

# Schema.org Update

Guha

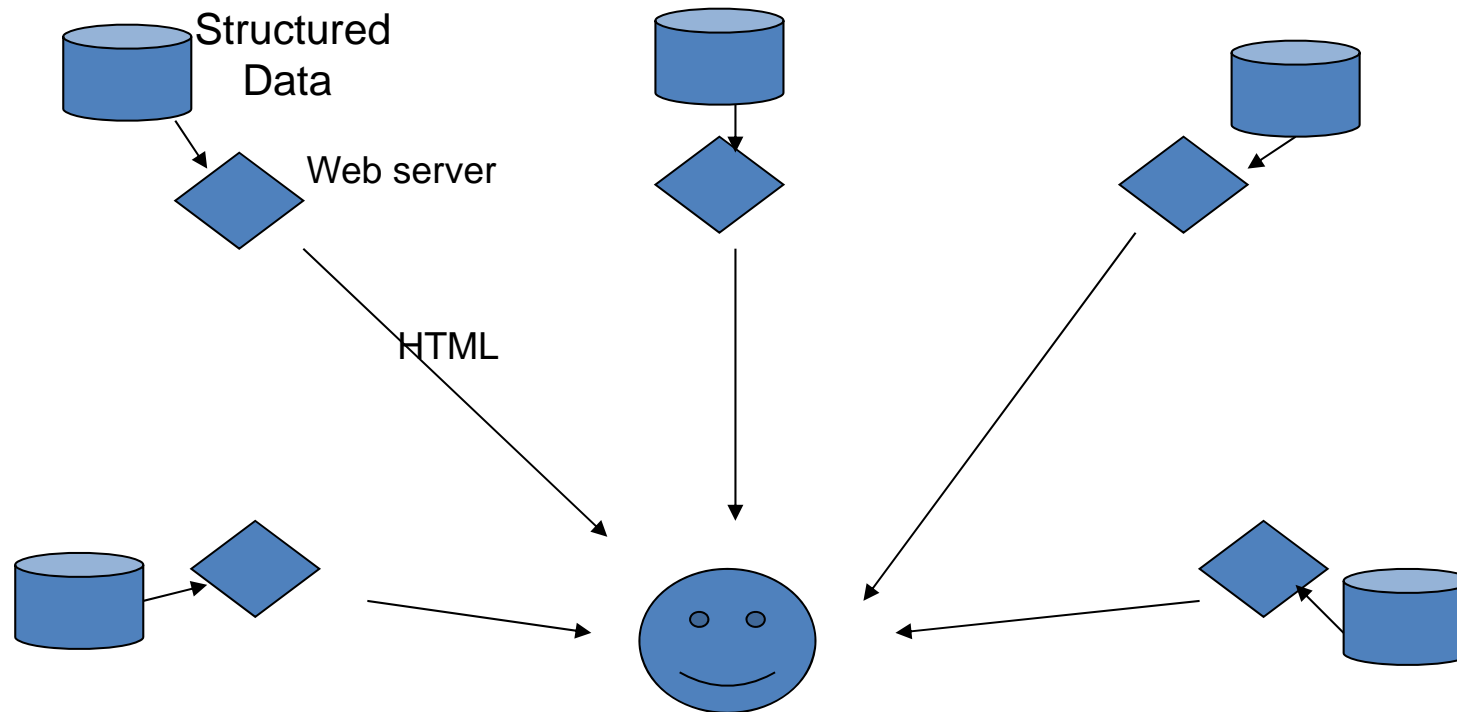
# Outline of talk

- The context
  - How did we end up where we are with the ‘Semantic Web’
- Schema.org
  - What it is, status of adoption
  - Interesting examples & applications
  - Schema.org principles, how does it work
  - Schemas in the pipeline
- Research problems/opportunities

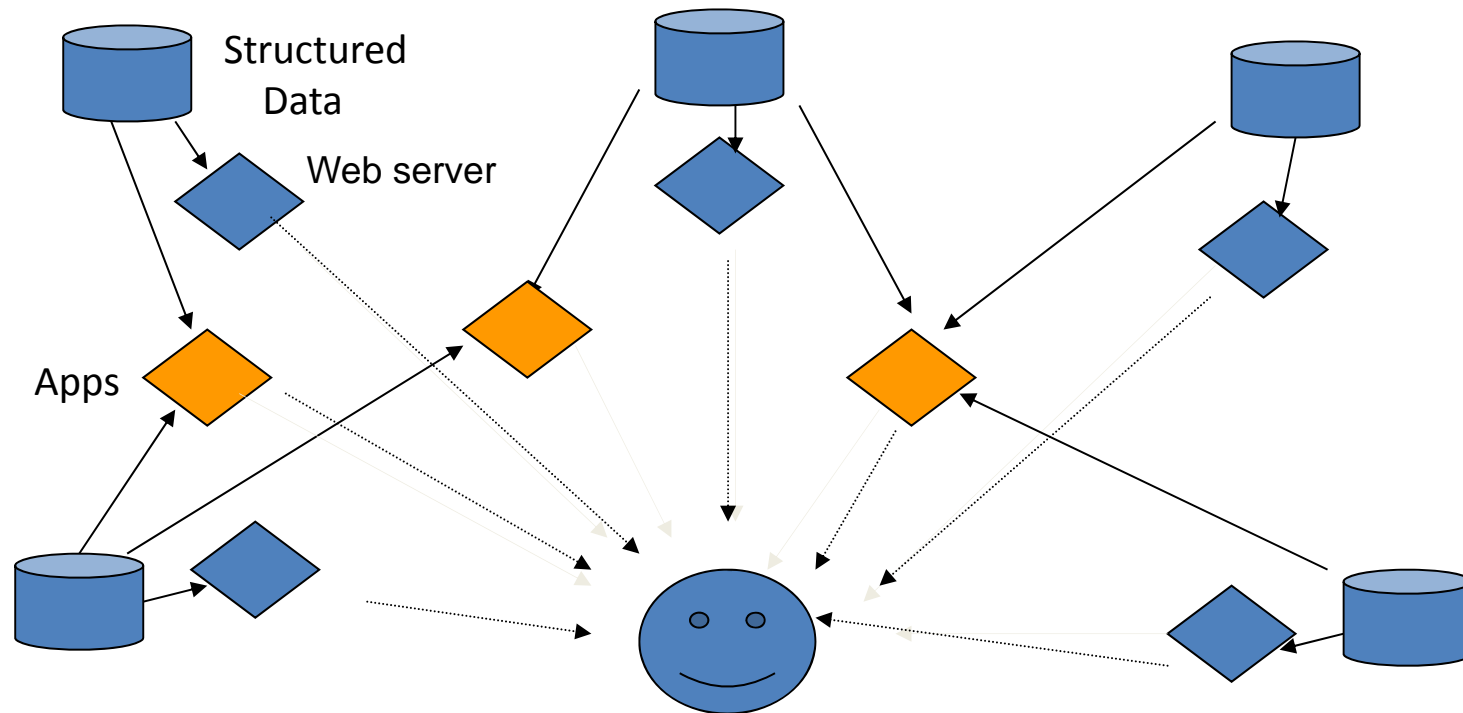
# About 17 years ago, ...

