UNIVERSITY OF TORONTO
Faculty of Arts and Science

APRIL/MAY EXAMINATIONS 2005

CSC 401H1 S and 2511H S

Duration - 3 hours

No aids allowed

Last Name:

First Name:

Student Number:

Rules:

1. This exam paper contains 11 pages. Legibly write your last name and student number on every page, as the pages of this exam will be separated during marking. *Unnamed pages will not be marked.*
2. There are 34 questions which you must solve, worth a total of 100 marks.
3. Write your answers directly on the exam in the space provided, or if you run out of space, on the back of the same page. For multiple choice questions, write the letter of the correct answer to the left of the question.

Do not write below this line.

MC:

SA:

BW:

WS:

TS:

Total:

Last Name: Student Number:

Multiple Choice questions (2 marks each)

Legibly write your answer in the left margin of each question.

M1. An allophone is _
- (a) a basic unit of speech found in all languages.
- (b) a psychologically distinctive unit of speech.
- (c) one of several units of speech that are indistinguishable to native speakers.
- (d) a person who is afraid of everyone else.

M2. _ is not a part of speech.
- (a) "Subject"
- (b) "Noun"
- (c) "Adverb"
- (d) "Participle"

M3. Hidden Markov models differ from other Markov models in that _
- (a) they emit symbols at states rather than on arcs.
- (b) not all of their output symbols are known.
- (c) their output emissions are not observable.
- (d) their state transitions are not observable.

M4. The [j] sound is a _ consonant.
- (a) low
- (b) fricative
- (c) velar
- (d) palatal

M5. Interpolated average precision is useful for _
- (a) comparing retrieval methods that differ only in the order of the documents returned.
- (b) evaluating retrieval with a high sensitivity to recall.
- (c) comparing only the precision of two methods.
- (d) all of the above.

M6. WordNet is _
- (a) a weighted graph that indicates semantic distances between concepts.
- (b) a large IS-A hierarchy.
- (c) a structured database of synonym classes.
- (d) all of the above.

M7. The intuition behind Witten-Bell smoothing is to _
- (a) waste no probability mass on unseen events.
- (b) base the probability of an unseen outcome on the probability of observing an outcome for the first time.
- (c) linearly interpolate between binomial and uniform distributions.
- (d) first select from among a uniformly distributed set of distributions.

M8. The intuition behind Maximum Likelihood Estimation is to _
- (a) waste no probability mass on unseen events.
- (b) base the probability of an unseen outcome on the probability of observing an outcome for the first time.
- (c) linearly interpolate between binomial and uniform distributions.
- (d) first select from among a uniformly distributed set of distributions.

M9. The forward-backward algorithm is used to _ in an HMM.
  (a) find the most probable state sequence given an output sequence
  (b) find the most probable output sequence given a state sequence
  (c) maximise the probability of training data by re-estimating parameters
  (d) calculate the probability of an output sequence

M10. European digital telephone networks use the _ encoding to warp the signal along the y-axis.
  (a) μ-law
  (b) A-law
  (c) Grimm's law
  (d) Murphy's law

M11. In part-of-speech tagging, Markov models are used to _
  (a) implement the Brill tagger.
  (b) combine syntagmatic and paradigmatic statistics on tag assignment.
  (c) iteratively improve upon an initial tag assignment.
  (d) outperform the Brill tagger.

M12. C4.5 greedily builds a decision tree by _
  (a) minimising the distance from the root to the leaves.
  (b) minimising entropy.
  (c) maximising information gain.
  (d) minimising information gain.

M13. The lowest formant, F_0, of a speech signal corresponds to _
  (a) pitch.
  (b) spectral tilt.
  (c) vowel quality.
  (d) word sense.

M14. Naive Bayes classification assumes _
  (a) the optimal class can be chosen by maximizing the joint probability of the class and features.
  (b) that Bayes' rule is false.
  (c) that using a single, reliable feature of context is better than combining evidence from all features.
  (d) none of the above.

M15. Yarowsky's algorithm is superior to Lesk's algorithm and all of its variants in that it can make use of _
  (a) bitexts.
  (b) thesauri.
  (c) unannotated texts.
  (d) cosine similarity scores.

M16. The perceptron algorithm is _
  (a) guaranteed to converge.
  (b) prone to converging to local optima.
  (c) a gradient descent algorithm.
  (d) none of the above.

M17. A disadvantage of using k-Nearest Neighbours for classification is that _
  (a) it is ineffective for problems that cannot be approached by greedily dividing the reference set.
  (b) it takes too long to train.
  (c) the optimal choice of k is often unclear.
  (d) all of the above.

M18. The backtrace is used during Viterbi decoding to _
- (a) ensure that no errors were made.
- (b) maximize the probability of the observed data.
- (c) reconstruct the optimal state sequence.
- (d) re-estimate the parameters of the model.

M19. Subject-verb agreement is observable in _
- (a) case, number and gender.
- (b) tense, number and mood.
- (c) person, number and gender.
- (d) time, place, manner, and degree.

