Lexical Semantics

Natural Language Processing: Jordan Boyd-Graber
University of Colorado Boulder

20. OCTOBER 2014

Adapted from material by Christiane Fellbaum

Roadmap

- The lexicon as a component of human language
- Theoretical models of the lexicon
- Computational models of the lexicon (WordNet)
- The lexicon in Natural Language Processing: Similarity and word sense disambiguation

The Lexicon

Definition

collection of all words of a language

- the component of grammar that includes speakers' knowledge of words
- Very large ($>$ 40000 words, dictionaries)
- cannot enumerate its entries
- never static
- words come and go
- words change their meaning, usage, pronunciation

Lexical Variation

- Acquisition, loss are continuous, lifelong (but other components of language are acquired by age five)
- No two speakers have exactly the same lexicon
- Terminology, regional differences, slang, ...

Components of Lexicon

- Sound (pronunciation)
- Morphology (e.g., plural formation: woman→women, house→houses)
- For many languages/speakers: written representation (spelling, mapping to graphemes)
- Syntax: subcategorization/selectional restrictions

Selectional Restriction Example

Direct objects of word "eat" can only be things that are considered food.

Selectional Preference Example

"strong tea", not "powerful tea"

Lexicon's Representation of Meaning

- Most important: meaning/concept behind the word
- Assumption: given the large number of words and concepts, there must be an organizing system

Lexicon's Representation of Meaning

- Most important: meaning/concept behind the word
- Assumption: given the large number of words and concepts, there must be an organizing system
- Categorization!
- Some languages overtly mark categories with class membership morphemes (e.g. counting markers in Chinese / Japanese)

The Lexicon's Apparent Disconnect

- Saussure's "Arbitrariness of the sign": Form has no inherent, motivated relation to the meaning
- (exceptions e.g., onomatopoeia: glisten, glimmer, gleam, glow)
- No two speakers associate precisely the same meaning(s) with a word form, but there is a core meaning that is agreed upon by convention within the speaker community
- (e.g., your concept of house may not be the same as mine but we are likely to call the same instances houses)

What enters the lexicon?

- All humans perceive same color spectrum (same visual processing)
  - But languages don't "cut up"/ label colors in the same way e.g., Russian has two words for blue
  - Interior designers have hundreds of words for red, blue, etc.
- Pirahã number words
- Lexicon is only an imperfect reflection of our concept inventory

Word Meanings

- How can a child (or any learner) figure out the meanings behind words?
- Not from dictionary definitions—they are made from words that are themselves "entries" (a closed dictionary is fiendishly hard to create)

Word Meanings

- How can a child (or any learner) figure out the meanings behind words?
- Not from dictionary definitions—they are made from words that are themselves "entries" (a closed dictionary is fiendishly hard to create)
- Infinite regress!
- Do speakers have a "categorizer" that detects features and assigns them to a category?
- Bootstrap from known concepts

Concepts Connecting Words to Meaning

Synonymy

Different words with the same meaning

Polysemy/Homonymy

Same word with multiple meanings.

- **Polysemy**: Slightly different meanings
- **Homonomy**: Completely different meanings

Synonymy

- One meaning/concept is expressed by several different word forms:
- `beat, hit, strike, shut, close`
- `car, motorcar, auto, automobile`
- `big, large, difficult, hard`

Polysemy / Homonomy

Homonomy

- Homonymy: multiple unrelated meanings (relation inaccessible to everyday speakers)
  - `pitch` (acoustic property/tar)
  - `bar` (saloon/stick)
  - `bass` (fish, musical instrument)
  - `bank` (money institution/elevated land)

Polysemy

- One word form expresses multiple meanings
- table: `tabular_array` `piece_of_furniture` `table_mountain, mesa` `postpone`
- bass: `(vocal range, singer, score)`
- newspaper: `(paper copy/institution/building)`
- Borderline between polysemy, homonymy can be fuzzy

Polysemy test

- Zeugma: conjoining of different senses of a word is odd
  - He left Rome and the bills on the table (bad)
  - He left Rome and later the country (fine)
  - He left the bills and an apple on the table (fine)

Polysemy test

- Zeugma: conjoining of different senses of a word is odd
  - He left Rome and the bills on the table (bad)
  - He left Rome and later the country (fine)
  - He left the bills and an apple on the table (fine)

