Semantic markup, microdata

Overview

- **Major deliverables**
  - Paper/presentation
  - Final project
- **HTML5 semantic markup**
  - Why?
  - Common tags
- **Custom markup**
  - Microdata
  - RFDa
  - Microformats

Possible paper topics

- Geolocation
  - Technical details about how it works
  - Location-based applications
  - Privacy issues

- Web services
  - Differences between SOAP and REST web services
  - Learn details, do something interesting with a specific API
    - e.g. What can be done with the Facebook API?

- What do high performance servers such as nginx and lighttpd do differently from Apache?
  - Discuss feature differences, advantages, disadvantages
  - Empirical performance comparison

Possible paper topics

- Web farms at extreme scale
  - "The Datacenter as a Computer", Google
- ????

Paper details

- Paper:
  - Due start of class, Mon Apr 9th
  - 1500 words minimum, 3000 words maximum
  - Format:
    - Technical paper with references
      - Abstract
      - Introduction
      - Content section(s)
      - Conclusions
      - References
  - Teach me something about the subject
  - 10% of your grade
  - No late days

Presentation details

- Presentations:
  - Following week, Wed Apr 11th, Fri Apr 13th
  - ~15 minutes, audio/visual aids as appropriate
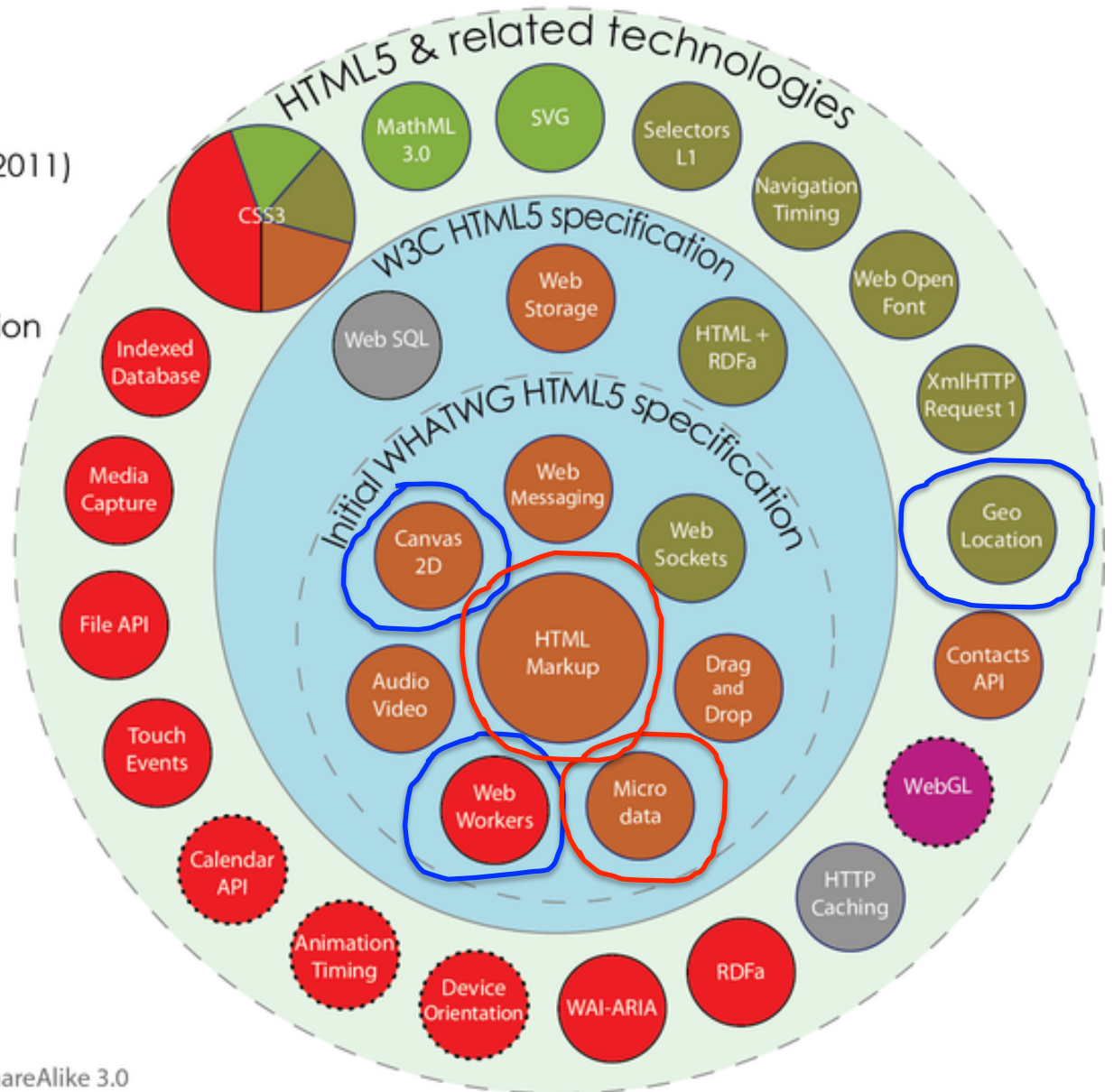 - 5% of your grade
  - No late days

