

I-WEST 2010

José Cordeiro, Dimitris Mitrakos and
Boris Shishkov (Eds.)

Enterprise Systems and Technology

Proceedings of the
4th International Workshop on
Enterprise Systems and Technology
I-WEST 2010

Athens - Greece, July, 2010

SciTePress

Organized by



In cooperation with



José Cordeiro
Dimitris Mitrakos
and Boris Shishkov (Eds.)

Enterprise Systems and Technology

Proceedings of the
4th International Workshop on
Enterprise Systems and Technology
I-WEST 2010

Athens, Greece, July 2010

SciTePress
Portugal

Volume Editors

José Cordeiro (Portugal)
Dimitris Mitrakos (Greece)
and Boris Shishkov (Bulgaria)

4th International Workshop on
Enterprise Systems and Technology
Athens, Greece, July 2010
José Cordeiro, Dimitris Mitrakos and Boris Shishkov (Eds.)

Copyright © 2010
SciTePress
All rights reserved

Printed in Portugal

ISBN: 978-989-8425-44-7
Depósito Legal: 320655/10

Foreword

This volume contains the proceedings of the Fourth International Workshop on Enterprise Systems and Technology (I-WEST 2010), held on July 24-25 in Athens, Greece.

The I-WEST workshop is a scientific event of IICREST, the *Interdisciplinary Institute for Collaboration and Research on Enterprise Systems and Technology*. I-WEST provides a platform to researchers and practitioners, from academia and industry, to discuss challenges, solutions, ideas, and experiences that relate to the broad field of enterprise systems and technology. Each year, a special theme is chosen within this broad field, in order to make presentations and discussions more focused. The theme of I-WEST 2010 is: **Collaborative Environments: Support for Complex Distributed Activities**.

Collaborative environments, representing heavily distributed service-oriented and context-aware ICT facilities, have been proving useful applicability concerning: (i) learning and training within virtual communities; (ii) governmental agencies' cooperation for environment protection and road traffic safety; (iii) social networking. This requires sometimes "externalized" IT systems and services as a potential cost-savings advantage by moving internally hosted IT services to external providers. We need to take a closer look into the main technical foundations and related challenges concerning collaborative environments, taking into consideration also the applicability in domains of high societal relevance. The two-day I-WEST event is to bring together researchers, academics and innovators to discuss the latest collaborative IT innovations intended to improve quality of life, create business opportunities, and stimulate economic growth.

Hence, the contribution of I-WEST 2010 is inspired by the goal of advancing collaborative enterprise technologies and applying them in new ways, to create solutions for the dynamic and networked enterprises of the future. Spontaneous collaboration, context awareness, economic sustainability, and adaptive behavior are just some desired properties of future enterprise systems. Such systems have to be realized in distributed, heterogenous, and dynamic environments. Further, creative solutions grounded in enterprise technologies are needed. I-WEST 2010 also has the intention to unite researchers from various communities, including researchers working on enterprise computing, enterprise interoperability, context-aware systems, IT infrastructures, model-driven development and service-oriented architectures.

Following the I-WEST'10 Call for Papers and received submissions, 10 papers were selected for a 30-minutes oral presentation during the workshop and for publication in these proceedings. The selected papers are a good illustration of different relevant topics that are currently under research: some papers are more oriented towards enterprise technology (considering it from the perspectives of data technology and information storage/retrieval) while other papers are directed more to the development of ICT applications related to this technology (considering human-actions-driven application specification, service-oriented application models, and some related computing paradigms, such as Autonomic Computing and Context-Aware Computing) as well as to innovative business models, enterprise information systems, and "externalization" of IT.

Taking this opportunity, we would like to express our gratitude to all who have contributed to I-WEST 2010, including the authors (who have provided the main content for these proceedings) as well as the program committee members (who have provided constructive comments contributing to the content quality). We would also like to thank Vitor Pedrosa for the brilliant work and support in preparing the proceedings. Last but not least, we thank the invited speaker, Joaquim Filipe, for his invaluable contribution. Finally, we tremendously appreciate the willingness of SciTePress to publish the proceedings, expressing special gratitude to INSTICC.

We wish all presenters and attendees an inspiring workshop and a nice stay in the beautiful city of Athens.

