LarKC

The Large Knowledge Collider
a platform for large scale integrated reasoning and Web-search

FP7 – 215535

D6.1

Requirements summary and initial data repository

Emanuele Della Valle, Irene Celino, Daniele Dell'Aglio (CEFRIEL), Zhisheng Huang (VUA), Kono Kim (Saltlux), Werner Hauptmann, Ralph Grothmann, Yi Huang, Volker Tresp (Siemens)

Document Identifier:	LarKC/2008/D6.1
Class Deliverable:	LarKC EU-IST-2008-215535
Version:	version 1.0
Date:	September 30, 2008
State:	final
Distribution:	public

Executive Summary

This document reports the results of the ongoing discussion about the requirements set by the Urban Computing field and the outcome of the data collection activity for the future realization work in this workpackage.

To address the Urban Computing requirements for LarKC technologies, we present the results of our cooperative work under the form of a survey of related works and two scientific papers we submitted to two relevant workshops in the field. For what regards the data collection, we present the outcomes of the ongoing gathering activity we are performing to identify possible data sources we could use in our Urban Computing environment.

Document Information

IST Project Number	FP7 – 215535	**Acronym**	LarKC
Full Title	Large Knowledge Collider		
Project URL	`http://www.larkc.eu/`		
Document URL			
EU Project Officer	Stefano Bertolo		

Deliverable	**Number**	6.1	**Title**	Requirements summary and initial data repository
Work Package	**Number**	6	**Title**	Urban Computing: Real Time City

Date of Delivery	**Contractual**	M6	**Actual**	30-Sept-08
Status	version 1.0		final ☒	
Nature	prototype ☐ report ☒ dissemination ☐			
Dissemination Level	public ☒ consortium ☐			

Authors (Partner)	Emanuele Della Valle (CEFRIEL), Irene Celino (CEFRIEL), Daniele Dell'Aglio (CEFRIEL), Zhisheng Huang (VUA), Kono Kim (Saltlux), Werner Hauptmann (Siemens), Ralph Grothmann (Siemens), Yi Huang (Siemens), Volker Tresp (Siemens)			
Resp. Author	Emanuele Della Valle (CEFRIEL)		**E-mail**	emanuele.dellavalle@cefriel.it
	Partner	CEFRIEL	**Phone**	+39 (02) 23954-266

Abstract (for dissemination)	This document reports the results of the ongoing discussion about the requirements set by the Urban Computing field and the outcome of the data collection activity for the future realization work in this workpackage.
Keywords	urban computing, traffic management, requirements, data collection

Version Log			
Issue Date	**Rev No.**	**Author**	**Change**
Sept 4, 2008	1	Irene	Creation
Sept 5, 2008	2	Irene	Inserted content into main sections and description of annexes
Sept 15, 2008	3	Irene	Addressed Georgina's feedbacks
Sept 22, 2008	4	Irene	Finalization for Frank's check and annexes addition
Sept 30, 2008	5	Irene	Final version

Project Consortium Information

Acronym	Partner	Contact
Semantic Technology Institute Innsbruck `http://www.sti-innsbruck.at`	STI	Prof. Dr. Dieter Fensel Semantic Technology Institute (STI) Innsbruck, Austria E-mail: dieter.fensel@sti-innsbruck.at
AstraZeneca AB `http://www.astrazeneca.com/`	ASTRAZENECA	Bosse Andersson AstraZeneca Lund, Sweden E-mail: bo.h.andersson@astrazeneca.com
CEFRIEL SCRL. `http://www.cefriel.it/`	CEFRIEL	Emanuele Della Valle CEFRIEL SCRL. Milano, Italy E-mail: emanuele.dellavalle@cefriel.it
CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O. `http://cyceurope.com/`	CYCORP	Michael Witbrock CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O., Ljubljana, Slovenia E-mail: witbrock@cyc.com
Hoechstleistungsrechenzentrum, Universitaet Stuttgart `http://www.hlrs.de/`	HLRS	Georgina Gallizo Hoechstleistungsrechenzentrum, Universitaet Stuttgart Stuttgart, Germany E-mail : gallizo@hlrs.de
Max-Planck-Institut fr Bildungsforschung `http://www.mpib-berlin.mpg.de/index_js.en.htm`	MAXPLANCKGESELLSCHAFT	Dr. Michael Schooler, Max-Planck-Institut fr Bildungsforschung Berlin, Germany E-mail: schooler@mpib-berlin.mpg.de
Ontotext Lab, Sirma Group Corp. `http://www.ontotext.com/`	ONTO	Atanas Kiryakov, Ontotext Lab, Sirma Group Corp. Sofia, Bulgaria E-mail: atanas.kiryakov@sirma.bg
SALTLUX INC. `http://www.saltlux.com/EN/main.asp`	Saltlux	Kono Kim SALTLUX INC Seoul, Korea E-mail: kono@saltlux.com
SIEMENS AKTIENGESELLSCHAFT `http://www.siemens.de/`	Siemens	Dr. Volker Tresp SIEMENS AKTIENGESELLSCHAFT Muenchen, Germany E-mail: volker.tresp@siemens.com
THE UNIVERSITY OF SHEFFIELD `http://www.shef.ac.uk/`	Sheffield	Prof. Dr. Hamish Cunningham THE UNIVERSITY OF SHEFFIELD Sheffield, UK E-mail: h.cunningham@dcs.shef.ac.uk
VRIJE UNIVERSITEIT AMSTERDAM `http://www.vu.nl/`	Amsterdam	Prof. Dr. Frank van Harmelen VRIJE UNIVERSITEIT AMSTERDAM Amsterdam, Netherlands E-mail: Frank.van.Harmelen@cs.vu.nl
THE INTERNATIONAL WIC INSTITUTE, BEIJING UNIVERSITY OF TECHNOLOGY `http://www.iwici.org/`	WICI	Prof. Dr. Ning Zhong THE INTERNATIONAL WIC INSTITUTE Mabeshi, Japan E-mail: zhong@maebashi-it.ac.jp

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER http://www.iarc.fr/	IARC2	Dr. Paul Brennan INTERNATIONAL AGENCY FOR RESEARCH ON CANCER Lyon, France E-mail: brennan@iarc.fr

Table of Contents

1 Introduction

This document is the first (and public) deliverable of the Urban Computing use case. It reflects the work done by WP6 partners in the first six months of the project. In this period, the discussion about the requirements and the possible outputs of this use case has been quite fervent and active.

Therefore, this document contains a summary of what we gained in these six months. We do not think that the knowledge or the conclusions we draw in this deliverable are fixed and immutable; indeed, we want to keep the discussion alive, with the twofold purpose of challenging the technical WPs of the LarKC project (namely, WP2, 3, 4 and 5) to address the problems and requirements coming from this interesting scenario and of providing innovative and valuable solutions to the Urban Computing community.

The two main topics of this document are the requirements (see Section 2) and the data collection (see Section 3). To address the former topic, we present the results of our cooperative work under the form of a survey of related works and two scientific papers we submitted to two relevant workshops in the field, namely the OneSpace workshop (`http://kmi.open.ac.uk/events/onespace08/`) – First International Workshop on Blending Physical and Digital Spaces on the Internet – at the Future Internet Symposium 2008 (in which our paper was accepted) and the NEFORS workshop (`http://nefors08.larkc.eu/`) – Second International Workshop on "New forms of reasoning for the Semantic Web: scalable, tolerant and dynamic" – at the 3rd Asian Semantic Web Conference (for which our paper is currently under review). For what regards the latter topic, we present the outcomes of the ongoing gathering activity we are performing to identify possible data sources we could use in our Urban Computing environment; since we think this activity will go on throughout the project, we attach in the Annexes the current version of this collection as on the project wiki at the delivery time of this document.

2 Urban Computing Requirements

The Urban Computing challenge is open and many efforts are needed to address it. As a starting point, we started the collection of material about related works (events, projects, papers, applications, etc.). The result of the first months of this activity are summarized in a document we attach in the Annexes. We however aim at continuing this activity throughout the project and we will update this survey of state of the art tools and technologies.

In order to compare our work with the open community and to receive feedbacks on the goodness of our research in this field, we decided to submit two papers to two relevant events for external review: OneSpace 2008 and NeForS 2008.

OneSpace 2008 is the First International Workshop on Blending Physical and Digital Spaces on the Internet[1]. It aims to present a high-quality forum of discussion about the identification and study of the complex relationship of the Internet with space, place, geography and distance, whether physical or virtual. It called for papers that present technologies as well as novel ideas, experiments, and insights originating from multi-disciplinary viewpoints, including humanities, social sciences and mathematics.

NeForS 2008 is the Second International Workshop on "New forms of reasoning for the Semantic Web: scalable, tolerant and dynamic"[2]. It focuses on scalability and robustness of reasoning on the Web, and furthermore to investigate alternative reasoning methods, which take into account incompleteness and distribution of data and knowledge as inherent properties.

In those papers, we present the challenging problem of realizing the Urban Computing vision and in particular we describe the requirements for future mobility management systems. We show that novel multi-disciplinary ideas are required to address the Urban Computing challenge and that only partial solutions can be found today.

Therefore, we prepared and submitted the two following papers. The first one was accepted while the second is under review at the time of writing.

1. E. Della Valle, I. Celino, K. Kim, Z. Huang, V. Tresp, W. Hauptmann, Y. Huang, *"Challenging the Internet of the Future with Urban Computing"*, submitted to OneSpace 2008 – First International Workshop on Blending Physical and Digital Spaces on the Internet, colocated with FIS 2008 – Future Internet Symposium, 28.09.2008 – Vienna, Austria (accepted).

2. E. Della Valle, I. Celino, D. Dell'Aglio, K. Kim, Z. Huang, V. Tresp, W. Hauptmann, Y. Huang, R. Grothmann, *"Urban Computing: a challenging problem for Semantic Technologies"*, submitted to NeFoRS 2008 – Workshop on New forms of Reasoning for the Semantic Web: scalable, tolerant and dynamic, colocated with ASWC 2008 – the 3rd Asian Semantic Web Conference, 08.12.2008 – Bangkok, Thailand (under review at the time of writing).

The first paper captures our work as it was in the end of July, whereas the second is a correct picture of our current developments. In particular in the second: (1)

[1] http://kmi.open.ac.uk/events/onespace08/
[2] http://nefors08.larkc.eu/

we present a revised story board that better captures our ideas, (2) we provide a description of our previous experience to better explain how we gathered the requirements, and (3) we better focus the requirement section introducing the problem of coping with scale.

For readers' benefit, we inserted the referenced papers as well as the related works' survey in the Annexes.

3 Urban Computing Initial Dataset

For a complete data repository for the Urban Computing use case, several typologies of sources of information must be taken into account. Therefore, in order to stimulate the requirements raised in the described scenarios and detailed in the referenced papers, as well as to prove LarKC technologies and tools, we started the collection of a list of pointers to potential, available and interesting data sources.

Since this gathering activity can be very effective if performed collaboratively, we created some pages on the project wiki to collect relevant links and information. Moreover, we do believe that this process will go on beyond the delivery date of this document and, for this reason, we decided not to crystallize the list of data sources in this deliverable. Indeed, we will regularly check and update those wiki pages to evolve and maintain a coherent and up-to-date "repository" of data for the Urban Computing use case.

The initial dataset comprises (but is not limited to):

- Some general-purpose services and sources (like geographic or mapping systems, event repositories, etc.). This kind of information is listed at `http://wiki.larkc.eu/UrbanComputing/OtherDataSources`;
- The list of datasets made available by various UK institutions for the "Show us a better way" initiative[1]. This kind of information is listed at `http://wiki.larkc.eu/UrbanComputing/ShowUsABetterWay`;
- Some institutions with which we can negotiate the use of their data. This kind of information is listed at `http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/InformationSources`.

For readers' benefit, in the annexes of this deliverable we inserted the printed version of the first two referenced wiki pages as they are at the time of writing this document (September 30th, 2008). The third wiki page is password-protected on the wiki, since it contains confidential information and therefore it is not included in this deliverable (released with a public dissemination level).

Moreover, always on the wiki, we started to describe some possible scenarios which make use of the identified data sources to provide added value services for urban citizens. Those scenarios are aimed at constituting an initial design of Urban Computing services and systems for citizens and to draw some hypothesis about the use (and usefulness) of LarKC technologies in those fields. We inserted also the current version of those scenarios/storyboards in the deliverable Annexes.

[1] `http://www.showusabetterway.co.uk/call/`

4 Annexes

In the following you will find:

1. the paper accepted at the OneSpace 2008 workshop ("Challenging the Internet of the Future with Urban Computing");
2. the paper currently under review at the NEFORS 2008 workshop ("Urban Computing: a challenging problem for Semantic Technologies");
3. the technical report containing our initial survey of related works (technologies, applications, projects, events, etc.);
4. the printed version of the content of the wiki page `http://wiki.larkc.eu/UrbanComputing/OtherDataSources` (at this document delivery date);
5. the printed version of the content of the wiki page `http://wiki.larkc.eu/UrbanComputing/ShowUsABetterWay` (at this document delivery date).
6. the printed version of the possible scenarios we are defining for the Urban Computing case study, from the following wiki pages:
   - Traffic Management Scenario
     `http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/StoryBoard`
   - Home Finder Scenario
     `http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/Storyboard-HF`
   - United Kingdom Travel Scenario
     `http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/Storyboard-UKT`
   - Personal Assistant Scenario
     `http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/Storyboard-PA`

Challenging the Internet of the Future with Urban Computing

Emanuele Della Valle[1,2], Irene Celino[1], Kono Kim[3], Zhisheng Huang[4], Volker Tresp[5], Werner Hauptmann[5], and Yi Huang[5]

[1] CEFRIEL – Politecnico of Milano, Via Fucini 2, 20133 Milano, Italy
{name.surname}@cefriel.it
[2] Dip. di Elettronica e Informazione, Politecnico di Milano, Milano, Italy
emanuele.dellavalle@polimi.it
[3] Saltlux Inc., Seul, Korea kono@saltlux.com
[4] Computer Science Department, Vrije Universiteit Amsterdam, De Boelelaan 1081, Amsterdam, The Netherlands huang@cs.vu.nl
[5] Corporate Technology, Siemens AG, Information and Communications, Munich, Germany {name.surname}@siemens.com

Abstract. Urban Computing is an emerging branch of Pervasive Computing that aims at integrating computing, sensing, and actuation technologies into everyday urban settings and lifestyles. Realizing the Urban Computing vision is challenging and requires novel ideas. In this paper we first present future mobility management systems as a special case of Urban Computing and then, based on our previous experience in the field, we present a set of challenging requirements of general interest for those that aim at addressing the Urban Computing challenge.

