

STRUKTURIRANI PODACI U WEB STRANICAMA: RDFa & MICRODATA

JELENA JOVANOVIĆ

JELJOV@GMAIL.COM

[HTTP://JELENAJOVANOVIĆ.NET](http://JELENAJOVANOVIĆ.NET)

PRIMERI

Podaci o proizvodima umetnuti u stranice Web sajtova za online prodaju

- Sears: [stranica Apple iPad3](#)
- KMart: [stranica Lego Ninjago Battle Box](#)
- BestBuy:
[stranica Canon PowerShot A2300 digital camera](#)
- Za dodatne primere pogledati:
<http://wiki.goodrelations-vocabulary.org/Datasets>

PRISTUPI ZA “UMETANJE” STRUKTURIRANIH PODATAKA U HTML STRANICE

- Meta tagovi
- Microformats (<http://microformats.org/>)
- RDFa (<http://www.w3.org/TR/xhtml-rdfa-primer/>)
- Microdata (<http://dev.w3.org/html5/md/>)

META TAGOVI

```
<html>
<head>
  <meta name="dcterms.identifier" content=
http://www.metoffice.gov.uk/public/weather/forecast/city-of-london>
  <meta name="dcterms.title" content="City of London forecast">
  <meta name="dcterms.creator" content="Met Office">
  <meta name="dcterms.created" content="2013-01-29">
  <meta name="dcterms.publisher" content="Met Office, FitzRoy Road,
  Exeter, EX1 3PB, 0870 900 0100, enquiries@metoffice.gov.uk">
  <meta name="dcterms.subject" content="Meteorology">
  <meta name="dcterms.subject" content="Climate and weather">
  <meta name="dcterms.subject" content="Weather forecasts">
  <meta name="dcterms.coverage" content="UK">
  <meta name="dcterms.language" content="eng">
  <!-- the rest of the page-->
</html>
```

MICROFORMATS

Neki od najpopularnijih mikroformata:

- hCard (<http://microformats.org/wiki/hcard>)
- hCalendar (<http://microformats.org/wiki/hcalendar>)
- hReview (<http://microformats.org/wiki/hreview>)
- xfn (<http://gmpg.org/xfn/>)

Primeri primene:

- Last.fm, Food.com,...
- Microformats in the Wild @ Microformats.org

RDFA

Primer: podaci o osobi

```
<div vocab="http://schema.org/" typeof="Person">
  <a property="image" href="http://manu.sporny.org/images/manu.png">
    <span property="name">Manu Sporny</span></a>,
  <span property="jobTitle">Founder/CEO</span>
  <div>
    Phone: <span property="telephone">(540) 961-4469</span>
  </div>
  <div>
    E-mail: <a property="email" href="mailto:msporny@digitalbazaar.com">
      msporny@digitalbazaar.com</a>
  </div>
  <div>
    Links: <a property="url" href="http://manu.sporny.org/">
      Manu's homepage</a>
  </div>
</div>
```

[Manu Sporny](#), Founder/CEO
Phone: (540) 961-4469
E-mail: msporny@digitalbazaar.com
Links: [Manu's homepage](http://manu.sporny.org/)

RDFa (IN META TAGS)

```
<html
  xmlns:og="http://ogp.me/ns#"
  xmlns:fb="http://www.facebook.com/2008/fbml">
  <meta property="og:url"
    content="http://www.imdb.com/title/tt1907668/" />
  <meta property='og:image'
    content="http://ia.media-imdb.com/images/M/MV5BMTUx
MjI1OTMxN15BM15BanBnXkFtZTcwNjc3NTY1OA@@._V1_SX32_CR0,0,32,44_.jpg"/>
  <meta property='og:type' content="video.movie" />
  <meta property='og:title' content="Flight (2012)" />
  <meta property='og:site_name' content='IMDb' />
  <meta property='fb:app_id' content='115109575169727' />

  <!-- the rest of the page-->
</html>
```

MICRODATA

Primer: podaci o filmu

```
<div itemscope itemtype="http://schema.org/Movie">  
<a itemprop="url" href="http://www.imdb.com/title/tt1674773/">  
  <div itemprop="name"><strong>Hannah Arendt</strong></div>  
</a>  
<div itemprop="description">A look at the life of philosopher and political  
theorist Hannah Arendt, who reported for The New Yorker on the war crimes  
trial of the Nazi Adolf Eichmann. </div>  
<div itemprop="director" itemscope itemtype=http://schema.org/Person>  
<b>Directed by</b>: <span itemprop="name">Margarethe von Trotta</span></div>  
<div><b>Starring</b>: <div itemprop="actors" itemscope itemtype=  
http://schema.org/Person> <span itemprop="name">Barbara Sukowa,</span></div>  
<div itemprop="actors" itemscope itemtype=http://schema.org/Person> <span  
itemprop="name">Axel Milberg</span></div>  
</div>
```

[Hannah Arendt](#)

A look at the life of philosopher and political theorist Hannah Arendt, who reported for The New Yorker on the war crimes trial of the Nazi Adolf Eichmann.

