Intelligent Synthesis and Real-time Response using Massive Streaming of Heterogeneous Data (INSIGHT) and its anticipated effect on Intelligent Transport Systems (ITS) in Dublin City, Ireland

Dermot Kinane[1*], François Schnitzler[2], Shie Mannor[2], Thomas Liebig[3], Katharina Morik[3], Jakub Marecek[4], Bernard Gorman[4], Nikolaos Zygouras[5], Yannis Katakis[5], Vana Kalogeraki[6] and Dimitrios Gunopulos[6]

[1]Intelligent Transport Systems, Dublin City Council, Republic of Ireland

[2]Faculty of Electrical Engineering, Israel Institute of Technology, Israel

[3]Computer Science VIII, TU Dortmund University, Germany

[4]IBM Research, Dublin, Republic of Ireland

[5]Department of Informatics and Telecommunications, University of Athens, Greece

[6]Department Informatics, Athens University of Economics and Business, Greece

Abstract

Intelligent traffic and transport management involves the use of large data streams to identify and effectively manage issues of congestion and quality of service. In particular, urban traffic has been in the eye of the storm for many years now and gathers increasing interest as cities become bigger, crowded and 'smart'. This paper examines the likely impact of the work of the EU-funded FP7 INSIGHT programme on traffic incident detection and its amelioration in Dublin city.

Keywords

Big Data, traffic management, crowdsourcing.

Introduction: A Data Revolution

Today's technological advances are fuelling a virtual explosion of the quantity, quality, and variety of information that has become available. Mobile computing combined with smart city infrastructures is generating massive, heterogeneous data and creating the opportunities for novel applications. In traffic monitoring, the data sources include traditional ones (sensors) as well as novel ones such as micro-blogging applications like Twitter; these provide a new stream of textual information that can be utilized to capture events, or allow citizens to constantly interact using mobile sensors. More specifically, we identify three recent revolutionary technologies that are having an impact on techniques for intelligent traffic and transport management:

a. Instrumentation of the world: Advances in sensor networking and the availability of low-cost sensor-enabled devices have led to the integration of sensors into vital sectors of transportation, healthcare, and emergency response. These sensors give us an unprecedented monitoring capability over our environment.

b. The widespread adoption of smartphones: With the proliferation of smartphones, tablets, and other mobile computing devices, we enter an era where people can instantaneously share aspects of their lives, creating virtual communities with direct connections to the real world, and become both data producers and data consumers at the same time. Today the number of smartphone/tablet users in the Republic of Ireland is approximately 2.5 million which represents over 55% of the population. (Commission for Communications Regulation, 2013)

c. The extent of social networks: Social networks, such as Twitter or Facebook, have significantly changed the way humans interact. Crowdsourcing is already used as a mechanism to build bridges between the real world and virtual communities. These networks are also used to disseminate news, advertisement, entertainment, and even to organize social action.

The challenge of heterogeneous data

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. Truly visionary applications however must combine these three technologies. To fully realise the benefits it is imperative to identify and work on those important applications where the concerted use of such technologies is the key for fundamental advances.

The vision of the INSIGHT project[1] is to develop the capabilities that the combined use of technologies can offer for fundamental advances in the Smart City paradigm, focusing especially on emergency situations. The long term goal of INSIGHT is to enable traffic managers to detect with a high degree of certainty unusual events throughout the road network. We recognise that a deep integration of data collection, knowledge discovery and information sharing capabilities offered by today's technologies is the cornerstone for the advances we seek. We present the general framework of a system that has been designed in this context, and address the challenges that arise from a real installation and application.

Dublin City Traffic Management Use Case

Dublin city is the capital city of the Republic of Ireland. Dublin County is comprised of four local authority areas whose populations are detailed inTable 1, below.

Dublin City Council (DCC) is the democratic body governing Dublin city. It is the largest Local Authority in the Republic of Ireland with a staff of approximately 6,000 and provides a wide range of services for its citizens and businesses.

