At http:// http://alfonseca.org/eng/research/wordsim353.html find human annotations for the similarity of two words. Here is a subset:

tiger	tiger	10.00
tiger	jaguar	8.00
tiger	cat	7.35
tiger	carnivore	7.08
tiger	animal	7.00
tiger	mammal	6.85
tiger	fauna	5.62
tiger	organism	4.77
jaguar	cat	7.42
jaguar	car	7.27
jaguar	stock	0.92

- Use the data provided at http://www.lsv.uni-saarland.de//data/distributional_semantics.zip . You can use this data to build a vector with context words (e.g. one left, one right) containing the frequency (or tf-idf value) for each context word. Wider context windows (e.g. five to the left and right) can also be explored.
- Compute the similarity between words by trying different metrics measuring the distance of the vectors representing the words. You can use e.g. a cosine or the Euclidian distance. Depending on what you pick, you might want to scale the result in order to make sure that the range is between 0 and 10.
- How does your model compare to the human annotation? You could for example do a scatter plot to visualize the results.

This is a very experimental task!

It is fully optional. Still it would be nice if some of you could do it and report the findings.