Requirements for Context-Dependent Mobile Access to Information Services*

Augusto Celentano†
Dipartimento di Informatica
Università Ca' Foscari di Venezia
auce@dsi.unive.it

Fabio A. Schreiber, Letizia Tanca
Dipartimento di Elettronica e Informazione
Politecnico di Milano
{schreibe,tanca}@elet.polimi.it

Abstract

In this position paper we discuss information systems access in a mobile context-dependent user environment, characterized by multiple information sources, in particular w.r.t. the relationships between context and multiple sources.

1 Introduction

The communication environment surrounding our daily experience is more and more characterized by mobile devices which can exchange multimedia information and provide access to multimedia services of complex nature. Access to complex information is characterized by a progressive shifting from form-based traditional interfaces to a unique, homogeneous information system, towards systems based on multiple sources of information and mobile devices, where navigation, context-dependence, adaptability, multichannel delivery and ubiquity are key concepts. Such a shifting affects both human-computer interaction and information management, i.e., information specification, classification, presentation and delivery.

Several issues have been studied, models and methodologies proposed, and tools and systems implemented. To mention but the most relevant issues, mobility and multichannel delivery have received great attention also due to the performance increase of devices such as cellular phones, PDAs and notebooks, which considerably augments the potentiality of use.

In recent years a consistent research effort has been devoted to the issues of adaptiveness and personalization in user interaction [22], hypermedia systems [7] and web applications [17]. The results of this extensive work allow designers and users to extend information access and efficiency, but require a deep revision of the paradigms of information management and use, in terms of issues such as the *context*, the *sensorial channels* involved in dialogue and the role of the *environment*. Terms like context-awareness, ubiquity and pervasiveness denote key concepts for the design and development of widely usable information services.

The term "context" has been extensively studied in the area of human-computer interaction; it is mainly associated to the concept of "location", but is far richer than that; some works have underlined different categories of context, such as computational, user, physical and temporal context [10, 12, 13, 19, 20]. In the area of databases and information systems, however, the role of the context has been largely unexplored until recent times.

Information integration in a multiple heterogeneous sources environment has been studied and published in the literature since the end of the 70's. Integration is conceived as the problem of combining data stored in different systems and providing the user with a unified view [3, 15, 18]. Issues which have been considered are both semantic and architectural: finding similar concepts in the different sources, solving structural as well as semantic data conflicts, integrated vs. federated and multidatabase architectures, and integration methodologies at DB design time vs. on-line gathering and filtering of information coming from the queried sources [2, 4, 9, 16].

While all these research topics have been individually studied in an extensive way, their interaction within mobile information systems raises new challenges, which constitute the focus of this position paper.

Context-dependence of information and mobility of the end users can be viewed in two different perspectives: the first one concerns their use in the *design of information sources* (databases) suited to fit into mobile and technology constrained devices; the second one is related to the advantages they offer in *accessing information* in a way which is focused on the particular user interests and actual environment. In the sequel we will deal with the second aspect while referring to [5, 6] for the design issues.

*This work is partially supported by the Italian National MIUR project FIRB-MAIS.

†Work done while on sabbatical leave at Politecnico di Milano.

2 An application scenario

In order to analyze the requirements of context-dependent mobile information systems, we consider an example related to a tourism application. Here a user who travels in a geographical area accesses, at different times, information services describing that area. The example is characterized by the following issues:

- Along his/her way the user comes across different information service providers, which deliver information about tourism related topics and issues. We focus on information services of practical interest such as finding a hotel, a restaurant, a museum, getting information about opening times, prices, events, and so on. Also services related to roads and locations are considered. Such services are characterized by levels of detail ranging from short schematic information to long comments and explanations, possibly with graphics (e.g., roadmaps) and multimedia (e.g., museum catalogs).

- The user is in principle unaware of the information providers, that he/she discovers progressively as the trip advances, therefore he/she cannot anticipate who will answer the queries and which will be the structure of the reply.

- Each information provider gives information according to a proprietary model and schema. However, given a specific application domain, the different schemata can be assumed "compatible" in some way. The user therefore receives an answer or a set of answers which are expressed in a structure decided by the provider, with a presentation compatible with his/her situation, device and preferences. From the user point of view that means to be able to interpret correctly different answers even if they differ in structure and content type.

From the user point of view, information access is mediated and adapted in order to mask the presence of multiple information providers and the differences in information content and shape. This view is discussed in the next Section.

3 Requirements for context-dependent mobile information access

In a scenario like the one illustrated in Section 2, the design of information systems and services is bound to a number of goals of primary relevance, related to the general problem of data integration, discussed in the following points.

