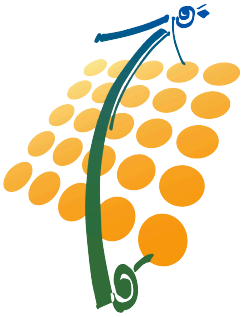

Sixth Framework Programme
Information Society Technologies
Future Emerging Technologies

Project Acronym:

GeoPKDD

Project full title:

Geographic Privacy-aware Knowledge Discovery and Delivery

Project Number:FP6-014915
Instrument: Specific Target Research Project

Project Deliverable D3.6

Consolidated report on geographic knowledge interpretation and delivery

Date of preparation: M6
Revised: M6
Revision: final

Operative commencement date of Contract: 01/11/2005
Project Coordinator: Fosca Giannotti
Project institute coordinator: Knowledge Discovery and Delivery-LAB / ISTI-CNR

Programme Name: .....IST
Project Number:.....014915
Project Title: .....GEOPKDD

Document Number: D3.6
Work-Package:.....WP1
Document Type: .....Document
Contractual Date of Delivery: M6
Actual Date of Delivery: M6
Title of Document: .....Technical Report
.....Deliverable 3.6 – Deliverable 3.6 – Consolidated Report on geographic knowledge interpretation and delivery

Author(s): .....Gennady Andrienko (FAIS),
..... Chiara Renso (ISTI-CNR),
..... Jose Macedo (EPFL),
..... Monica Wachowicz (Univ. Polytechnic Madrid)

Dissemination Level..... Public

GeoPKDD consortium partecipants

Partic. Role	Partic. N.	Participant name	Short name	Country
CO	1	**KDD Lab.** joint research group of **ISTI-CNR**, Istituto di Scienza e Tecnologie dell'Informazione, Pisa. http://www-kdd.isti.cnr.it/ and **Univ. Pisa**, Dept. of Computer Science http://www.di.unipi.it	**KDDLAB**	I
CR	2	**Univ. of Hasselt**, Theoretical Computer Science Group. http://alpha.uhasselt.be/research/groups/theocomp/	**HASSELT**	B
CR	3	**EPFL - Ecole Polytechnique Fédérale de Lausanne**, Lab. DB, Lausanne. http://lbdwww.epfl.ch/e/ **and** University of Milan - Computer Science and Communication Department http://www.dico.unimi.it/	**EPFL**	CH
CR	4	**Fraunhofer Institute for Autonomous Intelligent Systems**, Sankt Augustin. http://www.ais.fraunhofer.de/	**FAIS**	D
CR	5	**Wageningen UR,** Centre for GeoInformation. http://cgi.girs.wageningen-ur.nl/	**WUR**	NL
CR	6	**Research Academic Computer Technology Institute,** Research and Development Division. http://www.cti.gr/ and Univ. Piraeus, Dept. of Informatics http://www.unipi.gr	**CTI**	GR
CR	7	**Sabanci University**, Faculty of Engineering and Natural Sciences. http://www.sabanciuniv.edu/	**UNISAB**	TK
CR	8	**WIND Telecomunicazioni SpA,** Direzione Reti Wind Progetti Finanziati & Technology Scouting. http://www.wind.it	**WIND**	I
CR	9	**Polytechnic University of Madrid ,** Latingeo Laboratory http://www.latingeo.net	**UPM**	ES

Table of contents

1. DATA MINING QUERY LANGUAGE TASK

In this section we collected the main publications which have been produced within GeoPKDD data mining query language task.

The main contribution of the DMQL task within GeoPKDD is the introduction of DAEDALUS framework, First of all, a formal model to support progressive mining and querying, in the specifc context of geographic knowledge discovery, is proposed. Secondly, an effective implementation of the system has been done. DAEDALUS is based on the powerful object-relational formalism and relying on emergent technology on Moving Object Database, thus providing for free a number of basic primitives to manage spatio-temporal data. In this approach, to the best of our knowledge, we take a first step towards progressively mining and querying movement data. In this direction, we propose an innovative computational environment that provides effective support to the whole knowledge discovery process. DAEDALUS implements IAPYX data mining query language capable of supporting the user in specifying and refining mining objectives. The IAPYX language is based on an algebraic framework, called 2W Model, capable of accommodating and combining disparate mining tasks into a multi-step knowledge discovery process. DAEDALUS operates as a statement execution layer on top of the Hermes moving-object database.

