

Course Code	DA 2009			
Course Name	Data Collection Methods II			
Credit Value	2C			
Core/Optional	Core			
Prerequisites	None			
Hourly Breakdown	Theory	Practical	Independent Learning	Notional
	20	20	60	100
Course Aim: To provide critical awareness of fundamental concepts and tools relevant to data collection from the web.				
Intended Learning Outcomes: After the successful completion of the course, the student will be able to <ul style="list-style-type: none"> • CLO1: <i>demonstrate</i> awareness of fundamental concepts in data collection from the web • CLO2: <i>identify</i> and <i>apply</i> suitable tools and techniques to extract data from the web • CLO3: <i>practice</i> ethical standards and professional integrity in collecting data from the web 				
Course Content: Introduction to web based data collection; different forms of data available in the web; collect data using python: introduction to python modules, basic HTML syntax, practical examples of data collection, usage of JSON APIs; challenges in web scraping: ethical scraping, dynamic pages, handling captcha; data extraction from pdf: read pdf tables and other information of pdf with proper format, extract data from text/excel/csv file, introduction to OCR and its applications.				
Teaching / Learning Methods: Lectures, Videos, Discussions, Practical Sessions, Group Activities, Presentations, LMS based activities, Independent Learning				
Assessment Strategy:				
Continuous Assessment 60%		Final Assessment 40%		
Details: Quizzes - 20% (CLO 1, 2) Case Studies - 40% (CLO 1, 2, 3)		Theory (%)	Practical (%)	Other (%)
		-	40%	-
Recommended Reading: <ul style="list-style-type: none"> • Mitchell, R. (2018). Web scraping with Python: Collecting more data from the modern web. " O'Reilly Media, Inc." 				