[1] www.insight-ict.eu

Administrative area	Population 2011
Dublin County	1,270,603
of which	
Dublin city	525,383
Dun Laoighre-Rathdown	206,995
Fingal	273,051
South Dublin	265,174

Table 1 – Census population of Dublin County (Central Statistics Office, 2011)

Transport in Dublin

The Roads and Traffic Department of Dublin City Council develops, maintains and manages the City's road network for the benefit of pedestrians, cyclists, motorists and public service and commercial vehicles. The department is currently pursuing the following strategies:

- Providing alternatives to car commuting
- Developing, optimising and maintaining the city's road network
- Managing on-street parking
- Improving the city's environment

Dublin city faces huge challenges in the area of sustainable mobility. There has been considerable success in the provision of public transport for commuting over the past 10 years. However, the demand for travel and movement of goods continues to expand.

The city street network is finite and the historic street pattern does not lend itself to significant expansion. It is necessary to optimise the network to provide sustainable transport to meet the needs of business, residents, commuters, shoppers and disabled people.

The current public transportation system in Dublin comprises bus services mainly operated by Dublin Bus, heavy rail lines operated by Irish Rail and light rail transport comprising two Luas tram lines operated by Transdev.

Current Situation

The Dublin Regional Traffic Management Centre manages the traffic flow around the Dublin city region. The centre operates from a purpose-built control room and is staffed by a team of ten experts operating on a 24/7 – 365 rota and is managed by one supervisor. The control room staff monitors the current traffic situation via SCATS (Sydney Coordinated Adaptive Traffic System) which is the traffic management software in use throughout Dublin city and most of the main traffic corridors in the region. Through a network of approximately 250 CCTV cameras located throughout the region, the control room staff have the ability to view the current status of traffic conditions in real-time. The control room staff liaises with the Gardai Traffic Corps (Police Service) when unusual congestion is caused by such events as accidents, incidents or unusual weather events. This team is assisted by members of the LiveDrive radio team which monitor traffic during peak hours (07:00-10:00 and 16:00-19:00) and report their information to a 60,000+ listener base. During these peak-times the team is also assisted by members of AA Roadwatch who disseminate the main traffic issues to a wider audience.

In more recent times unusual weather events such as snow and flooding have severely impacted the city with, in some cases, fatal consequences. This has, in part, lead to the upgrade of the traffic control room to include an incident management centre, where city managers, roads authorities, the emergency services and other stake holders can meet during times of crises to coordinate responses. See Figure 1 to Figure 4 for a few incident examples.

Figure 1 – Example of flooding on M50 orbital motorway around Dublin

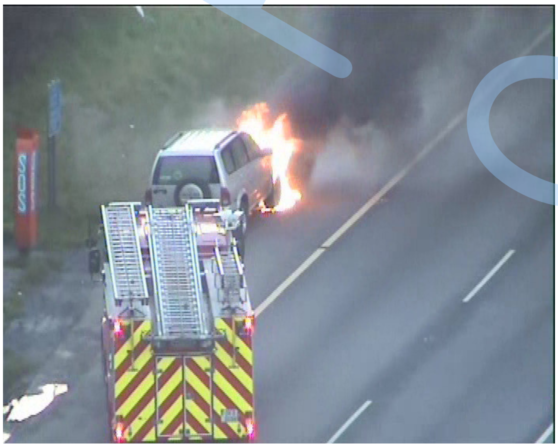

Figure 2 – Example of an incident on a roadway in Dublin

Figure 3 – Example of snow causing congestion in Parnell Square, Dublin

Figure 4 – Example of an accident involving two forms of public transport on O'Connell Street, Dublin

Data Sources

In pursuit of the delivery of INSIGHT for Dublin city, many forms of available heterogeneous data will be examined. These data sources are detailed below.

SCATS - Dublin City Council currently uses systems monitoring real-time traffic density through a system called SCATS. SCATS (Sydney Coordinated Adaptive Traffic System) is an innovative computerised traffic management system developed by Roads and Maritime Services (RMS) Australia. The evolution of SCATS started in the 1970s and is continually being improved by RMS to manage traffic signal networks around the world. SCATS has been distributed to 263 cities in 27 countries worldwide controlling more than 35,531 intersections (Roads and Maritime Services, 2011). Dublin City Council manages approximately 750 SCATS controlled intersections throughout the road network.