The papers collected here are:

ACMGIS 2008: R. Ortale, E. Ritacco, N. Pelekis, R. Trasarti, G. Costa, F. Giannotti, G. Manco, C. Renso, Y. Theodoridis The DAEDALUS Framework: A knowledge discovery analysis framework for movement data ACMGIS 2008, Irvine, USA, November 2008.

GeoPKDD_TR_WP3_20_2009: R. Ortale, R. Trasarti, F. Giannotti, G. Manco, C. Renso, A framework for progressive mining and querying of movement data. Submitted to IDA 2009. GeoPKDD Report N. GeoPKDD_TR_WP3_20_2009

SEBD2008: R. Ortale, E. Ritacco, N. Pelekis, R. Trasarti, G. Costa, F. Giannotti, G. Manco, C. Renso, Y. Theodoridis DAEDALUS: A knowledge discovery analysis framework for movement data SEBD 2008, Palermo, Italy.

In the following we report the abstract of each paper, whereas the full versions of the papers are available separately for private download.

ACMGIS 2008: R. Ortale, E. Ritacco, N. Pelekis, R. Trasarti, G. Costa, F. Giannotti, G. Manco, C. Renso, Y. Theodoridis The DAEDALUS Framework: A knowledge discovery analysis framework for movement data ACMGIS 2008, Irvine, USA, November 2008.

In this work we propose DAEDALUS, a formal framework and system, specifically focussed on progressive combination of mining and querying operators. The core component of DAEDALUS is the MO-DMQL query language that extends SQL in two respects, namely a pattern definition operator and the capability to uniform manipulating both raw data and unveiled patterns. DAEDALUS system is specifically focussed on movement data and has been implemented as a query execution layer on top of the Hermes Moving Object Database. The expressiveness and usefulness of the MODMQL language as well as the computational capabilities of DAEDALUS are qualitatively evaluated by means of a case study.

GeoPKDD_TR_WP3_20_2009: R. Ortale, R. Trasarti, F. Giannotti, G. Manco, C. Renso, A framework for progressive mining and querying of movement data. Submitted to IDA 2009. GeoPKDD Report N. GeoPKDD_TR_WP3_20_2009

In this work we propose a Data Mining Query Language framework, called DAEDALUS, specifically focussed on progressive combination of mining and querying movement data. The framework introduces an algebraic model, called 2W Model, capable of accommodating and combining disparate mining tasks into a multi-step knowledge discovery process. The core component of DAEDALUS is the IAPYX query language that extends SQL in two respects, namely a pattern definition operator and the capability to uniform manipulate both raw data and unveiled patterns. The expressiveness and usefulness of the IAPYX language as well as the computational capabilities of DAEDALUS are qualitatively evaluated by means of a case study.

SEBD2008: R. Ortale, E. Ritacco, N. Pelekis, R. Trasarti, G. Costa, F. Giannotti, G. Manco, C. Renso, Y. Theodoridis DAEDALUS: A knowledge discovery analysis framework for movement data

In this work we propose DAEDALUS, a knowledge discovery system that enables a uniform querying of movement data and patterns. Such system is founded on a formal framework that defines knowledge discovery execution process as a progressive combination of mining and querying operators. The core component of DAEDALUS is the query language that extends SQL in two respects, namely a pattern definition operator and the capability of uniformly manipulate both raw data and unveiled patterns. DAEDALUS is specifically focussed on movement data and has been implemented as a query execution layer on top of the Hermes Moving Object Database. A qualitative evaluation of the system will show the power of DAEDALUS in progressive mining task over movement data.

2. SPATIO-TEMPORAL REASONING TASK

In the present section we collected the main publications which have been produced within GeoPKDD in the context of the reasoning task.

The basic idea developed during GeoPKDD is to define an approach, concretized in the Athena system (SECOGIS2008,AGILE2009), to incrementally enrich raw trajectory data with context geoinformation. The first step is the definition of semantic trajectory as sequence of stops (movement suspension) and moves (the actual movement). A further enrichment step exploits the knowledge capabilities provided by the ontology to integrate context geoinformation with the semantic trajectory patterns. Eventually, a reasoning step allows to infer new knowledge (e.g. the inferred movement behaviour) that can be added back to the ontology. For example, semantic trajectories patterns showing that typically stops are happening in tourist places can be associated with the movement of tourists. In contrast, semantic trajectory patterns stopping at offices in the morning and residential areas in the evening can be identified as a home-work routing behaviour. This approach exploits the knowledge representation and reasoning capabilities of formal ontologies. In particular, Athena is based on Oracle11g Semantic Technologies.

