



Autumn Examinations 2019/2020

Course Instance Code(s)	1CSD1, 1CSD2, 1SPE1, 1MAO2, 1MAI1
Exam(s)	MSc in Computer Science (Data Analytics), MSc in Computer Science (Artificial Intelligence), MSc in Computer Science (Artificial Intelligence) - Online
Module Code(s)	CT5120, CT5146
Module(s)	Introduction to Natural Language Processing, 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 examination, please upload all answers as a single file

Duration	2 hours
No. of Pages	5
Discipline(s)	Computer Science
Course Co-ordinator(s)	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			
Graphic material in colour	Yes	<input type="checkbox"/>	No	<input checked="" type="checkbox"/>

Plagiarism policy

Please include the following statement in your submission:

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

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

12 Boris Johnson is a named

5 Marks

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

Boris Johnson has been the prime minister of the UK since last year. should be tokened as 1

Question 1B

10 Marks

Fill in the blanks in these statements:

government and govern are morphologically related through derivation

minister and ministers are morphologically related through inflection

road and roadmap are morphologically related through compound / derivation

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.

parallel → collection of ^{exact} translated documents
→ translations are referring to the same thing

PTO

comparable → collection of documents on same topic but in different language.

machine translation

Section 2: Parsing

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

Question 2A

10 Marks

Consider the following grammar:

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

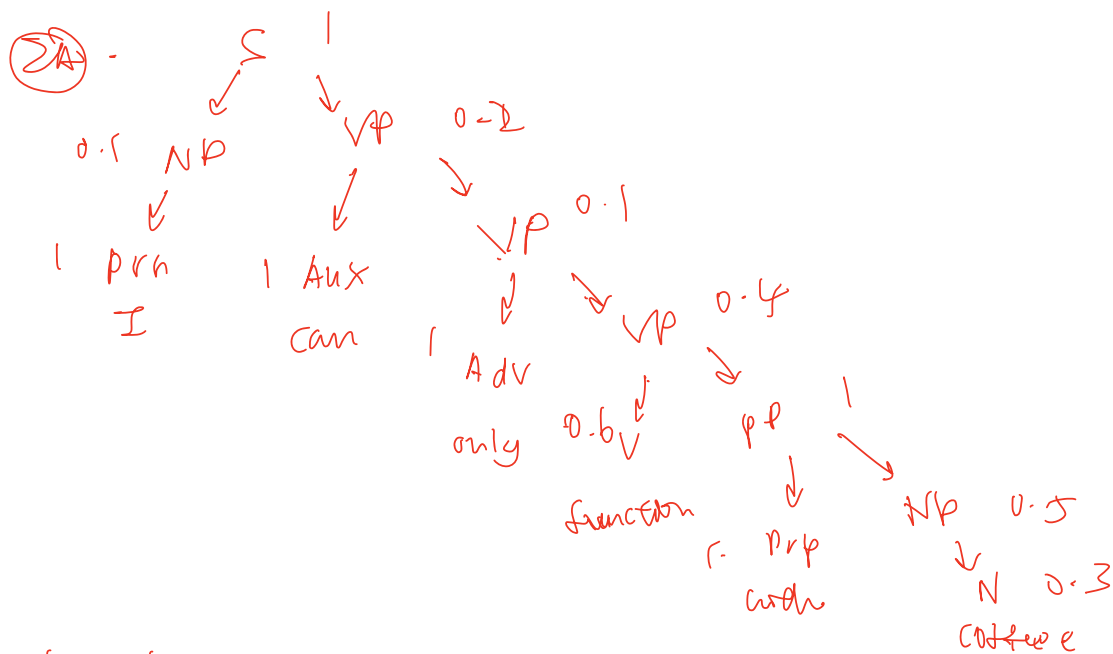
Using the grammar of Question 2A, find a sentence that is accepted by the grammar but is not grammatical in English and suggest a modification to the grammar so that this sentence is not generated.

Question 2C

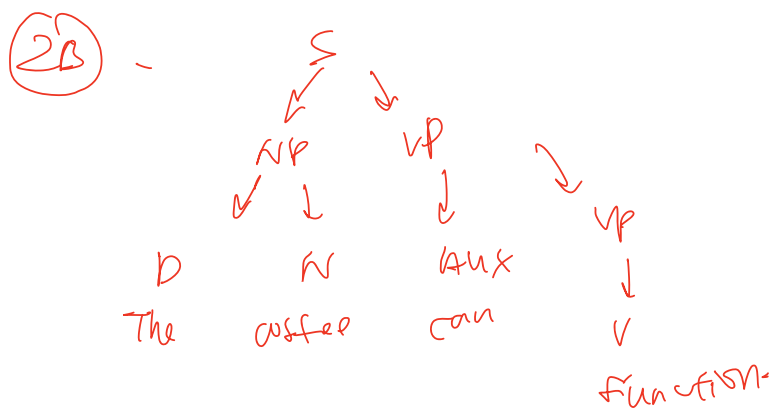
5 Marks

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

PTO



[x0.1 x0.2 x1x1x0.1 x1x0.6 x0.6x1x1x0.5
x0.3] = 0.00024.



2C - lexical dependencies - could not capture non-direct dependencies

Section 3: Semantics

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

Question 3A

Fill in the blanks in these statements on words that are semantically related by stating the kind of semantic relation:

government, cabinet, administration are synonym

light and dark are antonym

Question 3B

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

Question 3C

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

Question 3D

How can Wikipedia be used in word sense disambiguation?

considered nearly labeled data with wikipedia categories as.

PTO

wikipedia categories information is align with word net senses.

treated as semi-supervised method

as they are related data at a distance.

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

correlation / association measure that quantifies the likelihood of co-occurrence of 2 events
Given words a, b, c , explain how $\text{PMI}(a, b)$ for a given corpus can be higher than $\text{PMI}(a, c)$. [PMI = Pointwise Mutual Information] $\Rightarrow \log(p(a, b) / (p(a) \times p(b)))$

Question 4C

5 Marks

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

(4B) If co-occurrence of (a, b) is higher than
END

(a, c) & occurrence of $b \geq c$

, then $\text{PMI}(a, b) > \text{PMI}(a, c)$

lexicon may not be complete,
might have negation which ~~is~~ uses
positive words in negative sentiment.