

# The Puzzle of Semantic Technologies

Kathleen Dahlgren, Ph.D.  
Cognition Technologies, Inc.

**COGNITION**  
GIVING TECHNOLOGIES NEW MEANING™

# Approaches to Semantic Web ...

Owl and tagging

Top –down

Fixed set of categories

Like what a librarian does

Risk of missing relevant retrievals

Taxonomies and pull-downs

Top down

User browses taxonomy

Risk reduced with user assistance

**COGNITION**  
GIVING TECHNOLOGIES NEW MEANING™

# ...Approaches to Semantic Web

- NLP

Bottom up

Word by word

Structure by structure

Completely free and covering

– any content is indexed

More like what a person does

while reading the text --

forming a cognitive model of text meaning

# Google and senses

- it struck me that bats should fly at night when fans are quiet
- [Baseball - Bats Blog - NYTimes.com](#)
- Apr 12, 2009 ... "The outcry of Phillie *fans* had M.L.B. change it," Kalas told *me* last October. ... *Struck* him out! The Philadelphia Phillies are 2008 world champions of baseball!" ... on Sunday and had to *fly* across the country for this series. .... The Yankees planned to give Johnny Damon the *night* off Saturday, ... [bats.blogs.nytimes.com/](#) - 125k - [Cached](#) - [Similar pages](#)
- [New York Yankees' bats go quiet in 4-0 loss to Twins - NJ.com](#)
- MINNEAPOLIS -- Perhaps before Tuesday *night's* game, the Yankees can all go over to the ... Twins lefty Glen Perkins gave up four hits, walked three and *struck* out four in his ... After Richie Sexson walked, Ivan Rodriguez hit a *fly* ball to shallow center .... Bronx Bombers *fans*, post your photos. Yankees *fan* Photos ... [www.nj.com/yankees/index.ssf/2008/08/new\\_york\\_yankees\\_bats\\_go\\_quiet.html](#) - 51k - [Cached](#) - [Similar pages](#)
- [Heckledepot - Baseball Fans Submitted Heckling Stories 6](#)
- My very *quiet* and conservative sister got us tickets right behind home plate at .... Then, I said it for the rest of his at *bats* that *night*, and sadly, .... But the one that I heard recently that *struck me* funny was...when a ... I'd like to tell that guy at *bat* he *should* have swung but how do I pronounce his name? ... [www.heckledepot.com/submitted5.html](#) - 37k - [Cached](#) - [Similar pages](#)
- [C70 At The Bat: A St. Louis Cardinals Blog.](#)
- In fact, it made *me* wonder about how Cardinal *fans* would handle a .... It seemed to be effective, as he *struck* out two in his inning and I believe ... It's always a good *night* when you have to make up your mind who *should* get the Hero tag. .... piled on runs in the first inning, only to go *quiet* and lose the game. ... [www.cardinal70.com/](#) - 112k - [Cached](#) - [Similar pages](#)
- [Marlins rally to top Cubs, 'quiet those fans' - Baseball ...](#)
- Jul 25, 2008 ... Disable *Fly-out* What are flyouts? ... Fight *night* in the minor league ... Jeremy Hermida managed to *quiet* them with one swing of the *bat*. ... give confidence to the coaches that they can throw *me* out there in any situation. ... when he *struck* out pinch-hitter Alfredo Amezaga to lead off the seventh. ... [nbcsports.msnbc.com/id/25853732/](#) - 62k - [Cached](#) - [Similar pages](#)

# Google and senses

- **Workers struck each other at a union meeting**
- [Fire down below|4Dec99|Socialist Worker](#)
- OVER 1000 *workers struck* against Tory controlled Wandsworth council as anger over years ... They and Wandsworth *workers* plan to join *each other's* protests next week. ... Loads of us didn't even know that a *union meeting* had been called. ...  
[www.socialistworker.co.uk/art.php?id=1773](http://www.socialistworker.co.uk/art.php?id=1773) - 22k - [Cached](#) - [Similar pages](#)
- [Race riot: Chicago in the Red Summer of 1919 - Google Books Result](#)
- by William M. Tuttle - 1996 - Social Science - 305 pages  
peted for jobs; no *other* potentially divisive force more ominously ... and when the packers announced a wage reduction, 23000 packing house *workers struck*. ...  
[books.google.com/books?isbn=0252065867...](http://books.google.com/books?isbn=0252065867...)
- [The Militant - a socialist newspaper -- front page - Minnesota ...](#)
- *Workers* at Dakota Premium first *struck* the company for seven hours June 1 to protest ... firings, provocations, and *other* attacks, *workers* convinced *each other* of the need to ... The mood at the *union meeting* a few hours later was serious, ... The company claims the *union* offered \$10000 to *each worker* and that it ...  
[www.themilitant.com/2000/6433/643302.shtml](http://www.themilitant.com/2000/6433/643302.shtml) - 16k - [Cached](#) - [Similar pages](#)
- [Substance News - April Chicago Teachers Union meeting...](#)
- ... those representing *other workers* at the Chicago Public Schools]. ... *Other* information gleaned at the pre-*meeting* question period ... Koffman's answer *struck* many observers as another example of how out of touch the *union's* leaders are ... people who were sent *each other's* addresses and social security numbers. ...  
[www.substancenews.com/content/view/437/88/](http://www.substancenews.com/content/view/437/88/) - 36k - [Cached](#) - [Similar pages](#)

