

1 INSIGHT Annual Report



Intelligent Synthesis and Real-Time Response using Massive Streaming of Heterogeneous Data

www.insight.-ict.eu

Project INSIGHT, “Intelligent Synthesis and Real-Time Response using Massive Streaming of Heterogeneous Data”, is a Specific Targeted Research Project that runs since September 1st 2012 involving the following partners: i) National and Kapodistrian University of Athens (UoA), ii) IBM Ireland (IBM), iii) Fraunhofer Gesellschaft zur Foerderung der angewandten Forschung e.V IAIS (Fraunhofer), iv) Technische Universitaet Dortmund (TUD), v) Israel Institute of Technology (Technion), vi) Federal Office of Civil Protection and Disaster Assistance (BBK), vii) Dublin City Council (DCC). This documents summarizes the objectives of the project and highlights the accomplishments achieved during the first year of the project.

1.1 INSIGHT Context

Today’s Environment: Today’s technological advances are fuelling a virtual explosion on the quantity, quality, and variety of information that is becoming available. As a consequence, new challenges have risen due to three recent revolutionary technologies: a) Advances in sensor networking and the availability of low-cost sensor enabled devices in transportation, healthcare and emergency response, b) the widespread adoption of smartphones and c) the extent of social networks such as Twitter and Facebook. Each of these revolutionary technologies is driving the development and adoption of applications where mobile devices are used as tools for continuous data sensing, collection and analysis.

The INSIGHT goal: The goal of INSIGHT is to exploit and develop further the new capabilities that the combined use of these technologies is offering, in order to bring forth fundamental advances in the Smartcity paradigm, focusing especially on emergency situations. We aim to achieve a significant improvement in the utility of automated systems to manage resources and put new, more capable tools in the hands of disaster planners and city personnel when responding to emergencies in smart cities and countries.

Current systems are limited in several important elements: (i) There is a lack of methods for handling heterogeneous data streams in real-time; (ii) there is limited integration of big data analytics and social computing, (iii) real-time prediction and alarm capabilities have not yet been incorporated into the infrastructure for intelligent management.

To truly achieve quantum improvement, we have to integrate information that comes from the analysis of complex data of very different kinds, for example combine quantitative data from physical sensor with qualitative data from social networking media. The inherent complexities are immense, necessitating novel work both on algorithmic and systems aspects. We plan to create an infrastructure that will provide a long-term support for building, maintaining, and improving such systems. Our efforts will be realized in the context of two carefully selected use cases that aim to develop new capabilities in managing emergency situations:

1. **City Level Use Case:** The City of Dublin is a Smart City already equipped with an infrastructure of sensor that collects a wide variety of data from the urban environment. We leverage the availability of such rich data, in combination to exploiting social media data, to enable traffic managers to detect with a high degree of certainty unusual events throughout the network. For example, we aim to develop novel methods to ameliorate the impact of urban flooding on the capacity of the road system, through monitoring, analysing and combining sensor data and social data to improve emergency response and evacuation and provide the necessary information to stakeholders and end-users.
2. **Nation-wide Disaster Monitoring:** The aim is to achieve early warning of disasters with nation-wide impact, such as severe weather conditions, floods, earthquakes and their subsequent events, such as fires, power outages. The INSIGHT project will develop a system to provide big data monitoring and reliable detection of disasters in near real-time, operating on combined data from mobile and social sources to verify the existence and visualize all facets of a disaster, and manage its consequences, thus improving our capability to handle disasters. The German Federal Office of Civil Protection and Disaster Assistance (BBK) will channel INSIGHT's technology to evaluate their training activities.

1.2 INSIGHT Objectives

The INSIGHT project aims at creating an open, flexible and powerful solution and developing the capabilities for fundamental advances in coping with emergencies in the Smartcities of the future. To achieve our goal, the project has the following innovation objectives:

1. To develop an adaptive, scalable and dependable, real-time infrastructure for emergency monitoring. Our system comprises a hybrid storage and processing system that will be developed combining a streaming engine for fast online analysis of data streams with a big data warehouse for analysing and processing historic data (following lambda architecture principles), and a scalable crowdsourcing system that simplifies the process of sensing, collecting, sharing, sensor data on mobile phones through the development of a mobile, real-time distributed programming framework that offers MapReduce functionality. Our system provides end-to-end real-time and reliable delivery and analysis of continuous streams of data.
2. To develop radically new methodologies for monitoring, processing, analyzing and synthesizing massive amounts of heterogeneous data for improving our ability of coping with emergencies. INSIGHT focuses on investigating and developing methods and systems for data collection, knowledge discovery and information sharing that take into account the characteristics and needs of the applications, to provide insights into regional issues such as transportation planning, environmental sustainability and emergency response. Scientific advances sought by the project include: (i) dealing in real time with the massive volume and distributed nature of the data that typical city- or nation- wide sensor network applications can produce, (ii) analysing data from different sources and incorporating data from social networking sites in real-time to identify events of interest; understanding the implications coming from the different characteristics and properties of each data source, and complying with data-related legal restrictions, so that the right methods can be developed, (iii) addressing the (lack of) quality of the data that is inherent in both sensor and social applications, employing active learning and people-centric data gathering techniques.
3. A major objective of INSIGHT is to facilitate the adaptation and use of the technologies that we will develop: To ensure reusability and facilitate faster adaptation of the proposed methodology, we will provide appropriate customized user-interfaces, as well as release code in the public domain.

