H.F.R.I.
Hellenic Foundation for Research & Innovation

Description of the funded research project

1st Call for H.F.R.I. Research Projects to Support Faculty Members & Researchers and Procure High-Value Research Equipment

Title of the research project: **CHOROLOGOS:**
Semantic Spatio-textual Data Analysis and Processing

Principal Investigator: **Christos Doulkeridis**

Reader-friendly title: **CHOROLOGOS**

Scientific Area: **Mathematics & Computer Science**

Institution and Country: **University of Piraeus, Greece**

Host Institution: **University of Piraeus**

Collaborating Institution(s): **Norwegian University of Science and Technology (NTNU)**

Project webpage (if applicable): www.ds.unipi.gr/chorologos/

Budget: **171.000 €**

Duration: **01/12/2019 – 30/11/2021**

H.F.R.I.
Hellenic Foundation for Research & Innovation

With the widespread adoption of web-based services, mobile smartphones equipped with GPS capabilities, the Internet of Things (IoT), and social networks, an ever-increasing wealth of geotagged data is available for use on a daily basis. However, existing approaches for spatio-textual query processing largely rely on exact or syntactical matching techniques, which has a negative impact on the quality of results as well as on the expressiveness of query formulation. The proposed project, named CHOROLOGOS, aims at advancing the state-of-the-art in spatio-temporal-textual query processing, by introducing a novel framework that tightly combines spatio-textual and spatio-temporal querying with semantic retrieval, focusing on expressive query formulation beyond syntactical matching and towards similarity-based, pattern-based, and eventually semantic retrieval. Core research objectives of this framework include definitions of novel query types that incorporate semantic matching with advanced spatio-temporal constraints, effective indexing structures tailored for the joint organization spatio-temporal-textual data, novel filtering techniques that eagerly prune the search space, efficient query processing algorithms that capitalize on the available access methods, and scalable analysis of massive spatio-textual data by means of parallel processing. The application domains targeted by CHOROLOGOS include: (a) processing and location-based analysis of social data, such as geotagged tweets, and (b) enriched trajectory data of moving objects. In this way, CHOROLOGOS will support applications and services targeting the mobile tourist, by providing flexible and expressive retrieval of sets of points of interest in combination with complex spatial, temporal and textual constraints.

The originality of this research proposal aims to introduce a semantic spatio-textual processing framework for supporting novel query types over spatio-textual or spatio-temporal-textual data. More concretely, the research objectives of CHOROLOGOS include:

- **Formulation of expressive query types that enable selection of underlying spatio-temporal-textual data based on diverse information needs, going beyond exact or syntactical matching and towards semantic retrieval. Examples of such queries include similarity matching, pattern-based matching, as well as semantic similarity matching.**
- **Theoretical contributions in terms of properties and search bounds for the proposed query types, thus laying the foundations for efficient processing and search.**
- **Design of appropriate access methods that jointly index space, time, and text, in an appropriate way to support filtering of data that is irrelevant to the query at hand.**
- **Efficient query processing algorithms following well-established methodologies, including filter-and-refine and branch-and-bound, aiming at fast delivery of accurate query results.**
- **Parallel processing of the proposed query types, towards scalable algorithms that make the analysis of vast-sized data sets feasible in practice.**

Expected Results

CHOROLOGOS aims at advancing the state-of-the-art in spatio-temporal-textual query processing, by introducing a novel framework that tightly combines spatio-textual and spatio-temporal querying with semantic retrieval, focusing on expressive query formulation beyond syntactical matching, efficient indexing and query processing, and scalable analysis of massive spatio-textual data.

Impact

CHOROLOGOS promises to move the research frontier a step forward in the area of semantic spatio-textual data management. Effective and efficient retrieval of spatio-temporal-textual data is a challenging topic, which has attracted considerable attention recently, not only from the academia, but also from the industry. Search engines (such as Google, Yahoo and Bing) and social network providers (Twitter, Foursquare, etc.) either collect or own vast-sized spatio-textual data sets, and conduct research in new methods and technologies for advanced analytics, in order to provide personalized recommendations, targeted marketing, etc. By exploiting CHOROLOGOS the analysis of massive spatio-textual datasets, typically encountered in the aforementioned domains and especially in social networks, is going to be facilitated significantly.

The HFRI funding is very important as it enables the implementation of the proposed research, it supports a PhD student and a postdoctoral researcher, and it allows us to maintain and strengthen the research collaboration with a foreign university in Norway.

H.F.R.I.
Hellenic Foundation for Research & Innovation

COMMUNICATION

185 Syggrou Ave. & 2 Sardeon St. 2
171 21, N. Smyrni, Greece
+30 210 64 12 410, 420
communication@elidek.gr
www.elidek.gr