The papers collected here are:

AGILE 2009: M. Baglioni, J.Macedo, C. Renso, R.Trasarti, M. Wachowicz Towards Semantic Interpretation of Movement Behavior Agile 2009, Hamburg, Germany.

SECOGIS 2008: M. Baglioni, J.Macedo, C. Renso, M. Wachowicz An ontology-based approach for the semantic modelling and reasoning on trajectories SeCOGIS 2008, Barcelona, LNCS 5232.

MMV08: Monica Wachowicz, Daniel Orellana, Chiara Renso, Estefania Muoz Moraga, and Javier Parada. The spatial knowledge representation of players movement in mobile outdoor gaming. The Fourth International Conference on Monitoring and Management of Visitor Flows in Recreational and Protected Areas, Montecatini Terme, october 2008

GEOKDD09: Monica Wachowicz, Jose Macedo, Chiara Renso, and Arend Ligtenberg The role of a multi-tier ontological framework in reasoning to discover meaningful patterns of sustainable mobility in Geographic Data Mining and Knowledge Discovery, 2nd Edition

WISC08: Zhixian Yan ; Jose Macedo, Christine Parent, Stefano Spaccapietra, Trajectory Ontologies, In: Terra Cognita Workshop In Conjunction with the 7th International Semantic Web Conference. p. 80-92, 2008

TGIS08: Zhixian Yan ; Jose Macedo, Christine Parent, Stefano Spaccapietra,Trajectory Ontologies and Queries, In Transactions on GIS Journal, v. 12, issue s1, p. 75-91, December, 2008

TGIS08B M.Baglioni, E. Giovannetti, M.V. Masserotti, C. Renso, L. Spinanti. Ontology-driven querying of geographical databases Transactions in GIS Vol 12 Suppl 1 December 2008.

The main thread of these papers is the definition of a methodology (finally concretized in a running system in AGILE09) for the semantic enrichment of trajectories and mining patterns, accompanied by the automatic reasoning capabilities provided by the ontology engine of Oracle 11g.

Particularly, paper SECOGIS 2008 introduces the preliminary idea concerning the semantic enrichment process based on the concept of semantic trajectory, further enriched with a domain ontology expressing the basic background knowledge.

Paper AGILE 2009 improves previous work by defining in a systematic way, the semantic enrichment process for both trajectories and mining patterns. Furthermore, the methodology has been proved by means of an implemented system, called Athena, based on Oracle 11g Semantic Technologies.

As a collateral work, arising from the main thread ideas, paper MMV08 reports on an approach for developing a spatial knowledge representation based on the existence of multi tier spaces as a mental construction of human movement. The spatial knowledge representation was implemented as a computational ontology in Protégé, and it has been applied to provide new insight about the actual behavioural patterns of players within a recreation site, accordingly to checkpoints and similar players´ interactions.

Chapter GEOKDD09 proposes a multi-tier ontological framework to support a GKDD process. Three ontological tiers are described to provide the common base for the organisation of different nature and sources of knowledge as well as the reasoning tasks integrated within a spatio-temporal database. The potential of this approach is illustrated on reducing the semantic gap between an ontological tier and a database representation by using mappings between the ontology and conceptual model, and permitting to define spatio-temporal relationships within an ontology, which can be translated to spatio-temporal conceptual queries.

In MMV8, the reasoning approach is applied to an experiment consisted of an educational game in Amsterdam using mobile phones and GPS-technology for 200 students having the age of 12-14. The results demonstrate that different types of inferences play a different role accordingly to what a recreational planner needs to infer, that is, the location of interactions among players and the environment.

The Chapter in the book GEOKDD09 proposes a multi-tier ontological framework to support a GeoPKDD process. Three ontological tiers are described to provide the common base for the organisation of different nature and sources of knowledge as well as the reasoning tasks integrated within a spatio-temporal database. They are: Domain, Application and Data Ontology Tiers. The potential of this approach is illustrated on (a) reducing the semantic gap between an ontological tier and a database representation by using mappings between the ontology and conceptual model, and (b) permitting to define spatio-temporal relationships within an ontology, which can be translated to spatio-temporal conceptual queries.

In WISC08 and TGIS08 we described a framework for a semantic-oriented structuring, modeling and querying of trajectory data that relies on Oracle semantic technology. The framework counts on the definition of trajectory-related ontologies, addressing domain-independent and application-specific geometric and semantic facets. This work is more oriented towards trajectory knowledge integration and modularization.

