ICDAR'20: Intelligent Cross-Data Analysis and Retrieval

Minh-Son Dao
dao@nict.go.jp
National Institute of Information and Communications Technology
Tokyo, Japan

Morten Fjeld
Bergen University
Bergen, Norway
morten.fjeld@uib.no

Filip Biljecki
National University of Singapore
Singapore
filip@nus.edu.sg

Uraz Yavanoglu
Gazi University
Ankara, Turkey
uraz@gazi.edu.tr

Mianxiong Dong
Muroran Institute of Technology
Muroran, Japan
mx.dong@csse.muroran-it.ac.jp

ABSTRACT

The First International Workshop on "Intelligence Cross-Data Analytics and Retrieval" (ICDAR'20) welcomes any theoretical and practical works on intelligence cross-data analytics and retrieval to bring the smart-sustainable society to human beings. We have witnessed the era of big data where almost any event that happens is recorded and stored either distributedly or centrally. The utmost requirement here is that data came from different sources, and various domains must be harmonically analyzed to get their insights immediately towards giving the ability to be retrieved thoroughly. These emerging requirements lead to the need for interdisciplinary and multidisciplinary contributions that address different aspects of the problem, such as data collection, storage, protection, processing, and transmission, as well as knowledge discovery, retrieval, and security and privacy. Hence, the goal of the workshop is to attract researchers and experts in the areas of multimedia information retrieval, machine learning, AI, data science, event-based processing and analysis, multimodal multimedia content analysis, lifelog data analysis, urban computing, environmental science, atmospheric science, and security and privacy to tackle the issues as mentioned earlier.

The actual ICDAR'20 Proceedings are available in the ACM DL at: https://dl.acm.org/citation.cfm?id=3379174

CCS CONCEPTS

• **Computer systems organization → Embedded and cyber-physical systems**; • **Security and privacy → Human and societal aspects of security and privacy**; **Systems security**; • **Information systems → Information retrieval**; • **General and reference → Cross-computing tools and techniques**;

KEYWORDS

cross-data; data analytic; information retrieval; security and privacy; artificial intelligence; knowledge discovery

ICMR '20, June 8–11, 2020, Dublin, Ireland

ACM ISBN 978-1-4503-7087-5/20/06.
https://doi.org/10.1145/3372278.3388041

ACM Reference Format:
Minh-Son Dao, Morten Fjeld, Filip Biljecki, Uraz Yavanoglu, and Mianxiong Dong. 2020. ICDAR'20: Intelligent Cross-Data Analysis and Retrieval. In *Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20), June 8–11, 2020, Dublin, Ireland*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3372278.3388041

1 INTRODUCTION

Currently, people can collect data from themselves and their surrounding environment quickly due to the exponential development of sensors and communication technologies and social networks. The ability to collect such data opens the new opportunity to better understand the association between human beings and the properties of the surrounding environment. These associations can be utilized for intelligence, planning, controlling, retrieval, and decision making efficiently and effectively by governments, industries, and citizens. Wearable sensors, lifelog cameras, and social networks can report people's health, activities, and behaviors by the first-view perspective. In contrast, surrounding sensors, social networks interaction, and third-party data can give the third-view perspective of how their society activities look like. Several investigations have been done to deal with each perspective, but few investigations focus on how to analyze and retrieve cross-data come from different perspectives to bring better benefits to human beings. The target of the workshop is to attract researchers to work on the intelligent cross-data analysis and retrieval to bring the smart-sustainable society to human beings.

2 SCOPE

The domain of the research can vary from wellbeing, disaster prevention mitigation, mobility, to food computing, to name a few. Example topics of interest include but is not limited to the following

- Event-based cross-data retrieval
- Data mining and AI technology to discover and predict spatial-temporal-semantic correlations between cross-data.
- Complex event processing for linking sensors data from individuals, regions, to broad areas dynamically.
- Transfer Learning from one region to another region using locally-collected data to construct or customize similar analysis and prediction of events effectively and efficiently.
- Realization of a prosperous and independent region in which people and nature coexist.

- Applications leverage intelligent cross-data analysis for a particular domain.
- Hypotheses Development of the associations within the heterogeneous data contributes towards building good multimodal models that make it possible to understand the impact of the surrounding environment on human beings at the local and individual scale.
- Cross-datasets for Repeatable Experimentation.
- Security and Privacy for both data and systems of cross-data platforms.
- Sensing techniques that can help to monitor people passively without interfering directly to their life.