Shades of Polysemy: Metonymy

- Regular, systematic, productive polysemy
- Metonymy: conflating a part and the whole
  - The “White House” issued a statement
  - The “office” isnt answering the phone

Shades of Polysemy: Metaphor

- Conventionalized, systematic, productive
- "Time is Money": {Save/spend/invest/waste/lose/gain/buy } time (Lakoff and Johnson 1980)
- Unconventional, ad hoc (though readily interpretable):
    - My surgeon is a butcher
    - This restaurant is a zoo on weekends

Synonymy

- Difficult for learners, computer-generated text
- Must choose context-appropriate synonym (machine translation examples next week)
- Real example: Diner (non-native speaker) asks waiter for
  - vacant vessel
  - Wants empty bowl
- Real synonymy is rare (or non-existent?)
- Few (no?) two or more words are substitutable for one another in all contexts
- Selectional preference: “strong” vs. “powerful” tea, “powerful” “strong” computer

Computational Models of Word Meaning

- PMI: Measuring word association (e.g., selectional preference)
- Context Similarity
- Computational Lexicons (WordNet)

Computational Measures of Association

- Pointwise Mutual (Church and Hanks, Fano)
- Measures how strongly two words are with each other in a text
- Compares joint probability of two words probability of observing each word independently

$$PMI_{x,y} = \log \frac{p(x,y)}{p(x)p(y)} \tag{1}$$

  where $p(x)$ is the probability of a word appearing in a sentence (or n-window context).

- Interpretation: the number of bits of information obtained about probability of x given you've observed y (or vice-versa)

PMI examples

- If *PMI* close to 0, then x and y are independent
- If *PMI* is positive, theen seeing one tells you that you're more likely to see the other

Examples

Bette	Midler	18.38
videocassette	recorder	15.94
unsalted	butter	15.19
⋮	⋮	⋮
time	last	0.29

Comparing Word Meanings

- "You shall know a word by the company it keeps" (Firth 1957)
- Build a word vector by counting the words it shares contexts with

$$w_x = < c_{x,w_1}, c_{x,w_2}, c_{x,w_3}, c_{x,w_4}, \dots c_{x,w_n} > \tag{2}$$

  where c_{x,w_i} is the number of times word x appears next to w_i.
- The distance (e.g., cosine) between w_x and w_y corresponds to how different their meanings are

Word Sense Disambiguation

- Given a polysemous word in context, which meaning is correct?
- Necessary if we want a computerized, unambiguous representation of meaning.
- Measures of performance
  - Recall: $\frac{\text{true positives}}{\text{true positives}+\text{false negatives}}$
  - Precision: $\frac{\text{true positives}}{\text{true positives}+\text{false positives}}$
- Current state of the art: 70% precision and recall

Word Sense Disambiguation ("paper" example)

- Typical, say the critics, of a paper that pretends to lead, but actually follows.
- Make sure your house is thoroughly draughtPproofed by testing windows and doors by closing them on a piece of paper.
- Within a few years, the last of the Forest's paper-mills ceased operations.
- Manuscripts of the Tales, lots of undated lyrics on random little slips of paper, and of course all the revisions of Melusina, which she rewrote at least eight times, always changing it.
- Opinion, the paper reports, is moving in favour of the idea that in future self-management communities of interest should be regulated by general laws instead of by social compacts, which have practically no authority, legal or otherwise.
- There are clear signs that senior officials, such as Burke Trend, William Armstrong and Ian Bancroft, involved in drafting the 1970 White Paper went along with the Heath analysis.
- NO SOONER had we settled in our seats when the burly Arthur Lewis (West Ham) blew in, waving an order paper, shouting Point of Order, Mr Speaker!
- Lessingham made it a habit to leave his agenda behind, a small personal protest against the amount of paper which the formal weekly meeting generated.