M20. A word type that occurs once in a corpus is called a _
- (a) singular value.
- (b) hapax legomenon.
- (c) word token.
- (d) typographical error.

Short Answer questions (3 marks each)

A1. State Bayes' Rule.

A2. Define the entropy of a probability distribution, p:

$H(p) =$

A3. Classify the consonantal phonemes (3 in all) corresponding to the underlined letters in the following English words by their voicing, place and manner of articulation: $\underline{\text{c}}$ity, $\underline{\text{g}}$oa$\underline{\text{t}}$.

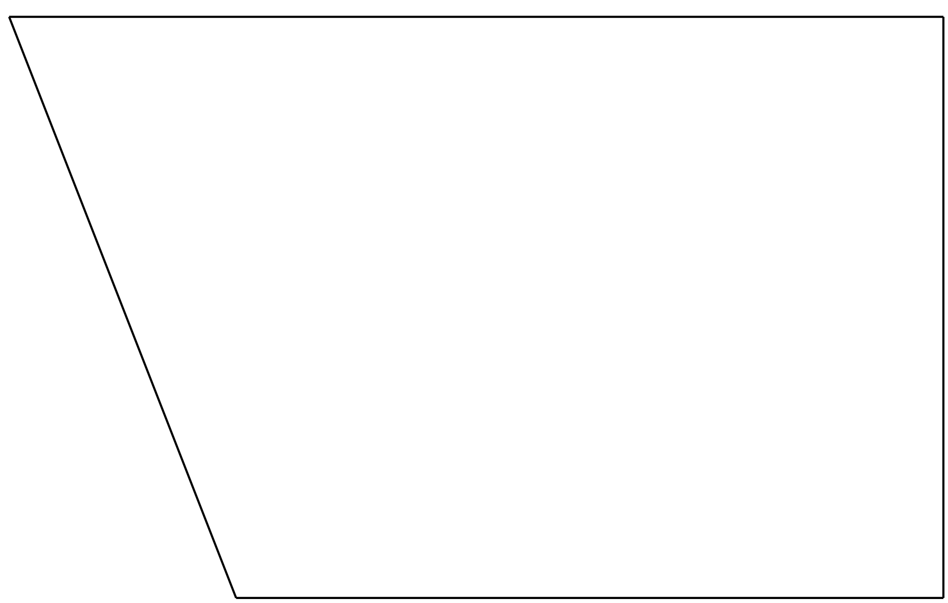
Last Name: Student Number:

A4. Compute the mutual information of X and Y, $I(X;Y)$, given the following joint distribution:

$p(x,y)$	1	2	3
1	0	$\frac{1}{16}$	$\frac{1}{16}$
2	$\frac{3}{8}$	$\frac{1}{16}$	$\frac{1}{16}$
3	0	$\frac{3}{8}$	0

A5. Place the following five vowel sounds in their appropriate position in the vocalic trapezoid below:

1. [i], as in sweet
2. [a], as in father
3. [æ], as in mat



A6. Informally state Zipf's First Law and explain its importance to corpus linguistics.

A7. Given n features of context, $\vec{f} = (f_1 \dots f_n)$, define how a naive Bayes classifier computes the most probable class. Your answer must explicitly refer to the features.

$\bar{c} = \arg\max_{c}$

A8. State (without proof) the Good-Turing Theorem.

A9. Define precision, recall, and F-measure as a function of the contingencies $\{$true,false$\} \times \{$positive,negative$\}$.
Note: true-negative (tn) refers to an outcome that is truly negative and was identified as such, not to one that is “true” in the sense of being positive.

A10. Name one advantage and two disadvantages of using decision trees relative to the other classification methods presented in class.

Baum-Welch Re-estimation

B1. (10 marks) Given an initial HMM, $\langle A, B, \Pi \rangle$, and an output sequence, $O = o_1 \ldots o_T$, define the next iteration of model, $\langle \hat{A}, \hat{B}, \hat{\Pi} \rangle$ that Baum-Welch re-estimation produces. You may assume that the forward probabilities, $\alpha_i(t)$, and backward probabilities, $\beta_j(t)$, have been defined and calculated for you.

Word-Sense Disambiguation

W1. (7 marks) Given a token sequence, $v_1 \ldots v_{m-1}\ w\ v_m \ldots v_n$, in which w is a sense-ambiguous word type, define Lesk's algorithm for disambiguating the sense of this occurrence of w.

W2. (3 marks) What linguistic resources does Lesk's algorithm require? Name two additional resources that have been used to improve the performance of Lesk's algorithm.

Text Summarization

T1. (10 marks) List and explain five other dimensions along which text summarization systems can differ from each other.

0. Extractive vs. Abstractive
Extractive summaries consist only of phrases from the original document; abstractive summaries generate new text so that the summary reads more fluently.

1. vs.

2. vs.

Last Name: Student Number:

3. vs.

4. vs.

5. vs.