



Publishing Linked Data in Web Pages

ME-E4300 Semantic Web, 27.1.2021

Eero Hyvönen

Aalto University, Semantic Computing Research Group (SeCo) <http://seco.cs.aalto.fi>

University of Helsinki, HELDIG

<http://heldig.fi>

eero.hyvonen@aalto.fi

Learning Objective

Understand how Linked Data is published on the Web to be used for practical applications





Distributed Approach: Embedding data in HTML pages

- Microformats
- Microdata
- RDFa

Centralized Approach: Standalone data services on servers



Embedding Data in HTML Pages: Microformats, Microdata and RDFa

Microformats

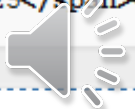
- For representing structured data embedded in HTML 4 web pages
- HTML tags are used for information representation
- Information represented with a set of tag attributes and their values
- Formats tailored at specific need
 - *E.g., hCard, hCalendar, hRecipe ...*
 - *Simplicity; but no common, cross-domain “semantics”*
- Structured data enables
 - *More focused search*
 - *More accurate classification and visualization of search results*
- Can be transformed into RDF with GRDDL (XSLT)
- Supeceded by Microdata in HTML 5



Example: hRecipe for food recipes

- hrecipe Metadata about a recipe
- fn Title of the recipe
- ingredient Ingredients

```
<p class="ingredient"><span class="value">125</span><span class="type">ml</span> milk</p>
```



- yield Quantity of production
- instructions Description of the process
- duration How long does it take
- photo Accompanying image
- ...



```
1279 <div class="ingredients" style="margin-top: 10px;">
1280 <h3>
1281   Ingredients</h3>
1282
1283   <ul>
1284
1285     <li class="plaincharacterwrap ingredient">
1286       1 pound uncooked pasta</li>
1287
1288     <li class="plaincharacterwrap ingredient">
1289       1 tablespoon olive oil</li>
1290
1291     <li class="plaincharacterwrap ingredient">
1292       2 cloves garlic, crushed</li>
1293
1294     <li class="plaincharacterwrap ingredient">
1295       1/2 cup chopped red onion</li>
1296
1297     <li class="plaincharacterwrap ingredient">
1298       1 pound skinless, boneless chicken breast meat - cut into bite-size pieces</li>
1299
1300     <li class="plaincharacterwrap ingredient">
1301       1 (14 ounce) can marinated artichoke hearts, drained and chopped</li>
1302
1303     <li class="plaincharacterwrap ingredient">
1304       1 large tomato, chopped</li>
1305
1306     <li class="plaincharacterwrap ingredient">
1307       1/2 cup crumbled feta cheese</li>
1308
1309     <li class="plaincharacterwrap ingredient">
1310       3 tablespoons chopped fresh parsley</li>
1311
1312     <li class="plaincharacterwrap ingredient">
1313       2 tablespoons lemon juice</li>
1314
1315
```


Searching "chicken pasta" with "lemon"



The screenshot shows a Google search for "chicken pasta" with approximately 92,000 results. The search filters are set to "Recipes with lemon". Two recipe results are visible:

- Easy Lemon Pasta with Chicken** (www.foodnetwork.com): 322 reviews, 35 mins. Ingredients include lemon, penne, chicken, pepper, garlic, olive oil, parsley, and grated cheese.
- Chicken Piccata Pasta Toss** (www.foodnetwork.com): 215 reviews, 30 mins. Ingredients include lemon, olive oil, chicken, pepper, butter, garlic, onion, and flour.

A red circle highlights the "Ingredients" filter section on the left, which includes a table:

Ingredients	Yes	No
lemon	<input checked="" type="checkbox"/>	<input type="checkbox"/>
penne	<input type="checkbox"/>	<input type="checkbox"/>
mint	<input type="checkbox"/>	<input type="checkbox"/>
...	<input type="checkbox"/>	<input type="checkbox"/>

Microdata



Extension of the HTML5 specification

- Primitives for information representation
- Simple semantics compared to the Semantic Web standards
- W3C working draft:
<https://www.w3.org/TR/microdata/>

<http://schema.org>

- Schemas for structured data mark-up on web
- Google, Yahoo, Microsoft
- Launched in June 2011