In a mobility oriented information environment different information providers may act; even if, ultimately, the same information is delivered, the format with respect to the design model and to the data schema may be different across different providers. This raises a problem of translation, hence of compatibility and coherence, when the user switches from one provider to the other.

An information management system tailored for such a scenario must allow a user to formulate queries which are (1) independent of the specific information provider, and (2) correct with respect to the information schemata, possibly integrating information coming from different providers and giving the user help for refining the query according to the possibility offered by the specific provider.

A more demanding requirement is the ability to enhance the information gathered from the different providers with the value added obtained by integrating the single sources.

Coherence and compatibility are also key issues due to the discontinuous nature of mobile communication; information access could be interrupted and resumed due to communication problems, and even at short time intervals and within the same information provider, the context in which the communication is resumed could change.

Finally, the ample range of physical device capabilities, joined to the large scope of user needs in terms of information type, from short news to long and detailed information, demand sophisticated adaptiveness.

In addition to what required by a static distributed information environment (e.g., the Web), the mobile environment asks for more demanding requirements about information integration and adaptation.

The user accesses information from different sources because they are managed by different providers, or because they are located in different places. The latter case is dominant in the case of GIS-based information, because it concerns a limited geographical area, and at the borderline between two cells information provided by each cell is incomplete. The same case may happen in mobile communication, at the border between two location-aware services.

Information, therefore, must be kept consistent with respect to mobility, i.e., a change in the information source should not give the user the perception of a discontinuity in its content.

Since variance in user situation is higher in the mobile scenario w.r.t. a static one, information adaptiveness is more complex, and covers not only the device features, but also aspects like attention, interest, capability of interaction, etc., and the very semantics of the query which can depend on the user's location and timing.

Mobile information is by nature multichannel and multimodal, due to the possible switching among communication channels and communication codes during the same session. The user should perceive a continuity (within rea-

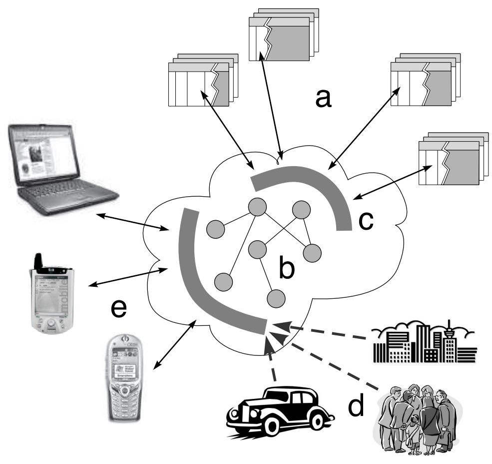

Figure 1. The components of a mobile heterogeneous information system

sonable limits) in information delivered even if the way it is delivered changes, e.g., from visual to acoustic.

In this paper we do not engage in technological issues concerning mobile communication such as quality of service, availability, etc., since the state of art technology can support our proposal at the modelling level independently from performance figures. Figure 1 illustrates the components of a mobile heterogeneous information system that will be discussed in the sequel.

Many databases with different schemata (a). The presence of several databases and different information providers raises a problem of compatibility and integration. The problem is well known in the database area, and can be approached, e.g., by mediators [1, 4, 9]. It is correct to assume that, being the databases designed for the same application domain, semantic compatibility exists to some degree. More precisely, some databases will share compatible subsets of their schemata, and it is possible to define mappings or transformations among the schemata in order to give the user a compatible subset of information across different providers.

Also databases with disjoint schemata will be present in the general case, since each provider can complete the common subset with different extensions, covering different topics of the application domain.

Common ontology (b). A common ontology must be defined in the application domain; problems come from the two ways this requirement can be approached: by referring to the intersection, or to the union of the different databases.

In the first case, a minimum common ontology exists which is managed by a mediator. Other possible interesting information can be passed in a transparent way to the user as links to the local data (and data structures). In the second case we fall into the next item.

Database mapping (c). The ontology is used for mapping user queries to the different databases. In principle two approaches can be taken: selection and integration. In the first case, the user query is directed to the "most suitable" database, according to a measure of correspondence between the query parameters and attributes, and the database schema. In the second case the query is directed to all the compatible databases, and the results are filtered and integrated.

In case of a "selection" approach the mediator is responsible for providing the relevant answers taking them from the "most suitable" local site on the ground of the knowledgeable context. In case of "integration" a common schema exists and the mediator supplies the user with the appropriate answer (and possibly with the DB structure info).