TRIPS - Travel-time Reporting and Integrated Performance System software is a sophisticated facility that directly integrates with and provides accurate travel time prediction information on road network performance and other important information on the operation of SCATS. TRIPS actively collects and processes travel time data every 60 seconds ensuring that TRIPS outputs are always consistent with current road demands (Advantech Design, 2011). DCC has configured approximately 45 routes throughout the city using TRIPS.

Bus Data - In Dublin city, the main mode of public transport is by bus; it far outweighs both the heavy and light rail systems both in passenger numbers and in coverage. The main operator is Dublin Bus, a semi-state owned company with a fleet of over a 1,000 buses, most of which either terminate or pass through the city centre area. In 2008, Dublin Bus began the process of fitting an Automatic Vehicle Location (AVL) system to their fleet. Real-time GPS traces from buses are collected at 30-second intervals and this data is available in real-time.

CCTV - Through a network of approximately 250 CCTV cameras, traffic can be monitored throughout the traffic network. INSIGHT will use CCTV, where possible, to confirm a prediction.

Live Drive – the Live Drive radio show[2] has been on air for over 7 years and has become part of the fabric of Dublin radio. Boasting a daily listenership in excess of 60,000, the show is a lively mix of music and the most accurate up-to-the-minute traffic information. Live Drive broadcast for 6 hours daily on 103.2 Dublin City FM - from 7am to 10am and from 4pm to 7pm - live from Dublin City Council's state-of-the-art traffic control centre. Through a combination of Twitter and text messages received from their listenership coupled with the access to CCTV, Live Drive provide the most up-to-date information to members of the travelling public in the Dublin city region. This data is recorded on a spreadsheet in real time and is available to INSIGHT.

Weather Data - Flooding is an important issue in Dublin due to a combination of weather events and tides. Weather is fast-changing: rainfall is frequent and very difficult to predict. Tides are easier to predict, but have a strong effect on the level of the river Liffey that bisects the city. (Kinane, 2013)

[2] http://dublincityfm.ie/livedrive

Future situation with INSIGHT

Flood Management

In October 2011, severe floods caused by heavy localised rainfall, over a period of a few days in the city of Dublin caused two deaths and severe damage to the city's traffic infrastructure. Such flooding are increasingly common and compromise public health and safety. Given the high density of people in cities, even small-scale floods can lead to considerable damage. A goal of the INSIGHT project is to develop novel methods to predict the impact of urban flooding on the capacity of the road system. INSIGHT will seek to tap into measurements from dedicated sensors, social media sources, as well as human sensors to give accurate estimates, and possible predictions of floods and other unusual weather events.

INSIGHT will use GPS traces and vehicle counts to infer the absence of flooding in some areas of the city. Twitter reports will be used to infer flooding in other areas, and selective querying will be used to obtain flooding measurements where neither dedicated sensor nor social media reports are available.

Traffic Management

At the moment, data streams (SCATS and CCTV) are monitored in a manual fashion by the control room staff. Interventions are carried out when unusual congestion is noticed. This is very much a reactive response.

INSIGHT will automatically monitor all data sources available to the system. INSIGHT will provide and update in real-time a probability that any particular type of event is taking place. A first prototype processing only two data streams is illustrated in Figure 7. When an event occurs, INSIGHT will notify the control room staff and inform them of the space, time and semantics of the event. A graphical user interface (GUI) will display, in real-time, the nature of all events throughout the road network. INSIGHT will enable the traffic managers to reduce the response time to unusual congestion events and reallocate human resources from monitoring to response implementation. This automated alert system will mean that resources can be deployed to the correct location in a timely manner, thus reducing the time taken to resolve the event.

To test the capabilities that will be developed and integrated in the system we are developing a realistic testing scenario that involves monitoring two main routes in the city of Dublin as described in (Kinane, 2013). The CCTV network will be used to evaluate and validate the results of the project.

The Initial INSIGHT System: Incident detection in SCATS and bus data and resolution of disagreement by crowdsourcing

In this section we report in more detail initial work that is part of the first prototype developed by the INSIGHT consortium (Artikis, et al., 2014). The fundamental advance of this work is the proposal to compliment SCATS and bus data streams analysis with crowdsourcing to detect and label traffic incidents without any manual operation by city operators. This system is schematized in Figure 7. The input of the system includes two types of sensor which are considered as the main event sources. Buses transmit information about their position and congestion and SCATS vehicle detectors are installed at intersections and report on traffic flow and density.

