Institute for the Management of Information Systems Athena Research Center

http://www.imis.athena-innovation.gr

1. INTRODUCTION

The Institute for the Management of Information Systems (IMIS) is part of the "Athena" Research Center, and was founded in 2007. IMIS is directed by Timos Sellis, a Professor at the School of Electrical and Computer Engineering in the National Technical University of Athens.

The mission of IMIS is to conduct research in the area of data management and large-scale information systems. This report presents an overview of its major research activities, which are currently headed by senior researchers Theodore Dalamagas, Dieter Pfoser, Yannis Stavrakas and Professor Panos Constantopoulos (Athens University of Economics and Business), and postdoctoral researchers George Papastefanatos, Dimitris Sacharidis, and Manolis Terrovitis. The institute members cooperate also with several other faculty members from universities in Greece.

2. DATA LIFECYCLES

Large data sets, and particularly curated scientific databanks, are often compared to living organisms, as they continuously evolve and introduce additional management requirements. A significant effort in IMIS is on handling the lifecycles of such evolving data collections.

Data Evolution. The key problem in reasoning about *data evolution* stems from the fact that information systems usually treat changes as distinct events. In reality, a number of changes that occur at disparate and seemingly unrelated pieces of data constitute conceptually a single "complex" change event. Such high-level changes are more meaningful than the individual changes they encompass, and offer a richer interpretation of the evolution process. In our approach, changes are discrete objects that have complex structure and retain their semantic and temporal characteristics, rather than being isolated low-level transformations on data. Consider the following simple example: the high-level change operation "move" is represented as a complex object composed by the atomic change objects "remove" and "add".

Evo-graph [37] is a novel model for data evolution that describes successive snapshots and treats changes as first-class citizens. A snapshot captures the state of the database at a specific time instance, and is represented as a rooted directed graph. A change can be compound, i.e., comprising disparate changes, and is associated with the data items it affects. Managing data evolution with evo-graphs is particularly useful when the provenance of the data needs to be traced, and past states need to be re-assessed. [38] specifies how an evo-graph can be reduced to the snapshot at a specified time instance. Furthermore, it introduces an XML representation of evo-graph, called evoXML, and presents a method to answer interesting evolution queries. A prototype, called *C2D*[1], records the history of data and the structure of changes step by step, as the current snapshot evolves [26]. Our current work is on a query language for evo-graph expressing conditions on both changes and data, and on evoXML storage optimization.

Schema Evolution. IMIS is also concerned with the management of *schema evolution* in data-centric ecosystems. Such systems, comprising a large number of applications and data stores, are highly vulnerable to schema changes. *Hecataeus* [27] is a set of tools for managing a data-centric ecosystem, i.e., the database schema along with its dependent views and queries. It satisfies the fundamental needs of the developer, administrator, and designer of a data-centric ecosystem. The developer can create evolution scenarios in order to evaluate the impact of a schema evolution event, and define rules so that both syntactical and semantic correctness is retained. The administrator can control the propagation of the event's impact to affected constructs. The designer is given an extensible suite of objective metrics that report crucial and vulnerable parts of the system regarding potential evolution events.

Digital Curation. The need to ensure adequate representation and long-term access to digital information as its context of use changes, and to counter the risk

[1] `http://web.imis.athena-innovation.gr/projects/c2d`

of repositories becoming "data mortuaries", introduces a grand challenge for digital curation research: to develop the conceptual and technological tools necessary for maintaining and adding value to a trusted body of digital information for current and future use, through the active questioning, dynamic co-evolution and adequate knowledge representation of its epistemic and pragmatic content and context.

To address this challenge, IMIS adopts (a) a *lifecycle approach* to the representation of curated information objects as these evolve in interaction with changing designated communities, (b) a *cross-disciplinary* scope, so as to cater adequately for differences in digital curation requirements between diverse scientific and functional (business, social, economic) contexts of use, (c) a broader notion of digital curation *actors*, including those involved in the production, public communication and utilization of knowledge, and (d) *event-centric* structural representations of digital information "life events".

