2023 Master internship at University of Cádiz

TITLE : INTELLICITIZEN. Building a framework for identifying problems in the city through social sensing

LAB & PEOPLE

- Name of the hosting lab: SPI&FM (Software Process Improvement and Formal Methods)
- General activities of the lab : Process and service engineering, Creative computing and learning assessment
- Website: https://produccioncientifica.uca.es/grupos/7918/detalle
- Number of staff / PhD: 17/ 15
- Supervisor name and contact: Andrés Muñoz Ortega (andres.munoz@uca.es)

TOPIC OF THE INTERNSHIP

- **Scientific context of the internship (max 20 lines)** : The research for the internship proposes the use of knowledge-based techniques to advance the boundaries of social sensing (or collective intelligence), where aggregating information from a group of people allows decisions to be made that are often better informed than the decision made by disaggregated information (Li et al., 2017). Citizens, through their actions in the virtual world, become social sensors that provide data about a given event. As a result, a large amount of information can nowadays be collected through social media platforms and applications. This information can be used by public institutions, companies or citizens themselves in different application domains (e.g. environment, health or transport) to know the status of situations or events of interest and to make decisions based on this knowledge.

    According to recent research (Xu et al. (2018), Wang et al. (2019) and Ghahramani et al. (2020)), collective intelligence techniques involve a number of challenges that have not yet been solved, such as (i) extracting fine-grained information from data obtained from social networks, (ii) aggregating and fusing data about the same event or situation from different publications, and (iii) obtaining explanatory reasoning about the causes of an event. This research aims to address these three challenges in a novel way using knowledge-based techniques. For the first challenge, the development of a complete multimodal analysis of the content published on social networks is proposed, as it is indeed increasingly common to find different types of information (e.g. text, audio, image) in the same post and, therefore, it is necessary to extract the most relevant semantic features of each type. The second challenge is addressed by adapting traditional semantic information fusion techniques to a collective intelligence data

scenario. This challenge is mainly due to the enormous amount of information that exists to combine and increase the knowledge about the context of the event of interest. Therefore, the efficient fusion of multimodal information obtained through collective intelligence opens a new line of research in semantic knowledge representation techniques. The third challenge lies in obtaining human-understandable descriptions of the causes of the detected event. To this end, the use of inference techniques of semantic reasoners and visualisation techniques will be studied.

- **Keywords** : Data fusion, Social Sensing, Crowdsensing, Text Analysis, Image Analysis, Machine Learning, Semantic Web

- **Tasks and duties entrusted to the student**: The student will be entrusted to study and use existing state-of-the-art techniques in the areas of Natural Language Processing (NLP) and Machine Learning/Deep Learning for the processing of multimedia data obtained from social networks and other open data sources. The data obtained from these techniques will be semantically characterised through ontology-based knowledge graphs for their subsequent fusion. Regarding such fusion techniques, they will be addressed through high-level fusion techniques (Hall & Llinas, 2017) to solve the problems that will arise in the information combination process, such as inconsistency management (e.g. conflicting/contradictory data about the same fact), lack of completeness (e.g. lack of context information, such as the location), incorrectness (e.g. typographical errors) or outdated data (e.g. data taken at different time periods referring to the same event), to name but a few (Achichi et al., 2019; Fischer et al., 2021).

- **Skills to be acquired or developed:** Techniques for data fusion, techniques for analysing text in social media, techniques for analysing video in social media, techniques for inferring relevant information from the data fusion of these elements

Bibliography

Achichi, M., Bellahsene, Z., Ellefi, M. B., & Todorov, K.(2019). Linking and disambiguating entities across heterogeneous RDF graphs. Journal of Web Semantics, 55, 108-121.

Fischer, Lukas, et al. (2021). AI System Engineering—Key Challenges and Lessons Learned. Machine Learning and Knowledge Extraction, 3(1), 56-83.

Ghahramani, M., Zhou, M., & Wang, G. (2020). Urban sensing based on mobile phone data: approaches, applications, and challenges. IEEE/CAA Journal of Automatica Sinica, 7(3), 627-637. Hall, D. L., & Llinas, J. (2017). Multisensor Data Fusion. Handbook Multisensor Data Fusion: Theory and Practice.

Li, W., Wu, W., Wang, H., Cheng, X., Zhou, Z. & Ding, R.(2017) Crowd intelligence in AI 2.0 era. Frontiers of Information Technology & Electronic Engineering 18, 15-43.

Lou, W., Pi, R., Wang, H., & Ju, Y. (2020). Low-cost similarity calculation on ontology fusion in knowledge bases. Journal of Information Science 46 (6), 823-836.

Osman, I., Yahia, S. B., & Diallo, G. (2021). Ontology integration: Approaches and challenging issues. Information Fusion, 71, 38-63.

Wang, D., Szymanski, B.K., Abdelzaher, T., Ji, H. & Kaplan, L. (2019) The age of social sensing. IEEE Computer 52 (1), 36-45.

Xu Z., Mei L., Choo K.K.R., Lv Z., Hu C., Luo X., Liu Y. (2018) Mobile crowd sensing of human-like intelligence using social sensors: A survey. Neurocomputing 279:3–10.ç

PROFILE OF THE DESIRED STUDENT

- Minimum level of study required: Level 6 - Bachelor's Degree

- Field(s) of study: Computer Science / Data Science

- Scientific skills : Data analysis, NLP techniques, Image analysis, Technical writing, Programming skills (preferably in Python)

- Language skills required : English required (B2), Spanish optional

THE INTERNSHIP ASSIGNMENT:

Desired duration of the internship (in months): 9

Desired Starting date of the mission: *01/01/2024*

Indicative weekly schedule: *25h / week*

Remuneration : This internship will be funded by the Erasmus+ internships programme.

Internship agreement: *an internship agreement will be signed.*

To SEA-EU students:

If you're interested please send your CV and letter of motivation to the scientist in charge, andres.munoz@uca.es before the date 30/09/2023.