Final project

- Project combining things we've learned
  - Exact problem TBD
  - Individual or in pairs
  - Due: Wed May 2nd (demo to class)
  - 20% of your grade
  - No late days

HTML5

Taxonomy & Status (December 2011)

- W3C Recommendation
- Candidate Recommendation
- Last Call
- Working Draft
- Non-W3C Specifications
- Deprecated W3C APIs

HTML5 & related technologies

W3C HTML5 specification

Initial WHATWG HTML5 specification

CSS3 · MathML 3.0 · SVG · Selectors L1 · Navigation Timing · Web Open Font · XmlHTTP Request 1 · Geo Location · Contacts API · WebGL · HTTP Caching · RDFa · WAI-ARIA · Device Orientation · Animation Timing · Calendar API · Touch Events · File API · Media Capture · Indexed Database

Web SQL · Web Storage · HTML + RDFa · Web Messaging · Web Sockets · Canvas 2D · HTML Markup · Drag and Drop · Audio Video · Web Workers · Micro data

By Sergey Mavrody 2011 | CC Attribution-ShareAlike 3.0

8

Semantic web

- The problem:
  - Web pages are hard for computer to parse
    - Everything in a <div> with a id/class
    - Is the <div> "menu" the navigation sidebar or a menu of a restaurant?
  - What part of the page is:
    - A blog post? The date of the post?
    - Header of the web site? Footer?
    - Figure? Caption of that figure?

Semantic web

- ## What is it?

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation"

– May 2001, Scientific American, Tim Berners-Lee

"The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications."

– W3C 2001

New HTML5 semantic tags

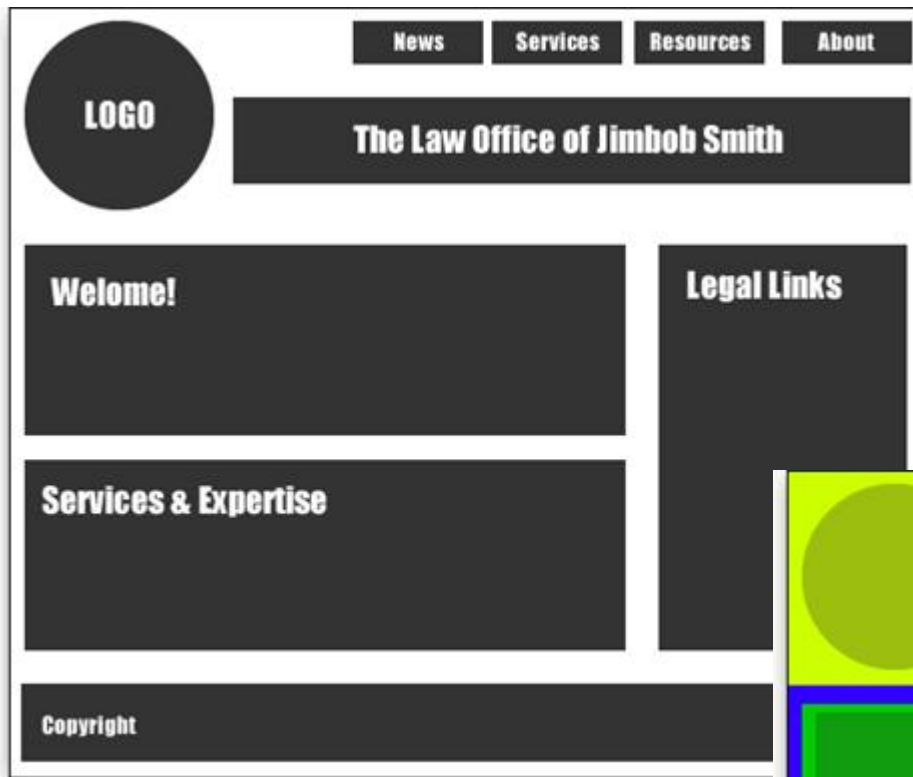| Tag	Description
<article>	Defines an article
<aside>	Defines content aside from the page content
<bdi>	Isolates a part of text that might be formatted in a different direction from other text outside it
<command>	Defines a command button that a user can invoke
<details>	Defines additional details that the user can view or hide
<summary>	Defines a visible heading for a <details> element
<figure>	Specifies self-contained content, like illustrations, diagrams, photos, code listings, etc.
<figcaption>	Defines a caption for a <figure> element
<footer>	Defines a footer for a document or section
<header>	Defines a header for a document or section
<hgroup>	Groups a set of <h1> to <h6> elements when a heading has multiple levels
<mark>	Defines marked/highlighted text
<meter>	Defines a scalar measurement within a known range (a gauge)
<nav>	Defines navigation links
<progress>	Represents the progress of a task
<ruby>	Defines a ruby annotation (for East Asian typography)
<rt>	Defines an explanation/pronunciation of characters (for East Asian typography)
<rp>	Defines what to show in browsers that do not support ruby annotations
<section>	Defines a section in a document
<time>	Defines a date/time
<wbr>	Defines a possible line-break

