

# CHOROLOGOS

## Newsletter

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:

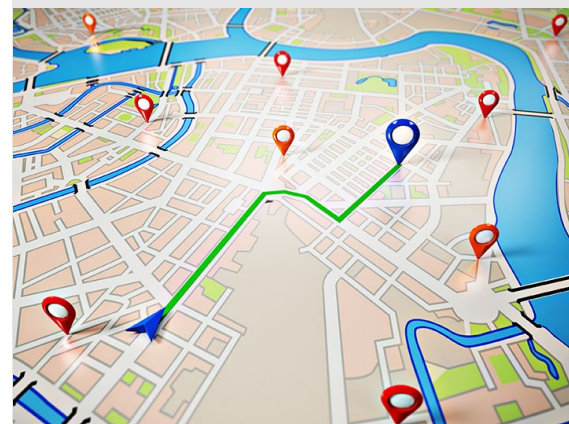
- Formulation of expressive query types that enable selection of underlying spatio-temporal-textual data based on diverse information needs.
- Theoretical contributions in terms of properties and search bounds for the proposed query types, thus laying the foundations for efficient processing and search.



### 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.

## WHAT WILL THE PROJECT OFFER IN PRACTICE?

---

CHOROLOGOS will offer a spatio-textual retrieval paradigm so as to be used in tandem with search engines. This will enable the formulation of complex queries with spatio-temporal constraints and associations between spatial objects of interest. An example of such query is *"which are the top-3 hotels that have the best combination of Italian restaurants and bars with nice cocktails in their close vicinity?"*

## WHAT'S NEW?

---

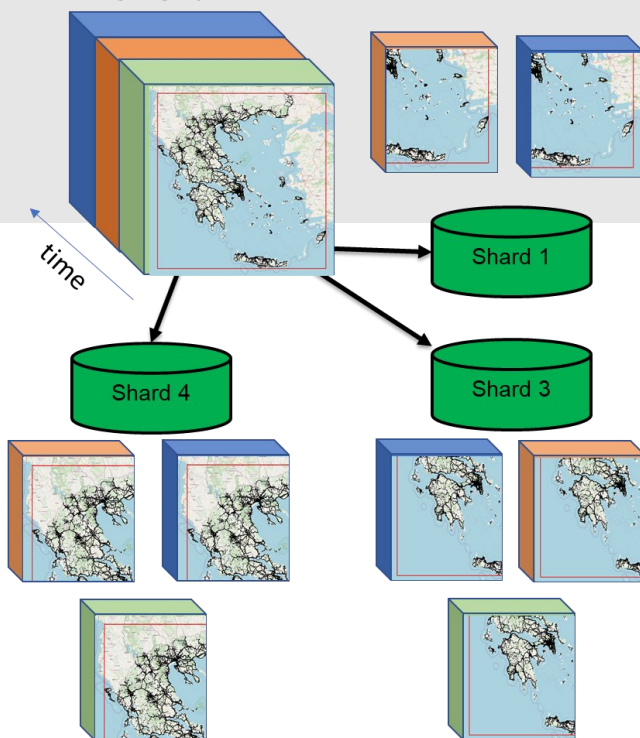
- Scalable Spatio-temporal Indexing and Querying over a Document-oriented NoSQL Store.

# Scalable Spatio-temporal Indexing and Querying over a Document-oriented NoSQL Store.

NoSQL systems are widely used by modern applications for scalability and high performance. However, existing NoSQL stores provide limited support for spatial data and (quite often) no native support for spatio-temporal data. As a result, the performance of query execution over large collections of spatio-temporal data is often suboptimal.

We present an approach for indexing spatio-temporal data, which is applicable to any NoSQL store that provides key-based access to data without modifications to its code, and we show how to generate data partitions that preserve data locality. The approach uses the Hilbert space-filling curve to generate one-dimensional (1D) keys, which facilitates indexing of spatio-temporal data, and is applied on a document-oriented NoSQL store, namely MongoDB.

Different temporal slices of the same geographical dataset

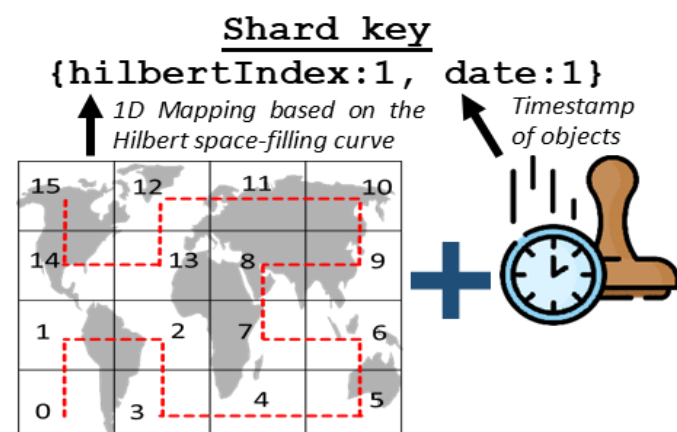


## Indexing & Sharding based on Hilbert SFC

Instead of relying on the built-in spatial indexes of MongoDB, our approach is based on the 1D mapping of data by means of the Hilbert space-filling curve. Specifically, for each document, the 1D Hilbert value is determined given its longitude and latitude value, and then it is included as a new field (`hilbertIndex`) that stores this value.

```
{
  _id: 1,
  location: {"type": Point,
            coordinates: [37.983810, 23.727539]},
  date: ISODate("2018-09-12T12:15:17.777Z"),
  hilbertIndex: NumberLong(10),
  ...
}
```

Given this new spatial field (`hilbertIndex`) and the date field, we set the shard key as `{hilbertIndex, date}`, thus imposing spatio-temporal partitioning of data to nodes. Consequently, the formation of the chunks is based both on spatial and temporal information and each of them contains documents that exist in specific spatial cells (1D values) for a certain time period. In case of spatio-temporal skewness in the data, chunks are unlikely to become jumbo because of the cardinality of the temporal field.



# 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. Panagiotis Tampakis, Dimitris Spyrellis, Christos Doulkeridis, Nikos Pelekis, Christos Kalyvas, Akrivi Vlachou, [A Novel Indexing Method for Spatial-Keyword Range Queries](#), SSTD'21, 2021.
5. Nikolaos Koutroumanis, Kousathanas Nikolaos, Christos Doulkeridis, Akrivi Vlachou, [Declarative Querying of Heterogeneous NoSQL Stores](#), SEADData'21, 2021.

## 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