Bringing the Research Lab into Everyday Life: Exploiting Sensitive Environments to Acquire Data for Social Research

Geert de Haan, Sunil Choenni, Ingrid Mulder, Sandra Kalidien and Peter van Waart

INTRODUCTION

This chapter discusses the concept of sensitive environments. Sensitive environments are essentially spatial and public areas like streets or school that are provided with an intelligent infrastructure that collects sensory information of users while they move and interact. Such sensitive environments offer researchers an environment which enables them to automatically collect lots of data. However, when data collection takes place through a sensitive environment instead of by interviews, questionnaires and structured observations, it consequently changes the nature of social research and imposes new types of questions: How to design the environment to fit the research questions. How do we deal with huge datasets in a sensible way? How can we interpret data that might be incomplete or uncertain in a meaningful and useful way? In the remainder of this work, we discus the use of sensitive environments as a tool for social research embedded within everyday life.

To describe where the concept of a sensitive environment comes from and why it is necessary, we first focus on the wider context of human centered design. Research and development in the field of human centered design evolve in different directions and levels, which are not necessarily divergent. There is a practical need to develop and implement systems according to the concept of human centered design, since many information and communication technology (ICT) systems appear to fail in real-life (Dobson, 2007). Dobson describes the classical and much sited case of the London Ambulance Service in which the ICT system turned out to be unusable in real-life. A core element of the ICT system, the ambulance dispatch system, was designed to ensure that an ambulance would be available at the scene of an accident within a certain amount of time. On paper the ICT system worked well and met all stated requirements. In real-life, however, it turned out that among others, certain human requirements had been overlooked in the transition of command from the human operator to the ICT dispatching system, such as, allowing ambulance personnel time to recover after some gruesome accidents or anticipating changes in the demand for ambulance services.

It has been pointed out that an important factor for this failure is that these systems do not meet the expectations and values of the users. In finding answers, several backgrounds and disciplines have joined the research field; however, understanding real-life behavior appears to be complex.

A promising direction to understand human behavior and values is to exploit sensitive environments. In a sensitive environment data is sensed, which makes it possible to derive information such as "who is where, and possibly, doing what", more or less like an anthropologist or a sociologist might do in an observation study of a particular group to establish a social network map of the inhabitants of a street, the power structure in a governing board, etc.

A sensitive infrastructure needs not be sophisticated; all that is necessary is that it has some sensing devices like a surveillance camera, an automatic light switch or an entrance gate to the metro which registers the number of passers-by. Sensors may be placed in a environment accidentally or for different purposes then research. For example, the automatic light switch is only put there to have the lights off when no one is there.

When the sensors are intentionally placed in a environment for research or development purposes, the term "Living Lab" may be used (eJOV, 2008) to indicate that the research

laboratory has moved into the everyday life. The automatic light switch with some added wiring may also be used to get a rough estimate of how many people are in a particular room at a certain time. Likewise, to create a social network map of the inhabitants of a street, a surveillance camera might be used with some software to analyze movement patterns between households to get a first idea about interactions between households. In a sensitive environment, computing and sensing capabilities are embedded in the environment by means of devices and creative tools. These tools and devices are focused on the (continuous) gathering of data about people; examples include RFID (Radio Frequency Identification) applications, Bluetooth, or mobile phones. For example, RFID and Bluetooth may be used to refine the example of creating a social network map since these wireless techniques can be used to provide people and devices with unique identify codes. With the proper placement of sensors the interaction patterns in a street may be refined from household interactions to interactions between individuals. As such, insights in these data can be used to obtain a better understanding of user behavior. It should be clear, however, that the amount of sensor data may be huge and difficult to handle and, besides, there is not always a clear understanding of how to interpret sensor data, consider, for example, how to deal with the interaction pattern of the postman. Furthermore, we expect that in the future even more data will be able to be gathered since emerging technologies are becoming increasingly available and affordable for people in more contexts for more purposes. Technologies for wireless networks, image capture, storage, and display get cheaper and more functional. According to Juniper Research (Juniper Research, 2009) sales of smart phones, mobile phones with multimedia and internet abilities, may help to cushion the downturn in the mobile handset market due to the recession. Juniper forecasts that in 2013 about 23% of all handsets sold will be smart phones, creating a 300 million annual market. Also GSA, the Global mobile Suppliers Association with members like Nokia and Sony Ericsson, shows figures in their market update statistics on GSM and mobile broadband subscriptions, indicating that GSM growth is slowing down in western countries but wireless internet is rapidly growing almost everywhere (GSA, 2009). Another dynamic characteristic of these emerging technologies is that they become personalized, they stay and go with one person, and they are consequently used in various contexts. In other words, sensing capabilities in these devices people carry wit them can be part of the sensing infrastructure as well.

Different types of devices and creative research tools may give rise to different types of data ranging from structured to unstructured data and from numerical to categorical data. The analysis of the gathered data can be done from different assumptions. In the so-called closed world assumption (Genesereth and Nilsson, 1987), it is assumed that collected data is true and complete, while in the case where this assumption does not hold, it is assumed that the collected data is incomplete and uncertain. This definition of the closed world assumption sounds fairly theoretical, however, in the context of ICT it generally comes down to the question whether or not the data used by computer applications is pre-defined. Within a database of a payroll application, for example, data fields named *address* or *balance* have well-defined meanings. This is because their specific meanings have been defined within the "closed world" of the database and the payroll application by software engineers when the database and application were created. In a different database, such as an office telephone register, an *address* field may have a very different meaning, interpretation and format. Most importantly, in general, there is a notable difference between the terms used within the closed worlds of ICT systems and the open world of everyday life. Consider, for example, that in plain English there are many different ways to use terms like *balance* or *address* and it is not always clear what is exactly meant when they are used.
Related to the closed world assumption is the structuredness of data. Within a closed world, the meaning of terms or data fields is well structured: not only the terms are defined in

advance but also the relations between them are. Within an art database, there is a structured relation between a *painter* and a *painting* but this relation does not have to be true for painters in everyday life. Consider, for example, that in text documents the co-occurrence of *painter* and *painting* may be completely accidental, when "the painter removed the painting to paint the wall" and, nevertheless, a search engine like Google will report a hit.

