



Egy W3C dokumentum

Szeretném felhívni a figyelmet egy W3C dokumentumra, amit néhány napja tett közzé az egyik szerző, az ismert könyvtáros szakértő *Karen Coyle*:

SHACL Core Abstract Syntax and Semantics
<https://www.w3.org/TR/2016/WD-shacl-abstract-syntax-20160825/>

A SHACL a Shapes Constraint Language rövidítése ami egy szabványtervezet, a fenti dokumentum ennek rövidített és olvashatóbb változata. A SHACL egy absztrakt nyelv amivel adatellenőrzéseket lehet definiálni, pl. egy bibliográfiai rekord adott mezője milyen értékeket vehet fel, lehet-e adott mezőt ismételni, milyen belső adatstruktúrát kell felvennie stb. A SHACL az RDF-re épül, ezért annak terminológiai készletét használja, de valójában nem csak szigorúan RDF alapú adatbázisok szerkezetése, hanem (néhány kivételtől eltekintve) relációs és NoSQL alapú adatszerkezetek esetében is remekül alkalmazható.

Egy példa:

szabály:

```
<NoActionIssueShape> sh:property [ sh:predicate ex:state; sh:in (ex:Resolved, ex:Rejected) ] .
```

magyarázat

<NoActionIssueShape> – ez a szabályunk azonosítója

sh:property – ... amire igaz a következő állítás

[;] – ez egy olyan állítás, aminek nincs azonosítója (ún. blank node) és két részből áll, amelyeket pontosvessző választ el

sh:predicate – a vizsgálandó elem az RDF triplet egyik predikátuma, vagyis az állítmány (relációs adatbázisban: a rekord egyik mezője)

ex:state – ez a kiválasztott predikátum (az példa névtér "state" nevű tulajdonsága)

sh:in – ez a tulajdonképpeni megkötés: a state tulajdonságértékének egy kötött listából kell származnia, minden ettől eltérő érték hibás

ex:Resolved, ex:Rejected – a példanévtérben definiált értékek.

vizsgált rekordelemek:

```
<issue1> ex:state ex:Resolved .
```

```
<issue2> ex:state ex:Unresolved .
```

Ez két egyszerű állításhármas. Az első esetben a state értéke Resolved, a másodikban Unresolved. A Resolved rendben is van, hiszen a szabályban definiált felsorolás egyik eleme, viszont a második, az Unresolved nincs benne a szabályunkban, tehát az ellenőrzés során ez hibát fog okozni.

Mi ebben az újdonság, hiszen különféle adatbázisokban létre lehet hozni ehhez hasonló szabályokat? A probléma ott van, hogy egyrészt a legtöbb adatbázisban ez csak elvi lehetőség és erősen technológiaspecifikus, másrészt amikor adatcseréről van szó (közös katalogizálás, aggregálás, rekordhonosítás stb.), akkor a rendszeren belül esetlegesen meghozott szabályokat szinte sosem viszik át az új rendszerekre. A SHACL egy nagyon jó és rugalmas alapszótárat biztosít tehát az adatellenőrzések definiálására, úgy, hogy azt a közgyűjteményi szakértő is, és a programozó is megértheti.

Király Péter közlése.

/Forrás:

<https://www.facebook.com/groups/312261069629/>

(Válogatta: B. Bné)