Current lines of work include: (a) modeling digital curation processes, (b) representing domain knowledge in the form of ontologies and reference models, (c) modeling scientific and scholarly information behavior, processes and user requirements, (d) developing and maintaining knowledge resources and knowledge organization systems, (e) streamlining the enrichment of these resources from text by extracting relevant entities and relations, (f) ontology-driven search and fact discovery, (g) automatically generating text from databases as a more human-oriented form of information, to be considered for preservation purposes in addition to communication, (h) preserving contextual, schema and operational information in conjunction with primary data, so as to enable the use of databases containing valuable data over time, (i) user community modeling and social tagging, (j) conceptualizations of epistemic discourse and communication genres (i.e., rhetorical structure) in specific disciplinary and pragmatic contexts, (k) grounded research on the motives, activities and contexts of appraisal, knowledge enhancement and use of digital resources by diverse interpretive communities, and (l) cost-benefit assessment of preservation policies.

A significant part of IMIS work has contributed to a number of European projects in digital humanities infrastructures[2] and good practices[3]. Some results to date are: a digital curation lifecycle model [10, 11], an evidence-based process model of scholarly activity [3], user requirements for digital scholarly research infrastructures [6, 2], an architecture and approach to asset representation for aggregating cultural content [28], a curation-aware repository system that supports semantically-enhanced preservation services [4, 17, 18], semantic integration of collection descriptions [24], text generation from ontologies [21], ontology-based development of query patterns and their exploitation in optimizing RDF queries [12, 13].

[2] `http://www.dariah.eu` and `http://www.ehri-project.eu`

[3] `http://www.carare.eu`

3. GEOINFORMATICS

The proliferation of location-aware devices, e.g., GPS receivers, smartphones, etc., has resulted in an abundance of spatiotemporal data, which, in turn, has renewed interest in geospatial data management methods. IMIS has a long and strong interest in geoinformatics, and its research activities include indices and algorithms for road networks, as well as crowdsourcing approaches.

Road Networks. Road networks represent an interesting showcase for geospatial data management algorithms. IMIS research focuses on provision of network attributes, e.g., dynamic travel times, and the design of efficient transportation algorithms, e.g., dynamic and hierarchical routing, TSP, etc.

We introduce a hierarchical tiling scheme [16] that benefits shortest-path computation. Somewhat related, the HBA* algorithm [30] explores network hierarchies to determine the shortest path. While typically GPS technology is used to track location, our work in [5] demonstrates that wireless networks can become a complementary source especially for mobile phone users.

Optimization problems is another area of interest. In the pickup and delivery with transfers problem, the goal is to assign of a set of transportation requests to a fleet of vehicles in a way that satisfies a number of constraints and at the same time minimizes a specific cost function. [7] studies its dynamic variant and proposes a solution based on a label-setting search algorithm.

Geospatial Crowdsourcing. A research ambition of IMIS is to utilize crowdsourcing and apply smart content extraction techniques to develop large, qualitative geospatial datasets. In particular, our work [29] focuses on (a) tools for intentional provision of user-contributed spatiotemporal information (geoblogging), (b) identifying and recording geospatial observation data contained in existing content, (c) developing uncertainty as a means to represent spatial and spatiotemporal relationship data, (d) fusing data to derive an integrated geospatial datasets, and (e) developing reasoning algorithms for an evolving geospatial knowledge base.

IMIS has developed a tool [22] for collaborative search and extraction of geospatial objects from point cloud datasets such as flickr. The work in [14] summarizes the content of travel blogs by means of user sentiment, i.e., by geocoding and sentiment mapping travel blog texts, we create a heat map indicating the user sentiment (ranging from positive to negative) in relation to

geographic areas. [23] explores the annotation of travel guide content with task and location information so as to increase its usability. Narratives, in the form of travel blog entries, provide a rich resource for geospatial data, and [15] describes a method to extract geospatial route information from texts. Rather than extracting geospatial content, [31] presents a geoblogging Web-based tool that provides a simple means to author geocoded travel blog entries consisting of texts, images, video and audio.

Indexing Techniques. For a variety of problems, IMIS has utilized specialized geospatial indices. One such case is *path queries* over a large collection of geospatial routes, which is a special case of reachability queries. In particular, the goal is to construct a path between two given locations, if one exists. To expedite query processing, we introduce two complementary indexing schemes that capture transitivity information among the routes [9]. Based on these indices, our proposed algorithms are significantly faster than conventional graph traversal techniques. A more recent work [8] performs a detailed asymptotic analysis of the indices, and presents a methodology for efficiently updating them as routes are added or deleted from the collection.

