2016 가을학기

소셜 네트워크 데이터마이닝과 분석

담당교수 이준환
email: joonhwan@snu.ac.kr
cell: 010-9212-4975
office: 64동 405호

시간 월요일 9:30-12:30

장소 83-601

수업 개요 소셜 컴퓨팅(social computing)과 라지데이터 분석(large data analysis) 등이 커뮤니케이션 분야에서도 중요한 이슈로 부상함에 따라 컴퓨터공학을 전공하지 않은 연구자들도 소셜 네트웍 시스템의 기술적, 구조적 특성을 이해할 필요가 있다. 이 수업에서는 스크립팅 프로그래밍 언어인 파이썬(Python)을 사용하여 컴퓨터 프로그래밍의 기초를 학습하고, 웹 기반 기술(web technology), 데이터베이스 등의 관련 기술에 대한 학습을 통해 실제로 소셜 네트워크 데이터를 수집하고 분석하는 방법을 배운다.

강의내용

1. 파이썬을 이용한 기초 프로그래밍
2. 웹 기반 기술(web technology)
3. 소셜 데이터 마이닝(social data mining)을 통한 데이터 분석

교재 (참고 서적)

- CodeCademy: Learn Python
- CodeCademy: Learn HTML & CSS
- 파이썬 라이브러리를 활용한 데이터 분석 (Python for Data Analysis)
- 기타 필요한 교재/논문 등은 강의 중 제공

수업진행 계획

1주 Introduction to Social Computing
- What is Social Computing?
- What is Computer Programming?
- Why use Python?

2주 Python Crash Course 1
- Python Basics
- Using Github

3주 Python Crash Course 2
- Python Basics

4주 Data Processing
- Data cleaning process

Topic 1: Social Computing Background
- Computer Networks as Social Networks by Wellman
- Computational Social Science by Lazer et al.

5주 Data Analysis Using numpy and pandas 1
- numpy와 pandas를 활용한 데이터의 분석

Topic 2: Socialogical Concepts
- The Strength of Weak Ties by Granovetter
- The Strength of Weak Ties: A Network Theory Revisited by Granovetter

6주 Data Analysis Using numpy and pandas 2
- numpy와 pandas를 활용한 데이터의 분석

Topic 3: SNS & Internet 1
- The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites by Ellison, Steinfield & Lampe
- Social translucence: an approach to designing systems that support social processes by Erickson & Kellogg

7주 Text Data Processing
- NLTK를 활용한 텍스트의 처리
- KoNLPy를 활용한 한글 텍스트의 처리

Topic 4: SNS & Internet 2
- Internet paradox. A social technology that reduces social involvement and psychological well-being? by Kraut et al.
- Internet Paradox Revisited by Kraut et al.

8주 Web Technologies 1
- History of data communication
- History of HTML & CSS
- Basic HTML & CSS

Topic 5: Network
- The Chat Circles Series: Explorations in Designing Abstract Graphical Communication Interfaces by Viégas & Donath
- Inferring relevant social networks from interpersonal communication by De Choudhury et al.
- The Political Blogosphere and the 2004 U.S. Election: Divided They Blog by Adamic & Glance

9주 Web Technologies 2
- Packet communication
- IP, Server/Client
- Web programming frameworks
- Data exchange formats: XML, JSON

Topic 6: Twitter & Facebook
- Twitter mood predicts the stock market by Bollen, Mao & Zeng
- Twitinfo: aggregating and visualizing microblogs for event exploration by Marcus et al.
- Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles by Quercia et al.
- Growing closer on facebook: changes in tie strength through social network site use by Burke & Kraut

10주 Social Data Mining 1
- Crawling data from websites
- Crawling data from complex websites

Topic 7: Crowdsourcing
- The future of crowd work by Kittur et al.
- Soylent: a word processor with a crowd inside by Bernstein et al.

11주 Social Data Mining 2
- Crawling data from Twitter

Topic 8: Human-Computation
- Human computation: a survey and taxonomy of a growing field by Quinn & Bederson
- Labeling images with a computer game by von Ahn & Dabbish

12주 Social Data Mining 3
- Crawling data from Facebook & Instagram
- Using external API for data analysis

Topic 9: Analysis Methods: Case Studies 1
- Predicting tie strength with social media by Gilbert & Karahalios
- Predicting postpartum changes in emotion and behavior via social media by De Choudhury, Counts & Horvitz

13주 Statistical Data Analysis
- 통계적 추론방법의 학습
- ANOVA, Regression의 학습
- 클러스터링 방법론 학습

Topic 10: Analysis Methods: Case Studies 2
- Widespread Worry and the Stock Market by Gilbert & Karahalios
- Experimental evidence of massive-scale emotional contagion through social networks by Kramer at al.

14주 Project idea presentation

15주 Team Meeting

16주 Final Presentation (Final Paper Submit)

과제 및 평가 Assignment 1: 2 CodeCademy course complete (5+5 points)
Assignment 2: Python coding exercise (10 points)
Assignment 3: Data cleanup exercise (10 points)
Assignment 4: Text data processing (10 points)
Assignment 5: Webpage crawling (15 points)
Assignment 6: Twitter crawling (15 points)
Assignment 7: Facebook crawling (or Instagram crawling) (15 points)

Team Project: Project & Paper (40+40 points)
Paper Review: 10 points each
Peer Review: 30 points
if (absent >= 5): fail