

deem

Formal Ontology for Sentiment Analysis

Adam Pease

Rearden Commerce

`adam.pease@reardencommerce.com`

code available at <http://www.ontologyportal.org>



Sentiment Analysis

- Emotional content of text
- Pilot project combining
 - _ Sentiment analysis (computational linguistics)
 - _ Concept extraction (linguistic semantics/ontology)
- Note this is just a pilot project and the computational linguistic method used is really basic, not state of the art
- Applications:
 - _ Fine grained search by features
 - _ Ratings by review, not by stars, and integrated across sources
 - _ Merge hotel ratings from different services that have different scales by using sentiment

sent-correct-format-US.xls - LibreOffice Calc

File Edit View Insert Format Tools Data Window Help

Find

Q1060 f(x) Σ =

	A	C	D	H	I	J	K	L	M	N	O	P	Q	R	
1	name	city	state/prov	avg sentime	House	Restaurant	Breakfast	Walking	Bed	Fireplace	City	Fabric	ShoreArea	Suite	Tra
1046	La Quinta Inn	South San Francisco	CA	8.00											
1047	Meadowood Resort	St. Helena	CA	95.00		10		7	1						
1048	Wine Country Inn	St. Helena	CA	81.00			0		3	3			3		
1049	El Bonita Motel	St. Helena	CA	65.00					1			1			
1050	Stinson Beach Motel	Stinson Beach	CA	49.00								4		5	
1051	Best Western Stockton Inn	Stockton	CA	28.00											
1052	La Quinta Inn	Stockton	CA	22.00		6									
1053	Stovepipe Wells Village	Stovepipe Wells	CA	-5.00											
1054	Best Western Trailside Inn	Susanville	CA	-2.00											
1055	High Country Inn	Susanville	CA	31.00			5	5			1				
1056	High Country Inn	Susanville	CA	51.00				10			1				
1057	The Foxes Inn of Sutter Creek	Sutter Creek	CA	10.00	5			0		5	0				
1058	Eureka Street Inn	Sutter Creek	CA	4.00	0						0	1			
1059	Sutter Creek Days Inn	Sutter Creek	CA	3.00						1					
1060	Mother Nature's Inn	Tahoe City	CA	22.00							20				
1061	River Ranch Lodge	Tahoe City	CA	-41.00				5			0				
1062	Sunnyside Steakhouse and Lodge	Tahoe City	CA	115.00					10		-1	10			
1063	Cottage Inn	Tahoe City	CA	21.00	0			0			1	0		0	
1064	Rustic Cottages	Tahoe Vista	CA	37.00	12						0			1	

Sheet 1 / 3 PageStyle_Sheet1 STD Sum=0 100%

Meadowood, St. Helena:

Restaurant: 10

“In recent years the elegant but unstuffy dining room has won rave reviews, becoming a destination restaurant.”

Marys Lake Lodge and Resort, CO:

Roadway: -8

“Not to mention it is very expensive and located in a place that doesn't get much sun so it's icy and cold; and the maintenance of roads is terrible in winter.”



By Amenities

- Laundry room (64)
- Free Meeting room broadband internet (95)
- Meeting room broadband internet (35)
- Free Meeting room wireless internet (127)
- Meeting room wireless internet (36)

[Check none](#)

Internet

Location (meeting rooms, lobby, rooms)

Cost (free, paid)

New Features

House, Restaurant, Breakfast, Walking, Bed, City, Fabric, ShoreArea, Suite, Transportation, WaterArea, RecreationOrExercise, ColorAttribute, Private, ArtWork, viewType, Roadway, Clothing, Near, Eating, Offering, Bathroom, Cooperation, NightTime, Nation, FloweringPlant, Lake, FluidContainer, enjoys, SocialInteraction, Game

- Linguistic elements, with multiple ways to say them, become canonicalized terms
- All concepts can become selectable “amenities”

Travel | Select Your Hotel

Your Search

[Modify search](#)

San Francisco
CA US
Check-in: Wed, Aug 24, 2011
Check-out: Thu, Aug 25, 2011
Guests: 1

Narrow Your Choices

Your search returned 309 Hotels
[Show all](#) | [Set to my Hotel preferences](#)

By Hotel name, Brand

Hotel name [Find](#)

Amenities/Features

- Bed
 - Food
 - Internet
 - Parking
 - Spa
 - Traffic
 - Vehicles
- [Check none](#)