To analyze a huge collection of structured data under the closed world assumption, both data mining and statistical tools can be used. A major difference between data mining and statistics is that data mining tools help to generate useful hypotheses, while statistics is focused on the rejection or acceptance of a pre-defined hypothesis (Choenni et al., 2005A). An example of data mining is comparing data in different databases related to welfare or unemployment benefits in order to find frauds, or to analyze airline passenger data to determine who might be a terrorist. To analyze semi-structured and/or unstructured data, we propose to use text mining and information retrieval tools. The best known example of an information retrieval tool is probably Google which looks for the occurrence of a particular search text in a collection of documents. More advanced information retrieval tools are also able to distinguish between the different meanings of terms, like the bank of the balance and the one in the park. Text mining tools go one step further and analyze the relations among text terms, as such text mining tools are able to consider the context in which terms are used and sometimes to "understand" what a document is about. Multi-sensor data fusion techniques will be tailored for personalized applications (Waltz and Llinas, 1990). In such applications, measurement by different sensors will be combined in order to serve the information needs of a single user. A good example of multi-sensor data fusion is the digital camera which "knows" where its pictures are taken, and the GPS positioning information is fused with the digital image. We note that contemporary analyzing tools are equipped with visualization modules for different types of users (Laudon and Laudon, 1999).

Analyzing data in the case that the closed world assumption does not hold is in its childhood. The processing of data in this case leads to scenario studies or how to exploit these data in such a way that it adds value to the data analysis processed under the closed world assumption.

In this chapter, we discuss the potentials and concerns that are entailed by sensitive environments. We expect that the exploitation of these environments may be beneficial for, amongst others, social scientists. Suppose a group of people is provided with wireless identity tags and camcorders in a sensitive environment within an office, a library or an urban area. When the sensitive environment is setup, is possible to track who is where, with whom and preferably doing what, the setting would enable researchers to collect data to study, for example, interaction patterns among people in more and less formal settings, without the need to interfere with the ongoing activities. Data collected in these environments may be complementary to data gathered by questionnaires or interviews. Observation data may be used to validate or support answers to questions or the questions may be used to give meaning to or resolve ambiguities in the observations. These data may give arise to a wide variety of innovative applications. However, if the collection and processing of these data are not submitted to rules and regulations it may violate the privacy law (Kalidien et al., 2009). We give an overview of analyzing methods and tools that may be applied to different data types. We also distinguish the role of the time dimension in analyzing data. Furthermore, we discuss how data that are uncertain and for which the closed world assumption does not hold can be collected and exploited for a better understanding of human behavior.

SENSITIVE ENVIRONMENTS

While most methods and tools for data collection are explicit, i.e., users are involved in an explicit way, emergent technologies also provide opportunities for getting insight into human behavior in an implicit and less obtrusive manner. By embedding a sensor network in an environment, it is possible to track users' patterns of interaction with and usage of appliances and movements from one location to the other. When such a sensitive environment is used for research purposes, and the research laboratory is moved into the real-life world, the sensitive environment is referred to as a "Living Lab" (eJOV, 2008).

The first applications benefiting emergent technology were mainly motivated by the desire to monitor elderly and disabled people i.e., observing people that are hard to reach for 'data collection', for example, using movement sensors to determine if these people are well. The next generation in emergent technology includes the adoption of biosensors, which measure phenomena such as skin temperature or heart beat frequency; in this way it is possible to infer stress and excitement levels, for instance while the user is playing with an interactive game. However, most examples of environments collecting context data concentrate on logging of usages or track changes in location. One of the first examples of a sensitive environment as well as an example of a Living Lab is the Active Badge System, developed around 1991 at Olivetti Research Lab (Want et al., 1992). In this system, researchers wore badges with a small transmitter whose location could be monitored by a number of receivers dispersed over the laboratory building. This system allowed people to determine the whereabouts of their colleagues and whether or not they could be approached. Even though the system was used on a voluntary basis and could be switched off, it did evoke a lot of discussion about privacy issues. A more recent and much more advanced example of a 'logging' environment is the intelligent coffee corner (Mulder, Lenzini, Bargh, & Hulsebosch, 2009), where people taking coffee can use a variety of services offered in the intelligent environment at the coffee corner's site. In the intelligent coffee corner, real-life data is collected by employees carrying detectable devices (e.g. Bluetooth-enabled mobile phones or PDA's and WLAN-enabled laptops) with them and a RFID-enabled badge, which is needed to open doors in order to access the different floors in the office building.

It might be clear that sensitive environments open a wealth of possibilities for real-life data as they enable researchers to come close to people. It moves research out of laboratories into real-life contexts and provides opportunities to non-intrusively study social phenomena in users' social and dynamic context of daily life.

Mulder et al. (2009) stress the need for methodological guidelines and tools that effectively combine the intelligent features of such environments with the strengths of methods and tools traditionally used in social science research, like interviews and focus groups. In the example to create a social map of the interaction patterns among inhabitants of a residential area, it is possible to use data from surveillance camera's or RFID tags to identify and measure the frequency of the interaction patterns. This would be fine when research questions can be answered using purely numeric data. However, it is more likely that researchers will also want qualitative data to investigate why particular interactions take place, and it takes an interview or questionnaire to find out if people are relatives of each other or perhaps postmen. Notwithstanding, it is necessary to have reliable data collection systems as well as (automatic) solutions for capturing and analyzing user behavior, taking into account people's sensitivity to privacy.

When the aim of trying to create a sensitive environment is not how to build a failsafe system but rather to utilize as much information from real-life as possible to serve its inhabitants, it may not be a good idea to create yet another closed system using certain and well-defined data but rather to strive for systems that provide "surplus value" for the users. To create a reliable electronic banking system requires the design of systems that provide absolute certainty about the identity of the bearer of a bankcard and the amount of money on

his or her balance. Furthermore, any cash withdrawal system should only allow for interactions guarded by valid and reliable means of identification, such as a PIN code to guarantee the relation between identity and balance. Although having to use a PIN code for identification is a system requirement following from the choice, or perhaps, the need to use certain ICT systems, it therefore is not something that users might need or even appreciate. In the context of banking systems, users might appreciate sound advice or transactions without hassle as surplus values of a banking system, for example, in comparison to other systems.

In order to design a sensitive environment in which any kind of information available is used to serve any kind of 'user' or 'inhabitant,' it is rather unnecessary to strive for a closed information system. In order to design a sensitive environment, a system should utilize as many sources of information as possible, much like the way in which we, human beings, function best in environments that provide information from all our different senses. In general, human beings don't function very well in environments that feature information that is restricted to visual or visual-spatial information only, as most ICT systems force us to do.

Information systems aiming to support sensitive environments should include any kind of information that is available, both in terms of whether it is available at all, as well as in terms of whether information is available in a structured, well-formed, reliable, deterministic, and explicit form. In the context of designing information systems, it is important to realize that human beings, in contrast to information systems, do not require structured information; it may be helpful but is not a prerequisite for functional behavior. Human beings do not fall silent when information is unknown; rather, they tend to predict or fill-in what is unknown or unreliable in the environment. In addition, most human knowledge remains tacit and thus implicit (Polanyi, 1958).

