



### **Semester 1 Examinations 2019/2020**

<b>Course Instance Code(s) Exam(s)</b>	1CSD1, 1CSD2, 1SPE1, 1MAO2, 1MAI1 MSc in Computer Science (Data Analytics), MSc in Computer Science (Artificial Intelligence), MSc in Computer Science (Artificial Intelligence) - Online
<b>Module Code(s) Module(s)</b>	CT5120, CT5146 Introduction to Natural Language Processing, Introduction to Natural Language Processing - Online
Paper No.	1
Repeat Paper	No
External Examiner(s)	Professor Pier Luca Lanzi
Internal Examiner(s)	Dr. Michael Madden *Dr. Paul Buitelaar, Dr. John McCrae

**Instructions:** Answer all parts of all questions. There are 4 sections; each section is worth 25 marks (100 marks total). **Use a separate answer book for each section answered.**

<b>Duration</b>	2 hours
<b>No. of Pages</b>	5
<b>Discipline(s)</b>	Computer Science
<b>Course Co-ordinator(s)</b>	Dr. Enda Howley, Dr. Michael Schukat, Dr. James McDermott

### **Requirements:**

Release in Exam Venue	Yes	<input checked="" type="checkbox"/>	No	<input type="checkbox"/>
MCQ	Yes	<input type="checkbox"/>	No	<input checked="" type="checkbox"/>
Handout	None			
Statistical/ Log Tables	None			
Cambridge Tables	None			
Graph Paper	None			
Log Graph Paper	None			
Other Materials	None	<input type="checkbox"/>		<input checked="" type="checkbox"/>
Graphic material in colour	Yes		No	



# Introduction to Natural Language Processing

Exam Duration: 2 Hours

**You must complete Sections 1 to 4**

## Section 1: Semantics

**Instructions:** Provide answers for questions 1A, 1B, 1C and 1D

### Question 1A

**10 Marks**

Consider the following sentences:

*The black horse jumped over the white fence.*  
*The black fence was too high for the black horse.*  
*The white horse could not have jumped over the black fence.*  
*The black horse jumped successfully over the black fence.*  
*The white fence was also too high for the white horse.*

Using cosine similarity, compute the distributional semantic distance between the vectors for 'black' and 'white' constructed on the basis of the sentences above with a context window of two tokens. Give the vectors as well as the similarity score.

### Question 1B

**5 Marks**

How can dependency parsing be used in a vector space model?

### Question 1C

**5 Marks**

What is the opposite of a hyponym?

### Question 1D

**5 Marks**

Give 2 examples of types of semantic roles and construct a sentence in which both of these can be used. Annotate the relevant words in the sentence for each of the 2 roles.

**PTO**

## Section 2: Part-of-speech tagging

**Instructions:** Provide answers for question 2A, 2B and 2C

### Question 2A

**10 Marks**

Consider a Hidden Markov Model with the following probabilities (Start designates the start state):

$p(w_i t_i)$	$w_i = \text{tag}$	$w_i = \text{this}$	$w_i = \text{text}$	$w_i = \text{herring}$
$t_i = N$	0.2	0.2	0.5	0.1
$t_i = V$	0.6	0.1	0.1	0.2
$t_i = O$	0.1	0.7	0.1	0.1

$p(t_i t_{i-1})$	$t_{i-1} = N$	$t_{i-1} = V$	$t_{i-1} = O$	$t_{i-1} = \text{Start}$
$t_i = N$	0.4	0.3	0.4	0.4
$t_i = V$	0.2	0.2	0.3	0.5
$t_i = O$	0.4	0.5	0.3	0.1

By using the Viterbi algorithm or otherwise, what is the most likely sequence of tags for the text “tag this text”?

### Question 2B

**5 Marks**

What is the advantage of the Viterbi algorithm over an exhaustive search of all possible part-of-speech tag combinations?

### Question 2C

**10 Marks**

Give one advantage and disadvantage of unsupervised learning of a Hidden Markov Model by means of the Baum-Welch algorithm as opposed to supervised learning of the model. Why might you wish to combine supervised and unsupervised learning for Hidden Markov Models?

**PTO**

### Section 3: Sentiment Analysis

**Instructions:** Provide answers for question 3A, 3B and 3C

Consider the following sentences from product reviews with sentiment scores, using a range of -1, 0, +1:

- 1      *The food was great, but the service in this restaurant was unfriendly.*
- +1      *The steak was cooked to perfection and the service was great.*
- 0        *The best part of the meal was the beer.*

#### Question 3A

**10 Marks**

Create a sentiment lexicon on the basis of the review sentences given above.

#### Question 3B

**10 Marks**

Recall that a count vector represents the proportions of negative and positive sentiment in a review text. Using the sentiment lexicon you created in question 3A, calculate the count vector for each review sentence given above.

#### Question 3C

**5 Marks**

Give an example of *aspect-based* sentiment analysis from the review sentences given above.

**PTO**

## Section 4: Information Extraction & Knowledge Graphs

**Instructions:** Provide answers for questions 4A, 4B and 4C

Consider the following sentence:

*David Robert Joseph Beckham is a former professional footballer who played for Manchester United and the England national team.*

### Question 4A

**10 Marks**

Annotate the sentence above for Named Entity types 'person' (PER), 'organization' (ORG) and 'location' (LOC) by the use of the IOB tagging scheme. Explain the reasoning behind your annotations.

### Question 4B

**10 Marks**

The sentence above provides a positive instance for extracting the 'played for' relation between a 'person' (footballer) and an 'organization' (football team). Give a negative instance for this relation.

### Question 4C

**5 Marks**

How can clustering be used in taxonomy extraction? What are potential problems with this approach?

**END**