# What's different

Linguistic approach

4 sources of semantics

Broad coverage of English

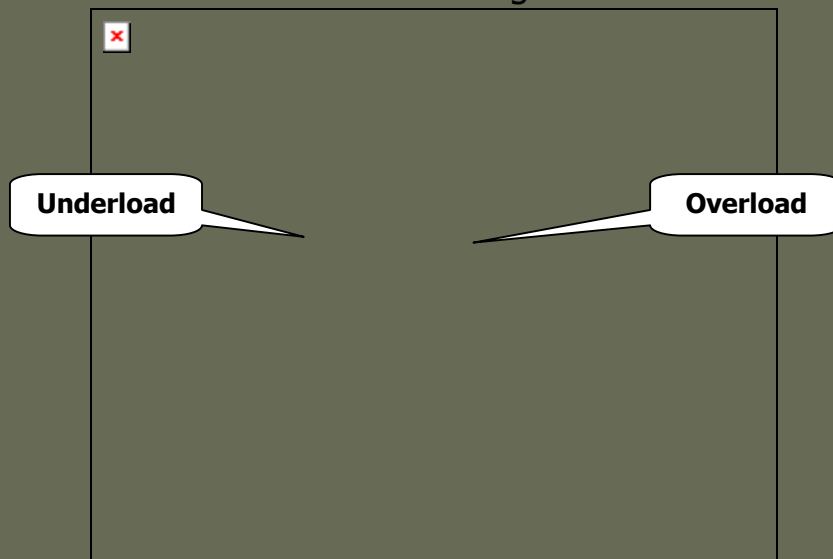
- high performance out of the box

# CognitionSearch

## High Precision and High Recall

### Traditional Search Engines

Miss the Target



#### Overload

- many false positives – ambiguity
- lack understanding that words have multiple meanings

#### Underload

- miss relevant information – synonymy
- lack understanding of multiple words with similar meanings