Example: Microdata



```
<section itemscope itemtype="http://data-vocabulary.org/Person">
  Hello, my name is
  <span itemprop="name">John Doe</span>,
  I am a
  <span itemprop="title">graduate research assistant</span>
  at the
  <span itemprop="affiliation">University of Dreams</span>.
  My friends call me
  <span itemprop="nickname">Johnny</span>.
  You can visit my homepage at
  <a href="http://www.JohnnyD.com" itemprop="url">www.JohnnyD.com</a>.
  <section itemprop="address" itemscope itemtype="http://data-vocabulary.org/Address">
    I live at
    <span itemprop="street-address">1234 Peach Drive</span>
    <span itemprop="locality">Warner Robins</span>
    ,
    <span itemprop="region">Georgia</span>.
  </section>
</section>
```

Embedding Data in web pages: <http://schema.org>



By Google, Yahoo, Bing, Yandex
Using Microdata and RDFa



The screenshot shows a web browser window displaying the schema.org website. The browser's address bar shows the URL <http://schema.org/docs/gs.html>. The page has a dark red header with the "schema.org" logo on the left and a search bar on the right. Below the header, there are navigation links for "Home", "Schemas", and "Documentation". The main content area features the heading "Getting started with schema.org" followed by an introductory paragraph. The text explains that while HTML tags tell browsers how to display information, they don't provide context, making it difficult for search engines to understand the content. It then states that schema.org provides a collection of shared vocabularies for webmasters to use. Below this, there is a list of topics for getting started with schema.org.

Getting started with schema.org

Most webmasters are familiar with HTML tags on their pages. Usually, HTML tags tell the browser how to display the information included in the tag. For example, `<h1>Avatar</h1>` tells the browser to display the text string "Avatar" in a heading 1 format. However, the HTML tag doesn't give any information about what that text string means—"Avatar" could refer to the hugely successful 3D movie, or it could refer to a type of profile picture—and this can make it more difficult for search engines to intelligently display relevant content to a user.

Schema.org provides a collection of shared vocabularies webmasters can use to mark up their pages in ways that can be understood by the major search engines: Google, Microsoft, and Yahoo!

You use the schema.org vocabulary, along with the [microdata format](#), to add information to your HTML content. While the long term goal is to support a wider range of formats, the initial focus is on Microdata. This guide will help get you up to speed with microdata and schema.org, so that you can start adding markup to your web pages.

1. [How to mark up your content using microdata](#)
 - a. [Why use microdata?](#)
 - b. [itemscope and itemtype](#)
 - c. [itemprop](#)
 - d. [Embedded items](#)
2. [Using the schema.org vocabulary](#)
 - a. [schema.org types and properties](#)
 - b. [Expected types, text, and URLs](#)
 - c. [Testing your markup](#)
3. [Advanced topic: Machine-understandable versions of information](#)
 - a. [Dates, times, and durations](#)
 - b. [Enumerations and canonical references](#)
 - c. [Missing/implicit information](#)
 - d. [Extending schema.org](#)

12

Schema.org type hierarchy (ontology) (only part)



Thing: description, image, name, url

CreativeWork: about, accountablePerson, aggregateRating, alternativeHeadline, associatedMedia, audio, author, awards, comment, contentLocation, contentRating, contributor, copyrightHolder, copyrightYear, creator, dateCreated, dateModified, datePublished, discussionUrl, editor, encodings, genre, headline, inLanguage, interactionCount, isFamilyFriendly, keywords, mentions, offers, provider, publisher, publishingPrinciples, reviews, sourceOrganization, thumbnailUrl, version, video

Article: articleBody, articleSection, wordCount

BlogPosting

NewsArticle: dateline, printColumn, printEdition, printPage, printSection

ScholarlyArticle

Blog: blogPosts

Book: bookEdition, bookFormat, illustrator, isbn, numberOfPages

ItemList: itemListElement, itemListOrder

Map

MediaObject: associatedArticle, bitrate, contentSize, contentURL, duration, embedURL, encodesCreativeWork, encodingFormat, expires, height, interactionCount, offers, playerType, regionsAllowed, requiresSubscription, uploadDate, width

AudioObject: transcript

ImageObject: caption, exifData, representativeOfPage, thumbnail

MusicVideoObject

VideoObject: caption, productionCompany, thumbnail, transcript, videoFrameSize, videoQuality

