The use of ontologies in contextually aware environments

Ian C. Millard, David De Roure, Nigel Shadbolt
School of Electronics and Computer Science
University of Southampton,
Hampshire, SO17 1BJ, UK.
{icm02r, dder, nrs}@ecs.soton.ac.uk

ABSTRACT
In this paper we outline work in progress related to the construction of contextually aware pervasive computing environments, through the use of semantic and knowledge technologies. Key to this activity is modelling both *where* and *what* a user is doing at any given time. We present a prototype application to illustrate this work and describe part of its implementation.

Keywords
Context, contextual awareness, ontologies, location, modelling, semantic technologies, knowledge technologies.

INTRODUCTION

During the course of the last decade, our lives have become increasingly dependent on the availability and exchange of information. In today's world, knowledge can be equated to people, money, flexibility, power, and competitive advantage. Knowledge is more relevant to sustained business than capital, labour or land, yet in most organisations it remains the most neglected asset. For large businesses and organisations, being able to access the right information at the right time is key to success.

The contextually aware environment aims to aid in this task, presenting the right information to the right users, at the right time and in the right place. In order to achieve this, a system must have a thorough understanding of its environment, the people and devices that exist within it, their interests and capabilities, and the tasks and activities that are being undertaken. Such a system must recognise that people often work on a number of different projects, collaborate with other colleagues, and have a unique set of skills and concerns. That is to say, the system must be able to identify where, and under what context each person is working.

Within the Intelligence, Agents, Multimedia Group at Southampton, there is significant research interest in both knowledge/semantic technologies and pervasive computing. The work discussed in this paper is a result of experiments in the combination of these two fields, utilising the expressive power and interoperability of semantic technologies as the basis of building blocks for the creation of a pervasive computing environment which exhibits some aspects of contextual awareness.

LOCATION AWARENESS

Sensing the location of people and devices within a building is non-trivial. In large open outdoor spaces, technologies such as GPS and RF signal strength analysis allow location to be determined accurately, typically with a resolution of a few metres. However, inside a building or in built up areas, these techniques become unreliable, if not dysfunctional, hence other methodologies are required.

Several research projects have been undertaken in the area of localised high resolution location tracking, with prototype systems including triangulation of position utilising infra-red [12] ultrasonic [9, 13] or radio signals [2, 3], location sensitive flooring [7], and body mounted accelerometers. However, deploying any of these systems on a large scale would require a significant investment in proprietary infrastructure.

Conversely, there are many off-the-shelf systems that can be used to sense the presence of a tag, fob or card in close proximity to a specific receiver unit. These include RFID, magnetic swipe cards, smart cards and iButtons, and are often used for access control or the identification of products in industrial applications. However, such technologies typically have an active range of only a few hundred millimetres, rendering them only suitable for presence detection at point locations, such as at a particular door, desk, or computer terminal for example.

The different forms of physical location technology offer presence detection with a variety of different accuracies, reliabilities, and ranges in which those observations can be made. In general, location information is subject to error which should itself be characterised and taken into consideration. A location system is likely therefore to have to take into account data from a number of different information sources, potentially including both spatial (coordinate) and symbolic point information. Combining these two fundamentally different types of system is non-trivial and is an important research topic. [5]

For applications and services to operate efficiently within a pervasive environment, some form of middleware is required to monitor these potential sources of location information, and present that data through an integrated interface. To achieve this, it is proposed that a common model is created, representing the environment in which the users and devices will be situated. This data should be recorded in RDF, mediated though an ontology, and stored in a repository, for reasons of flexibility, data reuse and automatic inference.

Investigative work has been undertaken in this area, with the construction of a location ontology, defining a common model which can be used to describe the environment in which the users and devices in an environment will be situated.

The model is not designed primarily from a spatial perspective, as it is thought that precise coordinate positioning is not critical for many pervasive or contextually aware applications. It is more important to represent the type and/or purpose of different areas, their relative proximity and positioning, and to be able to assert that objects are situated in such locations. These requirements come from experience across a number of projects within the Equator research programme (see www.equator.ac.uk).

ONTOLOGY

The ontology uses a hierarchical notion of locations and sub-spaces built from the principle concept of an Abstract-Space. Properties are defined such that spaces may be located in, be visible from, and/or adjacent to other spaces, owned by an organisation, and have a number of usual occupants. An Enclosed-Space is a subclass of Abstract-Space, representing any area which has some enclosing boundaries that do not permit free access into and out of that space, and one Enclosed-Space may permit access to other Enclosed-Spaces. Such areas are then further hierarchically subclassed to represent common features such as Buildings, Floors and Rooms.