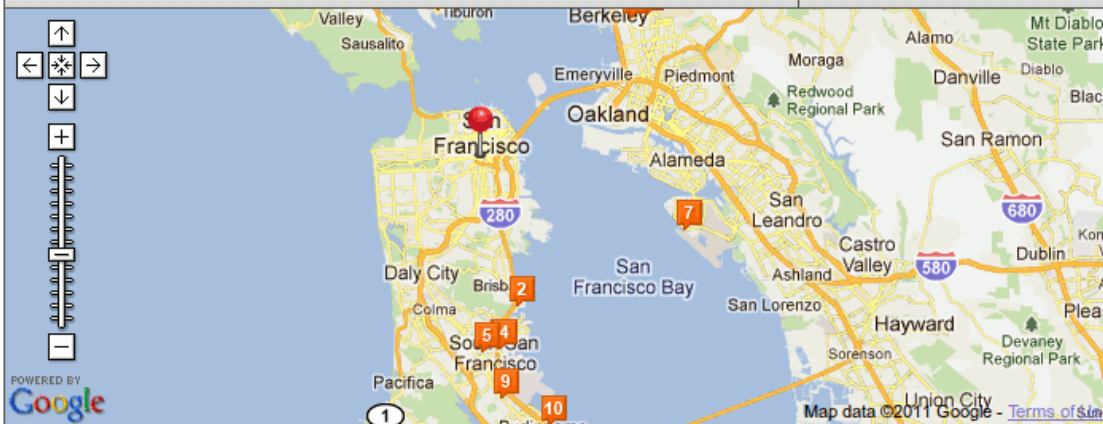
By Distance

Search within:

Hotel Search Map (collapse)

Show on map:

Company Locations



Sort by [Company Policy](#)

1 - 10 of 11

1 from \$160 per night

[See room rates](#)



Lodge at Tiburon

7.2 mi from San Francisco CA US ★★★★★

[Photos](#) | [Amenities](#)

Compare

[bed](#) reviews

[spa](#) reviews

[bar](#) reviews

2 from \$129 per night

No Rooms Available



Radisson Hotel SFO Arpt Sierra Point

7.2 mi from San Francisco CA US ★★★★★

[Photos](#) | [Amenities](#)

Compare

[spa](#) reviews

[bed](#) reviews



Travel | Select Your Hotel

Your Search

[Modify search](#)

San Francisco
CA US
Check-in: Wed, Aug 24, 2011
Check-out: Thu, Aug 25, 2011
Guests: 1

Narrow Your Choices

Your search returned 309 Hotels

[Show all](#) | [Set to my Hotel preferences](#)

By Hotel name, Brand

Hotel name

[Find](#)

Amenities/Features

- Bed
- Food
- Internet
- Parking
- Spa
- Traffic
- Vehicles

[Check none](#)

By Distance

Search within:

15 mi

1 from **\$160** per night[See room rates](#)

synXis

[Lodge at Tiburon](#)

7.2 mi from San Francisco CA US

[Photos](#) | [Amenities](#) Compare[bed](#) reviews[spa](#) reviews[bar](#) reviews