1 Introduction

Our cities must provide answer to very critical questions [6] and among others: "How can we reduce traffic congestion yet stay connected?"

Internet for sure cannot provide an answer on its own, but it is an enabling factor, if not the most important one. A sign that Internet for urban area is growing at a recognizable pace is the rise of the term Urban Computing [1–4] – the integration of computing, sensing, and actuation technologies into everyday urban settings and lifestyles.

Some years ago, due to the lack of data, solving Urban Computing problems looked like a Sci-Fi idea. Nowadays, a large amount of the required information can be found on the Internet at almost no cost (see the result of survey[7,8]).

For this reason we are challenging the LarKC project[9], which is aiming at a configurable platform for infinitely scalable Semantic Web reasoning [5, 6],

[6] http://www.uli.org/
[7] http://wiki.larkc.eu/UrbanComputing/ShowUsABetterWay
[8] http://wiki.larkc.eu/UrbanComputing/OtherDataSources
[9] http://www.larkc.eu

to support the realization of an innovative solution to traffic management. We have been working in this area for years and we can derive from our previous experiences challenging requirements not only for the LarKC project, but also for the entire community working on complex relationship of the Internet with space, places, people and content.

In the rest of the paper, we identify the problem we want to untangle (Section 2) from which we derive requirements for Urban Computing (Section 3). In Section 4, we provide a short description of the partial solutions we are working on, whereas, in the concluding Section 5, we briefly discuss the potential impact of Urban Computing.

2 Sustainable Mobility

Mobility demand has been growing steadily for decades and this growth is foreseen to continue in the future. For many years, the primary way of dealing with this increasing demand has been the increase of the roadway network capacity, by building new roads or adding new lanes to existing ones. However, financial and ecological considerations are posing increasingly severe constraints on this process. Hence, there is a need for additional intelligent approaches designed to meet the demand while more efficiently utilizing the existing infrastructure and resources.

2.1 A Challenging Use Case

We present an use case that shows the added value of (1) collecting a broad set of information about mobility, (2) integrating it and (3) using it to support a citizen that has to go to Milan from Varese (another city in Lombardy).

- Actors:
  - Carlo: a citizen living in Varese (60 km North-West of Milan).
  - MUCS: the fictitious Milan Urban Computing System.
- Story Board:
  1. Carlo arranged a meeting in Milan city center for the day after at 11.00.
  2. Willing to plan his travel, he accesses MUCS.
  3. Carlo fills in the required data:
     - FROM: Varese tomorrow after 8.00;
     - TO: Milan city center before 11.00;
     - USING: my private car.
  4. MUCS works as Google Maps does today and gives the resulting driving directions, but, instead of saying that such a travel requires 50 minutes, MUCS explains Carlo that he should leave home after 9.00 (when the usual commuters traffic on the A8 motorway is almost over).
  5. MUCS also asks Carlo if he wants to be informed via SMS about traffic conditions and possible alternatives.
  6. Carlo agrees and exits MUCS.

7. The day after an accident involving multiple vehicles takes place at 8.15 on the A8 motorway.
8. MUCS estimates that an accident of such kind will result in a congestion of A8 until 10.00, therefore it checks if any planned travel is at risk. It finds Carlo's travel.
9. MUCS checks if Carlo can take an alternative drive, but no alternatives are found to allow Carlo get to Milan in time for his meeting.
10. MUCS checks if Carlo can take public transportation instead. It founds two alternatives:
    (a) Railroad "LeNord" and Subway M3:
        - 8.39 Varese Casbeno - 10.03 Milano Repubblica;
        - take M3 from Repubblica[10] to Duomo (average waiting time 7 minutes, average duration 5 minutes);
    (b) Railroad "FS" and Subways M2 and M3:
        - 8:43 Varese Stato - 9.55 Milano Garibaldi;
        - take M2 from Garibaldi to Centrale (average waiting time 3 minutes, average duration 7 minutes);
        - take M3 from Centrale to Duomo (average waiting time 7 minutes, average duration 8 minutes).
11. MUCS sends an SMS to Carlo informing him that an accident is holding up A8 and he'd better use public transportation; two itineraries have been already prepared for him.
12. Carlo accesses the MUCS service and checks the alternatives. He chooses the first one and uses the ticket-less option to buy the train ticket.

2.2 Challenging Problems

Public authorities have taken steps in the direction to support this use case, but very complex problems have to be solved. Control centers for mobility management have to be connected to different devices (such as detectors on roads, cameras, traffic lights, etc.) and require sophisticated tools for traffic modeling, estimation, prognosis and decision support.

Traffic System Infrastructure Setup Today, in a typical information infrastructure for real-time traffic control that can be found in different cities usually the following basic components can be discriminated. There are sensors (e.g. loop detectors, cameras, traffic eyes, radar detectors) on major roads recording several traffic magnitudes such as vehicle speed (km/h), traffic flow (vehicle/h) and occupancy or traffic density, i.e., the percentage of time the sensor is occupied by a vehicle (vehicles/km). The distance between successive sensors on a freeway is typically in the order of about 500 meters. The information is periodically transmitted to a control center. The control center also receives information about the current state of control devices. Such control devices include traffic signals at

[10] Repubblica is both the name of the train station and of the subway station, but they are two different places.

intersections, traffic signals at sideways entry-ramps, variable message signs that can display different messages to drivers (e.g., warning about existing congestions, accidents or alternative path recommendations), radio advisory systems to broadcast messages to drivers, and reversible lanes (i.e., freeway lanes whose direction can be selected according to the current and expected traffic demand). In the control center, operators interpret the sensor data and detect the presence of problems and their possible causes. Problems are congested areas at certain locations caused by lack of capacity due to accidents, excess of demand, like rush hours, etc. In addition, operators determine control actions to solve or reduce the severity of existing problems. For instance, they can recommend to increase the duration of a phase (e.g. green time at a traffic signal) or they may suggest displaying certain messages on some variable message signs to divert traffic.

Recent developments not only consider stationary traffic data provided by standard detectors, but also allow to integrate floating car data (FCD), and an increasing number of operators of advanced traffic management systems also use mobile traffic data.

Traffic Modeling and Estimation An analysis of the current and predicted traffic state in the entire road network and the identification of reserve capacities comprise the basis for advanced city traffic management and navigation solutions. Mobile and stationary sensors collect the appropriate traffic data and transmit them to a central unit. Similarly to the weather forecast, the different and heterogeneous information sources are combined to obtain an estimation of the traffic state during a period ranging from minutes to hours or even longer. Thus, a comprehensive knowledge base is built up to support optimal individual route guidance.

Innovative technologies are required in order to process and integrate the resulting collection of distributed information bits within a complex, diverse information environment. Here, a major task is the provision of appropriate solutions for the integration and fusion of heterogeneous information sources, where each source of information can have distinct characteristics with respect to availability, precision, reliability, resolution and representation.

Reacting to a Changing Environment However, as the use case above shows, deploying an infrastructure, modeling and estimating traffic alone is not sufficient; reacting to changes and suggesting other possible solutions is also important. Traffic is just one aspect of mobility. Private cars are just one of the possible means of transportation. Sometimes public transportation can be by far the best choice.

In the storyboard, Carlo is presented by MUCS with an alternative solution that depends on the ability of MUCS to collect on-the-fly information about all means of transportation, estimating (based on historical data) that the resolution of the accident will take longer because it took place in the rush hours, comparing a solution using private car with others that use public transportation and proposing Carlo valid alternatives.

3 Requirements

In this section, we present requirements for Urban Computing that we derived from our previous experience in the field. As argued before, we are particularly interested in the reasoning requirements for LarKC, but we believe such requirements interesting for the entire community working on the complex relationship of the Internet with space, places, people and content.

3.1 Coping with Heterogeneity

Dealing with heterogeneous data has been called upon for a long time in many areas in computer science and engineering, which include database systems, multimedia application, network systems, and artificial intelligence. Here, we would like to propose a notion of heterogeneity processing specifically for semantic technologies. We distinguish the following different levels of heterogeneity: Representational Heterogeneity, Reasoning Heterogeneity, and Default Heterogeneity.

Representational Heterogeneity means semantic data are represented by using different specification languages. Systems supporting Representational Heterogeneity would allow for semantic data specified by multiple semantic languages, rather than using a single metadata or ontology language, like OWL or RDF/RDFS. However, note that different representation of semantic data does not necessarily mean that they have different semantics. The problem of merging and aligning ontologies is a structural problem of knowledge engineering and it is always considered when developing an application of semantic technologies.

Urban Computing-related data can come from different and independent data sources, which can be developed with traditional technologies and modeling methods (e.g., relational DBMS) or expressed with "semantic" formats and languages (e.g., RDF/S, OWL, WSML); for example, geographic data are usually expressed in some geographic standard[11], events details are published on the Web in a variety of forms, traffic data are stored in databases; etc. The integration and reuse of those data, therefore, need a process of conversion/translation for the data to become useful together.

Reasoning Heterogeneity means the systems allow for multiple paradigms of reasoners. For instance, many applications of Urban Computing may need different reasoners for temporal reasoning, spatial reasoning, and causal reasoning. However, it does not necessarily mean that we have to develop a single but powerful reasoner which can cover all of those reasoning tasks. A system which supports Reasoning Heterogeneity would find a way to allow multiple single-paradigm-based reasoners to achieve the result of Reasoning Heterogeneity.

Some data related to Urban Computing need precise and consistent inference; e.g., knowing if two roads are connected for a given kind of vehicle; telling that at a given junction all vehicles, but public means of transportation, must go straight; checking if private cars are allowed to enter a specific urban area. Other data need approximate reasoning or imperfect estimations; e.g., calculating the

[11] http://en.wikipedia.org/wiki/Geographic_Data_Files

probability of a traffic jam given the current traffic conditions and the traffic history.

Therefore, the requirement is for different kinds of techniques and reasoners to deal with those kinds of data; moreover, another requirement is for a system which dynamically selects and runs a specific reasoner on the basis of the available data and the desired processing tasks.

By **Default Heterogeneity**, we mean that systems support for various specification defaults of semantic data. Well-known specification defaults of semantic data are closed world assumption, open world assumption, unique name assumption and non-unique name assumption. In the Semantic Web community, it is widely accepted that semantic data for the Web should take the open world assumption and the non-unique name assumption, as taken by the popular ontology language OWL.

However, as we have observed in many applications of Urban Computing, we should not commit to any single specification default. Take the example of traffic and transportation ontologies: although in many cases we can take the open world assumption and non-unique name assumption, because of our limited knowledge and information about the data, sometimes it is much convenient to take a *local* closed world assumption. For example, for a time table of a bus station, it is well reasonable to assume that the information about the bus schedule in the time table is locally complete, in the sense that if you cannot find any information about a bus which is scheduled at specific time, it would mean that there is no bus scheduled for that time. The same scenario is also applied to a city map: if there is no information which states a road connects two streets directly on the map, that would mean that there is no road which connects those two streets directly.

The same applies to the unique name assumption. Consider the use case in Section 2 and in particular the fact that Repubblica is both the name of the train station and of the subway station, but they are two different places. If MUCS has to calculate a trip and Carlo is aware that MUCS will use multiple means of transportation then MUCS can ignore that the two Repubblica stations are not exactly the same one. If, on the contrary, Carlo wants only to use subways, then MUCS cannot assume that the two Repubblica station are one physical place.

The examples above show that the semantic systems of Urban Computing should support multiple specification defaults and should provide users or knowledge engineers a high freedom degree to state any data with any reasoning assumption.

3.2 Coping with time-dependency

Knowledge and data can change over time. For instance, in Urban Computing names of streets, landmarks, kind of events, etc. change very slowly, whereas the number of cars that go through a traffic detector in five minutes changes very fast. This means that the system must have the notion of "observation period", defined as the period when the system is subject to querying.

Moreover the system, within a given observation period, must consider the following four different type of knowledge and data:

- *Invariable knowledge*:
  - It includes obvious terminological knowledge (such as an address is made up by a street name, a civic number, a city name and a ZIP code) and
  - less obvious nomological knowledge that describes how the world is expected to be (e.g., given traffic lights are switched off or certain streets are closed during the night) or to evolve (e.g., traffic jams appears more often when it rains or when important sport events take place).
- *Invariable data*: they do not change in the observation period, e.g. the names and lengths of the roads.
- *Periodically changing data*: they change according to a temporal law that can be
  - Pure periodic law, e.g. the fact that every night at 10pm Milan west-side overpass road closes; or
  - Probabilistic law, e.g. the fact that a traffic jam is present in the west side of Milan due to bad weather or due to a soccer match is taking place in San Siro stadium.
- *Event driven changing data*: they are updated as a consequence of some external event and they can be further characterized by the mean time between changes:
  - Fast, as an example consider the intensity of traffic (as monitored by sensors) for each street in a city;
  - Medium, as an example consider roads closed due to traffic accidents or congestion;
  - Slow, as an example consider roads closed for scheduled works.

3.3 Coping with Noisy, Uncertain and Inconsistent Data

We distinguish the following different levels of data uncertainty and inconsistency.

- Noisy Data: part of data are completely useless or semantically meaningless.
- Inconsistent Data: parts of data are logically self-contradictory or semantically impossible.
- Uncertain data: the semantics of data are partial, incomplete, or they are conceptually arranged into a range with multiple possibilities.

Traffic data are a very good example of such data. Different sensors observing the same road area give apparently inconsistent information. For example, a traffic camera may say that the road is empty whereas an inductive loop traffic detector may tell 100 vehicles went over it. The two information may be coherent if one consider that a traffic camera transmits an image per second with a delay of 15-30 seconds, whereas an inductive loop traffic detector tells you the number of vehicles when over it in 5 minutes and the information may arrive to you 5-10 minutes later.

Moreover, a single data coming from a sensor in a given moment may have no certain meaning. For example, consider an inductive loop traffic detector, it tells you 0 car went over, what does it mean? Is the road empty? Is the traffic completely stuck? Did somebody park the car above the sensor? Is the sensor broken? Combining multiple information from multiple sensors in a given time window can be the only reasonable way to reduce the uncertainty.

4 Partial Solutions

This work is part of the on-going research project LarKC which aims at building very large-scale manipulation of information ("semantic computing at Web scale"). We are envisioning a set of partial solutions to address the challenges of Urban Computing including: Traffic Prediction using recurrent neural networks, Data Scheduling to address scalability and Stream Reasoning to address time-dependency.

4.1 Predicting Traffic Using Recurrent Neural Networks

Given that a forecast model should focus on the underlying dynamics of the traffic flow and external influences on the traffic volume should be incorporated in the model, we intend to use time-delay recurrent neural networks for the traffic predictions [7]. With this approach we presume that the traffic volume is the outcome of an open dynamical system which combines an autonomous development with external influences (e.g. calendar effects, special events etc.).