**CognitionSearch**  
Data ▶ Information ▶ Knowledge ▶ Understanding

Better Precision, Better Recall



#### Increased precision

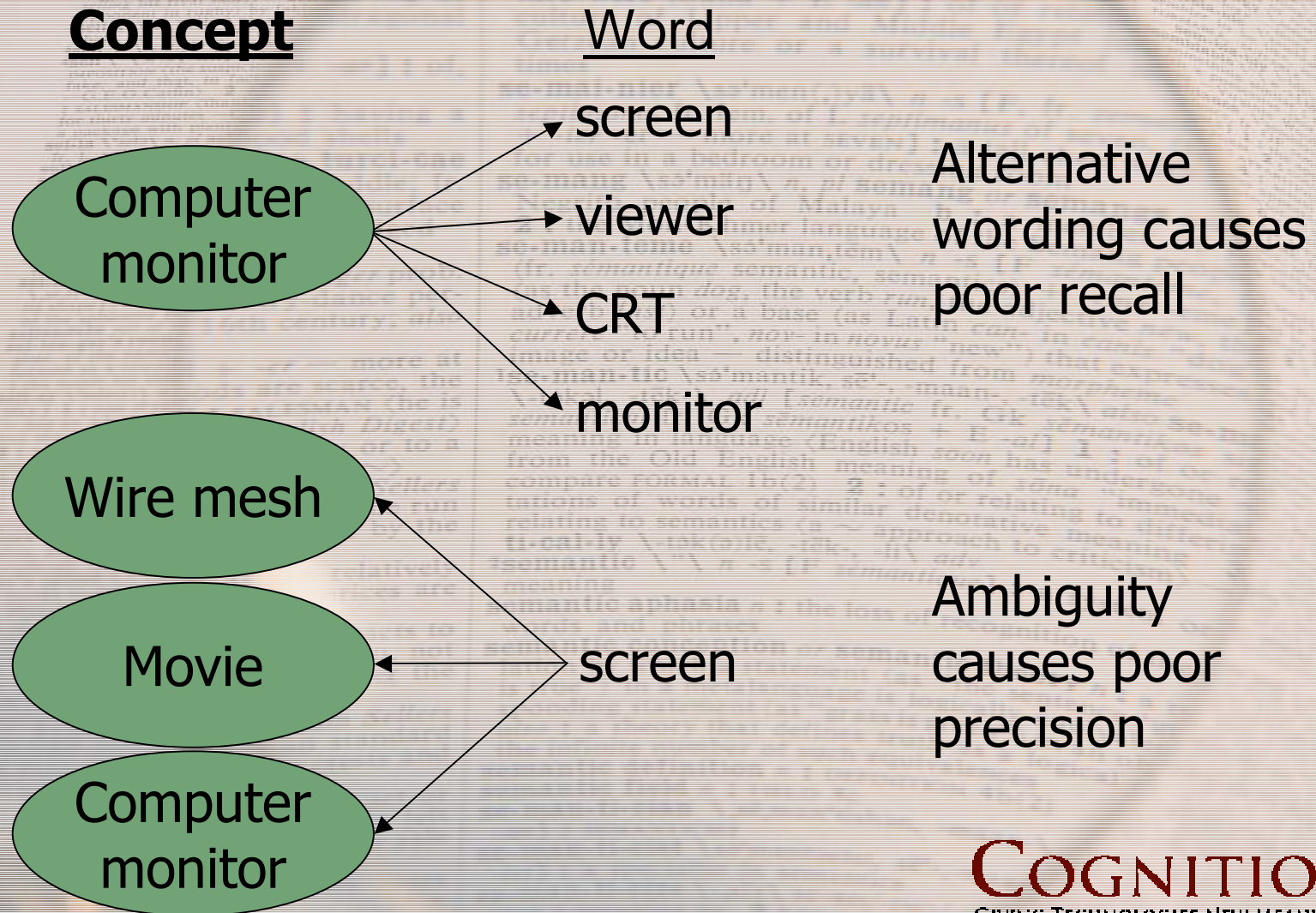
- retrieves fewer irrelevant documents – manages ambiguity
- understands meaning of words and phrases