July 2010

José Cordeiro
Dimitris Mitrakos
Boris Shishkov

Workshop Chairs

José Cordeiro
Polytechnic Institute of Setúbal / INSTICC / IICREST, Portugal
Dimitris Mitrakos
Aristotle University of Thessaloniki / IICREST, Greece
Boris Shishkov
IICREST, Bulgaria

Program Committee

Dumitru Dan Burdescu, University of Craiova, Romania
Kuo-Ming Chao, Coventry University, United Kingdom
Samuel Chong, Capgemini, United Kingdom
Jorge Cardoso, University of Coimbra, Portugal
Joaquim Filipe, Polyt. Inst. of Setúbal/INSTICC/IICREST, Portugal
Alexandre Girardi, Multitel ASBL, Belgium
Markus Helfert, Dublin City University, Ireland
Ilian Ilkov, IBM Nederland B.V., The Netherlands
Ivan Ivanov, SUNY Empire State College, United States
Dimitri Konstantas, University of Geneva / IICREST, Switzerland
Kecheng Liu, University of Reading, United Kingdom
Bart Nieuwenhuis, University of Twente / ExSer, The Netherlands
Erik Proper, Radboud University Nijmegen / Capgemini, The Netherlands
Dick Quartel, Novay, The Netherlands
Manfred Reichert, University of Ulm, Germany
Marten van Sinderen, University of Twente / IICREST, The Netherlands
Alexander Verbraeck, Delft University of Technology, The Netherlands

Supporting Organizations

CTIT - Centre for Telematics and Information Technology
(The Netherlands)

INSTICC - Institute for Systems and Technologies of Information,
Control and Communication (Portugal)

Table of Contents

Foreword.....	iii
Workshop Chairs	v
Program Committee	v
Supporting Organizations.....	vi

Invited Speaker

Multi-Agent Systems: Theory and Application in Organization Modelling	3
<i>Joaquim Filipe</i>	

Papers

NOMIS: A Human Centred Modelling Approach of Information Systems	17
<i>José Cordeiro, Joaquim Filipe and Kecheng Liu</i>	
Methodological Support for the Design of Enterprise Information Systems with SDBS: Towards Distributed, Service-Oriented, and Context-Aware Solutions	36
<i>Boris Shishkov</i>	
Service-Oriented Architecture for Household Energy Management <i>Marten van Sinderen, Martijn Warnier and Boris Shishkov</i>	51
Hovering Information: A Self-Organising Infrastructure-Free Information Storage and Retrieval Service	69
<i>Alfredo A. Villalba Castro, Giovanna Di Marzo Serugendo and Dimitri Konstantas</i>	

Business Modeling and Value Network Design Case Study for a Tele-Rehabilitation Service	88
<i>Lambert J. M. Nieuwenhuis</i>	
Design of a Service-Oriented e-Learning Environment	98
<i>Dumitru Dan Burdescu, Marian Cristian Mibaescu, Costel Marian Ionascu and Bogdan Logofatu</i>	
Model Driven Service-Oriented Approach for Smart Building Energy Management.....	116
<i>Ding-Yuan Cheng, Nasaraf Shah, Chen-Fang Tsai, Kuo-Ming Chao and Chi-Chun Lo</i>	
Adaptive Knowledge Representation for a Self-Managing Home Energy Usage System	131
<i>Martijn Warnier, Marten van Sinderen and Frances M. T. Brazier</i>	
A Multimedia Database Server for Content-Based Retrieval and Image Mining.....	141
<i>Cosmin Stoica Spahiu, Liana Stanescu and Dumitru Dan Burdescu</i>	
Contextualising Information Quality: A Method-Based Approach ..	149
<i>Markus Helfert and Owen Foley</i>	
Author Index	163

**INVITED
SPEAKER**

Multi-Agent Systems: Theory and Application in Organization Modelling

Joaquim Filipe

Escola Superior de Tecnologia do Instituto Politécnico de Setúbal
Rua Vale de Chaves, Estefanilha, 2910 Setúbal, Portugal
joaquim.filipe@estsetubal.ips.pt

1 Introduction

Organisations are multi-agent systems, eventually including both human and artificial agents. Organisations are also seen as multilayered Information Systems (IS) themselves, including an informal subsystem, a formal subsystem and a technical system as shown in figure 1.

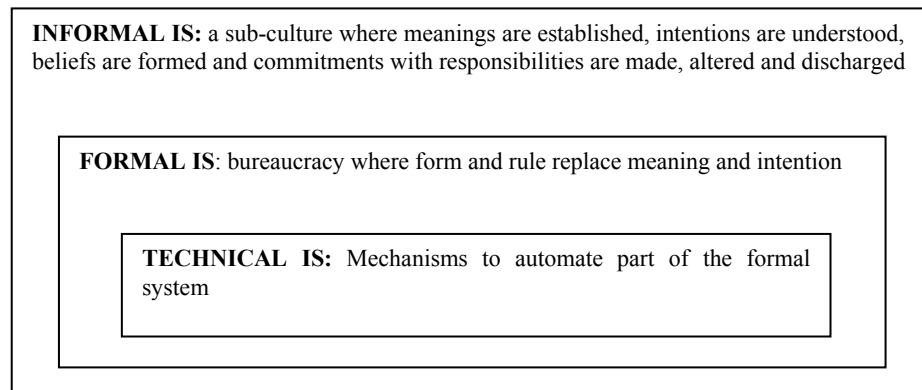


Fig. 1. Three main layers of the real information system [10].

