

Welcome to INFO319: Research Topics in Big Data

Vimala Nunavath

<Vimala.nunavath@uia.no>

What are we going to do today?

- Some practical information
- Expectations for the course
- **About the themes**
 - what are we going to focus on?
- About the programming projects
 - how about a coordinated task?
- **About the essay**
- **Introduction to Emergency Management**
- Plans for the next session?

Themes:

- Possible themes:
 - Introduction to Emergency management
 - Introduction to Big data
 - Big Data sources for EM
 - Sensors/IOT for EM
 - Social media for EM
 - NLP for EM and visualization/Dashboards
 - Machine learning for EM

Sessions:

- 9 sessions
 - Thursdays and Fridays from 1015
 - seminar room 548 (SV:S548) in this building
 - also a BDEM workshop with international partners
 - in October 2018
 - the sessions will become highly interactive
 - you present, critique and discuss
 - chapters and papers
 - your essays and projects
 - 80% attendance is required
 - that is: 80% active participation

Prerequisites:

- Need some programming skills.
- Formal requirements:
 - active participation in seminars (80%)
 - graded individual essay (30%)
 - graded group programming project (40%)
- Final exam (30%).

- Grades are given as follows:
 - *Written exam (30%)*: An individual
 - Theoretical essay with thoughtful research and discussion of an assigned topic (30%).
 - In addition, there is a practical assignment in groups (40%).
- Both assignments must be done in the teaching semester.
- Dates:
 - Submission of group assignment: 27.11.2018
 - Submission of individual essay: 5.12.2018
 - Exam date: December 17th 2018.

Curriculum:

- Textbooks:
 - Rob Kitchin (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences*. Sage, 208 pages.
 - Carlos Castillo (2016). *Big Crisis Data – Social Media in Disasters and Time-Critical Situations*. Cambridge, 212 pages.
- Research papers in the wiki
 - will appear in wiki.uib.no/info319 – possibly quite a few!
 - we will make a selection of core paper
- mitt.uib.no for announcements / discussion
- More information about the course reading list can be found on:
<https://wiki.uib.no/info319/index.php/Sessions>

The Themes

Central themes

- Central themes:
 - Introduction to Emergency management
 - Introduction to Big data
 - Big Data sources for EM
 - Sensors/IOT for EM
 - Social media for EM
 - Natural language processing (NLP) for EM and visualization/Dashboards
 - Machine learning for EM

Introduction to Emergency management

- Lots of emergencies such as natural and technological and man-made emergencies have been happening all around the world, every day and even every single hour.
- What is emergency/disaster.
- What are the phases of EM.
- Who are involved.
- What are the Information needs of stakeholders.
- What are the challenges and obstacles for EM.
- Which new tools and techniques are needed for EM?

Introduction to Big data

- Big (and open) data are becoming enormously important
 - making sense of them is a major challenge for EM.
- What is big data?
- What are existing big data sources?
- What are the dangers of big data?
- Privacy, surveillance, societal impact of big data;
- What are big data challenges?
- Which new tools and techniques are needed?

Emergency Big Data sources for EM

- Big (and open) data are becoming enormously important...
- What are the available EM big data sources?
- How to deal with EM big data sources for making sense of it?
 - semantics can help
- What are the characteristics of big data sources?
- What technical architectures for big data management are available and needed?

Sensors/IOT for EM

- Lots of sensors, actuators and other (smart and dumb) things are on the net.
 - closely related to big data as a source.
 - examples: Radio-frequency identification (RFID), sensor web, wireless sensor network etc..
- What are the characteristics of the IoT?
- Which research efforts target the use of IoT data for EM?
- Which new developments are needed?



Social media for EM

- Social media have played an increasing role in emergencies and disasters. It has become a central source of big data for EM.
- How to collect (process) social media data and present it to needed stakeholder during EM?
- How are social media streams being analyzed?
- How can the analysis results be organized as social network graphs?
- How can the graphs be used, e.g., to detect events and provide information?

NLP for EM and visualization/Dashboards

- Social networks offer a wealth of information for capturing additional information on people's behavior, trends, opinions and emotions during disasters.
- What is NLP? Why sentiment analysis?
- Why it is important?
- Who are beneficiaries?
- How to train a sentiment classifier to categorize messages, which was then used to examine the level of concern and anxiety in the people before and after the disaster?
- How to visualize the processed data on dashboard??

Machine learning for EM

- Knowledge building from large data sets
 - available data are used to automatically build predictions, models, explanations(?), theories...
 - machine learning techniques.
- Collect data and test them.

Which themes should we focus on?

- Go to docs.google.com

(<https://docs.google.com/forms/d/1UaOM10WwxLv8lClpJEL65-hhQ7ha2xodEuVZku4ctzg/edit>)

- What themes do you like (so far)?
- What type of course do you want?
- And how much do you know about each theme?

The Essays

Individual essays

- The essay shall present and discuss selected theory, technology and tools related to semantic technologies, backed by scholarly and other references
 - counts 30% of final grade
 - presentations: November 22nd
 - deadline: December 5th 1400
- Encouraged:
 - more than a paper
 - social media contrib's (Wikipedia, Wikidata...)

The student group Programming Project

Group programming project

- The project shall develop an application that can be used for emergency management. Development and run-time platform is free choice, as is programming language. The project should be carried out in groups of three and not more. Working individually or in pairs is not recommended.
- Counts 40% of final grade.
- Final presentation: Friday November 23rd
- Submission deadline: Tuesday November 27th, 1400

Project Topics

- 1) Integrating different social media applications for providing information awareness either to victims/first responders.
- 2) Developing an application to detect emerging hot topics over the social media before, during and after emergency management.
- 3) Social media data analytics for disaster management.
- 4) Deep Learning for emergency management Using Social Media Information.
- 5) Analysis of Post-disaster Twitter Communication: A case study of....
- 6) Develop an information visualization tool to identify the disaster risk.

Integrating different social media data sources for Information awareness.

- Developing an application for integrating different social media data sources for information awareness.

Developing an application to detect emerging hot topics over the social media before, during and after emergency management.

- Developing an application for detecting the disaster trends in various regions.

Social media data analytics for emergency management

- Develop a system capable of processing, analyzing, and extracting useful information from Twitter during an emergency to understand the public sentiment.
e.g. Usecase: Kerala floods 2018.

Deep Learning for emergency management Using Social Media Information.

- Classification of social media information by using different machine learning algorithms.

Analysis of Post-disaster Twitter Communication: A case study

- Develop an application to understand/classify the communication tasks that are performed by using Twitter during post-disaster phase.

Information visualization tool to identify the disaster risk.

- Develop an information visualization tool to identify the disaster risk.

Should we find a coordinated project task?