Word Sense Disambiguation ("paper" example)

- Typical, say the critics, of a paper that pretends to lead, but actually follows.
- Make sure your house is thoroughly draughtPproofed by testing windows and doors by closing them on a piece of paper.
- Within a few years, the last of the Forest's paper-mills ceased operations.
- Manuscripts of the Tales, lots of undated lyrics on random little slips of paper, and of course all the revisions of Melusina, which she rewrote at least eight times, always changing it.
- Opinion, the paper reports, is moving in favour of the idea that in future self-management communities of interest should be regulated by general laws instead of by social compacts, which have practically no authority, legal or otherwise.
- There are clear signs that senior officials, such as Burke Trend, William Armstrong and Ian Bancroft, involved in drafting the 1970 White Paper went along with the Heath analysis.
- NO SOONER had we settled in our seats when the burly Arthur Lewis (West Ham) blew in, waving an order paper, shouting Point of Order, Mr Speaker!
- Lessingham made it a habit to leave his agenda behind, a small personal protest against the amount of paper which the formal weekly meeting generated.

What's a meaning?

- To answer this problem, we need an inventory of meanings
- Most common solution: WORDNET

What's a meaning?

- To answer this problem, we need an inventory of meanings
- Most common solution: WordNet

Outline

WordNet

WN Similarity

WSD Task

History of WordNet

- Electronic dictionary of words and meaning
- George Miller and Christiane Fellbaum (Princeton)
- Interesting organizational structure
- Included as part of `nltk`
- http://wordnet.princeton.edu
- Includes most English nouns, verbs, adjectives, adverbs
- Inspired WordNets in many other languages (70)

Design Principles of WordNet

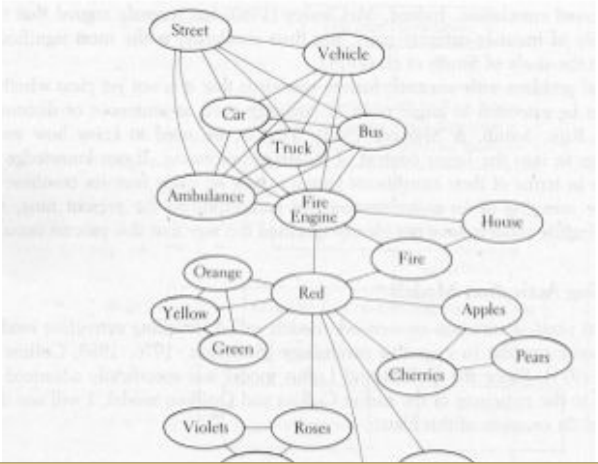
- Psychologically real: consistent with experimental data
- Universally applicable (mostly)
- Testable
- Elegant/simple
- Amenable to computational analysis, applications

Organization

- Existing (western) dictionaries: organize by sound
- Syntax-based organization
  - Levin (1993): syntactic properties of English
  - VerbNet (Kipper and Palmer)
- The WordNet model
  - Semantics-based
  - Words are interconnected by meaning relations

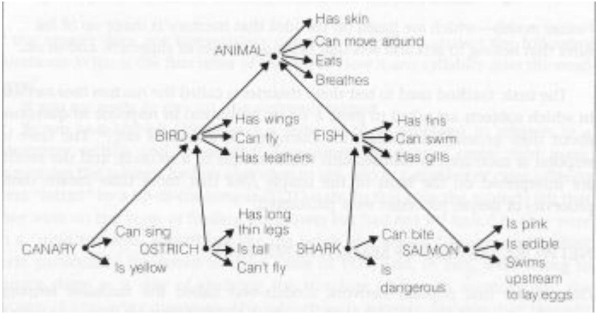
Semantic Network (Psychological Foundation)

Semantic network representation (Collins and Quillian 1969, 1970, 1972)



Semantic Network (Psychological Foundation)

Semantic network representation (Collins and Quillian 1969, 1970, 1972)



Claim: we know that "canaries fly" because "birds fly" and "canaries are a kind of bird"

Semantic Network

- Collins & Quillian (1969) measured reaction times to statements involving knowledge distributed across different "levels"
- Responses to statements like
  - Do birds move?
  - Do canaries move?
  - Do canaries have feathers?
  - Are canaries yellow?
- Reaction times varied depending on how many nodes had to be traversed to access the information

WordNet Inspiration

- What would such a network look like exactly and on a large scale?
- Can most/all of the lexicon (of any language?) be represented as a semantic network?
- Would some words be unconnected?
- Later: crosslingual perspective

What are the connections in WordNet?