Movie: actors, director, duration, musicBy, producer, productionCompany, trailer

MusicPlaylist: numTracks, tracks

MusicAlbum: byArtist

MusicRecording: byArtist, duration, inAlbum, inPlaylist

Painting

Photograph

Recipe: cookTime, cookingMethod, ingredients, nutrition, prepTime, recipeCategory, recipeCuisine, recipeInstructions, recipeYield, totalTime

Review: itemReviewed, reviewBody, reviewRating

Sculpture

TVEpisode: actors, director, episodeNumber, musicBy, partOfSeason, partOfTVSeries, producer, productionCompany, trailer

TVSeason: endDate, episodes, numberOfEpisodes, partOfTVSeries, seasonNumber, startDate, trailer

TVSeries: actors, director, endDate, episodes, musicBy, numberOfEpisodes, producer, productionCompany, seasons, startDate, trailer

WebPage: breadcrumb, isPartOf, mainContentOfPage, primaryImageOfPage, significantLinks

RDFa



- W3C's recommended way to embed semantic information in an HTML page
- Schema.org supports RDFa 1.1
 - *RDF data can be embedded also as JSON-LD*
- Open Graph Protocol (OGP)
 - *Significant user of RDFa*
 - *Used in Facebook*
- Yahoo! research (Mika, Potter, 2012) found that 25% of web pages contain RDFa markup



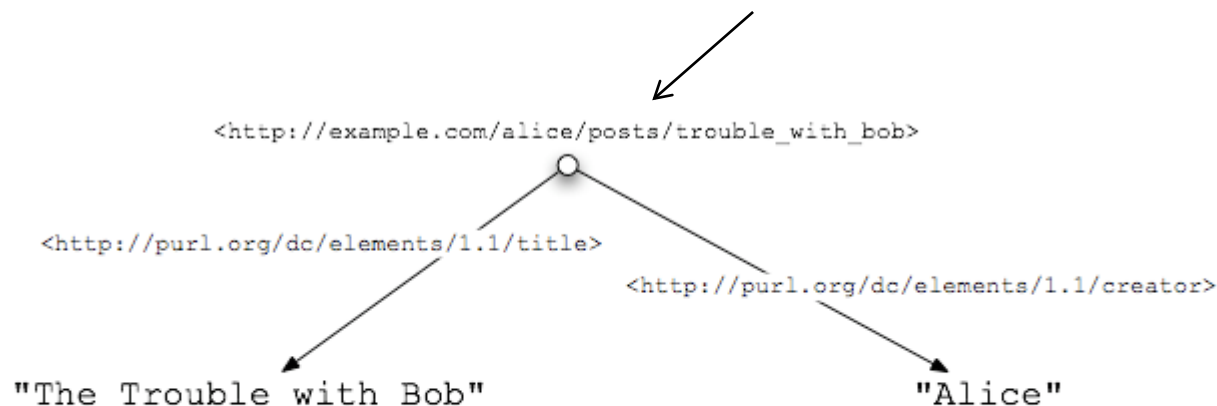
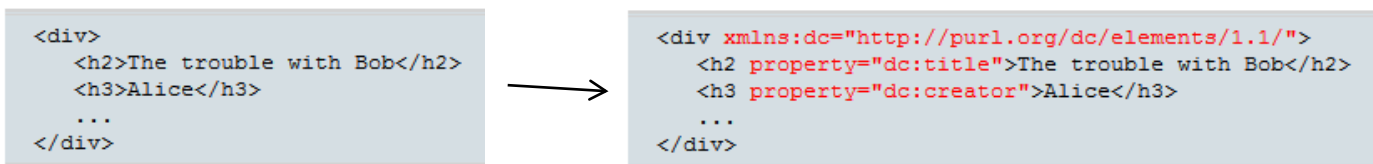
RDFa: bridging the human and data webs



On the left, what browsers see. On the right, what humans see. Can we bridge the gap so browsers see more of what we see?

- Bridge by embedding RDF in XHTML
- Software can then extract the RDF graph for the machine
- See also interactive RDFa “playground”

Example of using RDFa



Literal Properties: RDFa lets Alice connect not just one URL to another—for example to connect her blog entry URL to the Creative Commons license URL— but also to connect one URL to a string such as "The Trouble with Bob". All arrows are labeled with the corresponding property name, which is also a URL.