Recurrent neural networks offer a new way to model (nonlinear, high dimensional) open dynamical systems based on time series data. Our recurrent neural networks are formulated as state space models in discrete time to identify the traffic dynamics and the impact of the external influences [8].

In state space formulation a recurrent neural network is described by a hidden state-transition- and an output-equation. The temporal equations are transformed into a spatial neural network architecture using shared weights (so-called unfolding in time) [9].

Prior knowledge about the application (e.g. topology of the traffic network or the temporal structure of the traffic flows) can be easily incorporated in the neural network architecture. For instance, an error correction mechanism can be used to consider the impact of unplanned construction sites, traffic accidents or holdups. This is also the key for robust forecasting [10].

4.2 Data Scheduling

The idea of data scheduling takes inspiration from memory management techniques developed and adopted in computer systems and software engineering (e.g., garbage collection, memory caching and direct memory access).

Large scale data are organized at different memory levels based on their relevance and on the context of applications: working data, which should be accessed by systems immediately without any over-heading cost; neighboring data, which can be accessed by the system with a moderate cost; and remote data, which can be accessed by the system with a significant amount of cost.

The research problem is finding automatic ways to move data from higher access cost memory into lower access cost memory and vice versa. Such memory shift should take place in parallel with reasoning.

4.3 Stream Reasoning

Periodically changing data and event driven changing data are best represented as data streams, unbounded sequences of time-varying data elements.

Data streams occur in a variety of modern applications, such as network monitoring, traffic engineering, sensor networks, RFID tags applications, telephone call records, financial applications, Web logs, click-streams, etc. The very nature of traffic management can be explained by means of data streams, representing real objects that are monitored at given locations: cars, trains, crowds, ambulances, parking spaces, and so on.

Processing of data streams has been largely investigated in the last decade [11] and specialized systems have been developed. While reasoners are year after year scaling up in the classical, time invariant domain of ontological knowledge, reasoning upon rapidly changing information has been neglected or forgotten. By coupling reasoners with powerful, reactive, throughput-efficient stream management systems, we introduce the concept of Stream Reasoning [12]. We expect future realization of such a concept to have a strong impact on Urban Computing because it enables reasoning in real time, at a throughput and with a reactivity not obtained in previous works.

5 Conclusions

In this paper we focus on presenting the Urban Computing challenge and in particular some requirements for future mobility management systems. We also presented some novel multi-disciplinary ideas about ways to address the Urban Computing challenge by partially satisfying one or more requirements. More solutions and, in particular, broader ones should be explored. As a matter of fact, if we were able to cope with requirements present in Section 3 we would be able to solve a broad range of Urban Computing problems. Such problems include:

– City Planning: Urban Computing applications can extract statistics and synthetic descriptions of citizens' movements, habits and opinions in order to position new housing complex, office buildings, shops, parking lots, green areas and to optimize public and private transportation routes and timetables. The City Planning can also lower pollution and enhance energy savings.

– Tourism and Culture: Urban Computing applications analyze tourists' movements and enhance the appeal of current places of interest and create targeted promotional campaigns to increase tourism.
– Public Safety: Urban Computing applications can perform continuous statistical analysis of people movements to find abnormal behavior and correlate them with the ones coming from law enforcement and public protection forces to enhance city safety.

Acknowledgments

This research has been partially supported by the LarKC project (FP7-215535).

References

1. Kindberg, T., Chalmers, M., Paulos, E.: Guest editors' introduction: Urban computing. IEEE Pervasive Computing **6**(3) (2007) 18–20
2. Arikawa, M., Konomi, S., Ohnishi, K.: Navitime: Supporting pedestrian navigation in the real world. IEEE Pervasive Computing **6**(3) (2007) 21–29
3. Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. IEEE Pervasive Computing **6**(3) (2007) 30–38
4. Bassoli, A., Brewer, J., Martin, K., Dourish, P., Mainwaring, S.: Underground aesthetics: Rethinking urban computing. IEEE Pervasive Computing **6**(3) (2007) 39–45
5. Fensel, D., van Harmelen, F., Andersson, B., Brennan, P., Cunningham, H., Della Valle, E., Fischer, F., Huang, Z., Kiryakov, A., il Lee, T.K., School, L., Tresp, V., Wesner, S., Witbrock, M., Zhong, N.: Towards larkc: a platform for web-scale reasoning, IEEE International Conference on Semantic Computing (ICSC 2008) (2008)
6. Fensel, D., van Harmelen, F.: Unifying reasoning and search to web scale. IEEE Internet Computing **11**(2) (2007)
7. Zimmermann, H.G., Neuneier, R.: Modeling dynamical systems by recurrent neural networks. In Ebecken, N., Brebbia, C., eds.: Data Mining II, WIT Press (2000) 557–566
8. Zimmermann, H.G., Neuneier, R.: Neural network architectures for the modeling of dynamical systems. In Kolen, J., Kremer, S., eds.: A Field Guide to Dynamical Recurrent Networks, IEEE Press (2001) 311–350
9. Haykin, S.: Neural Networks. A Comprehensive Foundation. Macmillan College Publishing, New York (1994)
10. Zimmermann, H.G., Neuneier, R., Grothmann, R.: Modeling of dynamical systems by error correction neural networks. In Soofi, A., Cao, L., eds.: Modeling and Forecasting Financial Data, Techniques of Nonlinear Dynamics, Kluwer Academic Publishers (2002) 237–263
11. Garofalakis, M., Gehrke, J., Rastogi, R.: Data Stream Management: Processing High-Speed Data Streams (Data-Centric Systems and Applications). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
12. Della Valle, E., Ceri, S., Barbieri, D.F., Braga, D., Campi, A.: A first step towards stream reasoning. In: Proceedings of the Future Internet Symposium. (2008)

Urban Computing: a challenging problem for Semantic Technologies

Emanuele Della Valle[1,2], Irene Celino[1], Daniele Dell'Aglio[1], Kono Kim[3], Zhisheng Huang[4], Volker Tresp[5], Werner Hauptmann[5], Yi Huang[5], and Ralph Grothmann[5]

[1] CEFRIEL – Politecnico of Milano, Via Fucini 2, 20133 Milano, Italy
{name.surname}@cefriel.it
[2] Dip. di Elettronica e Informazione, Politecnico di Milano, Milano, Italy
emanuele.dellavalle@polimi.it
[3] Saltlux Inc., Seul, Korea kono@saltlux.com
[4] Computer Science Department, Vrije Universiteit Amsterdam, De Boelelaan 1081, Amsterdam, The Netherlands huang@cs.vu.nl
[5] Corporate Technology, Siemens AG, Information and Communications, Munich, Germany {name.surname}@siemens.com

Abstract. In this paper we present the Urban Computing challenge and in particular we exemplify it in the context of traffic management. From our previous experiences in the field we draw requirements in terms of capacity to cope with heterogeneity in representation, semantics and defaults; with scale; with time-dependency of data; and with noisy, uncertain and inconsistent Data. Existing reasoning techniques fail to fulfill all those requirements at once.

1 Introduction

Our cities face many challenges well expressed by the following questions posed to the international community by the Urban Land Institute (http://www.uli.org/):

- How can we redevelop existing neighborhoods and business districts to improve the quality of life?
- How can we create more choices in housing, accommodating diverse lifestyles and all income levels?
- How can we reduce traffic congestion yet stay connected?
- How can we include citizens in planning their communities rather than limiting input to only those affected by the next project?
- How can we fund schools, bridges, roads, and clean water while meeting short-term costs of increased security?

The Information and Communication Technology (ICT) for sure cannot provide an answer to those question on its own, but it is one of the most important enabling factor. A signal that ICT for Urban area is growing at a recognizable size was the publication in 2007 of a special issue of IEEE Pervasive Computing dedicated to Urban Computing [1–4] – the integration of computing, sensing, and actuation technologies into everyday urban settings and lifestyles.

Urban settings range from our own cars, while we drive them in town, to public spaces such as streets and squares including semipublic ones like cafés and tourist attractions. Urban lifestyles are even broader and include people living, working, visiting and having fun in those settings. Not surprisingly, people constantly enter and leave urban spaces, occupying them with highly variable densities and even changing their usage patterns between day and night [3].

Some years ago, due to the lack of data, solving Urban Computing problems with ICT looked like a Sci-Fi idea. Nowadays, a large amount of the required information can be made available on the Internet at almost no cost: computerized systems contain maps with the commercial activities and meeting places (e.g., Google Earth), events scheduled in the city and their locations, positions and speed information of public transportation vehicles and of mobile phone users [3], parking availabilities in specific parking areas, and so on.

However, current ICT technologies are not up to the challenge of solving Urban Computing problems: this requires the combination of a huge amount of static knowledge about the city (i.e., urbanistic, social and cultural knowledge) with an even larger set of dynamic data (originating in real time from heterogeneous and noisy data sources) and reasoning above the resulting time-varying knowledge. A new generation of reasoners is clearly needed.

For this reason we are running the LarKC project, which is aiming at a configurable platform for infinitely scalable semantic web reasoning [5, 6], to address one of the Urban Land Institute open questions: "how can we reduce traffic congestion yet stay connected?". We select such question because we have been working in this area for years and we can derive from our previous experiences challenging requirements not only for the LarKC project, but also for the entire community working on scalable, tolerant and dynamic reasoning.

In the rest of the paper, we present a Story Board to make our vision more explicit (Section 2). We identify traffic management (Section 3) as a special case of Urban Computing and we present state-of-the-art in this field. Then, we describe our past implementation experiences (Section 4), from which we derive requirements (Section 5) for LarKC project and the entire community. In Section 6, we provide a short description of the solutions we are working on, claiming that they are only partial and that new reasoning techniques are needed to fulfill all the requirements of Urban Computing at once. Finally, in Section 7 we give a broader vision of the potential impact of solving the Urban Computing challenge.

2 A Story Board: Getting to Milano

This use case shows the added value of creating a Urban Computing System (UCS) that collects a broad set of information around traffic, integrates it and use them to support a citizen that has to go to Milano from another city nearby.

1. Carlo lives in Varese 60 km from Milano. The day after, he has to go to the Lombardy Region premises in Milano where he arranged a meeting at 11.00.
2. He opens the "plan a travel" service of UCS and fills in the required data
   - FROM: via Luigi Sacco, 1 Varese after 8.00 tomorrow

  - TO: via Taramelli, 20 Milano before 11.00 tomorrow
  - USING: any means of transportation
3. UCS provides Carlo with three alternatives, two using public transportations (i.e., A and B) and one by using his private car (i.e., C)
   (a) Using railroad "LeNord" and Metro M3; leaving home at 8.30 and arriving between 10.15 and 10.30.
   (b) Using railroad "Ferrovie dello Stato" (alternative to "LeNord") and Metro M3; leaving home at 8.20 and arriving between 10.05 and 10.20.
   (c) Using private car; leaving home after 9.30 (when the commuters traffic on motorway A8 is almost over) and arriving between 10.10 and 10.40.
4. Carlo is tempted by option C, he could sleep a little longer, but while traveling by train he could complete the presentation for the meeting, so he chooses the option A and he uses the ticket-less option to buy the train ticket.
5. Before exiting UCS, Carlo asks to be alerted if the option he chose will no longer be the best one (e.g., due to problems to the railroads).
6. The day after (25.9.2008) at 7.14 UCS learns from the railroad information system of railroad "LeNord" that a technical problem is causing an average delay of 45 minutes to all "leNord" trains from Varese to Milano.
7. UCS estimates that an accident of such kind will not be solved before 11.00, therefore it checks if any planned travel is at risk. It finds Carlo's travel.
8. UCS checks if the other options it proposed to Carlo are still valid. Apparently they are, so UCS sends an SMS to Carlo informing him that a accident is causing 45 minutes delay for all trains on railroad "LeNord" and he can either use the the railroad "Ferrovie dello Stato" (option B) or take his private car, in this case Carlo can convert his train ticket into a daily parking ticket for one of the parking lots of the sub-urban metro stations in Milano.
9. Carlo receives the SMS, he enters UCS and checks the two alternatives. He can take option B, but he knows that when problem of this kind happens on "LeNord", all commuters take the railroad "Ferrovie dello Stato" and he will never be able to find a sit. On the other hand, UCS (taking into account weather data through the route and real time traffic congestion status on top of historical traffic congestion statistics) predicts that being a rainy day the traffic on A8 will be slower and he should leave around 9.00.
10. Carlo decides to take his car, in this way he has all the time to complete the presentation before leaving. He leaves home around 9.00 and instructs its GPS Navigator to interact with UCS traffic service and to find the cheapest gas station along the road. During the driving Carlo receives the instructions for the gas station and refuels the car.
11. At a certain moment his GPS Navigator receives alert from UCS: Milano North-West area is hit by heavy showers and the traffic is getting slower. Instead of going to a parking lot to the planned North-West parking lot, the GPS suggests to go to one in the South-West; the metro from there will only take 10 minutes more than from the planned parking lot, but the estimated time to the planned parking lot is 25 minutes more than the planned one.
12. Carlo considers the option and decides to follow it.
13. Carlo parks the car and taking the metro arrives in time to his appointment.

3 Traffic Management

Traffic demand has been growing steadily for decades and this growth is foreseen to continue in the future. For many years, the primary way of dealing with this increasing demand has been the increase of the roadway network capacity, by building new roads or adding new lanes to existing ones. However, financial and ecological considerations are posing increasingly severe constraints on this process. Hence, there is a need for additional intelligent approaches designed to meet the demand while more efficiently utilizing the existing infrastructure.

Public authorities have taken steps in this direction through the installation of traffic management systems intended to equalize traffic demand both temporally (by spreading out trips in time) and spatially (by redistributing demand). The domain of traffic management solutions has recently experienced a significant demand of advanced information technology. Control centers for traffic management are online connected to different devices (such as detectors on roads and traffic lights) making it feasible for operators to supervise the state of the road network by consulting data bases with recent information from detectors and to modify the state of control devices. The use of such traffic monitoring and management facilities requires sophisticated tools for traffic modeling, estimation, prognosis and decision support for online operators to help them in dealing with the information complexity and diversity of sensor data and control devices.