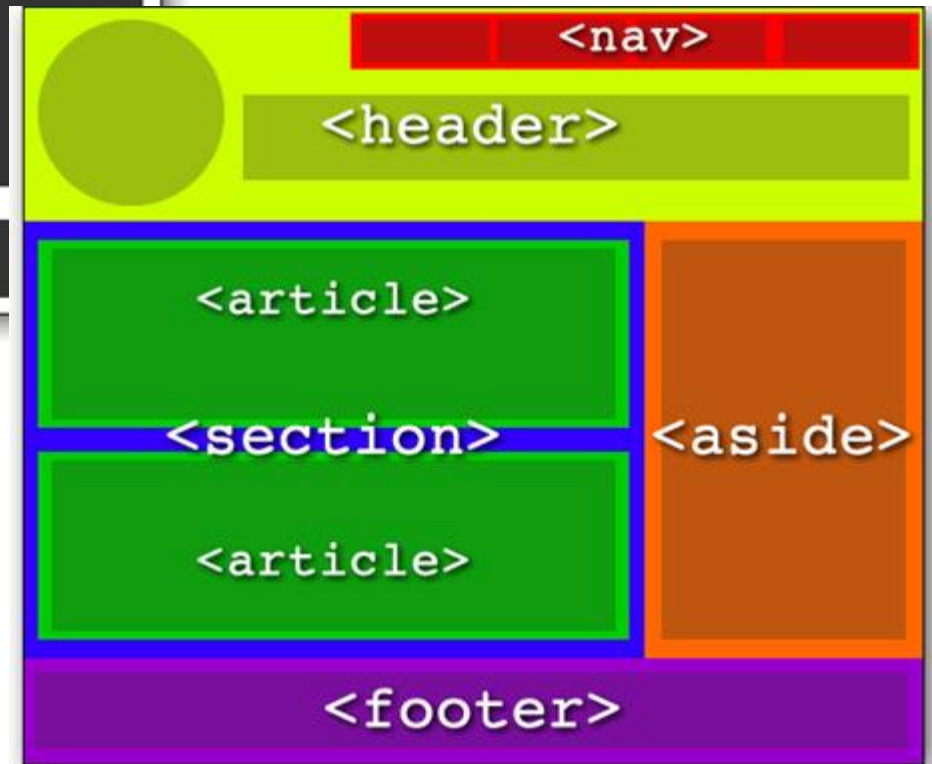Figure 1 - General browser layout for fictitious home page

http://msdn.microsoft.com/en-us/scriptjunkie/gg454786

Figure 2 - New HTML5 semantic elements applied to layout for fictitious site

Semantic element support

	*Usage stats:	Global
Support:		66.31%
Partial support:		6.55%
Total:		72.86%

New semantic elements - **Working Draft**

HTML5 offers some new elements, primarily for semantic purposes. The elements include: section, article, aside, hgroup, header, footer, nav, figure, figcaption, time, mark.

