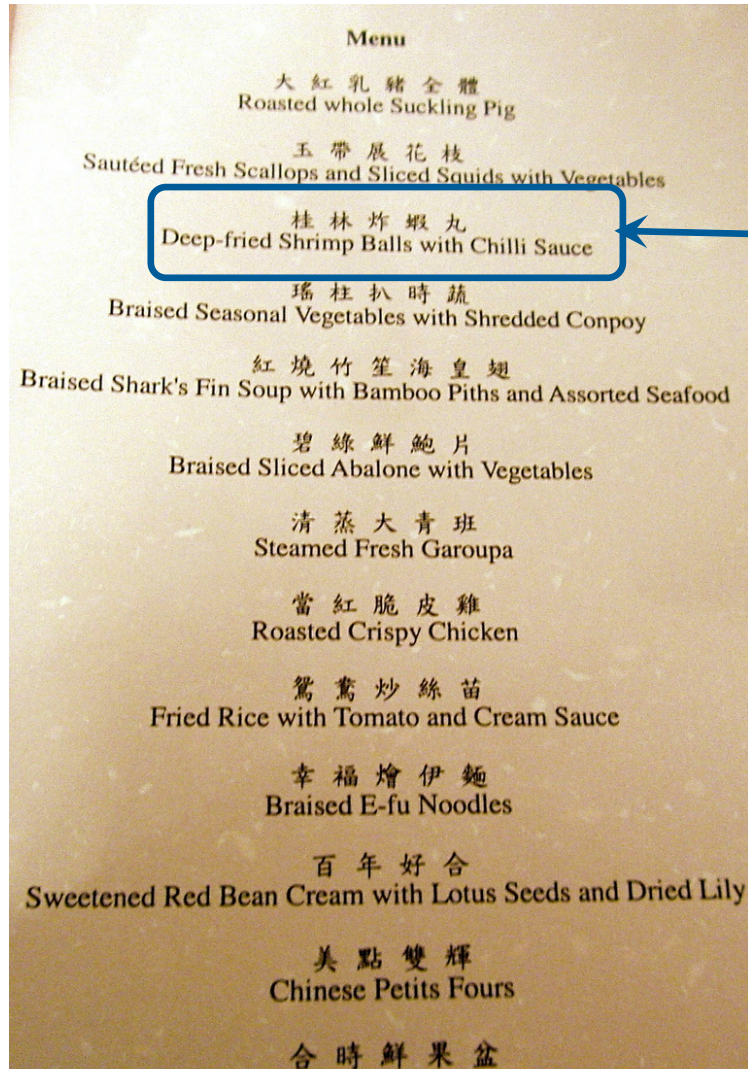


KRATKO PODSEĆANJE NA KLJUČNE IDEJE I KARAKTERISTIKE WEB-A PODATAKA

Osnovna ideja:

omogućiti programima razumevanje Web sadržaja time što će se ti sadržaji opisati mašinski razumljivim jezikom



I nama je potrebno da imamo opis dat na poznatom jeziku

Web of Data

- Glavne karakteristike:
 - Podaci (na Web-u) su **strukturirani** i **interlinkovani**
 - **Semantika** podataka i linkova je **eksplicitno data**
 - Omogućeno je izvršavanje **složenih upita** nad više izvora
 - Vizija Web-a kao jedne **gigantske globalne baze podataka**

Teme za današnji čas

- Dodavanje strukturiranih podataka u Web stranice
 - strukturirani podaci = podaci sa dobro definisanom strukturom + eksplicitno definisanim značenjem (semantikom)
 - RDFa, Microdata, JSON-LD
 - Schema.org, Open Graph Protocol
- Linkovanje podataka na Web-u
 - Principi publikovanja linkovanih podataka na Web-u
 - Linked data star scheme
 - Linked Open Data (LOD) Cloud
 - Linked Open Vocabularies (LOV)

DODAVANJE STRUKTURIRANIH PODATAKA U WEB STRANICE

Dodavanje strukturiranih podataka u Web stranice

- Veliki broj Web stranica već sadrži strukturirane podatke sa eksplicitno definisanim značenjem
- Pogledaćemo nekoliko primera koristeći Google-ov [Structured Data Testing tool](#); na primer:
 - filmove na [RottenTomatoes.com](#)
 - muzičare i muzičke grupe na [Last.fm](#)
 - proizvode na [BestBuy.com](#)
 - recepte na [AllRecipes.com](#)