For the case of distributed geospatial data, we propose a novel inherently multidimensional index, termed MIDAS [46]. MIDAS implements a distributed k-d tree, whose leaves correspond to peers, and its internal nodes dictate message routing. MIDAS requires that peers maintain little network information, and features mechanisms that support fault tolerance and load balancing. For a network of n peers, MIDAS is shown to process point and range queries in $O(\log n)$ hops in expectance.

4. WEB OF DATA

An important research area of IMIS is Web technologies, and particularly the Web of Data, described as "a Web of things in the world, described by data on the Web". IMIS has developed technologies related to Linked Open Data, RDF stores, and Web Services.

Linked Open Data. IMIS is the premium national R&D pole in Linked Open Data management and governance. The Linked Data paradigm involves practices to publish, share, and connect data on the Web, and offers a new way of data integration and interoperability. Linked Data technologies enables the Web of Data, which extends current Web to a global data space connecting data from diverse domains, and is impelled by the current trend towards an open Web. The open data movement is a significant and emerging force towards this direction. Open data is public data which are easily discoverable, accessible, and available to people without any restriction. *Linked Open Data* (LOD) serve a great cause, enabling transparency, accountability and good governance for public administrations.

IMIS has developed and launched various Open Data services. The most significant project is geodata.gov[4], a leading effort on providing services for spatial data, done in collaboration of the PM's eGov office, and the Ministry of Environment, Energy, and Climate Change. Also, IMIS has designed Linked Open Data services for microRNA databanks[5] maintained by the "A. Fleming" Biomedical Sciences Research Center (BSRC)[6], where information about biological entities are provided for the scientific community of life sciences.

The goal of IMIS is to provide innovative technologies for best governance and curation practices to produce sustainable LOD ecosystems. Our research handles the full lifecycle of LOD ecosystems, from data extraction, storage and maintenance, to monitoring, protection and repair. We focus on the following scientific and technological activities: (a) effective methods for exposing large volumes of structured and unstructured data as LOD, (b) efficient storage solutions for large volumes of LOD, (c) query methods, retrieval algorithms and ranking techniques, (d) methods for interlinking and fusing LOD from different Web data sources, (e) models and query languages to represent and query changes in LOD spaces, (f) provenance models and methods to trace the origins and transformations in LOD spaces, (g) design principles and best practices to expose LOD with anonymity guarantees, and (h) models and methods to ensure privacy for publishing LOD.

RDF Stores. In line with its LOD initiative, IMIS has developed expertise in RDF storage engines. In particular, [47] proposes a distributed storage system for RDF triples. The store is based on the distributed index of [46] and utilizes a labeling scheme to encode transitivity information within the RDF subgraphs. As a result, our store efficiently processes RDF pattern queries with known performance guarantees, and also features a forward chaining mechanism that supports RDF Schema reasoning. To exemplify its functionality, a distributed semantic content-based publish-subscribe service is implemented on top of the distributed RDF store.

Semantic Web Services. Research in IMIS also addresses the matchmaking semantic Web services, i.e., how to retrieve the best service given a user's request. If the requested service is too specific and/or the number of matching services is large, selecting the most appropriate service is a challenging task. There exist many criteria to consider for ranking Web services, such as the degree of match, quality of service, and moreover there exist many methods to quantify them. However, no sin-

[4] `http://geodata.gov.gr`
[5] `http://diwis.imis.athena-innovation.gr/mlod`
[6] `http://www.fleming.gr`

gle criterion or metric is ideal. In [35], we propose a methodology that is able to combine multiple criteria in an objective manner. To increase the utility and diversity of the returned Web services, the work in [36] extends this framework and proposes clustering method to better capture trade-offs among criteria.

5. PERSONALIZED DATA MANAGEMENT

Personalization refers to data management techniques that are centered around the user's interests, preferences, and intentions. IMIS has studied several aspects, including search and ranking, skyline queries, and indices.

Personalized Search. The general goal of personalized search is to exploit historic search data and re-rank/filter retrieved documents so as to better serve individual information needs. Traditionally, search engines have focused on improving the accuracy of results. However, accuracy does not tell the whole story. Real user experience shows that search should take into account user's extent of interest in various topics when presenting the search results. For example, this is the case when the answer set includes many documents which are similar to each other, and the resulting ranked list seems poor due to lack of diversity. A better approach is to present results in a way that covers a more diverse set of topics capturing all possible interpretations of the user's intent.