Paper TGIS08B introduces a methodology to extract and enrich geospatial ontologies from geographical databases in a context of enhanced query answering system.

The abstracts of the papers follows:

SECOGIS 2008: M. Baglioni, J. Macedo, C. Renso, M. Wachowicz An ontology-based approach for the semantic modelling and reasoning on trajectories SeCOGIS 2008, Barcelona, LNCS 5232.

In this paper we present a methodology for the semantic enrichment of trajectories. The objective of this process is to provide a semantic interpretation of a trajectory in term of behaviour. This has been achieved by enhancing raw trajectories with semantic information about moves and stops and by exploiting some domain knowledge encoded in an ontology. Furthermore, the reasoning mechanisms provided by the OWL ontology formalism have been exploited to accomplish a further semantic enrichment step that puts together the different levels of knowledge of the domain. A final example application shows the added power of the enrichment process in characterizing people behaviour.

MMV8: Monica Wachowicz, Daniel Orellana, Chiara Renso, Estefania Muñoz Moraga, and Javier Parada The spatial knowledge representation of players movement in mobile outdoor gaming

This paper describes an innovative approach for developing a spatial knowledge representation based on the existence of multi tier spaces as a mental construction of human movement. The three "spaces" paradigm has been proposed to support the reasoning process in terms of sensing, symbolic, and social spaces. The spatial knowledge representation was implemented as a computational ontology in Protégé, and it has been applied to provide new insight about the actual behavioural patterns of players within a recreation site, accordingly to checkpoints and similar players´ interactions. This first experiment consisted of an educational game in Amsterdam using mobile phones and GP technology for 200 students having the age of 12-14. The results demonstrate that different types of inferences play a different role accordingly to what a recreational planner needs to infer, that is, the location of interactions among players and the environment.

GKDD09 Monica Wachowicz , Jose Macedo , Chiara Renso, and Arend Ligtenberg. The role of a multi-tier ontological framework in reasoning to discover meaningful patterns of sustainable mobility

The successful applications of Geographical Knowledge Discovery in Databases (GKDD) are not common, despite the vast literature on knowledge discovery in databases. Although it is relatively straightforward to find patterns in very large spatio-temporal databases, establishing their relevance and explaining their causes are both very complex problems. In practice, spatiotemporal databases are not adequate to handle geographical knowledge in an ad-hoc manner, and as a result, most of the patterns found in a GKDD process may well already be background knowledge, which refers to the common sense reasoning of a geographical knowledge domain. Addressing these issues requires considering a geographical knowledge discovery process as a multitier ontological process, in the sense that more complex reasoning modes can be used to help the comprehension of what makes one pattern structurally and meaningfully different from another. Towards this goal, this Chapter proposes a multi-tier ontological framework to support a GKDD process. Three ontological tiers are described to provide the common base for the organisation of different nature and sources of knowledge as well as the reasoning tasks integrated within a spatio-temporal database. They are:

Domain, Application and Data Ontology Tiers. The potential of this approach is illustrated on (a) reducing the semantic gap between an ontological tier and a database representation by using mappings between the ontology and conceptual model, and (b) permitting to define spatio-temporal relationships within an ontology, which can be translated to spatio-temporal conceptual queries. The implementation has been carried out as a proof-of-concept and the specific information metaphor of movement-as-trajectory has been used to illustrate the implementation of the Data Ontology Tier. The preliminary results are pointing out that geographical knowledge of sustainable mobility must come from a global and systemic view of patterns within a GKDD process, and the important role of a multi-tier ontological framework on the integration of semantic abstractions (concepts), reasoning tasks, and patterns. They had also drawn attention to the fact that combining ontological representation with database querying mechanisms is fundamental for the use information metaphors in GKDD processes.

AGILE2009: Miriam Baglioni, José Antônio Fernandes de Macêdo, Chiara Renso, Roberto Trasarti, Monica Wachowicz. Towards Semantic Interpretation of Movement Behavior

In this paper we aim at providing a model for the conceptual representation and deductive reasoning of trajectory patterns obtained from mining raw trajectories. This has been achieved by means of a semantic enrichment process, where raw trajectories are enhanced with semantic information and integrated with geographical knowledge encoded in an ontology. The reasoning mechanisms provided by the chosen ontology formalism are exploited to accomplish a further semantic enrichment step that gives a possible interpretation of discovered patterns in terms of movement behaviour. A sketch of the realised system, called Athena, is given, along with some examples to demonstrate the feasibility of the approach.