#### Increased recall

- more relevant documents retrieved
- understands synonymy

**COGNITION**  
GIVING TECHNOLOGIES NEW MEANING™

# Opacity of Language



# Formal Semantics

# Word Senses

Word string

Senses

lemon

lemon1 – a citrus fruit

lemon2 – a poorly manufactured item

yellow

yellow1 – a color property

yellow2 – a behavioral property

World

Word Sense

lemon

lemon1

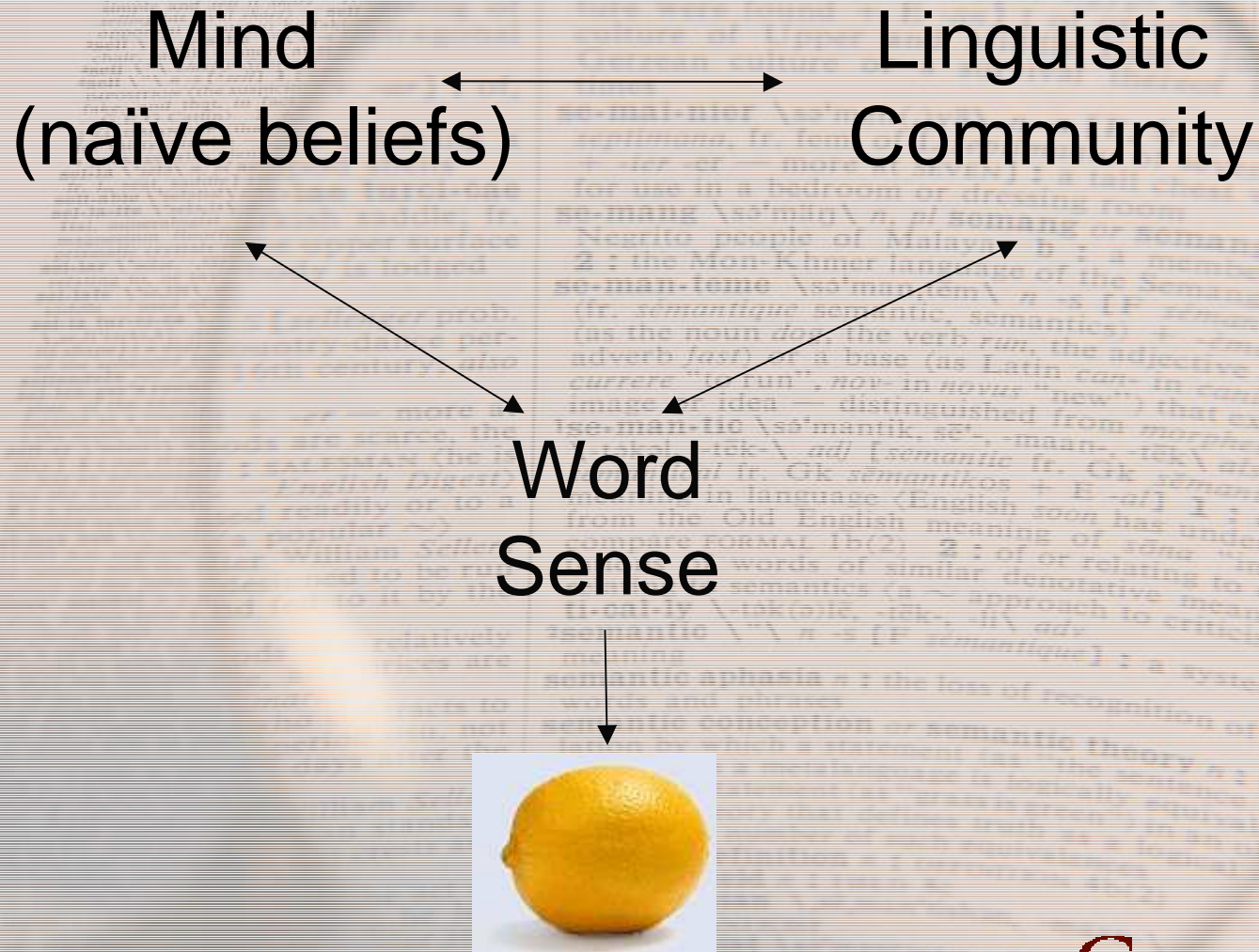
# Concepts Don't Guarantee Reference

Typicality – Lemons are yellow.

Personal ignorance – Pythons are dangerous.

Social ignorance – The sun shines on the Earth.

# Meaning Contract



# Linguistic Semantics

Word	Word Sense	Ontology	Naïve Semantics
lemon	lemon1	citrus fruit	yellow, tart taste
lemon	lemon2	mechanical device	has many defects, in the repair shop a lot

# Shallowness

## Shallow Naïve Semantics

lemon1 – ISA citrus fruit – typically yellow, typically tart taste

## Deeper Knowledge

lemons grow on low trees with shiny deep green leaves

lemon juice flavors lemonade

lemon pie sometimes has meringue, sometimes has whipped cream

# Linguistic Semantics of *base*

Word sense	Ontology -mother	Ontology-daughters	Synonyms	Meaning contexts
base3	physical part		bottom1 Foot2	mountain1 statue1 cliff1, support3
base4	military complex	Fort-Dix, Fort-Bragg ...	post2 station2 Installation1	tank5 soldier1 headquarters
base5	chemical	purine adenine guanine	alkali1 acidity1 basicity1	titration1 ion1 oxidation1
base6	baseball area	first2 second2 home5	bag8	runner2, teammate1 baseball1