Research in IMIS includes orthogonal approaches for re-ranking search results that do not share the above limitation, and benefit all users equally [19, 20]. They are based on the observation that existing techniques focus on the retrieved content and on users search histories, but leave an important aspect unaddressed: the analysis of user search behavior. This behavior is directly observable by user feedback, by means of clicks on the results, and allows reasoning about the user's intent.

Our approach does not rely on user-specific models, but captures the user intent by grouping queries entailing similar behavior. Thus, the presented search results consist of different document types (e.g., specifications, promotion, reviews, etc.) that have been associated with products in the past. In other words, our method re-ranks the retrieved results, so that they represent the broad spectrum of user behavior for a given query. The aim is to diversify the search results in order to further enhance the exploration of the document information space. The underlying technique builds models for user intent by clustering queries with respect to the user intent, and then learns a ranking function for every cluster. The clustering and the ranking function are jointly optimized due to their dependencies.

Skyline Queries. The skyline operator has received considerable attention over the past few years as a means to retrieve tuples that are objectively good for a large set of meaningful ranking functions. The vast majority of methods are designed to efficiently process the case when every ranking attribute is totally ordered. On the other hand, our framework [34] allows the extension of almost any existing technique to handle attributes that are only partially ordered. This is quite useful for preferences specified on nominal attributes, e.g., the Make of a product, or on hierarchical attributes, e.g., an ontology. Moreover, our framework naturally extends to the dynamic case, where, for the same attribute, users may have different and conflicting preferences.

Depending on the context or underlying task, the preferences of a user may differ considerably. However, explicitly asking the user to specify her/his preferences for each context is burdensome. Instead, [32] takes a different approach. Based on the information collected so far about various contexts and user preferences, we propose a simple mechanism that uses probabilities to define preferences for the current context. Subsequently, we appropriately redefine the skyline operator so as to capture this uncertainty among preferences.

Indexing Techniques. Personalized data management can benefit from low-level performance optimizations of commonly encountered queries. One such example is boolean containment queries, motivated by the case of large set-valued tuples. Our research addresses the problem from a traditional database perspective, and focuses on data management systems that use the secondary storage to keep the indices and the data. The work in [41] proposes methods that reduce the I/O cost in basic containment queries, like subset and superset.

6. BIOLOGICAL DATA MANAGEMENT

Advances in various life sciences' applications have led to an explosive growth of experimental and computational data. IMIS establishes collaborations with biologists and designs tailor-made tools to aid their research.

Biological Analysis Tools. Biologists used to consider proteins and DNA as movers and shakers in genomics, seeing RNA as nothing more than a messenger to carry information between the two. This has changed, in early 2000s, after the discovery of the key role played in gene expression by small RNA molecules, called microRNAs. MicroRNAs can completely silence proteins by binding themselves to complementary sequences on mRNA transcripts, called *targets*. However, there is a lack of high-throughput experimental methods for identifying microRNA targets. Thus, computational methods have become increasingly important, and led to the experimental identification of many microRNA targets.

IMIS and the DNA Intelligent Analysis group (DIANA) of BSRC have designed and implemented an ad-

vanced IT infrastructure for genomic data management, oriented to processing, analysis and visualization of computationally predicted microRNA targets [25, 1]. Furthermore, IMIS and DIANA have developed a set of advanced Web applications to provide access to computationally predicted microRNA targets, including two core services: (a) DIANA microT[7] and (b) DIANA mirGen[8]. Since its original launch, these services have been one of the most widely used tool for microRNA analysis, counting more than 1,500 users per month.

Sequence Alignment. Following the collaboration between IMIS and DIANA, we have worked on a novel sequence alignment problem [48]. It has been observed that a chemical association, known as binding, of a non-coding RNA sequence, the pattern, with a gene usually occurs around a key location of the pattern, called the core (typically around the nucleotides near the start of the non-coding RNA). The Approximate Regional Sequence Matching (ARSM) problem models this chemical phenomenon and is able to predict if and where bindings occur. We have proposed an efficient algorithm that outperforms existing techniques used by biologists.

Furthermore, IMIS examines next-generation high-performance platforms for sequence alignment problems. Particularly, we employ cloud computing to efficiently detect approximate matching of pattern sequences on data sequences. This research is related to subcontracted pilot activities carried out for VENUS-C[9], an FP7 Research Infrastructures project.