WISC08: Zhixian Yan ; Jose Macedo, Christine Parent, Stefano Spaccapietra, Trajectory Ontologies.

Many real world applications build today on analyses of movement and related features. Examples of such applications include transportation management, urban planning, tourism services, and animal migration monitoring, just to name a few. Recent database modeling and management research prototypes have the capability to store and manipulate movement data in terms of point or region geometries that evolve over time (moving point or moving and deforming region). This captures the spatio-temporal trace left by a moving object, but ignores its links with non-geometric information that enable a semantic interpretation of the movement of moving objects. The concept of trajectory has been introduced to express a more semantic understanding of movement, taking it closer to the perception of applications. This paper describes a framework for a semantics-oriented structuring, modeling and querying of tra jectory data. The framework relies on the definition of trajectory-related ontologies, addressing domain-independent and application-specific geometric and semantic facets. A system architecture to implement the framework is also proposed.

TGIS08: Zhixian Yan, Jose Macedo, Christine Parent, Stefano Spaccapietra, Trajectory Ontologies and Queries

Many real world applications today are built on analyses of movement and related features. Examples of such applications include transportation management, urban planning, tourism services, and animal migration monitoring, just to name a few. Recent database modeling and management research prototypes have the capability to store and manipulate movement data in terms of point or region geometries that evolve over time (moving point or moving and deforming region).

This captures the spatio-temporal trace left by a moving object, but ignores its links with non-geometric information that enable a semantic interpretation of the movement of moving objects. The concept of trajectory has been introduced to express a more semantic understanding of movement, taking it closer to the perception of applications. This paper describes a framework for a semantics-oriented structuring, modeling and querying of trajectory data. The framework relies on the definition of trajectory-related ontologies, addressing domain-independent and application-specific geometric and semantic facets. Last we briefly discuss how the proposed approach has been applied for a traffic management application.

TGIS08B: Miriam Baglioni, Maria Vittoria Passerotti, Laura Spinanti, Emiliano Giovanetti, Chiara Renso Ontology-supported Querying of Geographical Databases

Querying geographical information systems has been recognized as a difficult task for non-expert users. Furthermore, user queries are often characterized by semantic aspects not directly managed by traditional spatial databases or GIS. Examples of such semantic geospatial queries are the use of implicit spatial relations between objects, or the reference of domain concepts not explicitly represented in data. To handle such queries, we envisage a system that translates natural language queries into spatial SQL statements on a database, thus improving standard GIS with new semantic capabilities. Within this general objective, the contribution of this article is to introduce a methodology to handle semantic geospatial queries issued over a spatial database. This approach captures semantics from an ontology built upon the spatial database and enriched by domain concepts and properties specifically defined to represent the localization of objects. Some examples of the use of the methodology in the urban domain are presented.

3. GEOVISUALIZATION AND VISUAL ANALYTICS TASK

Within this task, a general visual analytics framework for trajectory analysis has been proposed (earlrly version has been described in SIGKDD07). The framework is supported by the visual analytics toolkit which includes visually-driven database aggregation of trajectories (VAST08, extended versions submitted to The Cartographic Journal), progressive clustering methods (IVS08) based on various distance functions (TIME07), cluster-based classification (VAST09 submitted), and analysis of interactions (GIScience08). The toolkit has been widely used within the project consortium for analysis of various data sets in combination with several database and data mining methods.

The papers collected here are:

VAST08. Gennady Andrienko, Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In IEEE Visual Analytics Science and Technology (VAST 2008). Proceedings, IEEE Computer Society Press, 2008, pp.51-58

VAST08a. Natalia Andrienko, Gennady Andrienko. Evacuation Trace Mini Challenge Award: Tool Integration. Analysis of Movements with Geospatial Visual Analytics Toolkit. In IEEE Visual Analytics Science and Technology (VAST 2008) Proceedings, IEEE Computer Society Press, 2008, pp.205-206

GIScience08. Natalia Andrienko, Gennady Andrienko, Monica Wachowicz, Daniel Orellana. Uncovering Interactions between Moving Objects. In Thomas J. Cova, Harvey J. Miller, Kate Beard, Andrew U. Frank, Michael F. Goodchild (Eds.) GIScience, 5th international conference, Proceedings, pp.16-26

