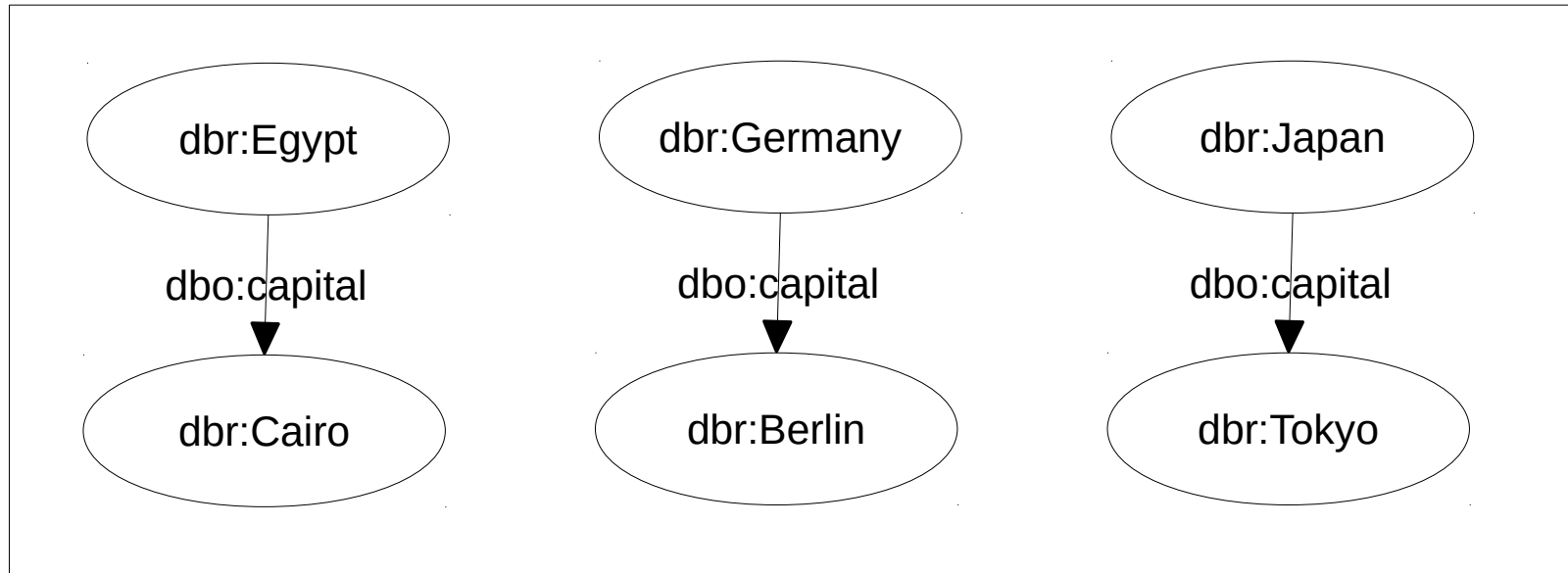
Gráfminták

- Egy gráfminta egy hármasokból áll halmaz, pont mint egy RDF gráf.
 - Ezek a hármasok azonban tartalmazhatnak változókat.
- A legegyszerűbb gráfminta egy olyan hármas, melyben az alany, az állítmány és a tárgy bármelyike lehet egy változó.
 - A változók pozíciója alapján az ilyen hármasoknak nyolc fajtája van.
 - A hármasok kombinálásával összetettebb gráfminták hozhatók létre.
 - A hármasok között közös változókkal valósítható meg kapcsolat.

Részgráfillesztési feladat

- Részgráfillesztés egyetlen hármas mintából álló gráfminta esetén:
 - Ha a hármas mintában nincs változó, akkor azt kell meghatározni, hogy a gráf tartalmazza-e a hármaszt.
 - Ha a hármas mintában egy vagy több változót tartalmaz, akkor azok lehetséges helyettesítési értékeit kell megadni.
- A részgráfillesztési feladatok megoldásai olyan függvények, melyek a gráfminták minden egyes változójához egy helyettesítési értékeket (egy RDF termet) rendelnek hozzá.
 - Egy megoldás helyettesítési értékeinek a gráfmintába helyettesítése az eredeti RDF gráf egy részgráfjával ekvivalens RDF gráfot kell, hogy adjon.
 - Egy változóból és helyettesítési értékéből álló rendezett elempárt **kötésnek** (*binding*) nevezünk.
 - A megoldások kötésekből álló halmazoknak tekinthetők.