Our starting point has been the reference ontology developed in the Advanced Knowledge Technologies program (see www.aktors.org) which already contains a model of location, primarily intended to represent geographical addresses. Classes such as Country, City and Town are defined, again in a hierarchical relationship from an abstract Location. To make best advantage of these existing classes, the Abstract-Space and Location classes have been declared as equivalent, permitting crossover of properties and reuse of existing entities.

Using OWL to define the structure of the location model offers several useful benefits. In addition to the usual subclassing and properties offered by RDFS, OWL permits the use of transitive properties. By defining the is-located-in property as transitive, in the situation where RoomA is located in FloorB and FloorB is located in BuildingC, then a reasoner can automatically infer that RoomA is located within BuildingC. Similarly, by defining the is-adjacent-to property to be symmetric, then if RoomA is adjacent to RoomB, then there is no need to state explicitly that RoomB is adjacent to RoomA. These kinds of inferences provide a powerful technique for populating the model, as location data from different sources can be easily combined, and queries can be performed at different granularities of location. For example we can easily find who is in a room, on a floor of a building, or all of the towns in a Country. In addition, through the use of the permits-access-to property, it may be possible to use a recursive search to find navigable routes between two given locations.

In addition to this hierarchical notion of inter-connected symbolic spaces, the ontology also supports coordinate or point locations, through the class Coordinate-Location. These are subclassed as either a GPS-Coordinate-Location, or a Relative-Coordinate-Location which is defined as a three dimensional displacement from the origin of an Abstract-Space-With-Origin. The origin of such a space is again a Coordinate-Location, being of either GPS or relative variety. This arrangement permits different location sensing schemes to operate at different resolutions, without requiring knowledge of the location system in which it is situated.

Although the ontology offers these coordinate features, it is very hard to perform any inference over them using a normal reasoner, as they are usually not spatially aware. For example, an OWL reasoner cannot determine whether coordinate x,y lies within the bounds of $\{x_1,y_1\},\{x_2,y_2\}, \ldots ,\{x_n,y_n\}$. One potential solution to this which has been discussed is to use an external spatially aware application or service in conjunction with an RDF repository and reasoning engine (see future work).

CONTEXTUAL MODELLING

Central to this activity is the need to represent and collect a wealth of information such as current activities, skills, interests, personal preferences, privacy requirements and the relationships between people. However, representing a generic notion of context is a hugely complex and difficult task. In addition, there is some doubt to the usefulness of such a generic representation, should one be created, as it is likely that it would not afford the level of detail required by many applications. Instead, we put forward the case that activity or task based modelling within a given domain would be more appropriate in many situations.

Activity Based Modelling

Most activities carried out by people are task-oriented, hence the creation of more specific contextual representations is likely to be both easier and more productive. For example, most working days can be conceptualised as a sequence of different tasks, such as a project meeting, document review, teleconference, patient consultation, or student supervision. Each of these tasks may have a variety of important properties, such as meaningful relationships with other people and/or resources, which would be impossible to represent in a generic context ontology.

It is proposed, therefore, that contextually aware systems can be more usefully constructed in a fashion that is tailored to a specific environment, such as that of an academic research department, hospital ward, or lawyers practice, through the application of domain specific ontologies and rules which may be applied to common infrastructure components. This may be achieved through multiple ontologies.

To represent the activities of an individual the concept of a schedule is perceived, containing a reference to each of the tasks or events which that person will be undertaking. For a given domain, it is likely that a high level classification may be made of the typical tasks carried out. These tasks, or events, may be identified as complying with one or more templates, implying that common properties, features and/or resources are required, in a similar fashion to implementing an interface in an object oriented programming language. These high level templates may provide the basis for a hierarchy of more specialised instances of those tasks, defined specifically for the domain in which they are applicable. They may additionally include data such as the location, persons attending, topics of interest, importance and ‘interruptability’ relating to a particular instance of an event or class of events.

As an example, we have looked to the academic domain, focusing on the activities of a senior faculty member. Often having a variety of different research interests, busy schedules which may include significant travelling, and a number of different teaching, research, and supervisory duties, the life of the academic can be hectic and diverse. Throughout a typical day, an academic may switch context on a frequent basis as they attend different meetings and work on different papers or projects, and each of these tasks are likely to address differing research interests, and require a different set of resources to be at hand.

Base-level templates, such as ‘meeting’, ‘document review’, and ‘presenting information’ have been defined, and specialised hierarchically down to such events as a business meeting for project X, or a particular undergraduate lecture course.