Springer08c. Gennady Andrienko and Natalia Andrienko. A Visual Analytics Approach to Exploration of Large Amounts of Movement Data. In M. Sebillo, G. Vitiello, and G. Schaefer (Eds.) Visual Information Systems, 10th International Conference, VISUAL 2008, Volume 5188 of LNCS, Springer, 2008, pp.1-4

AVI08. Natalia Andrienko, Gennady Andrienko. Supporting Visual Exploration of Massive Movement Data. In Working Conference on Advanced Visual Interfaces AVI 2008. Proceedings, ACM Press, 2008, pp.474-475

Springer08b. Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melancon. Visual Analytics: Definition, Process, and Challenges. In Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.) Information Visualization – Human-Centered Issues and Perspectives. Volume 4950 of LNCS State-of-the-Art Survey, Springer, 2008, pp.154-175

IVS08a. Gennady Andrienko, Natalia Andrienko, Jason Dykes, Sara Irina Fabrikant, Monica Wachowicz. Geovisualization of Dynamics, Movement and Change: Key Issues and Developing Approaches in Visualization Research. In Information Visualization, 2008, v.7 (3/4), pp. 173-180

In the following the paper abstracts:

IVS08a. Gennady Andrienko, Natalia Andrienko, Jason Dykes, Sara Irina Fabrikant, Monica Wachowicz. Geovisualization of Dynamics, Movement and Change: Key Issues and Developing Approaches in Visualization Research.

The work presented here represents a selection of the contributions made to a workshop coordinated by the *International Cartographic Association (ICA) Commission on Geovisualization* and the *Association of Geographic Information Laboratories in Europe (AGILE)* on the Geovisualization of Dynamics, Movement and Change. Theoretical and methodological approaches for exploring and analyzing large datasets with spatial and temporal components were presented, discussed and developed at the meeting in Girona, Catalunya which was held on 5th May 2008 one day before AGILE's 11th International Conference on Geographic Information Science.
The high level of interest raised by the open call for contributions and the ultimate participation of more than 40 scientists suggests that this theme is timely and of relevance to many researchers and research groups. It would also seem to indicate that spatiotemporal data pose plenty of interesting and unsolved research problems. The workshop, and the work subsequently reported here suggest that many of these are complex and can benefit from the application of cross-disciplinary approaches. Cross-disciplinarity has been reflected not only in the contents of the submissions, but also in the composition of the workshop delegates, which included scientists from a variety of nations with backgrounds in geography, geographic information science, information visualization, data mining and other cognate disciplines. We reflect upon some of these trends in this introduction to the papers. Integration of approaches from multiple disciplines is a characteristic feature of geovisualization - a research domain addressing the visual exploration, analysis, synthesis, and presentation of geographic data, information, and knowledge (Dykes et al. 2005). The *ICA Commission on Geovisualization* works to develop, promote and communicate advances in this multidisciplinary domain - http://geoanalytics.net/ica. One way of so doing is to attract researchers with various disciplinary backgrounds to themed workshops that showcase current multidisciplinary approaches whilst allowing participants to learn about relevant theories and methods existing in related fields. They also create new opportunities for considering problems from different perspectives, and for starting new cross-disciplinary collaborations.

IVS08b. Salvatore Rinzivillo, Dino Pedreschi, Mirco Nanni, Fosca Giannotti, Natalia Andrienko, Gennady Andrienko. Visually–driven analysis of movement data by progressive clustering.

The paper investigates the possibilities of using clustering techniques in visual exploration and analysis of large numbers of trajectories, i.e. sequences of time-stamped locations of some moving entities. Trajectories are complex spatio-temporal constructs characterized by diverse non-trivial properties. To assess the degree of (dis)similarity between trajectories, specific methods (distance functions) are required. A single distance function accounting for all properties of trajectories, first, is difficult to build, second, would require much time to compute, third, might be difficult to understand and to use. We suggest the procedure of progressive clustering where a simple distance function with a clear meaning is applied on each step, which leads to easily interpretable outcomes. Successive application of several different functions enables sophisticated analyses through gradual

refinement of earlier obtained results. Besides the advantages from the sense-making perspective, progressive clustering enables a rational work organization where time-consuming computations are applied to relatively small potentially interesting subsets obtained by means of "cheap" distance functions producing quick results. We introduce the concept of progressive clustering by an example of analyzing a large real dataset. We also review the existing clustering methods, describe the method OPTICS suitable for progressive clustering of trajectories, and briefly present several distance functions for trajectories.

