C H O R O L O G O S
N e w s l e t t e r

Semantic Spatio-textual Data Analysis and Processing

The CHOROLOGOS project is a research project that is funded by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Innovation (GSRI), under grant agreement No [HFRI-FM17-81]. The instrument aims to support research of Academic Staff and Researchers, and the principal investigator of CHOROLOGOS is Christos Doulkeridis. The project is hosted at the Department of Digital Systems in the University of Piraeus.

Objectives and Challenges (Space + Time + Text)

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 daily. Interestingly, geotagged data originating from mobile users is not only **multidimensional** but also **unstructured**; apart from its spatial and temporal dimension, or numerical information (ratings, likes, retweets, etc.), textual descriptions are also available.

The combination of spatio-textual data with spatio-temporal data arises several challenges which are related to their collection, representation, processing, analysis, mining, and interpretation. This opens up new research directions, while at the same time challenges existing data processing solutions.

The research objective of CHOROLOGOS include:

- Formulation of expressive query types that enable selection of underlying spatio-temporal-textual data based on diverse information needs.
- Theoretical contributions in terms of properties and search bounds for the proposed query types, thus laying the foundations for efficient processing and search.

PROJECT'S MAIN GOAL

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.
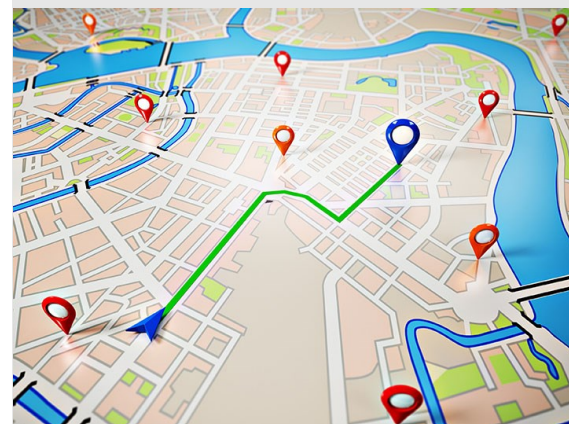
- Design of appropriate access methods that jointly index space, time, and text, in an appropriate way to support filtering of data that is irrelevant to the query at hand.
- Efficient query processing algorithms following well-established methodologies, including filter-and refine and branch-and-bound, 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.

Impact

By exploiting CHOROLOGOS, the analysis of massive spatio-textual datasets (especially in social networks), is going to be facilitated significantly. In consequence, applications will be able to query and analyze larger quantities of spatio-textual data in more expressive ways exploiting semantic retrieval, thus speeding up the analysis and interpretation of heterogeneous data (spatial or multidimensional data and unstructured textual data).

The expected impact of CHOROLOGOS to economy is non-negligible, as it relates to one of the main pillars of Greek economy, namely tourism. CHOROLOGOS promises to deliver innovative location-based services and applications, which would benefit the local tourism business considerably. The research results will be directly applicable as web/mobile search application for tourist-related information, using advanced and more expressive querying, which can be extremely useful to a tourist visiting an unknown city.
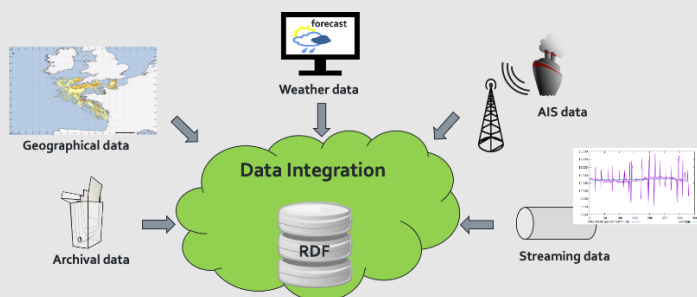
RDF-Gen: Generating RDF Triples from Big Data Sources

Transforming disparate and heterogeneous data sources that provide large volumes of data in high velocity into a common form, allows integrated and enriched views on data and thus, provides further opportunities to advance the effectiveness and accuracy of data analysis and prediction tasks.


