SUMMARY OF RESEARCH ACTIVITIES

PROF. PANOS VASSILIADIS

SUMMARY

Prof. Panos Vassiliadis received his Diploma in Electrical Engineering and his PhD in Computer Science from the National Technical University of Athens in 1995 and 2000, respectively. In 2002, Prof. Vassiliadis joined the Department of Computer Science of the University of Ioannina. In the context of his teaching and supervising duties, Prof. Vassiliadis has taught several programming, software engineering and database courses at both the undergraduate and the graduate level and (co)supervised several students at the PhD, MSc and diploma thesis level. Prof. Vassiliadis has conducted research in the general area of data management with a notable focus on data warehousing (metadata management, OLAP, and quite emphatically, ETL) for long. He has also worked on top-k and context-based querying, as well as in the areas of software engineering and web services. Following a common thread in his work, his current research focuses on the uncovering of the mechanisms of database evolution and the introduction of rigorous models of data, software and their interdependence for the design, visualization and evolution management of data-intensive software ecosystems. Prof. Vassiliadis has several publications in international journals and conferences, including co-editing the book “Fundamentals of Data Warehouses” (Springer). Moreover, he has an active service to the scientific community as a reviewer and PC chair.

More information is available at http://www.cs.uoi.gr/~pvassil.

RESEARCH ACTIVITIES

Since his PhD studies, Prof. Vassiliadis has worked in the area of data warehousing; the main corpus of Prof. Vassiliadis’ early work (and the most well-known, at the same time) can be found in this area, and specifically in the rigorous definition of metadata models, and the modeling and optimization of ETL data flows. The latest developments in this area concern the re-definition of OLAP in terms of (a) redefining the traditional OLAP operators with operators operating at a higher level of abstraction and (b) data storytelling, presenting richer results to users. Since the late ‘00s, Prof. Vassiliadis has been pursuing a research agenda in the area of data-intensive ecosystems, with increasing emphasis over the years, esp., on the parts of uncovering the links between software applications and database, their visualization, the measurement of the quality of this interrelationship and ultimately the management of their evolution. Lately, in the ‘10s, Prof. Vassiliadis has been involved in a line of work that extracts patterns and internal mechanisms behind schema evolution in databases. In parallel to these major lines of research, Prof. Vassiliadis has collaborated with colleagues and students in the Univ. of Ioannina in the areas of software engineering, web services, top-k views and p2p systems. Finally, a fairly large body of research results has been produced in the general area of data management, in a variety of topics.

RESEARCH IN DATA WAREHOUSING

Metadata Management and Quality in Data Warehouses. Prof. Vassiliadis' involvement with the area of Data Warehouses starts back in the late '90s , and during his PhD studies, when the research area of data warehousing was at its infancy. The first topic that Prof. Vassiliadis explored concerned the problem of managing meta-information and quality in data warehouses. The results of Prof. Vassiliadis' research were published in a series of highly cited papers in CAiSE and Information Systems between 1998 and 2001 and the emphasis was posed on the introduction of a comprehensive and coherent model for managing metadata and quality in a principled fashion for data warehouses. This line of research was completed with (a) the PhD thesis of Prof. Vassiliadis, and, (b) one of the first books on data warehousing, "Fundamentals of Data Warehouses", edited by Springer-Verlag, which was co-edited by Prof. Vassiliadis.

OLAP. Prof. Vassiliadis has also been active in the area of On-Line Analytical Processing (OLAP). Specifically, two highly cited results include (a) a survey of available models published in SIGMOD record (1999), and, (b) the first model for multidimensional data and operations that could formally define all the popular OLAP operations – and especially drill-down – (SSDBM 1998, CAiSE 2000). Later, during the 00's, as a co-supervisor of A. Maniatis' PhD (supervisor: Y. Vassiliou), Prof. Vassiliadis was involved in the extension of the model with visualization metadata (resulting in the so-called Cube Presentation Model) that facilitates the visualization of data cubes in non-traditional ways (by adopting the Table Lens technique, that emphasizes interesting parts of the presented data).

