Plan for today...

• Recap + Open Graph Protocol

• Making Use of Structured Data

• Useful Tools and Services

• Reference Solution Homework 6
RECAP
What is Schema.org?

This site provides a collection of schemas, i.e., html tags, that webmasters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google and Yahoo! rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, Bing, Google and Yahoo! have come together to provide a shared collection of schemas that webmasters can use.

We invite you to get started!

Last Updated: 27 May 2011
Pirates of the Carribean: On Stranger Tides (2011)

Jack Sparrow and Barbossa embark on a quest to find the elusive fountain of youth, only to discover that Blackbeard and his daughter are after it too.

Director: Rob Marshall
Writers: Ted Elliott, Terry Rossio, and 7 more credits
Stars: Johnny Depp, Penelope Cruz, Ian McShane
8/10 stars from 200 users. Reviews: 50.
<div itemscope itemtype="http://schema.org/Movie">

<h1 itemprop="name">Pirates of the Carribean: On Stranger Tides (2011)</h1>

<span itemprop="description">Jack Sparrow and Barbosa embark on a quest to find the elusive fountain of youth, only to discover that Blackbeard and his daughter are after it too.</span>

Director:
<span itemprop="director" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Rob Marshall</span>
</span>

Writers:
<span itemprop="author" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Ted Elliott</span>
</span>

, and 7 more credits
Stars:
<span itemprop="actor" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Johnny Depp</span>,
</span>

<span itemprop="actor" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Penelope Cruz</span>,
</span>

<span itemprop="actor" itemscope itemtype="http://schema.org/Person">
  <span itemprop="name">Ian McShane</span>
</span>

<span itemprop="aggregateRating" itemscope itemtype="http://schema.org/AggregateRating">
  <span itemprop="ratingValue">8</span>/<span itemprop="bestRating">10</span> stars from <span itemprop="ratingCount">200</span> users.
</span>

Reviews: <span itemprop="reviewCount">50</span>.
schema.org

• Defines
  – a number of types (e.g., person), organized in an inheritance hierarchy
  – a number of properties (e.g., name)

• Extension mechanisms to extend the schemas

• OWL representation:
  http://schema.org/docs/schemaorg.owl

• http://schema.rdfs.org/index.html
Open Graph Protocol
Graph API

API Reference / Graph API

At Facebook's core is the social graph; people and the connections they have to everything they care about. The Graph API presents a simple, consistent view of the Facebook social graph, uniformly representing objects in the graph (e.g., *people*, *photos*, *events*, and *pages*) and the connections between them (e.g., friend relationships, shared content, and photo tags).

Every object in the social graph has a unique ID. You can access the properties of an object by requesting [https://graph.facebook.com/ID](https://graph.facebook.com/ID). For example, the official page for the Facebook Platform has id 19292868552, so you can fetch the object at [https://graph.facebook.com/19292868552](https://graph.facebook.com/19292868552):

```json
{
    "name": "Facebook Platform",
    "website": "http://developers.facebook.com",
    "username": "platform",
    "founded": "May 2007",
    "company_overview": "Facebook Platform enables anyone to build...",
    "mission": "To make the web more open and social.",
    "products": "Facebook Application Programming Interface (API)...",
    "likes": 449921,
    "id": 19292868552,
    "category": "Technology"
}
```

Alternatively, people and pages with usernames can be accessed using their username as an ID. Since "platform" is the username for the page above, [https://graph.facebook.com/platform](https://graph.facebook.com/platform) will return what you expect. All responses are JSON objects.

All objects in Facebook can be accessed in the same way:

- Users: [https://graph.facebook.com/btaylor](https://graph.facebook.com/btaylor)
name: "Coca-Cola",
is_published: true,
website: "http://www.coca-cola.com",
username: "coca-cola",
founded: "1886",
description: "Created in 1886 in Atlanta, Georgia, by Dr. John S. Pemberton, Coca-Cola was first offered as a fountain beverage at Jacob's Cascades in Atlanta. Coca-Cola was patented in 1887, registered as a trademark in 1893 and by 1895 it was being sold in every state and territory in the United States. Coca-Cola might owe its origins to the United States, but its popularity has made it truly universal. Today, you can find Coca-Cola in more than 200 countries and territories around the world."