- People started thinking about structured data on the web
  - A few people from Netscape, Microsoft and W3C got together @MIT
- Trying to make sense of a flurry of activity/proposals
  - XML, MCF, CDF, Sitemaps, ...
- There were a number of problems
  - PICS, Meta data, sitemaps, ...
- But one unifying idea

# Context: The Web for humans



# Goal: Web for Machines & Humans



# What does that mean?

IMDb Find Movies, TV shows, Celebrities and more... All

Movies TV News Showtimes Community

**Chuck Norris**  
Actor | Music Department | Producer

Chuck Norris attended North Torrance High School in September, 1955 until his graduation in 1958. He is one of several storied alumni from the school. Other NHS alumni include Bob Hite (1943-1981), the lead singer of "Canned Heat," Chris Demaria, a professional baseball player in the Kansas City and Milwaukee Brewers organizations... [See full bio](#)

**Born:** Carlos Ray Norris  
March 10, 1940 in Ryan, Oklahoma, USA

---

Celebrities » Chuck Norris

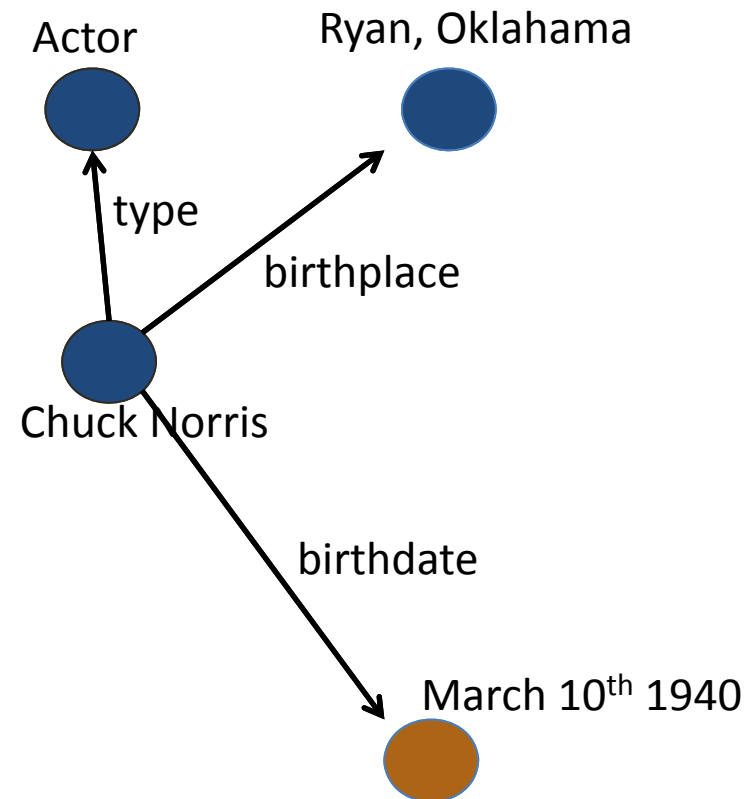
**Chuck Norris**

Highest Rated: 🍅 100% [Return of the Dragon \(The Way of the Dragon\) \(1973\)](#)  
Lowest Rated: 🍌 0% [Firewalker \(1986\)](#)

Birthdate: Mar 10, 1940  
Birthplace: Ryan, Oklahoma, USA

Bio: American action star Chuck Norris first learned martial arts while serving in the U.S. Marine Corps. From 1968 through 1974, he held the world's middleweight karate championship. During this period, he made his film debut in *The Wrecking Crew* (1968) and his TV bow on *Kung Fu* in 1972. Thanks...

[Full Chuck Norris Bio](#)



# How do we get there?

- How does the author give us the graph
  - Data Model: Graph vs tree vs ...
  - Syntax
  - Vocabulary
  - Identifiers for objects
- Why should the author give us the graph?

# Going depth first

- Many heated battles
  - Lot of proposals, standards, companies, ...
- Data model
  - Trees vs DLGs vs Vertical specific vs who needs one?
- Syntax
  - XML vs RDF vs json vs ...
- Model theory anyone
  - We need one vs who cares vs what's that?



# Timeline of 'standards'

- '96: Meta Content Framework (MCF) (Apple)
- '97: MCF using XML (Netscape) → RDF, CDF
- '99 -- : RDF, RDFS
- '01 -- : DAML, OWL, OWL EL, OWL QL, OWL RL
- '03: Microformats
- And many many many more ... SPARQL, Turtle, N3, GRDDL, R2RML, FOAF, SIOC, SKOS, ...
- Lots of bells & whistles: model theory, inference, type systems, ...

