Requirements for Context-dependent Mobile Access to Information Services

Augusto Celentano  Fabio Schreiber, Letizia Tanca
Università Ca’ Foscari di Venezia  Politecnico di Milano

MIS 2004, College Park, MD, August 25, 2004

Background research

Relevant research effort in three areas

- adaptiveness & personalization
- context awareness
- information integration

How to integrate them into a single comprehensive framework?
**Background: adaptiveness & personalization**

Research effort in several fields
- user interaction, hypermedia systems, web applications, …
- improve information access and understanding

New emerging paradigms of information management and use
- context (context-awareness, ubiquity)
- sensorial and communication channels (multimodality, multichannel delivery)
- environment (mobility, pervasiveness)
- …

**Background: context-awareness**

Mainly associated to the concept of location…
- computational context
- user context
- physical context
- temporal context

Goal of information adaptation to context is to make information use easier
- relevant information delivery
- no information overload
- “calm” interaction
Background: information integration

Studied since the end of the 70’s
- combine data stored in heterogeneous systems
- provide the user with a unified view

Issues concern architecture and semantics
- concept similarity (ontologies)
- data conflicts
- integrated vs. federated DB architecture
- integration at design time vs. run time

Towards an integrated view of the problem

Background issues individually studied
- context-awareness and adaptation have been addressed in the database and IR worlds only recently (and marginally)
- the interaction with mobility-based information systems raises new challenges

Three main issues
- context dependence of information
- mobility of the end user
- information design vs. information access

Focus on information identification (a DB p.o.v.)
An application scenario (1)

A traveling user meets different providers offering information services related to tourism issues

- information of practical interest at different levels of detail, possibly with multimedia content
- information related to, and limited by, location
- user is unaware of providers: progressive discovery
- each provider gives information according to a proprietary schema
- “plausible” commonalities exist among schemata
- the user can understand and combine information

An application scenario (2)

Information of practical interest at different levels of detail...

- different providers = different information schemata
- no common schema exists, or it is incomplete / redundant
- structured / semi-structured / unstructured information
- similar to Web search environment...
An application scenario (3)

Information related to, and limited by, location
- the user location is (explicit) part of the query
- the relevant location can be the one at current time or at another time...
- ...extends the “query/data validity” concept of temporal databases
- location identification is mostly relevant in “strong” mobility situations

An application scenario (4)

User is unaware of providers: progressive discovery
- the user knows his/her needs, but cannot ask a correct query at first instance
- the user refines search as previous queries results expose information on accessible schemata
- schemas may be incompatible, user makes partial queries
An application scenario (5)

“Plausible” commonalities exist among schemata, the user can understand and combine information

- commonalities refer to a “common sense” understanding of the application domain (not guaranteed...)
- user is able to “semantically integrate” information structured according to different schemata...
- …but automatic integration is an open research issue

Requirements for CM-IS (1)

The information management system should

- allow the user to formulate queries independent from the provider …
- … but correct with respect to the schemata
- integrate information from different sources …
- … helping the user to contextually refine the query according to the specific provider
- keep track of the context (mobility, time, environment, device) …
- … giving the user a feel of continuity
**Requirements for CM-IS (2)**

Components of a mobile heterogeneous information system

- many DBs with different schemas
- common ontology
- database mapping
- context processing
- device adaptation

**Many DBs with different schemas**

A problem of compatibility and integration

- semantic compatibility exists at some degree, bound to the application domain
- a known problem in DB area, solved by mediators when a common ontology exists
- not (yet) approached in IR/Web area, where user makes integration (...)
- some degree of user participation is unavoidable
**Common ontology**

Necessary for result integration

- usually approached as intersection or union of the different DBs
- the intersection defines a “minimum common ontology”, i.e., the minimum knowledge necessary to understand the application domain
- the union defines a dynamic “multiple valued” ontology
- minimum ontology can be managed by a mediator, supplementary information can be passed to the user as links to local data

**Database mapping**

A union-based ontology maps queries to the DBs

- a selection approach sends queries to the “best” DB
- an integration approach sends queries to all DBs, filtering and integrating the results
- selection improves data coherence, integration improves data completeness (~ precision and recall concepts in IR?)
- how to select the “best” DB?
- how to filter and integrate the results?
**Context processing**

Context is a system knowledge able to change its behavior
- in information management, context also denotes some information which can augment or modify a query
- space and time references
- additional parameters (implicit for the user)
- context may also participate to result filtering (e.g., user preferences)
- different stages of processing for separate context features

**Device adaptation**

Almost independent from the specific scenario, may be impacted by data integration choices
- a DB union approach generates large amount of data at once
- a DB intersection approach generates many links to detail data
- the range of possible structure and content adaptation may increase dramatically
**Progressive discovery of information (1)**

A progressive discovery system is based on a mediator and a set of wrappers:
- a core ontology is initially known to the system...
- ...based on common understanding of the application domain...
- ...formally defined as the minimum schema shared by information sources...
- ...which must be known in advance

**Progressive discovery of information (2)**

Case 1: the mediator returns the information covering the core ontology:
- returns links to non-common local information
- can return local schema information
- each wrapper wraps only the common part of ontology
- the mediator is a transparent channel for local information
Progressive discovery of information (3)

Case 2: the mediator returns also additional information based on the user context

- how?
- what about data coherence?
- the mediator knows a global, growing ontology

Progressive discovery of information (4)

From user point of view the progressive discovery can be managed in two ways

- the system maintains a single core ontology
- extensions are accumulated on a per user basis
- the initial core ontology is increased as users query the system
- users benefit from other users access
**Conclusions (1)**

Context-dependent mobile access to information systems raises new problems in DB design and DB access

- at design time, schemata self-description is required to build common/integrated ontologies at query time (meta-description)
- at run-time, a new type of mediator is required, able to work on a dynamically evolving ontology
- context capturing has not been discussed but is a key problem
- multimedia boosts data heterogeneity problems

**Conclusions (2)**

Despite a large research effort in data integration, in our framework this is still an open issue

- *traditional* data integration technology assumes that local schemata and/or global schema are known (GAV / LAV)
- in our case a GAV approach seems better...
- ...but schemata are progressively discovered, can be loosely compatible...
- ...then a global schema could be inconsistent
- framework close to P2P information processing
Future work

P2P information systems are close to the problems discussed in our framework, but...

- context, mobility and integration are considered non relevant issues...
- ... since they only approach the problem of “finding” something
- so far, only files, but proposals for P2P databases are emerging
- consistency and validity in time are key requirements due to the dynamics of the environment

Bibliografía

- Kumar V., Zdonik S. B. - NSF Context-Aware Mobile database Management (CAMM), Final report - 2002