Directed by: Margarethe von Trotta

Starring: Barbara Sukowa, Axel Milberg

VOKABULARI

- Schema.org (Google, Yahoo, Bing, Yandex)
 - RDF verzija: <http://schema.rdfs.org/>
 - Proširenja Schema.org:
 - [GoodRelations](#)
 - [rNews](#)
 - [lista proširenja Schema.org](#)
- Open Graph protocol (Facebook): <http://ogp.me/>
- RDFS vokabulari
 - [Dublin Core](#)
 - [Friend Of A Friend \(FOAF\)](#)

ALATI ZA EKSTRAKCIJU PODATAKA IZ WEB STRANICA

- Alati za uvid u podatke
 - [Google's Structured data testing tool](#)
 - [Microdata Parser](#)
 - [Green Turtle RDFa](#) (Chrome extension)
 - [Semantic Inspector](#) (Chrome extension)

ALATI ZA EKSTRAKCIJU PODATAKA IZ WEB STRANICA

- Alati za programski pristup podacima
 - [RDFa and Microdata Distiller](#) (RESTful)
 - [W3C Microdata to RDF Parser](#) (RESTful)
 - [Java-RDFa](#) (Java; Jena)
 - [RDFa ekstraktor iz Semarql biblioteke](#) (Java; Jena)
 - [Any23](#) (RESTful; Java)
 - [Parseri za Microformats](#)

ALATI ZA DODAVANJE STRUKTURIRANIH PODATAKA U WEB STRANICE

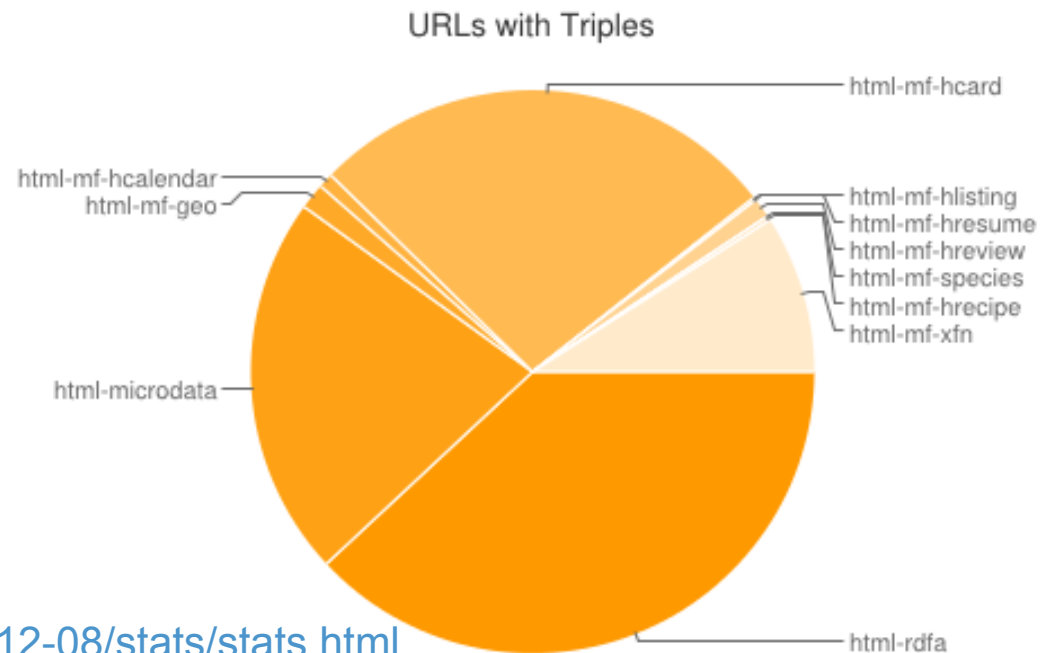
- [Drupal 7](#) - native podrška za RDFa
- Proširenja (extensions/plugin) za CMS sisteme poput Joomla, WordPress ; primeri:
 - [proširenja za popularne e-commerce platforme](#)
 - [Schema Creator WordPress plugin](#)
- Templejti za generisanje HTML koda sa umetnutim podacima; primeri:
 - [Schema Creator](#)
 - [Microdata Generator](#)
 - [FOAF RDFa Card Creator](#)

DODATNI PRIMERI PODATAKA U WEB STRANICAMA

- Novinski članci na sajtu [New York Times](#)-a (microdata)
- Video snimci na YouTube-u (RDFa, Microdata)
- Recepti i nutritivni podaci o jelima na sajtu Food.com (microformats, microdata, RDFa)
- Muzičke grupe, albumi, pesme na sajtu [Last.fm](#) (microdata, RDFa)
- Filmovi na sajtovima [Rotten Tomatoes](#) i [IMDB](#) (microdata, RDFa)
- ...

WEB DATA COMMONS

- [Web Data Commons](#) - pruža uvid u zastupljenost strukturiranih podataka u Web stranicama
- Poslednje statistike o strukturiranim podacima u Web stranicama ([link](#)) dobijene analizom CommonCrawl dataset-a iz Avgusta 2012. god



PRIMERI PRIMENE

- Google Rich Snippets
- Structured (faceted) search
 - primer: Google Recipes search ([link](#))
- Sistemi preporuke
- Generalno, sistemi zasnovani na podacima

Primer Rich Snippet-a:

[Norah Jones' YouTube channel](#)

www.youtube.com/user/norahjones

Norah Jones & Ted 'Everybody Needs A Best Friend' **norahjones** 58,178 views 2 weeks ago **Norah Jones & Ted** perform the Oscar-nominated song 'Everybody ...

▶ 2:11 [Norah Jones & Ted 'Everybody Needs A Best Friend'](#) 11 Feb 2013

▶ 3:22 [Norah Jones - Little Broken Hearts \(Album Trailer\)](#) 23 Apr 2012