An example application

An example application has been created, demonstrating the potential use which can be made from the description of actual events and the templates to which they conform. The client side application has been prototyped through a web browser, as shown in Figure 1. Based around a calendar or diary interface, events in the user’s schedule for the date in question are presented in a list form. Upon selecting a particular task or event, a central repository is queried to find resources that are of interest or relevance to that particular task, such as resources explicitly defined within the template hierarchy, or documents or events which have matching interests or research topics to those prescribed to be of significance to the event.

Back-end support is provided by a Jena-based [6] dynamic repository and inference engine, operating on a number of OWL ontologies in addition a set of custom rules. Developed by the authors, this ‘OwlSrv’ software provides a server which combines storage, inference, and querying capabilities with facilities for content browsing and dynamic updates. Presenting a similar interface and function as 3Store [4], OwlSrv is not as efficient or as massively scalable, yet it provides the crucial additional features of OWL support, inference and dynamic updates.

Testing showed the experimental OWL reasoner, shipped as part of the Jena toolkit, to be rather inefficient when running on the complex ontologies. The main cause is thought to be the overheads introduced in satisfying the subsumption clauses inherent in OWL Full, and as these features are not essential those rules concerned with the propagation of restriction, equivalence and disjointness have been removed. In addition, facts which are instantiated as a result of an inferential consequence are not considered for further inference, unless properties relating to them are explicitly tabled for processing by the backward chaining pass of the hybrid reasoner.

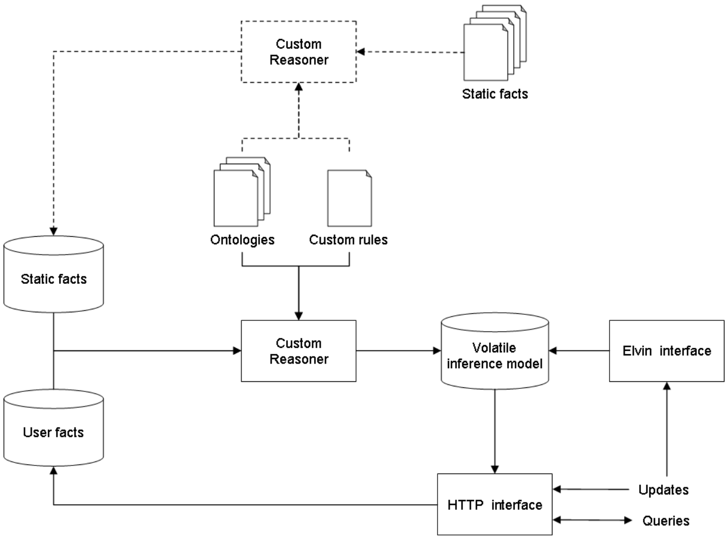

Figure 2: OwlSrv data storage and inference model

Figure 1: Example application

OwlSrv takes this modified set of OWL rules, and combines them with a set of custom inference rules, created in a domain specific manner for each contextually aware environment. These rules may make assertions such as flagging a conference as being of interest to a person if both have matching research interests, or asserting the position of a person if preconditions are satisfied which associate a user to a device and that device to a known location.

Storage model

The Jena toolkit provides support for inference models, created through the application of a reasoner and an ontology to another model containing the data on which the inference is to be performed. Built-in reasoners are provided for RDFS and an experimental forward/backward OWL implementation, both of which are effectively a complex rule set applied to a generic rule reasoner.

The combined rule set is used to create a generic rule reasoner, which is then bound to any number of ontologies, loaded at runtime as determined by a configuration file. The reasoner, following the description of the domain as prescribed by the ontologies, is then available to make inference passes over a model of known facts.

Jena supports models which are backed by an external database, making them both persistent and more efficient than holding all of the data in memory. OwlSrv uses two different persistent models, and a third volatile in-memory model, arranged as in Figure 2.

The model of static facts contains data which changes infrequently, such as details of institutions or businesses, the layout of a building, classifications of research interests, or personal details (name, email address, affiliations, etc). These facts are loaded from disk, the inference engine passed over them, and then the both the initial facts and resultant deductions saved in a database model. This base model then remains as-is in a 'read-only' fashion until such time that the underlying facts require updating, when it is recomputed as an optional stage of the OwlSrv start-up.

The second persistent model contains 'user' facts, which may change on a frequent basis. These may include, for example, schedule information, items of interest, and notes or memos. This model is operated in a 'read-write' mode, permitting dynamic updates to be made to the dataset, and contains only explicitly asserted statements. Deductions resulting from inference over the data are not saved back into the model, permitting facts and any further consequences to be removed safely at a later stage.