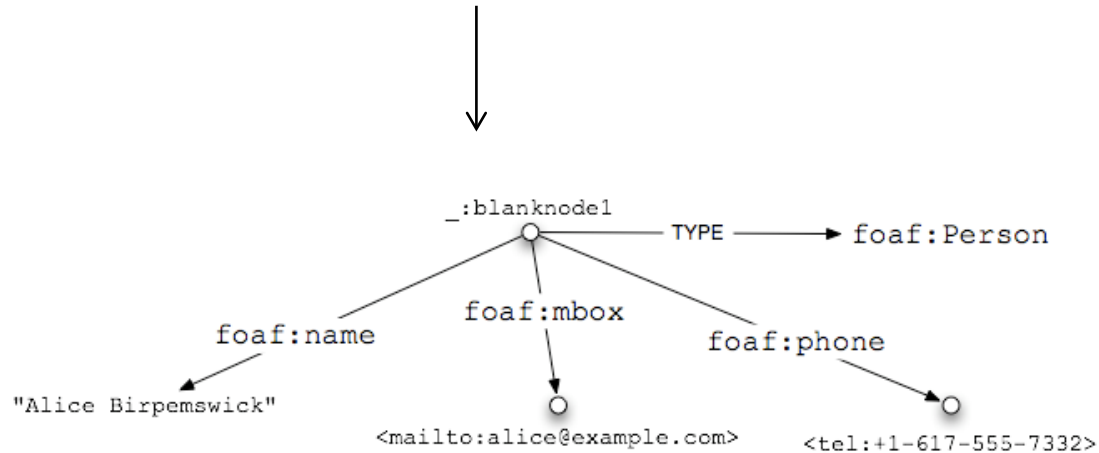
Another example



```
<div typeof="foaf:Person" xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <p property="foaf:name">
    Alice Birpemswick
  </p>

  <p>
    Email: <a rel="foaf:mbox" href="mailto:alice@example.com">alice@example.com</a>
  </p>

  <p>
    Phone: <a rel="foaf:phone" href="tel:+1-617-555-7332">+1 617.555.7332</a>
  </p>
</div>
```



Real-life example



RDFa 1.1 Distiller and Parser

- <http://www.w3.org/2012/pyRdfa/>
- Try it, e.g., with Svenska YLE news:
- <http://svenska.yle.fi/artikel/2014/01/21/krav-pa-egen-laptop-leder-till-ojamlikhet>



RDFa 1.1 Distiller and Parser

Warning: This version implements [RDFa 1.1 Core](#), including the handling of the [Role Attribute](#). The distiller can also run in XHTML+RDFa 1.0 mode (if the incoming XHTML content uses the RDFa 1.0 DTD and/or sets the `version` attribute). The [package available for download](#), although it may be slightly out of sync with the code running this service.

Distill by URI Distill by File Upload Distill by Direct Text Input

Distill by URI

URI:

Output Format:

Returned content:

Expand vocabularies:

Generate warnings for non RDFa 1.1 Lite usage:

▶ More (non-standard) options



Real-life example: web page



yle SVENSKA.YLE.FI ARENAN YLE FEM RADIO VEGA X3M BUU Google™ Anpassad sökning yle.fi

Huvudstadsregionen Svenska Yle

Inrikes Utrikes Sport Kultur och nöje Huvudstadsregionen Västnyland Åboland Österbotten Östnyland Väder Kontakta oss

"Krav på egen laptop leder till ojämlikhet"

Publicerad för ungefär 5 timmar sedan. Uppdaterad för ungefär 3 timmar sedan



Copyright YLE/Ted Urho

Hem och Skolas verksamhetsledare Micaela Romantschuk-Pietilä tror att kravet på att nya gymnasieelever [själva ska bekosta bärbar dator](#) kan leda till att en del väljer bort gymnasiet.

Nya artiklar

- [Berghäll blommade och fick pris!](#)
- [Nya ambulanser kom aldrig – Helsingfors säger upp kontrakt](#) -4
- [Nojor för eld i Tisdagssnacket](#)
- ["Krav på egen laptop leder till ojämlikhet"](#) -15
- [Bastu efterlyst av polisen](#)
- [Nyhetsöversikt svenska.yle.fi](#)

Relaterat

- [I år måste alla gymnasieelever ha egen laptop](#)
- [Lite bättre för Merituuleleverna](#)
- [Skrivstilen snart ett minne blott](#)
- [Sex alternativ för skolor och dagvård i södra Borgå](#)
- [Inget intresse för Solbergs skola](#)

