Natural Language Engineering

Monday 26th January 2009    Time: 09:45 – 11:45

Please answer any THREE Questions from the FIVE questions provided

The use of electronic calculators is NOT permitted
1. a) The distinction in meaning between arguments and modifiers accounts for ambiguous sentences and dictates the geometry of a phrase structure tree. What is the main difference between arguments and modifiers? Provide an example of a phrase where both an argument and a modifier occur, illustrating the difference of attachment between the two. Draw the tree and show the attachments. (6 marks)

b) What do we mean by subcategorisation frame of a verb? Provide an example of a transitive, an intransitive and a ditransitive verb. (4 marks)

c) What is morphology? Provide an example of a morphologically complex word. (3 marks)

d) Between the words 'eat' and 'find', which would you expect to be more effective in selectional restriction-based sense disambiguation? Why? (3 marks)

e) How would you express the probability estimate of the sequence [He runs a company] using a uni-gram language model? (2 marks)

f) Provide a formula to calculate the probabilities in the unigram model. (2 marks)

2. a) Critically discuss the difference between transfer based machine translation and interlingua? Provide examples and discuss advantages and disadvantages of each design. (5 marks)

b) Describe the difference between rule based vs example based MT (EBMT) systems. Provide the definition of EBMT. (5 marks)

c) What syntactic knowledge can be used in resolving the ambiguity in word senses? Give an example and explain how it can help and also its limitations. (2 marks)

d) A machine learning based Word Sense Disambiguation (WSD) system essentially looks at a vector of collocational or co-occurrence features extracted from the context of an occurrence of a word, and classifies it into one of its possible senses. Suppose word 'w' has a set of senses S, where S=\{s_1, s_2, ..., s_n\}, and we have defined a set of features V, where V=\{v_1, v_2, ..., v_m\}. Given an input vector of feature values, describe how one can predict the most probable sense, using the naive Bayes classification. (8 marks)
3. a) What is the scope of automatic term recognition? Provide the definition of ‘a term’. Enumerate some of the applications of automatic term recognition in NLP and related problems. (6 marks)

b) Define evaluation metrics 'precision', 'recall', and 'F1' score, given the true positives (TP), false positives (FP), and false negatives (FN). (3 marks)

c) Suppose you were asked to provide an Information Extraction system for joint ventures. Consider the following sentence:

'Bridgestone Sports Co. said Friday it had set up a joint venture in Taiwan with a Japanese trading house and Nike Inc.'

i) A typical IE system like FASTUS first recognizes complex words and basic phrases. Show which parts of the sentence should be recognized as complex words and basic phrases. (2 marks)

ii) Ambiguities such as PP attachment and scope of coordination are the major difficulties in NLP. Show where such ambiguities appear in the above sentence. (3 marks)

iii) We would like to extract information concerning which companies establish or disband a joint venture and where. Design a template for the IE system and show an event pattern by which information can be extracted from the above sentence. Show the result of IE. (6 marks)
4. a) Explain, with examples, how using a feature-based framework lets you write more compact grammars than is possible with simple context-free rules. (6 marks)

b) The grammar below uses the category-valued feature slash to account for the movement of the phrases on the bus and an old man from their normal positions in the sentence On the bus sat an old man. Show either a top-down or a bottom-up derivation for this sentence. You should assume that SNP+SVP in the first rule denotes the concatenation of the two lists SNP and SVP. (10 marks)

c) Explain why the notion of sponsorship cannot be used to control the generation of slashed items in this case. (4 marks)

\[
\begin{align*}
\text{s[slash=(SVP+SNP)]} & \Rightarrow \text{np[slash=SNP], vp[slash=SVP]} \\
\text{s[slash=[SH \mid ST]]} & \Rightarrow \text{s[slash=ST], SH} \\
\text{s[slash=[SH \mid ST]]} & \Rightarrow \text{SH, s[slash=ST]} \\
\text{SLASH[slash=[SLASH]]} & \Rightarrow \\
\text{np[slash=[]]} & \Rightarrow \text{det, nn} \\
\text{nn} & \Rightarrow \text{adj, nn} \\
\text{nn} & \Rightarrow \text{noun} \\
\text{pp[slash=[]]} & \Rightarrow \text{prep, np[slash=[]]} \\
\text{vp} & \Rightarrow \text{iverb} \\
\text{vp[slash=SPP]} & \Rightarrow \text{vp, pp[slash=SPP]} \\
\text{det} & \Rightarrow \text{an} \\
\text{det} & \Rightarrow \text{the} \\
\text{noun} & \Rightarrow \text{bus} \\
\text{noun} & \Rightarrow \text{man} \\
\text{adj} & \Rightarrow \text{old} \\
\text{prep} & \Rightarrow \text{on} \\
\text{iverb} & \Rightarrow \text{sat}
\end{align*}
\]

5. a) How would you express the probability estimate of the sequence [He runs a company] using a bi-gram model? (4 marks)

b) What is the advantage of using tri-gram models instead of bi-gram models? What is the disadvantage? (4 marks)

c) Explain very briefly the transition and emission probabilities in a first-order hidden Markov model for part-of-speech tagging. (4 marks)

d) What is the merit of using the Viterbi algorithm? (4 marks)

e) Why is term weighting important in IR and text classification? Define the inverse document frequency term weight. (4 marks)

END OF EXAMINATION