SPADES Newsletter

May 2021, Issue 2

Spatio -Textual Data Exploration at Scale

Web: https://www.ds.unipi.gr/spades/

The SPADES project is a research project that is funded by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No [1667]. The instrument aims to support post-doctoral research, and the principal investigator of SPADES is Akrivi Vlachou. The project is hosted at the Department of Digital Systems in the University of Piraeus.

Description of Work

Several research challenges are associated with the efficient support of spatio-textual processing at scale. Two main factors are critical for location-based services, thus determining their overall performance: the efficiency of query processing for spatio-textual queries and the quality of the retrieved points of interest. the efficiency of query processing directly influences the query throughput, which is very important in the context of scalable applications. Consequently, SPADES identifies a set of research and technological challenges that need to be effectively addressed, in order to support spatio-textual and spatio-temporal queries over massive data:

1. Support of advanced spatio-textual query types

2. Novel distributed index structures for spatio-textual search

3. Abstractions for parallel spatio-textual data processing

Objectives

- Novel rank-aware query types complying with the paradigm of spatial-keyword search that cover a wide variety of information needs, targeting at the mobile user.
- Provision of more expressive querying mechanisms for points of interest that combine spatial information and textual relevance with temporal information and user preferences.
- Advanced distributed indexing structures capable to support complex spatial-keyword queries effectively, by means of harnessing the merits of spatial data structures and text indexes.
- Novel partitioning mechanisms and load balancing techniques for spatio-textual queries, aiming at efficient parallel data processing.

Impact

SPADES aims to address the limitations of spatio-textual data analysis and processing when applied in the context of Big Spatial Data, as witnessed by the lack of existing systems and techniques for this purpose. Achievement of this goal constitutes a substantial step forward in dealing with challenges emerging from management of Big Spatial Data. At a practical level, the research outcome will benefit applications such as spatio-textual search and retrieval, mining of spatio-textual data, next generation location-based services, and tourism-oriented applications to name a few.

By exploiting SPADES the analysis of massive spatio-textual datasets (typically encountered in the aforementioned domains and especially in social networks) is going to be accelerated significantly. In consequence, applications will be able to query and analyze larger quantities of spatio-textual data in shorter time, thus speeding up the process of making new scientific discoveries.

Results

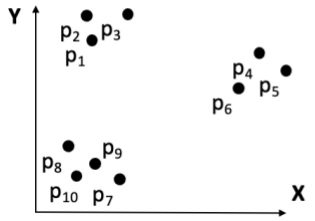
The results of the project are expected to be exploited by the local tourism business that could benefit from the provision of innovative location-based services over vast quantities of data to tourists. Such services entail queries that are processing-intensive and typically run for minutes rather than seconds, and often produce results that are not truly useful to the end user. SPADES promises to facilitate the access to location-based information, a task of particular importance to tourists. In this respect, the research results of SPADES are expected to benefit also society at large.

Research Highlight: A Novel Mapping of Spatio-textual Data

We propose a novel indexing method for spatio-textual data that supports efficient processing of spatial-keyword range queries, where given a query ***q*** that consists of a location (***q.x***, ***q.y***) and a set of keywords ***Q***, the objective is to retrieve all spatio-textual objects within distance ***r*** from ***q*** and having keyword set similarity above a user-specified threshold ***τ***.This query type offers more flexibility than boolean spatial-keyword queries that impose exact matching on the keywords that describe each object.

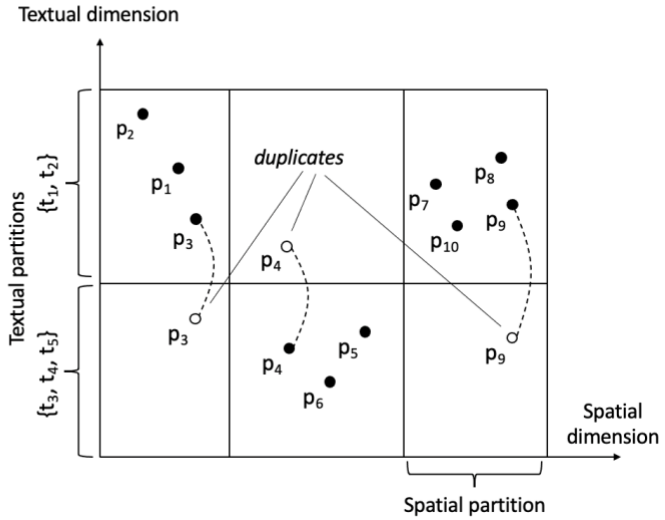
Nevertheless, efficient processing of spatial-keyword queries requires the use of specialized access methods that combine spatial and text indexing techniques in a joint index, a challenging topic due to the high dimensionality of spatio-textual representations. Unfortunately, these index structures typically have high memory or disk requirements due to the integration of spatial with textual information, and moreover they are not supported in existing database management systems.