In the mid '10s, Prof. Vassiliadis revisited the area of OLAP and Business Intelligence with the idea of *data storytelling* (DOLAP 2013, Inf. Systems 2015). Data storytelling, supported by the Cinecubes prototype system proved that it is possible to aid analysts starting with an original query to receive a story, packaged as a "data movie" that encompasses results of auxiliary queries, results of data analytics called "highlights", as well as automatic text and audio extraction. The redefinition of OLAP operators like roll-up and drill-down with more abstract operators, like, e.g., explain, predict, etc., has been proposed in DOLAP 2018.

Latest results also include the mining of hierarchical relationships in RDF data (EDBT 2016, Inf. Systems 2018).

ETL. One of the most well-known areas of Prof. Vassiliadis research is the area of data warehouse maintenance via Extract-Transform-Load (ETL) workflows. Since the early 00's, Prof. Vassiliadis has been involved in establishing the groundwork and expanding this research area. The goal has been to provide methods that (a) allow administrators to design ETL workflow at conceptual and logical levels, (b) implement and tune these workflows at the physical level, and, (c) come up with efficient algorithms that can be incorporated in ETL tools to allow the efficient execution of ETL workflows. Prof. Vassiliadis' research contributed towards several results, and specifically, (a) conceptual and logical models (DOLAP 2002, CAiSE 2003, ER 2004, Information Systems 2005 to name a few highly cited publications); (b) design automation (DSS 2008) and taxonomies for ETL activities (DOLAP 2009), (c) efficient algorithms for ETL optimization (ICDE 2005, TKDE 2005), efficient lookup joins for streaming incoming data (ICDE 2007, TKDE 2008), scheduling for ETL workflows (QDB 2011, Inf. Systems 2013), and benchmarking (DOLAP 2007, TPC-TC 2009). Prof. Vassiliadis has co-authored the lemma on ETL in the Springer Encyclopedia of Databases (2009), several lemmas and chapters on the topic in other encyclopedias and collections, and a survey for ETL research efforts (IJDWM 2010). In the early stages of this research effort, Dr. A. Simitsis obtained his PhD, co-supervised by Prof. Vassiliadis (supervisor: T. Sellis). The cooperation of Dr. Simitsis with Prof. Vassiliadis has continued afterwards with a significant amount of results.

MODELING, VISUALIZATION AND EVOLUTION OF DATA-INTENSIVE ECOSYSTEMS

One of the main areas of Prof. Vassiliadis' research since the late 00's has been the modeling, visualization and evolution of data-intensive ecosystems. Data-intensive ecosystems involve the capturing and modeling of complex, heterogeneous configurations of databases along with the applications that interact with them. Prof. Vassiliadis has introduced the idea of *Architecture Graphs* (CAiSE 2006, JODS 2009) to model such ecosystems. Architecture Graphs include both the internals of the databases of the ecosystem (relations, views, constraints) as well as the graph-theoretic modeling of client applications that access the databases. In the latter case, software modules that interoperate with underlying databases are characterized as collections of queries and each query is represented by a graph that reflects both the query semantics and the structural dependence of the query from the underlying database constructs. This modeling facilitates several design and administration tasks; the research results that reflect this ability can be briefly listed as follows:

- Architecture Graphs can be annotated via a declarative language (ICEIS 2008) in order to *tune the reaction of the ecosystem to changes via policies*: the main idea is that we can pre-register our agreement or veto to potential changes and regulate the flooding of the impact of a change to the database (JoDS 2009, ER 2013, JoDS 2015).

- Architecture Graphs can be used for *assessing the design quality* (expressed as the likelihood to suffer from evolution events) of such ecosystems (ER 2008, JoDS 2012), by measuring not only direct, but also transitive dependence of client modules from the underlying databases.

- Architecture Graphs can be used for the *visualization* of such ecosystems with different methods –primarily concentric circles or arcs (DaLi 2011, ER 2014).

