COM5507: SOCIAL MEDIA DATA ACQUISITION AND PROCESSING

Effective Term
Semester B 2024/25

Part I Course Overview

Course Title
Social Media Data Acquisition and Processing

Subject Code
COM - Media and Communication
Course Number
5507

Academic Unit
Media and Communication (COM)

College/School
College of Liberal Arts and Social Sciences (CH)

Course Duration
One Semester

Credit Units
3

Level
P5, P6 - Postgraduate Degree

Medium of Instruction
English

Medium of Assessment
English

Prerequisites
Nil

Precursors
Nil

Equivalent Courses
Nil

Exclusive Courses
Nil

Part II Course Details

Abstract

The course trains students of communication and new media to collect and process content data from social media using computational social science methods, tools, and algorithms. Special emphasis will be placed on web sampling, crawling, storage, and text processing based on a combination of tailor-made tools and open source resources. Through interactive learning sessions including in-class tutorials, individual exercises, group-based projects, etc., the students are expected to become proficient to collect big data from social media for a variety of basic and applied research purposes such as theory-driven studies, data-driven reporting, news visualization, social media user recommendation systems, etc. Issues of policy and research ethics such as privacy protection, data integrity, and open access will also be explored along with technical challenges and solutions.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Demonstrate the capacity for self-directed learning to understand the principles and procedure of collecting and processing social media data.				x
2	Explain the basic methodologies and techniques of social media data collection and processing to recognize the strengths and weaknesses of computational approaches to social media analytics.			x	
3	Interpret numerical and textual data to systematically assess the characteristics and patterns of user generated content and behaviour on social media.		x	x	
4	Value ethical and socially responsible actions in data collection and processing.		x		
5	Demonstrate critical thinking skills in planning and implementing plans for studying social media content.		x	x	x

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

Learning and Teaching Activities (LTAs)

	LTAs	Brief Description	CILO No.	Hours/week (if applicable)
1	Online lectures	Explain key concepts, such as procedures and methods for data collection and processing.	1, 2, 3, 4, 5	1.5 hours/week

2	Tutorial and individual exercises	Students develop and test customized algorithms individually to collect and process social media data.	1, 2, 3, 4, 5	1.5 hours/week
3	Group projects	Students work in teams to collect, process, and analyze social media data and present their findings in data product and an oral presentation.	1, 2, 3, 4, 5	3 hours/week for 5 weeks

Additional Information for LTAs

Online lectures - CILO 2*,3*,5*
Tutorial and individual exercises - CILO 1*,4*
Group projects - CILO 3*

* indirectly

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Class participation	1, 3, 4	30	
2	Tutorials and individual exercises	2, 3, 5	40	
3	Group project and presentation	1, 2, 4, 5	30	

Continuous Assessment (%)

100

Assessment Rubrics (AR)

Assessment Task

Class participation (30%) (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion

Demonstration of self-directed learning to the procedure and methods of social media data collection and processing

Excellent

(A+, A, A-) High

Good

(B+, B, B-) Significant

Fair

(C+, C, C-) Moderate

Marginal

(D) Basic

Failure

(F) Not even reaching marginal levels

Assessment Task

Tutorial and individual exercises (40%) (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion

Ability to replicate the procedure and methods of social media data collection and processing based on given examples

Excellent

(A+, A, A-) High

Good

(B+, B, B-) Significant

Fair

(C+, C, C-) Moderate

Marginal

(D) Basic

Failure

(F) Not even reaching marginal levels

Assessment Task

Group project and presentation (30%) (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion

Ability to demonstrate and explain with technical details, accuracy and clarity, the process and results of collecting and analyzing social media data

Excellent

(A+, A, A-) High

Good

(B+, B, B-) Significant

Fair

(C+, C, C-) Moderate

Marginal

(D) Basic

Failure

(F) Not even reaching marginal levels

Assessment Task

Class participation (30%) (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Demonstration of self-directed learning to the procedure and methods of social media data collection and processing

Excellent

(A+, A, A-) Thoroughly study the materials given before class and actively participate in class discussion

Good

(B+, B) Study the materials given before class and participate in class discussion

Marginal

(B-, C+, C) Attend class discussion

Failure

(F) Fail to learn independently

Assessment Task

Tutorial and individual exercises (40%) (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Ability to replicate the procedure and methods of social media data collection and processing based on given examples

Excellent

(A+, A, A-) Actively participate in the lecture sessions, and fully complete all the tutorial tasks

Good

(B+, B) Attend the lecture sessions, and complete the basic tutorial tasks

Marginal

(B-, C+, C) Attend most of the lecture sessions, and complete most of the tutorial tasks

Failure

(F) Do not attend the lecture sessions, or do not hand in tutorial tasks

---

Assessment Task

Group project and presentation (30%) (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Ability to demonstrate and explain with technical details, accuracy and clarity, the process and results of collecting and analyzing social media data

Excellent

(A+, A, A-) Demonstrate creativity in applying knowledge learnt in class and outside class in the project

Good

(B+, B) Able to apply knowledge learnt in the class to the project

Marginal

(B-, C+, C) Able to deliver a project with some technical elements

Failure

(F) Fail to use the knowledge taught in the course in the project

---

Part III Other Information

Keyword Syllabus

Computational social science, web sampling, web scraping, application programming interface (API), digital traces, open source tools, social media, online social networks, user generated content, web data format, text mining

Reading List

Compulsory Readings

	Title
1	Hanretty, C. (2013). Scraping the web for arts and humanities U of East Anglia. [http://www.essex.ac.uk/ldev/documents/going_digital/scraping_book.pdf]
2	Russell, M. A. (2013). Mining the social web. O' Reilly. [http://shop.oreilly.com/product/0636920030195.do]
3	Fredheim, R. & Zabala, A. (2014) Web scraping using R. [http://quantifyingmemory.blogspot.hk/2014/02/web-scraping-part2-digging-deeper.html]
4	Feinerer, I. (2013). Text mining in R. [http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf]
5	Grun, B., & Hornik, K. (2011). Topicmodels: An R package for fitting topic models. Journal of Statistical Software 40(13), 1-30. [http://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf]

Additional Readings

	Title
1	Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. Annual Review of Sociology, 40(1), 129-152. [http://www.annualreviews.org/doi/abs/10.1146/annurev-soc-071913-043145]
2	Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. Science, 346(6213), 1063. [http://www.sciencemag.org/content/346/6213/1063.summary]
3	Zhu, J. J. H., Mo, Q., Wang, F., & Lu, H. (2011). A Random Digit Search (RDS) Method for Sampling of Blogs and Other User-Generated Content. Social Science Computer Review 29 (3), 327-339. [http://ssc.sagepub.com/content/29/3/327]