# But something was missing ...

- Fewer than 1000 sites were using these standards
- Something was clearly missing and it wasn't more language features
- We had forgotten the 'Why' part of the problem
- The RSS story

# '07 - :Rise of the consumers

- Yahoo! Search Monkey, Google Rich Snippets, Facebook Open Graph
- Offer webmasters a simple value proposition
- Search engines to webmasters:
  - You give us data ... we make your results nicer
- Usage begins to take off
  - 1000x increase in markup'ed up pages in 3 years

# Yahoo Search Monkey

- Give websites control over snippet presentation
- Moderate adoption
  - Targeted at high end developers
  - Too many choices



# Google Rich Snippets: Reviews

Google

xanh mountain view

About 8,810,000 results (0.26 seconds)

Search


Advanced search

Everything  
Images  
Videos  
News  
Shopping  
More

Mountain View, CA  
Change location

Show search tools

**XANH RESTAURANT**  
xanhrestaurant.com/ - Cached - Similar



**Xanh Restaurant**  
Place page  
110 Castro Street  
Mountain View, CA 94041  
(650) 964-1888  
Tram: Mountain View Station (1)  
Get directions - Is this accurate?  
Open Mon-Thu 11:30am-2:30pm, 5pm-10pm;  
Fri-Sat 11:30am-2:30pm, 5pm-11pm; Sun  
5pm-10pm  
★★★★★ 1587 reviews - Write a review  
"Pros: Good decor Cheap food ( \$12 lunch)  
Fast service ... Cons: 1. Parking is a ..."  
- yelp.com

→ **Xanh Restaurant - Mountain View, CA**  
★★★★★ 913 reviews - Price range: \$\$  
913 Reviews of Xanh Restaurant "I've been here for dinner and lunch a few times. The dinner is good but I heart their lunch buffets.  
www.yelp.com > Restaurants > Vietnamese - Cached - Similar

→ **XANH Restaurant - Mountain View, CA | OpenTable**  
★★★★★ 228 reviews - Price range: \$30 and under  
228 Reviews for XANH in Mountain View. "Nice/fancy presentation but food is so so. The fish is overdone. In general, it is lacking the freshness..."  
www.opentable.com/xanh - Cached - Similar

→ **Xanh - Mountain View | Urbanspoon**  
★★★★★ 12 reviews - Price range: \$25 on up per entree  
Xanh, Vietnamese Restaurant in Mountain View. See the menu, 3 photos, 7 critic reviews, 1 blog post and 4 user reviews. Reviews from critics, food blogs and ...  
www.urbanspoon.com > ... > Peninsula > Mountain View - Cached - Similar

# Google Rich Snippets: Events

Web Images Videos Maps News Mail more -

Google


britney spears concert


About 42,800,000 results (0.26 seconds) Advanced search


Everything  
Images  
Videos  
News  
Shopping  
Realtime  
More


Mountain View, CA  
Change location

Any time  
Latest  
Past 24 hours  
Past 2 days  
Past week  
Past month  
Past year  
Custom range...

**Britney Spears Concert - Cheap Britney San Jose Tickets.**  Ads  
Check Our Prices. Save 15% or More.  
ticketliquidator.com is rated ★★★★★ 80 reviews  
[www.ticketliquidator.com](http://www.ticketliquidator.com)


**Britney Spears Tickets - Britney Spears 2011 concert tour.**   
Tickets to **Britney Spears** concerts.  
[britney.concertticketcenter.com](http://britney.concertticketcenter.com)

**Britney Spears tickets, concerts and tour dates. Official ...**   
Results 1 - 10 of 34 ... Find and buy **Britney Spears** tickets at Ticketmaster.com. ... **Britney Spears** Tickets and **Concert** Dates ...  
[www.ticketmaster.com/Britney-Spears-tickets/artist/723427](http://www.ticketmaster.com/Britney-Spears-tickets/artist/723427) - Cached - Similar

**Britney Spears Tickets, Britney Spears Concert - Britney Tickets ...**   
The widest selection of **Britney Spears** Tickets - GOtickets is your source for premium **Concert** Tickets. so BUY with confidence and KNOW your order is secure ...

Thu, Jun 16	Power Balance Pavilion, Sacramento, CA
Sat, Jun 18	HP Pavilion at San Jose, San Jose, CA
Mon, Jun 20	Staples Center, Los Angeles, CA

[www.gotickets.com/concert/britney\\_spears.php](http://www.gotickets.com/concert/britney_spears.php) - Cached - Similar

**Wet Weather Sends Britney Spears Concert Indoors - News Story ...**   
Mar 23, 2011 ... SAN FRANCISCO -- A nationally televised **concert** featuring singer **Britney**



# Google Rich Snippets: Recipe View



ina garten cakes



Search

About 69,700 results (0.22 seconds)

[Advanced search](#)

Everything

Images

Videos

News

Shopping

Recipes

More

Ingredients Yes No

coffee ☐ ☐

coconut ☐ ☐

sour cream ☐ ☐

lemon ☐ ☐

chocolate ☐ ☐

vanilla extract ☐ ☐

baking soda ☐ ☐

maple syrup ☐ ☐

Any cook time

[Less than 15 min](#)

[Less than 30 min](#)

[Less than 60 min](#)

Any calories

[Less than 100 cal](#)

[Less than 300 cal](#)

[Less than 500 cal](#)

## [Coconut Cake](#)



★★★★★ 324 reviews - 1 hr 55 min

Food Network invites you to try this Coconut **Cake** recipe from **Ina Garten**.

Ingredients: butter, sugar, eggs, vanilla extract, almonds, flour, baking ...

[www.foodnetwork.com](#) > [Recipes](#) > [Baking](#) - [Cached](#) - [Similar](#)

## [Sour Cream Coffee Cake](#)



★★★★★ 246 reviews - 1 hr 40 min

Food Network invites you to try this Sour Cream Coffee **Cake** recipe from **Ina Garten**.

Ingredients: butter, sugar, eggs, vanilla extract, sour cream, flour, baking ...

[www.foodnetwork.com](#) > [Recipes](#) > [Baking](#) - [Cached](#) - [Similar](#)

## [Lemon Yogurt Cake](#)



★★★★★ 433 reviews - 1 hr 20 min

Food Network invites you to try this Lemon Yogurt **Cake** recipe from **Ina Garten**.

Ingredients: flour, baking powder, yoghurt, sugar, eggs, lemon, vanilla extract ...

[www.foodnetwork.com](#) > [Recipes](#) > [Baking](#) - [Cached](#) - [Similar](#)

## [Beatty's Chocolate Cake](#)



★★★★★ 1103 reviews - 1 hr 35 min

Food Network invites you to try this Beatty's Chocolate **Cake** recipe from **Ina Garten**.

Ingredients: butter, flour, sugar, cocoa, baking soda, baking powder, milk ...

[www.foodnetwork.com](#) > [Recipes](#) > [Baking](#) - [Cached](#) - [Similar](#)

## [Lemon Cake](#)



★★★★★ 223 reviews - 1 hr 30 min

Food Network invites you to try this Lemon **Cake** recipe from **Ina Garten**.