Coca-Cola Page House Rules: http://CokeURL.com/q28a",
about: "The Coca-Cola Facebook Page is a collection of your stories showing how people from around the world have helped make Coke into the world's most loved drink."

- location: {
  latitude: 19.2,
  longitude: -96.1333
},
checkins: 127,
talking_about_count: 1687218,
category: "Food/beverages",
id: "40796308305",
link: "http://www.facebook.com/coca-cola",
likes: 53165520,

- cover: {
  cover_id: "10151948639733396",
  offset_y: 0
}
<https://graph.facebook.com/40796308305#>

page:about "The Coca-Cola Facebook Page is a collection of your stories showing how people from around the world have helped make Coke into what it is today."

page:category "Food/beverages"
page:checkins 127;
page:cover
   :cover_id 10151948639733306;
   :offset_y 0;

page:description """"Created in 1886 in Atlanta, Georgia, by Dr. John S. Pemberton, Coca-Cola was first offered as a fountain beverage at Jacob's Pharmacy by mixing Coca-Cola syrup with carbonated water.

Coca-Cola was patented in 1887, registered as a trademark in 1893 and by 1895 it was being sold in every state and territory in the United States. In 1899, The Coca-Cola Company began franchised bottling operations in the United States.

Coca-Cola might owe its origins to the United States, but its popularity has made it truly universal. Today, you can find Coca-Cola in virtually every part of the world.

Coca-Cola Page House Rules: http://CokeURL.com/q28a"""";
page:founded 1886;
page:id "40796308305";
page:is_published true;
page:likes 53165803;
page:link <http://www.facebook.com/coca-cola>;
page:location
   :latitude 19.2;
   :longitude -96.1333
page:name "Coca-Cola"
page:talking_about_count 1687218
page:username "coca-cola"

rapper: Parsing returned 20 triples
Open Graph
Core Concepts > Open Graph

At Facebook’s core is the social graph: people and the connections they have to everything they care about. Historically, Facebook has managed this graph and has expanded it over time as we launch new products (ex: photos, places). In 2010, we introduced an early version of Open Graph, an extension of the social graph, via the Open Graph protocol, to include third-party websites and pages that people liked throughout the web. We are now extending the Open Graph to include arbitrary actions and objects created by third-party apps and enabling these apps to integrate deeply into the Facebook experience.
The Open Graph protocol

Introduction
The Open Graph protocol enables any web page to become a rich object in a social graph. For instance, this is used on Facebook to allow any web page to have the same functionality as any other object on Facebook.

While many different technologies and schemas exist and could be combined together, there isn’t a single technology which provides enough information to richly represent any web page within the social graph. The Open Graph protocol builds on these existing technologies and gives developers one thing to implement. Developer simplicity is a key goal of the Open Graph protocol which has informed many of the technical design decisions.

Basic Metadata
To turn your web pages into graph objects, you need to add basic metadata to your page. We’ve based the initial version of the protocol on RDFa which means that you’ll place additional <meta> tags in the <head> of your web
protocol builds on these existing technologies and gives developers one thing to implement. Developer simplicity is a key goal of the Open Graph protocol which has informed many of the technical design decisions.

Basic Metadata

To turn your web pages into graph objects, you need to add basic metadata to your page. We've based the initial version of the protocol on RDFa which means that you'll place additional <meta> tags in the <head> of your web page. The four required properties for every page are:

- **og:title** - The title of your object as it should appear within the graph, e.g., "The Rock".
- **og:type** - The type of your object, e.g., "video.movie". Depending on the type you specify, other properties may also be required.
- **og:image** - An image URL which should represent your object within the graph.
- **og:url** - The canonical URL of your object that will be used as its permanent ID in the graph, e.g., "http://www.imdb.com/title/tt0117500/".