In short, a sensitive environment can be seen as an intelligent infrastructure that collects sensory information of users while they move and interact. Although such an environment eases data collection processes as lots of data can be captured automatically, benefits for data analysis are not often that obvious. How do we deal with huge datasets in a sensible way? How can we interpret data that might be incomplete or uncertain in a meaningful and useful way? In the remainder of this work, we focus on making sense of real-life datasets.

DATA ANALYZING TOOLS

Equipping our environment with different types of measurement tools, such as sensors, interactive white-boards, cameras, etc., results in the collection of a vast amount of data. Depending on the nature of a measurement tool, different kinds of data will be collected. At a high abstraction level, we distinguish three types of data: structured, semi-structured, and unstructured (Bocij, Greasley and Hickie, 2009).

Structured data can be regarded as numbers or facts that can be conveniently stored and retrieved in an orderly manner. This is due to the fact that the semantics of the data are well defined. Database and data warehouse systems are developed to facilitate the storage of structured data. The database of a bank-account system, for example, describes the amounts of currency in bank-accounts, identified by account numbers, connected to (legal) persons, who reside at postal addresses, which in turn consist of a street name, an address number, a city name, a postal code, and possibly a state or country code. Such an account system allows one to select the postal addresses of all the people whose name is 'Smith' or 'Jones' and who own more than a certain amount of money. The bank-account system allows for operations like selection by name or address because the elements of the database—account numbers, names, and address elements—relate to each other by means of pre-defined relations. Therefore, these systems contain, in general, structured data. The opposite of structured data is unstructured data.

Unstructured data refers data that lacks pre-defined relations. Unstructured data will most often refer to textual documents; data in these documents are ambiguous and therefore not well defined. Information retrieval systems such as Google are developed to facilitate, store, and retrieve unstructured data. Since the data used by Google has no pre-defined relations between bank-account numbers, owner names and addresses, Googling for 'Smith' or 'Jones' will result in a long list of banks, bank-account owners, addresses, museums, horses, and whatever, each with 'Smith' or 'Jones' in its name. The fundamental differences between these systems are summarized in Table 1.

Aspect	Structured data	Unstructured data
Matching	Exact	Partial & best
Model	Deterministic	Probabilistic
Query language	Formal	Natural
Answers to questions	Exact	Relevant
Output sensitivity to errors	No	Yes

Table 1: Difference in handling structured versus unstructured data

As Table 1 shows, for the handling of structured data, a question needs to be formulated in a formal query language, which in turn is used to search for data that exactly match to the question. Therefore, a data retrieval system is capable to return exact answers without errors to the user. Let us assume that a patent office has automated the applications of patents, and information with regard to patents can be obtained via a website. A question like "give me the patents that have been submitted by Dr. Knuth" will be typically handled by systems that facilitate the retrieval of structured data. Whereas a question like "give me the patents that significantly contributed to the development of information retrieval systems?" will be typically handled by systems that facilitate the retrieval of unstructured data, for the following reasons: the phrase "significantly contributed" is subjective to a certain extent and therefore not well-defined; and to answer the question, a large set of patents, which are textual documents, have to be examined.

The third type of data we distinguish is semi-structured data, which is in between structured and unstructured data and has some regular structure. For example, in a book we recognize, besides unstructured data, some regular structure in the make-up. A book is divided into chapters, a chapter into sections, and a section into paragraphs. Today XML (eXtended Markup Language) is becoming the standard to model semi-structured data (Elmasri and Navathe, 1994). If a book is understood to be a long string of characters, like the file of a word-processor, then XML may be used to indicate which parts of the string contain book elements like the names of the authors, the title, the chapters, the textual data, etc. by means of MARKUP tags like <TITLE> ... <\TITLE>, <AUTHOR> ... <\AUTHOR>, or <TEXT> ... <\TEXT>. XML is used for a particular document type, to describe which markup tags there are and how they relate. In HTML, the language to markup documents for the World-Wide Web, XML is used to describe the tags that determine how documents are presented in web-browsers. Note that HTML is concerned with the presentation rather then the structure of documents. Although XML enables the structured application of data, it will only in very simple and trivial cases allow for a complete structural definition of the available data. Generally, the greater part of the data, like the text in sections of a book, will remain unstructured.

Although the need for analyzing tools for semi-structured and unstructured data is widely recognized, the majority of data analyzing tools pertain to structured data. In the following, we discuss the concepts behind analyzing tools that pertain to structured and

unstructured data. The concepts behind semi-structured analyzing tools can be considered a mix of the concepts applied to analyze structured and unstructured data.

ANALYZING STRUCTURED DATA

A wide variety of analyzing tools is reported in the literature (Fayyad et al., 1996). The suitability of each tool depends on the analysis that should be performed on the data that is collected. Consider for example a shop that attaches an RFID sensor to its shopping baskets that is equipped with several sensor readings. Suppose that customers are requested to take a basket before entering the shop. In such a setting, we may collect RFID sensor data that pertain to the position of a customer at a certain time *t*, which can be modeled as a triple (*basket_id, position, time*) (Farina and Studer, 1985). Furthermore, we may collect data with regard to items in the basket, which may be modeled as a sequence of [$item_1$, $item_2$, …$item_n$]. Note that time plays a crucial role in the data collected by the sensors, while this is not the case with regard to the data in the basket. The former collection of data will be referred to as time dependent data, while the latter will be referred to as time independent.

As said before, two approaches can be used to analyze time independent data: a statistical or a data mining approach (Choenni et al., 2005A). The starting point of the statistical approach is to collect data and analyze data that might be relevant in rejecting or accepting a hypothesis. The starting point of data mining is collecting data that might not be related to a specific problem at hand and searching for interesting hypotheses. Note that a hypothesis can be regarded as a model of the real world. Although both approaches have a different starting point, the goal is to come up with a useful and adequate model of the real world for a problem given at hand. We illustrate the differences and similarities between data mining and statistics by means of an example. Let us assume that we have collected and stored several items in the baskets of each customer. A typical data mining question on the collected data is: “Search for an interesting hypothesis for me that might be relevant to my shop fitting.” Such a question might result in an association between diapers and beer such as, the selling of diapers leads to the selling of beer. On the basis of this result, a shop owner may decide to store beer and diapers next to each other for the sake of convenience for customers. A typical statistical question on the collected data might be, “Is there an association, e.g. correlation, between beer and diapers?” In both questions, the interest in an association between beer and diapers is expressed. However, in the data mining question the interest is expressed in an implicit manner, while in the statistical question it is done in an explicit way.