Hecataeus, a tool for managing the above tasks (CSMR 2008, ICDE 2010) has been the main platform for supporting the research in this area. Prof. Vassiliadis has initiated the exploration of this research area that has so far led to a PhD thesis by G. Papastefanatos (supervisor: Y. Vassiliou, co-supervised by Prof. Vassiliadis) and an on-going PhD thesis, by P. Manousis, in Univ. Ioannina.

SCHEMA EVOLUTION

During the mid '10s, Prof. Vassiliadis actively pursued a research agenda towards uncovering the internal mechanics of schema evolution, and specifically patterns, regularities and reasons behind schema evolution events. The effort was facilitated by the availability of a set of schema histories of database-encompassing open-source software projects that were collected and analyzed.

In CAiSE 2014 and Inf. Systems 2015, the compatibility of database schema evolution with Lehman's laws was studied: the results showed that whereas the essence of Lehman's laws holds, the specific mechanics have important differences when it comes to schema evolution. In CAiSE 2016, in a similar line of research, the lives of web services were investigated for the same laws, with similar findings. In ER 2015 & Inf. Systems 2017, the object of study shifted from entire schemata to individual tables. This line of research continued with papers in CAiSE 2017 & JoDS 2017. These papers were the first to investigate how evolution-related properties, like the possibility of deletion, life duration, or the amount of updates that a table undergoes, are related to observable table properties like the number of attributes or the version of birth of a table. The evolution of foreign keys was first investigated in ER 2017. Out of the several patterns observed in all these works, we highlight a guiding force, "gravitation to rigidity", i.e., the striking scarcity of maintenance and evolution activities in the lives of schemata, which is attributed to their potential impact to the code that surrounds the database.

SOFTWARE ENGINEERING & WEB SERVICES

Prof. Vassiliadis has cooperated for long Prof. A. Zarras, leading the Software Engineering group of the Department and this collaboration has produced research results in several areas.

Web service and middleware management. DOA 2004 and IJBPIM 2008 present a mechanism to assess the reliability of workflows of web services via Block Diagrams and Markov Models. DOA 2005 presents a scheduling mechanism that decides the acceptance or rejection of user requests for mobile peers on the basis of the server load, the lifetime of servers and clients and the periodicity of the client requests. SEIW 2007 and JDM 2010 concern the deployment of workflows of web services to clusters of servers taking response time and load balancing into consideration on the basis of different server topologies. IJACI 2010 presents and extension of the traditional 2 Phase Commit mechanism for the case of peers that export their schedules and itineraries to each other. ICSE 2011 presents a method for clustering web services into "service abstractions" to facilitate the construction of easily maintainable client programs that are exploit the presence of cluster representatives and inter-cluster mappings for their evolution and adaptation in the presence of changes. FSE 2012 complements this structuring with groupings of services according to their quality characteristics.

Refactoring Interfaces. IEEE TSC 2015 proposes a method that progressively decomposes a given web service interface into more cohesive interfaces via a suite of cohesion metrics that rely solely on the specification of the service interface without access to its internals. ICSME 2015 provides an interface extraction method that accommodates both the need for fitness to client needs and flexibility to cope with client evolution. A principled method for organizing and visualizing refactoring patterns is presented in FSE 2015, where Fowler's refactoring patterns are organized in a map, called *map of the archipelago of refactorings,* which organizes the possible refactorings in families and also links them with succession, part-of or alternation relationships.

RESEARCH IN THE BROADER FIELD OF DATA MANAGEMENT

Apart from the aforementioned major areas of interest, Prof. Vassiliadis has contributed to research results in the field of data management. Prof. Vassiliadis has cooperated for long with the E. Pitoura, leading the Laboratory for the Distributed Management of Data in the areas of preference-based, context-aware and p2p-based query processing.