As an example, the following is the Open Graph protocol markup for The Rock on IMDB:

```html
<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>
```

Optional Metadata

The following properties are optional for any object and are generally recommended:
Object Types

In order for your object to be represented within the graph, you need to specify its type. This is done using the `og:type` property:

```html
<meta property="og:type" content="website" />
```

When the community agrees on the schema for a type, it is added to the list of global types. All other objects in the type system are CURIES of the form

```html
<head prefix="my_namespace: http://example.com/ns#">  
<meta property="og:type" content="my_namespace:my_type" />
</head>
```

The global types are grouped into verticals. Each vertical has its own namespace. The `og:type` values for a namespace are always prefixed with the namespace and then a period. This is to reduce confusion with user-defined namespaced types which always have colons in them.

Music

- Namespace URI: http://ogp.me/ns/music#

`og:type` values:

music.song

- `music:duration` - integer $\geq 1$ - The song's length in seconds.
- `music:album` - music.album array - The album this song is from.
- `music:album:disc` - integer $\geq 1$ - Which disc of the album this song is on.
- `music:album:track` - integer $\geq 1$ - Which track this song is.
- `music:musician` - profile array - The musician that made this song.

music.album
MAKING USE OF STRUCTURED DATA
Google Knowledge Graph

• Enables search for things (people, places) that Google knows about

• Rooted in public sources such as Freebase, Wikipedia, CIA World Factbook, etc.
  – augmented to 500M objects, 3.5B facts and relationship

• Next generation search (semantic index)
http://www.youtube.com/watch?v=mmQl6VGvX-c
John Joseph "Jack" Nicholson (born April 22, 1937) is an American actor, film director, producer, and writer. He is known for his often dark portrayals of neurotic characters. His twelve Oscar nominations make him the second most nominated actor of all time, tied with Katharine Hepburn, and behind only Meryl Streep. Nicholson was nominated for an Academy Award twelve times and won the Academy Award for Best Actor twice: for One Flew Over the Cuckoo's Nest and As Good as It Gets.

**Date of birth:** Apr 22, 1937 (age 75 years)

**Place of birth:** New York City, New York, United States of America

**Height:** 1.77 m (5.81 ft)

**Religion:** Roman Catholicism

**Also known as:** John Joseph Nicholson, Jack Nicholson, Nick, Mulholland Man, The Great Seducer, John Joseph "Jack" Nicholson

**Place of birth:** New York City, New York, United States of America

**Country of nationality:** United States of America

**Gender:** Male

**Profession:**
- Actor
- Screenwriter
- Film Director
- Film Producer

**Ethnicity:** Irish descent
Jack Nicholson - Wikipedia, the free encyclopedia

John Joseph "Jack" Nicholson (born April 22, 1937) is an American actor, film ... roles such as Jack Torrance in The Shining, Frank Costello in The Departed, and ..., in the movie adaptation of Ken Kesey's novel One Flew Over the Cuckoo's Nest, ... The 1989 Batman movie, wherein Nicholson played the psychotic murderer ... Lorraine Nicholson - As Good as It Gets - List of awards - About Schmidt

Jack Nicholson - IMDb

www.imdb.com/name/nm0000197/


IMDb: jack nicholson movies - a list by linda-smeltzer

www.imdb.com/list/SLTo8jTN9TY/
Jack Nicholson's Bacon number is 1


Google 'Bacon Number' And Say Goodbye To The Rest Of Your Day
moviesblog.mtv.com/2012/09/13/bacon-number-google/
Sep 13, 2012 – We learned that Kristen Stewart has a Bacon number of two because ... and Jack Nicholson both appeared in "A Safe Place," while Bacon and ...

