Mining the Social Web

Dr. Mikhail Klassen
Data Scientist, Astrophysicist, Co-Founder and
Chief Technology Officer at Paladin AI
May 21-22, 2019
9:00am – 4:30pm

Summary

The world is awash in data and hidden therein are the patterns that make, connect, and define our world. In response to this, the much-hyped profession of data scientist has emerged in recent years, who use data mining techniques to separate signal from noise.

One of the most valuable sources of data is social media. Over the last decade we have seen many platforms emerge and play pivotal roles in world-shaping events, from the Arab Spring to the US election.

In this workshop, participants will learn some of the basic techniques of data mining, and how they can be applied to social media. We will discuss the ethical uses of data mining, citing examples of how these platforms have been used in both positive and negative ways, and we will talk briefly about some of the efforts that have been made to implement reforms and address data privacy.

No prior programming experience is required, but participants will be learning from code examples written in the Python programming language. Examples will be presented in an accessible way using Jupyter Notebooks, which enable easy experimentation and tinkering. The goal of this workshop is to equip participants with the tools they need to conduct their own research and data mining experiments.

All students are encouraged to bring a laptop to this workshop.

Outline

Tuesday, May 21

Morning	Afternoon
* Welcome / What is Data Mining / Setting up Jupyter Notebook	* Lecture: Introduction to Data Mining
* Lecture: Introduction to Python	* Tutorial
* Tutorial	* Lecture: Mining Twitter Data
* Lecture: Introduction to Working with Data in Python	* Tutorial
* Tutorial	

Wednesday, May 22

Morning	Afternoon
* Lecture: Working with Text Data	* Lecture: Mining Websites
* Tutorial	* Tutorial
* Lecture: Mining Text Files	* Lecture: Data Ethics / Data for Good
* Tutorial	* Tutorial

Resources

Russell, Matthew A. and Mikhail Klassen. 2019. *Mining the Social Web* (3rd edition). O'Reilly Media, Inc. https://learning.oreilly.com/library/view/mining-the-social/9781491973547/

Chapter 1. Mining Twitter: Exploring Trending Topics, Discovering What People Are Talking About, and More

Chapter 5. Mining Text Files: Computing Document Similarity, Extracting Collocations, and More

Chapter 6. Mining Web Pages: Using Natural Language Processing to Understand Human Language, Summarize Blog Posts, and More

ACCESS O'REILLY FOR THE FIRST TIME THROUGH THE FOLLOWING LINK:

http://www.safaribooksonline.com/library/view/temporary-access/ (Note: you must be logged into Concordia library and you will need to enter your Concordia email address ex: @mail.concordia.ca)

GitHub online compendium for Mining the Social Web (3rd edition): https://github.com/mikhailklassen/Mining-the-Social-Web-3rd-Edition

Gryd: https://gryd.us/