This difference in the formulation of questions has major consequences for the development of technology used for data mining and statistics. To answer the data mining question, all associations need to be inspected in order to find the interesting ones, while in the case of statistics the association between diapers and beer needs to be computed. We note that an exponential complexity is involved in answering the data mining question, i.e., if we have n variables then 2^n possible associations need to be inspected. Therefore, an important issue in data mining is how to control complexity. Furthermore, we observe that traditional statistical techniques are tailored towards the handling of data mining questions. For example, classification and clustering concepts are extended to determine the profiles of entities, such as customer, reserearchers, etc.

Main goal for tools that focus on time dependent data is to track or predict the location of a moving object, such as an airplane or a car, at a certain time or within a time frame. For tools and concepts in the field of multi-sensor data fusion the goal is to determine the location of an object on the basis of data that is obtained from different sources, e.g. sensors. In our shopping basket example, we may be interested in what departments of a shop, e.g. the food or fashion department, a customer is interested in. Suppose that the shop is

equipped with a sensor network such that whenever a customer enters or leaves a certain department of the shop, this data is read by the sensor readings and stored in a database. To compute the time that a customer has spent in a department, data of the sensors that noted that a customer has entered and has left the department can be processed. Moreover, data from the sensor reading may also be used to determine the popular routes in the shop.

Today, we observe also an emerging trend to exploit comprehensive sensor networks to serve for answering "navigational" queries such as "what is the closest best restaurant?" Mobile location-based services employ information from wireless networks such as GPS and WIFI to establish one's location relative to some beacons such as route-maps as in car-navigation systems, places of interest like touristic sites, buildings or works of art as in e-guide systems. Navigational queries are not limited to establishing locations relative to stationary beacons; the beacons may also be other people carrying GPS or WIFI devices, such as in mobile mixed-reality games (Benford et al., 2006) or the GRINDR iPhone application to 'meet guys near you' (Grindr, 2009).

ANALYZING UNSTRUCTURED DATA

The explosive growth of the web has entailed a boost in the development systems that are capable to handle unstructured and semi-structured data. Today, the web has become a huge knowledge resource, containing information about many subjects. The challenge the user faces is finding the information he/she needs and generating useful knowledge from a vast amount of semi- or unstructured data. Information retrieval systems facilitate users in finding documents that might be relevant for them (Croft, 1993; Baeza-Yates and Ribeiro-Neto, 1999), while text mining systems search for useful knowledge, such as associations and regularities within the data (Tan, 1999; Berry, 2004). We briefly discuss the issues and potentials behind these systems.

In the field of information retrieval, effort is put into building systems that are capable of handling the information needs of a user. Information needs formulated by a user are not necessarily exact, as they are in traditional (database) applications, but rather vague and incomplete (Choenni et al., 2005A). Often an information need is expressed by a set of keywords. Suppose that we have a system containing a digital library and a user needs to gain some information about information systems. Therefore, he/she consults this system by typing the keywords "information systems" in order to find all relevant documents that deal with this subject. It should be clear that there are many documents about this subject, and therefore it is not trivial to select the proper documents for this user. Furthermore, a document dealing with information systems can be of interest for one user but not for another, even though both express their information need by the same keywords.

Another issue that should be taken care of by information retrieval systems is to recognize that keywords not explicitly mentioned by users may still be of interest to them. For example, since a database is a major component of an information system, documents about databases also may be of interest for a user who used "information systems" as a keyword and did not explicitly mention databases. By means of an interactive session with the user, an information retrieval system attempts to discover what precisely the information need of a user is and to meet this need. The data generated in the interactive session may be stored and analyzed to better understand users. This understanding can be exploited to serve users better whenever they use the system again (Choenni et al., 2005B). Text mining may be applied as a tool to analyze the data such that it contributes to a better understanding of users.

In the field of text mining, we may distinguish roughly two directions. The first takes as its starting point a database that contains structured data; the goal is to extend this database with unstructured data from other sources. The general approach is to transform the

unstructured data into structured data, and then to apply the data mining process. Suppose that we distribute a questionnaire to the customers in our shop example. The goal of the questionnaire is to find out to what extent customers are satisfied by the services in the shop. Customers are asked to answer a set of open and closed questions. In general, the answers of each customer to the open questions can be added to an existing database, as long as the database makes a distinction between different customers. To add the unstructured data for mining purposes in the database requires significantly more effort. The answers to each open question should be classified into a limited number of subjects with which the database is extended. Techniques for information retrieval and extraction are used to facilitate this step. We note that information extraction has as its goal to extract facts from a vast amount of texts. For each subject, the answers of the customers should be clustered into a limited number of clusters. Then, each customer's answer should be mapped on one of these clusters, which will be stored in the databases. Suppose that a questionnaire is sent out to customers and one of the questions pertains to complaints about the items in a shop. Let us assume that the items in the shop are supplied by a limited number of suppliers S1 to S7 and we want to know from the answers on the questionnaire to which suppliers the complaints pertain mostly. Then, for each supplier, we have to make a list of items that are delivered by this supplier. From the questionnaires, we have to extract the items and have to map each item to a supplier.

A generalized approach is to model documents in databases (Blok et al., 2004). To keep them manageable, not all the terms of a document can be stored in databases. Therefore, a selection of words should be made that will be recorded in a database. In literature, it is proposed to make this selection on the basis of the so-called document and term frequencies. Term frequency is defined as the occurrence of a term in a document related to the term with the highest occurrence in the document. The document frequency of a term is defined as the number of documents in which this term occurs related to the total number of documents. The higher these frequencies for a term, the more important this term is and should be therefore recorded in the databases. Such a database can then be mined with data mining tools.

To conclude, one direction of text mining focuses on extending databases with the unstructured data by processing the data in such a way as to create neatly categorized and structured data.

The second direction of text mining is focused on the analysis of documents. The goal is to determine to what extent two sets of documents are associated with each other or to what extent these sets differ from each other. For example, association may be measured in the amount of overlapping terms between documents. Results from this approach might be useful in the examination of patents, books that will be purchased for a library, etc.

PROCESSING UNCERTAIN AND INCOMPLETE DATA

Earlier, a number of distinctions were made between different types of data, e.g. between structured and unstructured data and between knowledge or information that may be derived from data mining operations versus knowledge or information that required information processing or text mining. In reality as opposed to data modeling, no such strict distinctions do exist, and for this reason we propose a combined approach both to think about data modeling in sensitive environments as well as to make optimal use of existing data sources.