Original Spatio-textual Data Space



Object	Text	Object	Text
p_1	t_1, t_2	p_6	t_3, t_5
p_2	t_1, t_2	p_7	t_1, t_2
p_3	t_1, t_3	p_8	t_1, t_2
p_4	t_2, t_4	p_9	t_1, t_4
p_5	t_3, t_4, t_5	p_{10}	t_2

2D Transformed Space



In the figure, an illustrative example of our mapping is shown, with the original data set ***D*** that consists of 10 objects {p_i} associated with keywords {t_j} depicted on the left, whereas the transformed 2D space is shown on the right. In summary, we map the location information in one dimension (horizontal axis), and the textual information in another dimension (vertical axis). Essentially, in the transformed 2D space, the objects form spatial partitions based on their pairwise distances, as well as textual partitions (in the example {t_1, t_2} and {t_3, t_4, t_5}) based on grouping together subsets of frequently co-occurring keywords. While each object belongs to a single spatial partition, it may be assigned to multiple textual partitions. For example, objects p_1, p_2 and p_3 are assigned to the same spatial partition because they form a spatial cluster in the original space. On the other hand, objects p_3, p_4 and p_9 are duplicated to both textual partitions depicted, because they contain keywords from both textual partitions.

Interested to find our more? Details on:

- Deriving appropriate search bounds for spatial-keyword range queries in the transformed space
- Two efficient query processing algorithms that rely on this mapping
- An implementation in PostgreSQL 13 and PostGIS

can be found in our paper "***A Novel Indexing Method for Spatial-Keyword Range Queries***" which will apper in the ***17th International Symposium on Spatial and Temporal Databases (SSTD'21)*** in August 2021.

Publications

Journals

1. Christos Doulkeridis, Akrivi Vlachou, Nikos Pelekis, Yannis Theodoridis: A Survey on Big Data Processing Frameworks for Mobility Analytics. In SIGMOD Record, 2021 (in press).
2. Panagiotis Nikitopoulos, Georgios A. Sfyris, Akrivi Vlachou, Christos Doulkeridis, Orestis Telelis: Pruning Techniques for Parallel Processing of Reverse Top-k Queries. In Distributed and Parallel Databases Journal, Springer, Vol. 39, Issue 1, pages 169-199, March 2021.

Conferences & Workshops

1. Nikolaos Koutroumanis, Nikolaos Kousathanas, Christos Doulkeridis, Akrivi Vlachou. A Demonstration of NoDA: Unified Access to NoSQL Stores. In Proceedings of the 47th International Conference on Very Large Data Bases (VLDB'21), Copenhagen, Denmark - August 16-20, 2021.
2. Panagiotis Tampakis, Dimitris Spyrellis, Christos Doulkeridis, Nikos Pelekis, Christos Kalyvas and Akrivi Vlachou. A Novel Indexing Method for Spatial-Keyword Range Queries. In Proceedings of the 17th International Symposium on Spatial and Temporal Databases (SSTD'21), August 2021.
3. Andreas Tritsarolis, Christos Doulkeridis, Nikos Pelekis, Yannis Theodoridis. ST_VISIONS: A Python Library for Interactive Visualization of Spatio-temporal Data. In Proceedings of 22nd International Conference on Mobile Data Management (MDM'21) - demo track, Toronto, Canada (Virtual), June 15-18, 2021
4. Akrivi Vlachou, Christos Doulkeridis, Nikolaos Koutroumanis, Dimitrios Poulopoulos, Kjetil Norvag. The SPADES Framework for Scalable Management of Spatio-textual Data. In Proceedings of 24th Pan-Hellenic Conference on Informatics (PCI'20), Athens, Greece, November 2020
5. Stella Maropaki, Sean Chester, Christos Doulkeridis, Kjetil Norvag: Diversifying Top-k Point-of-Interest Queries via Collective Social Reach In Proceedings of 29th ACM Conference on Information and Knowledge Management (CIKM'20), October 19-23, 2020.
6. Nikolaos Koutroumanis, Panagiotis Nikitopoulos, Akrivi Vlachou, Christos Doulkeridis: NoDA: Unified NoSQL Data Access Operators for Mobility Data. In Proceedings of the 16th International Symposium on Spatial and Temporal Databases (SSTD'19), Vienna, Austria, August 2019.
7. Georgios M. Santipantakis, Apostolos Glenis, Christos Doulkeridis, Akrivi Vlachou, George A. Vouros: stLD: Towards a Spatio-temporal Link Discovery Framework. In Proceedings of the International Workshop on Semantic Big Data (SBD'19) (workshop held in conjunction with SIGMOD'19), Amsterdam, The Netherlands, July 2019, pp.4:1-4:6.
8. Panagiotis Nikitopoulos, Georgios A. Sfyris, Akrivi Vlachou, Christos Doulkeridis, Orestis Telelis: Parallel and Distributed Processing of Reverse Top-k Queries. In Proceedings of the 35th IEEE International Conference on Data Engineering (ICDE'19), Macau SAR, China, April 2019, pp.1586-1589
9. Akrivi Vlachou, Christos Doulkeridis, Apostolos Glenis, Georgios M. Santipantakis, George A. Vouros: Efficient Spatio-temporal RDF Query Processing in Large Dynamic Knowledge Bases. In Proceedings of the 34th ACM/SIGAPP Symposium On Applied Computing (SAC'19), Limassol, Cyprus, April 2019, pp.439-447.

Contact Us

For more information please contact Akrivi Vlachou (avlachou@unipi.gr).
For more information about the research group and the department, please visite the respective home pages:
Department of Digital Systems

SPADES
SPAtiotextual Data Exploration at Scale

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

GSRT
GENERAL SECRETARIAT FOR RESEARCH AND TECHNOLOGY