Dodavanje strukturiranih podataka u Web stranice

- Structured data testing tool
 - omogućuje web administratorima uvid u strukturirane podatke koji su dostupni programima koji pristupaju datoj Web stranici
 - ali, ne omogućuje direktan, programski pristup tim podacima, tj. njihovu ekstrakciju iz date Web stranice
- Microdata Distiller alat
 - omogućuje programski pristup strukturiranim podacima umetnutim u Web stranice
 - osnovni benefit: jednostavna ekstrakcija podataka iz Web stranice – bez screen scraping-a ili nekih sličnih pristupa – radi korišćenja tih informacija u vašem programu
 - može se pozivati kao RESTful servis ili preuzeti i instalirati na lokalnoj mašini

Dodavanje strukturiranih podataka u Web stranice

- Za dodavanje strukturiranih podataka u Web stranice, potrebni su nam:
 - RDFS vokabulari koji će omogućiti opisivanje sadržaja Web stranica u mašinski razumljivom formatu
 - Način da proširimo HTML jezik tako da mašinski razumljivi opisi podataka budu sastavni deo Web stranice
- Da bi odgovorili na 1. zahtev, možemo koristiti Schema.org ili neki drugi RDFS vokabular
- Da bi odgovorili na 2. zahtev, možemo koristiti RDFa, Microdata ili JSON-LD – W3C preporuke za proširenje HTML jezika mašinski razumljivim opisima podataka

Schema.org

- Vokabular za opisivanje podataka u mašinski razumljivom obliku; trenutno, najzastupljeniji vokabular na Web-u
- Inicijativa potekla od velikih Web kompanija: Google, Yahoo, Microsoft (Bing), Yandex
- Dalje se razvija kao community effort u okviru Web konzorcijuma: <http://www.w3.org/wiki/WebSchemas>
- Inicijalno omogućavao opis svega nekoliko osnovnih tipova sadržaja, vremenom se taj broj značajno uvećao
 - lista svih tipova koje Schema.org trenutno podržava: <http://schema.org/docs/full.html>

Schema.org

- Par osnovnih statistika o korišćenju Schema.org (odnose se na početak 2014. god):
 - ~15% Web stranica koje indeksiraju glavni Web pretraživači sadrže podatke opisane Schema.org vokabularom;
 - preko 5M Web sajtova koristi ovaj vokabular;
 - za dodatne statistke i ostale informacije o Schema.org, pogledati [ove slajdove](#)

Schema.org

■ Preporuka:

- pogledati keynote Ramanathan Guha-e – lidera [W3C WebSchemas](#) grupe – na temu strukturiranih podataka na Web-u, Schema.org, kao i razvoja, primene i značaja otvorenih tehnologija za semantičko opisivanje Web sadržaja:

http://videolectures.net/iswc2013_guha_tunnel/

- takođe, interesantan i koristan može biti i intervju sa Guha-om publikovan na SemanticWeb.com blogu:

http://semanticweb.com/schema-org-chat-googles-r-v-guha_b40607

RDFa, Microdata, JSON-LD

- W3C preporuke (standardi) za dodavanje strukturiranih podataka u HTML stranice:
 - RDFa:
 - Specifikacija: <http://www.w3.org/TR/xhtml-rdfa-primer/>
 - Relevantne informacije, primeri koda, dodatni materijali: <http://rdfa.info/>
 - Microdata:
 - Specifikacija: <http://dev.w3.org/html5/md/>
 - JSON-LD:
 - Specifikacija: <http://www.w3.org/TR/json-ld/>
 - Relevantne informacije, primeri koda, dodatni materijali: <http://json-ld.org/>
 - Schema.org sajt za svaku klasu daje bar jedan (obično i više) primera Schema.org opisa korišćenjem ova 3 standarda