Ingredients: butter, sugar, eggs, lemon, flour, baking powder, baking soda ...

[www.foodnetwork.com](#) > [Recipes](#) > [Baking](#) - [Cached](#) - [Similar](#)

## [Flag Cake](#)



★★★★★ 132 reviews - 1 hr 15 min

Food Network invites you to try this Flag **Cake** recipe from **Ina Garten**.

# Google Rich Snippets

- Multi-syntax
- Adhoc vocabulary for each vertical
- Very clear carrot
- Lots of experimentation on UI
- Moderately successful: 10ks of sites
- Scaling issues with vocabulary



# Situation in 2010

- Too many choices/decisions for webmasters
  - Divergence in vocabularies
    - Too much fragmentation
    - N versions of person, address, ...
- A lot of bad/wrong markup
  - ~25% for micro-formats, ~40% with RDFa
  - Some spam, mostly unintended mistakes
- Absolute adoption numbers still rather low
  - Less than 100k sites

# Schema.org

- Work started in August 2010
  - Google, Yahoo!, Microsoft & then Yandex (Baidu, sort of)
- Goals:
  - One vocabulary understood by all the search engines
  - Make it very easy for the webmaster
- It is A vocabulary. Not The vocabulary.
  - Webmasters can use it together other vocabs
  - We might not understand the other vocabs. Others might

# Schema.org: Major sites

- News: [Nytimes](#), [guardian.com](#), [bbc.co.uk](#),
- Movies: [imdb](#), [rottentomatoes](#), [movies.com](#)
- Jobs / careers: [careerjet.com](#), [monster.com](#), [indeed.com](#)
- People: [linkedin.com](#),
- Products: [ebay.com](#), [alibaba.com](#), [sears.com](#), [cafepress.com](#), [sulit.com](#), [fotolia.com](#)
- Videos: [youtube](#), [dailymotion](#), [frequency.com](#), [vinebox.com](#)
- Medical: [cvs.com](#), [drugs.com](#)
- Local: [yelp.com](#), [allmenus.com](#), [urbanspoon.com](#)
- Events: [wherevent.com](#), [meetup.com](#), [zillow.com](#), [eventful](#)
- Music: [last.fm](#), [myspace.com](#), [soundcloud.com](#)

# Schema.org: categories

- Most used categories by occurrence
  - Person, Offer, Product, PostalAddress, VideoObject, ImageObject, BlogPosting, WebPage, Article, AggregateRating, LocalBusiness, Place, Organization, MusicRecording, JobPosting, Recipe, Book, Movie, Blog, Photograph, ImageGallery
- Most used categories by domains
  - ImageObject, WebPage, PostalAddress, BlogPosting, Product, Person, Offer, Article, LocalBusiness, Organization, Blog, AggregateRating, Review, VideoObject, Place, Event, Rating, AudioObject, MusicRecording, Store

# Schema.org: properties

- Top properties by occurrence
  - name, url, image, description, offers, author, price, thumbnailUrl, datePublished, addressLocality, address, itemOffered, duration, streetAddress, isFamilyFriendly, priceCurrency, playerType, paid, regionsAllowed, postalCode, hiringOrganization, jobLocation,
- Top properties by domain
  - Name, description, url, image, contentURL, address, author, telephone, price, postalCode, offers, ratingValue, priceCurrency, datePublished, addressRegion, availability, email, bestRating, creator, review, location, startDate

# Applications

- Applications drive adoption
- First generation of applications
  - Rich presentation of search results
- Many new applications are coming up
  - On search page and beyond

# Newer Applications: Knowledge Graph

proxy

Maps Shopping Videos More Search tools

its (0.28 seconds)

[r Proxy \(1994\) - IMDb](#)  
[le/tt0110074/](#) ▼  
7.3/10 - 48442 votes  
en. With Tim Robbins, Jennifer Jason Leigh, Paul Newman, Charles  
isiness graduate is installed as president of a ...  
- [The Hudsucker Proxy Trailer](#) - [Trivia](#) - [Quotes](#)

[r Proxy - Wikipedia, the free encyclopedia](#)  
[wiki/The\\_Hudsucker\\_Proxy](#) ▼  
oxy is a 1994 screwball comedy film co-written, produced, and  
d Ethan Coen. Sam Raimi co-wrote the script and served as ...  
pment - [Soundtrack](#)

[r Proxy - Rotten Tomatoes](#)  
[es.com/m/hudsucker\\_proxy/](#) ▼  
56% - 39 reviews  
and Ethan Coen concocted this stylish screwball comedic amalgam  
Howard Hawks. Tim Robbins stars as Norville ...

[r Proxy Official Trailer #1 - Charles Durning Movie ...](#)  
[www.youtube.com/watch?v=dBa8p0NFwM8](#) ▼  
Jan 10, 2012 - Uploaded by oldhollywoodtrailers  
The Hudsucker Proxy Trailer - Directed by Joel Coen and starring



**The Hudsucker Proxy**  
1994 Film

★★★★☆ 7.3/10 - IMDb  
★★★★☆ 56% - Rotten Tomatoes  
★★★★☆ 53% - Metacritic



The Hudsucker Proxy is a 1994 screwball comedy film co-written, produced, and directed by Joel and Ethan Coen. Sam Raimi co-wrote the script and served as second unit director. Wikipedia

**Release date:** March 11, 1994 (USA)  
**Directors:** [Joel Coen](#), [Ethan Coen](#)  
**Initial DVD release:** May 18, 1999  
**MPAA rating:** PG  
**Screenplay:** [Joel Coen](#), [Ethan Coen](#), [Sam Raimi](#)

**Cast**



# Newer Applications: Knowledge Graph



[Web](#) [Images](#) [Maps](#) [Shopping](#) [News](#) [More ▾](#) [Search tools](#)

About 113,000,000 results (0.95 seconds)

Ad related to **of monsters and men**

**Of Monsters and Men - My Head Is an Animal**  
[itunes.apple.com/OfMonstersAndMen](https://itunes.apple.com/OfMonstersAndMen)  
Featured video: [King and Lionheart](#)  
In stores now  
Feat. Little Talks, Mountain Sound, & King and Lionheart  
iTunes \$7.99

**Of Monsters and Men** | The Official website for **Of Monsters and Men**  
[www.ofmonstersandmen.com/](http://www.ofmonstersandmen.com/) ▾  
The Official website for **Of Monsters and Men**.

