

VerticalSearchWorks

Assuring
Broad Quality in Large-Scale Ontologies
Ontology Summit 2013
2013-02-14

Gavin Matthews, VP of Research

Assuring broad quality in large-scale ontologies



- Who are we and what do we do?
- Our ontology and ontology management tools
- Our search engine
- Ontology and product evaluation

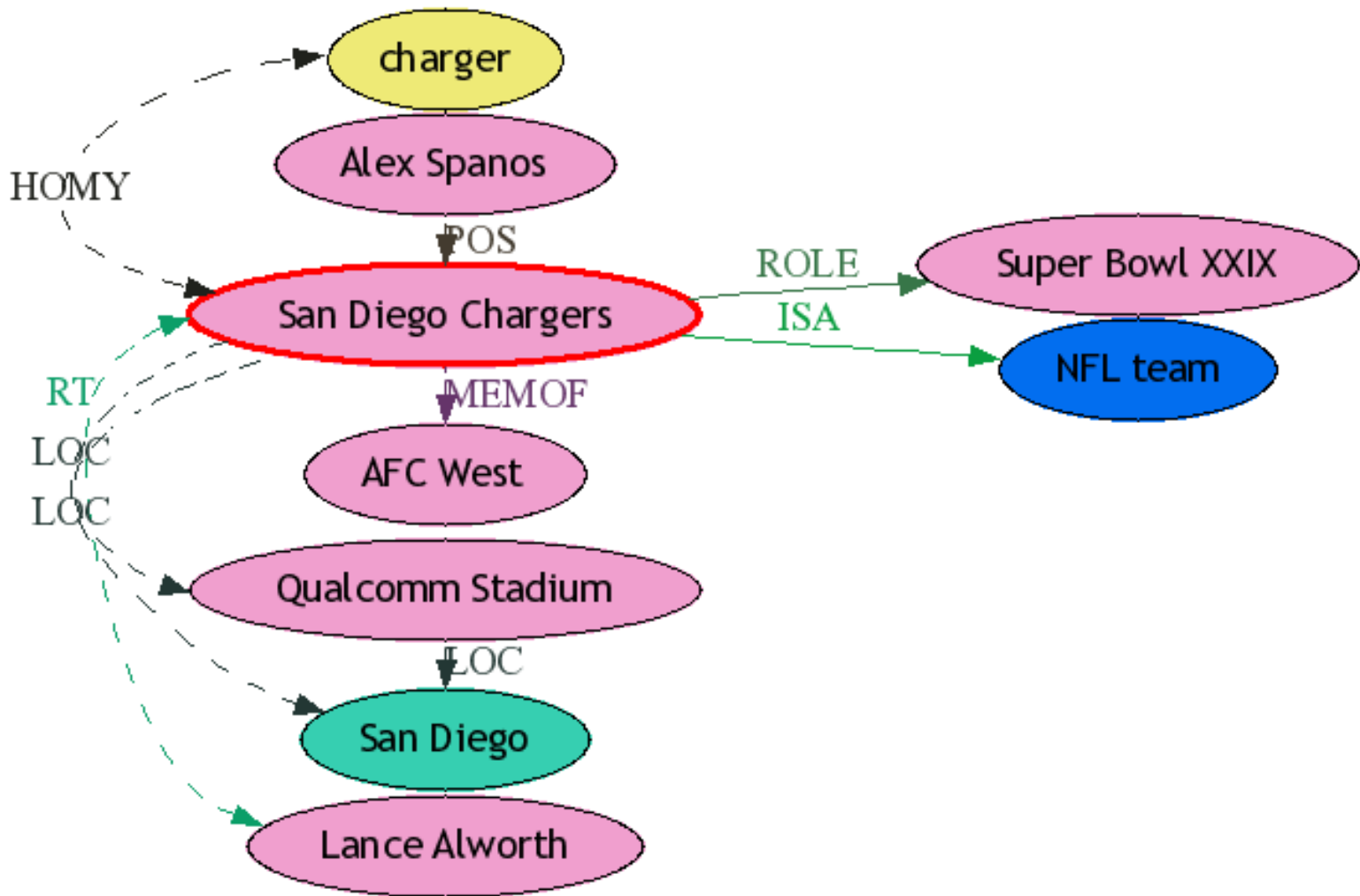


Who We Are

- Vertical Search Works was created in 2009
- Merger from First Light ERA and Convera (formerly Excalibur Technologies)
- Around 150+ employees
- 17 engineers (engineering & ontology group)
- Engineering located in Carlsbad, Sales in NY
- Two core businesses:
 - Semantic vertical search
 - Contextual advertising