Resources: Workaround for IE Alternate workaround Article on structural elements

Show all versions	IE	Firefox	Chrome	Safari	Opera	iOS Safari	Opera Mini	Opera Mobile	Android Browser
								10.0	2.1
	6.0	3.6				3.2		11.0	2.2
	7.0	8.0				4.0-4.1		11.1	2.3
	8.0	9.0	16.0	5.0		4.2-4.3		11.5	3.0
Current	9.0	10.0	17.0	5.1	11.6	5.0	5.0-6.0	12.0	4.0
Near future	10.0	11.0	18.0	6.0	12.0				
Farther future		12.0	19.0						

Note: Partial support refers to missing the default styling. This is easily taken care of by using display:block for all new elements (except time and mark, these should be display:inline anyway).

Feedback

http://caniuse.com/#feat=html5semantic

13

<section>

- According to W3C:
  - "represents a generic section of a document"
  - "a thematic grouping of content, typically with a heading"
- Not a generic container for styling
  - That is <div>'s job
- Avoid if <article>, <aside>, <nav> more appropriate
- Examples:
  - Chapters
  - Numbered section of a thesis

<section> example

```html
<article>
 <hgroup>
  <h1>Apples</h1>
  <h2>Tasty, delicious fruit!</h2>
 </hgroup>
 <p>The apple is the pomaceous fruit of the apple tree.</p>
 <section>
  <h1>Red Delicious</h1>
  <p>These bright red apples are the most common found in many
  supermarkets.</p>
 </section>
 <section>
  <h1>Granny Smith</h1>
  <p>These juicy, green apples make a great filling for
  apple pies.</p>
 </section>
</article>
```

<article>

- Self-contained chunk of content
  - Something you may want to share
  - Independently distributable or reusable (syndication)
- Examples:
  - Forum post
  - Magazine/newspaper article
  - Blog entry
  - User-submitted comment

```html
<article itemscope itemtype="http://schema.org/BlogPosting">
 <header>
  <h1 itemprop="headline">The Very First Rule of Life</h1>
  <p><time itemprop="datePublished" datetime="2009-10-09">3 days ago</time></p>
  <link itemprop="url" href="?comments=0">
 </header>
 <p>If there's a microphone anywhere near you, assume it's hot and
 sending whatever you're saying to the world. Seriously.</p>
 <p>...</p>
 <section>
  <h1>Comments</h1>
  <article itemprop="comment" itemscope itemtype="http://schema.org/UserComments" id="c1">
   <link itemprop="url" href="#c1">
   <footer>
    <p>Posted by: <span itemprop="creator" itemscope itemtype="http://schema.org/Person">
     <span itemprop="name">George Washington</span>
    </span></p>
    <p><time itemprop="commentTime" datetime="2009-10-10">15 minutes ago</time></p>
   </footer>
   <p>Yeah! Especially when talking about your lobbyist friends!</p>
  </article>
  <article itemprop="comment" itemscope itemtype="http://schema.org/UserComments" id="c2">
   <link itemprop="url" href="#c2">
   <footer>
    <p>Posted by: <span itemprop="creator" itemscope itemtype="http://schema.org/Person">
     <span itemprop="name">George Hammond</span>
    </span></p>
    <p><time itemprop="commentTime" datetime="2009-10-10">5 minutes ago</time></p>
   </footer>
   <p>Hey, you have the same first name as me.</p>
  </article>
 </section>
</article>
```

Headers and footers

- **<header>, for the tops of:**
  - Sections, articles
  - Top of body for main header of your page
  - You can have multiple

- **<footer>, for the bottoms of:**
  - Sections, articles
  - Anywhere you need footer content
  - e.g. Author of document, copyright, privacy policy
  - You can have multiple

Header/footer example

```
<header>
<h1>Scientist discover way to reduce headaches</h1>
<b><p>Sleeping with your shoes strongly correlated with
waking up with a headache</p></b>
</header>
<article>
<p>Blah blah blah blah blah</p>
<p>Blah blah blah...</p>
</article>
<footer>
Copyright 2012 by Author
</footer>
```

<hgroup>

- Represents the headings of a section
  - Group set of h1-h6 elements when heading has multiple levels
  - e.g. subheadings, alternative titles, taglines

```
<hgroup>
 <h1>The reality dysfunction</h1>
 <h2>Space is not the only void</h2>
</hgroup>
<hgroup>
 <h1>Dr. Strangelove</h1>
 <h2>Or: How I Learned to Stop Worrying and Love the Bomb</h2>
</hgroup>
```

