

Department of Computer Science & Informatics

CSIA7915

Natural Language Processing



Module Guide (2023) **Version 1**

NQF level 9
20 Credits

Compiled by Prof Eduan Kotzé
© 2023

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA



UFS·UV
NATURAL AND
AGRICULTURAL SCIENCES
NATUUR- EN
LANDBOUWETENSKAPPE

COMPUTER SCIENCE AND INFORMATICS
REKENAARWETENSKAP EN INFORMATIKA

Table of Contents

1. Contact Details	3
2. Module Objectives	3
3. Module Outcomes	3
4. Prerequisites	3
5. Study Material.....	4
6. Additional Study Material	4
7. Contact Sessions	4
8. Communication.....	5
9. Practical Assignment	5
10.Assessment.....	5
ASSESSMENT GUIDE: MARK ALLOCATION FOR CSIS7915	6
11.Module Schedule.....	7

1. Contact Details

Lecturer and Academic Department Head:	Prof Eduan Kotzé
Office:	WWG 313A
Telephone:	051 4013707
E-mail:	KotzeJE@ufs.ac.za
Consulting Hours:	Per appointment
Secretary:	Ms Sanah Tabanyane
Office:	WWG 312
Telephone:	051 4012754
E-mail:	TabanyaneMS@ufs.ac.za
Blackboard URL:	http://learn.ufs.ac.za

2. Module Objectives

To understand and apply the theoretical concepts of data science (data analysis and practical data visualization from large document collections), natural language processing (practical applications on textual data), and machine and deep learning.

3. Module Outcomes

After the successful completion of the module the student will have:

1. Learnt to apply the latest machine learning, deep learning and natural language processing techniques and approaches.
2. Acquire hands-on experience with software for text categorisation, language understanding, and generation.
3. Produced a short research proposal (minimum length of 20 pages) that should broadly include an Introduction, Literature Study, Proposed Research Idea(s) and a Conclusion. The student is assessed on this document.
4. Made a presentation at the end of the semester. The student is assessed on this presentation.

4. Prerequisites

None.

5. Study Material

Prescribed textbooks for this module:

1. Jurafsky & Martin. *Speech and Language Processing*. Third Edition if available, otherwise the online draft chapters: <https://web.stanford.edu/~jurafsky/slp3/>

6. Additional Study Material

Supplementary notes will periodically be posted on the Blackboard module website during the semester. Please note, all material which is discussed during lectures and/or placed on Blackboard is relevant for evaluation purposes and not only the contents of the prescribed text book. It is the responsibility of the student to ensure that you are aware of all content which is covered during contact sessions or which is placed on Blackboard.

7. Contact Sessions

Being at post-graduate level, students are expected to do a lot of work on their own – this involves preparation of the work **before** the contact sessions and to do research on certain topics. The spoon-feeding idea of lectures at pre-graduate level does not apply at postgraduate level. The purpose of contact sessions is to explain problematic issues and to discuss problems that the students identified with the work **while preparing the designated chapters in the text book**. A student who does not prepare the work beforehand, will not gain the benefit of these contact sessions and will, in effect, be a burden to the lecturer and fellow-students, as they will not be able to contribute in a meaningful way. It is in the student's own interest to ensure that contact sessions are attended. It will be presented in an interactive format. Please refer to the schedule at the end of the module guide for more information regarding the planned contact sessions.

English		Time	Venue
	Monday	14:30 – 16:30	Online/BB Colab

Please note that only registered students will be allowed to attend the CSIA7915 contact sessions.

8. Communication

It is your own responsibility to ensure that you receive all the information in connection with this module provided during the theory classes and practical sessions, via e-mail, on the Blackboard module web site, on the notice boards in the Mathematical Sciences Building (third floor) and in the Faculty yearbook (calendar).

Communication with the lecturer via e-mail is preferred (kotzeje@ufs.ac.za). Daily visits to the Blackboard module website are necessary to keep up to date with what is going on in CSIA7915. Students are welcome to schedule appointments with the lecturer via email.

9. Practical Assignment

Late submissions will be penalised with **20% per calendar day**, starting from the date and time of submission, for the first three calendar days. Thereafter, you will receive a zero for the practical or assignment. However, you must still submit a practical or assignment which meets all the stipulated requirements otherwise you will receive an incomplete for the module. If your practical or assignment is more than seven calendar days late, you will receive an incomplete. Your module will be marked incomplete if you do not complete all the assignments to the satisfaction of the lecturer.

No co-operation whatsoever between students is allowed. Students are also not allowed to submit same/similar assignments/projects as a fellow student(s) OR an assignment/project that was previously submitted by another student for the same/different module. You may only submit your own work. Students found guilty of any transgression in this regard will be severely punished. Action will also be taken against students who help or allow other students to copy their work. The only exception to this rule is when assignments are classified as group assignments. (Also refer to the Departmental Disciplinary Code on the Blackboard module website.)

It is your own responsibility to keep backup copies of your assignments. Lost or corrupted computer files, network problems, etc., cannot be used as an excuse for late submissions.

10. Assessment

In this module, **you will choose an NLP topic from a closed list**. You need to produce a short proposal (minimum length of 20 pages) that should broadly include an Introduction, Literature Study, Proposed Research Idea(s) and a Conclusion. Students will also be expected to make **a presentation** about their work at the end of the semester.

ASSESSMENT GUIDE: MARK ALLOCATION FOR CSIS7915

• LITERATURE STUDY (Maximum: 30 marks)

- adopts a thematic approach
- is properly and logically structured
- avoids source-by-source treatment
- avoids literally reproducing other authors' words and phrases
- consistently acknowledges all sources according to an acceptable referencing technique
- has consulted enough recent and relevant literature sources
- includes definitions
- consistently address the focus(es) of the literature review topic
- shows/illustrates a thorough understanding of research area

Mark allocation: _____ /30

• PROPOSED RESEARCH IDEAS (Maximum: 10 marks)

- feasibility of the potential research idea
- applicability
- research idea link properly with the evidence accumulated from literature

Mark allocation: _____ /10

• TECHNICAL ASPECTS (Maximum: 10 marks)

- language usage; spelling
- quality of the layout
- punctuation marks
- use of appropriate headings, sub-headings
- acceptable referencing techniques
- complete list of references
- title page, declaration, table of contents, appendices, etc.
- illustrations, graphic representations etc.

Mark allocation: _____ /10

• OVERALL IMPRESSION (Maximum: 10 marks)

- Is the selected literature review topic relevant to the specific research area?
- Does the literature review sufficiently address/discuss the stated topic?
- Does the literature review have overall coherence according to a logically justified structure?
- Length of document?
- Are there unsubstantiated assertions?
- Has the candidate shown the ability to evaluate and interpret findings from other research studies?

Mark allocation: _____ /10

• PRESENTATION (Maximum: 10 marks)

- Professionally done?
- Has the candidate shown the ability to convey the findings?

Mark allocation: _____ /10

TOTAL: _____ /70 _____ %

11. Module Schedule

Note: This schedule is subject to change and it remains the sole responsibility of the student to ensure they have the correct information pertaining to the course.

For a detailed outline of the module schedule, please visit:

<https://github.com/eduankotze/BITM>