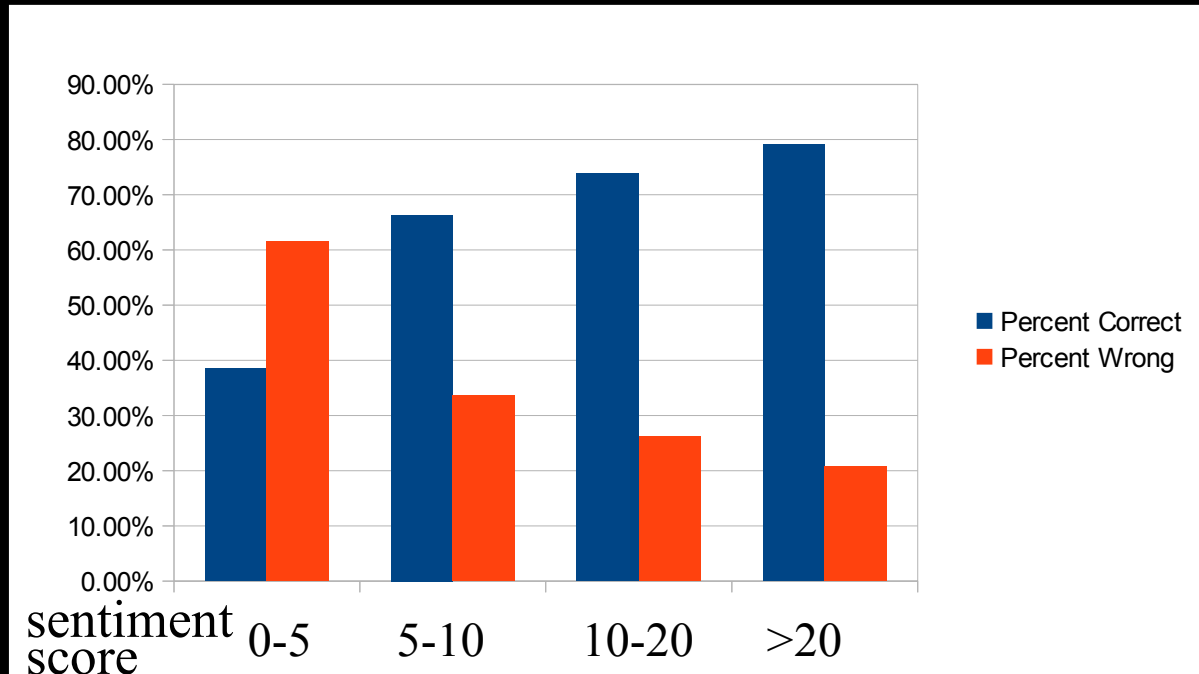
REVIEWS: bed

Bob771: This is an older hotel; however, the **bed** was very comfortable.
Sarah2: It was no 5 star hotel, but the **bed** was comfortable, it was ...
JoeSmith: ... made to order breakfast and the **bed** was very comfortable.

Testing

- Tested algorithm against a standard test corpus for sentiment analysis
 - Movie reviews
- ~10,000 sentences with manually-created sentiment score

Language is tough



Algorithm works well on strongly positive or negative reviews.
Not so well on mild or ambiguous ones.

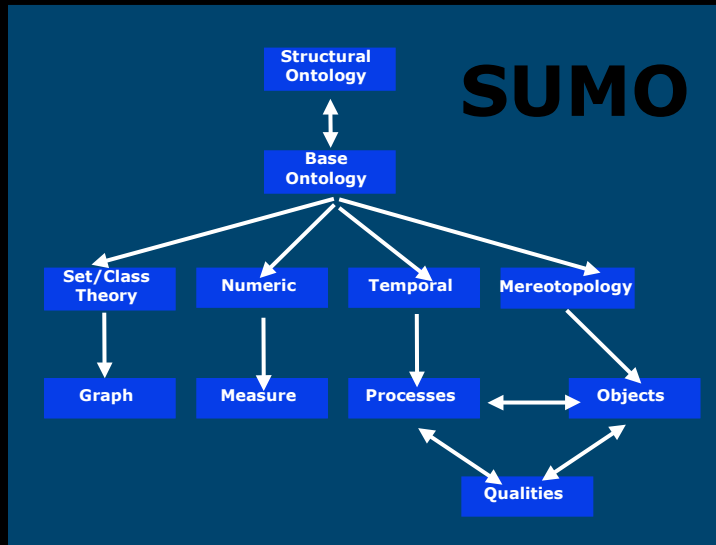
Tough for humans too

- “A disturbing and frighteningly evocative assembly of imagery and hypnotic music composed by Philip Glass.”
- Positive or negative?
- ...human raters typically agree about 70% of the time. Thus, a 70% accurate program is doing as well as humans... [Wikipedia]

Suggested Upper Merged Ontology

- Initial versions: 1000 terms, 4000 axioms, 750 rules
- Mapped by hand to all of WordNet 1.6
 - then ported to 3.0 and continually updated
- Associated domain ontologies totalling 20,000 terms and 80,000 axioms
 - Now linked with factbases including YAGO for millions of facts
 - New ontologies of Hotels and Dining
 - If-then rules, not just a taxonomy or semantic web structure
- Free
 - SUMO is owned by IEEE but basically public domain
 - Domain ontologies are released under GNU
 - www.ontologyportal.org