As the OwlSrv initialises, the pre-computed static facts are combined with the user facts, and passed through the custom reasoner. It is this combined inference model which is then used as a basis for further analysis or manipulation.

An Elvin [11] interface is used to accept updates of single RDF statements. Elvin is a distributed content-based messaging system facilitating a publish/subscribe model to notifications which consist of a 'bag' of key value pairs. This mechanism is primarily intended for use by a range of sensor inputs which require only simple interactions with the model, such as asserting or retracting the presence of an identifier at a given location. Such updates are never persistent, and will be lost when OwlSrv restarts.

Interpersonal awareness

The importance of *what* task a user is engaged with is also complemented by that of *with whom* they are undertaking it. The relationship between colleagues may be of great use in a contextual system, both in terms of groups of people with similar interests, i.e. communities of practice [1], or in terms of line management hierarchies or organisational status. For example if an academic is engaged in an undergraduate supervision, a brief interruption from another colleague is likely to be acceptable, whereas in the situation where that academic is entertaining a visitor a non-urgent disruption would be most undesirable.

Ontological analysis of data relating to people and events may permit the automatic deduction of importance, though it should be noted that these may be significantly different for the same event when looked at from the perspective of different participants.

Contextual indicators

A number of sources of contextual information may be drawn from the office environment. Most notably, through the combination of suitably rich schedule data and a presence detection system as described earlier, it should be possible to infer in many cases which task a user is currently undertaking. Further data can be assimilated through monitoring of computing devices, such as ascertaining an idle state and/or availability through observing screensavers or instant messaging clients.

FUTURE WORK

The main area of the future work envisaged is to expand on the somewhat simplistic functionality of the example application, and to immerse this and some of the other features discussed into a pervasive environment. Through the combination of pervasive infrastructure, contextual awareness, and the utilisation of semantic representation, data, and services, a next-generation proactive environment should be achievable which exhibits a new level of 'ambient intelligence'.

There are many issues to be resolved in achieving this aim, including the representation of the requirements and applicability of different applications and services at different times and in different situations, the discovery of suitable interfaces for a range of different purposes, and the orchestration of content and applications running on those devices.

Other work has resulted in the creation of a 'device' ontology, permitting the representation of details relating to physical display interfaces, and their capabilities. One can imagine a range of different services, or applications, running within the environment which may assimilate and present information to users, based on their interests and affiliations. Given that applications can be matched with suitable displays, both in terms of their capabilities and location, a dynamic and adaptive presentation mechanism may be formed.

A potential demonstrator currently under discussion for the contextually aware pervasive environment is that of a message delivery system. Applications wishing to deliver notifications to users should be able to present the desired content, along with descriptive metadata, to a system service which will then set about delivering that information to the specified user(s). Messages should be passed to their recipient at an appropriate time and on an appropriate device, according to the current context of the user and the message. As discussed previously, this should take into account the importance of the message, the relationship between the sender and recipient, and the task in which the recipient is currently engaged, to prevent unwarranted interruptions.

A secondary research area lies with the adaptation of content. Techniques for authoring content to enable visual alteration for display on different classes of device have been well researched, though the contextually aware system may additionally provide an opportunity for automatically adapting content depending on the context in which it is presented. Taking the proposed message delivery or notification system as an example, it is likely that a user will wish to be informed of an important missed call, or a change in their schedule. However, if these notifications are displayed in a public setting, then it may be appropriate to restrict the detail of the content of these messages. Additionally, in a scenario in which potential clients are being briefed about a project or product, the context of that meeting is different from when that document is being reviewed by employees only. As a result, it may be deemed that certain aspects of the information are to be automatically withheld in such situations.

On the subject of combining coordinate and symbolic location representation systems, one system which is particularly suited to this type of application is the Multi-User Domain (MUD), which has previously been used to model children moving around woodland in the Ambient Wood project [10]. MUDs are typically made with locations that describe the area they represent, and modern MUD technology is capable of accurately representing real space, where each location can have specific dimensions. Alternatively, a MUD may dispense with strict containment rules, and allow a more freeform method of object visibility and manipulability, similar to the ontologically mediated model described. People and objects may move around inside the MUD, where permitted, and rules can be defined to trigger actions when certain conditions are met.

We propose that by combining the ontology and a MUD engine, a powerful location modelling system can be created. Firstly, the physical properties of an environment are described using the ontology, and stored in a repository. This model can then be imported into a MUD, which acts a location-based event trigger. With a range of sensory inputs, the MUD can position people and devices within the environment, combining the symbolic and coordinate models as required.

