

EXAMINATIONS — 2012

MID-YEAR

COMP 423
INTELLIGENT AGENTS

Time Allowed: 3 Hours

Instructions: Attempt all questions.

The exam will be marked out of 180.

Calculators and non-electronic foreign language dictionaries are permitted.

Clean copies of the papers will be distributed for the exam.

Questions

- | | |
|------------------------|------|
| 1. Basic Concepts | [20] |
| 2. Text Representation | [20] |
| 3. Wikipedia | [20] |
| 4. Clustering | [40] |
| 5. Two Papers | [60] |
| 6. Your Project | [20] |

Question 1. Basic Concepts

[20 marks]

Briefly explain the following concepts in the context of information retrieval.

- (a) [5 marks] Vector space model
- (b) [5 marks] Tf-idf
- (c) [5 marks] Feature selection and feature extraction
- (d) [5 marks] Relevance feedback

Question 2. Text Representation

[20 marks]

The “bag-of-words” model for text representation does not consider “semantics”. In your opinion, how would you represent semantics and integrate it into text representation?

Question 3. Wikipedia

[20 marks]

Recent research suggests that Wikipedia is useful for many tasks such as document clustering/classification, information retrieval, query expansion, semantic relatedness calculation, disambiguation, etc. Choose one task and explain how you would use Wikipedia to improve the current technology.

Question 4. Clustering

[40 marks]

- (a) [10 marks] Explain how a single-link Hierarchical Agglomerative Clustering algorithm works? What are the main limitations of this algorithm?
- (b) [10 marks] HAC and K-means are two well known clustering algorithms. Name **TWO** situations when HAC is better than k-means and justify your answer.
- (c) [20 marks] Suffix Tree clustering is one of the algorithms for web search results clustering. In your opinion, what are the main challenges and problems for web search results clustering and how you might solve these problems?

Question 5. Two papers

[60 marks]

We discussed the following seven papers in this course. Copies of all these papers should be available during the exam.

1. Clustering documents using Wikipedia-based concept Representation.
2. Computing Semantic Relatedness using Wikipedia-based Explicit semantic analysis.
3. Concept-Based Feature Generation and Selection for Information Retrieval.
4. Thumbs up or Thumbs down? Semantic Orientation applied to Unsupervised classification of Reviews.
5. Learning to Cluster Web Search Results.
6. Query dependent Pseudo-Relevance Feedback based on Wikipedia
7. Open Information Extraction from the Web

Choose **TWO** papers and answer the following questions for each paper:

(a) What are the main contributions of this paper?

(b) What are the main limitations of this paper? Your answers should include any solutions you may have for addressing these limitations.

Please clearly label your answers, for example, if you choose paper 1 and 7, then label your answers as 5.1.a, 5.1.b, 5.7.a, and 5.7.b.

Question 6. Your Project

[20 marks]

What are the main ideas and the main contributions of your COMP423 project? What are the main limitations of your current system? If it were a Masters project and you had one year to work full time on your project, how would you improve your system?