- If the (English) lexicon can be represented as a semantic network (a graph), what are the links that connect the nodes?
- WN distinguishes two kinds of links
- Links among nodes (concepts) are conceptually semantic (e.g., bird-feather)
  - Hyponomy
  - Meronomy
  - Synonomy
- Links among specific word forms are lexical (e.g., feather-feathery)
- Lexical links subsume conceptual-semantic links

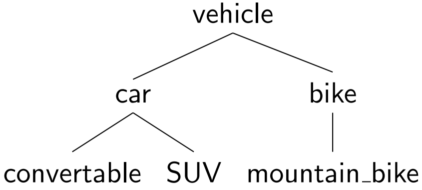
WordNet Stats

- The "Net" part of WordNet
- Synsets are interconnected Bi-directional arcs express semantic relations
- Result: large semantic network (directed acyclic graph/DAG)

Part of speech	Word forms	Synsets
noun	117,798	82,115
verb	11,529	13,767
adjective	21,479	18,156
adverb	4,481	3,621
total	155,287	117,659

Hypo/hypernym

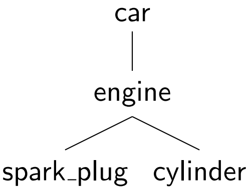
- Hypo-/hypernymy relates noun synsets
- Relates more/less general concepts
- Creates hierarchies, or “trees”



- “A car is is a kind of vehicle”
- Noun hierarchies have up to 16 levels

Meronymy / holonymy

- Meronymy/holonymy: part-whole relation



- "An engine has spark plugs"
- "Spark plus and cylinders are parts of an engine"
- Inheritance:
  - A finger is part of a hand
  - A hand is part of an arm
  - An arm is part of a body
  - A finger is part of a body
  - ("a fingernail is a part of an arm" seem odd though)

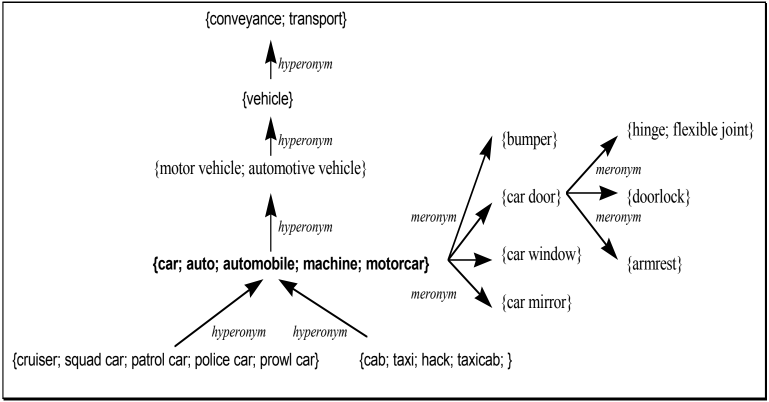
Antonymy

- Adjective relations: antonymy
- Strong mutual association between members of antonymous adjective pairs:
- hot-cold, old-new, high-low, big-small
- Distributional overlap (shared selectional restrictions)
- Highly frequent, polysemous
- Statistically high co-occurrence in the same sentence (Justeson and Katz 1991)
- Members of antonymous pairs are acquired together by children, learners

Verb Connections

- Relations among verbs
- Manner relation (troponymy): to x is to y in some manner (move-walk, whisper-talk, smack-hit, gobble-eat)
- Can construct trees (not as deep as nouns): move-run-jog-run, communicate-talk-whisper
- Other relations among verbs reflect temporal or logical order between two events
    - divorce-marry (backward presupposition)
    - snore-sleep, pay-buy (inclusion)
    - kill-die (cause)
- One event unidirectionally entails the other
- Entailment also holds among troponyms

WN Structure



Constructing new wordnets

- Princeton WordNet (English) was manually built
- Many new wordnets don't have resources for fully manual construction
- Use bootstrapping methods for building new and augmenting existing WordNet

Outline

One Application: Word Similarity

- WordNet allows one to measure the semantic similarity or distance among words
    - Counting edges
    - Weighting edges based on corpus statistics
    - Implemented in `nltk`

WN Similarity: Leacock and Chodorow

- Things further away in WordNet are less similar

$$sim_{\mathrm{LC}}(s_1, s_2) = -\log \text{path-len}(s_1, s_2) \tag{3}$$

- Problem: not all edge lengths are equal
  - Scottish Terrier—Terrier
  - Mammal—Vertebrate