Traffic System Infrastructure Setup. Today, in a typical information infrastructure for real-time traffic control, as it can be found in different cities, usually the following basic components can be discriminated. There are sensors (e.g. loop detectors, cameras, traffic eyes, radar detectors) on major roads recording several traffic magnitudes such as vehicle speed (km/h), traffic flow (vehicle/h) and occupancy or traffic density, i.e., the percentage of time the sensor is occupied by a vehicle (vehicles/km). The distance between successive sensors on a freeway is typically in the order of about 500 meters. The information is periodically transmitted to a control center, which also receives information about the current state of control devices. Such control devices include traffic signals at intersections, traffic signals at sideways entry-ramps, variable message signs that can display different messages to drivers (e.g., warning about existing congestions, accidents or alternative path recommendations), radio advisory systems to broadcast messages to drivers, and reversible lanes (i.e., freeway lanes whose direction can be selected according to the current and expected traffic demand). In the control center, operators interpret the sensor data and detect the presence of problems and their possible causes. Problems are congested areas at certain locations caused by lack of capacity due to accidents, excess of demand, like rush hours, etc. In addition, operators determine control actions to solve or reduce the severity of existing problems (e.g. extended green-phase of a traffic signal, switch of variable message signs).

Traffic management systems must be reactive to the different states of traffic flow in the controlled network. In the early systems, the approach was based on a library of signal plans applied online in different predefined situations according to some time-based criteria or to the traffic data collected by roadside sensors. However, this precalculated-plan approach usually lacked the conceptual granularity required by the system to be adaptive enough to the variety of situations, in time and space, which may occur in the network. Later, more adaptive systems were introduced, where an intelligence for understanding traffic situations in real time was designed and integrated with a model for decision making.

In the last years, a considerable amount of work has concentrated on the fields of traffic modeling and estimation and the analysis and forecast of traffic conditions. One of the main tasks in traffic management systems is to model the current traffic condition for the entire road network. Traffic flow is normally only measured at certain points along the roadway. Employing appropriate models this data is used to estimate traffic conditions for the major part of the network. Here, different approaches are applied: methods for statistical evaluations and visualization of stationary and mobile traffic data, methods for evaluating and projecting traffic correlations from current and historical traffic flows and propagation methods for calculating the current traffic condition on the basis of origin-destination matrices as well as on statistical analysis of traffic data surveyed online. In general, the propagation method is based on the assumption that the traffic volume measured at a cross section is a superposition of different traffic flows. They branch out before and after the cross section within the network. From the assignment calculation the operator knows the different traffic flows which amount to the measured value. As a result, it is possible to allocate the percentage of each traffic flow to the routes within the network. The propagation method allows the user to dynamically visualize congestion impacts which are plausibly running upstream over several time slices.

Recent developments not only consider stationary traffic data provided by standard detectors but also allow to integrate so-called floating car data, as an increasing number of operators of advanced traffic management systems also use mobile traffic data. With this mobile data the level of detail of analysis and forecast methods can be considerably improved and information about areas not covered by roadside detectors can be provided.

Evolution of Traffic Management Solutions. Despite the increasing sophistication of the traffic management and control infrastructures run by public authorities, such collective systems for traffic management suffer from several limitations. One aspect is that they are unable to provide continuous, up-to-the-minute information to drivers. Another aspect is that it is impossible to restrict an advice or control action to a targeted subgroup of drivers, for example, those with a particular destination area. The options in collective route guidance are essentially "all or nothing". On the other hand, private service operators will in the future be providing an increasing level of traffic information services targeted to the user's need for the fastest or shortest route. Modern navigation systems are designed to take delays due to incidents or congestions dynamically into account, provided that these delays have been previously reported and transmitted by some means to the device, e.g., by broadcast media such as TMC (Traffic Message Channel) in Europe. The reaction of present route guidance systems to delays and incidents is a short-term and/or small-scale strategy. Comprehensive optimization of dynamic routing strategies is not provided. In addition, current systems are unable to include the influence of public traffic management strategies, such as traffic signal coordination. Therefore, the integration of commercial route guidance recommendations with public policy and collective interests as well as advancing the further development of vehicle navigation are primary issues for the evolution of traffic management solutions. Particularly, new approaches and algorithms have to be provided for dynamically incorporating diverse available sources of traffic data and information as well as public traffic management and control strategies and priorities into integrated comprehensive systems. These kinds of integrated route guidance systems are also referred to as third-generation or traffic-responsive navigation solutions. They are to enable services targeted to the drivers' needs such as recommendation of routes with coordinated traffic signals in urban areas or congestion-free alternative routes on motorways.

Possible Approaches for Future Real-time Traffic Management. An analysis of the current and predicted traffic state in the entire road network and the identification of reserve capacities comprise the basis for advanced city traffic management and navigation solutions. Mobile and stationary sensors collect the appropriate traffic data and transmit it to a central unit. Similarly to the weather forecast, the different and heterogeneous information sources are combined to obtain an estimation of the traffic state during a period ranging from minutes to hours or even longer. Thus, a comprehensive knowledge base can be built up to support optimal individual route guidance.

In order to obtain a high-resolution picture of the current traffic state as well as of weather conditions and other environmental factors, current research activities are directed towards both utilizing existing stationary detection facilities, such as loop detectors, and advanced vehicle-based data source known as XFCD (eXtended Floating Car Data). Floating cars act as mobile sensors and can collect a range of information including speed and position data. During a trip, XFCD-vehicles perform an analysis using position, speed, and other data that gives important information on the local traffic state as well as the traffic context and surroundings (e.g., dynamic control systems, rain sensor, driver assistance systems, braking activity). If there is relevant information available it is transmitted anonymously to a traffic center and fused with other data sources. The advantage of this distributed data source is that measurements of traffic occurrences are possible in principle within the entire road network without the requirement of expensive stationary infrastructure.

The quality of dynamic route guidance crucially depends on the quality of the available dynamic traffic state reconstruction in the road network. The higher the quality of the reconstruction, the more reliable the traffic prediction. As mentioned above, one important contribution to this quality could be provided by comprehensive measurement of XFCD. In addition, further concepts envision the integration of the origins, destinations and route informations as planned by the individual onboard navigation systems of vehicles, which are provided to the traffic management center. On the other hand, innovative technologies are required in order to process and integrate the resulting collection of distributed information bits within a complex, diverse information environment. Here, a major task is the provision of appropriate solutions for the integration and fusion of heterogeneous information sources, where each source of information can have distinct characteristices with respect to availability, precision, reliability, resolution and representation (see Section 5).

4 Implementation experiences

u-City – Intelligent Transportation System. u-City is a new intelligent real-time city project currently in progress in Korea[6]. The goal of the project is to build a ubiquitous computing based environment. Roads, cars and buildings that physically exist, plus all things that take up electronic space, cell phones, PDA, DMB devices are modeled in a formal manner and are interlinked. As a result a new 3D city space is created. u-City offers an infrastructure through which any information can be access anytime, from anywhere, using any device without any obstruction in a seamless connection.

One of u-City objective is to provide a transportation system that allows for an increased efficiency in the whole city as well as for decreasing the city operation fees. In particular we noticed that many users of GPS navigators complain about the inability of such systems to anticipate road conditions. At the same time, Intelligent Transportation System (ITS) are able to accumulated data from the sensors located on city roads and analyzes them in real time, but it is currently not possible to embed an ITS in a GPS navigator.

So we decide to realize a u-City service that interlinks ITS with GPS navigators and exploits semantic technology to provide a new type of transportation service that increases the satisfaction of the urban inhabitants (see Figure 1). Such service addresses is able to change course, to recommend different road courses, and to send information to ITS for further automatic actions.

We developed rules in F-logic for ITS, by expressing all situations of the car, road and traffic conditions as ontology assertions and axioms. Our ontology is similar, in terms of capability of supporting the computing of the best trip, to the one [7] developed by the European FP6 Project REWERSE (IST-506779), but it is less focused on general terms and it has ore sophisticated features that allows for real-time traffic management. To exemplify how we use rules, in the following listing we show the one we used for calculating expected traffic resolution time when a traffic accident occurred.

[6] See for example `http://www.udongtan.or.kr/english/cyber/cyb_01_7.aspx`

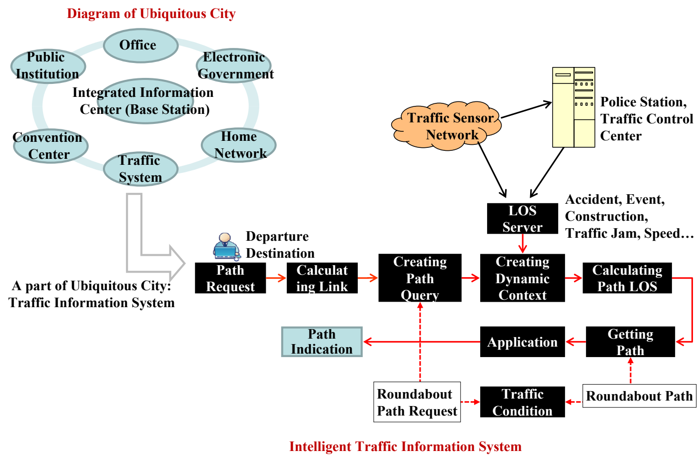

Fig. 1. Overall system flow of ITS

IF takeIn(TrafficAccident, Link) AND hasCarSituation(TrafficAccident, CarSituation(fire)) AND hasRelatedTrafficAccidentStat(TrafficAccident, TrafficAccidentStat) AND hasCarSituation(TrafficAccidentStat, CarSituation(fire)) AND solvedTime(TrafficAccidentStat, X) AND necessaryTime(TrafficAccidentStat, Y) AND add(X,Y,Z)
THEN ofEvent(TrafficAccidentSolvedTime, TrafficAccident) AND relatedLink(TrafficAccidentSolvedTime, Link) AND hasEventTime(TrafficAccidentSolvedTime, TrafficEventTime) AND hasRelatedTrafficAccidentStat(TrafficAccidentSolvedTime, TrafficAccidentStat) AND solvedTime(TrafficAccidentSolvedTime, Z)

The u-City navigation system takes the departure and destination points and, in real time, suggests to the driver alternative routes based on the real situation. The system proved to be effective in various traffic conditions and it expanded the possibilities of a traditional navigation service.

Our conclusions are twofold; on the one hand we are convinced that our ontology based system is an extremely effective navigation method; on the other hand we face significant technological challenges in terms of scalability and reasoning performance.

Siemens Previous Experiences in the Field. The department of learning systems at Siemens Corporate Technology (https://www.ct.siemens.com/) has a lot of research and project experience in the field of traffic modeling and forecasting. Since 15 years they have been using machine learning techniques, e.g., fuzzy clustering, neural networks and reinforcement-learning to model, predict and optimize traffic flows. The following papers represent only a brief summary of our work.

In 1995 Hellendorn and Baudrexl combined fuzzy methods and feedforward neural networks to control and to forecast traffic [8]. They built a fuzzy system for traffic flow control and incident recognition that has been in use for some time. Furthermore, the system was used in forecasting whether a particular parking garage is full or not.

An approach to traffic modeling is presented by Wagner et al. in 1996 [9]. The authors derived a second-order traffic flow model from microscopic equations. The model incorporated different driver characteristics on the microscopic level called the desired velocity. The authors explored dynamical quantities for the mean and variance of the desired velocity, and the covariance between actual and desired velocity. Through these quantities an alternative explanation for the onset of traffic clusters can be given, i.e., a spatial variation of the variance of the desired velocity can cause the formation of a traffic jam.

Lenz et al. extended the microscopic car-following model by Bando et al. by incorporating multi-vehicle interactions [10]. The authors showed that the reaction to more than one vehicle ahead leads to a stabilization of the dynamical behavior, i.e., the stable region increases. The fundamental macroscopic properties of traffic, namely free flow and congested flow, were still described. Due to the multi-anticipative driving behaviour driving in narrow platoons is forced such that a third fundamental property of traffic flow, the so-called synchronized flow is modelled as well. Related work on traffic modelling can be found in [11]. Here, anticipative schemes are used to switching between speed limits based on the density of the downstream segment.

Stutz and Runkler classified and predicted road traffic by using an application specific fuzzy clustering approach [12]. The authors used fuzzy methods for traffic data analysis. The results of the data analysis were classification and prediction systems. The work was focused on fuzzy clustering methods. The known clustering models were extended to: constrained prototypes, the use of a mix of different prototypes for one data set, partial supervision of the clustering, and the estimation of the number of clusters by cluster merging. A successful application example was given for the classification of traffic jam on a German motor highway.

Within the public funded research project LEONET[7] Appl and Sollacher used adaptive learning systems for intelligent traffic control. In the paper [13] the authors described the application of modern reinforcement-learning approaches for the automatic control of groups of traffic lights. As an extension to this work Sollacher and Klein applied feedforward neural networks in combination with information of origin-destination traffic flows to forecast upcoming traffic volumes in short-term (up to one week in 15 min. time buckets). The feedforward neural network incorporated a so-called bottleneck coordinate transformation to cluster the daily traffic variation curves.

The bottleneck enables us to identify the non-linear principal components of the traffic variation curves. In order to predict the future development of the traffic volume the neural network concentrates on the principal components. The dimensionality of the forecasting problem is therefore dramatically reduced. The principal components are forecasted by taking into external influences. The reconstruction of the traffic variation curve for every time step of the future is assured by the reuse of the bottleneck network [14]. A long-term forecast models

[7] `http://www.inb.uni-luebeck.de/research/leonet`

has also been developed on the basis of this clustering approach of daily traffic variation curves and the estimation of origins and destinations.

5 Requirements

In this section, we will investigate requirements of Urban Computing. As argued before, we are particularly interested in the reasoning requirements for LarKC, but we believe such requirements interesting for the entire community working on the complex relationship of the Internet with space, places, people and content.

Coping with Heterogeneity. Dealing with heterogeneous data has been appealed for long time in many areas in computer science and engineering, which include database systems, multimedia application, network systems, and artificial intelligence. Here, we would like to propose a comprehensive notion of heterogeneity processing for semantic technologies. We distinguish the following different levels of heterogeneity: Representational Heterogeneity, Semantic Heterogeneity, and Default Heterogeneity.

Representational Heterogeneity means semantic data are represented by using different specification languages. Systems supporting Representational Heterogeneity would allow for semantic data specified by multiple semantic languages, rather than using a single metadata or ontology language, like OWL or RDF/RDFS. However, note that different representation of semantic data does not necessarily mean that they have different semantics. The problem of merging and aligning ontologies is a structural problem of knowledge engineering and it is always considered when developing an application of semantic technologies.

Urban Computing-related data can come from different and independent data sources, which can be developed with traditional technologies and modeling methods (e.g., relational DBMS) or expressed with "semantic" formats and languages (e.g., RDF/S, OWL, WSML); for example, geographic data are usually expressed in some geographic standard[8], events details are published on the Web in a variety of forms, traffic data are stored in databases; etc. The integration and reuse of those data, therefore, need a process of conversion/translation for the data to become useful together.