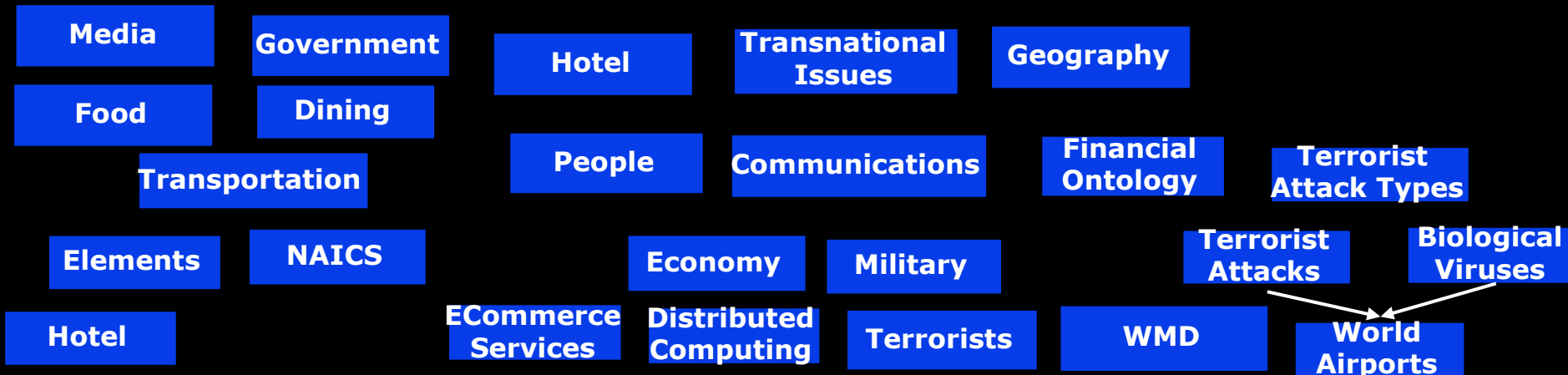
SUMO+Domain Ontology



Total Terms **Total Axioms** **Total Rules**
20977 **88257** **4730**

Relations: 1280

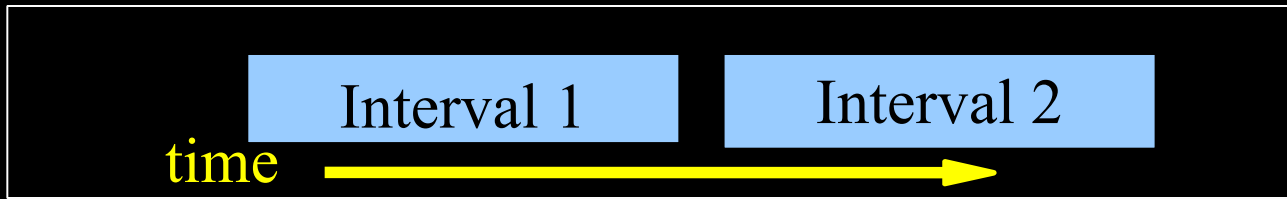
Mid-Level



WordNet

- A dictionary for computational linguistics applications
- 100,000 word senses, hand-created
- Mapped by hand to SUMO
- Open source
- Semantic links
 - Aid in computation
 - Verification of meaning during construction

Formal Ontology



(\Leftrightarrow)

```
(earlier ?INTERVAL1 ?INTERVAL2)
(before
  (EndFn ?INTERVAL1)
  (BeginFn ?INTERVAL2))
```

- WordNet has synsets for “earlier” etc
- But nothing in WordNet would allow a computer to assert that the end of one event precedes the start of another if one event is earlier than the other
- This is not a criticism of WordNet

Internationalization

- Translation of SUMO paraphrases to diverse multiple languages
 - Some confidence there's no cultural or linguistic bias
 - Chinese, Hindi, Tagalog, Czech, German, Italian, Korean, Romanian, Arabic
- SUMO is linked to multiple very large lexicons (Euro WordNet, Balkanet, HowNet etc)
 - English, Chinese, Italian, Arabic
- Ability to report review concepts in many languages, even though review is in English

Word Sense Disambiguation

- Brown corpus of English
 - One million words
 - “balanced” corpus from many sources – newspapers, novels etc.
- WordNet SemCor
 - Marked up Brown corpus by hand with parts of speech and word senses

WordNet SemCor

- “Bed” sense 1
 - air_mattress curtain sleep sleeping_bag slipper
 - SUMOTerm: Bed
- “Bed” sense 2
 - compost decayed manure pansy spade spread_over yard
 - SUMOTerm: CultivatedLandArea
- “Bed” sense 3
 - dry face homely river tilt
 - SUMOTerm: GeographicArea

Sentiment words

<u>type</u>	<u>word</u>	<u>POS</u>	<u>polarity</u>
weak	abandon	verb	negative
weak	abate	verb	negative
weak	abdicate	verb	negative
strong	aberration	adj	negative
weak	able	adj	positive

Created at U. Pittsburgh by having a team of students score words in WordNet, (and then validated by inter-subject agreement measures).

Sentiment Finding

"The guest rooms, ...
are somewhat
shopworn or
downright
shabby ..."

shabby -5
shopworn -1
...

Word
Sentiment
List

Stop
Word
List

Reviews

Scorer

Sentiment
Score

Concept
Sentiment

SF Drake: Room (-6)

Associate word
senses with
frequency of co-
occurring words

Disambig.
Corpus

Word Sense
Disambig.

Word
Senses

SUMO-
WordNet
Mapping

SUMO
Concepts

room (in a building) – hall 12, guest 10...
room (space) – grow 22,

room (in a building) →
Room