WN Similarity: Leacock and Chodorow

- Things further away in WordNet are less similar

$$sim_{\text{LC}}(s_1, s_2) = -\log \text{path-len}(s_1, s_2) \tag{3}$$

- Problem: not all edge lengths are equal
  - Scottish Terrier—Terrier
  - Mammal—Vertebrate
- Solution: Use corpus statistics to find how specific concepts

WN Similarity: Resnik

- Probability of a concept: need all the words contained in the concept
- E.g.: w(mammal) $\equiv$ { hampster, monkey, horse, ... }

$$P(c) = \frac{\sum_{w \in w(c)} \text{count}(w)}{N} \tag{4}$$

WN Similarity: Resnik

- Probability of a concept: need all the words contained in the concept
- E.g.: w(mammal) $\equiv$ { hampster, monkey, horse, ... }

$$P(c) = \frac{\sum_{w \in w(c)} \text{count}(w)}{N} \tag{4}$$

- Given a corpus with N words (double count words with multiple meanings)
- Define the information content of a concept as $\text{IC} = -\log P(c)$
  - "entity" concept has 0.0 information content
  - "dog" has much higher information content

WN Similarity: Resnik

- Probability of a concept: need all the words contained in the concept
- E.g.: w(mammal) $\equiv$ { hampster, monkey, horse, ... }

$$P(c) = \frac{\sum_{w \in w(c)} \text{count}(w)}{N} \tag{4}$$

- Given a corpus with N words (double count words with multiple meanings)
- Define the information content of a concept as $\text{IC} = -\log P(c)$
  - "entity" concept has 0.0 information content
  - "dog" has much higher information content

WN Similarity: Resnik

- Least-Common Subsumer: Lowest (most-specific) node in WordNet containing a common ancestor
  - lcs(cat, dog) = mammal
  - lcs(wolf, puppy) = canine
- Resnik similarity

$$sim_{\text{Res}}(s_1, s_2) = \text{IC}\left[\text{lcs}(s_1, s_2)\right] \tag{5}$$

Outline

WordNet

WN Similarity

WSD Task

Word Sense Disambiguation

I	saw	a	98	year	old	man	who	still	walks	and	tells	jokes
	25			4	8	11		4	10		8	4

- Now that we have an inventory (WordNet), we can apply it to text
- Tens of millions of combinations! (Using WordNet)

WSD Methods

- Unsupervised: Maximize similarity within a sentence
- Supervised: Tag a bunch of sentences and train a classifier (e.g. MaxEnt)

WSD Methods

- Unsupervised: Maximize similarity within a sentence
  - Requires heuristics: e.g. one sense per discourse
  - Inspired machine learning methods (Yarowsky Algorithm)
- Supervised: Tag a bunch of sentences and train a classifier (e.g. MaxEnt)

Supervised WSD

- Difficulty: Getting training data
  - People don't agree (unlike parsing)
  - Requires skilled annotators
- Requires carefully engineered features
  - Part-of-speech in context
  - Topic of document
  - Words around context

Supervised WSD

- Difficulty: Getting training data
  - People don't agree (unlike parsing)
  - Requires skilled annotators
- Requires carefully engineered features
  - Part-of-speech in context
  - Topic of document
  - Words around context
- 70% accuracy

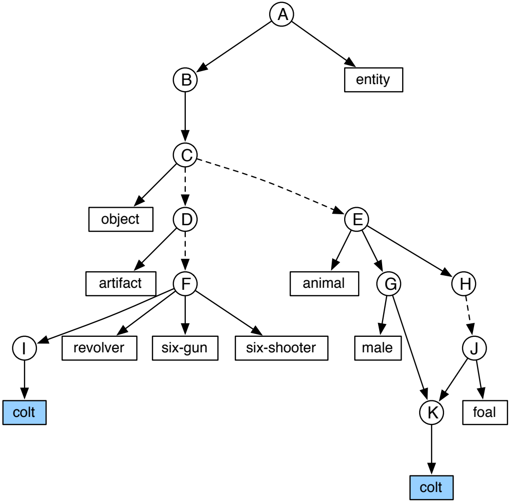
Wrapup

- Lexical Semantics
- Computational approaches to representing word meaning
- WordNet: a resource for representing word meaning
- WordNet similarity: more in class
- The task of word sense disambiguation