Reasoning Heterogeneity means the systems allow for multiple paradigms of reasoners. For instance, many applications of Urban Computing may need different reasoners for temporal reasoning, spatial reasoning, and causal reasoning. However, it does not necessarily mean that we have to develop a single but powerful reasoner which can cover all of those reasoning tasks. A system which supports Reasoning Heterogeneity would find a way to allow multiple single-paradigm-based reasoners to achieve the result of Reasoning Heterogeneity.

Some data related to Urban Computing need precise and consistent inference; e.g., knowing if two roads are connected for a given kind of vehicle; telling that at a given junction all vehicles, but public transportation ones, must go straight;

[8] http://en.wikipedia.org/wiki/Geographic_Data_Files

checking if private cars are allowed to enter a specific urban area. Other data need approximate reasoning or imperfect estimations; e.g., calculating the probability of a traffic jam given the current traffic conditions and the past history. Therefore, the requirement is for different kinds of techniques and reasoners to deal with those kinds of data; moreover, another requirement is for a system which dynamically selects and runs a specific reasoner on the basis of the available data and the desired processing tasks.

By **Default Heterogeneity**, we mean that systems support for various specification defaults of semantic data. Well-known specification defaults of semantic data are closed world assumption, open world assumption, unique name assumption and non-unique name assumption. In the Semantic Web community, it is widely accepted that semantic data for the Web should take the open world assumption and the non-unique name assumption, as taken by the popular ontology language OWL.

However, as we have observed in many applications of Urban Computing, we should not commit to any single specification default. Take the example of traffic and transportation ontologies: although in many cases we can take the open world assumption and non-unique name assumption, because of our limited knowledge and information about the data, sometimes it is much convenient to take a *local* closed world assumption. For example, for a time table of a bus station, it is well reasonable to assume that the information about the bus schedule in the time table is locally complete, in the sense that if you cannot find any information about a bus which is scheduled at specific time, it would mean that there are no bus scheduled for that time. The same scenario is also applied to a city map: if there is no information which states a road connects two streets directly on the map, that would mean that there is no road which connects those two streets directly.

The examples above show that the semantic systems of Urban Computing should support multiple specification defaults. It should allow users or knowledge engineers feel free to state any data with any reasoning assumption. Some part of semantic data may be based on the open world assumption, and some part may be based on the closed world assumption.

Coping with Scale. The advent of Pervasive Computing and Web 2.0 technologies led to a constantly growing amount of data about urban environments, like information coming from multiple sensors (traffic detectors, public transportation, pollution monitors, etc.) as well as from citizens' observation (black points, commercial activities' ratings, events organization, etc.). The result, however, is that the amount of data available to be used and integrated is not manageable by state-of-the-art technologies and tools and a severe focus on scalability issues must be taken into account. For example, intelligent methods for data sampling or selection should be adopted before employing traditional reasoning techniques, e.g. to select traffic data to employ in predictions.

Although we encounter large scale data which are not manageable, it does not necessary mean that we have to deal with all of the data simultaneously.

Usually, there are only very limited amount data which are relevant for a single query/processing at a specific application. For example, when Carlo is driving to Taramelli, Milano, only part of the Milano map data are relevant. Furthermore, it is impossible and unnecessary to store large scale data at the same memory level which are easily accessible for the processing. The idea of the data scheduling is to move relevant data from a memory level to another memory level in advance to make them easier accessible. For example, when Carlo concern the railroad information during the travel, the local parking information may become active by a prediction of the causal relation between abortive driving and parking. We consider this idea of the data scheduling as a partial solution for the scalability, which will be discussed in the next section.

Coping with time-dependency. Knowledge and data can change over the time. For instance, in Urban Computing names of streets, landmarks, kind of events, etc. change very slowly, whereas the number of cars that go through a traffic detector in five minutes changes very fast. This means that the system must have the notion of "observation period", defined as the period when we the system is subject to querying.

Moreover the system, within a given observation period, must consider the following four different type of knowledge and data:

- *Invariable knowledge*:
  - it includes obvious terminological knowledge (such as an address is made up by a street name, a civic number, a city name and a ZIP code) and
  - less obvious nomological knowledge that describes how the world is expected to be (e.g., given traffic lights are switched off or certain streets are closed during the night) or to evolve (e.g., traffic jams appears more often when it rains or when important sport events take place).
- *Invariable data* do not change in the observation period, e.g. the names and lengths of the roads.
- *Periodically changing data* change according to a temporal law that can be
  - Pure periodic law, e.g. every night at 10pm Milano overpasses close.
  - Probabilistic law, e.g. traffic jam appear in the west side of Milano due to bad weather or when San Siro stadium hosts a soccer match.
- *Event driven changing data* are updated as a consequence of some external event. They can be further characterized by the mean time between changes:
  - Fast, e.g. the intensity of traffic for each street in a city;
  - Medium, e.g. roads closed for accidents or congestion due to traffic;
  - Slow, e.g. roads closed for scheduled works.

Coping with Noisy, Uncertain and Inconsistent Data. We distinguish the following different types of data:

- Noisy Data: part of data are useless or semantically meaningless.
- Inconsistent Data: parts of data are in logical contradiction with each another, or are semantically impossible.

– Uncertain data: the semantics of data are partial, incomplete, or they are conceptually arranged into a range with multiple possibilities.

Traffic data are a very good example of such data. Different sensors observing the same road area give apparently inconsistent information. For example, a traffic camera may say that the road is empty whereas an inductive loop traffic detector may tell 100 vehicles went over it. The two information may be coherent if one consider that a traffic camera transmits an image per second with a delay of 15-30 seconds, whereas a traffic detector tells the number of vehicles that went over it in 5 minutes and the information may arrive 5-10 minutes later.

Moreover, a single data coming from a sensor in a given moment may have no certain meaning. For example, consider an inductive loop traffic detector, it it tells you 0 car went over, what does it mean? Is the road empty? Is the traffic completely stuck? Did somebody park the car above the sensor? Is the sensor broken? Combining multiple information from multiple sensors in a given time window can be the only reasonable way to reduce the uncertainty.

6 Partial Solutions

Within the LarKC, We are envisioning a set of partial solutions to address the challenges of Urban Computing including: Traffic Prediction using recurrent neural networks, Data Scheduling to address scalability and Stream Reasoning to address time-dependency.

Predicting Traffic Using Recurrent Neural Networks. Given that a forecast model should focus on the underlying dynamics of the traffic flow and external influences on the traffic volume should be incorporated in the model, we intend to use time-delay recurrent neural networks for the traffic predictions [15]. With this approach we presume that the traffic volume is the outcome of an open dynamical system which combines an autonomous development with external influences (e.g. calendar effects, special events etc.). Recurrent neural networks offer a new way to model (nonlinear, high dimensional) open dynamical systems based on time series data. Our recurrent neural networks are formulated as state space models in discrete time to identify the traffic dynamics and the impact of the external influences. In state space formulation a recurrent neural network is described by a hidden state-transition- and an output-equation. The temporal equations are transformed into a spatial neural network architecture using shared weights (so-called unfolding in time). Prior knowledge about the application (e.g. topology of the traffic network or the temporal structure of the traffic flows) can be easily incorporated in the neural network architecture. For instance, an error correction mechanism can be used to consider the impact of unplanned construction sites, traffic accidents or holdups. This is also the key for robust forecasting [16].

Data Scheduling. The idea of data scheduling takes inspiration from memory management techniques developed and adopted in computer systems and software engineering (e.g., garbage collection, memory caching and direct memory access). Large scale data are organized at different memory levels based on their relevance and on the context of applications: working data, which should be accessed by systems immediately without any over-heading cost; neighboring data, which can be accessed by the system with a moderate cost; and remote data, which can be accessed by the system with a significant amount of cost. The research problem is finding automatic ways to move data from higher access cost memory into lower access cost memory and vice versa. Such memory shift should take place in parallel with reasoning.

Stream Reasoning. Periodically changing data and event driven changing data are best represented as data streams. Processing of data streams has been largely investigated in the last decade [17] and specialized systems have been developed. While reasoners are year after year scaling up in the classical, time invariant domain of ontological knowledge, reasoning upon rapidly changing information has been neglected or forgotten. By coupling reasoners with powerful, reactive, throughput-efficient stream management systems, we introduce the concept of Stream Reasoning [18]. We expect future realization of such a concept to have a strong impact on Urban Computing because it enables reasoning in real time, at a throughput and with a reactivity not obtained in previous works.

7 Vision and Conclusions

In this paper we focus on presenting the Urban Computing challenge and in particular some requirements for future mobility management systems. We also presented some novel multi-disciplinary ideas about ways to address the Urban Computing challenge by partially satisfying one or more requirements. More solutions and, in particular, broader ones should be explored. As a matter of fact, if we were able to cope with requirements present in Section 5 we would be able to solve a broad range of Urban Computing problems.

City Planning. Urban Computing applications can extract statistics and synthetic descriptions of citizens' movements, habits and opinions in order to position new housing complex, office buildings, shops, parking lots, green areas and to optimize public and private transportation routes and timetables. The City Planning can also lower pollution and enhance energy savings.

Tourism and Culture. Urban Computing applications analyze tourists' movements and enhance the appeal of current places of interest and create targeted promotional campaigns to increase tourism.

Public Safety. Urban Computing applications can perform continuous statistical analysis of people movements to find abnormal behavior and correlate them with the ones coming from law enforcement and public protection forces to enhance city safety.

Acknowledgments

This research has been partially supported by the LarKC EU-funded project (FP7-215535). For more information visit `http://www.larkc.eu`.

References

1. Kindberg, T., Chalmers, M., Paulos, E.: Guest editors' introduction: Urban computing. IEEE Pervasive Computing **6**(3) (2007) 18–20
2. Arikawa, M., Konomi, S., Ohnishi, K.: Navitime: Supporting pedestrian navigation in the real world. IEEE Pervasive Computing **6**(3) (2007) 21–29
3. Reades, J., Calabrese, F., Sevtsuk, A., Ratti, C.: Cellular census: Explorations in urban data collection. IEEE Pervasive Computing **6**(3) (2007) 30–38
4. Bassoli, A., Brewer, J., Martin, K., Dourish, P., Mainwaring, S.: Underground aesthetics: Rethinking urban computing. IEEE Pervasive Computing **6**(3) (2007) 39–45
5. Fensel, D., van Harmelen, F., Andersson, B., Brennan, P., Cunningham, H., Della Valle, E., Fischer, F., Huang, Z., Kiryakov, A., il Lee, T.K., School, L., Tresp, V., Wesner, S., Witbrock, M., Zhong, N.: Towards larkc: a platform for web-scale reasoning, IEEE International Conference on Semantic Computing (2008)
6. Fensel, D., van Harmelen, F.: Unifying reasoning and search to web scale. IEEE Internet Computing **11**(2) (2007)
7. : Ontology of Transportation Networks. Technical report, REWERSE Project (2005)
8. Hellendoorn, H., Baudrexl, R.: Fuzzy neural traffic control and forecasting. In: Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE International Conference on. Volume 4. (1995) 2187–2194
9. Hoffmann, C., Sollacher, R., Wagenhuber, J., Schürmann, B.: Second-order continuum traffic flow model. In: Phys. Rev. E 54. (1996) 5073 – 5085
10. Lenz, H., Wagner, C.K., Sollacher, R.: Multi-anticipative car-following model. In: Eur. Phys. J. B 7. (1998) 331–335
11. Lenz, H., Sollacher, R., Lang, M.: Standing waves and the influence of speed limits. In: Proceedings of the European Control Conference 2001, Porto, Portugal (2001)
12. Stutz, C., Runkler, T.: Classification and prediction of road traffic using application-specific fuzzy clustering. In: IEEE Transactions on Fuzzy Systems. Volume 10. (2002) 297–308
13. Appl, M., Sollacher, R.: Intelligent traffic control with adaptive, learning systems. In: Automatisierungstechnik. Volume 49. (2001) 512
14. Zimmermann, H.G., Neuneier, R., Grothmann, R.: Modeling of the german yield curve by error correction neural networks. In Leung, K., Chan, L.W., Meng, H., eds.: Proceedings Intelligent Data Engineering and Automated Learning 2000 (IDEAL 2000), Hong Kong (2000) 262–267
15. Zimmermann, H.G., Neuneier, R.: Modeling dynamical systems by recurrent neural networks. In Ebecken, N., Brebbia, C., eds.: Data Mining II, WIT Press (2000) 557–566
16. Zimmermann, H.G., Neuneier, R., Grothmann, R.: Modeling of dynamical systems by error correction neural networks. In Soofi, A., Cao, L., eds.: Modeling and Forecasting Financial Data, Techniques of Nonlinear Dynamics, Kluwer Academic Publishers (2002)
17. Garofalakis, M., Gehrke, J., Rastogi, R.: Data Stream Management: Processing High-Speed Data Streams (Data-Centric Systems and Applications). Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
18. Della Valle, E., Ceri, S., Barbieri, D.F., Braga, D., Campi, A.: A first step towards stream reasoning. In: Proceedings of the Future Internet Symposium. (2008)

Urban Computing: State of the Art

Daniele Dell'Aglio[1], Irene Celino[1], Emanuele Della Valle[1,2], Yi Huang[3], and Kono Kim[4]

[1] CEFRIEL – Politecnico of Milano, Via Fucini 2, 20133 Milano, Italy
Daniele.Dell'aglio@students.cefriel.it
{Irece.Celino|Emanuele.Dellavalle}@cefriel.it
[2] Dip. di Elettronica e Informazione, Politecnico di Milano, Milano, Italy
emanuele.dellavalle@polimi.it
[3] Corporate Technology, Siemens AG, Information and Communications, Munich, Germany yi.huang@siemens.com
[4] Saltlux Inc., Seul, Korea kono@saltlux.com

Abstract. Cities are alive: they born, grow, evolve like living beings. The state of a city changes continuously, influenced by a lot of factors, both human (people moving in the city or extending it) and natural ones (precipitations or climate changes).
Cities are potentially huge sources of data of any kind and in our era, the information age, some technologies are ready to be used to create and extract these source. These data will allow the development of new services that will improve the quality of life of citizens. These are the main purposes of Urban Computing.
In this work we will present the state of the art of Urban Computing, deepening on what kinds of fields it involves and works done about this topic (like applications, conferences, and projects).

1 Introduction

Urban Computing proposes a set of different issues, from technological to social ones, to create a new kind of cities where they can actively interact with citizens [1, 2]. In this section we will present these topics and how they are connected with Urban Computing.