Ontology: Visualization and Browsing





Ontology: Flat file Representation

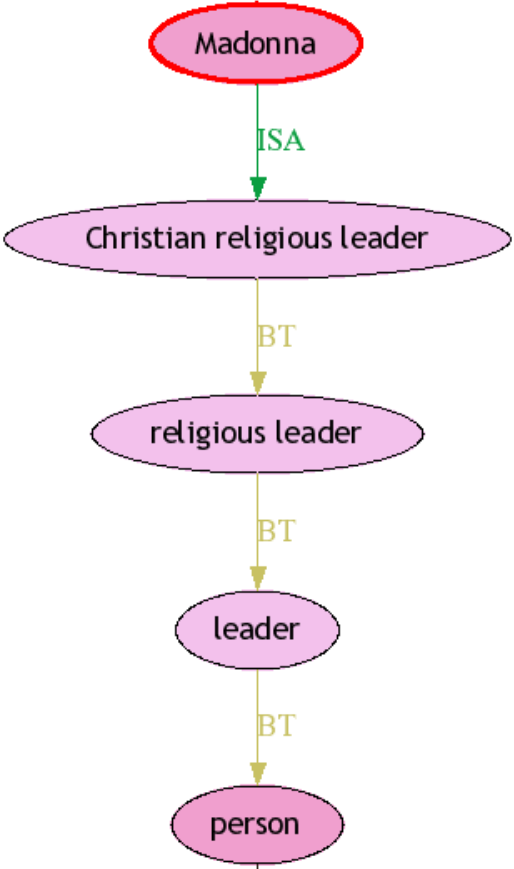
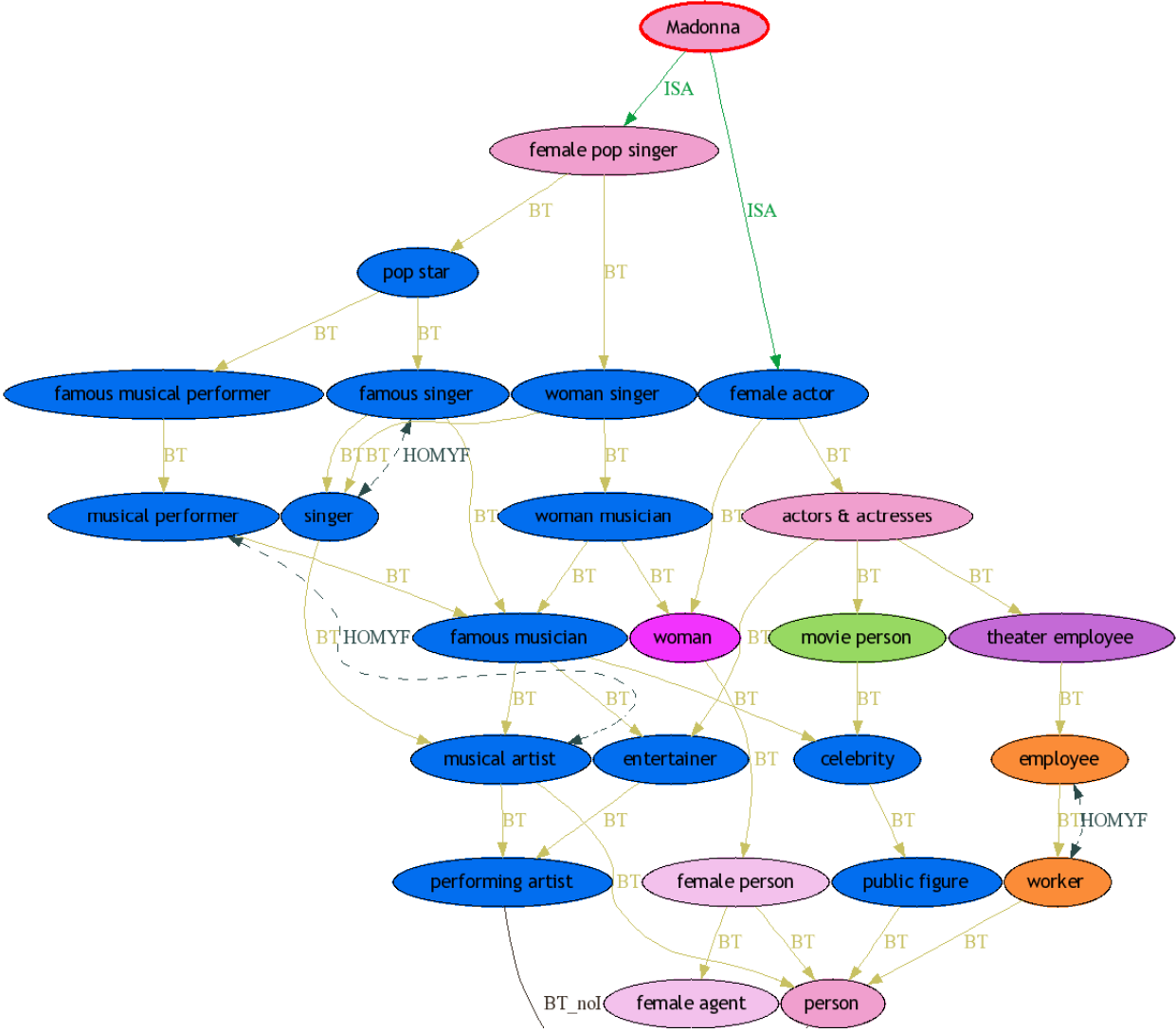
SYNSET:kec.004KN (San Diego Chargers)
HOMY:cpr.006RC (charger)
ISA:pct.01QPV (NFL team)
LOC:kec.00MEM (Qualcomm Stadium)
LOC:usg.1661377 (San Diego)
MEMOF:kec.CCI8S (AFC West)
ROLE:kec.CCFFX (Super Bowl XXIX)
RT:kec.005ZQ (Lance Alworth)