7. PRIVACY PRESERVATION

Protection of privacy in information systems has attracted significant interest in the last years. By combining research and applied activities, IMIS aims at transferring its expertise to industry and academia.

Research in IMIS focuses on privacy protection methods based on data anonymization. Unlike traditional approaches, where privacy is protected by regulating access to the data, anonymization regulates the amount of information that is released in each shared dataset. The goal is to transform a dataset in a way that specific personal characteristics can no longer be discerned.

An important line of work in IMIS is privacy preservation for highly dimensional data. Examples of such data are transactional logs [40, 42, 45, 44], user movement traces [43, 39], XML files and multi-relational databases. To identify privacy threats to individuals or other entities, their respective records have to be treated as points in a multidimensional data space. The more information these records contain, the higher the dimensionality. Anonymization in this context is challenging, since adversaries can easily detect outliers, by exploiting any dimension where a record has a unique value.

Another research direction is anonymization in the presence of publicly available external knowledge. Privacy always comes at the expense of data usefulness. An interesting result of our work [33] is that it is possible to release anonymized data that are less generalized than they would be using conventional methods. We propose a generic framework that extends any existing k-anonymity algorithm so as to take into account publicly available datasets in a manner that increases the utility of the released anonymized data.

The privacy related work in IMIS is not limited to basic research, but also includes consulting services for the public and private sector and the creation of anonymization tools. IMIS is supporting the public sector in its effort to release its data to the public, by providing expertise and solutions on data anonymization.

[7] `http://diana.cslab.ece.ntua.gr/DianaTools/index.php?r=microtv4`

[8] `http://diana.cslab.ece.ntua.gr/?sec=databases`

[9] `http://www.venus-c.eu`

8. ACKNOWLEDGEMENTS

The research results presented here were made possible thanks to the dedication and hard work of the entire IMIS research team. Acknowledgments are due to Prof. Ion Androutsopoulos, Stavros Angelis, Anastasios Arvanitis, Spiros Athanasiou, Agiatis Benardou, Prof. Costis Dallas, Prof. Antonios Deligiannakis, Vicky Dritsou, Euthymios Drymonas, Alexandros Efentakis, Dimitris Gavrilis, Panos Georgantas, Giorgos Giannopoulos, Prof. Yannis Kotidis, John Liagouris, Prof. Christos Papatheodorou, Dimitrios Skoutas, and Thanasis Vergoulis.

9. REFERENCES

[1] P. Alexiou, T. Vergoulis, M. Gleditzsch, G. Prekas, T. Dalamagas, M. Megraw, I. Grosse, T. Sellis, and A. Hatzigeorgiou. mirgen 2.0: a database of microrna genomic information and regulation. *Nucleic Acids Research (Database issue)*, 38, 2010.

[2] S. Angelis, A. Benardou, P. Constantopoulos, and C. Dallas. Defining user requirements for holocaust research infrastructures and services in the ehri project. In *Proc. of the iConference*, 2012.

[3] S. Angelis, A. Benardou, P. Constantopoulos, C. Dallas, and D. Gavrilis. A conceptual model for scholarly research activity. In *Proc. of the iConference*, 2010.

[4] S. Angelis, A. Benardou, P. Constantopoulos, C. Dallas, and D. Gavrilis. A curation-aware repository supporting information-intensive scholarly research. In *Proc. of the Intl. Digital Curation Conf.*, 2010.

[5] S. Athanasiou, P. Georgantas, G. Gerakakis, and D. Pfoser. Utilizing wireless positioning as a tracking data source. In *Proc. of the Intl. Symposium on Spatial and Temporal Databases*, 2009.

[6] A. Benardou, P. Constantopoulos, C. Dallas, and D. Gavrilis. Understanding the information requirements of arts and humanities scholarship: implications for digital curation. *Intl. Journal of Digital Curation*, 5(1), 2010.

[7] P. Bouros, D. Sacharidis, T. Dalamagas, and T. Sellis. Dynamic pickup and delivery with transfers. In *Proc. of the Intl. Symposium on Spatial and Temporal Databases*, 2011.

[8] P. Bouros, D. Sacharidis, T. Dalamagas, S. Skiadopoulos, and T. Sellis. Evaluating path queries over frequently updated route collections. In *IEEE Transactions on Knowledge and Data Engineering*, 2011.