In traditional ICT, data is generally explicitly defined and used for the purpose of the systems being designed. For example, in order to design a public transport billing system, one might use a personal balance card as a means to keep track of the distance traveled and the amount charged. As such, the notions of the personal identity of the traveler and the balance, be it in miles, clicks, or euros, are to be represented by means of the balance card. Such a

card is, of course, not a natural thing but rather a burden to its keeper. Assuming, furthermore, that the card is dedicated to the local or national public transport system or even to the transport company, it follows that the public transport billing system and the balance card constitute a closed system. In general, closed information systems only have a few well-defined possible relations to the real world in which they operate

It may be noted that in order to function properly, information systems, in contrast to human beings, require exactly the type of information that is well-structured, well-formed, reliable, deterministic, and explicit. When the aim is to design ICT systems to support human beings, rather then to design ICT systems for administration or business purposes, in these circumstances, the question then is how to utilize unstructured, ill-formed, unreliable, etc. information in systems that require the opposite.

Earlier in this chapter, we argued that it is sometimes possible to turn implicit and uncertain data into explicit information, as in the case of the relation between diaper and beer sales. In that example, statistics or data mining established the relation. In advance, the data about beer sales and diaper sales were structured explicitly. The relation between the two categories was not explicitly known but could be established by using the drag-net of data mining: calculating statistical correlations between (sales) categories and filtering out those that surpass a certain level of significance. After this, further analysis may reveal the nature of the relation, for example, that there is a certain category of buyer that links diaper and beer sales.

The interesting point here is that uncertain data should not be avoided but rather used to make deterministic data more interesting or useful. Consider for example that when buying books, people might be interested in other books that are in some respect similar to the ones they already know, as is exemplified by Amazon's "people who bought the book also bought..." To answer a question like this, it is not necessary to know exactly, reliably, and in well-defined ways what people's interests are. The only thing that needs be known is the purchasing outcome of others. Of course, if customer's interests were known, in addition to the data about other people's purchasing behavior, the recommendation system would be even better able to advise a customer.

SOME ILLUSTRATING EXAMPLES

As an example, consider the recommendation system behind the Amazon online bookstore. What is required for a good recommendation system is some sort of basic data set consisting of other people's opinions and choices that is large enough to analyze the data and organize it into interesting or otherwise significant patterns in behavior. These patterns, in turn, may be interpreted by a process of sense making to yield directly applicable results. In (Davenport and Glaser, 2002), an advanced knowledge management system is described in a medical context. On the basis of knowledge in the medical field and the experiences of medical colleagues, the system advised what drugs might be prescribed for a certain disease.

Outside the well-defined ICT environment we might essentially do the same, except that the data or information may not be derived from the well-defined deterministic environment of the information system but from the outside world. In this case, the information is not the result of an analysis process that is fed into a sense making process, but rather the opposite: a process of sense making is applied to the outside world such that phenomena that may not be observed directly may be predicted from observable phenomena such as data available about past behavior. An example of such a 'sensible relation' is that - other things being equal - a person's interest in an object, for example a painting in a museum, a dress in a shop window, or a stereo in a car, may be established by measuring the time spent inspecting the object by art lovers, shopping addicts, or car burglars, respectively.

Naturally, a 'sensible' relation between things like interests and time spent may not be very reliable; nevertheless, it points in the right direction and might reduce the amount of data to be taken into account. In addition, predicting a phenomenon like 'interest' may not be very reliable using a single predictor, but reliability may significantly be increased using multiple predictors. A recommendation system in an online bookshop is presumably better able to make good recommendations if it looks at the purchases of other customers of a particular item and, additionally looks at things like customers' purchases in real bookshops. Note that, in the area of shopping, only positive recommendations matter, in other areas however one also has to consider costs and benefits of negative recommendations. Consider, for example, security surveillance. A person in a parking lot spending some time observing a particular car is not necessarily a car thief; after all, either the person may own the car or he/she may genuinely be interested in the make or the design of the car. Only when a person is showing interest in a number of different cars, while looking suspiciously around and trying several door-handles, there may be reason to raise suspicion. With respect to designing systems that support people in their everyday lives, a main question is not how to establish and use interesting relations between pre-given data elements or how to utilize sensible relations between phenomena in the outside world in a reliable manner. Instead, the main question should be how to create surplus value from a combination of pre-given data with less certain data relations in the real-life world.

A major disadvantage of the type of data that is laid down in predefined ICT systems is its limited utility: this type of data is defined, collected, and put into databases with specified purposes. Even if such data is brought together from multiple sources, it is generally very difficult to use such data for any other purpose then the one underlying the raison d'etre of the ICT systems.

Looking at the interesting types of data in the real-life world, a major disadvantage of data outside ICT systems is that it is, unfortunately, less reliable. It is one thing that common sense may yield a number of interesting or sensible relations between items in the outside world, but it is quite another thing to ask for the validity and reliability of these relations.

Given the limited utility of the one type of data and the limited reliability of the other, it may be interesting to ask what may be gained from bringing the two types of data together. In this case, the idea is to start with the most reliable data and add observed data from the outside world to create additional information.

As an example, consider a museum that provides electronic touring guides to its visitors. The example is loosely based on two European Community IST (Information Society Technology) projects: COMRIS and i-MASS. Both of these projects concern research on 'ubiquitous computing,' a term coined by Mark Weiser (Weiser, 1991), to expand the utilization of computers outside the working context and support people in a natural way in everyday life. The i-Mass project concerned, among other things, the technical information infrastructure to adapt the presentation and content of information in museums to the characteristics of its visitors (de Haan, 2002). The COMRIS project concerned the design of a ubiquitous device to present information from a range of heterogeneous sources to support attending a large conference or exhibition (de Haan, 1999).

When the visitor arrives at an interesting piece of art, he or she types in some number connected to the specific item and the touring guide provides the visitor with information about the item, such as biographic information about the artist, the materials used, and so on. A slightly more advanced electronic touring guide might provide visitors with the opportunity to select a level of explanation that is most suitable to his or her level of expertise. In both these cases, it is the visitor who has to interact with the e-guide device. This may not always be possible or desirable. In addition, one might ask why it is the visitor who has to make decisions.

As an alternative, it may be possible to build some intelligence into the device or into the environment to sense the location of the visitor and to help decide which level of expertise is appropriate for the particular visitor. Suppose that the identity of a museum visitor is known in advance from the ID data on his or her museum card. The ownership of the museum card may be taken to support the idea that this person is more knowledgeable about art than the average museum visitor, which may then be used to instruct the electronic guide to present information at a more advanced level. In addition, the card might be used to store a visitor's personal settings and preferences.