INSIGHT Architecture

Both the SCATS measurements and the bus GPS traces can be used to detect congestion or other interesting traffic related events. However, monitoring such a large stream of data requires an efficient use of resources. We develop a sophisticated architecture, described in more detail in (INSIGHT Deliverable 2.1) and in (Artikis et al., 2013) that allows us to deal with large data volume and data velocity in a scalable fashion. Keys to our architecture are: (i) the use of the concept of the lamda architecture (Marz, 2013) which allows the concurrent use of a streaming component for real-time monitoring and prediction, and a batch component for analytics, (ii) the use of novel sophisticated streaming components that can scale with the volume of the data, in particular the STREAMS Framework (Bockerman, 2012) and Storm[3], (iii) a novel mechanism, Intelligent Sensor Architecture (ISA) (INSIGHT Deliverable 2.1) for abstracting the different input sources, thus facilitating the introduction of new data sources in the system and making the integration of heterogeneous data sources easier, and (iv) a flexible mobile infrastructure that allows the seamless integration of several mobile end users in the system (INSIGHT Deliverable 3.1). Figure 5 describes the main elements of the architecture.

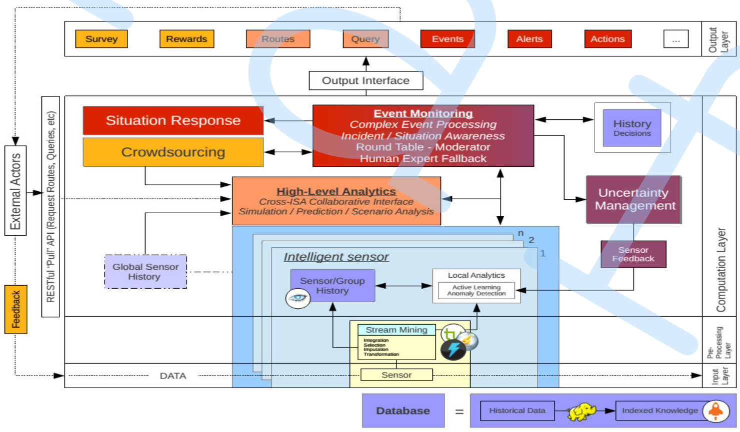

Figure 5 – The INSIGHT Architecture

Data Analysis Steps

One of our goals is to develop mechanisms that automatically flag possible events as soon as possible. To this end we develop and use logical rules to automatically detect interesting events in both data streams. Event detection is performed online, on successive time windows in the first step of the complex event processing component. There can be some overlap

[3] http://storm.incubator.apache.org/

between the windows, which allow the system to take into account measurements arriving late, a not so unlikely occurrence due to the size of the system.

We stress that one of the first obstacles in analysing and using the data is that, although very rich and potentially very useful, data streams are very noisy and therefore difficult to understand. For example Figure 6 shows that a bus stop with the same identifier can be located in many different locations in the city. This problem led to a new technique that groups together bus stops which are nearby, taking into account the buses' direction and therefore reduce the noise in the input data. As another example, Traffic Flow values for several sensors display large fluctuations and therefore we designed and applied streaming Moving Average filters. We also perform statistical analysis (in batch mode) to improve our understanding and find patterns in the data. We measured the mean and the standard deviation of traffic flow values for specific sensors at a particular day of week (Mon-Sun) and at different hours (0-23), as it is described in (INSIGHT Deliverable 5.1). Also we did the same analysis in the bus data, calculating the mean and the standard deviation of the changes of delay values for particular bus lines, bus stops and areas.