RDF-Gen enables the transformation of data provided by archival and streaming data sources, provided in various formats, into RDF triples. RDF-Gen introduces a generic mechanism which supports high throughput and low latency on data transformation, even in cases where the velocity of data presents high peaks. A parallel implementation of RDF-Gen is also available.
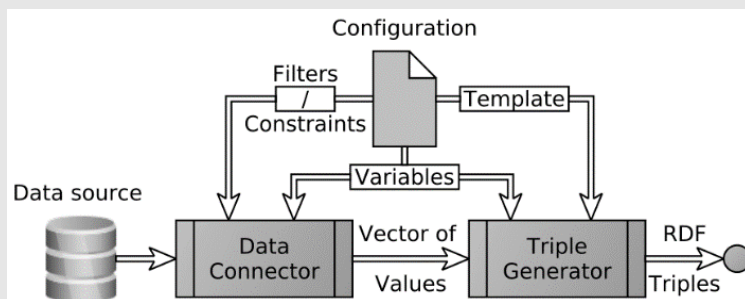
Main Features

- Inherently supports the RDF generation of both streaming and archival datasets.
- Provides facilities for close-to-source data processing tasks, e.g. for data cleansing, conversion of values, and generation of URIs.
- Scalable: necessary for the transformation of big data.
- Extensible: in the sense that (i) it can integrate custom data processing and manipulation functions, and (ii) it can be instantiated to new data formats.
- Efficient: in terms of high throughput and low data-generation latency.
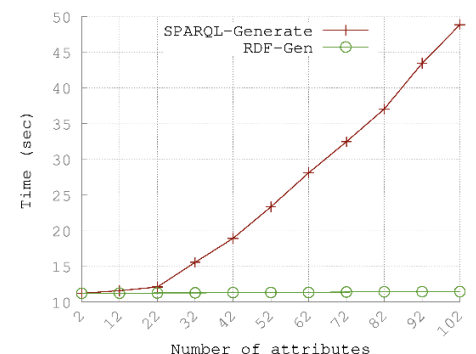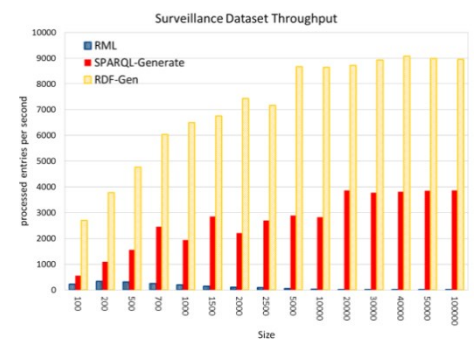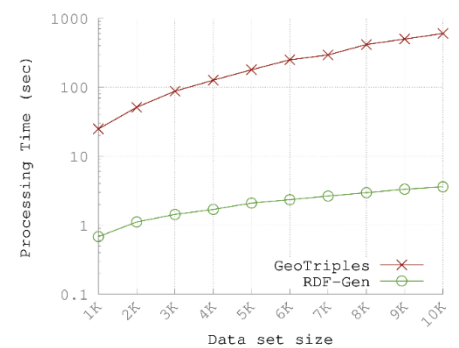
Publications

Book Chapters

1. Georgios Santipantakis, Christos Doulkeridis, Akrivi Vlachou, George Vouros, [Integrating Data by Discovering Topological and Proximity Relations Among Spatiotemporal Entities](), Big Data Analytics for Time-Critical Mobility Forecasting: From Raw Data to Trajectory-Oriented Mobility Analytics in the Aviation and Maritime Domains, "Springer International Publishing", (pp. 155-179), 2020.

2. Georgios Santipantakis, Christos Doulkeridis, George Vouros, [Link Discovery for Maritime Monitoring](), Guide to Maritime Informatics, Springer, (pp. 201-227), 2021.

Conferences

1. Georgios Santipantakis, George Vouros, Christos Doulkeridis, [Coronis: Towards Integrated and Open COVID-19 Data](), EDBT'21 demo track, 2021.

2. Nikolaos Koutroumanis, Christos Doulkeridis, [Scalable Spatio-temporal Indexing and Querying over a Document-oriented NoSQL Store](), EDBT'21, 2021.

3. Nikolaos Koutroumanis, Nikolaos Kousathanas, Christos Doulkeridis, Akrivi Vlachou, [A Demonstration of NoDA: Unified Access to NoSQL Stores](), VLDB'21 demo track, 2021.

4. Nikolaos Koutroumanis, Christos Doulkeridis, Akrivi Vlachou, [Tearing Down the Tower of Babel: Unified and Efficient Spatio-temporal Queries for NoSQL Stores](), MDM'22, 2022.

5. Fragkiskos Gryllakis, Nikos Pelekis, Christos Doulkeridis, Iraklis Varlamis, Yannis Theodoridis, [Social Spatio-temporal Keyword Pattern (S²KP) Queries in Multiple Aspect Trajectories Databases](), SSDBM'22, 2022.

Archive Reports

1. Georgios Santipantakis, George Vouros, Christos Doulkeridis, [Towards Integrated and Open COVID-19 Data](), CoRR, 2020.

Contact Us

Principal Investigator: Associate Professor Christos Doulkeridis

University of Piraeus, Department of Digital Systems

Phone: +30 210 414 2545

Email: cdoulk at unipi.gr