In a similar vein, when an e-guide might wirelessly sense that it is near a certain location, a beacon or a piece of art, the vicinity information may be utilized in different ways. In such a case, the visitor is relieved from the task of instructing the e-guide about his or her whereabouts or the required level of complexity in the guidance information. In addition, when a visitor spends considerable time in the vicinity of a particular item, the e-guide might provide additional or more comprehensive information about the item than the standard message. In this case, the e-guide system might infer from the visitor’s hanging-around behavior that additional explanation is appropriate. In both these examples, information is used from a single source: the amount of time that the e-guide receiver senses that it is in the vicinity of the transmitter associated with a certain location or piece of art.

In a slightly more complex design, it may be possible to use data from different sources to create additional information. Visitors who spend a considerable amount of time near impressionist paintings and who spend, in a consistent manner, relatively little time near naive or abstract paintings may reveal a particular interest in impressionism as an indication that some sort of expert explanation might be appropriate for this visitor. It may be noted that a relation between time near something and amount of interest may not always exist. Even if it seems that people are interested in impressionist art, it may be that in reality people are merely interested in the availability of seats in the room where the paintings are on display. However, at least in the context of visiting museums, there is presumably little harm done in presenting the "wrong" information, especially compared to the benefits of those visitors who are served better.

The utilization of information from multiple sources may become even more advanced when information is used from outside the particular context of use. As an extension of the museum example, consider that through the visitor's museum card or some other publicly accessible identity token, it may be possible to establish a link to information about the visitor on the internet, such as his or her homepage, information from social-networking sites such as Orkut or Facebook, or professional information from sources such as LinkedIn or company websites.

Internet information from sources like these is often and negatively associated with marketing, fraud, and security purposes, often notably different from the personal intentions and wishes of those whom it concerns. Public information, however, may also be used to empower people by using the information in such a way as to adapt the environment according to their own purposes. In the context of the running example, suppose that, upon entering a museum, visitors allow the e-guide system to utilize the information that is available about them on the internet. This time, however, the purpose is not to perform some security check or try and sell something to the visitor but rather to enhance the person's visiting experience by providing guiding information that is optimally adapted to his or her personal interests and experience.

The e-guide system might utilize the internet for information about the visitor in various ways, using more or less advanced techniques, ranging from simple keyword matching, to personal profiling and agent-based metadata analysis using semantic web techniques. When a visitor has a homepage, for example, finding a keyword expression such

as 'impressionist painting' under a heading 'interests' may be taken as a direct reference to a specific interest or a specific level of expertise. More often, keyword analysis will yield indirect references to interests and expertise when sets of words like 'Pisarro,' 'beer,' or 'knitting' may either strengthen, weaken, or be neutral in their relation to certain interests or levels of expertise.

Some problems are true for individual keyword-interest relations which may not be very strong or valid or may not even be reliable. Keyword matching, profiles, and semantic analysis, however, are hardly never concerned with single keyword matches. On the contrary, since concern is with large numbers of such relations and value networks, an analysis will yield useful relations, if only by sheer number. This is especially true when concern is not with a single data source but when there are many homepages, profiles, or function descriptions at stake. Indeed, whereas ICT systems under the closed world assumption tend to be restricted to single database systems featuring a certain level of reliability and validity, the idea to create useful information from a multitude of different data and different data sources inherently supports the notion to rely on sound data patterns.

To summarize this section, first, it is necessary to open up the closed systems in order to make the trapped data inside available for different purposes. In principle, the data inside closed systems is only meaningful to the purposes for which those systems were built in the first place, and this may be done in different ways. Having the data readily available, they can be related to different sets of less structured and less reliable data, such as unstructured text on the internet or semi-structured XML data. Finally, by relating the different types of data using a sufficiently large number of sources, where each source provides a different perspective on environment, it will not only be possible to increase the reliability and certainty of the information, but it will also be possible to create more utility and meaningfulness than was available in the resources.

BENEFITS AND POTENTIALS IN DATA COLLECTION

In the previous sections, we briefly mentioned some of the benefits of sensitive environments. In this section, we discuss the impact of data collected in sensitive environments in social research and some innovative applications that might be entailed by the data. To gain insight in social phenomena, a standard tool that is used by researchers is surveys. Valuable data can be collected form surveys and interesting hypotheses can be answered from these data (as illustrated before in the shop fitting example).

However, using surveys for collecting data in order to answer explicit as well as implicit questions also has restrictions leading to some challenges. One restriction is that the reliability of surveys is dependent on the number of participants in a survey. Often the participation in a survey is voluntary. One challenge is that there are not enough participants in a sample to make reliable statistical statements. Secondly, composing a questionnaire can be quite time-consuming. Specific items on the research subject must be chosen carefully at hand and often a pre-test is needed to see how effective a certain item is. The processing and analyzing of the data is often also quite time-consuming (athough nowadays many surveys are made by the internet). Nevertheless, analyzing questionaires may cost a lot of time, especially in the case of ‘open’ questions, e.g. “What do you think of this product?” instead of ‘closed’ questions, answered with e.g. yes, no, don’t know. In the case of open questions, there is a wide range of interpretation, while in the case of closed questions the interpretation is fixed. Another challenge may be that different interviewers interview in a different way; this means that the interrealibility amongst interviewers may not be optimal. By training interviewers, one may increase the interreliability amongst interviewers; however this also costs time and money.

Yet another challenge of surveys is that there are different ways of sampling. For instance, participants may be recruited by taking a sample of the central population register in one survey. In another survey, the sampling may be done by means of postal codes or telephone registers. Some interviews may be held at the participants home, while others are made by telephone or email. Comparing the results of different surveys about the same object may not be evident in this way. Also, comparing the results of surveys over different periods may not be evident when different sampling methods have been used.

In addition, the restricted memory of a respondent may influence the results, especially in the case of questions about the past, e.g. "Did you drink more beer compared to 5 years ago?" From psychological research it is known that human memory is not a simple recorder of events but an active process, influenced by irrelevant external factors like emotions, suggestions, intentions, how quetions are formulated, etc. As a result it is very difficult to have people act as eyewitnesses, particularly about their own lives (Loftus, 1996).

By exploiting (implicit) data by means of technology, one may profit in a more efficient way from the data. For example, in the case of sampling, a sensor in the shop basket automatically registers all clients visiting a shop. In this way, no interviewers are needed and also a broad variety of possible hypotheses may be captured. In the case of a survey, specific items must be thought of at hand. Another example is that in some large shops costumers can scan all products themselves and pay with their credit card instead of waiting in the line at the cash register. In this way, implicit data about consumer behaviour is registered without the need for composing questionnaires and sampling. Using the same technique in different regions may lead to the comparison of different regions. Also, by exploiting technology the challenge of restricted memory is overcome, since data is registered each time a customer enters the shop. As illustrated above, in many cases one may benefit from replacing survey data with register data, in the case that the enviroment is adjusted adequately with existing measurement technology.