Mest kommenterade

- [Staten satsar flera hundratusen euro på Guggenheim-tävling](#) -32
- [Minister om Guggenheimstöd: Finns ingen skyldighet att informera](#) -29
- [Guggenheim ser ut att få tomten i Södra kajen](#) -28

Real-life example: extracted RDF



```
1 http://svenska.yle.fi/artikel/2014/01/21/krav-pa-egen-laptop-leder-till-ojamlikhet<?xml version="1.0" encoding="utf-8"?>
2 <rdf:RDF
3   xmlns:foaf="http://xmlns.com/foaf/0.1/"
4   xmlns:sioc="http://rdfs.org/sioc/types#"
5   xmlns:dc="http://purl.org/dc/terms/"
6   xmlns:content="http://purl.org/rss/1.0/modules/content/"
7   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
8   xmlns:ns1="http://www.w3.org/1999/xhtml/vocab#"
9   xmlns:sioc="http://rdfs.org/sioc/ns#"
10 >
11 <sioc:Comment rdf:about="http://svenska.yle.fi/artikel/2014/01/21/krav-pa-egen-laptop-leder-till-ojamlikhet#comment-102347">
12   <dc:title xml:lang="sv">Sluta göra beslut som trycker</dc:title>
13   <sioc:reply_of>
14     <foaf:Document rdf:about="http://svenska.yle.fi/artikel/2014/01/21/krav-pa-egen-laptop-leder-till-ojamlikhet">
15       <dc:title xml:lang="sv">Krav på egen laptop leder till ojämlikhet</dc:title>
16       <dc:date rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-01-21T11:39:21+02:00</dc:date>
17       <sioc:num_replies rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">15</sioc:num_replies>
18       <rdf:type rdf:resource="http://rdfs.org/sioc/ns#Item"/>
19       <content:encoded xml:lang="sv">Hem och Skolas verksamhetsledare Micaela Romantschuk-Pietilä tror att kravet på att nya gym
20 I höst förutsätts alla som börjar gymnasiet att ha en egen bärbar dator, i och med den nya elektroniska studentexamen.
21 Micaela Romantschuk-Pietilä från Hem och Skola konstaterar att gymnasiestudier redan nu kostar en hel del för gymnasisterna, och at
22 - Det finns uträkningar om att läroböckerna för hela gymnasietiden kostar ungefär 1500 euro, och det kommer ju andra utgifter på. I
23 Ojämlikhet
24 Att gymnasieutbildningen innebär så höga kostnader kan påverka valet av andra stadiets utbildning, tror Romantschuk-Pietilä. Spela
25 - Redan nu är yrkesskolans utgifter för en elev mycket lägre, så jag är lite rädd för att de ganska höga kostnaderna i vissa fall l
26 Hem och Skolas verksamhetsledare konstaterar att det finns ett starkt samband mellan föräldrars akademiska bakgrund och elevernas r
27 - Jag tycker att vi borde komma ifrån att man på grund av ekonomiska orsaker i familjer väljer bort ett alternativ - det vill säga
28 "Borde få gratis dator!"
29 Romantschuk-Pietilä tycker inte att det är orimligt att tänka sig att gymnasieeleverna gratis skulle få en bärbar dator, som de ju
30 - Man vet att gymnasiestudier för staten och kommunen är ett billigare alternativ. En gymnasieelev kostar ungefär sjutusen euro, me
31 Ett argument som tas upp när man motiverar varför det är rimligt att gymnasieelever har egen laptop är att det på sikt blir billiga
32 - Det finns än så länge väldigt lite e-läromedel på finlandssvenskt håll. Dessutom är de inte helt gratis heller, utan eleverna sk
33 Läs också I år måste alla gymnasieelever ha egen laptop
34 </content:encoded>
35   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p15018"/>
36   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p9729"/>
37   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p11320"/>
38   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p5509"/>
39   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p32775"/>
40   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p48358"/>
41   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p18749"/>
42   <dc:subject rdf:resource="http://www.yso.fi/onto/koko/p510"/>
43   <dc:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-01-21T11:39:21+02:00</dc:created>
44   </foaf:Document>
45   </sioc:reply_of>
46   <content:encoded xml:lang="sv">Sluta göra beslut som trycker ner fattiga familjer ännu mera, en laptop är en utgift på flere h
47 </content:encoded>
48   <rdf:type rdf:resource="http://rdfs.org/sioc/ns#Post"/>
49   <sioc:has_creator>
```

