Semantic Integration between Context-awareness and Domain Data to Bring Personalized Queries to Legacy Relational Databases

Vinícius Maran[1,4], Alencar Machado[2], Iara Augustin[3] and José Palazzo M. de Oliveira[4]
[1]*Coordination Office, Federal University of Santa Maria, Av. Presidente Vargas, 1958, Cachoeira do Sul, Brazil*
[2]*Polytechnic School, Federal University of Santa Maria, Santa Maria, Brazil*
[3]*Center of Technology, Federal University of Santa Maria, Santa Maria, Brazil*
[4]*Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil*

Keywords: Database, Context-awareness, Ubiquitous Computing, Query, Semantic Web, Ontology.

Abstract: Context-awareness is a key feature in ubiquitous middleware. Mainly, it is applied to adapt services and interfaces of applications that use ubiquitous features. The application of context information to personalize data queries is a recent topic in computing and still presents a large number of challenges. One of the main gaps evidenced by this research field is the lack of integration between context information, which is designed and used by ubiquitous middleware, and domain data, which are frequently persisted in relational databases. This integration is necessary because context can be used as a filter for content query. This position paper presents a motivational scenario that clarifies the necessity of the integration between context, used by ubiquitous middleware, and relational data, a comparison between the state of the art of the field, a list of research opportunities in the field, and a proposal of a framework that uses ontologies to integrate context and domain data, modeled and stored in relational databases.

1 INTRODUCTION

Mark Weiser (1991) presented a series of scenarios where computing acts to help users in their daily tasks, without even they are able to notice the use of computers in these tasks. These scenarios originated the ubiquitous computing research field. Recently, ubiquitous computing area involves a set of technologies and methodologies of implementation.

One of the key concepts used in ubiquitous computing is called context-awareness. In recent definitions (Makris et al., 2013) (Perera et al., 2014), context information is defined as a measured and inferred data about current state of entities present in the environment. This information can be used in systems for adaptation in query of content and execution of services. Context is modeled and used in many forms by ubiquitous systems. Recent surveys (Bettini et al., 2010) (Strang; Popien, 2004) show that the representation of context based on ontologies presents a series of benefits over other forms of representation, like the existence of patterns to define ontologies and high expressiveness of them. Ontology can be defined as a formal and explicit specification of a shared conceptualization (Borst, 1997) and it can be represented in several languages. These languages are classified as: (i) based on logic, or (ii) serialized in XML languages. The second group is the most currently used because there are standards for representation, managed by W3C (2016).

Context information can also be applied in many forms in ubiquitous systems. The most frequent usage of context are (Dey et al., 2001): (i) dynamic adaptation of services, (ii) personalization of user interfaces, (iii) search of resources, and (iv) content and data querying. Content and data, related to the domain of the application, must be queried in a personalized form in ubiquitous systems, mainly using context information to filter it. This feature differs from the process of querying data in traditional systems in two ways. First, ubiquitous systems must query heterogeneous sources of data, including legacy relational databases. Second, the filtering information that is used in the query is not informed by the user explicitly, because the context information is automatically collected and inferred by ubiquitous middleware and then it is used in the query (Maran et al., 2015a). So for these domain data, persisted in relational databases, to be recovered in a contextualized manner, it is necessary to create forms of interconnection between context information, which are often represented in ontologies and domain

Maran, V., Machado, A., Augustin, I. and Oliveira, J.
Semantic Integration between Context-awareness and Domain Data to Bring Personalized Queries to Legacy Relational Databases.
In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016) - Volume 1*, pages 238-243
ISBN: 978-989-758-187-8

information, persisted in relational databases (Bolchini et al., 2013). This paper presents a motivational scenario of the data access based on context, an overview of the state of the art in the data access based on context research area, a list of research opportunities in the area, and a proposal of a framework to link legacy relational databases and context models, used in ubiquitous middleware.

