City University of Hong Kong
Course Syllabus

offered by Department of Media and Communication
with effect from Semester B, 2015/16

Part I   Course Overview

Course Title:	**Social Media Data Acquisition and Processing**
Course Code:	**COM5507**
Course Duration:	**1 Semester**
Credit Units:	**3**
Level:	**P5**
Medium of Instruction:	**English**
Medium of Assessment:	**English**
Prerequisites: *(Course Code and Title)*	**Nil**
Precursors: *(Course Code and Title)*	**Nil**
Equivalent Courses: *(Course Code and Title)*	**Nil**
Exclusive Courses: *(Course Code and Title)*	**Nil**

Part II   Course Details

1.   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.

2.   Course Intended Learning Outcomes (CILOs)
(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs	Weighting (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	*A2*	*A3*
1.	Demonstrate the capacity for self-directed learning to understand the principles and procedure of collecting and processing social media data.				√
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.			√	
3.	Interpret numerical and textual data to systematically assess the characteristics and patterns of user generated content and behaviour on social media.		√	√	
4.	Value ethical and socially responsible actions in data collection and processing.		√		
5.	Demonstrate critical thinking skills in planning and implementing plans for studying social media content.		√	√	√
		100%			

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 self-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.

3. Teaching and Learning Activities (TLAs)
(TLAs designed to facilitate students' achievement of the CILOs.)

TLA	Brief Description	CILO No.						Hours/week (if applicable)
		1	2	3	4	5		
Lectures and tutorials	Explain key concepts, such as procedures and methods for data collection and processing.	√	√*	√*	√	√*		3 hours/week
Individual exercises	Students develop and test customized algorithms individually to collect and process social media data.	√*	√	√	√*	√		2 hours/week for 8 weeks
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.	√	√	√*	√	√		3 hours/week for 5 weeks

√* indirectly

4. Assessment Tasks/Activities (ATs)
(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.						Weighting	Remarks
	1	2	3	4	5			
Continuous Assessment: 100%								
Class participation and tutorial tasks	√		√	√			30%	
Individual exercises		√	√		√		40%	
Group project and presentation	√	√		√	√		30%	
Examination: ____% (duration: , if applicable)								
							100%	

5. **Assessment Rubrics**
(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Adequate (C+, C, C-)	Marginal (D)	Failure (F)
1. Class participation and tutorial tasks (30%)	Ability to replicate the procedure and methods of social media data collection and processing based on given examples	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Individual exercises (40%)	Capacity for self-directed learning to understand the procedure and methods of social media data collection and processing	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Group project and presentation (30%)	Ability to demonstrate and explain with technical details, accuracy and clarity, the process and results of collecting and analyzing social media data	High	Significant	Moderate	Basic	Not even reaching marginal levels

Part III   Other Information (more details can be provided separately in the teaching plan)

1.   **Keyword Syllabus**
(An indication of the key topics of the course.)
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

2.   **Reading List**
2.1   **Compulsory Readings**
(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

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]

2.2   **Additional Readings**
(Additional references for students to learn to expand their knowledge about the subject.)

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]