**Of Monsters and Men** - Wikipedia, the free encyclopedia  
[en.wikipedia.org/wiki/Of\\_Monsters\\_and\\_Men](http://en.wikipedia.org/wiki/Of_Monsters_and_Men) ▾  
**Of Monsters and Men** is a six-piece, English-language, indie folk/pop band from Iceland, formed in 2010. The members are lead singer/guitarist Nanna Bryndís ...  
[My Head Is an Animal](#) - [Little Talks](#) - [Nanna Bryndís Hilmarsdóttir](#) - [Mountain Sound](#)


**Of Monsters and Men** - YouTube  
[www.youtube.com/artist/of-monsters-and-men](http://www.youtube.com/artist/of-monsters-and-men) ▾  
After winning the nationwide 2010 battle-of-the-bands competition Musiktíraunir in their native Iceland, six-piece chamber pop group **Of Monsters and Men** wer...

**Of Monsters and Men** | Facebook  
[www.facebook.com/ofmonstersandmen](http://www.facebook.com/ofmonstersandmen) ▾  
**Of Monsters and Men**. 784817 likes · 52452 talking about this.

**Of Monsters and Men** – Free listening, concerts, stats, & pictures at ...  
[www.last.fm/music/Of+Monsters+and+Men](http://www.last.fm/music/Of+Monsters+and+Men) ▾  
Top tracks from **Of Monsters and Men**: [Little Talks](#), [Dirty Paws](#) & more. **Of Monsters and Men** are an indie folk band which formed in 2010 in Reykjavík, Iceland.

**Of Monsters and Men** | Free Music, Tour Dates, Photos, Videos  
[www.myspace.com/ofmonstersandmenmusic](http://www.myspace.com/ofmonstersandmenmusic) ▾  
**Of Monsters and Men**'s official profile including the latest music, albums, songs, music videos and more updates.

**Of Monsters and Men** Play 'Little Talks' on 'SNL' - Video | Rolling St...

[More images](#)

## Of Monsters and Men

Band

Of Monsters and Men is a six-piece, English-language, indie folk/pop band from Iceland, formed in 2010. [Wikipedia](#)

**Origin:** [Iceland](#)

**Albums:** [My Head Is an Animal](#), [Little Talks](#)

**Members:** [Nanna Bryndís Hilmarsdóttir](#), [Brynjar Leifsson](#), [More](#)

**Lead singers:** [Nanna Bryndís Hilmarsdóttir](#), [Ragnar Þórhallsson](#)

**Record labels:** [Universal Music Group](#), [Republic Records](#), [Universal Republic Records](#)

### Upcoming events

May 16 Thu	<a href="#">Of Monsters and Men</a> Oakland, CA (near you)
May 19 Sun	<a href="#">Live 105's BFD 2013</a> Mountain View, CA (near you)
May 13 Mon	<a href="#">Of Monsters and Men</a> Portland, OR

### Songs

<a href="#">Little Talks</a>	2011	<a href="#">Into the Woods</a>
<a href="#">Mountain Sound</a>	2012	<a href="#">My Head Is an Animal</a>



# Non web search Applications

- Searching for Veteran friendly jobs

The screenshot shows the 'National Resource Directory' website, which is a platform for finding veteran-friendly jobs. The header features the site's name and mission: 'Connecting Wounded Warriors, Service Members, Veterans, Their Families and Caregivers with Those Who Support Them'. Below the header, the 'Veterans Job Bank' section is highlighted, with a 'Beta' label. The search interface includes a 'Keyword(s)' field with 'engineer' entered, a 'MOS/MOC(s)' field with a placeholder 'Enter a code or title', and a 'Location' field with 'San Jose, CA' selected. A list of nearby cities is shown below the location field, including Sunol, Campbell, Mount Hamilton, Sunnyvale, and Newark, all of which are checked. To the right of the search filters, there are options to 'Sort listings by: Relevancy' and 'Date Posted'. Two job listings are displayed: 'Engineering Equipment Operator' and 'Network Software Engineer (Libraries and System Software)'. Each listing includes a job summary, a 'View More Details' link, and a 'Feedback' link. The 'Engineering Equipment Operator' listing also shows the employer as 'Department Of Agriculture, Forest Service' and the salary as '43680.00'.

**National Resource Directory**  
Connecting Wounded Warriors, Service Members, Veterans,  
Their Families and Caregivers with Those Who Support Them

[Home](#) » Veterans Job Bank

**Veterans Job Bank**  
Beta

Keyword(s) [Help](#)  
engineer

MOS/MOC(s)  
Enter a code or title [Add](#)

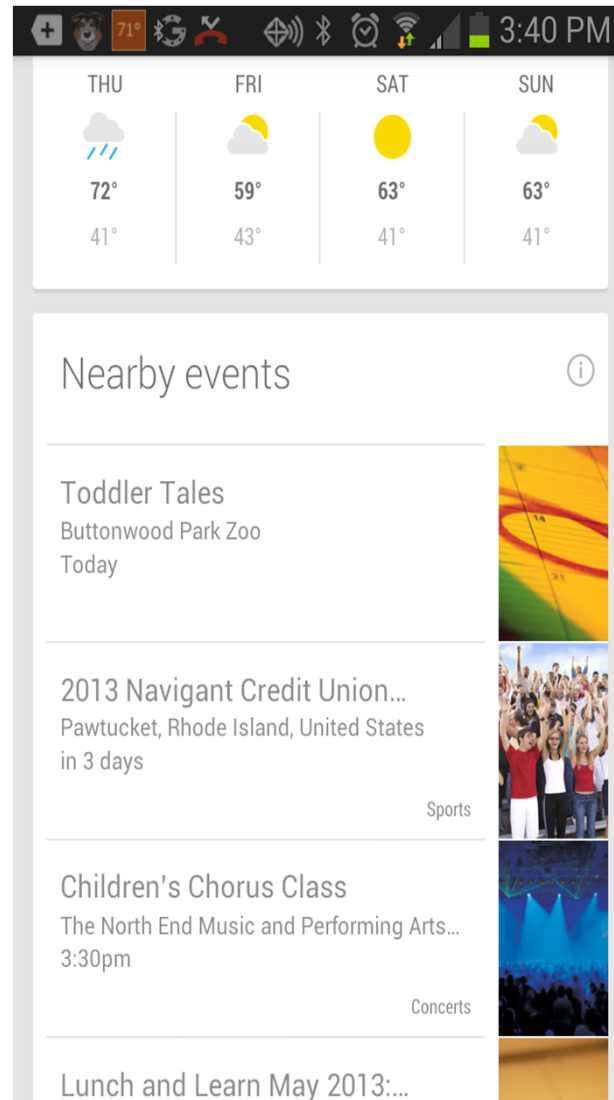
Location  
San Jose, CA  
☒ Include nearby cities  
☒ Sunol, CA  
☒ Campbell, CA  
☒ Mount Hamilton, CA  
☒ Sunnyvale, CA  
☒ Newark, CA