1.1 Technological issues

Urban Computing uses methodologies and instruments produced by a set of other disciplines, and in the meantime it helps them to grow further. The main important contribution in Urban Computing (also knows as Pervasive Computing and Ubicomp) is given by Ubiquitous Computing [3]; this term, conied 30 years ago by Mark Weiser [4], refers to a new vision of machines: no longer a "traditional" computer with keyboards (and mouses) to give inputs and a monitor to watch the outputs, but a new kind of devices integrated in everything.

Ubiquitous Computing develops projects to find new kind of interfaces (i.e. voice or gestures) and how to use them in ordinary contexts: from houses (the field of domotic) to cities [5]. A fundamental requirement of Urban Computing is that cities should have input/output devices to interact with environment and citizens [6], therefore Ubiquitous Computing studies how to make it possible. Scenarios with these kinds of devices may sound as sci-fi but are closer to reality than we may think: nowadays mobile devices, cameras, screens are everywhere. Then communication wireless technologies, like Bluetooth, UMTS and 802.11 allow all these devices to interact each others [7].

Another important input is given by research in the field of Geoinformatics [8]. Two important fields for Urban Computing development are Geocoding and Geolocation: Geocoding [9] is the process to join generic data with information about location (i.e. coordinates); Geolocation is a set of methodologies and instruments to discover the geographic location of different devices (often mobile devices like PDAs, notebooks).

1.2 Modeling issues

One of the most relevant problems in Urban Computing is how to model the city. The first (and one of the simplest) idea is to represent it with a map: streets, squares, rivers, rails, etc., their names and their location. But such models are very approximative: there are one-way streets, traffic lights, traffic islands, etc. As we can expect, it is always possible to increase considerably the level of detail: it is possible to add information about traffic, locals, bus stops, weather, pollution. However, whenever dealing with models, no perfect model of a city can be drawn, but only models that are good enough for a given purpose [10].

It's important to have a model of the city to enable the collection and the processing of data [11]. Ubiquitous Computing uses instruments similar to sensors and actuators in robotics: with them it is possible to get data from the environment (including people) and to interact with it; but without an appropriate model, it is not possible to store information retrieved and to process them in an efficient way.

1.3 Social issues

The goal of Urban Computing is to improve citizens' quality of life. They are the final users, so it's necessary to take them into account when designing and choosing what kind of applications to develop. This fact influences the modeling of the city: there are a lot of points of view to describe the city and there is no easy way to identify the best one [12]. Urban Computing should give flexible services that customize their behaviors according with the person (or the kind of person) invoking them.

Urban Computing will change the way in which people feel the city [13]. New classes of services and applications will be available to citizens, to allow the communication among citizens, for example exchanging data with people around them [14] or querying the city like a database (for urban navigation or discovering).

1.4 Problems

Urban Computing carries new kinds of issues to be faced. There are problems of both social and technological nature.

The most relevant social trouble is that stored and publicly distributed data about city could be used in bad ways. In fact there can be people that will use those services for illegal purposes: it will be very important to foresee and consider this kind of issue to develop applications for Urban Computing.

Some examples of technological problems are unavailable services and the difficulties in processing of data. Devices in Urban Computing are pervasive and they entwine every object around us. This kind of troubles in Urban Computing are thornier than ones in Web: what should a service do if it invokes another service and it is unavailable? Or if a device is broken and it gives wrong data, what will happen? A Web site could publish an HTTP 404 message, but a door controller can't warn about the problem and blocks the door until it will be repaired.

Another issue is about the amount of data: the development of technologies to archive data improves everyday and storage devices are becoming more and more capable. The real problem is how to process data: Urban Computing requires high performance techniques to compute large amounts of data (and real-time generated data) with low response time.

2 Community projects and events

In this section we will present some initiatives and projects related to the Urban Computing field.

2.1 City Ware

The goal of City Ware[5] is to develop theory, principles, tools and techniques for the design, implementation and evaluation of city-scale pervasive systems as integral facets of the urban landscape. The City Ware project integrates the disciplines of Architecture, Human-Computer Interaction and Distributed Systems, building on our previous work to develop principles, tools and techniques for designing, implementing and evaluating city-scale pervasive systems as integral facets of urban design. City Ware addresses the challenges of scaling up the design and implementation of pervasive systems to long-term, city-scale systems and evaluating these systems and their relationships with urban space and society through both targeted and longitudinal studies. Outcome includes the followings:

– Making space - This project explores the relationship between the spaces created by urban architecture and the interaction spaces created by artifacts such as digital devices.

[5] `http://www.cityware.org.uk/`

- Radio city - The aim of this project is to survey the "mobile computing landscape", by which it mean (1) the disposition of wireless communications signals in the city, and (2) the human behaviors associated with the presence of those radio signals.
- Movement Flow in Bath City Center - This project has carried out spatial analyses and observation studies of the city of Bath using Space Syntax methods
- Digital Footprints - It seeks to investigate how tourists and visitors use the spaces created by urban architecture. Urban spaces are frequently populated by tourists with a very different agenda to those that live there.
- Multimedia blogging - This project focuses on how the Internet is used in our private lives, in the form of sharing photographs across the Internet.

Project Details. City Ware is a multidisciplinary research project, integrating the disciplines of architecture and urban design, human-computer interaction and distributed systems. City Ware is funded through the UK Engineering and Physical Sciences Research Council's WINES programme, with support from the City Ware industrial partners. City Ware began in October 2005 and runs until March 2009.

2.2 ICiNG - Intelligent Cities of the Next Generation

ICING[6] researches a **multi-modal**, **multi-access** concept of **e-Government**. The model - thin-skinned City - is sensitive to both the citizen and the environment through the use of mobile devices, universal access gateways, social software and environmental sensors. Intelligent infrastructure enables a Public Administration Services layer and a Communities layer.

ICING researches into **e-Community** and Usability and also into two-way interaction with the physical environment. The research focuses on the areas of embedded intelligence, tighter integration of operator platforms and city infrastructure to enable novel services, empowerment of citizens to evolve systems of interaction with the city via social software, input from citizens and sensors for management systems and decision modeling, and a combination of city systems and multi-modal, multi-device communications to provide enhanced services. The technology platform gathers indicators from the City, processes the information, proposes actions to be taken with human intervention and supervision and connects the City with its constituency. Services and information are delivered on a range of commodity devices, providing greater reach and accessibility to local government and communities.

Solutions are tested in - City Laboratories - in strategic city regeneration districts, 22@ in Barcelona, Grangegorman in Dublin and Arabianranta in Helsinki, where users will trial and evaluate technologies and services. Outcomes include the following:

[6] http://cordis.europa.eu/fetch?CALLER=FP6_PROJ&ACTION=D&DOC=2&CAT=PROJ&QUERY=011c6aa58204:3f69:162ba07e&RCN=79299

– Vision Model of a more sensitive, accessible city;
– Technology Models and Open Source Tools for multi-modal access;
– Communications Gateways and Location Based Services that interact with the citizen and the environment;
– Urban Mediator system for citizen-led services;
– Research roadmap for the future.

Project details.

Project Reference:	026665
Contract Type:	Specific Targeted Research Project
Start Date:	2006-01-01
End Date:	2008-06-30
Duration:	30 months
Project Status:	Completed
Project Cost:	4.65 million euro
Project Funding:	2.9 million euro

2.3 Urban Information Systems in MIT

Urban Information Systems[7] is a cross-cutting group in MIT's Department of Urban Studies and Planning. Research focuses on the use of information technologies to understand the relationships underlying **urban spatial structure** and on the use of technology to facilitate broader and deeper participation in the **planning of urban futures**. The participates are interested in applying **computing technology** or techniques in order to understand the ripple effects of computing, communications, and digital spatial information on urban and regional planning processes and on the methods for shaping and nurturing metropolitan areas.

One of the current research interests is SENSEable City Lab[8] - studying and anticipating the **real-time city**. A recent project is Real Time Rome.

Real Time Rome. The project Real Time Rome[9], contribution to the 2006 Venice Biennale, aggregated data from cell phones (obtained using Telecom Italia's innovative Lochness platform), buses and taxis in Rome to better understand **urban dynamics in real time**. By revealing the pulse of the city, the project aims to show how technology can help individuals make more informed decisions about their environment. Some important issues are as following:

– aggregate records collected from communication networks
– the visualizations of Real Time Rome
– neighborhoods used in the course of a day, distribution of buses and taxis correlated with densities of people, goods and services
– individual privacy

[7] `http://web.mit.edu/dusp/uis/www/`
[8] `http://senseable.mit.edu/`
[9] `http://senseable.mit.edu/realtimerome/`

2.4 GeoPKDD - Geographic Privacy-aware Knowledge Discovery and Delivery

A flood of data about **moving objects** is becoming available, due to the automated collection of telecom data from mobile phones and other location-aware devices, from sensor networks, from ubiquitous computing. The goal of the GeoPKDD project[10] is to develop theory, techniques and systems for **geographic knowledge discovery**, based on new privacy-preserving methods for extracting knowledge from large amounts of raw data referenced in space and time.

The project aims at devising **data warehousing** and **data mining** methods for trajectories of moving objects; such methods will be designed to preserve the **privacy** of the source sensitive data.

Obtaining the potential benefits by means of a **trustable technology**, designed to protect individual privacy, is a highly challenging goal; if fulfilled, it would enable a wider social acceptance of many new services of public utility that find in geographic knowledge a key driver - e.g., in **sustainable mobility**, urban planning, environmental monitoring, and risk management.

Project details.

Project Reference:	014915
Contract Type:	Specific Targeted Research Project
Start Date:	2005-12-01
End Date:	2008-11-30
Duration:	36 months
Project Status:	Execution
Project Cost:	3.45 million euro
Project Funding:	2.2 million euro

3 Workshops/Conferences

Urban Computing doesn't have anymore dedicated conferences. Actually his topics are treated in workshops and conferences about Geoinformatics and Ubiquitous Computing.

3.1 LocWeb 2008: First International Workshop on Location and the Web, April 22, Beijing

The workshop[11] focuses on all the geospatial aspects that are related to the Web. The main objective is to look into the fields of how to **extract**, **index**, **mine**, **find**, **exploit**, **mashup**, and **visualize** Web content with respect to its **location semantics**. The workshop aims to bring together researchers in the fields of geographic information retrieval, location-based media search on the Web, Web 2.0

[10] http://www.geopkdd.eu/

[11] http://medien.informatik.uni-oldenburg.de/LocWeb2008/index.html

and user generated content, core Web technologies, and geographical information systems both from academia and industry labs to discuss and present the latest results and trends in all facets of the **relationship between a physical location and Web information**.

Orgnaizers. Susanne Boll, the Department of Computing Science at the University of Oldenburg, Germany, et al.

Topics of interest.

- Spatial Web information retrieval (geographical search, ranking, and annotation; spatial indexing; locative spam)
- Understanding and modeling location (deriving, harvesting and mining location; core location concepts and formal ontologies; location as first-level Web concept; location syntax and semantics)
- Visualizing geographically referenced data (user interfaces; visualizing and interacting with location-driven query results and localized data on mobile devices)
- Location, mobility and the user (mobile localized search; mobile sensor data fusion and location)
- Geospatial media and applications (geo-driven media and content; location in and from Web 2.0 communities; location-aware social software)

3.2 The Ambient Information Systems Workshop (W9) at Ubicomp 2008 (related), September 21, COEX, Seoul, South Korea

The workshop[12] is about research in pervasive and ubiquitous computing: context-aware computing, wireless connectivity, multi-sensor platforms, smart materials, and location-tracking technologies. In addition this workshop focuses on information movement between the periphery and the center of one's attention, and highly transparent technologies .

The workshop will bring together researchers working in the areas of ambient displays, eripheral displays, slow technology, glanceable displays, and calm technology, to discuss and collaborate on developing new design approaches for creating ambient information systems, to identify problems in the design, development, and evaluation of AIS and to derive fundamental challenges of AIS research.

Orgnaizers. William R. Hazlewood, School of Informatics, Indiana University, Bloomington, et al.

[12] `http://ambientinformation.org/`

Topics of interest.

- differences from other information technologies
- examples for the implementation of ambient information
- merit in Ambient Noise, Ambient Smells, Tactile Ambience, and Ambient Taste
- perceivable and comprehensible ambient information
- interaction methods for information devices
- good placement of AIS to improve their chances of being used
- sorts of information conveyed by an ambient display
- methods for evaluating ambient information systems
- the values of these particular technologies in our everyday lives
- use of existing technologies (e.g. smart materials, wearable systems, etc.)
- knowledge from other domains (e.g., from art, cognitive science, design, psychology, sociology)

3.3 Digital Geography in a Web 2.0 World

Visualization[13] is essential to make sense of very large geographic data sets quickly, and where it is necessary to predict many alternative spatial patterns, and where the system is too complex to reduce all analysis to numbers.

Three different ways of understanding geographic patterns in space and time

- traditional methods for urban simulation based on land use transport modeling
- iconic digital models based on 3D GIS and CAD
- demonstrate how patterns such as residential land use at the macro level can occur through the actions of individuals at the micro scale. Examples from the evacuation of a building to traffic jams through the use of Web 2.0 technologies, especially that of Second Life.

GENESIS Project[14]. Generative e-Social Science for Scio-Spatial Simulation:

- Collaboration of the Centre for Advanced Spatial Analysis (CASA) at University College London and the Centre for Spatial Analysis and Policy (CSAP) at the University of Leeds
- agent-based simulation + geo-spatial visualisation
- model agents using census and survey data (large-scale);
- simulate their interactions under specified conditions;
- visualise the emergent patterns;
- deliver via the Internet as an easy-to-use web service.

[13] `http://www.casa.ucl.ac.uk/barbican/` Februar 20, 2008, Barbican, London, UK `http://www.ncess.ac.uk/events/DigitalGeography/` September 15, 2008, Urbis Museum in Manchester, UK

[14] `http://www.genesis.ucl.ac.uk/`

Orgnaizers. University College London in collaboration with National Centre for e-Social Science[15] (NCeSS). About the latter:

- NCeSS investigates how innovative and powerful computer-based infrastructure (commonly known as the Grid) and tools can benefit the social science research community.
- e-Social Science is a term used to describe collaborations between computer scientists and social scientists. The aim is that the computer scientists design and develop what is known as middleware - the software that makes sharing easy for non-experts - in order to address the social scientists' substantive research problems. Sharing resources refers to undertaking enormous calculations or processing huge amounts of data.
- One of the research activities is PolicyGrid that brings together social scientists with interests in rural policy development and appraisal with computer scientists who have experience in Grid and **Semantic Web technologies**.
- growing artificial societies in silico from the bottom up

4 Existing Urban Computing Applications.

Recently Urban Computing started to develop some applications for citizens and it shows that it is potentially useful and how it could allow to develop new kinds of services.