[9] P. Bouros, S. Skiadopoulos, T. Dalamagas, D. Sacharidis, and T. Sellis. Evaluating reachability queries over path collections. In *Proc. of the Intl. Conf. on Scientific and Statistical Database Management*, 2009.

[10] P. Constantopoulos and C. Dallas. Aspects of a digital curation agenda for cultural heritage. In *Proc. of the IEEE Intl. Conf. on Distributed Human-Machine Systems*, 2008.

[11] P. Constantopoulos, C. Dallas, I. Androutsopoulos, S. Angelis, A. Deligiannakis, D. Gavrilis, Y. Kotidis, and C. Papatheodorou. DCC&U: An extended digital curation lifecycle model. *Intl. Journal of Digital Curation*, 3(1), 2009.

[12] P. Constantopoulos, V. Dritsou, and E. Foustoucos. Developing query patterns. In *Research and Advanced Technology for Digital Libraries*, 2009.

[13] V. Dritsou, P. Constantopoulos, A. Deligiannakis, and Y. Kotidis. A curation-aware repository supporting information-intensive scholarly research. In *Proc. of the Intl. Digital Curation Conf.*, 2010.

[14] E. Drymonas, A. Efentakis, and D. Pfoser. Opinion mapping travelblogs. In *Proc. of the Terra Cognita Workshop*, 2011.

[15] E. Drymonas and D. Pfoser. Geospatial route extraction from texts. In *Proc. of the ACM SIGSPATIAL Intl. Workshop on Data Mining for Geoinformatics*, 2010.

[16] A. Efentakis, D. Pfoser, and A. Voisard. Efficient data management in support of shortest-path computation. In *Proc. of the ACM SIGSPATIAL Intl. Workshop on Computational Transportation Science*, 2011.

[17] D. Gavrilis, A. Angelis, and C. Papatheodorou. Mopseus - a digital repository system with semantically enhanced preservation services. In *Proc. of the Intl. Conf. on Preservation of Digital Objects*, 2010.

[18] D. Gavrilis, C. Papatheodorou, P. Constantopoulos, and S. Angelis. Mopseus - a digital library management system focused on preservation. In *Proc. of the European Conf. on Research and Advanced Technology for Digital Libraries*, 2010.

[19] G. Giannopoulos, U. Brefeld, T. Dalamagas, and T. Sellis. Learning to rank user intent. In *Proc. of the Intl. Conf. on Information and Knowledge Management*, 2011.

[20] G. Giannopoulos, T. Dalamagas, and T. Sellis. Search behavior-driven training for result re-ranking. In *Proc. of the Intl. Conf. on Theory and Practice of Digital Libraries*, 2011.

[21] G. Karakatsiotis, D. Galanis, and I. Androutsopoulos. Naturalowl: Generating texts from owl ontologies in protege and in second life. In *Proc. of the European Conf. on Artificial Intelligence*, 2008.

[22] G. Lamprianidis and D. Pfoser. Jeocrowd: collaborative searching of user-generated point datasets. In *Proc. of the ACM SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems*, 2011.

[23] J. Liagouris, S. Athanasiou, A. Efentakis, S. Pfennigschmidt, D. Pfoser, E. Tsigka, and A. Voisard. Mobile task computing: beyond location-based services and ebooks. In *Proc. of the Intl. Conf. on Web and wireless geographical information systems*, 2011.

[24] I. Lourdi, C. Papatheodorou, and M. Doerr. Semantic integration of collection description: Combining CIDOC/CRM and dublin core collections application profile. *D-Lib Magazine*, 15(7/8), 2009.

[25] M. Maragkakis, T. Vergoulis, P. Alexiou, M. Reczko, K. Plomaritou, M. Gousis, K. Kourtis, N. Koziris, T. Dalamagas, and A. Hatzigeorgiou. Diana-microt web server upgrade supports fly and worm mirna target prediction and bibliographic mirna to disease association. *Nucleic Acids Research (Webserver issue)*, 39, 2011.

[26] G. Papastefanatos, Y. Stavrakas, and T. Galani. Capturing the history and change structure of evolving data. *Submitted*.

[27] G. Papastefanatos, P. Vassiliadis, A. Simitsis, and Y. Vassiliou. Hecataeus: Regulating schema evolution. In *Proc. of the IEEE Intl. Conf. on Data Engineering*, 2010.