1.3 Activities and Results of the first year of the project

From Sept 2012 to August 2013, the work of the consortium focused on the development of the framework of the INSIGHT system, the development of the first integrated demonstrator, the development of advanced algorithms and software components for event detection and event recognition, as well as on disseminating early project results. The work of the first year demonstrated the strong collaborations between the partners, and the focus on designing realistic use-cases and evaluation scenarios. In summary, the following are the main activities and results during this first yearly period of the project:

- 1. Algorithmic and Systems Progress:** The consortium made important progress at all algorithmic and system components that will be employed in realizing INSIGHT. Novel techniques were developed for: (i) Traffic modelling; recent work has focused on scaling novel modelling techniques to city- and country-scale. (ii) Complex event processing; investigated the scalability of complex event processing techniques to the size of the datasets we expect, and the ability to monitor for events online. (iii) Twitter and social media monitoring and analysis; developed novel techniques for geo-locating twitter messages, techniques for monitoring emotional response in social media in real time, novel visual analytics techniques for spatio-temporal analysis of geo-located tweet datasets. (iv) Monitoring heterogeneous data sources; investigated techniques for merging heterogeneous data sources, both merging traffic data from different sensors in real-time, and by combining social media data with static data to provide real-time feedback. (iv) Active learning and crowdsourcing; we developed novel algorithms for user selection and an approach for using crowdsourcing to discover and track events. First versions of these components have been developed. It is important to note that progress in the development of the majority of these components has been the result of collaborative work between INSIGHT partners. The individual software components and the integrated infrastructure will be constantly improved in terms of functionality and performance for the rest of the project.
- 2. Integration and Collaboration:** A major outcome of the first year is the definition of framework of the INSIGHT architecture, including the INSIGHT architecture, the INSIGHT communication mechanism, and the evaluation mechanisms through the use of concrete use cases. This work was done by a collaboration of all partners, leading to a common understanding of the use cases from end-users, algorithmic and system perspective. This process provided guidance for the architectural design phase through the mapping of the framework to the use cases, and led to the development of a proof of concept of the INSIGHT system that focuses in the city-level use case. In addition, this process has highlighted the extensive collaboration between the INSIGHT partners. This process will be used as a basis for the next development cycle of the system.
- 3. Dissemination Efforts:** The significant progress made is also reflected in the 22 scientific publications in top conferences and journals, in addition to several talks in scientific and technical conferences and forums. Fraunhofer presented INSIGHT at leading forums for public safety and security in Germany targeting officials, politicians and public. The integration of the streams framework of TUD and the complex event processing frameworks used by Technion was presented at the DEBS 2013 Grand Challenge and it received very encouraging feedback by the community, where it won the audience award.
- 4. Data Collection:** In order to develop and test event recognition and event processing algorithms it has been crucial to collect sufficient amounts of real-world data from the application domains. Therefore, significant effort has been directed at data collection, data cleaning, data understanding, and data annotation for traffic data as well as microblogging data.

1.4 Expected final results and their potential impact and use

The expected final exploitable results of the project include:

1. The first exploitation direction is directly linked to the city-wide and the nation-wide monitoring scenarios. Should the project be successful, completely new approaches for improved disaster management become possible. In the first step, Fraunhofer aims to transform the INSIGHT research prototype in a fully working system, jointly with BBK. Having a working prototype at a federal level is a very important strategic prerequisite for achieving this goal. Simultaneously, centers at the European level (and beyond) will be targeted. We plan to explore the use of the system to other cities as well, taking advantage of the similarity of the traffic management problem, while building an infrastructure with the flexibility to address differences. We plan to create an infrastructure that will provide a long-term support for building, maintaining, and improving such systems. Building on our partnership with the Dublin City Council and the German Federal Office of Civil Protection we will use our experience in bringing this system online in other cities as well.
2. Another direction of exploitation is based on the collaboration with Vodafone in Mobile Analytics. Research results of INSIGHT will greatly improve Fraunhofer's abilities for real-time analysis of mobile phone data. Fraunhofer and Vodafone will be in a position to offer advanced data and analytics services. Already today, performance measures in the out-of-home-media industries in Germany and Switzerland are centrally based on Fraunhofer mobility mining solutions. These services would be crucially enhanced by offering near-real-time capabilities based on mobile phone activity data.
3. The techniques developed in the project will advance the state-of-the-art in monitoring individual data sources as well, and in exploitation we will build stand-alone modules for monitoring one or more data sources. These will provide reinforced ability for a wide range of innovators to tap data infrastructures and to add value beyond the original purpose of the data through data analysis. Concrete examples of such results and their impact are: (i) Improved ability to mine small chunks of textual data, e.g. twitter. As an example of the benefits of such improved methods, Fraunhofer is currently offering sentiment analysis services to German industry (partly in co-operation with IBM Germany). The ability to mine large collections in almost real-time will greatly enhance the quality of such a service. (ii) Develop mature crowdsourcing technologies by considering the set of problems from scalability to privacy and legal issues. (iii) Open source infrastructure that allows the development of systems that analyse massive heterogeneous data.

Key expected results for the following one year-period of the INSIGHT project include:

1. The development of techniques and algorithms for the analysis of the data; the focus of the second year period is on the scientific advances in analysis, monitoring, system building, that will be fundamental in developing the actual system in the third year.
2. The continuation and validation of the integration work.
3. Work on the exploitation and adoption of the developed technologies, including the development of easy to use ontologies.

1.5 Public web site and additional resources

www.insight-ict.eu