The data collected in sensitive environments may also open directions for new applications in the near future. This might be complete new directions that are based on tracking data or existing applications that are enriched with a tracking dimension. An example of the latter has already been given in the context of navigational queries. Database systems are now able to handle additional queries such as "give me the closest best restarurant from the position where a user is." An example of new directions that are completely based on tracking data is to manage crowds in cities and traffic jams. Big events always attract many people, which often lead to congestions and traffick jams. By keeping track of the mobile phones of people, one may follow the developments of the traffick and movement of people. Authorities may use this knowledge to control the movement of people and/or traffic whenever congestions arise.

Data from sensitive environments may also have an impact in the field of direct marketing, also referred to as pervasive adverstising (e.g. van Waart & Mulder, in press). Since the movement of people can be tracked at individual level, it becomes easy to find out their travelling behaviour. Therefore, more effective offers by marketing departments may be made to them, especially when this data can be combined or integrated with other data, such as the reading interest, preferences for food and so on. Market researchers still use a small palette of traditional techniques to get insights in what their customers want. These methods, such as interviews, focus groups or surveys, usually focus on what people say they do. Differently put, insights gained will be those that can be expressed by people themselves, and the data consequently do not reveal customer insights. Creative research tools promise to be different and yield therefore different insights as well. In the current chapter, we elaborated on how to deal with uncertain data in a meaningful way.

It should be clear that the collection of data in sensitive environments has potential in several domains. However, the counterpart of this potential is the possible violation of people's privacy. How to deal with privacy issues in innovative applications that may expose the identity of individuals is considered as a major challenge (see a.o Broder, 2000; Choenni and van Dijk, 2009).

CONCLUSIONS

This chapter started with the observation that an important factor in the failure of information systems in real life is that these systems do not meet the expectations and values of the users. The concepts of sensitive environments and Living Labs, in which information systems continuously collect and process data about their users, may provide a better understanding of the needs and limitations of them. Consequently, this understanding may be incorporated in the design of contemporary information systems, which increases the acceptance of information systems by its users, and therefore reduces the chances of failure of an information system. Since the understanding of the behavior, needs and limitations of people is also a cornerstone in social research, the collection and processing of data in the context of sensitive environments may become an effective and efficient vehicle to answer questions raised by social researchers.

A widely accepted method to study a real-world phenomenon is to capture the phenomenon in a model, which is the subject of further analysis and reasoning. Data collected in real-life are used to derive models of a phenomenon or to underpin a model that is built on the basis of theories. However, for a single phenomenon, one may devise many models. We recall that in general a model for a phenomenon can be regarded as a simplified map of the real-world phenomenon. Depending on the research questions, one model may be marked as better than another model. To study the behavior, needs, and limitations of people, interactions with and amongst people have been proved to be indispensable to collect the proper set of data for building suitable models. Therefore, we propose to create sensitive environments that also focus on the support of interactions.

In order to create sensitive environments that support people's interactions with each other, the general goal of Human Centered ICT approaches, it will be necessary to make the data in closed ICT systems available for other purposes than those for which the specific systems were built in the first place. As illustrated in the foregoing, the data from ICT systems for surveillance purposes might also be used for, for instance, keeping in touch with one's neighbors. Mere access to data does not ensure that it can be used meaningfully. Meaning might be added by means of some analytical process. In this chapter, we propose to seek external data to facilitate changing streams of low-level data into meaningful information. Using external information, from the internet, for example, in combination with well-structured data also helps to create information that is sufficiently reliable for further processing in ICT systems, in order to shape a meaningful context containing meaningful information. An example of these new opportunities is illustrated, in the foregoing, by means of an electronic guiding system for museums that uses a combination of well-defined and less well-defined data sources to allow for automatic adaptation of the subject and of the level of explanation to the interests, wishes, and experience of museum visitors, thus enhancing the experience of visiting a museum. The electronic guidance systems have become more and more sophisticated in the past recent years due to the success of PDA's, GPRS enabled phones and Smart Phones, in combination with technologies like Java, RFID and Wifi (Santoro et al., 2007). Most e-guides in actual use still require the user to indicate the item of interest and other settings, but location-aware e-guides are commercially available. E-guides featuring automatic adaptation to the type of user or usage have -to our knowledge- not yet left the research phase of development (Baus et al., 2005).

Only when we succeed in crossing the border between structured and unstructured data will we be able to create true ubiquitous computing system, supporting everyday life using ICT.

We have argued that to design such systems requires the utilization of a range of different types of data, such as structured, semi-structured and unstructured data, each requesting specific methods for analysis, including data mining, statistics, multi-sensor data-fusion, and techniques as advanced as reasoning with uncertainty and incomplete information. Applying some general principles behind human behavior in the natural world, it follows that the most interesting and useful types of information are those which are presently not used or underutilized in ICT systems that tend to depend on well-defined, reliable and valid information, thus creating closed-world systems: systems with few and well-defined interfaces to the outside world, such that they are only usable for the specific purpose that they are designed for. Although data mining technologies has as goal to exploit collected data also for other purposes than data is initially collected for, these technologies are primarily focused on structured data. In order to provide sensitive environments to support people in everyday life, including personal wishes and intentions, it is necessary to look beyond the technologies of the closed world assumption and utilize less well-structured data and information. The sheer abundance of less well-structured and less reliable data sources will not only create new opportunities to design more useful systems answering the needs and demands of everyday life, but also increase the reliability and utility of the data that is already available and used in information systems. In general, extracting models from these wide variety types of data will lead to better and more reliable models that may be used to understand people behavior and social phenomena in society. In many cases, social researchers will be relieved from issues that are involved to data collection, such as the design and processing of questionnaires and model building. Therefore, it is possible that social researchers will shift their efforts towards the interpretation of models in the future.

Research and development related to sensitive environments evolves in several directions, which are not necessarily divergent. Two main streams may be distinguished, the so-called technical and application oriented stream. The technical oriented stream, mainly consisting of computer scientists, mathematicians, and electrical engineers, primarily focuses on proper ways to represent different type of data in computer systems and to reason/combine these data in an efficient and robust way. The application oriented stream primarily focuses on the set up of sensitive environments and devising and implementing innovative applications with the rich set of different type of data that is nowadays available. We note that new technology is also deployed for efficiency purposes e.g. in case of product development rather then research purposes. During this process they often have to face technical challenges and shortcomings. These challenges and shortcomings are communicated to the technical stream, which may use them as a source of inspiration for their research agenda. It should be clear that in such a setting both streams are complementary to each other.