3 OBJECTIVE

The goal of the workshop is to attract researchers and experts in the areas of multimedia information retrieval, machine learning, AI, data science, event-based processing and analysis, multimodal multimedia content analysis, lifelog data analysis, urban computing, environmental science, and atmospheric science to tackle the intelligent cross-data analysis and retrieval issue.

4 COMMITTEE MEMBERS

We are very grateful to each of our Programme Committee members for their continued willingness to provide reviews of submissions. Each submission received at least three reviews.

- **Steering Committee**
  - Minh-Son Dao (National Institute of Information and Communications Technology, Japan)
  - Morten Fjeld (Bergen University, Norway)
- **Program Chairs**
  - Mianxiang Dong (Muroran Institute of Technology, Japan)
  - Filip Biljecki (National University of Singapore)
  - Uraz Yavanoglu (Gazi University, Turkey)
- **Program Committee**
  - Nitta Naoka (Osaka University, Japan)
  - Nicola Conci (Trento University, Italy)
  - Anindya Nag (Dongguan University of Technology, China)
  - Zhihan Lv (Qingdao University, China)
  - Xiaoyi Yu (Peking University, China)
  - Ankur Bist (Govind Ballabh Pant University of Agriculture Technology, India)
  - Thanh-Binh Nguyen (University of Science at HCM City, Vietnam)
  - Hung Tran (Fenikaa University, Vietnam)

5 INVITED SPEAKER

We have the honor to welcome Koji Zettsu as our invited speaker with the exciting talk of "*From Data Collection Merit to Data Connection Merit for Smart Sustainable Cities*" that gives an overview of how the cross-data platform can bring the sustainable for smart cities.

Koji Zettsu is a Director General of Big Data Integration Research Center of National Institute of Information and Communications Technology (NICT). He has been doing research and development of data analytics technology in NICT, and now leading Real Space Information Analytics Project since 2016 to implement smart data platform based on data mining and AI. For promoting industry-academia-government collaboration on the platform, he is also a leader of Cross-Data Collaboration Project of Smart IoT Acceleration Forum in Japan. Dr. Zettsu worked for IBM Japan (1992-2003). He was a Vice Chair of IEICE Data Engineering Technical Group (2011). His research interests are databases, data mining, information retrieval and software engineering. He has serviced on numerous academic societies, conference committees and working groups.

6 ACCEPTED PAPERS

We selected six original works (over seventeen submissions) whose content reflects different aspects of the intelligence cross-data analytic and retrieval. These researches can be considered as a piece of puzzles to build a big picture of intelligence cross-data analytic and retrieval.

The work introduced in "*Duplicate Identification Algorithms in SaaS Platforms*" written by **Nguyen et al.** first analyzes the common errors created by users in a SaaS application and then utilizes the dataset provided to study a character-based triplet Siamese model to act as a deep learning metric, which computes the similarity between two given addresses.

The paper "*Malware detection using system logs*" presented by **Nguyen et al.** proposes a new method for malware detecting using system logs. This approach can have multiple applications, especially for constructing an autonomous system to detect malware in IoT systems.

Nguyen-Tai et al. wrote the paper "*MNR-HCM Data: A Personal Lifelog and Surrounding Environment Dataset in Ho-Chi-Minh City, Viet Nam*" that introduced the data collection system (both hardware and software) for cross-data collection activities focusing on supporting to understand the association between air pollution, urban nature, and human health. This dataset and system can be reproduced for a bigger scale and different domains.

Qiu et al. introduce a paper "*Residence and Workplace Recovery: User Privacy Risk in Mobility Data*" that discusses the user privacy risk in mobility data. They prove that even encryption is used to protect data, with large-scale factual mobile data and long-time tracking data captured from a big city, the identification of residence and workplace of user can be revealed.

The paper written by **Nguyen and Tran** "*Microwave Doppler radar sensing system for vital sign detection - From Evaluated Accuracy models to the Intelligent system*" presents the research about the application of radar sensing systems in detecting vital signs (breathing rate) of humans. The authors study the mathematical model to evaluate the accuracy of the system when it operates in different media. Besides, they introduce a combination of the radar sensing system and AI techniques to make this system smarter. This system has a high potential in predicting breathing/sleeping disorders in humans.

The last paper "*A Digital Insight Provider From Financial Documents In Banking*" discusses a structured information retrieval and data analysis system that automatically and recursively digitalizes documents of the bank's customers to provides new sales opportunities. **Baydar et al.** give a different perspective on how to utilize analyze and retrieve data in a cross-domain.