In Class

- Information Content
- Similarity

Concepts


A
B
entity
C
object
D
E
artifact
F
animal
G
H
I
revolver
six-gun
six-shooter
male
J
colt
K
foal
colt

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	
artifact	2000	D	
revolver	500	F	
six-gun	100		
six-shooter	100		
colt	100	I	
animal	1000	E	
male	800	G	
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	
artifact	2000	D	
revolver	500	F	
six-gun	100		
six-shooter	100		
colt	100	I	
animal	1000	E	
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	
artifact	2000	D	
revolver	500	F	
six-gun	100		
six-shooter	100		
colt	100	I	
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	
artifact	2000	D	
revolver	500	F	
six-gun	100		
six-shooter	100		
colt	100	I	$p(I) = \frac{100}{5000} = 0.02$
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	
artifact	2000	D	
revolver	500	F	$p(F) = \frac{800}{5000} = 0.16$
six-gun	100		
six-shooter	100		
colt	100	I	$p(I) = \frac{100}{5000} = 0.02$
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	
artifact	2000	D	$p(D) = \frac{2800}{5000} = 0.56$
revolver	500	F	$p(F) = \frac{800}{5000} = 0.16$
six-gun	100		
six-shooter	100		
colt	100	I	$p(I) = \frac{100}{5000} = 0.02$
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	
object	100	C	$p(C) = \frac{4800}{5000} = 0.96$
artifact	2000	D	$p(D) = \frac{2800}{5000} = 0.56$
revolver	500	F	$p(F) = \frac{800}{5000} = 0.16$
six-gun	100		
six-shooter	100		
colt	100	I	$p(I) = \frac{100}{5000} = 0.02$
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	
		B	$p(B) = \frac{4900}{5000} = 0.98$
object	100	C	$p(C) = \frac{4800}{5000} = 0.96$
artifact	2000	D	$p(D) = \frac{2800}{5000} = 0.56$
revolver	500	F	$p(F) = \frac{800}{5000} = 0.16$
six-gun	100		
six-shooter	100		
colt	100	I	$p(I) = \frac{100}{5000} = 0.02$
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Information Content

Word Counts		Probability	
entity	100	A	$p(A) = \frac{5000}{5000} = 1.0$
		B	$p(B) = \frac{4900}{5000} = 0.98$
object	100	C	$p(C) = \frac{4800}{5000} = 0.96$
artifact	2000	D	$p(D) = \frac{2800}{5000} = 0.56$
revolver	500	F	$p(F) = \frac{800}{5000} = 0.16$
six-gun	100		
six-shooter	100		
colt	100	I	$p(I) = \frac{100}{5000} = 0.02$
animal	1000	E	$p(E) = \frac{2000}{5000} = 0.4$
male	800	G	$p(G) = \frac{1000}{5000} = 0.2$
foal	100	J	$p(J) = \frac{200}{5000} = 0.04$
colt	100	K	$p(K) = \frac{100}{5000} = 0.02$

Similarity

1. K and J?
2. I and K?
3. E and D?
4. G and J?

Similarity

1. K and J?

$$IC[J] = -\log(0.04) = 3.2 \tag{6}$$

2. I and K?
3. E and D?
4. G and J?

Similarity

1. K and J?

$$IC[J] = -\log(0.04) = 3.2 \tag{6}$$

2. I and K?

$$IC[C] = -\log(0.96) = 0.04 \tag{7}$$

3. E and D?
4. G and J?

Similarity

1. K and J?

$$IC[J] = -\log(0.04) = 3.2 \tag{6}$$

2. I and K?

$$IC[C] = -\log(0.96) = 0.04 \tag{7}$$

3. E and D?

$$IC[C] = -\log(0.96) = 0.04 \tag{8}$$

4. G and J?

Similarity

1. K and J?

$$IC[J] = -\log(0.04) = 3.2 \tag{6}$$

2. I and K?

$$IC[C] = -\log(0.96) = 0.04 \tag{7}$$

3. E and D?

$$IC[C] = -\log(0.96) = 0.04 \tag{8}$$

4. G and J?

$$IC[E] = -\log(0.4) = 0.92 \tag{9}$$