AVI08. Natalia Andrienko, Gennady Andrienko. Supporting Visual Exploration of Massive Movement Data.

To make sense from large amounts of movement data (sequences of positions of moving objects), a human analyst needs interactive visual displays enhanced with database operations and methods of computational analysis. We present a toolkit for analysis of movement data that enables a synergistic use of the three types of techniques.

GIScience08. Natalia Andrienko, Gennady Andrienko, Monica Wachowicz, Daniel Orellana. Uncovering Interactions between Moving Objects.

Movement data resulting from tracking positions of various moving objects (people, animals, goods, etc.) have recently got close attention of researchers. Methods for analysis of movement data have been developed in the areas of geographic information science (Buliung and Kanaroglou 2004), geovisualization (Dykes and Mountain 2003), information visualization (Kapler and Wright 2005), data mining (Giannotti and Pedreschi 2007), and visual analytics (Andrienko et al. 2007). The methods intended for analysis of large datasets include computational techniques, which either aggregate and summarize the data (Dykes and Mountain 2003, Buliung and Kanaroglou 2004, Nanni and Pedreschi 2006, Andrienko et al. 2007) or extract specific features, e.g. occurrences of certain types of relationships between moving objects (Laube et al. 2005). Our recent research on analyzing movement data has lead towards the need to uncover the interactions between objects in the process of their movement. Movement data usually consist of time-stamped position records and do not contain any explicit information about interactions; hence, it is only possible to detect indications of possible interactions. One indication is spatial proximity between two or more objects at some time moment or during a time interval. The notion of spatial proximity depends of a number of factors; some of them are listed in Tab.1. Hence, the spatial proximity might be defined using a specific threshold of distances in each case of analysis. Furthermore, it may be insufficient to use only a distance threshold for identifying possible interactions. For example, it may be important to account for the duration of the spatial proximity between objects, or to look whether the objects stopped or continued their movement. The main research challenge relies on the definition of a theoretical and methodological framework for the detection and analysis of possible interactions between moving objects and between objects and the environment in which they move. This research challenge has two main foci: (1) the development of a theoretical foundation, including a formal definition of interaction and its indications, an inventory of factors influencing these interaction indications, and a typology of interactions; (2) the support of a methodology for visual exploration and analysis of possible interactions in large sets of movement data. This paper describes our first attempt towards uncovering interactions between moving objects by combining visual and filtering techniques.

Springer08b. Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melancon. Visual Analytics: Definition, Process, and Challenges.

We are living in a world which faces a rapidly increasing amount of data to be dealt with on a daily basis. In the last decade, the steady improvement of data storage devices and means to create and collect data along the way inflenced our way of dealing with information: Most of the time, data is stored without filtering and refinement for later use. Virtually every branch of industry or business, and any political or personal activity nowadays generate vast amounts of data. Making matters worse, the possibilities to collect and store data increase at a faster rate than our ability to use it for making decisions. However, in most applications, raw data has no value in itself; instead we want to extract the information contained in it. The information overload problem refers to the danger of getting lost in data which may be - irrelevant to the current task at hand - processed in an inappropriate way - presented in an inappropriate way Due to information overload, time and money are wasted, scientific and in- dustrial opportunities are lost because we still lack the ability to deal with the enormous data volumes properly. People in both their business and private lives, decision-makers, analysts, engineers, emergency response teams alike, are often confronted with massive amounts of disparate, conflicting and dynamic information, which are available from multiple heterogeneous sources. We want to simply and effectively exploit and use the hidden opportunities and knowledge resting in unexplored data sources. In many application areas success depends on the right information being available at the right time. Nowadays, the acquisition of raw data is no longer the driving problem: It is the ability to identify methods and models, which can turn the data into reliable and provable knowledge. Any technology, that claims to overcome the information overload problem, has to provide answers for the following problems: - Who or what defines the \relevance of information" for a given task?

- How can appropriate procedures in a complex decision making process be identified?
- How can the resulting information be presented in a decision- or task-oriented way?
- What kinds of interaction can facilitate problem solving and decision making?

