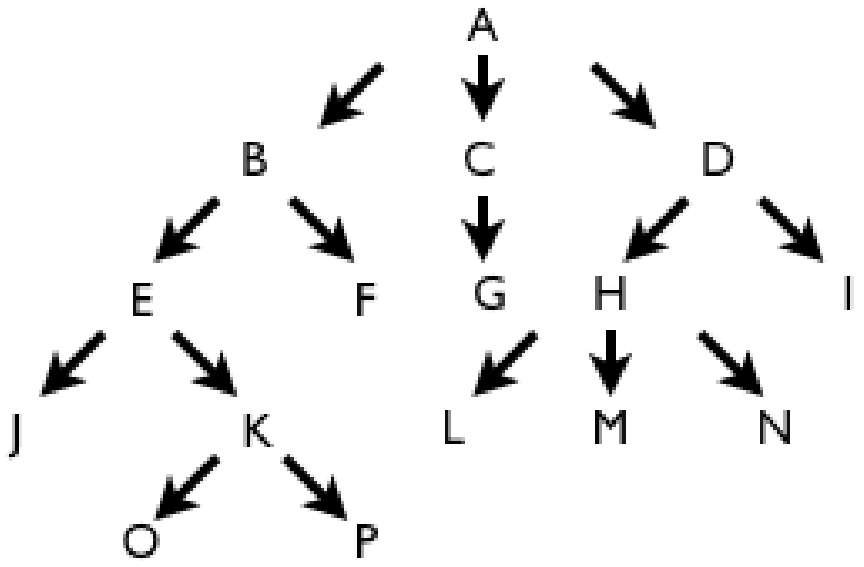
Feedback — Lexical Semantics and Question Answering

Help

You submitted this quiz on **Mon 30 Apr 2012 3:33 AM PDT**. You got a score of **5.00** out of **5.00**.

Question 1

Assume we have a corpus of 1000 words and the following WordNet Hierarchy:



Now assume we collect the following count data for each of the words:

J = 200

P = 200

F = 100

G = 50

L = 50

M = 100

N = 100

I = 200

Assuming that all other words do not appear in the corpus, what is $\mathrm{sim}_{\mathrm{resnick}}(F, E)$? Round your answer to two decimal places (e.g. 0.35, 0.60, 1.43) and use natural log in your calculation.

You entered:

0.69

Your Answer		Score	Explanation
0.69	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

We can solve this problem by first taking the lowest common subsumer of the two words. We can calculate the probability of that lowest common subsumer by summing up the counts of words in the subtree rooted at the lowest common subsumer (which in this case is simply summing over all the leaves) and dividing that number by 1000, the number of words in the corpus. We then take the negative log of that probability to obtain our answer.

Question 2

Consider the following list of words and their hypernyms, which are listed in order from the word itself to the root of the WordNet hierarchy:

shoe, plate, shield, protection, covering, artifact, whole, object, physical entity, entity
shoe, footwear, covering, artifact, whole, object, physical entity, entity

drive, propulsion, act, event, psychological feature, abstract entity, entity
drive, mechanism, device, instrumentality, artifact, whole, object, physical entity, entity

disc, sound recording, recording, memory device, device, instrumentality, artifact, whole, object, physical entity, entity
disc, round shape, shape, attribute, abstract entity, entity

chair, seat, furniture, furnishing, instrumentality, artifact, whole, object, physical entity, entity
chair, position, occupation, activity, act, event, psychological feature, abstract entity, entity

Calculate wordsim(shoe, drive) and round your answer to two decimal places.

You entered:

0.13

Your Answer		Score	Explanation
0.13	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

Because there are two senses for each word, we must calculate the pairwise distance between four different pairs of senses. We then take the shortest of those path lengths and use that to compute the wordsim by taking the reciprocal of that number. We can calculate the path length between two senses by examining where along the list of each sense's hypernyms the two senses intersect.

Question 3

Let's say we've calculated $ppmi(\text{Stanford}, \text{University})$, that is the positive pointwise mutual information for the word "Stanford" in the context of "University", and found that to be 2.3219. The particular context we are examining is one in which "University" was the next word following "Stanford", though for this problem, you don't need to be concerned with how the specific context is defined. Your professor now wants you to find how many of the sentences you examined contained the word "Stanford". Rather than running through the entire corpus and searching for the word "Stanford", you instead attempt to calculate this count using numbers you noted from before.

You remember looking at a corpus of 100,000 sentences, and of those sentences, there was a 50% chance that you saw "Stanford" right before the word "University" in the sentence when a sentence contained "University". For the sake of simplicity, also assume that each sentence contained at most one instance of the word "Stanford" or "University". How many times did the word "Stanford" appear in your corpus? Assume that the ppmi was calculated using a log of base 2 and round your answer to the nearest integer. Also, if your integer involves more than three digits, please do not include commas in your response, (so 1,234 -> 1234).

You entered:

10000

Your Answer		Score	Explanation
10000	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

Using the formula for ppmi, we have that,

$$2.3219 = \frac{P(w|c)P(c)}{P(w)P(c)} = \frac{0.5}{x/100000}$$

which we get from the fact that the $P(c)$s cancel out. Solving for the equation of x gives us our

solution of 10000.

Question 4

Assume we have the following co-occurence vectors for the words, "fish", "bird", "ant".

"fish"
subj-of-A 3
mod-of-B 2
obj-of-B 4
mod-of-C 2

"bird"
subj-of-A 3
subj-of-D 1
mod-of-C 4

"ant"
subj-of-A 5
mod-of-C 2
subj-of-D 1
obj-of-B 1

The numbers above represent the count for the context relation to the left.
Assuming that counts not listed are 0, calculate the cosine similarity (using counts, not PMI) for "bird" and "ant". Round your answer to two decimal places (e.g. 0.35, 0.60, 1.43).

You entered:

0.85

Your Answer		Score	Explanation
0.85	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

Question explanation

Question 5

Consider the following queries and the results that follow:

What is the capital of California?

1. New York City
2. Alabama
3. Sacramento*
4. Stanford

What is the capital of Georgia?

1. Altanta*
2. Wyoming
3. Chicago
4. Washington DC
5. China
6. Steve Jobs

What is the capital of Texas?

1. Denver
2. Obama
3. New York City

What is the capital of Massachusetts?

1. Bush
2. Japan
3. Massachusetts City
4. Boston*

What is the capital of Maryland?

1. Facebook
2. Africa
3. Los Angeles
4. Annapolis*
5. England

What is the capital of Nevada?

1. Flag
2. Baton Rouge
3. Montana
4. Disney World
5. New Zealand

Note that answers which are followed by an asterisk are the correct answers to the query. What is the mean reciprocal rank for the set of queries over the capital of the following set of states:

Texas, Massachusetts, Maryland, Nevada

Round your answer to two decimal places (e.g. 0.35, 0.60, 1.43).

You entered:

0.13

Your Answer		Score	Explanation
0.13	✔	1.00	
Total		1.00 / 1.00	

Question Explanation

As an example, consider the query set for California, Georgia, Texas, and Massachusetts. In these queries, the correct answer appeared in position 3, 1, never, and 4, respectively. Taking the reciprocals of these values yields 1/3, 1, 0, and 1/4, where a list that does not contain the correct result receives a reciprocal rank of 0. Averaging the reciprocals by taking the sum and dividing by 4 gives 19/48 $\approx$ 0.396