

LINKED OPEN DATA

VÎRLAN Ion

Universitatea Tehnică a Moldovei

Abstract: În acest articol este descrisă importanța procedurii de publicare a datelor într-un format deschis și accesibil, care contribuie la dezvoltarea profesională a multor specialiști din diferite domenii, rezultatul utilizării și distribuirii acestuia și motivarea de a utiliza și dezvolta acest procedeu cât mai mult. Se descrie lista celor cinci nivele de calitate a datelor, diferență dintre Linked Open Data și Open Data.

Cuvinte cheie: Linked Open Data, Open Data, Sisteme de Gestiune a Bazelor de Date (SGBD), SGBD-SQL, SGBD-NoSQL, URI, URL, URN, RDF.

1. Introducere

Multe organizații și instituții de pe întreg globul, odată cu începutul sec. XXI au început să publice datele într-un format deschis și accesibil pentru utilizatorii serviciilor web. Aceste acțiuni efectuate au creat pentru unii multe neclarități și întrebări de tipul: ”Din ce cauză întreprinderile, instituțiile de cercetare și alte organizații au descoperit la toți secretele lor personale obținute cu greu?”. Pe când acestea au fost făcute în urma multor analize și elaborări a diferitor strategii de transparență și economie pentru dezvoltări profesionale și economice.

Deci, în urma publicărilor informațiilor secrete pentru studiere și dezvoltare, au apărut un șir de probleme de tipul: structurarea datelor prin legări și relații între ele, specificarea tipurilor de acces, diferențierea datelor cu același nume, etc. În continuarea acestui articol vor fi relatate aceste probleme împreună cu soluțiile lor.

2. Relatarea temei

Ideea de date înlănțuite deschise (Linked Open Data)

Primul care a propus ideea de date înlănțuite deschise a fost Tim Berners-Lee conform a 4 principii, care stimulează principiile de bază web, pentru accesul la date:

- Aplicarea universală a unui identificator al resurselor (URI) în calitate de numele obiectelor
- Aplicarea identificatorilor HTTP URI pentru realizarea apelului după aceste nume
- Punerea la dispoziție a informației utile necesare și actuale utilizatorilor care se adresează prin URI cu ajutorul standardelor (RDF*, SPARQL)
- Includerea link-urilor la alte identificatoare de resurse universale pentru a obține mai multe detalii

Ce este Linked Open Data?

Este o metoda de publicare a datelor structurate, în așa fel încât acestea să fie înlănțuite între ele în formatul RDF conform semanticii web (Semantic Web) pentru a devenii ușor accesibile și utile. Datele înlănțuite, de fapt, sunt principalele posibilități propuse de semantica web. Acestea utilizează resursă unică de identificare Uniform Resource Identifier (URI), acesta la rândul său este format din URL și URI.

Lista celor 5 nivele de calitate a datelor

- a) Informația este disponibilă pe internet sub o licență deschisă.
- b) Informația este disponibilă ca date structurate care pot fi citite automat.
- c) Ca și la (b) plus: Informația este disponibilă în formatul non-proprietar.
- d) Ca și la (c) plus: Informația este disponibilă în standardele deschise (RDF și SPARQL) destinată pentru identificarea datelor.
- e) Ca și la (d) plus: În informație mai există și link-uri la altă informație luând în considerare contextul acesteia pentru a propune mai multe detalii care tot la fel sunt structurate.

Diferența dintre Linked Open Data și Open Data

Open Data - datele pot fi publicate ca să fie în mod public sub o licență deschisă, fără posibilitatea de conectare la alte surse de date. *Date clasificate cu 1, 2 sau 3 stele sunt denumite de obicei date deschise (Open Data)*

Linked Open Data - datele care propun posibilitatea de conectare prin URI la alte surse de date utilizând standardele semanticii RDF fără ai fi disponibilă licența deschisă publicului, în caz de necesitate. Datele clasificate cu 4 sau 5 stele sunt numite date deschise legate.

Necesitatea procedurii Linked Open Data

1. Prezența diversității datelor cu același nume
2. Multiplitate a datelor (limbă, tip, format, etc.)
3. Necesitatea de a accesa datele actuale: rapid și util
4. Eficiența gestionării datelor
5. Accesarea diferitor date din orice loc de pe glob cu permisiunea necesară
6. Diversitatea necesităților de accesare, editare, adăugare și ștergere.

Soluțiile acestor probleme

Soluțiile pentru problemele care au apărut în urma multiplicității datelor diverse dar cu anumite caracteristici coincidente, prevăd diversificarea tuturor datelor după anumite standarde, regulamente, teorii, identificatori unici și structurarea acestora după diferite tipuri de acces. Aceste lucruri se realizează de către diferite sisteme de gestiune care respectă necesitățile impuse pentru realizarea procedurii de date înlănțuite deschise. La începutul aparițiilor acestor probleme, ca soluționare se utilizau sistemele de gestiune a bazelor de date care au la bază limbajul relațional SQL bazat pe tabele și relații între ele, iar mai apoi au apărut și se utilizează la nivelurile înalte (Motorul de căutare Google, Google Maps, Facebook, etc.) pentru Big Data, SGBD-uri NoSql (mai întâi numit "Nu SQL", dar mai apoi "Nu doar SQL") care au la bază teoria grafurilor în sistemul Neo4j. Dar ambele aceste mari compartimente de SGBD-uri SQL și NoSQL se completează una pe alta și în continuare va fi necesară cunoașterea ambelor pentru a le utiliza în combinație la necesități speciale.

3. Concluzii

Acest procedeu, care deja funcționează la nivel mondial este foarte necesar, util, eficient și actual, de aceea consider că este una din responsabilitățile de bază a fiecărui stat, întreprinderi, firme, organizații să contribuie la dezvoltarea, îmbunătățirea acestuia, deoarece el este foarte important și strategic la formarea și dezvoltarea societății contemporane. Proverbul: "Unde-i unul – nu-i putere, unde-s mulți – puterea crește", sugerează importanța afirmației de mai sus, fiindcă munca de unul singur este grea și anevoioasă cu rezultate greu (uneori greșit) obținute, în timp ce studierea, colaborarea științifică și informațională la nivel global aduce rezultate frumoase, relativ rapide, deosebite și de succes.

Bibliografie

1. Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., & Tullo, C. (2007). Unlocking the potential of public sector information with Semantic Web technology. In: The 6th International Semantic Web Conference (ISWC), 11-15 Nov 2007, Busan, Korea.
2. Berners-Lee, T. (2006). Linked Data - Design Issues. Retrieved June 5, 2013, <http://www.w3.org/DesignIssues/LinkedData.html>
3. Blanchard, B.S. & Fabrycky, W.J. (2006). Systems Engineering and Analysis, Fourth Edition. Prentice Hall. p. 19. Brink, van den, L. Overbeek, H., Brentjens, T. (2013). Designing A URI Strategy For The Dutch Public Sector, in Pilot Linked Open Data – Part 2.