

H8TA: Text Analytics

Module Code:	H8TA
Long Title	Text Analytics APPROVED
Title	Text Analytics
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	10
Module Coordinator:	
Module Author:	Isabel O'Connor
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	Master's / PhD degree in a computing or cognate discipline. May have industry experience also.
Learning Outcomes	
<i>On successful completion of this module the learner will be able to:</i>	
#	Learning Outcome Description
LO1	Rationalise and defend methodological choices in pre-processing methods for text analytics
LO2	Build and critically evaluate text analytics models in a variety of contexts
LO3	Execute and document corpus-based case studies
LO4	Evaluate and discuss the impact machine learning models applied to text corpora
Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Entry requirements	Learners should have attained the knowledge, skills and competence gained from stage 3 of the BSc (Hons) in Data Science

H8TA: Text Analytics

Module Content & Assessment			
Indicative Content			
Introduction Introduction to:. Text Analytics, . Key Domain, Methods and Ethics. and software libraries / packages, and Web APIs. E.g.: NLTK, LIWC, S4, GATE, Alchemi API, Natural Language API, Mallet, tm/tidyverse/tidytext etc.			
Vector and Document Spaces Elementary Methods:. Bag(s) of Words, . Ngrams . Document and Language Classification via Vector spaces and the Zipfian Distribution. Dictionary-based approaches			
Vector and Document Spaces Vector Spaces: Term Document / Document Term Matrices, TF-IDF, Word2Vec, Doc2Vec			
Text Understanding and Semantics Topic Modelling:. Latent Dirichlet Allocation. Explicit Semantic Analysis. Latent Semantic Analysis. Hierarchical Dirichlet Process. And associated methods, e.g.. Singular Value Decomposition. Non-negative Matrix Factorisation			
Text Understanding and Semantics Part of Speech Tagging Entity Extraction / Identification, SPARQL and Linked Data, Aspect-based Reasoning			
Knowledge Graphs and Network Analysis Introduction to graph-based models for document corpora, Introduction to network analysis for graph-based models			
Computational Linguistics Interrogating structure, intent, language use independent of content; key use cases:. Affect Analysis. Deception Detection. Psychometric Profiling. Author fingerprinting			
Applied Machine Learning Case Studies in applying (un)supervised machine and/or deep learning to text analytics.			
Assessment Breakdown			%
Coursework			100.00%
Assessments			
Full Time			
Coursework			
Assessment Type:	Continuous Assessment	% of total:	Non-Marked
Assessment Date:	n/a	Outcome addressed:	1,2,3,4
Non-Marked:	Yes		
Assessment Description: Ongoing independent and group problem solving activities and feedback.			
Assessment Type:	Project	% of total:	50
Assessment Date:	n/a	Outcome addressed:	1,2
Non-Marked:	No		
Assessment Description: Students will submit a report (4000 words) on a case study where they will encompass 3 methods covered in the first 6 teaching weeks as outlined in the indicative structure above. The report should discuss the preparation of the corpora for each method, and rationalise the use and effectiveness of each method applied. It should also discuss related work in the area covering the context of the text data as well as studies applied to similar data sets			
Assessment Type:	Project	% of total:	50
Assessment Date:	n/a	Outcome addressed:	3,4
Non-Marked:	No		
Assessment Description: Students will submit a report (4000 words on a case study where they will encompass a further 2 methods from the teaching weeks 7-10 and a further 2 methods not yet included applied in conjunction with a selection of machine learning models: at least 1 unsupervised, and at least 1 supervised. The report should discuss the preparation of the corpora for each method, and rationalise the use and effectiveness of each method applied. It should also discuss related work in the area covering the context of the text data as well as studies applied to similar data sets.			
No End of Module Assessment			
No Workplace Assessment			
Reassessment Requirement			
Repeat examination <i>Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.</i>			
Reassessment Description Should learners not achieve a 40% pass mark, they will either sit a repeat terminal exam, or undertake an assessment that assesses all learning outcomes.			

H8TA: Text Analytics

Module Workload				
Module Target Workload Hours 0 Hours				
Workload: Full Time				
Workload Type	Workload Description	Hours	Frequency	Average Weekly Learner Workload
Lecture	Classroom & Demonstrations (hours)	24	Per Semester	2.00
Tutorial	Other hours (Practical/Tutorial)	24	Per Semester	2.00
Independent Learning	Independent learning (hours)	202	Per Semester	16.83
Total Weekly Contact Hours				4.00

Module Resources	
<i>Recommended Book Resources</i>	
Bird, S., Klien, E. & Loper, E.. (2009), Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit, O'Reily. Goldberg, Y.. (2017), Neural Network Methods in Natural Language Processing (Synthesis Lectures on Human Language Technologies), Morgan & Claypool Publishers. Silge, J.. (2017), Text Mining with R: A Tidy Approach, O'Reily. Rodrigues, M., & Teixeira, A. (2015), Advanced Applications of Natural Language Processing for Performing Information Extraction, Springer.	
<i>Supplementary Book Resources</i>	
Biemann, C. & Mehlner, A.. (2014), Text Mining, Springer. Pennebaker, J.. (2013), The Secret Life of Pronouns: What Our Words Say About Us, Bloomsbury Press. Sankar, D.. (2016), Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data, Apress. Wachsmuth, H.. (2015), Text Analysis Pipelines, Springer.	
<i>This module does not have any article/paper resources</i>	
<i>Other Resources</i>	
[Website], http://words.live [Website], https://liwc.wpengine.com [Website], https://developer.aylien.com/ [Website], https://gate.ac.uk [Website], http://mallet.cs.umass.edu [Website], http://www.nltk.org	
Discussion Note:	