Sort listings by: Relevancy | [Date Posted](#)

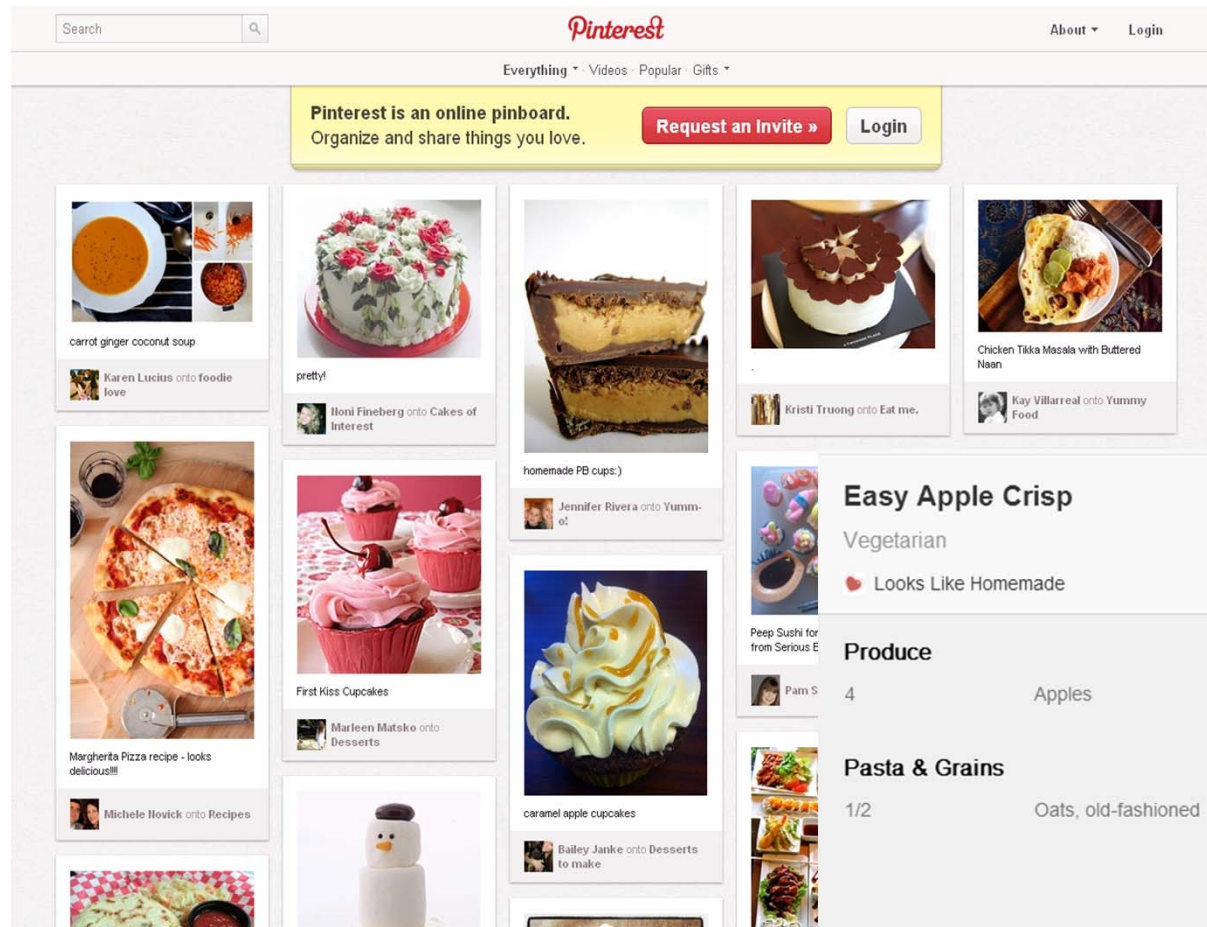
[Engineering Equipment Operator](#) [Feedback](#)  
JOB SUMMARY: A career with the Forest Service will challenge you to manage and care for more than 193 million acres of our nation's most magnificent lands, conduct research through a network...  
[View More Details](#)  
Department Of Agriculture, Forest Service • Multiple Locations • 43680.00  
Posted 315 days ago

[Network Software Engineer \(Libraries and System Software\)](#) [Feedback](#)  
Job Description The Intel Fabric Software development team is looking for a software development engineer to work in our Santa Clara design center to facilitate the delivery of the next generation...  
[View More Details](#)  
Intel Corporation • Santa Clara, California  
Posted 474 days ago