Google builds Six Degrees of Kevin Bacon into its search system ...
www.guardian.co.uk › Culture › Film
Sep 13, 2012 – By way of example, typing "Bacon number Simon Pegg" reveals that Bacon ..... who co-starred with Jack Nicholson appeared in A Safe Place.

Google builds Bacon Number functionality | MetaFilter
www.metafilter.com/.../Google-builds-Bacon-Number-functio...
100+ posts - 88 authors - Sep 13
Tippi Hedren and Kevin Bacon appeared in Jayne Mansfield's Car. Orson Welles's Bacon number is 2. Orson Welles and Jack Nicholson ...

Six Degrees Of Kevin Bacon - Television Tropes & Idioms
tvtropes.org/pmwiki/pmwiki.php/.../SixDegreesOfKevinBacon
A description of tropes appearing in Six Degrees Of Kevin Bacon. ... Jack Nicholson, and Jack Nicholson was in A Few Good Men with Kevin Bacon. ... Mathematicians
Coreference / entity resolution

- Cluster mentions of entities extracted from a body of text such that two mentions belong to the same cluster if and only if they refer to the same entity.
USEFUL TOOLS AND SERVICES
Google Refine

• A tool for working with messy data
  – cleaning, transforming, extending, linking data

• [http://code.google.com/p/google-refine/](http://code.google.com/p/google-refine/)

• Since 10/2012 OpenRefine
  ([http://github.com/OpenRefine](http://github.com/OpenRefine))
Google Refine

http://www.youtube.com/watch?v=B70J_H_zAWM

http://www.youtube.com/watch?v=5tsyz3ibYzk
Google Refine / LD Extension

RDF Refine - a Google Refine extension for exporting RDF

Main features

- Reconcile & interlink
  - Reconcile against SPARQL endpoints, RDF dumps
  - Search the Web for related RDF datasets

- Export RDF
  - CUI for defining the shape of the RDF graph
  - Use your own vocabulary or import existing ones
  - Autocomplete for property and class names

Getting started...

1. Make sure "extensions" folder exists in your Google Refine workspace

News

2012-07-27: version 0.8.0 released
Support for reconciling using Apache Stanbol is added. This feature has been
DBpedia Spotlight

• Named entity detection
  – analyzes natural language text
  – identifies named entity matches

• Named entity disambiguation
  – use heuristics to find the best match
  – similar to coreference resolution problem in Google Knowledge Graph

• https://github.com/dbpedia-spotlight/dbpedia-spotlight
President Obama on Monday will call for a new minimum tax rate for individuals making more than $1 million a year to ensure that they pay at least the same percentage of their earnings as other taxpayers, according to administration officials.

You should know:

- This interface has been tested with Firefox 6.0.2 and Chromium 12.0.
- We have a cute [bookmarked](#) that you should try out!
- DBpedia Spotlight is currently only available for English. However, it could be adapted for any language for which there is a DBpedia. Contact us if you’d like to help!

This demonstration uses the [DBpedia Spotlight jQuery Plugin v0.3](#) and the Web Service for [DBpedia Spotlight v0.5](#). For the latest version, please visit [http://spotlight.dbpedia.org](http://spotlight.dbpedia.org).
lucene-SKOS

• Query expansion module for Lucene / Solr
• Takes SKOS vocabularies
• Expands terms either at
  – query time OR
  – indexing time
• https://github.com/behas/lucene-skos
Overview

R is a programming language designed for statistics.
R Sparql allows you to run SPARQL Queries inside R and store it as a R data frame.
The main objective is to allow the integration of Ontologies with Statistics.
It requires Java and rJava installed.
Example (in R console):

```r
> library(sparql)
> data <- query("<your SPARQL query>","RDF file or remote SPARQL Endpoint")
```

Getting started

Quick start here:

http://code.google.com/p/r-sparql/wiki/GettingStarted