<nav>

- Navigation and links
  - Used for groups of links, not a single link
  - e.g. Links to all articles in a forum thread
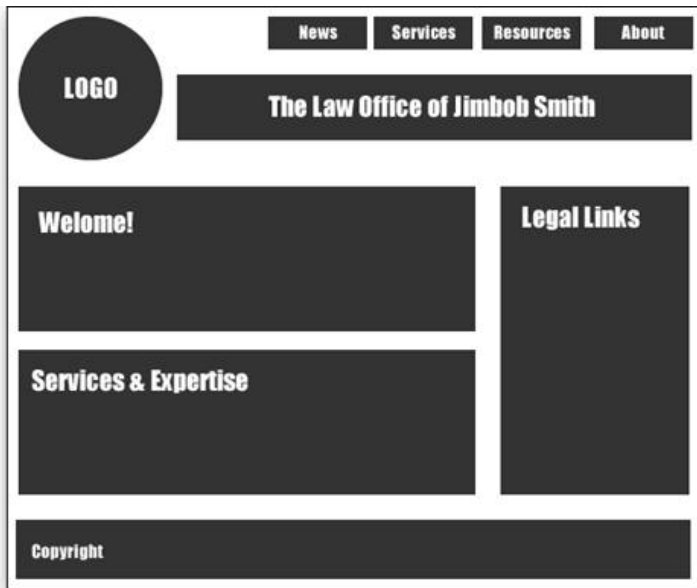 - Not needed for links in <header>, <footer>

Figure 1 - General browser layout for fictitious home page

Figure 2 - New HTML5 semantic elements applied to layout for fictitious site

http://msdn.microsoft.com/en-us/scriptjunkie/gg454786

```html
<header>
 <h1>Wake up sheeple!</h1>
 <p><a href="news.html">News</a> -
    <a href="blog.html">Blog</a> -
    <a href="forums.html">Forums</a></p>
 <p>Last Modified: <span itemprop="dateModified">2009-04-01</span></p>
 <nav>
  <h1>Navigation</h1>
  <ul>
   <li><a href="articles.html">Index of all articles</a></li>
   <li><a href="today.html">Things sheeple need to wake up for today</a></li>
   <li><a href="successes.html">Sheeple we have managed to wake</a></li>
  </ul>
 </nav>
</header>
<div>
 <article itemprop="blogPosts" itemscope itemtype="http://schema.org/BlogPosting">
  <header>
   <h1 itemprop="headline">My Day at the Beach</h1>
  </header>
  <div itemprop="articleBody">
   <p>Today I went to the beach and had a lot of fun.</p>
  </div>
  <footer>
   <p>Posted <time itemprop="datePublished" datetime="2009-10-10">Thursday</time>.</p>
  </footer>
 </article>
</div>
 <footer>
  <p>Copyright 2010 The Example Company</p>
  <p><a href="about.html">About</a> -
     <a href="policy.html">Privacy Policy</a> -
     <a href="contact.html">Contact Us</a></p>
 </footer>
```
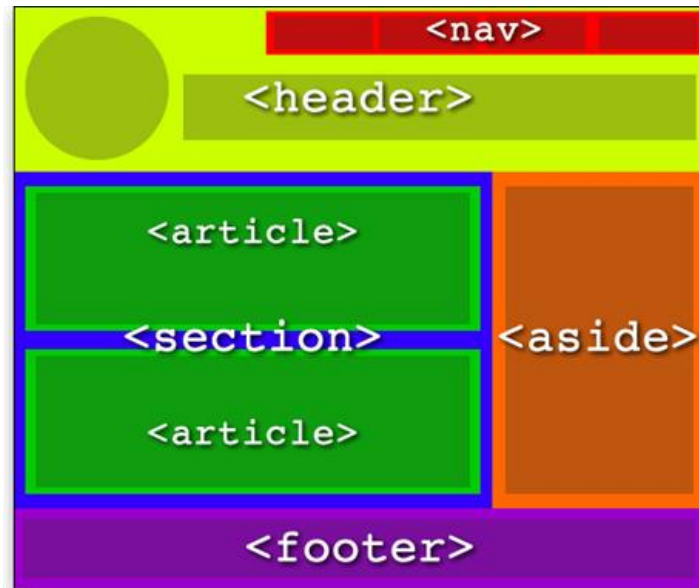
Some other elements

- **<aside>**
  - Chunks of content outside main flow of text
  - Sidebar, quote, after-though
- **<time>**
  - Machine readable time/date
- **<abbr>**
  - Expansion of an abbreviation
- **<mark>**
  - Mark words, for highlighting or editing

Even more meaning

- **How does the W3C actually know what I need?**
  - I want to markup:
    - Names: companies, first names, last names, pet names, …
    - Sarcastic comments in forum posts
    - Info about the CDs in my library
  - Extend HTML in some way so I can add my own tags or attributes

