Social Media Data

a new trend of corpus-based research?

Outline

- What is social media
- Why it is important
- The use of social media
- Collecting data from social media
- Big data
- Research for linguistics

What is social media

- What is new media? (http://en.wikipedia.org/wiki/New_media)
- media : analog => digital
- New Media as Computer Technology Used as a Distribution Platform
- New Media as Digital Data Controlled by Software
- “new media” are digital, often having characteristics of being manipulated, networkable, dense, compressible, and interactive

What is social media

- “a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.”
- Social media technologies take on many different forms including magazines, Internet forums, weblogs, social blogs, microblogging, wikis, social networks, podcasts, photographs or pictures, video, rating and social bookmarking
- six different types of social media: collaborative projects (for example, Wikipedia), blogs and microblogs (for example, Twitter), content communities (for example, YouTube and DailyMotion), social networking sites (for example, Facebook), virtual game worlds (e.g., World of Warcraft), and virtual social worlds (e.g., Second Life)
SOCIAL MEDIA 2013: STATISTICS AND TRENDS
(https://www.youtube.com/watch?v=5yxuljH1X09l)

Why it is important

- business: transform the marketing industry
- society: forum for support and safety information during a crisis
- political power: give a true voice to the people.
- globalization: easy communication worldwide

Social media revolution
(https://www.youtube.com/watch?v=x0EnhXnlboM)

Socialnomics: How Social Media Transforms the Way We Live and Do Business. Erik Qualman
Twitter

- Twitter is an online social networking service and microblogging service that enables its users to send and read text-based messages of up to 140 characters, known as "tweets".
- The service rapidly gained worldwide popularity, with over 500 million registered users as of 2012, generating over 140 million tweets daily and handling over 1.6 billion search queries per day.
- @username – Reply. Tweets that start with an @ and then a username is a reply to that user.
- #tag – Hashtag. Hash that is immediately followed by a tag (hashtag) in Twitter is a community-driven convention for adding additional context and metadata to your tweets.

The use of social media

- Using Twitter to Track the Flu: Researchers Find a Better Way to Screen the Tweets. ([http://releases.jhu.edu/2013/01/24/using-twitter-to-track-the-flu/](http://releases.jhu.edu/2013/01/24/using-twitter-to-track-the-flu/))
- Using Twitter to Map and Measure Online Cultural Diffusion ([http://www.oii.ox.ac.uk/research/projects/?id=89](http://www.oii.ox.ac.uk/research/projects/?id=89))
- Twitter for Historical Research? e.g. 2008 US President election
- Business: ask opinions, survey opinions, branding, customer relations, etc.

Researchers Create “Hate Map” of the U.S. With Twitter Data

Candian researchers have created a “hate map” that splashes a U.S. map with coloring showing the proportion of all Tweets that contain common hateful words about race, sexual orientation, or disabilities.
Collecting twitter data

- download ข้อมูล tweet ตัวเองได้จาก setting
- ข้อมูลคนอื่นๆ ได้จากการ search แล้วบันทึกไว้
- สามารถสร้าง widget ติด js script ไว้ในเว็บตัวเองได้ สำหรับ ‘ฝังการค้นหา’ แล้ว copy-paste code ไปใส่ในหน้าเว็บที่ต้องการแสดง
- สามารถช่วยผ่าน API : RESTful API or Streaming API ต้องการ Authentication (tutorial เก่าๆ ใช้ไม่ได้กับ API 1.0)
- ตัวอย่างการใช้ tweepy คือโปรแกรม Python (http://nbviewer.jupyter.org/url/raw.githubusercontent.com/raynach/hse-twitter/master/docs/Collecting%20Twitter%20data%20from%20the%20API%20with%20Python.ipynb)
- ข้อมูลเป็น JSON จะต้องนำมาแปลงเป็น csv อีกกระหน่ำ

What is Data Visualization
(https://www.youtube.com/watch?v=YaGqOPxHfYk)

Big Data Analytics: Answers from Big Data
(https://www.youtube.com/watch?v=LrNlZ7-SMFk)

The Art of Data Visualization
(https://www.youtube.com/watch?v=AdSZJzb-aX8)
Big data & Visualization

- Big data (http://en.wikipedia.org/wiki/Big_data)
- Analyzing Big Data with Twitter: Course at UC Berkeley (http://blogs.ischool.berkeley.edu/1290-abdi-t-s12/)
- Data Visualization (http://en.wikipedia.org/wiki/Data_visualization)
- Many Eyes (http://www-958.ibm.com/software/analytics/manyeyes/)
- Visualisation of text corpora: A case study of the PCEEC (http://www.helsinki.fi/vareng/series/volumes/07/sirtola_et_al/)

Research for linguistics

- Analysis of opinion on twitter
  - "Identifying Sarcasm in Twitter: A Closer Look"
  - "Mining Sentiments from Tweets"
  - "Sentiment Analysis of Twitter Data"
  - "Twitterology: A New Science?"
- Languages on twitter
  - language variation: male/female, demographic, discourse
  - "Using Twitter for Demographic and Social Science Research: Tools for Data Collection"
  - "Linguistic markers of secrets and sensitive self-disclosure in Twitter"
  - "Twitter Register Variation"
- Languages on twitter
  - "Discourse of Twitter and Social Media How We Use Language to Create Affiliation on the Web."
  - "Mark my words! Linguistic style accommodation in social media"
  - "What Kind of #Conversation is Twitter? Mining #Psycholinguistic Cues for Emergency Coordination"
  - "Geo-linguistic fingerprint and the evolution of languages in Twitter"
  - "Cambridge University Linguists Use Twitter To Study How Welsh Language Use Is Changing"
- research on other fields
  - Bibliography of Research on Twitter & Microblogging (http://www.danah.org/researchBibs/twitter.php)