There is extensive work on the relationship between toplogical and spatial querying in Geographical Information Systems, and in other projects we are making use of Postgres with GIS extensions [8]. This could also provide a basis for spatial queries. Similar requirements also emerge in Virtual Reality systems and CAD systems.

Rules can be defined to assert certain facts, for example if a group of people gather in a meeting room, then it is fair to imply that a meeting is taking place. Such information, in addition to more trivial arrival and departures, can then be dynamically re-entered into the repository, hence forming a stateful representation of the environment. A common model expressed in RDF and supported by the ontology provides an ideal platform for a query interface, or direct manipulation by other applications.

Although the ontological model used in the proposed location-sensing system allows for the flexible combination of information from a number of different sources, as there is no central analysis of the location data as it is fused together it is quite possible that data arriving from different sensors is inherently conflicting. For example, to a human it is immediately obvious that a person may not be located in two different rooms at the same time, however different sensor subsystems may assert that the same person is in multiple locations, with one system not being aware of the implications of the other. The resolution of such inconsistencies is non-trivial, other than in the case whereby all information relating to that individual is rejected, which although 'safe', is generally not helpful. In addition, the concept of uncertainty over time may be introduced as any of the indicators of position are based around point observations, rather than continuous monitoring. It could be argued that the value or certainty of a location identified by such an observation decreases in some proportion to the time since that reading or observation occurred. Whilst this may complicate the querying of a location, it may prove to be a valuable asset to the problem of conflict resolution. These are all areas for future consideration.

CONCLUSIONS

In this paper we have presented our experimental work in using ontologies to support contextual modelling. Our experiences in building infrastructure components and an exemplary application have been described, the problems we encountered discussed, and we have proposed avenues for our future work.

ACKNOWLEDGMENTS

Part of this work has been supported under the Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC) in conjunction with the Equator IRC, which are sponsored by the UK Engineering and Physical Sciences Research Council under grant numbers GR/N15764/01 and GR/N15986/01 respectively.

REFERENCES

1. Alani, H., Dasmahapatra, S., O'Hara, K. and Shadbolt, N. Identifying Communities of Practice through Ontology Network Analysis. *IEEE Intelligent Systems*, *18* (2). 18-25.
2. Bahl, P. and Padmanabhan, V.N., RADAR: an in-building RF-based user location and tracking system. in *IEEE Conference on Computer Communications (INFOCOM)*, (2000), 2:775-784.
3. Castro, P., Chiu, P., Kremenek, T. and Muntz, R., A probabilistic room location service for wireless networked environments. in *UbiComp 2001*, (2001), 18-34.
4. Harris, S. and Gibbins, N., 3store: Efficient Bulk RDF Storage. in *1st International Workshop on Practical and Scalable Semantic Web Systems*, (Sanibel Island, Florida, USA, 2003).
5. Leonhardt, U. Supporting Location-Awareness in Open Distributed Systems, Imperial College of Science, Technology and Medicine, University of London, 1998.
6. McBride, B., Jena: Implementing the RDF Model and Syntax specification. in *2nd International Semantic Web Workshop*, (2001).
7. McElligott, L., Dillon, M., Leydon, K., Richardson, B., Fernstrom, M. and Paradiso, J.A., 'ForSe fIElds' - Force Sensors for Interactive Environments. in *UbiComp 2002*, (2002), 168-175.
8. Oosterom, P.v. and Vijlbrief, T., Building a GIS on top of the open DBMS Postgres. in *Proceedings of EGIS'91*, (Brussels, Belgium, 1991), 775-787.
9. Priyantha, N.B., Chakraborty, A. and Balakrishnan, H., The Cricket location support system. in *Sixth International Conference on Mobile Computing and Networking (ACM MobiCom)*, (2000), 32-43.
10. Rogers, Y., Price, S., Fitzpatrick, G., Fleck, R., Harris, E., Smith, H., Randell, C., Muller, H., O'Malley, C., Stanton, D., Thompson, M. and Weal, M.J., Ambient Wood: Designing New Forms of Digital Augmentation for Learning Outdoors. in *Third International Conference for Interaction Design and Children (IDC 2004)*, (Maryland, USA, 2004), 1-9.
11. Segall, B., Arnold, D., Boot, J., Henderson, M. and Phelps, T., Content Based Routing with Elvin4. in *AUUG2K*, (Canberra, Australia, 2000).
12. Want, R., Hopper, A., Falcão, V. and Gibbons, J. The Active Badge Location System. *ACM Transactions on Information Systems*, *10* (1). 91-102.
13. Ward, A., Jones, A. and Hopper, A. A New Location Technique for the Active Office. *IEEE Personal Communications*, *4* (5). 42-47.