The paper is organized as follows: In Section 2 a motivating scenario and the main concepts related areas of context-awareness, ontologies, and contextualized data querying are presented. In Section 3 a qualitative comparison between recent research and a list of research opportunities are presented. In Section 4 a framework to integrate legacy relational databases and context models based on ontologies is presented. In Section 5 the conclusions and future work are presented.

2 WHY UBIQUITOUS COMPUTING NEEDS TO GET CLOSER TO RDBMS?

Relational Database Management Systems (RDBMSs) traditionally support SQL queries. For a system make a query in the database, it is necessary that this system and the user inform in an explicit form the terms and conditions that will filter the results over the relations. This works well with traditional system and database design process, mainly because the database schema is designed to fit the application and domain information. But with recent advances in systems and the large increase of the information size in databases, some well-known problems became relevant about queries in RDBMSs field: **(i)** To perform a query is required to system and user to inform filtering criteria (the selection and projection criteria). A known problem is that often a few users report, or do not report filtering criteria. So the amount of information retrieved is large and users need to filter the information manually (Bolchini et al., 2013); **(ii)** The same instance of the database must be accessed by multiple systems, modelled in different manner, although of being designed for the same domain. This fact induces a known problem about semantics of the data, frequently supported by the use of ontologies (Dey et al., 2001).

Ubiquitous middleware frequently use ontologies to describe context models (Bettini et al., 2010). This way, applications that use ubiquitous features must use these context models to perform actions in a personalized form. These context information, previously measured and inferred by ubiquitous middleware, can be informed to the application or the database at the moment of the query (Perera et al., 2014). However, context information is not informed all in the same way every time. For example, in a specific moment, a body sensor can send some important information, but in other moment, this sensor cannot be available, and the context associated to this sensor cannot be used (Makris et al., 2013). This problem is related to the two previously mentioned problems in RDBMS queries (i and ii), mainly because the context information is not informed in a complete and in the same form in every query, and because the context is frequently modelled in ontologies. Furthermore, the design process of a relational database and context modeling occur in isolation from each other.

2.1 Motivational Scenario

Ubiquitous middleware have been designed and applied in several fields. Figure 1 presents an overview of the motivational scenario, described below. Recent researches (Stavropoulos et al., 2013) (Machado et al., 2014) describe smart university environments, which have a wide variety of educational resources that are managed by a ubiquitous middleware. Some of these resources can be recommended to students, which study in these universities (Stavropoulos et al., 2013).

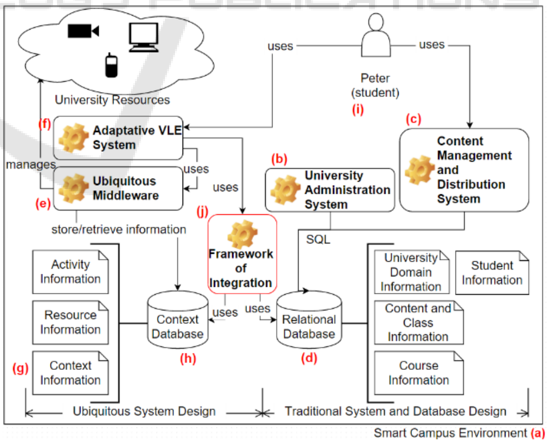

Figure 1: Motivational Scenario.

Let us imagine that a university campus **(a)** uses an university administration system **(b)**, and a content management system **(c)**, which allows teachers to provide supplementary materials, exercises and activities for students, which in turn access the system to get the materials. These systems store and retrieve information from a relational database **(d)**, using SQL queries. In a determined moment, the university starts to use a ubiquitous middleware **(e)** with an adaptive Virtual Learning System (VLE) (Maran et al., 2015) **(f)** to show content and resources of the university to students. To adapt its execution and content to the user, the middleware manages the context information, captured from the environment **(g)**. To represent the context, the middleware uses an ontology, which represents the context after it being measured and inferred, and stores and retrieves this context from a database **(h)**.

