Exam May 2014, Questions

Natural Language Processing (University of Sheffield)

The University Of Sheffield.

Data Provided:
None

DEPARTMENT OF COMPUTER SCIENCE **Spring Semester 2013-2014**

NATURAL LANGUAGE PROCESSING **2 hours**

Answer THREE questions.

All questions carry equal weight. Figures in square brackets indicate the percentage of available marks allocated to each part of a question.

1. a) Suppose you wish to create a simple bigram part-of-speech tagger that assigns to each word its most probable tag given the word and the tag assigned to the preceding word. When the tagger has no information about the current word given the preceding tag, it will simply "backoff" and assign the most probable tag for the current word.

   (i) Give general equations specifying which tag the bigram tagger assigns to a word and how the probabilities required for the bigram tagger could be estimated from a corpus. [20%]

   (ii) Suppose that the following counts are observed in a corpus:

   (A) unigram word/tag counts (e.g. the word *We* occurs tagged as PP 21 times);

   (B) bigram tag word counts (e.g. *saw* is observed following a PP tag 15 times);

   (C) counts of occurrences of a word with a particular part-of-speech tag following a particular part-of-speech tag, (e.g. there were five occurrences where *saw* was tagged as NN following a PP tag).

	Count
We/PP	21
saw/NN	7
saw/VBD	3
her/PP	95
duck/NN	13
duck/VB	40

(A)

	Count
PP saw	15
NN her	14
VBD her	12
PP duck	20

(B)

	Count
PP saw/NN	5
PP saw/VBD	10
NNS her/PP	13
PP duck/NN	5
PP duck/VB	15

(C)

All counts involving these words and tags not listed here may be assumed to be 0. The tags are: *PP = personal pronoun*, *NN = noun, singular, common*, *VBD = verb, past tense*, *VB = verb, base*.

Determine the tagging that would be output by a bigram tagger for the sentence *We saw her duck*. [Hint: Work through the sentence one word at a time computing the relevant bigram probabilities for each step.] [40%]

b) Why do zero counts cause problems for n-gram language models? Explain why smoothing solves this problem and describe the smoothing approach known as add-one/Laplace smoothing. [20%]

c) Give the equation for calculating the perplexity of a language model for a given test set, making sure you explain the terms in the equation. Explain why measuring perplexity is considered to be a good way to evaluate language models. [20%]

2. a) (i) Explain what is meant by dictionary-based approaches to Word Sense Disambiguation. What are the advantages and disadvantages compared to alternative approaches? [20%]

(ii) Describe one dictionary-based approach to Word Sense Disambiguation. [10%]

(iii) Explain the difference between "all words" and "lexical sample" in the context of Word Sense Disambiguation problems. Which approach to Word Sense Disambiguation would be most suitable for an "all words" Word Sense Disambiguation problem? Explain your answer. [20%]

b) The remainder of this question uses the fragment of the WordNet noun hierarchy shown in Figure 1 (note that this has been simplified for clarity):

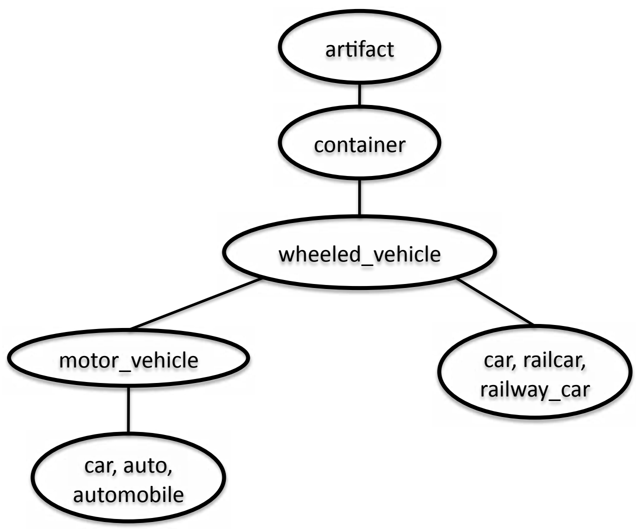

Figure 1: Fragment of WordNet hierarchy

(i) Explain the terms *synset*, *hypernym*, *hyponym* and *polysemy* in the context of WordNet. Use examples from the figure to illustrate your answer. [20%]

(ii) Similarity measures based on Information Content, such as Resnik's, make use of corpus counts to estimate the Information Content of each node in a hierarchy. Use the corpus counts below to determine the similarity between *auto* and *railcar* in the WordNet fragment using Resnik's similarity measure. Explain your working. (Note: You may leave your answers in the form of arithmetic expressions involving logarithms.)

	Count
auto	2
railcar	3
automobile	5
container	10

[30%]

3. a) The following four sentences each contain an instance of a natural language 'generalised quantifier' expression (shown underlined). For each such expression, explain how its meaning can be characterised by a formula that includes a condition on sets, in accordance with Generalised Quantifier Theory.

   (i) *At least one student will pass.*

   (ii) *John read no more than five books.*

   (iii) *Most lecturers drink coffee.*

   (iv) *Few students who study fail.* [20%]

   b) Consider the sentence

      Every child likes Santa.

   Show how a logical form for the sentence can be derived in an approach that accords with the rule-to-rule hypothesis by: (i) specifying the syntax and semantics of appropriate grammar rules and lexical entries, and (ii) providing a detailed derivation. [40%]

   c) Now consider the question

      What does every child like?

   Suggest a logical form for this question, and extend the grammar and lexicon developed for 3(b) so as to be able to derive this logical form, explaining the additions that you make [Hint: you should use a slash or gap feature in your phrase structure grammar]. Give a complete derivation for the entire example question. [40%]

4. Consider the sentence:

The girl mended the shirt with a patch

This sentence is syntactically and semantically ambiguous, in that the prepositional phrase *with a patch* may modify either the preceding noun or the verb.

a) Define a phrase structure grammar, which uses simple atomic categories (not feature representations), that will allow alternative analyses of the sentence above corresponding to the two alternative meanings. Draw the structures that the grammar allows for the sentence. [20%]

b) One basic approach to automated parsing is to use a naive top-down, depth-first strategy.

   (i) Describe the parsing strategy of a naive top-down, depth-first recursive descent parser illustrating your answer with reference to the example sentence and the grammar you have specified in 4(a). You do **not** need to provide a full derivation of the sentence. [20%]

   (ii) What are the main disadvantages of the approach to parsing described in 4(b)(i)? Illustrate your answer with reference to the example sentence. [10%]

c) Chart parsing addresses the principal weaknesses of the naive approach discussed in 4(b).

   (i) Describe an algorithm for bottom-up chart parsing. Your answer should include an account of the data structures used. Illustrate your answer with reference to the example above and the grammar you have specified for it in 4(a). A full chart derivation is not required. [40%]

   (ii) How does the chart parsing algorithm of 4(c)(i) address the problems of the naive top-down, depth-first strategy you identified in 4(b)(ii)? [10%]

END OF QUESTION PAPER