[28] C. Papatheodorou, C. Dallas, C. Ertmann-Christiansen, K. Fernie, D. Gavrilis, M. Masci, P. Constantopoulos, and S. Angelis. A new architecture and approach to asset representation for europeana aggregation: The CARARE way. In *Proc. of the Metadata and Semantic Research Conf.*, 2011.

[29] D. Pfoser. On user-generated geocontent. In *Proc. of the Intl. Symposium on Spatial and Temporal Databases*, 2011.

[30] D. Pfoser, A. Efentakis, A. Voisard, and C. Wenk. A new perspective on efficient and dependable vehicle routing. In *Proc. of the ACM SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems*, 2009.

[31] D. Pfoser, C. Lontou, E. Drymonas, and S. Georgiou. Geoblogging: user-contributed geospatial data collection and fusion. In *Proc. of the ACM SIGSPATIAL Intl. Conf. on Advances in Geographic Information Systems*, 2010.

[32] D. Sacharidis, A. Arvanitis, and T. Sellis. Probabilistic contextual skylines. In *Proc. of the IEEE Intl. Conf. on Data Engineering*, 2010.

[33] D. Sacharidis, K. Mouratidis, and D. Papadias. k-anonymity in the presence of external databases. *IEEE Transactions on Knowledge and Data Engineering*, 2(3), 2010.

[34] D. Sacharidis, S. Papadopoulos, and D. Papadias. Topologically sorted skylines for partially ordered domains. In *Proc. of the IEEE Intl. Conf. on Data Engineering*, 2009.

[35] D. Skoutas, D. Sacharidis, A. Simitsis, V. Kantere, and T. Sellis. Top-k dominant web services under multi-criteria matching. In *Proc. of the Intl. Conf. on Extending Database Technology*, 2009.

[36] D. Skoutas, D. Sacharidis, A. Simitsis, and T. Sellis. Ranking and clustering web services using multi-criteria dominance relationships. *IEEE Transactions on Services Computing*, 2010.

[37] Y. Stavrakas and G. Papastefanatos. Supporting complex changes in evolving interrelated web databanks. In *Proc. of the Intl. Conf. on Cooperative Information Systems*, 2010.

[38] Y. Stavrakas and G. Papastefanatos. Using structured changes for elucidating data evolution. In *Proc. of the ICDE Workshop on Managing Data Throughout its Lifecycle*, 2011.

[39] M. Terrovitis. Privacy preservation in the dissemination of location data. *ACM SIG KDD Explorations*, 13(1), 2011.

[40] M. Terrovitis. Privacy preserving publishing of set-values. *Privacy Observatory Magazine*, 2, 2011.

[41] M. Terrovitis, P. Bouros, P. Vassiliadis, T. Sellis, and N. Mamoulis. Efficient answering of set containment queries for skewed item distributions. In *Proc. of the Intl. Conf. on Extending Database Technology*, 2011.

[42] M. Terrovitis, J. Liagouris, N. Mamoulis, and S. Skiadopoulos. Privacy by disassociation. *Submitted*.

[43] M. Terrovitis and N. Mamoulis. Privacy Preservation in the Publication of Trajectories. In *Proc. of the Intl. Conf. on Mobile Data Management*, 2008.

[44] M. Terrovitis, N. Mamoulis, and P. Kalnis. Privacy-preserving Anonymization of Set-valued Data. *Proc. of the VLDB Endowment*, 1(1), 2008.

[45] M. Terrovitis, N. Mamoulis, and P. Kalnis. Local and global recoding methods for anonymizing set-valued data. *The Intl. Journal on Very Large Data Bases*, 2011.

[46] G. Tsatsanifos, D. Sacharidis, and T. Sellis. Midas: Multi-attribute indexing for distributed architecture systems. In *Proc. of the Intl. Symposium on Spatial and Temporal Databases*, 2011.

[47] G. Tsatsanifos, D. Sacharidis, and T. Sellis. On enhancing scalability for distributed rdf/s stores. In *Proc. of the Intl. Conf. on Extending Database Technology*, 2011.

[48] T. Vergoulis, T. Dalamagas, D. Sacharidis, and T. Sellis. Approximate regional sequence matching for genomic databases. *The Intl. Journal on Very Large Data Bases*, 2012.