Peter is a student **(i)** in an Electrical Engineering course in this smart campus. Currently, Peter is on the 2nd semester of the course and is attending the AL101EE discipline, called *Algorithms and Data Structures 1*. In a particular class, the context management middleware informs the context information about the user (1), about the educational context (2), information about location of the user (3), information about the device used by the user, and temporal and activity context information (5) to the adaptive VLE:

Context_User = {Peter, Student, 2ndSem, EEngineering} (1)

Context_Educational = {Peter,AL101EE,Class4,CBranch} (2)

Context_Location = {Country, City, SmartUniv} (3)

Context_Device = {AndroidBasedPhone} (4)

Context_Activity = {in_class, Init_time, End_time} (5)

This context information is modelled in the ontology, and it is used by the ubiquitous middleware. Currently, Massive Open Online Courses (MOOC) are offered in VLE **(f)** to an audience of the university as an opportunity to expand their knowledge. In a class of *Algorithms and Data Structures 1*, the teacher introduced the concept of Conditional Branch. At the end of the lesson, the recommender system showed to Peter in his smartphone the recommendation of the MOOC about algorithms. Peter is interested in more information about this topic, as recorded in MOOC.

By the time, the student enrols in MOOC, the ubiquitous middleware transfers context information (with the permission of the student) to the VLE system, and this information is used by the MOOC. Thus, profile and devices information associated with the student can be used by MOOC to filter course information to the student. The MOOC about Algorithms that Peter signed up presents algorithms and questions to assess student understanding on the basic algorithm structures applied to the student's course. Currently, the completion rate MOOCs varies between 5 and 10% (Pretz, 2014). Recent work (Pretz, 2014) (Quinn et al., 2014) attributed as a cause for this low completion rate the fact that the information related to courses are presented in the same way to all students, regardless of context information and information related to the student's profile. Originally, the MOOC platform was designed to use a relational database, with a pre-defined structure (EDX, 2015). As the context model is represented in ontologies and is managed by the ubiquitous middleware, it can not be informed explicitly at query time. Thus, there is a need to use up a framework **(j)** that implements a model that allows that context information used by the ubiquitous middleware to be used in information filtering. This scenario presents two relevant features regarding to recent research related to context:

- There is an exchange of context information between the middleware that manages context information on university campus and an external system (represented by the VLE). This is a research problem addressed in recent work (Makris et alk., 2013) (Perera et al., 2014);
- There is a gap for binding between context information and field data regarding persistence and recovery of data using context information in the query process. This gap is directly related to two characteristics: (i) information relating to the scope are usually modeled relationally and persisted in relational databases. This is evidenced by the wide adoption of RDBMS (DBEngines, 2016) and (ii) information related to context used by ubiquitous systems are often modeled on ontologies (Bettini et al., 2010). In recent work related to ubiquitous middleware, application independent ontologies were used as the basis for context modeling.

2.2 Context-awareness and Ontologies

Context is a broad term and has a set of definitions, according to specific fields. Context can be defined as *"any information that might be used to characterize the situation of entities that are considered relevant to the interaction between a user and an application"* (Dey et al., 2001). Context can be modeled in various ways. Some of the most common forms to represent context are: key-value pairs, object orientation model, logic based model, ontologies, and mark-up languages. According to Bettini et al. (2010), ontological models have greater capacity for representation and inference. Spatial models are more efficient if compared to ontologies and object oriented modeling, but they do not have as much representation capacity compared to the ontological models. This way, ontologies are the most used form to represent context information in ubiquitous architectures. Context-Aware Data Query states that the data retrieval and filtering operations should be based on context information reported to the system time of information query. Thus, stored information, whether in structured or unstructured form, can be adapted according to the given context. As the context involves large sets of information, defined in fields, it can also be considered a document in a collection. Thus, contexts should be available for query in the same way that a common document related to the domain. Therefore, the context can be used in two ways in information retrieval (Bolchini et al., 2013): (a) To derive a query that returns the documents that best fit the required context; (b) To treat context as a document, i.e., the context becomes the source of information being queried.