HOMEPAGE:<http://www.chargers.com/>
*,url
UF:*San Diego Chargers*
*,PN,gui
UF:*Chargers*
*,PN,case

SYNSET:pct.CBOA4 (waffle)
BT:cpr.CBIEJ (breakfast food)
BT:pct.0009L (baked good)
HOMY:men.00UB7 (geomys)
RT:cpr.003N4 (waffle iron)
RT:kec.CA9SS (Gaufres d'or)
UF:*waffle*
en,gui
UF:*gaufre*
fr,gui
UF:*gaufrette*
fr
UF:*wafle*
es,gui
UF:*gofre*
es



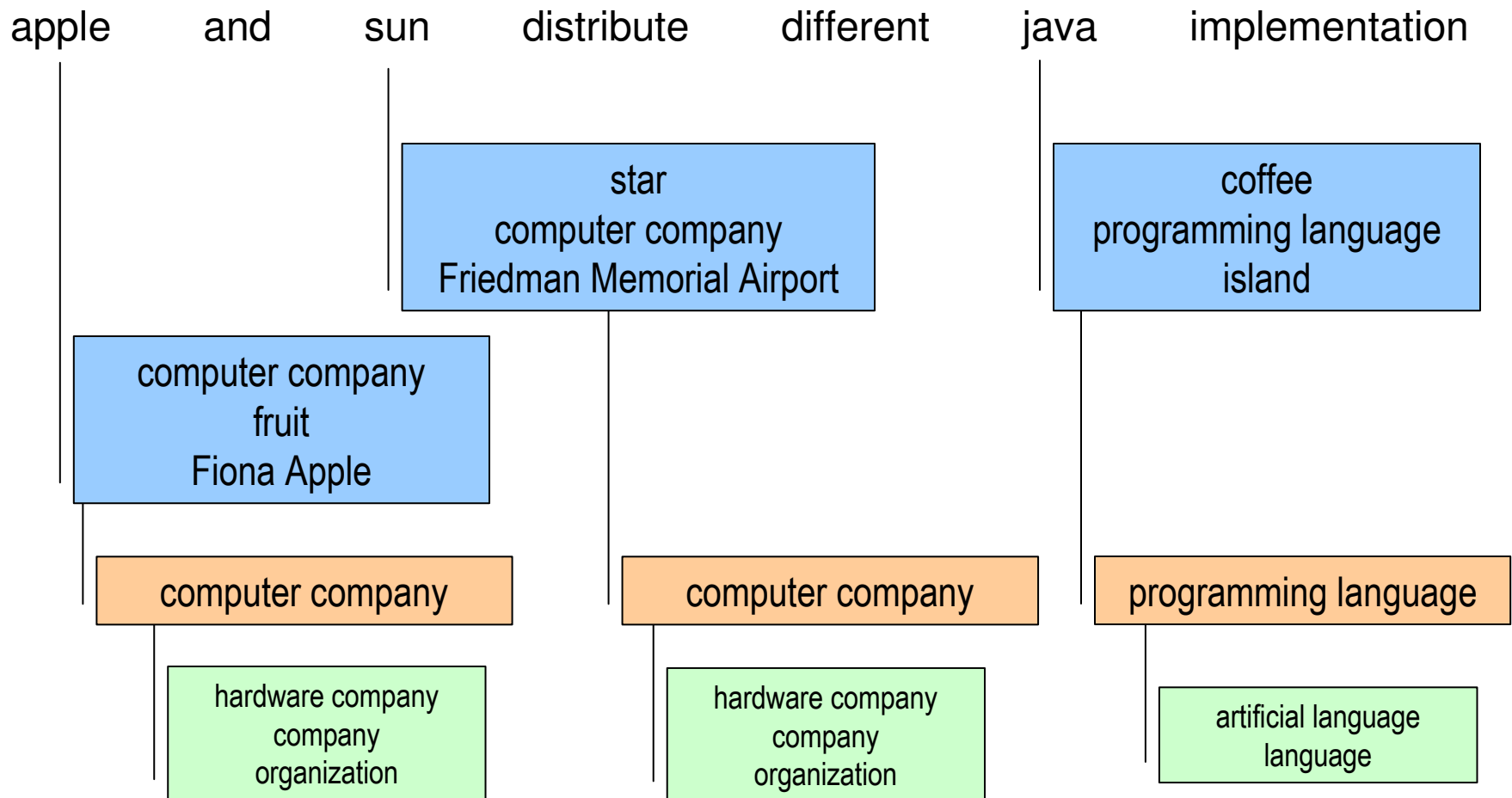
Two synsets with the expression “Madonna”





Using the ontology to disambiguate


“Apple and Sun distributed different Java implementations.”






Using the ontology for search



java 

▼ Narrower ▼ Broader ▼ Related

- espresso
- latte
- cappuccino
- instant coffee
- bean coffee
- iced coffee
- organic coffee
- flavored coffee
- decaffeinated coffee



[More ▶](#)

Categories:

Web Results

[Coffee, Perfect Cup of Coffee, How To Make Perfect Coffee, How To Make Coffee, Coffee Cupping, How To Cup Coffee](#)


Coffee - Organic Coffee for Your Health and the Environment. To gain a better understanding of what the term "**Organic Coffee**" really means we have put together this guide ...

whatscookingamerica.net


[How To Make Great Espresso](#)

college as a mean of keeping awake long enough to study. He used to drink **instant coffee** and drip **coffee**. ... An interesting fact is that hazelnut **flavored coffee** is one of the most popular flavors among consumers followed by Irish cream. ...


reluctantgourmet.com



The Perfect Cup of Coffee
Stop by today for high quality beans roasted in small batches to just the right degree of roast.
petalumacoffee.com



Organic Coffee
Enjoy a delicious cup of all natural coffee every morning! Click here to shop online today!
fourgreensteps.com



Wedding Special Coffee Roast
Let our specialty wedding coffee be the life of the party! Only \$3 per person with no minimum!
viennacoffeecompany.com