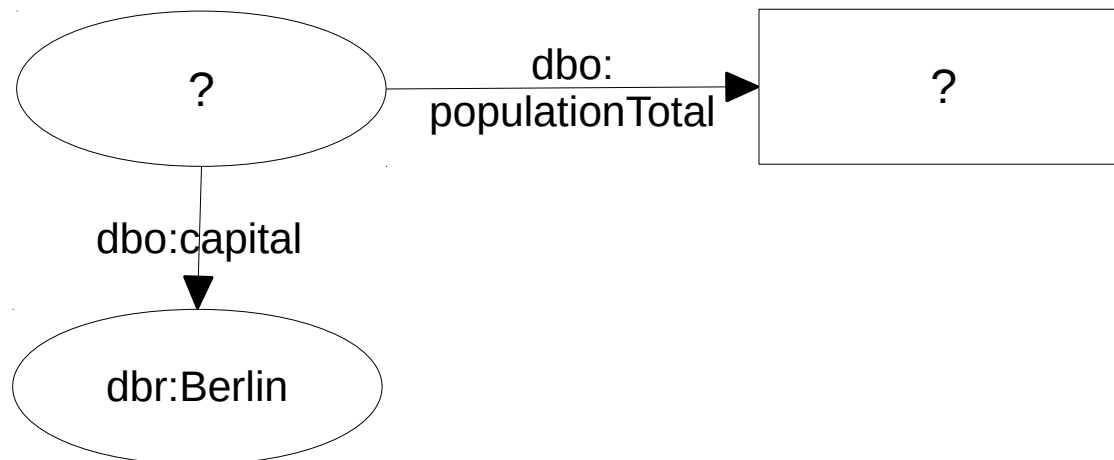
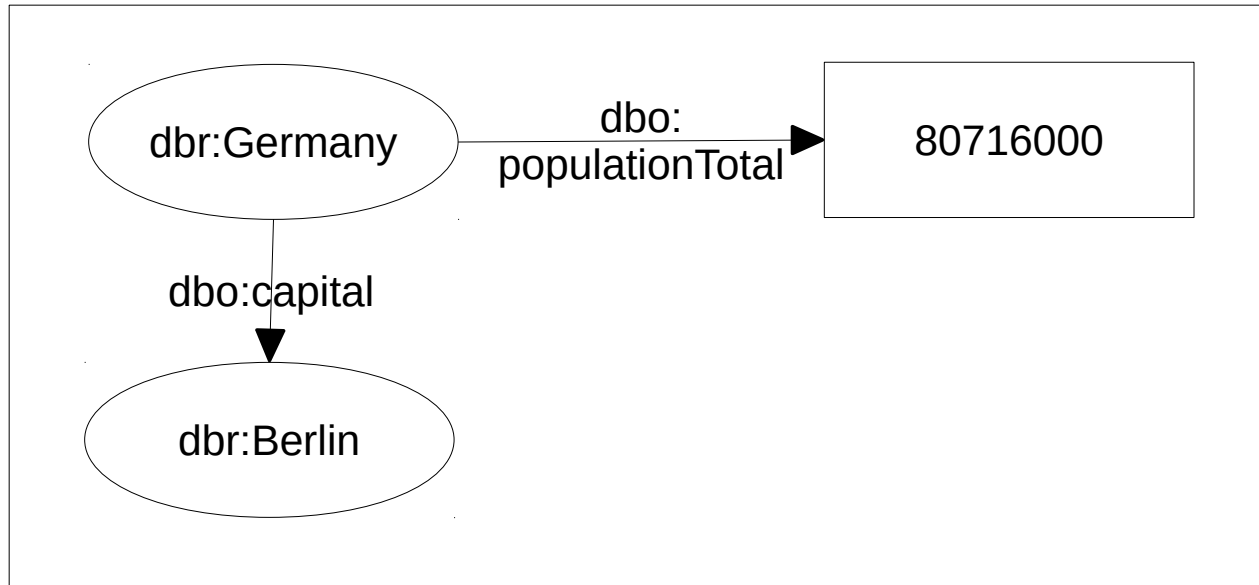
Példa (1)



Példa (2)

- Az ábrán látható gráfminták használata SPARQL lekérdezésekben:
 - PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT ?z
WHERE {
 dbr:Germany dbo:capital ?z .
}
 - PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT ?x ?z
WHERE {
 ?x dbo:capital ?z .
}
- Kipróbálás: <https://dbpedia.org/sparql>

Példa (1)



Példa (2)

- Az ábrán látható gráfminták használata SPARQL lekérdezésben:
 - ```
PREFIX dbr: <http://dbpedia.org/resource/>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT ?x ?z
WHERE {
 ?x dbo:capital dbr:Berlin .
 ?x dbo:populationTotal ?z .
}
```
  - A gráfmintát alkotó hármas minták között konjunkció van és közös változókon keresztül kapcsolódnak egymáshoz.
- Kipróbálás: <https://dbpedia.org/sparql>

# SPARQL implementációk

- *W3C Wiki – SparqlImplementations*  
<https://www.w3.org/wiki/SparqlImplementations>
  - *Apache Jena* (programozási nyelv: Java; licenc: *Apache License 2.0*) <https://jena.apache.org/>