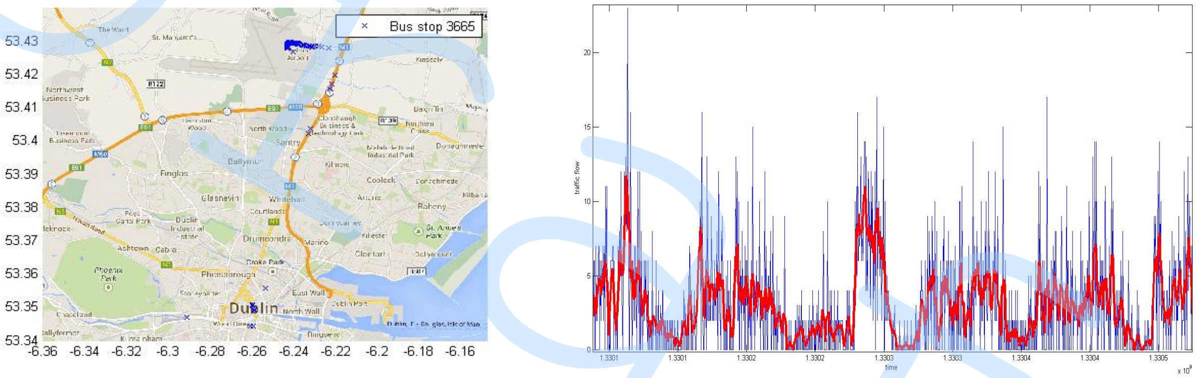

Figure 6 - (left) GPS Location where buses reported that they were at Bus Stop 3665, (right) Time series of raw Traffic Flow values (Blue) and Traffic Flow values after applying Moving Average (red)

Event monitoring

Essentially, the first step of our system is to perform a data transformation, with the objective to obtain data in a format that can be used efficiently to identify problems in the traffic. For example, for the BUS data source we extracted the following attributes from the raw bus data that are useful for further analysis: bus direction, delay increase, distance covered, time elapsed from the previous measurement and vehicle speed.

In order to identify problems in the traffic condition we use several rules (for examples see Artikis et al., 2013, Artikis et al., 2014), using the bus and SCATS data. Such rules take into account that each sensor, bus stop, bus line or area had different thresholds. Examples of such rules include:

2013). Recently, systems such as Gigwalk[5] and FieldAgent[6], illustrate that they can incorporate crowdsourcing in time-sensitive tasks such as traffic monitoring, location-aware surveys, points-of-interest suggestions, real-time entertainment recommendations, price checks, etc.

In our case, a disagreement between sensors will trigger queries to participants close to the location of the disagreement, typically through a dedicated smartphone application. The answers to these queries are collected and aggregated. The result will then be used for the resolution of the disagreement and sent to the city operator for verification or further action.

Crowdsourcing can also be used to obtain additional information about congestion, for example the cause of the congestion (accident, flood etc.). This can be particularly useful for anomalies occurring in locations not covered by CCTV and speed up the implementation of a relevant response to the incident.

In order to recruit and motivate participants, a reward is typically provided for answering questions correctly. In the city of Dublin, such rewards could consist of parking vouchers, already awarded to informative calls to LiveDrive. An advantage of crowdsourcing over voluntary reporting is that it can invalidate an event generated by sensor noise.

Given the massive amount of data that can be generated in such situations, it is imperative to develop efficient mechanisms for contacting a large number of users in real-time and with high reliability, as well as efficient mechanisms for filtering and ranking the results. And since users of the Insight system are often mobile, it is important to be able to reach them and gather their answers in a mobile setting. In our recent work (Boutsis, Kalogeraki, 2013) we have presented an approach that aims to stimulate user participation and handle dynamic task assignment and execution in the crowdsourcing system. The goal is to determine the most appropriate workers to assign incoming tasks, in such a way so that the real-time demands are met and high quality results are returned.

We have developed the Insight mobile app (INSIGHT Deliverable 3.1) that executes on the user mobile devices. During normal operation the Insight app would run in the background. It will not only enable human users to use the Insight system to receive real-time information about emergency events (i.e., floods, traffic congestions, incidents in the city-level use case) and updates, but it will also provide the Insight system with the capability to send queries to the users to question them about these events in real-time for verification for investigating the event further.This prototype has been developed based on and tested on historical records from April 2013 (except crowdsourcing, which was simulated). DCC and the rest of INSIGHT are currently tackling technical and other issues before large-scale deployment.