Ontology is defined as a formal and explicit specification of a group of concepts in a shared form (Borst, 1997). As context model is a type of knowledge about the environment and the entities that composes the environment of the user and system, it can be represented in ontologies. Research have modeled context ontologies. These ontologies vary in multiple aspects, for example the number of defined concepts, domain of application, validation methods and language of representation. Rodríguez et al. (2014) conducted a comparison between ontologies that represent context information and activities. It was found that PiVOn Ontology (Hervás et al., 2010) was the ontology that better attends the comparison criteria. This ontology has been used in a set of works related to ubiquitous computing (Rodriguez et al., 2014). Context information are essential for ubiquitous systems because the treatment of this information and its use allows that ubiquitous system be able to adapt itself to the needs of users and other systems. These adjustments must be made in real time and can result in both application behaviour change and in information retrieval.

3 STATE OF THE ART

Recent research propose models and extensions of existing tools and models to integrate context models and domain data querying. HyConSC (Anderson et al., 2006) is a framework that allows context-based consultations to be integrated to applications that use relational databases. To realize the extension of queries, the framework uses its own context model, represented in graphical model. The context information are persisted as notes in documents. The Context-Relational Algebra was proposed as an extension of relational algebra, which supports a logical model that allows to integrate contextual information to relational databases (Martinenghi et al., 2009). A photo recommendation based on context tool was proposed by Viana et al. (2011). This tool uses semantic context information to recommend photos based on similarity calculation.

The CARVE methodology (Context-Aware Automatic View Definition over Relational Databases) was defined as a proposal for integration between relational databases and context information in the form of a process of automatic generation of views based on context (Bolchini et al., 2013). The implementation of the methodology is performed in a number of phases, some of which must be manually set. The context information is modelled in Context Dimension Trees. HARE (Time-Aware Location-Aware and Health-Aware Recommender) is a content recommendation application that uses time, location and health data of the patient to perform recommendations (López-Nores et al., 2013). To perform the content recommendation, ontologies were previously used to describe the metadata about that content. An architecture was proposed by Hahm et al. (2014) to perform the custom recovery of engineering documents based on analysis of user profile. To conduct a qualitative analysis on the state of the art in data query based on context, it was made a list of important features for analysis:

Context Model: The way that context information are modeled. It can be modeled in relational form (BDR), object-oriented (OO), based on logic (BL), graphical models (G) or ontologies (Onto); **Domain Model:** The way the domain information are modeled and integrated with context information. It can be modeled in relational form (BDR), as documents with semantic annotations (DAS), or as domain ontologies (Onto); **Integration Mode:** The way the integration of context information and domain data is made. It can be based on ontologies alignment (Alin), relational-algebra expressions (RA), or integration by algorithms by property analysis (Alg); **Query Language:** Data recovery can be defined by relational algebra expressions (AR), defined in SQL, SPARQL or SQWRL; **Database Model:** Some researches do not specify (NE) using DBMS, others use models like Triple Stores (TS) or Relational databases (BDR). Some researches do not use databases (NU). Based on the list set up to carry out analysis of related work. The result of the features analysis is shown in Table 1. As can be seen, none of the research models context based on a generic context ontology. This contributes to the context share issue (Makris et al., 2013) (Perera et al., 2014) between ubiquitous systems. Generic ontologies are often used in ubiquitous systems generally to adapt the execution of services and to infer contexts. However, there is the existence of a

Table 1: Qualitative analysis of state-of-the-art.

Features / Work	**(Anderson et al., 2006)**	**(Martinenghi et al., 2009)**	**(Viana et al., 2011)**	**(Bolchini et al., 2013)**	**(López-Nores et al., 2013)**	**(Hahm et al., 2014)**
Context Modeling	BDR	BDR	Onto	G	Onto	Onto
Context Modeling based on a generic ontology	-	-	-	-	-	-
Domain Modeling	BDR	BDR	Onto / DAS	BDR	Onto	Onto / DAS
Integration Mode	Alg	AR	Alin	AR	Alin	Alin
Query Language	SQL	AR	SPARQL	AR	SPARQL	SPARQL
Specific for a Domain	No	No	Yes	No	Yes	Yes
Database Model	NE	BDR	NE	BDR	TS	NU

gap for binding contexts and ubiquitous systems domain data (Perera et al., 2014).