# Non search applications: Google Now



# Pinterest: Schema.org for Rich Pins



## Easy Apple Crisp

Vegetarian

♥ Looks Like Homemade

### Produce

4

Apples

### Pasta & Grains

1/2

Oats, old-fashioned

### Cooking & Baking

1/2 cup

All-purpose flour

3/4 cup

Brown sugar, packed

3/4 tsp

Cinnamon, ground

1/4 tsp

Cloves, ground

1/2 tsp

Nutmeg, ground

### Dairy

1/3 cup

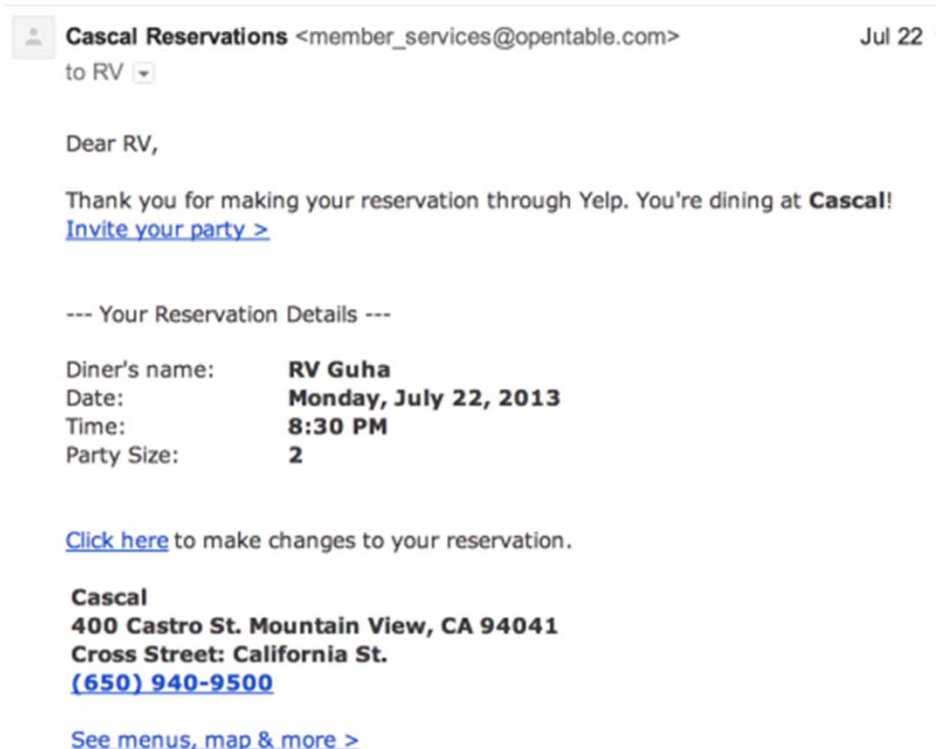
Butter

Make this on Looks Like Homemade



# Non search Applications

- Open Table website → confirmation email → Android Reminder



```
<span itemscope  
  itemtype="http://schema.org/Restaurant"  
  itemid="/restaurant">  
  Cascal  
</span>  
  
....  
<span itemprop="address"  
  itemscope  
  itemtype="http://schema.org/PostalAddress">  
  <span itemprop="streetAddress">  
    400 Castro St. Mountain View, CA 94041  
  </span>  
</span>  
  
...
```

# Schema.org principles: Simplicity

- Simple things should be simple
  - For webmasters, not necessarily for consumers of markup
  - Webmasters shouldn't have to deal with N namespaces
- Complex things should be possible
  - Advanced webmasters should be able to mix and match vocabularies
- Syntax
  - Microdata, usability studies
  - RDFa, json-ld, ...

# Schema.org principles: Simplicity

- Can't expect webmasters to understand Knowledge Representation, Semantic Web Query Languages, etc.
- It has to fit in with existing workflows
- Avoid KR system driven artifacts
  - domainIncludes/rangeIncludes
  - No classes like 'Agent'
  - Categories and attributes should be concrete

# Schema.org principles: Simplicity

- Copy and edit as the default mode for authors
  - It is not a linear spec, but a tree of examples
- Vocabularies
  - Authors only need to have local view
  - But schema.org tries to have a single global coherent vocabulary

# Schema.org principles: Incremental

- Started simple
  - ~ 100 categories at launch
- Applies to every area
  - Add complexity after adoption
  - now ~1200 vocab items
  - Go back and fill in the blanks
- Move fast, accept mistakes, iterate fast

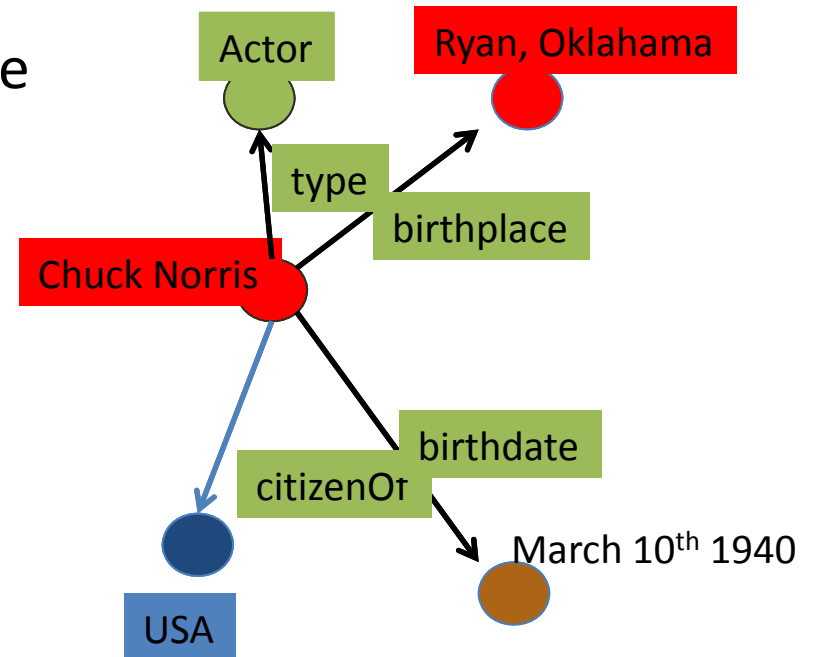


# Schema.org Principles: URIs

- ~1000s of terms like Actor, birthdate
  - ~10s for most sites
  - Common across sites

- ~10ks of terms like USA
  - External enumerations

- ~1b-100b terms like Chuck Norris and Ryan, Oklahoma
  - Cannot expect agreement on these
  - Reference by description
  - Consumers can reconcile entity references



# Schema.org Principles: Collaborations

- Most discussions on public W3C lists
- Work closely with interest communities
- Work with others to incorporate their vocabularies
  - We give them attribution on schema.org
  - Webmasters should not have to worry about where each piece of the vocabulary came from
  - Webmasters can mix and match vocabs

# Schema.org Principles: Collaborations

- IPTC /NYTimes / Getty with rNews
- Martin Hepp with Good Relations
- US Veterans, Whitehouse, Indeed.com with Job Posting
- Creative Commons with LRMI
- NIH National Library of Medicine for Medical vocab.
- Bibextend, Highwire Press for Bibliographic vocabulary
- Benetech for Accessibility
- BBC, European Broadcasting Union for TV & Radio schema
- Stackexchange, SKOS group for message board
- Lots and lots and lots of individuals

# Schema.org Principles: Partners

- Partner with Authoring platforms
  - Drupal, Wordpress, Blogger, YouTube
- Drupal 8
  - Schema.org markup for many types
    - News articles, comments, users, events, ...
  - More schema.org types can be created by site author
  - Markup in HTML5 & RDFa Lite
  - Come out early 2014



# Recent/Upcoming Vocabularies

- Actions, Fleshing out Events
- Commerce: Orders, Reservations, ...
- Communication: Fleshing out TV, Radio, Email, Q&A, ...
- Media: Scholarly works, Comics, Serials
- Sports
- and many many more ...

