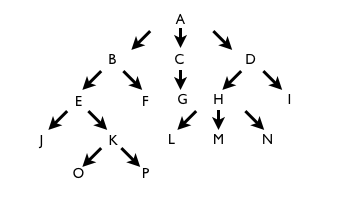
**Question 1**

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 = 100   
O = 100   
P = 100   
F = 100   
G = 100   
L = 100   
M = 100   
N = 100   
I = 200   
  
  
Assuming that all other words do not appear in the corpus, what is simresnick(N,L)? Round your answer to two decimal places (e.g. 0.35, 0.60, 1.43) and use natural log in your calculation.

simresnik(c1,c2) = -log P( LCS(c1,c2) )

LCS(c1,c2)= The lowest node in the hierarchy that subsumes both c1 and c2

P(c) = (for each w of c sum(count(w) ))/N

Then:

simresnik(N, L) = -log P( H )

P(H) = (count(L)+count(M)+count(N))/N = 300/1000 = 0.3

simresnik(N, L) = -log 0.3 = 1,2039728043259359926227462177618

**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(disc, chair) and round your answer to two decimal places.

wordsim(disc, chair)

Ecuations:

pathlen(c1, c2) = 1+ number of edges in the shortest path in the hypernym graph between sense nodes c1 and c2

Simpath(c1, c2) = 1/pathlen(c1, c2)

Wordsim(w1, w2) = max simpath(c1, c2) c1 Є senses (w1) y c2 Є senses (w2)

simpath(c1, c2 ) = 1/13

simpath(c1, c2 ) = 1/12

wordsim(disc, chair) = 1/12 = 0,0833

**Question 3**

Let's say we've calculated ppmi(Stanford, 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).

Number of Standford = P(stanford) x Corpus size.

PPMI(Stanford, University) = 2.3219.

So

2.3219 = log2(P(Stanford, University) /P(Stanford)xP(University))

4.9999 = P(Stanford, University) /P(Stanford)xP(University)

4.9999 = 0.5 / P(Standford)

**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).

= (3+1+4)/(sqrt(9+1+16)xsqrt(25+4+1+1)) = 8/(sqrt(26)xsqrt(31)) = 0,28178798147053875865570569253418