Još neki relevantni vokabulari

■ Schema Actions

- Skup klasa i atributa (properties) definisanih u okviru Schema.org sa ciljem da se omogući
 - opisivanje akcija koje neki Web sajt omogućuje svojim korisnicima, i kako se te akcije mogu programski inicirati
 - integrisanje podataka o akcijama korisnika na različitim sajtovima na Web-u
- Za više informacija, pogledati:
 - dokument koji opisuje Schema Actions i obezbeđuje instrukcije za njihovo korišćenje ([link](#))
 - članak koji objašnjava značaj Schema Actions ([link](#)), kao i članak koji ilustruje primenu u domenu muzike ([link](#))

Još neki relevantni vokabulari

■ GoodRelations

- Vokabular za opisivanje proizvoda, ponuda, prodavnica i sl.
- Ima široku primenu u domenu elektronske trgovine
 - koristeći Google-ov Structured Data Testing Tool pogledati primenu ovog vokabulara u stranicama Kmart.com, Sears.com, BestBuy.com
- Razvijeni su brojni alati koji omogućuju jednostavno opisivanje podataka ovim vokabularom
 - pogledati: <http://wiki.goodrelations-vocabulary.org/Tools>
- Ovaj vokabular je takođe sastavni deo Schema.org
 - <http://schema.org/Product> ; <http://schema.org/Offer> ...

Još neki relevantni vokabulari

■ Open Graph Protocol (OGP)

- Vokabular koji je uveo Facebook kako bi omogućio prikupljanje dodatnih informacija o sadržajima koje korisnici Like-uju na Web-u
 - OGP vokabular u kombinaciji sa RDFa standardom za proširenje HTML-a, obezbeđuje eksplicitnu semantiku Like-ovanih sadržaja
 - Tako prikupljene informacije Facebook koristi kao input za dalji razvoj svog Entity Graph-a
- OGP omogućuje opis različitih tipova sadržaja popularnih među korisnicima Web-a, kao što su muzika, knjige, video zapisi, profili korisnika i sl

Alati za rad sa strukturiranim podacimana Web-u

- Google je razvio više alata namenjenih
 - dodavanju strukturiranih podataka u Web stranice
 - praćenju korišćenja stanica sa umetnutim podacima,
 - detektovanju grešaka u podacima
- Ti alati su:
 - Structured Data Dashboard ([link](#))
 - Data Highlighter ([link](#))
 - Structured Data Markup Helper ([link](#))
- Video sa Google IO 2013 konferencije ([link](#)) opisuje ove alate i objašnjava njihovu namenu i korišćenje

Alati za rad sa strukturiranim podacima na Web-u

- Popularne Web platforme koje podržavaju RDFa/Microdata
 - Drupal
 - podrška za RDFa je deo Drupal-ovog core modula (od v.7);
 - naredna verzija (v.8) će uključiti Schema.org kao jedan od svojih osnovnih modela podataka
 - Webnodes
 - obezbeđuju punu podršku za rad sa Microdata i Schema.org (pogledati [ovaj članak](#))
 - WordPress
 - Obezbeđuje više proširenja za rad sa RDFa, Microdata, Schema.org (pogledati, npr., [ovu listu](#))

Nekoliko primera primene

■ Rich Snippets

- Bogatiji prikaz rezultata pretrage na Google-u za stranice koje sadrže umetnute strukturirane podatke
- npr., potražite na Google.com bilo koji film, ili aplikaciju za mobilni uređaj ili projekat sa Sourceforge-a

■ Interactive Snippets

- nova opcija koja se nudi u rezultatima pretrage Yandex pretraživača; npr., moguće je odmah uraditi check-in na avio letu; pogledati [ovaj članak](#) za više informacija

■ Pinterest's Rich Pins

- Pins sa dodatnim mogućnostima; npr., za proizvode obezbeđuju informacije o trenutnoj ceni, raspoloživosti, aktuelnim popustima
- pogledati, na primer, kako se strukturirani podaci koriste za generisanje rich pins za proizvode ([link](#))

Web Data Commons

- Publikuje godišnje izveštaje o raspoloživosti strukturiranih podataka na Web-u
- Izveštaji se zasnivaju na analizi ogromnog, javno dostupnog Web korpusa koji obezbeđuje Common Crawl
- Poslednji izveštaj je raspoloživ na sledećem linku:
<http://webdatacommons.org/structureddata/index.html#results-2013-1>