NUI Galway
OÉ Gaillimh

Autumn Examinations 2019/2020

Course Instance Code(s)	1CSD1, 1CSD2, 1SPE1, 1MAO2
Exam(s)	MSc in Computer Science (Artificial Intelligence) - Online
Module Code(s)	CT5146
Module(s)	Introduction to Natural Language Processing - Online
Paper No.	1
Repeat Paper	Yes
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). This is an open book exam.

Duration	2 hours
No. of Pages	6
Discipline(s)	School of Computer Science
Course Co-ordinator(s)	Dr. James McDermott

Requirements:

Release in Exam Venue	Yes	X	No	
MCQ	Yes		No	X
Handout	None			
Statistical/ Log Tables	None			
Cambridge Tables	None			
Graph Paper	None			
Log Graph Paper	None			
Other Materials	None			
Graphic material in colour	Yes		No	X

Exam Integrity Statement

By opening this exam paper, you agree to be bound by the following exam integrity statement:

In submitting this work, I confirm that it is entirely my own. I acknowledge that I may be invited to online interview if there is any concern in relation to the integrity of my exam, and I am aware that any breach will be subject to the University's Procedures for dealing with breaches of Exam Regulations:

https://www.nuigalway.ie/media/registry/exams/QA230---Procedures-for-Dealing-with-Breaches-of-Examination-Regulations.pdf

Introduction to Natural Language Processing

Exam Duration: 2 Hours

You must complete Sections 1 to 4

Section 1: Linguistics

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

Question 1A **5** Marks

How many tokens are there in this sentence? Explain your reasoning.

Boris Johnson has been the prime minister of the UK since last year.

Question 1B **10** Marks

Fill in the blanks in these statements:

government and *govern* are morphologically related through ___

minister and *ministers* are morphologically related through ___

road and *roadmap* are morphologically related through ___

Question 1C **10** Marks

Describe in your own words the difference between a parallel and comparable corpus. Give an example of an NLP application that uses such corpora.

PTO

Section 2: Parsing

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

Question 2A **10** Marks

Consider the following grammar:

Rule	Probability	Rule	Probability
S → NP VP	1.0	D → the	0.5
NP → D N	0.4	D → a	0.5
NP → N	0.5	N → coffee	0.3
NP → Prn	0.1	N → function	0.7
VP → V	0.3	V → function	0.6
VP → Aux VP	0.2	V → can	0.4
VP → Adv VP	0.1	Prn → I	1.0
VP → V PP	0.4	Aux → can	1.0
PP → Prp NP	1.0	Adv → only	1.0
		Prp → with	1.0

What is the probability of the following sentence in this grammar?

I can only function with coffee

Show which rules in the grammar were used in the parse tree of this sentence.

Question 2B **10** Marks

Using the grammar of Question 2A, find a sentence that is accepted by the grammar but is not grammatical in English and suggest a modification to the grammar so that this sentence is not generated.

Question 2C **5** Marks

Why do lexical dependencies cause an issue with a simple Probabilistic Context-Free Grammar (PCFG) approach to parsing?

PTO

Section 3: Semantics

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

Question 3A **5** Marks

Fill in the blanks in these statements on words that are semantically related:

government, cabinet, *administration* are ___

light and *dark* are ___

Question 3B **5** Marks

Explain in your own words how word senses are represented in WordNet. Give an example.

Question 3C **5** Marks

Explain in your own words how word senses are represented in FrameNet. Give an example.

Question 3D **10** Marks

How can Wikipedia be used in word sense disambiguation?

PTO

Section 4: Applications

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

Question 4A **10** Marks

Explain in your own words how a knowledge model can be used in information extraction. Give an example.

Question 4B **10** Marks

Consider Pointwise Mutual Information (PMI). Given words *a,b,c*, explain how PMI(a,b) for a given corpus can be higher than PMI(a,c).

Question 4C **5** Marks

Discuss a limitation of a lexicon-based approach to sentiment analysis.

END