Privacy VSW FeatureLink

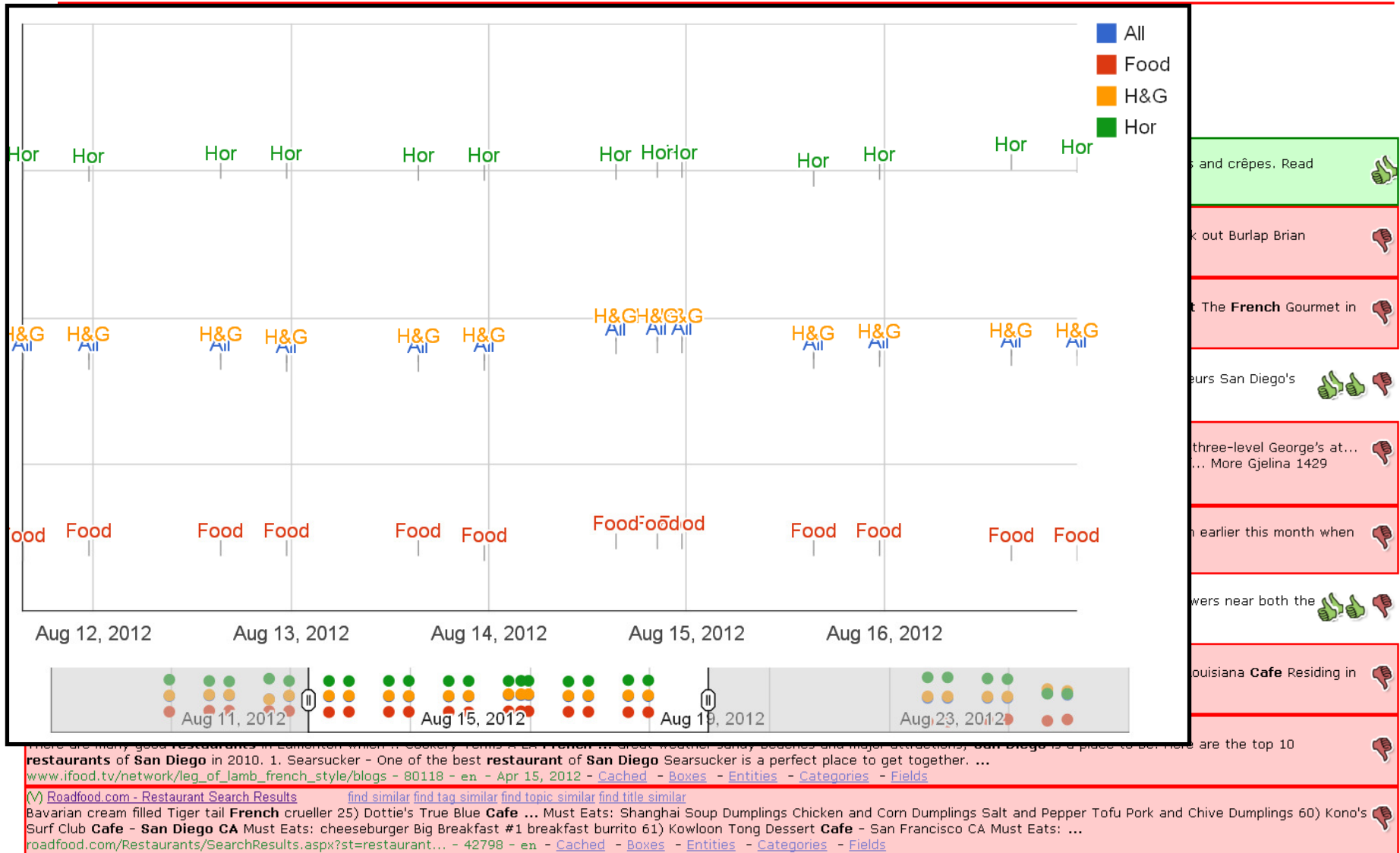


Ontology and Architecture Evaluation Cycle

- **Before check-in, locally**
 - Validation: Automated, syntax and inherent checks
- **Within the hour, from source control (CruiseControl)**
 - Validation
- **Nightly build and index**
 - Validation
 - Search Quality: Automated, black box, TREC methodology
 - Unit tests: Automated, glass box, semantic, regression and functional
- **Fortnightly push to production**
 - Validation, search quality, unit tests
- **Monthly QA cycle**
 - Manual desk check



Measuring search quality





Conclusion

- Large-scale ontology
 - Hyper-linked browser
 - Flexible visualization
- Ontology used for
 - Disambiguation
 - Search drilldown
 - Contextual advert matching
- Throughout development cycle
 - Intrinsic validation
 - Black box testing based on TREC
 - Glass box regression and function tests
 - Manual desk check

Where can you find our search portals?



VS4Food.com

VS4Family.com

VS4Entertainment.com

VS4Gardening.com

VS4HomeandGarden.com

VS4Woodworking.com