4.1 CitySense

CitySense[16] is a mobile application for local nightlife discovery and social navigation, answering the question, "Where is everybody?". CitySense (Figure 1) shows the overall activity level of the city, top activity hot spots, and places with unexpectedly high activity, all in real-time. Then it links to Yelp and Google to show what venues are operating at those locations. CitySense is a free demonstration of the Macrosense platform. Using a billion points of GPS and WiFi positioning data from the last few years - plus real-time feeds - CitySense sees S.F. from above and puts the top live hot spots on mobile devices.

How does it work? CitySense is an application that operates on the Sense Networks CitySense platform, which analyzes massive amounts of aggregate, anonymous location data in real-time. CitySense is already being used by business people for things like selecting store locations and understanding retail demand.

[15] `http://www.ncess.ac.uk/`
[16] `http://www.citysense.com/home.php`

Fig. 1. CitySense running on a smartphone

Application scenarios.

– Live overall activity & top hot spots - First of all see if it's a good night to go out. The city is 21% busier than normal for right now? Let's go. But where to? Check out the top hotspots in real-time and head out.
– What's at hot spot #1? - Click over to Yelp or Google and find out what's going on at the #1 hot spot: Bars? Clubs? Restaurants? Then check out what's at #2. All with one-handed navigation: no typing needed.
– Show me where the unusually high activity is - Even if you're a local, CitySense can give you the live details you need. When the Mission or Soma is busier than normal - you'll know immediately.

Future Plan. In its next release, CitySense will not only tell each person where everyone is right now, but where everyone like YOU is right now. The application will compare user's history and preferences with those of other users, and show user where he or she are most likely to find people with similar tastes at that moment. So each person's nightlife map will look a little different, and will display a unique top hot spot list.

4.2 Sense Networks

Sense Networks[17], Inc. indexes the real world using real-time and historical location data for predictive analytics across multiple industries. The company's platform, Macrosense, receives streaming location data in real-time, analyzes

[17] `http://www.sensenetworks.com/`

and processes the data in the context of billions of historical data points, and stores it in a way that can be easily queried to better understand aggregate human activity. The company's first consumer application, CitySense, is for local nightlife discovery and social navigation. Sense Networks, headquartered in New York City's SoHo neighborhood, was founded in 2003 and incorporated in early 2006. The founding team is composed of top computer scientists from MIT and Columbia University.

Technology. Analyzing locations in the real world requires an ability to process hundreds of thousands of data dimensions. Due to the temporal sensitivity of the number and type of people frequenting a place (which can be far more dynamic than a static web page on the Internet), the raw data describing places in the real world requires a staggering number of dimensions. Sense Networks attributes 487,500 dimensions to every place in a city, thus identifying a unique and complex "DNA" which describes it completely. The dimensions are based on the movement of people in and out of that place over time, and the places those people visit before and afterwards. Proprietary MVE (Minimum Volume Embedding) algorithms reduce the dimensionality of location and temporal data to 2 dimensions while retaining over 90% of the information. This allows for visualizations of data that allow humans to better understand key dimensions and data relationships. This also allows the Macrosense platform to extract key relationships in the flow of people in a city, such as the flow of those shopping, commuting to and from work, or socializing.

Additionally, Sense Networks applies advanced statistical algorithms to normalize activity based on years of historical data combined with demographic, weather, and other variables. Once a broad understanding of the spatial behaviors in a city is available, companies and investors can leverage the continuously updating framework to better understand their own customers from sparse location data, discover trends in aggregate consumer behavior for correlation with financial indicators, and predict demand for services and places.

The Minimum Volume Embedding Algorithm. Minimum Volume Embedding (referred to herein as MVE) is a state-of-the-art method for summarizing large, high-dimensional data compactly (Shaw and Jebara, 2007). Subsequently, processing complex datasets with MVE makes it easy to perform many interesting analytics operations. For instance, finding clusters, subgroups, correlations, and categories become much easier when data is low-dimensional - as do making predictions and forecasts. Since such operations are difficult to perform with high-dimensional data (the so-called "curse of dimensionality") MVE is a vital step in creating a data analytics platform. In addition, MVE makes data more compact, thereby permitting efficient data storage and data retrieval. Assume we have N high-dimensional items or objects. Each object may be a person, with their own behavioral data or movement trail.

Alternatively, the object may be a place, with its historical traffic density distributed over the work-week. The input to MVE is N large complex objects of dimensionality D. The output of MVE is a set of N low-dimensional summaries of the objects, for instance of dimensionality d. In most applications, the desired dimension d is much smaller than the data's original dimensionality D. For instance, the inputs may have hundreds of thousands of dimensions while the output of MVE is a two- or three-dimensional summary of each object.

4.3 Journey Planner

Any information held by Transport for London and its subsidiary companies is accessible by the public. Especially, Journey Planner[18] on TfL[19] website utilizes maps, timetables, and route information for helping people find optimal route.

Application Scenarios

- mobile alerts and services - After setting up personalized route, SMS alerts and email warnings of Tube and DLR service delays will be delivered. All information including Journey Planner, Travel news, timetables, maps are accessible by website and mobile phone via WAP (mobile Internet) and SMS.
- Planning Journey by text message - One can plan a journey by text message. Just text A to B to 60TFL* (that's 60835), where A and B can be postcodes, full station or full bus-stop names, in any combination

- Tube line service updates - One can check for the latest updates of any delays to your Tube service by texting TUBE [Tube line] to 60835*. For example text TUBE NORTHERN to 60835 to receive a text update on services on the Northern line.

- Find a taxi or Private Hire Operator - TfL has launched a new service to help get home. Just text HOME to 60835* and the telephone numbers of two nearest licensed minicab operators will be sent directly to the person's mobile phone. The person will also be sent the Taxi One-Number details.

[18] http://journeyplanner.tfl.gov.uk/
[19] http://journeyplanner.tfl.gov.uk/user/XSLT_TRIP_REQUEST2

5 Conclusions and Future Work

In this Technical Report we have discussed some initiatives (such as research activities, funded projects and real world applications) related to Urban Computing. New conferences about this field are scheduled for next months and their number is growing more and more, suggesting that this topic is becoming very important.

While we were surveying Urban Computing, we found some projects related with Semantic Web. For example there is a work of Great Britain's national mapping agency, Ordnance Survey, that has published some information about UK cities in RDF with the best practice of linked data[20]. Another example is the Ontology of Transportation Networks [15] developed by researchers of REWERSE project.

In our future works we will continue to investigate the state of the art of Urban Computing and we will study some works developed in the context of Semantic Web concerning Urban Computing.

References

1. Giussani, B.: LiftAsia08: Massively parallel, networked, talking cities. Running notes from LiftAsia08 in Jeju, Korea. Launch over IP. Bruno Giussani's running notes on people, places, technologies and ideas (2008) `http://www.lunchoverip.com/`.
2. Nova, N.: "Networked cities" session at LIFT Asia 2008. Pasta & Vinegar (2007) `http://www.liftlab.com/think/nova`.
3. Greenfield, A.: No Boundaries: The challenge of ubiquitous design. Adobe Design Center (2006)
4. Weiser, M.: The computer for the twenty-first century. Scientific American **265**(3) (1991) 94–104
5. Nova, N.: Adam Greenfield at PicNic 2007. Pasta & Vinegar (2007) `http://www.liftlab.com/think/nova`.
6. Fatah gen. Schieck, A., Lopez de Vallejo, I., Penn, A.: Urban Space and Pervasive Systems. The Seventh International Conference on Ubiquitous Computing (2005)
7. O'Neill, E., Kostakos, V., Kindberg, T., gen. Schieck, A.F., Penn, Danae Stanton, F., Jones, T.: Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape. In Dourish, P., Friday, A., eds.: Ubicomp. Volume 4206 of Lecture Notes in Computer Science., Springer (2006) 315–332
8. Davis, C.A.j.e., Vieira Monteiro, A.M.e.: Advances in geoinformatics. VIII Brazilian symposium on geoinformatics, GEOINFO 2006, Campos do Jordão (SP), Brazil, November 19–22, 2006. Berlin: Springer. xxiv, 315 p. (2007)
9. Markowetz, A., Chen, Y.y., Suel, T.: Design and implementation of a geographic search engine. In: WebDB. (2005)
10. Hill, D.: The Adaptive City. cityofsound. cities. design. architecture. media. music. etc. (2008) `http://www.cityofsound.com/`.
11. Hill, D.: Two or three recent diagrams. cityofsound. cities. design. architecture. media. music. etc. (2008) `http://www.cityofsound.com/`.

[20] `http://os.rkbexplorer.com/`

12. Dourish, P., Anderson, K., Nafus, D.: Cultural Mobilities: Diversity and Agency in Urban Computing. In Baranauskas, M.C.C., Palanque, P.A., Abascal, J., Barbosa, S.D.J., eds.: INTERACT (2). Volume 4663 of Lecture Notes in Computer Science., Springer (2007) 100–113
13. Girardin, F.: Systems are for Cities, but Cities are for People. 7.5 Floor (2008) `http://www.liftlab.com/think/fabien`.
14. Bassoli, A., Brewer, J., Martin, K., Dourish, P., Mainwaring, S.: Underground Aesthetics: Rethinking Urban Computing. IEEE Pervasive Computing **6**(3) (2007) 39–45
15. Lorenz, B.: Ontology of transportation networks. REWERSE - Reasoning on the Web with Rules and Semantics (2005)

Other Data Sources

Maps

Archive	Source	Data	Format	Other links
Yahoo! Maps	Yahoo!			http://developer.yahoo.com/maps/
Google Maps	Google			http://code.google.com/apis/maps/
Google Earth	Google			http://code.google.com/apis/earth/
Multimap	Microsoft			http://www.multimap.com/openapi/
ViaMichelin	Michelin	Maps, routes, hotels etc.	SOAP	http://dev.viamichelin.com/wswebsite/gbr/jsp/vmdn/VMDNHomePage.jsp

Geographical information

Archive	Source	Data	Format	Other links
GeoPlanet™	Yahoo!	It gives the WOEID (Where on Earth ID) of a generic place on the Earth.	REST	http://developer.yahoo.com/geo/guide/
UK location	Webservicex.NET	Postcode Town and County UK Address Validator	SOAP	

Events

Archive	Source	Data	Format	Other links
Upcoming	Yahoo!	An archive of events published by users	REST (XML)	http://upcoming.yahoo.com/services/api/
EVDB	Eventful	Events about sport, politics, music, etc. published by users	REST (XML, JSON, YAML)	http://api.eventful.com/
Last.fm		A social music website with musician profiles, web radio, events information, etc.	REST XML-RPC	http://www.last.fm/api/intro

Multimedia

Archive	Source	Data	Format	Other links
YouTube	Google	Videos		http://code.google.com/apis/youtube/overview.html
Flickr	Yahoo!	Photos	REST (XML, JSON, PHP) XML-RPC SOAP	http://www.flickr.com/services/api
Geograph British Isles	Yahoo!	Photos	REST RSS feed CSV export	http://www.geograph.org.uk/help/api

General informaton about cities and towns

Archive	Source	Data	Format	Other links
Weather	Yahoo!	Weather information	RSS feed	http://developer.yahoo.com/weather/
Traffic	Yahoo!	Traffic information, accidents, obstructions, etc.	REST (XML) RSS feed	http://developer.yahoo.com/traffic/rest/V1/index.html http://developer.yahoo.com/traffic/rss/V1/index.html
HotJobs	Yahoo!	Service to manage public and private job ads	REST (XML)	http://developer.yahoo.com/hotjobs/
Local	Yahoo!	Local business information and user reviews	REST (XML, JSON, PHP)	http://developer.yahoo.com/local/
Airport Information Webservice	Webservicex.NET	Airports information (location, name, code etc.)	SOAP	
Holydays	Holyday service	Holydays of different nations	SOAP	
Craiglist (London)		A Web site with free classified advertisements published by users		

Travels

Archive	Source	Data	Format	Other links
Super Break		A service to book hotels	REST	http://help.webservices.superbreak.com/
Travel	Yahoo!	A service to find airline tickets, hotels, car rental etc.	REST (XML, JSON, PHP)	http://developer.yahoo.com/travel/

Other

Archive	Source	Data	Format	Other links
Answer	Yahoo!		REST	http://developer.yahoo.com/answers/
Address Book	Yahoo!	A service to build user-location based applications	REST (XML, JSON)	http://developer.yahoo.com/addressbook/ http://developer.yahoo.com/addressbook/guide/

Mortgage Index	Webservicex.NET	This Web service Provides following monthly ,weekly and Historical Mortgage Indexes	SOAP	

Developement

Project	Source	Data	Format	Other links
BOSS	Yahoo!	Platform to build Web search engine		http://developer.yahoo.com/search/boss/boss_guide/
KML		A standard format to display maps in mobile device		http://code.google.com/apis/kml/ http://code.google.com/apis/kml/documentation/kml_tut.html
Fire Eagle	Yahoo!	A service to build user-location based applications	REST	http://developer.yahoo.com/fireeagle/ http://fireeagle.yahoo.net/developer
User Interface Library	Yahoo!	A library to build rich user Web application		

UrbanComputing/OtherDataSources (last edited 2008-09-08 14:51:47 by ?DanieleDellAglio)

Show Us A Better Way

List of available web sources:

Archive	Source	Data	Category of data				Format	Note
			Geo	Soc	Pol	Ent		
Notices	London Gazette	# State # Parliament # Ecclesiastical #Public Finance # Transport # Planning # Health # Environment # Water # Agriculture & Fisheries # Energy # Post & Telecom # Competition # Corporate Insolvency # Companies & Financial Regulations # Partnerships # Societies Regulation	v	v	v	v	XML	
Public Spending Information	HM Treasury	# Public expenditure overview # Analyses of budgets # Differences from previous plans # Trends in public sector expenditure # Functional and economic category analyses of public sector expenditure # Central government own expenditure # Local authority expenditure # Public corporations # Analysis of public expenditure by country and region # Country expenditure by sub-function		v	v		Excel CSV	
List of schools	Department of Children, Schools and Families	List of all the schools in England and Wales (from Edubase)	v	v			Excel SQL	
Neighbourhood Statistics (NeSS)	Office of National Statistics	# 2001Census # Access to Services # Community wellbeing/Social Environment # Crime and Safety # Economic Deprivation # Education # Skills and Training # Health and care # Housing, Indicators # Indices of Deprivation # People and Society # Physical Environment # Work Deprivation		v			SOAP	Registration required
Transport Information	Transport Direct	# Location of bus stops, stations etc and descriptions of the locations # The National Car Park Register # Journey planning information # A carbon calculator # The National Public Transport Data Repository feed	v				API FTP	Requires a registration by email to get the access to the data
Health care services and information	NHS	Information about health care service providers, and “live well” health information from the NHS.	v	v			REST (XML)	Registration required
Address Data	Royal Mail	Sample raw data from the Postcode Address File. This will give you a comprehensive overview of the type, structure and format of the address data held by the Royal Mail, as well as the different data sets Royal Mail maintain.	v					The data is available on CD and it should be required by e-mail