With every new “real-life" application, procedures are put to the test possibly under circumstances completely different from the ones under which they have been established. The awareness of the problem how to understand and analyse our data has been greatly increased in the last decade. Even as we implement more powerful tools for automated data analysis, we still face the problem of understanding and \analysing our analyses" in the future: Fully-automated search, filter and analysis only work reliably for well-defined and well-understood problems. The path from data to decision is typically quite complex. Even as fully- automated data processing methods represent the knowledge of their creators, they lack the ability to communicate their knowledge. This ability is crucial: If decisions that emerge from the results of these methods turn out to be wrong, it is especially important to examine the procedures. The overarching driving vision of visual analytics is to turn the information overload into an opportunity: Just as information visualization has changed our view on databases, the goal of Visual Analytics is to make our way of processing data and information transparent for an analytic discourse. The visualization of these processes will provide the means of communicating about them, instead of being left with the results. Visual Analytics will foster the constructive evaluation, correction and rapid improvement of our processes and models and - ultimately - the improvement of our knowledge and our decisions. On a grand scale, visual analytics solutions provide technology that combines the strengths of human and electronic data processing. Visualization becomes the medium of a semi-automated analytical process, where humans and machines cooperate using their respective distinct capabilities for the most effective results.

The user has to be the ultimate authority in giving the direction of the analysis along his or her specific task. At the same time, the system has to provide effective means of interaction to concentrate on this specific task. On top of that, in many applications different people work along the path from data to decision. A visual representation will sketch this path and provide a reference for their collaboration across different tasks and abstraction levels. The diversity of these tasks can not be tackled with a single theory. Visual analytics research is highly interdisciplinary and combines various related re- search areas such as visualization, data mining, data management, data fusion,

statistics and cognition science (among others). Visualization has to continuously challenge the perception by many of the applying sciences that visualization is not a scientific discipline in its own right. Even if the awareness exists, that scientific analysis and results must be visualized in one way or the other, this often results in ad hoc solutions by application scientists, which rarely match the state of the art in interactive visualization science, much less the full com- plexity of the problems. In fact, all related research areas in the context of visual analytics research conduct rigorous, serious science each in a vibrant research community. To increase the awareness of their work and their implications for visual analytics research clearly emerges as one main goal of the international visual analytics community. Because visual analytics research can be regarded as an integrating discipline, application specific research areas should contribute with their existing procedures and models. Emerging from highly application-oriented research, dispersed research communities worked on specific solutions using the repertoire and standards of their specific fields. The requirements of visual analytics introduce new dependencies between these fields.

Springer08c. Gennady Andrienko and Natalia Andrienko. A Visual Analytics Approach to Exploration of Large Amounts of Movement Data.

Data about movements of various objects are collected in growing amounts by means of current tracking technologies. Traditional approaches to visualization and interactive exploration of movement data cannot cope with data of such sizes. In this research paper we investigate the ways of using aggregation for visual analysis of movement data. We define aggregation methods suitable for movement data and find visualization and interaction techniques to represent results of aggregations and enable comprehensive exploration of the data. We consider two possible views of movement, traffic-oriented and trajectory-oriented. Each view requires different methods of analysis and of data aggregation. We illustrate our argument with example data resulting from tracking multiple cars in Milan and example analysis tasks from the domain of city traffic management.

VAST2009: Gennady Andrienko, Natalia Andrienko, Salvatore Rinzivillo, Mirco Nanni, Dino Pedreschi, Fosca Giannotti. Interactive Visual Clustering of Large Collections of Trajectories. Submitted to IEEE VAST 2009

One of the most common operations in exploration and analysis of various kinds of data is clustering, i.e. discovery and interpretation of groups of objects having similar properties and/or behaviors. In clustering, objects are often treated as points in multi-dimensional space of properties. However, structurally complex objects, such as trajectories of moving entities and other kinds of spatio-temporal data, cannot be adequately represented in this manner. Such data require sophisticated and computationally intensive clustering algorithms, which are very hard to scale effectively to large datasets not fitting in the computer main memory. We propose an approach to extracting meaningful clusters from large databases by combining clustering and classification, which are driven by a human analyst through an interactive visual interface.

KDD2009 Gennady Andrienko, Natalia Andrienko. Interactive Spatio-Temporal Cluster Analysis of VAST Challenge 2008 Datasets. Submitted to ACM KDD 2009 workshop on visual analytics.

We describe a visual analytics method supporting the analysis of two different types of spatio-temporal data, point events and trajectories of moving agents. the method combines clustering with

interactive visual displays, in particular, map and space-time cube. we demonstrate the use of the method by applying it to two datasets from the vast challenge 2008: evacuation traces (trajectories of people movement) and landings and interdictions of migrant boats (point).