Selecting a specific database has the advantage of returning the user a coherent set of data, both syntactically and semantically (e.g., a homogeneous description of hotel facilities, or a consistent range of restaurant quality judgements), but has the disadvantage of returning only data contained in the selected database. Conversely, integration has the advantage of returning the widest set of relevant information, that however can be (partially) incompatible both syntactically (attributes) and semantically (such as a different set of criteria for evaluating the same restaurant). A compromise could be to integrate the information leaving visible its origin; e.g., by grouping the results by database, and leaving duplicates visible. However, such a compromise gives the user the burden to filter and synthesize the information for practical use, therefore it is somehow in contrast with the goal of the system.

Context processing (d). Context is used to change and refine the user queries according to his/her situation and the available devices. The user context is a multi-facet description, whose components can have different relevance in different situations. E.g., the choice between a synthetic and a detailed answer depends on the user wish, the possibility to pay the needed attention, the time available to read the answer, the device capabilities, and the communication infrastructure. Generally speaking, context analysis is an inference process that, taking elementary data (coming from sensors, user settings, system parameters, QoS analysis, etc.) asserts a complex situation which affects both the user query and the result delivery.

Device adaptation (e). The final adaptation step is device dependent, and tunes the presentation of the query result to the device (channel) capabilities, and also to the part of user situation which directly involves the device; e.g., a user driving a car in heavy traffic can read information on a navigator display if it is short and its size is sufficiently large to capture it at a glance.

4 Discussion

The most interesting key points of such a scenario are the integration of multiple sources, the mobility and the context dependence. We focus our discussion on multiple sources integration; mobility and context dependence will be discussed to a lesser extent and only in relation to information management, since the state of art provides many useful insights.

4.1 Integration of multiple information sources

A classical solution to the problem of integrating multiple heterogeneous information sources in the framework of DB systems consists in defining a global schema on which the local sources map themselves as views (*Local As View* approach). Such an approach is fitted when the DB designer has full control over the entire distributed system, a situation far from the scenario we are considering. A second approach, called *Global As View*, produces a common schema moving from the schemata of the local sources. In this case, it is possible that the common schema does not cover all the information which is locally available, but only a common subset of it, from which the core ontology can be derived. This approach is mostly suited to highly heterogeneous systems, producing a federated DB architecture [16], and seems the most promising in an environment in which even the local schemata can be discovered only at query time.

The presence of many information sources with partial overlap of schemata, and the presence of many information providers with different subjects (even in a same application domain) leads to an information system whose content is largely unknown to the user. We assume that the user be aware of the domain ontology, in a more or less precise way, in order for him/her to be able to ask queries with the appropriate attributes and parameters[1]. In any case, the user does not know in which database the query will be solved, if all the queried attributes will be returned, to what degree of detail the answer will be, etc.; the system must direct the query to the appropriate database or set of databases, and collect the result matching as close as possible the user request.

Due to the unanticipated nature of user queries, the system appears to the user with a "progressive disclosure" behavior, which is typical of Web-based searches on large information spaces. The user becomes aware of the system

[1] We do not discuss about the query language syntax and interface.

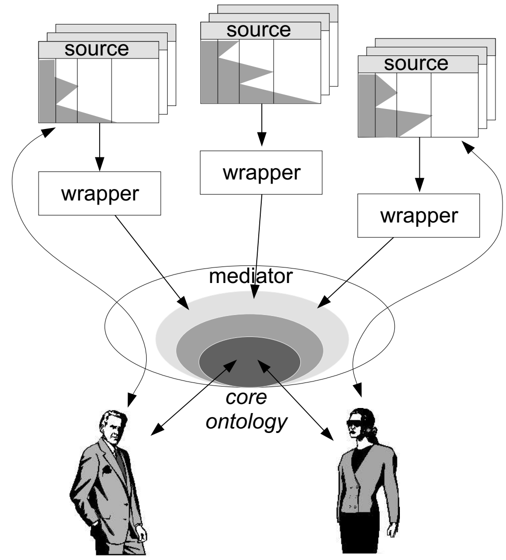

Figure 2. The architecture of a progressive disclosure information system

ontology while he/she proceeds in retrieving information [11]. The user can also discover what databases provide the most useful information, according to some measure of his/her satisfaction that can hardly be formalized, and can change in time. Some degree of "uncertainty" is observed and must be managed by the system, which should be able to adapt itself not only to a static profile of the user, and to an observable user situation (e.g., by sensors or user setting) but also to a dynamic change in user needs which is manifested through his/her interaction with the information system.

