

# A QUICK RECAP OF THE KEY IDEAS AND FEATURES OF WEB OF DATA

## The overall idea:

make the content of the Web 'legible' to computers,  
by presenting it in the language they 'understand'



We also need to have  
content presented in a  
known language

!image source:  
[http://chaxiubao.typepad.com/  
photos/uncategorized/pb060002.JPG](http://chaxiubao.typepad.com/photos/uncategorized/pb060002.JPG)

# Web of Data

- Main features:
  - Data (on the Web) is **structured** and **interlinked**
  - The **semantics** of data and links are made **explicit**
  - Allows for performing **complex queries** over **multiple sources**
  - The vision of the Web as a **gigantic global database**

# Today's topics

- Embedding structured data in Web pages
  - structured data = data with well-defined structure + explicitly defined meaning (semantics)
  - RDFa, Microdata, JSON-LD
  - Schema.org, Open Graph Protocol
- The current state of Linked (Open) Data
  - Principles of linked data publishing
  - Linked data star scheme
  - Linked Open Data (LOD) Cloud
  - Linked Open Vocabularies (LOV)

# EMBEDDING STRUCTURED DATA IN WEB PAGES

# Structured data embedded in Web pages

- Let's first take a look at some examples of structured data embedded in Web pages
- We'll first use [Google's Structured Data Testing tool](#)
- Using this tool, check e.g.,
  - some movies on [RottenTomatoes.com](#)
  - or, some artists on [Last.fm](#)
  - or, some products on [BestBuy.com](#)
  - or, some recipes on [AllRecipes.com](#)

# Structured data embedded in Web pages

- We'll have more use of tools that allow us to pull the structured data from a page programmatically
- W3C offers the [Microdata Distiller](#) tool
  - let's take a look at the same example(s) with this tool;
  - it could be called as a RESTfull service or installed and run locally,
  - it allows you to easily pull data from Web pages – without page scraping or any other similar efforts – and use them in your program

# Structured data embedded in Web pages

- To embed structured data in Web pages, we need:
  - vocabularies for describing the content of the page in a machine-processable format
  - a way to extend HTML to make those machine-processable descriptions an integral part of the Web page
- To address the 1<sup>st</sup> requirement, we can use Schema.org or some other RDFS vocabulary
- To address the 2<sup>nd</sup> requirement, we can use RDFa, Microdata, or JSON-LD – W3C recommendations for extending HTML with machine processable descriptions



# Schema.org

- Developed and maintained by Google, Yahoo, Bing, Yandex
- Started with only a handful of types, and significantly evolved over time through a W3C supported community process
- [Dan Brickley](#) – leading engineer on the project
  - author of widely used [FOAF \(Friend of a Friend\)](#) vocabulary and well known in Semantic Web research community
- Some stats about Schema.org (beginning of 2014):
  - about 15% of Web pages crawled by the major search engines have schema.org markup;
  - over 5M websites are using it;
  - for more information, see [these slides](#)

# Schema.org

- Recommendation:

- watch keynote talk by Google's Ramanathan Guha on the topic of Microdata, Schema.org, and development, application and benefits of these and associated open technologies:

[http://videolectures.net/iswc2013\\_guha\\_tunnel/](http://videolectures.net/iswc2013_guha_tunnel/)

- alternatively, or in addition, read an interview with Guha published at the SemanticWeb.com blog:

[http://semanticweb.com/schema-org-chat-googles-r-v-guha\\_b40607](http://semanticweb.com/schema-org-chat-googles-r-v-guha_b40607)

# RDFa, Microdata, JSON-LD

- W3C recommendations (standards) for embedding structured data in HTML pages:
  - RDFa:
    - Relevant info, code, materials, etc. about RDFa: <http://rdfa.info/>
    - Specification: <http://www.w3.org/TR/xhtml-rdfa-primer/>
  - Microdata:
    - Specification: <http://dev.w3.org/html5/md/>
  - JSON-LD
    - Relevant info, code, materials, etc. about JSON-LD: <http://json-ld.org/>
    - Specification: <http://www.w3.org/TR/json-ld/>
  - Good source of examples is Schema.org site where for each class, there is at least one example in each of the 3 standards

# More about vocabularies

## ■ Schema Actions

- one of the latest features of Schema.org
- allow websites to describe the actions they enable and how these actions can be invoked
- also, allow for integrating data about users' actions from different websites
- to learn how to use this feature, read the following articles:
  - document describing Schema.org actions and offering instructions for their use ([link](#))
  - an article explaining why this feature is relevant ([link](#)), and another one illustrating its use in the music domain ([link](#))

# More about vocabularies

## ■ GoodRelations

- Vocabulary for describing products, offers, shops, and the like
- Already in wide use in the e-commerce domain
  - use Google's Structured Data Testing tool to take a look at the data embedded in pages of Kmart.com, Sears.com, BestBuy.com
- A number of tools have been developed to facilitate the use of this vocabulary for describing products and related items
  - check: <http://wiki.goodrelations-vocabulary.org/Tools>
- This vocabulary has been integrated into Schema.org
  - <http://schema.org/Product> ; <http://schema.org/Offer> ...

# More about vocabularies

## ■ Open Graph Protocol (OGP)

- Introduced by Facebook to obtain more information about the things people ‘Like’ outside the Facebook’s domain
  - RDFa + OGP data embedded in the page provide a formal description of the “liked” item
  - Thus obtained information is used for further extending Facebook’s Entity Graph
- OGP supports the description of several popular domains including music, video, articles, books, websites and user profiles

# Tools for working with embedded structured data

- Google offers a number of tools:
  - Structured Data Dashboard ([link](#))
  - Data Highlighter ([link](#))
  - Structured Data Markup Helper ([link](#))
  - video from Google IO 2013 conference ([link](#)) introduces and describes these tools

# Tools for working with embedded structured data

- Popular Web platforms that support RDFa/Microdata
  - Drupal
    - support for RDFa is a part of Drupal's core functionalities (from v.7);
    - the upcoming version (v.8) will include Schema.org as a foundational data type
  - Webnodes
    - offers fully integrated dynamic support for Microdata and Schema.org (check [this article](#))
  - Wordpress
    - Offers a number of extensions for working with RDFa, Microdata and Schema.org (check, e.g., [this list](#))



# A few application examples

## ▪ Rich Snippets

- richer display of Google search results for pages with embedded structured data
- e.g., search Google for the JWNL Sourceforge project or any movie or any mobile app

## ▪ Interactive Snippets

- currently available in Yandex search results (“Islands”); see [this article](#) for more information

## ▪ Pinterest’s Rich Pins

- Pins with additional information/functionality; e.g., product rich pins provide current price, availability, location, even available discounts
- see, for instance, where product rich pins originate from, that is how structured data is used to generate rich pins ([link](#))