Example: Open Graph Protocol RDFa



```
<html prefix="og: http://ogp.me/ns#">
<head>
<title>The Rock (1996)</title>
<meta property="og:title" content="The Rock" />
<meta property="og:type" content="video.movie" />
<meta property="og:url" content="http://www.imdb.com/title/tt0117500/" />
<meta property="og:image" content="http://ia.media-imdb.com/images/rock.jpg" />
...
</head>
...
</html>
```

Data in HTML pages is utilized in e.g. Google search & browsing: Documents vs. entities



Google search results for "barack obama". The search bar shows "barack obama" and a magnifying glass icon. Below the search bar are tabs for "All", "Images", "Videos", "News", "Maps", "More", and "Search tools". The results show "About 177 000 000 results (0,87 seconds)".

Barack Obama - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Barack_Obama
"Barack" and "Obama" redirect here. For his father, see Barack Obama, Sr. For other uses of "Barack", see Barack (disambiguation). For other uses of "Obama", ...
Michelle Obama - Barack Obama, Sr. - Ann Dunham - Early life and career of Barack

In the news

Exclusive: Obama on Iowa, Clinton, Sanders and 2016
Politico - 23 hours ago
Barack Obama, that prematurely gray elder statesman, is laboring mightily to remain neutral ...

Barack Obama: Why we must rethink solitary confinement
Washington Post - 9 hours ago

Obama's Tortured Love Affair With Gadgets | MIT Technology Review
MIT Technology Review - 13 hours ago

More news for barack obama

Barack Obama (@BarackObama) | Twitter
<https://twitter.com/BarackObama>

11 hours ago - [View on Twitter](#)
Time's running out to #GetCovered before the January 31 deadline. Check out your options: [ofa.bo/g9gW](https://t.co/g9gW)

13 hours ago - [View on Twitter](#)
Take a look at President Obama's latest common-sense step to help fight climate change: [ofa.bo/h9qt](https://t.co/h9qt) #ActOnClimate

Barack Obama
44th U.S. President · barackobama.com

Barack Hussein Obama II is an American politician currently serving as the 44th President of the United States, the first African American to hold the office. [Wikipedia](#)

Born: August 4, 1961 (age 54), Honolulu, Hawaii, United States
Spouse: Michelle Obama (m. 1992)
Office: President of the United States since 2009
Parents: Ann Dunham, Barack Obama, Sr.
Education: Harvard Law School (1988–1991), more
Children: Malia Ann Obama, Natasha Obama

Profiles

[Twitter](#) [Facebook](#) [Google+](#) [Instagram](#) [Myspace](#)



More Information – Questions?



Semantic Web & Linked Data Standards

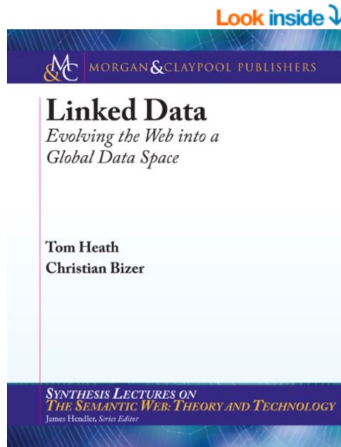
<http://www.w3.org/standards/semanticweb/>

T. Heath, C. Bizer: **Evolving the Web into a Global Data Space**

Free online version: <http://linkeddatabook.com/editions/1.0/>

WarSampo project homepage:

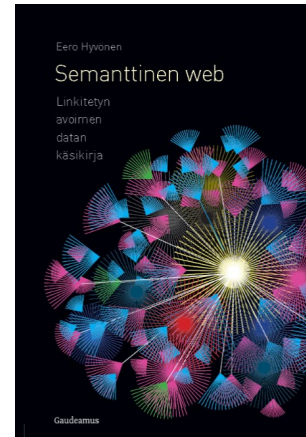
<https://seco.cs.aalto.fi/projects/sotasampo/en/>



In English

2011

<https://www.amazon.com/Linked-Data-Evolving-Global-Space-ebook/dp/B009KC1YM2>



In Finnish

2018

<https://www.gaudemus.fi/semanttinen-web/>