We assume that a *core ontology* is known to the system, which corresponds to the minimum knowledge necessary to deal with the application domain. The *core ontology* is based on some common understanding among the users, and can be defined with an empirical rather than a formal approach. However, it can also be defined formally as the minimum schema shared by the information sources (possibly with synonym translation), even if such a definition assumes that the schema have been defined in an intrinsically coherent way. We shall come to this issue later on.

Figure 2 shows the architecture of such a system, based on a mediator and a series of wrappers of the information sources. Two approaches can be adopted for deciding what part of the information sources schemata must be initially revealed to the user, assuming that only further requests discover larger parts of the schemata.

1. The mediator returns only the information covering the core ontology, plus a set of *links* to the sources where

other information is available. Since each source has its own schema, the mediator can return also them to the user (in some synthetic or partial form). In this way the user is able to query directly the different sources for receiving supplementary information.

2. The mediator knows the context in which the user has asked the query, therefore can infer what supplementary information is relevant to him/her. Since such information could exist only in a subset of sources, the answer returned to the user may be not homogeneous: e.g., some records will bear more attributes than others. If the context is correctly interpreted (surely a non trivial task) this anticipatory strategy could be more comfortable for the user.

In the second case the system could lead the context interpretation still further, by refusing to reply with information which is judged incomplete for that context, hence not relevant for the user. How the system evaluates the appropriateness of the information is matter for deeper investigation: as a first suggestion, since the user history is part of the context, a learning mechanism based on repeated patterns of user queries could be the correct approach.

From the design point of view, the two approaches lead to two different logical architectures:

In case 1, each information source wrapper wraps only the common part of the ontology (represented with a grey shade in Figure 2). The mediator dispatches and integrates information pertinent to the common ontology, but acts as a transparent channel for further information owned by the sources.

In case 2, the system knows an integrated, global schema which starts with the core ontology and incrementally grows as users interact with the system. The presence of a common schema guarantees that information common to several sources is given to the user in a common, consistent way, while information owned by a specific source and not by other sources is still rendered according to the source schema (which remains external to the global schema). The mediator task is more complex, since it depends on how users discover and access further information, and is also more seriously affected by the dynamics of the information sources, which may be high, depending on the application domain of interest.

From the user point of view, the progressive disclosure behavior can be managed in two different ways:

1. The *core ontology* remains the only ontology known to the system, but for a specific user (or class of homogeneous users) it is enriched as progressively users make queries and discover information. According to this view, there is no cross-knowledge among users of different classes besides the core ontology.

2. The knowledge that users discover by querying the system is accumulated and increases the initial knowledge. Users benefit from other users, and the system builds incrementally a system ontology based on the activity of the users. Figure 2 illustrates this case.

4.2 Mobility

Mobility is relevant in this framework because the environment of the user changes continuously. Traditionally, context-related issues have been studied with a focus on interaction rather than on information access. In our case, information can change due to mobility. Information sources themselves can be part of the context since the same kind of information can be provided by different owners, depending on the user status (e. g., road traffic information providers on a highway vs. the local sources providing the same kind of information in a large city).

We can notice that the problem of the instability of information sources has been solved on the Web — where users are static while sources can appear and vanish within a few days — by means of crawlers. In the case of mobile users the converse problem has still to be solved since, within the interaction time, sources are static, while it is the user who changes his/her context features [14].

Referring to the tourism example, let's think of a user moving along a road, looking for information about restaurants. His/her requests can be oriented to an immediate target, such as finding a restaurant in the neighborhood for a break in the trip, or to a deferred goal such as finding a restaurant for the evening. The databases which are accessed could be owned by tourist guide publishers, hence related to a wide area, or by the local tourist offices, therefore bound to a smaller area which is of interest or not according to the user trip schedule. In general, we can assume that a query has a temporal and spatial range of validity, and time and space are related to the dynamics of the user.

4.3 Context-dependence

Context-dependence is deeply analyzed in literature about HCI, therefore we approach here only aspects relevant to data source heterogeneity. In this framework, context has two meanings:

- a first meaning refers to something known to the system, which can work on its environmental knowledge;
- a second meaning is related to some information a query can bring with itself, which in some case can augment the system knowledge, while in others can substitute it; among this information we mention spatio/temporal references.

Context refers globally to the so-called "user situation", which has many facets. A set of (almost) static properties directly depends on the user, and is usually called "user profile": language, interest, culture level, goal w.r.t. the information use, etc.. Such properties are relevant in presence of multiple data sources, because they are mostly used to select the appropriate information content and shape among different versions sharing the same semantics.

User properties are related to the user independently of the environment, e.g. a user alone vs. a user in a tourist group. Such properties are used to classify the user information needs in a number of pre-defined classes which limit the variance of the queries and of the results, and may also contribute to the selection of the more appropriate data sources.

