ECIS 2017
25TH EUROPEAN CONFERENCE ON INFORMATION SYSTEMS
5-10 JUNE / GUIMARÃES, PORTUGAL

Workshops

Workshop description

Title of the workshop

Similarity Detection in Digital Trace Data

Main Contact Details

Mahmood Shafeie Zargar
Assistant Professor
Vrije Universiteit Amsterdam
Netherlands
m.shafeiezargar@vu.nl

Organizers

Mahmood Shafeie Zargar
Assistant Professor
Vrije Universiteit Amsterdam
Netherlands
m.shafeiezargar@vu.nl

Harris Kyriakou
Assistant Professor
IESE Business School
Spain
hkyriakou@iese.edu

Yegin Genc
Assistant Professor
Pace University
USA
ygenc@pace.edu

Theme of the workshop

The increasing availability of digital trace data and recent advances on both methods and tools to analyze data in unprecedented scale opens up new venues for information systems research. These trends permit the implementation of novel analysis methods and open up new arenas for theorization. Yet, the sheer volume of data means that the traditional research methods, whether qualitative or quantitative, may not be able to fully benefit from the depth and the breadth of the data available. That is unless the data is made accessible through intermediary transformation steps. Traditionally this accessibility has been achieved by limiting the scale of the data through

sampling strategies. But advances in computational methods have made other strategies possible that may provide additional or holistic insights. Among others, these methods include dimensionality reduction methods used to "summarize" the data by focusing on its salient features and the methods capable of detecting patterns and similarities at a large scale. The combination of these computationally-powered tools with traditional research methods open new possibilities for both inductive and deductive research.

This workshop focuses on similarity detection methods. An interactive workshop will be held where the participants will have the chance to use similarity detection methods on precollected data. The workshop will include presentations and interactive sessions focusing on the following methods:

1. **Fuzzy hashing for text similarity detection (Mahmood Zargar).**
2. **Spherical harmonic representation of 3D shapes for product design comparison (Harris Kyriakou).**
3. **Semantic transformations with ontology extraction using Wikipedia (Yegin Genc).**

Format

Hands-on workshop

Suitability to ECIS audience

The need to take stock of extant methods and emerging practices on collecting, handling, cleaning and analysing digital data has been identified in the IS community. This workshop will explore some of the latest advances in computational methods that can be nicely coupled with the commonly used qualitative and quantitative IS methods. Participants will be trained on these methods and have the chance to see practical examples of such methods. In turn, they will gain insights on potential applications of such methods in their research.

Program committee members

Mahmood Shafeie Zargar
Assistant Professor
Vrije Universiteit Amsterdam
Netherlands
m.shafeiezargar@vu.nl

Harris Kyriakou
Assistant Professor
IESE Business School
Spain
hkyriakou@iese.edu

Yegin Genc
Assistant Professor
Pace University
USA
ygenc@pace.edu

Plan for publications

The organizers will create and maintain a website for the workshop. Participants will submit a brief description of their current research work and how they think that this workshop can help them with their research. The workshop will provide an opportunity for participants to network and seek opportunities for future collaboration. Organizers are considering scheduling a special issue in an information systems journal. Then, participants will be able to submit their work utilizing methods focusing on similarity detection at a later date. Organizers will also be seeking collaboration opportunities with participants to publish in IS journals. The schedule as well as targeted journals will be finalized after discussion with the workshop participants.

Target audience and expected attendance

The organizers have already approached and will continue targeting researchers studying online communities, social media, peer production, crowds, open source, and any other digital platform generating publicly accessible data.

Planned acceptance rate

N/A

Workshop schedule

Deadline to Submit a Proposal to Present in the Interactive Workshop:
Notification of Acceptance for Presenters:
Deadline to Register as Participant at the Workshop:
Notification of Acceptance for Participants:
Workshop Day: June 6th.

09:00-09:15 Introduction & Overview of the Workshop
09:15-10:00 Showcase 1: Similarity hashing on programming code
10:00-10:15 Break and Discussion about Showcase 1
10:15-11:00 Showcase 2: Shape comparison on representations of 3D printed product designs
11:00-11:15 Break and Discussion about Showcase 2
11:15-12:00 Showcase 3: Ontology extraction using Wikipedia
12:00-12:15 Break and Discussion about Showcase 3
12:15-12:30 Feedback & Concluding Remarks

Special equipment needed

The organizers will need a projector and internet access. The participants are expected to bring their laptops in order to make the most out of the workshop.

Prior experience of the proposers in workshops organisation

Mahmood Zargar has organized and taught similar method workshops at McGill University. His workshops attracted the interest of many doctoral students, and professors. Yegin Genc chaired the Business Analytics Challenge at Manhattan College and organized method workshops at Stevens Institute of Technology. In addition, Yegin Genc organized two international Data Analytics Competition workshops for IEEE's Big Data Analytics Initiative.