Approaches

- **Microdata**
  - WHATWG HTML specification
    - Web Hypertext Technology Working Group
      - Development of HTML and APIs, formed by Apple, Mozilla, Opera
      - Response to "W3C's direction with XHTML, lack of interest in HTML and apparent disregard for the needs of real-world authors"
  - Now a W3C working draft

- **RDFa** (Resource Description Framework with attributes)
  - W3C recommendation
  - Set of attribute extensions to XHTML

- **Microformats**
  - Grassroots effort, not a standards body
  - 34 published formats

Microdata person example

```html
<section itemscope itemtype="http://data-vocabulary.org/Person">
        <h2 itemprop="name">Keith Vertanen</h2>
        <img itemprop="photo"
          src="http://www.keithv.com/pics/kv21.jpg"
          alt="What a handsome guy" />
        <p>I am an assistant professor at
        <span itemprop="affiliation">Montana Tech</span>.</p>
        <address itemprop="address" itemscope
          itemtype="http://data-vocabulary.org/Address">
        <b>Address:</b>
        <p itemprop="locality">Butte, Montana</p>
        </address>
</section>
```

http://www.google.com/webmasters/tools/richsnippets

Microdata product example

```html
<div itemscope itemtype="http://data-vocabulary.org/Product">
  <span itemprop="brand">ACME</span> <span itemprop="name">Executive Anvil</span>
  <img itemprop="image" src="anvil.png"/><br />
  <span itemprop="description">Sleeker than ACME's Classic Anvil, the
    Executive Anvil is perfect for the business traveler
    looking for something to drop from a height.
  </span><br />
  Category:
  <span itemprop="category" content="Hardware > Tools > Anvils">Anvils</span><br />
  Product #: <span itemprop="identifier" content="mpn:925872">925872</span><br />
  <span itemprop="review" itemscope
   itemtype="http://data-vocabulary.org/Review-aggregate">
    <span itemprop="rating">4.4</span> stars, based on
    <span itemprop="count">89</span> reviews <br />
    $<span itemprop="price">119.99</span>
  </span>
</div>
```

http://www.google.com/webmasters/tools/richsnippets

Microdata details

- Microdata vocabularies
  - Meaning for an item
  - Design your own custom one, or link to one
    - http://data-vocabulary.org
- Attributes
  - itemscope
    - Creates the item, descendants of this element has the information
  - itemtype
    - URL to the vocabulary that describes the item
  - itemprop
    - Value of a particular property of the item

Google rich snippets

- What custom semantic tags should we use?
  - The ones supported by Google

- Google rich snippets
  - "designed to give users a send of what's on the page and why it's relevant to their query"
  - Microdata (recommended), microformats, RDFa
  - Content types:
    - Reviews, People, Products, Businesses and organizations
    - Recipes, Events, Music

Property	Description
name (fn)	The name of the dish.
recipeType (tag)	The type of dish: for example, appetizer, entree, dessert …
photo	Image of the dish being prepared.
published	The date the recipe was published, in ISO date format.
summary	A short summary describing the dish.
review	A review of the dish. Can include nested review information.
prepTime	The length of time it takes to prepare the recipe for dish, in ISO 8601 duration format. Can use min, max as child elements to specify a range of time.
cookTime	The time it takes to actually cook the dish, in ISO 8601 duration format. Can use min, max as child elements to specify a range of time.
totalTime (duration)	The total time it takes to prepare the cook the dish, in ISO 8601 duration format. Can use min, max as child elements to specify a range of time.
nutrition	Nutrition information about the recipe. Can contain the following child elements: servingSize, calories, fat, saturatedFat, unsaturatedFat, carbohydrates, sugar, fiber, protein, cholesterol. These elements are not explicitly part of the hRecipe microformat, but Google will recognize them.
instructions	The steps to make the dish. Can contain the child element instruction, which can be used to annotate each step.
yield	The quantity produced by the recipe (for example, number of people served, number of servings, etc).
ingredient	An ingredient used in the recipe. Can contain child items name (name of the ingredient) and amount. Use this to identify individual ingredients.
author	Creator of the recipe. Can include nested Person information.

Summary

- **Semantic markup**
  - Using tags/attributes to describe meaning
  - Separates presentation from semantics
  - Makes computer processing easier
    - Very hard to parse meaning from arbitrary page: AI complete
  - Multiple standards
    - Microdata
    - RDFa
    - Microformats