The history of the user is bound to discover repeating patterns and similarities that can be used to assist the user in his/her task by anticipating and automatically completing some actions [8, 21].

Other properties describe the user capabilities in the specific location or time where the system is accessed; this is the widest and most variable area, and includes issues such as the degree of attention, the environment physical properties, etc., which can be used to further tailor the query with additional clauses, or to filter the answer with some context-specific processing.

5 Conclusions

In this paper we have traced a framework for accessing distributed and heterogeneous information sources from a mobile and context-dependent user environment. We have introduced requirements about the way information should be integrated or selected (in general mediated), and adapted to the user context. The requirements we have discussed concern system execution time; proper methodologies exist for system design which support information mediation according to the user and environment context [1, 4, 9].

References

[1] B. Amann, C. Beeri, I. Fundulaki, and M. Scholl. Querying XML sources using an ontology-based mediator. In *CoopIS, DOA, and ODBASE 2002 Proceedings*, LNCS 2519, pages 429–448. Springer-Verlag, 2002.

[2] C. Batini, S. Ceri, and S. B. Navathe. *Conceptual database design: an Entity-relationship approach*. Benjamin-Cummings Publishing Co., Inc., 1992.

[3] C. Batini, M. Lenzerini, and S. B. Navathe. A comparative analysis of methodologies for database schema integration. *ACM Comput. Surv.*, 18(4):323–364, 1986.

[4] S. Bergamaschi, S. Castano, M. Vincini, and D. Beneventano. Semantic integration of heterogeneous information sources. *Data Knowl. Eng.*, 36(3):215–249, 2001.

[5] C. Bolchini, F. A. Schreiber, and L. Tanca. A context-aware methodology for very small data base design. *SIGMOD Rec.*, 33(1):71–76, 2004.

[6] C. Bolchini, F. A. Schreiber, and L. Tanca. VSDB: A design methodology. Technical report, MAIS Project R5.2.1, 2004. Submitted for publication.

[7] P. Brusilovsky. Adaptive hypermedia. *User Modeling and User Adapted Interaction*, 11(1–2):87–110, 2001.

[8] A. Celentano and F. Pittarello. Observing and adapting user behavior in navigational 3d interfaces. In *Proc. of the Working Conf. on Advanced Visual Interfaces*, pages 275–282, 2004.

[9] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. D. Ullman, and J. Widom. The TSIMMIS project: Integration of heterogeneous information sources. In *16th Meeting of the Information Processing Society of Japan*, pages 7–18, Tokyo, Japan, 1994.

[10] G. Chen and D. Kotz. A survey of context-aware mobile computing. Technical Report TR2000-381, Dartmouth College, Department of Computer Science, 2000.

[11] H. Chen, T. Finin, and A. Joshi. An ontology for context-aware pervasive computing environments. *Special Issue on Ontologies for Distributed Systems, Knowledge Engineering Review*, 18(3):197–207, 2004.

[12] A. K. Dey and G. D. Abowd. Towards a better understanding of context and context-awareness. In *Proc. CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness*, The Hague, The Netherlands, 2000.

[13] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The anatomy of a context-aware application. *Wireless Networks*, 8(2/3):187–197, 2002.

[14] V. Kumar and S. B. Zdonik. NSF workshop on context aware mobile database management (CAMM). Final Report, 2002. http://www.cs.brown.edu/nsfmobile/report.pdf.

[15] M. Lenzerini. Data integration: a theoretical perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 233–246. ACM Press, 2002.

[16] M. T. Ozsu and P. Valduriez. *Principles of distributed database systems*. Prentice-Hall, Inc., 1991.

[17] M. Perkowitz and O. Etzioni. Adaptive web sites. *Communications of the ACM*, 43(8):152–158, 2000.

[18] E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4):334–350, 2001.

[19] W. N. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *IEEE Workshop on Mobile Computing Systems and Applications*, pages 85–90, Santa Cruz, CA, US, 1994. IEEE Computer Society Press.

[20] A. Schmidt, M. Beigl, and H.-W. Gellersen. There is more to context than location. *Computers & Graphics*, 23(6):893–901, 1999.

[21] F. A. Schreiber, F. Barbic, and S. Madeddu. Dynamic user profiles and flexible queries in office document retrieval systems. *Decision Support Systems*, 5(1):13–28, 1989.

[22] C. Stephanidis. Adaptive techniques for universal access. *User Modeling and User Adapted Interaction*, 11(1–2):159–179, 2001.