# Big initiatives underway

- Representing time
  - Lot of triples with associated time interval
- Tabular / CSV data
  - Census data, Scientific data, etc.
  - Need mechanisms for external specification of the meaning of these tables

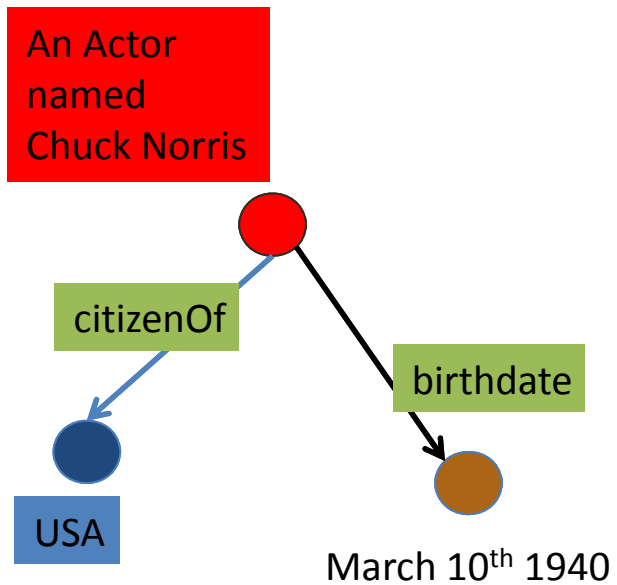
# Research ideas

- There are a large number of projects (e.g., Nell@cmu) that are trying to extract triples from the web
- Schema.org markup == Very large training set

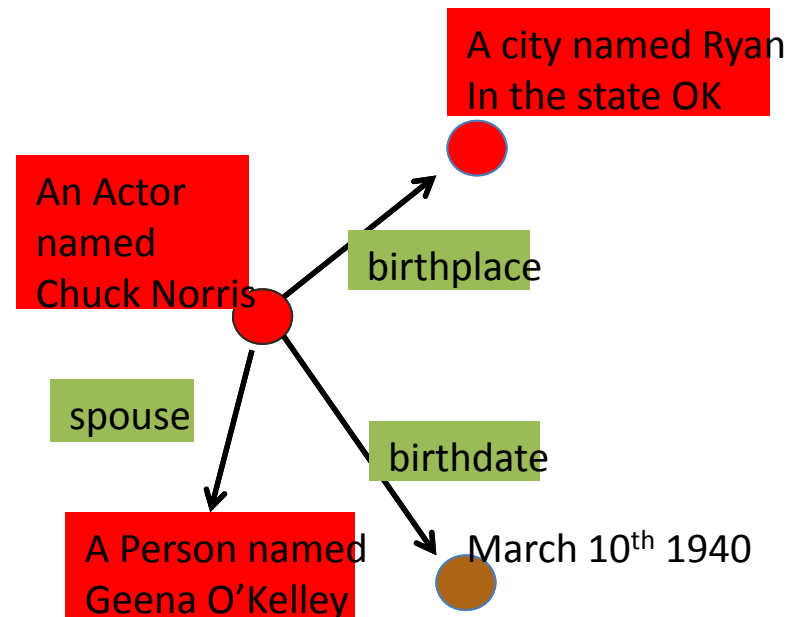
# Research Idea: Stich

- Billions of triples sharded across millions of sites
- Lots of common entities, but no cross pointers
- Need to put together the graph
  - Like solving the puzzle

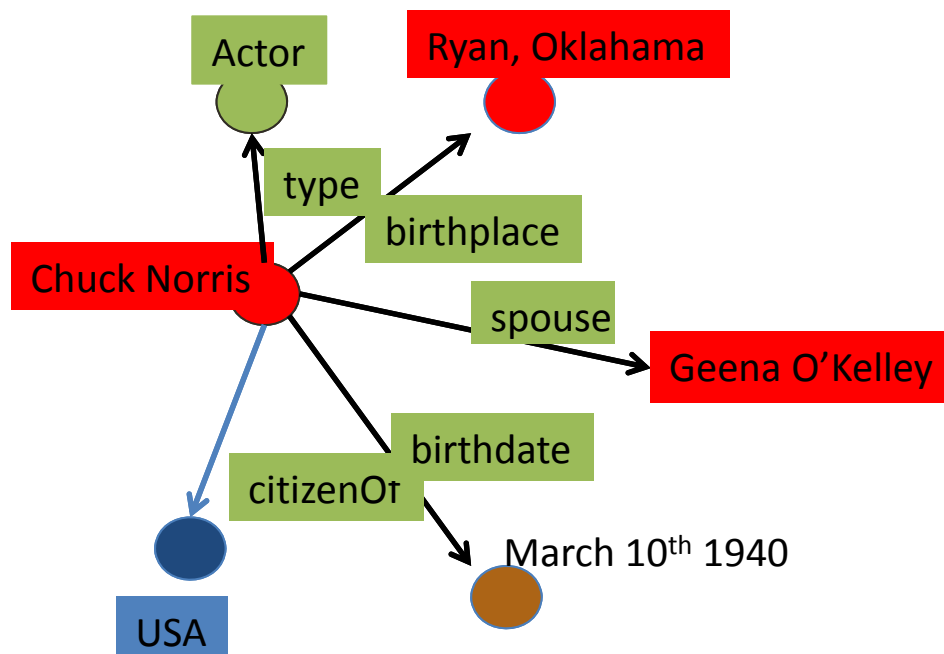




+



=



# Lessons from schema.org

- Schema.org has succeeded beyond anyone's imagination
- Make sure you have your carrot!
  - Carrots work much better than sticks!
- Find the right initial level of generality
- Start simple and iterate fast

This program was brought to you by the  
generous support from folks at Microsoft,  
Yandex, Yahoo!, Google and over five million  
webmasters

Questions?