# Baroqueness

Fixed stock of semantic features

Fixed stock of semantic tags

Baroque naïve semantics –

behavior, flies

taste, tart

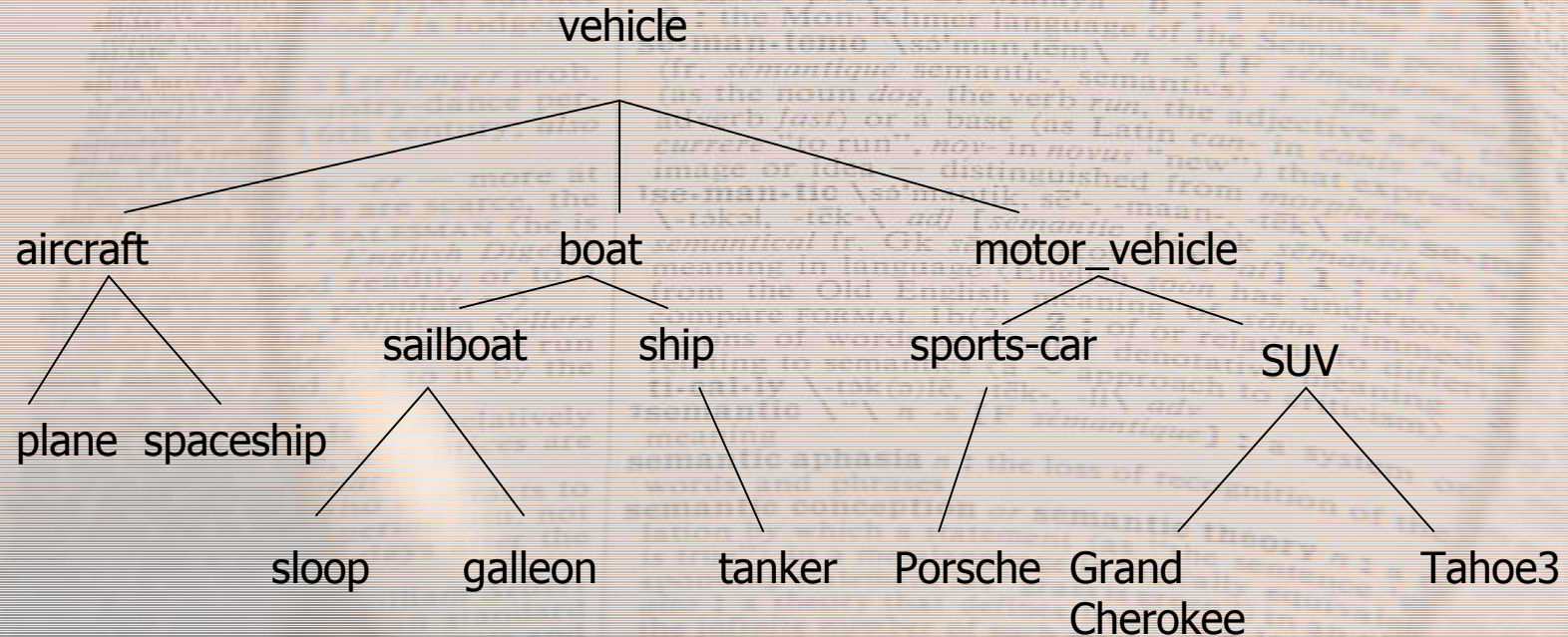
haspart, carburetor

# Linguistic processing

- Phrases
  - Bill of Rights
  - Securities and Exchange Commission – SEC
  - Mr. Xang Xi
  - December 13, 1942
- Morphology
  - Inflectional
    - bat – bats
    - catch – caught
    - baby – babies
  - Derivational
    - run – rerun
    - operate – inter-operate
  - Stems
    - wound/wind
    - Jaguar/jaguar
    - Bill/billing/Billings/bill/billed

# Ontology

## Vehicle Taxonomy – simplified



# Meaning Thesaurus

catch1 – bust3 – arrest1-snare3 – collar5

catch4 – come19 – contract5

catch7 – fascinate – grab3 – rivet2

catch9 – listen1 – overhear1

catch12 – lock1 – clasp2

SEC – Securities and Exchange Commission

DKA – diabetic ketoacidosis

# Semantic Contexts

catch a thief – catch1

catch a cold – catch4

catch one's fancy – catch7

catch what someone was saying – catch9

catch on a door – catch12

# Wide Coverage Semantic Databases

Lexical Database	User Benefits
<p align="center"><b>506,000 Words Stems</b></p>	<p align="center">Precision and Recall</p>
<p align="center"><b>536,000 Word Senses with Naïve semantics</b></p>	<p align="center">Precision</p>
<p align="center"><b>17,000 Ambiguous Word Entries</b></p>	<p align="center">Precision</p>
<p align="center"><b>191,000 Phrases</b></p>	<p align="center">Precision</p>

# Semantic databases

Semantic database	User Benefits
<b>7,500 nodes and 536,000 terms Ontology</b>	Recall
<b>8,300,000 Semantic Contexts</b>	Precision
<b>75,000 Thesaural Concept Groups (word senses &amp; phrases)</b>	Recall

# Combination of Rules and Statistics

- Teach the computer a shallow broad layer of semantics
- Bootstrap that knowledge with statistics

# Scalability

- Very wide coverage semantic databases
- Semi-automated machine learning
- Compilation of relationships into the index
- Extreme data compression
- Fully distributed
- No limit to document base size
- Incremental indexing