References

Baeza-Yates, R., and B. Ribeiro-Neto. 1999. *Modern Information Retrieval*. New York: Addison Wesley/ACM Press.

Baus, J., K. Cheverst and C. Kray. 2005. "A Survey of Map-based Mobile Guides". In *Map-based Mobile Services: Theories, Methods and Implementations*, ed. L. Meng, A. Zipf and T. Reichenbacher, 197-216. Springer Verlag.

Benford, S., A. Crabtree, M. Flintham, A. Drozd, R. Anastasi, M. Paxton, N. Tandavanitj, M. Adams and J. Row Farr. 2006. Can You See Me Now? *ACM Transactions on Computer-Human Interaction* 13 (1): 100-133.

Berry, M.W., ed. 2004. *Survey of Text Mining, Clustering, Classification and Retrieval*. New York: Springer.

Blok, H.E., R. Choenni, H. Blanken, and P. Apers. (2004). A Selectivity Model for Fragmented Relations: Applied in Information Retrieval. *IEEE Trans. Knowl. Data Eng.* 16(5): 635-639.

Bocij, P., A. Greasley and S. Hickie, ed. 2009. *Business Information Systems: Technology, Development and Management*, 4th edition. Pearson Education Limited.

Broder, A.J., ed. 2000. "Data Mining, the Internet, and Privacy". In *Proceedings of International WEBKDD'99 Workshop*, August 15, 1999, in San Diego, CA., USA. Springer LNCS 1836, 56-73. Berlin/Heidelberg: Springer Verlag.

Choenni, R., and J. van Dijk. 2009. "Towards privacy preserving data reconciliation for criminal justice chains". In *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, DG.O 2009, 223-229. ACM.

Choenni, R., R. Bakker, H.E. Blok and R. de Laat. 2005A. "Supporting Technologies for Knowledge Management". In *Knowledge Management and Management Learning: Extending the Horizon of Knowledge-Based Management*, ed. W. Baets. New York: Springer Verlag.

Choenni, R., S. Harkema and R. Bakker. 2005B. "Learning and Interaction via ICT Tools for the Benefit of Knowledge Management". In *Knowledge Management and Management Learning Extending the Horizons of Knowledge-Based Management,* ed. W. Baets. New York: Springer Verlag.

Croft, W.B. 1993. Knowledge-Based and Statistical Approaches to Text Retrieval. *IEEE Expert* 8(2): 8-12.

Davenport, T.H. and L. Glaser. 2002. Just in Time Delivery Comes to Knowledge Management. *Harvard Business Review*, July, 2002.

Dobson, J. 2007. "Understanding Failure: The London Ambulance Service Disaster". In *Responsibility and Dependable Systems*, ed. G. Dewsbury and J. Dobson, 130-161. London: Springer Verlag.

eJOV (2008). *eJOV – The Electronic Journal for Virtual Organizations and Networks 10, Special Issue on Living Labs*, (November), http://www.ejov.org/apps/pub.asp?Q=2993&T=eJOV%20Issues&B=1 (accessed July, 2009).

Elmasri, R., and S. Navathe. 1994. *Fundamentals of Database Systems*. 2nd Edition. Benjamin/Cummings.

Farina, A., and F.A. Studer. 1985. *Radar Data Processing, Vol.1 Introduction and Tracking, Research Studies*, Press Wiley Press.

Fayyad, U., G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, ed. 1996. *Advances in Knowledge Discovery and Data Mining*. AAAI/The MIT Press.

Genesereth, M. R., and N. Nilsson. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann Publishers.

Grindr 2009. Grindr - Meet_Guys_Near_You_on_your_iPhone. http://www.grindrguy.com/Grindr_iPhone_App/ (accessed June 28, 2009).

GSA 2009. GSM/3G Stats & Market Update. http://www.gsacom.com/ (accessed July 17, 2009).

de Haan, G. 1999. "The Usability of Interacting with the Virtual and the Real in COMRIS". In *Proceedings of Interactions in Virtual Worlds, TWLT 15*, ed. A. Nijholt, O. Donk and D. Van Dijk. May 19-21, in Enschede, the Netherlands.

de Haan, G. 2002. "The Design and Evaluation of Intelligent Access to Mankind's Collective Memory". In *Proceedings of the 11th. European Conference on Cognitive*

Ergonomics, ECCE-11 - Cognition, Culture and Design, ed. S. Bagnara, S. Pozzi, A. Rizzo and P. Wrigth, 47-53. September 8-11 2002, in Catania, Italy.

Juniper Research. 2009. Next Generation Smart phones: Players, Opportunities & Forecasts 2008-2013. http://www.juniperresearch.com/shop/viewreport.php?id=171 (accessed July 20, 2009).

Kalidien, S., R. Choenni and R. Meijer. 2009. Towards a Tool for Monitoring Crime and Law Enforcement, In *Proceedings ECIME 2009, 3rd European Conference on Information Management and Evaluation*, in Gothenburg, Sweden. Academic Publishing Limited.

Laudon, K.C., and J.P. Laudon. 1999. *Essentials of Management Information Systems*, 3rd edition. New Jersey: Prentice Hall.

Loftus, Elizabeth F. 1996. *Eyewitness Testimony. Cambridge*, 2nd edition. MA: Harvard University Press.

Mulder, I., G. Lenzini, M.S. Bargh, and B. Hulsebosch. 2009. Reading the tea leaves in an Intelligent Coffee Corner: Challenges for understanding behavior. *Behavior Research Methods* 41(3): 820-826.

Polanyi, M. 1958. *Personal Knowledge. Towards a Post Critical Philosophy*. London: Routledge.

Santoro C., F. Paternò, G. Ricci and B. Leporini. 2007. "A multimodal mobile museum guide for all". In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services. Mobile HCI'07*, 21-25. September 9-12 2007, in Singapore.

Tan, A.,-H. 1999. "Text Mining: The state of the art and the challenges". In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*. April 26 1999, in Beijing, China.

van Waart , P., and Mulder, I. (in press). Meaningful Advertising: Pervasive advertising in the experience economy. Forthcoming in *proceedings of Pervasive Advertising 2009*.

Waltz, E., and J. Llinas. 1990. *Multi Sensor Data Fusion*. Boston: Artech House Radar Library.

Want, R., A. Hopper, V. Falcão and J. Gibbons. 1992. The Active Badge Location System. *ACM Transactions on Information Systems*, 10(1): 91-102.

Weiser, M.D. 1991. The Computer for the Twenty-First Century. *Scientific American*. September 1991: 94-104.

this is a draft version of 21 july 2009