Carbon Footprint Information	Defra (via AMEE - a 3rd party aggregator)	An application to calculate energy consumption and CO2 emission for classes of objects (for example the set of average sized European cars driven for 5000 km)					REST (XML, JSON, HTML)	Registration required
UK Road Travel Data	BBC	Traffic information, accidents, obrstuctions etc.	v				XML (TPEG-RTM)	The XML file has validation problems
UK Public Transport Data	BBC	Public transport information, delays, warns (closed stations, replacement bus services, etc.)	v				XML (TPEG-PTI)	The XML file has validation problems
3 Day Weather Forecast	BBC	Weather information of the most important cities	v				XML (OPML)	
National Public Transport Access Node db (NaPTAN)		Information about stations, coach terminus, airports, ferry terminals, bus stops, etc.	v				XML CVS	Registration required
National Public Transport Gazetteer (NPTG)		An archive with links among aliases of towns and places	v				XML CVS	Registration required
They Work For You		Information about Members of Parliament (MPs), their votes and their speeches			v		REST (XML, JS, PHP, RABX)	Registration required
Parliament Parser		Structured versions of publicly available data from the UK parliament, and the source code that was used to generate the data			v		XML	
Downing Street Says		Answers of the Prime Minister's Spokesman in response to the questions made by the Lobby (a group of selected politic journalists)			v		RSS feed	
Tell Them What You Think		A list of current goverment consultations			v		REST (XML, PHP)	
Fix My Street		Local problems inserted by citizens	v	v			GeoRSS feed	
The Government Says		Search engine to find news published by Government			v		RSS feed	
Planning Alerts		Data about planning applications from local authority websites	v				REST (GeoRSS)	
BBC Backstage	BBC	BBC News feeds, BBC Search API, TV & Radio schedules				v	REST (xml, json, ics)	
BBC Programmes data	BBC	Information about radio and TV programmes				v	REST (xml, json, yaml)	
Open Space	Ordnance Survey	Maps, name of places etc.	v				API	Registration required Data published in Linked Data

UrbanComputing/ShowUsABetterWay (last edited 2008-09-08 09:09:40 by ?IreneCelino)

Getting to Milano

Motivation

This use case shows the added value of collecting a broad set of information around traffic, of integrating it and of using it to support a citizen that has to go to Milano from another city in Lombardy region.

Preconditions

- All relevant sources of information are known when the citizen uses the system, but data can be obtained at execution time.
- Citizens are recorded in the Milano Urban Computing System (MUCS), they authorized MUCS to notify them information by sending SMS to their mobile phone.
- Some integrated business model among the various actors is in place.

Story board

1. Carlo leaves in Varese 60 km North-West of Milano
2. Carlo has to go to the Lombardy Region premises in via Taramelli, 20 where he arranged a meeting at 11.00 for 25.9.2008
3. The day before the meeting he enters MUCS and opens the "plan a travel" service
4. Carlo fills in the required data
   - FROM: via Luigi Sacco, 1 Varese
   - WHEN: after 8.00 - 25.9.2008
   - TO: via Taramelli, 20 Milano
   - WHEN: before 11.00 - 25.9.2008
   - USING: any means of transportation
5. MUCS provides Carlo with three alternatives, two using public transportations (i.e., A and B) and one by using his private car (i.e., C)
   A. Using railroad Ferrovie Nord Milano *and Metro M3, leaving home at 8.30 and arriving between 10.15 and 10.30.*
      - *Walk from Via Sacco, 1 Varese to Varese Casbeno train station, it takes 7 minutes*
      - Railroad ?LeNord: 8.39 Varese Casbeno - 10.03 Milano Piazza Repubblica
      - Metro M3: Repubblica - Zara, every 5 minutes and takes 3 minutes (Repubblica is the name of the metro station colocated with the train station di Milano Piazza Repubblica)
      - Walk from Zara to Via Taramelli, 20, it takes 5 minutes

   B. Using railroad Ferrovie dello Stato (alternative to the previous one) and Metro M3; leaving home at 8.20 and arriving between 10.05 and 10.20.
      - Walk from Via Sacco, 1 Varese to Varese Stato train station, it takes 12 minutes
      - Railroad Ferrovie dello Stato: 8:43 Varese Stato - 9.55 Milano Piazza Repubblica
      - Metro M3: Repubblica - Zara, every 5 minutes and takes 3 minutes (Repubblica is the name of the metro station colocated with the train station di Milano Piazza Repubblica)
      - Walk from Zara to Via Taramelli, 20, it takes 5 minutes

   C. Using Private Car; leaving home around 9.30 and arriving between 10.10 and 10.40.
      - In this case MUCS performs as Google Maps does today and give the resulting

driving directions, but instead of saying that such a travel requires 49 minutes, MUCS explains Carlo that he should leave home after 9.30 (when the usual commuters traffic on A8 is almost over)

6. Carlo is tempted by option C, he could sleep a little longer, but while travelling by train he could complete the presentation for the meeting so he chooses the option A since he has to walk for a shorter distance, and he uses the ticket-less option to buy the train ticket.
7. MUCS also asks Carlo if he wants to be alerted via SMS if the option he chooses is no longer the best one (e.g., due to problems to the railroads)
8. Carlo agrees and exits MUCS
9. The day after (25.9.2008) at 7.14 MUCS learns from the railroad information system system of railroad "?LeNord" that a technical problem is causing an average delay of 45 minutes to all "leNord" trains from Varese to Milano.
10. MUCS estimates that an accident of such kind will not be solved before 11.00, therefore it checks if any planned travel is at risk. It finds Carlo's travel.
11. MUCS checks if the other options it proposed to Carlo are still valid. Apparently they are so MUCS sends an SMS to Carlo informing him that an accident is causing 45 minutes delay for all trains on railroad "?LeNord" and he can either use the railroad "Ferrovie dello Stato" (option B) or take his private car, in this case Carlo can convert his train ticket into a daily parking ticket for one of the parking lots of the sub-urban metro stations in Milano.
12. Carlo receives the SMS, he enters the MUCS and checks the two alternatives. He can take option B, but he knows that when problem of this kind happens on ?LeNord, all commuters take the other railroads and he will never be able to find a sit. On the other hand, MUCS (taking into consideration weather data through the route and real time traffic congestion status on top of historical traffic congestion statistics) predicts that being a rainy day the traffic on A8 will be slower and he could arrive in time only leaving around 9.00.
13. Carlo decides to take his car, in this way he has all the time to complete the presentation before leaving. He leaves home around 9.00 and instructs its GPS Navigator to interact with MUCS traffic service and to find the cheapest gas station along the road.
14. During the driving Carlo receives the instructions for the gas station and refuels the car.
15. At a certain moment his GPS Navigator receives alert from MUCS: Milano North-West area is hit by heavy showers and the traffic is getting slower. Instead of going to the planned North-West parking lot, the GPS suggests to go to one in the South-West; the metro from there will only take 10 minutes more than from the planned parking lot, but the estimated time from the planned parking lot is 25 minutes more than the originally planned time.
16. Carlo considers the option and decides to follow it.
17. Carlo parks the car and taking the metro arrives in time to his appointment.

Further extensions to this story board

- MUCS may handle several other information sources
  - Personal information for the meeting including other participants email, cellular, and location
  - Historic and tourist attaction points
  - Emergency policy for infrastructure including flooding, poison gas leakage, avalanche, etc.
- Possible scenarios with extended information sources
  - MUCS asks Carlo if he wants to let other participants know the same information he received. (like weather condition or traffic accident, or may be late for the meeting)
  - MUCS automatically connects to other relevant systems for quick resolutions. (like fire station or hospital)
  - MUCS let Carlo know other participants' location if they are very close in distance
  - MUCS informs Carlo the upcoming historic/tourist place while driving if Carlo's preference is met (when he has time for!)

LarkcProject/WP6/WorkInProgress/StoryBoard (last edited 2008-09-05 15:40:04 by ?YiHuang)

Home Finder

Motivation

Home Finder is a service to search for real estates (apartment, house, etc.) for sale or to be rent. The service takes the offers from a mashup done on data of estate agencies and it processes proposals to rank them on the base of user profiles (obtained with a form) and information coming from different sources (school location, pollution statistics, public safety, etc.).

Story Board

1. Carlo has to move to London to live with his family
2. He opens Home Finder Web site (HF)
3. He fills the search form with information about him and his family:
    - Number of people: 3
    - He has a a daughter (5 years-old)
    - He will work in North-West London (he inserts address)
    - He wants to spend less than x£
    - etc.
4. He submits the form to HF
5. HF processes the data:
    - It considers user information and other data, i.e. the paths from houses to Carlo's workplace, the school locations and public transportations:
        - He has a child, therefore he could be interested in a house near a school
        - He has a job, therefore he could be interested in a home near the job
        - HF checks if previous conditions can be both satisfied
        - There aren't schools near the workplace
        - HF tries to find a house with public transportation that link it to workplace/school
        - etc.
    - It ranks the offers from the most interesting one for Carlo to the worst one
6. HF returns the results on a map service (Google Maps), with different color indicators based on the rank; HF could explain (completely or partially) why it considers an offer better than another one
7. Carlo watches the offers and choose the one he prefers
8. HF redirects Carlo to the Web site of estate agency that published that offer

...

Kinds of data sources

- Real estate information (Real estate agencies)
- Maps (Google Maps, Yahoo! Maps, etc.)
- Public Transportation (NaPTAN, NPTG, BBC public transport information, etc.)
- Statistics (NeSS, etc.)
- Cities information: planning applications, hospitals, entertainments, services

Alternative option: finding a vacation house

A service like HF, but with real estates to be booked for a vacation instead of places where to live permanently. In this variation there are some kinds of data already seen to consider:

- Real estate information (Real estate agencies)
- Maps (Google Maps, Yahoo! Maps, etc.)
- Public Transportation (NaPTAN, NPTG, BBC public transport information, etc.)

And other kinds of source:

- Price (it could change in different seasons)
- Scheduling of availability
- Review of users that spent time in that house
- Events and entertainment (Upcoming, Eventful, etc.)
- Proximity to touristic places (e.g., monuments, museums, etc.)

LarkcProject/WP6/WorkInProgress/Storyboard-HF (last edited 2008-09-10 15:08:14 by ?DanieleDellAglio)

UK Travel

Motivation

UK Travel helps users that want to organize a trip in the UK. Firstly, it can be used when someone wants to plan the journey, with a Web site to look for air tickets, a place to stay (with a service similar to the one described in Home Finder). Then, while travelers are in the city, UKT offers features for mobile devices (smartphones, PDAs), i.e. routes, monuments and arts work descriptions, getting data from his archive and from sources like Wikipedia.

Story Board

Before the travel (1) - planning the trip

1. Carlo and his friends want to spend a vacation in London.
2. He opens UKT and fills the form with useful information to look for an available apartment
    - They are 5 people
    - They want to spend two vacation weeks in August
    - etc.
3. Carlo submits the form to UKT
4. UKT looks for available apartment and shows them to Carlo
5. Carlo books an apartment
6. UKT redirects him to the estate agency that published the offer

Before the travel (2) - planning a city tour

1. Carlo logs on UKT to plan a route to visit the city
2. UKT shows a list of default routes
3. Carlo chooses a route from the list
4. UKT shows some lists with museum, monuments, events that will take place in London in those days, etc.
5. Carlo customizes the route:
    - He adds a monument that he wants to see
    - He remove two museums
    - He adds some events
    - etc.
6. UKT tries to create a new route to satisfy Carlo requests
    - UKT realizes that Carlo and his friends can't arrive in time for the beginning of an event because it's a long way to reach the place
    - UKT warns Carlo about it
7. Carlo wants to watch that event, even if they get late
8. UKT replans the route

During the travel

1. Carlo and his friends are in London, and with Carlo's smartphone they connect to UKT
2. Carlo login to UKT mobile service
3. UKT sends information about the route Carlo planned, and it shows it in a map
4. Carlo starts to follow the route to visit the city

Kinds of data sources

- Real estate information (Real estate agencies)
- Maps (Google Maps, Yahoo! Maps, etc.)
- Public Transportation (NaPTAN, NPTG, BBC public transport information, etc.)
- Weather (BBC, Yahoo! etc.)
- Multimedia (?YouTube, Flickr, etc.)
- Travels (Yahoo! Travel etc.)
- Information about cities/mouments/museums... (Wikipedia, etc.)
- Restaurants (Yahoo!Local, Beer Mapping, Cafè Spot etc.)
- Events (Yahoo! Upcoming, EVDB, etc.)
- Review sites (Revyu, ciao, etc.)
- Points of interest

LarkcProject/WP6/WorkInProgress/Storyboard-UKT (last edited 2008-09-10 14:22:33 by ?IreneCelino)

Personal Assistant

Motivation

Personal Assistant is a service to give tips to users. It learns user preferences with an analysis of user profiles and personal calendar; then it starts to give him different kinds of news such as traffic update, information about events that user could be interested in, etc.

Story Board

1. Carlo usually goes to the cinema on Friday evening
2. PA notices that on next Friday evening Carlo hasn't planned anything yet
3. PA looks for upcoming events that will be on Friday evening
4. PA suggests what movies there are on cinemas near Carlo's home with related informative resources (trailers, pictures, reviews)
5. PA checks weather prediction and it founds that it will likely rain
6. In addition to the list of movies, PA adds the movies that will be played on TV
7. Carlo watches the list with options proposed by PA
8. Carlo chooses to stay at home and watch a movie on TV

Kinds of data sources

- Maps (Google Maps, Yahoo! Maps, etc.)
- Public Transportation (NaPTAN, NPTG, BBC public transport information, etc.)
- Weather (BBC, Yahoo! etc.)
- Multimedia (?YouTube, Flickr, etc.)
- Travels (Yahoo! Travel etc.)
- Restaurants (Yahoo!Local, Beer Mapping, Cafè Spot etc.)
- Events (Yahoo! Upcoming, EVDB, etc.)
- Review sites (Revyu, ciao, etc.)
- Public Holidays
- Information about user (i.e. calendar, profiles, etc.)

LarkcProject/WP6/WorkInProgress/Storyboard-PA (last edited 2008-09-10 14:25:36 by ?IreneCelino)