Another important feature that was observed in relation to the related work is that none of the tools modeled context information in ontologies - to be the most complete and extensible, and integrated it with relational databases - the most widely used format for domain data, and persisted only in a database format. Thus, it is proposed in this paper a framework of integration and recovery of domain information based on context, modeled in ontologies, and relational databases, which represent domain information.

4 A FRAMEWORK TO INTEGRATE CONTEXT AND RDBMS

This paper presents a framework for domain information query. This information is modeled and persisted in relational databases, and context information are modeled on ontologies. This way, the model uses an ontology-based model to perform information filtering based on context information. So even without changing the structure of the database that represents the application domain information, systems can use context information modeled with high expressiveness for querying this information. Figure 2 presents a high-level view of the proposed model to integrate context modeling with relational data querying. The framework is divided in five main levels, named:

(a) External Entities: Ubiquitous middleware perform management of context and environmental resources. Even in these environments, external systems can use resources from middleware. As ubiquitous middleware manages context information, they may be required to inform the current context of the environment when performing a query. External systems in turn hold queries to data using the model, which returns a data set that comply with the filters informed by external systems, and context informed by ubiquitous middleware;

(b) Interfaces: To communicate with the framework, ubiquitous middleware use REST interfaces, informing the context through representations in JSON-LD language. The process of serialization of context ontologies in JSON-LD was previously presented in (Maran et al., 2015). External systems in turn carry the information query using SQL language;

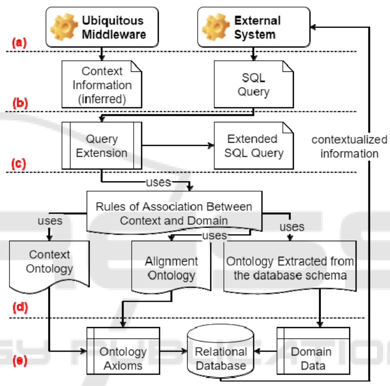

Figure 2: Overview of the framework.

(c) Query Extension: The query carried out by systems through SQL format is extended through a process. This process performs the query in association with rules that define links between context and domain data and checks for relationships that can be used to query. In this process, the definitions of ontologies are used to describe context, domain database schema, and the alignment of definitions of concepts;

(d) Conceptual Layer: This layer represent conceptual schemas and ontologies used by the model. In this paper, conceptual model is a modeling on a specific area, where this model was the basis for the logic model for a database that stores data about the domain. Some authors classify this type of conceptual modeling as a lightweight ontology (Maran et al., 2015a). To perform context-based information query, the model uses three distinct sets of settings: (i) an ontology that describes context based on PiVOn (Hervás et al., 2010), a context ontology independent of application, (ii) A conceptual model that describes the domain database., and (iii) An ontology defining generic concepts and relationships to allow the alignment of context and domain information. The alignment of the ontologies was previously presented in (Maran et al., 2015);

(e) Persistence: Instances of conceptual models and ontologies used by the proposed model are persisted in a relational database. An initial implementation of the serialization of the definitions was presented in (Maran et al., 2015a).

5 CONCLUSIONS

Ontologies have been used by ubiquitous architectures for representing context information. Furthermore, inference rules have been used for making inferences about the context, which according to current definitions is measured and inferred knowledge about the status of entities.

Relational databases are used in most applications. As shown by motivating scenario, the context of use in the structured information retrieval in relational databases is relevant. In this work, an overview of the field was presented, as an study about the state of the art and the proposal of a model of integration between context, modeled on ontologies and domain information, modeled in relational databases. The framework is in implementation phase. As the future work, we pretend to evaluate it in a scenario based in the motivational scenario presented in this work.

REFERENCES

Anderson, K. M.; Hansen, F. A.; Bouvin, N. O., Templates and queries in contextual hypermedia. In: *Proceedings of the seventeenth conference on Hypertext and hypermedia. ACM, 2006*. p. 99-110.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., Riboni, D. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing, v.6*, n.2, p.161-180, 2010.