View management for top-k queries. Top-k query processing involves the efficient answering of ranking queries, i.e., queries whose results are ranked according to a ranking function. Typically, this function presents a linear equation for the score, over the attributes of the queried relation. Out of all the tuples of a query result, we are interested in returning only a subset of a fixed cardinality (k), with the tuples having the highest score. To efficiently answer top-k queries, Prof. Vassiliadis' research has pursued the idea of exploiting materialized views over the queried relation. The produced research results include theoretical guarantees on when a materialized view can suffice to answer a query, along with algorithms for this task (DOLAP 2009), as well as techniques for addressing the problem of top-k materialized view refreshment, especially in the presence of bursts (MPref 2007, DAPD 2010). Close to these deliberations, the exploration of the problem of cube similarity (ICDE 2011) was conducted: in this research effort, a taxonomy of distance functions for data in a hierarchical, multidimensional space along with a user study have been presented. An interesting result of this effort has been the discovery that although users prefer hierarchically-based distances for individual values, their preference is less strong and possibly ambivalent to Hausdorff distances when it comes to sets of cube cells (i.e., cuboids). This line of research resulted in the successful completion of the PhD of Ms. E. Baikousi, supervised by Prof. Vassiliadis in the Univ. of Ioannina.

Preference-based and context aware query processing. Query personalization involves adapting queries to the preferences of users. In the context of K. Stefanidis' PhD (supervised by E. Pitoura), in the Univ. of Ioannina, Prof. Vassiliadis has contributed in the modeling of user preferences under a multidimensional, hierarchical view was pursued (MCMP 2005, JPCC 2007). This allows the modeling of user preferences with more expressiveness. This modeling facilitates the possibility of introducing context-aware query reformulation on the basis of user preferences. Specifically, the idea is that (a) preferences are expressed with respect to specific contexts, (b) each time a query occurs, its current context is resolved, and, (c) once this is achieved, the query is rewritten accordingly (ICDE 2007, IS 2011). The above has also been considered in the context of approximate results (ADBIS 2006).

Miscellaneous endeavors. Finally, there are several contributions in various lines of research: ontology-based query rewriting (DOLAP 2017, Inf. Systems 2018), privacy and data publishing (SSDBMS 2012), peer-to-peer query processing (Pervasive and Mobile Computing 2008, ICC 2007, Computer Networks 2012), indexing and query processing over set-valued attributes (CIKM 2006, EDBT 2011), management of cardinal relations (EDBT 2004, TKDE 2005), and pattern-base management (ER 2003, SSDBM 2004, DKE 2006).

RESEARCH INTERESTS & FUTURE PLANS

1. Evolution of data and software is in the epicenter of Prof. Vassiliadis research at the moment. From the scientific point of view, research is focused on the study of the available schema histories for the extraction of patterns, mechanisms and reasons of change. From the engineering point of view, the management of evolution via the rigorous modeling of data-intensive ecosystems is another line of research that aims at the invention of methods that allow us to design these ecosystems on the grounds of metrics and patterns, and manage their evolution effectively and efficiently.

2. OLAP & data analytics, where novel OLAP operators are combined with data storytelling methods over both traditional and non-traditional data.

3. No-DB Querying, which involves the querying over scientific data collections over the raw files, without the usage of a DBMS, along with the appropriate visualization of results.

PROFESSIONAL ACTIVITIES

Prof. Vassiliadis has served as a reviewer for various journals (including VLDBJ, Information Systems, IEEE TKDE, DKE, DAPD, DSS, JODS, IST and others) and as a PC member in numerous conferences. Prof. Vassiliadis is a member of the Editorial Board of DKE since 2018. Prof. Vassiliadis has also served as a PC chair for DMDW 2003, ParMa 2004 and 2006, DOLAP 2006, HDMS 2009, QDB 2010 and a track co-chair for ICDE 2009 ("DW, OLAP and data grids" track) and CIKM 2009 in the "Knowledge Management" Track (responsible for the areas "Data pre- and post-Processing", "Information Filtering and Recommender Systems", "Knowledge Synthesis and Visualization"). Moreover, he has served as Workshops co-chair for ER 2012 and Tutorials co-chair for ER 2016.

In parallel to his service in several Program Committees and Journals, Prof. Vassiliadis has served in several PhD committees, as an expert in data warehousing and ETL in particular, both in Greece and internationally. He is also a member of the steering committee of the DOLAP workshop since 2006, after his service as PC chair.