[5] http://gigwalk.com/
[6] http://www.fieldagent.net/

Towards Creation of a Smart City

An important additional step towards the creation of a smart city is to dynamically predict traffic conditions, and act upon such predictions. As with all cities, Dublin can get congested during peak times. The INSIGHT project will develop a dynamically updated traffic prediction model, that can not only be used to predict the future traffic conditions and therefore alert operators of possible problems but can also be used to generate trip planning recommendations with the objective to avoid future jams. The traffic prediction system and the traffic planner system utilize data recorded by the SCATS system and estimates future traffic using their spatio-temporal dependencies and correlations (Liebig et al., 2014, Schnitzler et. al., 2014) A regression is used to extrapolate the traffic flow values at unobserved locations. This traffic model is updated regularly with incoming SCATS data. Dataflow in the system is managed by the streams framework (Bockerman, 2012), a fast and reliable real-time processing framework suitable for fast real time event processing and data mining tasks.

Conclusions

We presented an initial report on a system for heterogeneous stream processing and crowdsourcing supporting intelligent urban traffic management. Complex events related to traffic congestions are detected from heterogeneous sources involving fixed sensors mounted at traffic intersections and mobile sensors mounted on public transport vehicles. To deal with the inherent data veracity, a crowdsourcing component handles and resolves source disagreement. Current work deals with the data sparsity problem by developing a traffic modelling component to make congestion estimates in areas with low or non-existent sensor coverage. The incorporation of heterogeneous data streams and crowdsourcing in this traffic prediction model are our next steps. Additional future work includes optimizing the individual routing suggestions to avoid causing a traffic load since the trip planner (aiming to avoid jams) may cause traffic jams itself.

Acknowledgements

This work is funded by the EU FP7 INSIGHT project (318225).

References

Advantech Design. (2011). TRIPS User Manual. Melbourne.

Artikis A., Weidlich M., Gal A., Kalogeraki V. and Gunopulos D. (2013), Self-Adaptive Event Recognition for Intelligent Transport Management, IEEE Int. Conference on Big Data, Oct. 2013, Santa Clara.

A. Artikis, M. Weidlich, F. Schnitzler, I. Boutsis, T. Liebig, N. Piatkowski, C. Bockermann, K. Morik, V. Kalogeraki, J. Marecek, A. Gal, S. Mannor, D. Gunopulos, and D. Kinane. (2014). "Heterogeneous Stream Processing and Crowdsourcing for Urban Traffic Management," in Proc. 17th International Conference on Extending Database Technology (EDBT), Athens, Greece, March 24-28, 2014, 2014, pp. 712-723.

Bockerman C., Blom, H. (2012) The streams Framework. TU Dotmund University, Technical Report.

Central Statistics Office. (2011). Census 2011 Preliminary Results.

Commission for Communications Regulation. (2013). Irish Communications Market - Quarterly Key Data Report.

Good, B., & Su, A. (2013). Crowdsourcing for Bioinformatics. Bioinformatics Journal.

INSIGHT Consortium (2013), Deliverable 6.1: Use Cases

INSIGHT Consortium (2013), Deliverable 5.1: Report on end-user requirements, test data and on prototype definitions

INSIGHT Consortium (2013), Deliverable 2.1: System requirements spec, standards and guidelines for development and architecture

INSIGHT Consortium (2013), Deliverable 3.1: Synthesis of Massive Heterogeneous Streaming Data: The INSIGHT Communication Infrastructure

Kinane, D. (2013). INSIGHT Work Package 6 Document. Dublin.

Land, K., Slosar, A., Lintott, C., Andreescu, D., Bamford, S., Murray, P., Nichol, R., Raddick, J., Schawinski,K., Szalay, A., Thomas, D., and Vandenberg, J. (2008). Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society , 388, 1686-1692.

T. Liebig, N. Piatkowski, C. Bockermann, and K. Morik (2014). Predictive Trip Planning – Smart Routing in Smart Cities, in Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014), Athens, Greece, March 28, 2014, 2014, pp. 331-338.

Marz. N. (2013), Big Data: Principles and best practices of scalable realtime data systems.

Roads and Maratime Services. (n.d.). Retrieved 2014, 07-01 from http://www.rms.nsw.gov.au/environment/cleanerair/roadnetworkplanning/

F. Schnitzler, T. Liebig, S. Mannor, and K. Morik, (2014), Combining a Gauss-Markov model and Gaussian process for traffic prediction in Dublin city center, Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference (EDBT/ICDT 2014), Athens, Greece, March 28, 2014, 2014, pp. 373-374.