Bolchini, C; Quintarelli, E; Tanca, L. CARVE: Context-aware automatic view definition over relational databases.*Information Systems,v.38*,n.1,p.45-67, 2013.

Borst, W., Construction of engineering ontologies for knowledge sharing and reuse. Universiteit Twente, 1997.

DBEngines. Knowledge Base of Relational and NoSQL Database Management Systems. Website. Available at: http://db-engines.com/en/ranking.

Dey et al., A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction, v.16*, n.2, p.97-166, 2001.

Edx Website. Available at: code.edx.org/.

Hahm, G. J. et al. A personalized query expansion approach for engineering document retrieval. *Advanced Engineering Informatics, v.28*, n.4, p.344-359, 2014.

Hervás, R.; Bravo, J.; Fontecha, J., A Context Model based on Ontological Languages: a Proposal for Information Visualization. *JUCS, v. 16*, n. 12, p. 1539-1555, 2010.

López-Nores, M. et al. Context-Aware Recommender Systems Influenced by the Users' Health-Related Data. In: User Modeling and Adaptation for Daily Routines. *Springer London, 2013*. p. 153-173.

Makris, P.; Skoutas, D. N.; Skianis, C. A Survey on Context-Aware Mobile and Wireless Networking: On Networking and Computing Environments' Integration. Communications Surveys & Tutorials, *IEEE, v. 15*, n. 1, p. 362-386, 2013.

Maran, V., Palazzo M. de Oliveira, J., Pietrobon, R., Augustin, I. Ontology Network Definition for Motivational Interviewing Learning Driven by Semantic Context-Awareness. In Computer-Based Medical Systems, *2015 IEEE 28th I.S.* on. p. 264-269.

Maran, V., Machado, A., Augustin, I., Wives, L. K., & de Oliveira, J. P. M. (2015a). Proactive Domain Data Querying based on Context Information in Ambient Assisted Living Environments. In: *17th International Conference on Enterprise Information Systems*.

Martinenghi, D.; Torlone, R., Querying context-aware databases. In: Flexible Query Answering Systems. *Springer Berlin Heidelberg, 2009*. p. 76-87.

Machado, G. M.; Palazzo, J., Context-aware adaptive recommendation of resources for mobile users in a university campus. In: Wireless and Mobile Computing, Networking and Communications, 2014 *IEEE 10th International Conference* on. p. 427-433.

Perera, C., S. Member, A. Zaslavsky, e P. Christen. Context aware computing for the internet of things: A survey. Communications Surveys & Tutorials, *IEEE, v. 16*, n. 1, p. 414-454, 2014.

Pretz, K. "Low Completion for MOOCs". IEEE Roundup. Available at: http://theinstitute.ieee.org/ieee-roundup/opinions/ieeeroundup/low-completion-rates-for-moocs. 2014.

Quinn, S., Bond, R., & Nugent, C. D. "An Ontology Based Approach to the Provision of Personalized Patient Education". In: Ambient Assisted Living and Daily Activities. p. 67-74. *Springer. 2014*.

Rodriguez, N. D. et al., A survey on ontologies for human behavior recognition. *ACM Computing Surveys, v. 46*, n. 4, p. 43, 2014.

Stavropoulos, T. G. et al. aWESoME: A web service middleware for ambient intelligence. *Expert Systems with Applications, v. 40*, n. 11, p. 4380-4392, 2013.

Strang, T.; Linnhoff-Popien, C. A context modeling survey. In: *Workshop Proceedings. 2004*.

Viana, W. et al. Towards the semantic and context-aware management of mobile multimedia. *Multimedia Tools and Applications, v. 53*, n. 2, p. 391-429, 2011.

W3C Website. Available at: www.w3.org/.

Weiser, M. The computer for the 21st century. Scientific american, v. 265, n. 3, p. 94-104, 1991. Available at: http://doi.acm.org/10.1145/329124.329126.