We aim at improving the technical subsystem within the constraints defined by the other two. Organisational information systems are inherently distributed, nowadays, thus communication and coordination are major problems in this kind of information systems. Perhaps motivated by the difficult problems there is currently a strong interest on this area, which is an active research field for several disciplines, including Distributed Artificial Intelligence (DAI), Organisational Semiotics and the Language-Action Perspective, among others. Our approach integrates elements from these three perspectives.

The Epistemic-Deontic-Axiologic (EDA) designation refers to the three main components of the agent structure described in this paper. Here we propose an agent model which, contrary to most DAI proposals, not only accounts for intentionality but is also prepared for social interaction in a multi-agent setting. Existing agent models emphasise an intentional notion of agency – the supposition that agents should be

understood primarily in terms of mental concepts such as beliefs, desires and intentions. The BDI model [9] is a paradigm of this kind of agent.

We claim that the cognitive notions that constitute the basis of intentional models show only one face of the coin, the other one being social agency. It is required that an adequate agent model, able to function in a multi-agent setting, should emphasise social agency and co-ordination, within a semiotics framework. Our approach focuses on organisational agents who participate in organisational processes involving the creation and exchange of signs, *i.e.* knowledge sharing, in a way that is inherently public, thus depending on the agent's social context, *i.e.* its information field [10], to support co-ordinated behaviour. An information field is composed by all the agents and objects involved in a certain activity and sharing a common ontology.

A realistic social model must be normative: both human agents and correctly designed artificial agents ought to comply with the different kinds of norms that define the information field where they operate, although exceptions may occur. Private representations of this shared, normative, knowledge are translations of the public knowledge into specific agent mental structures. When designing an artificial agent, the designer must adopt a private knowledge representation paradigm to set up the agent's internal knowledge in such a way that it fits the normative shared ontology.

We postulate that norms are the basic building blocks upon which it is possible to build co-ordination among organised entities and co-ordinated actions are the crux of organised behaviour. We claim that although organised activity is possible either with or without communication, it is not possible without a shared set of norms. Therefore, these socially shared norms define an information field that is a necessary condition for heterogeneous multi-agent co-ordination including both artificial agents and humans.

2 The Normative Structure of the EDA Model

Norms are typically social phenomena. This is not only because they stem from the existence of some community but also because norms are multi-agent objects [5]:

- They concern more than one individual (the information field involves a community)
- They express someone's desires and assign tasks to someone else
- Norms may be regarded from different points of view, deriving from their social circulation: in each case a norm plays a different cognitive role, be it a simple belief, a goal, a value, or something else.

Social psychology provides a well-known classification of norms, partitioning them into perceptual, evaluative, cognitive and behavioural norms. These four types of norms are associated with four distinct attitudes, respectively [10]:

- Ontological – to acknowledge the existence of something;
- Axiologic – to be disposed in favour or against something in value terms;
- Epistemic – to adopt a degree of belief or disbelief;
- Deontic – to be disposed to act in some way.

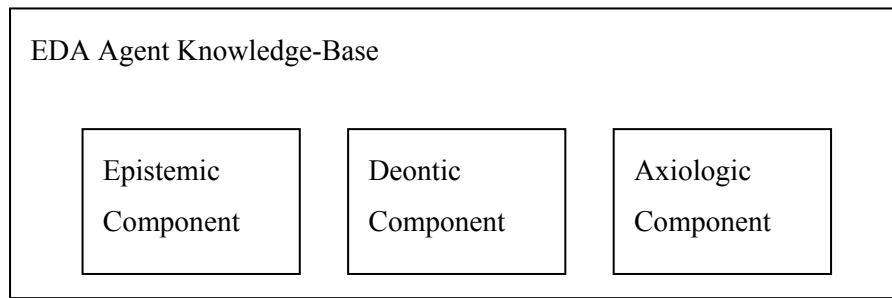


Fig. 2. The EDA Agent Knowledge-Base Structure.

An EDA agent is a knowledge-based system whose knowledge base structure is based on the following three components: the Epistemic, the Deontic and the Axiologic.

The epistemic model component is where the knowledge of the agent is stored, in the form of statements that are accepted by that agent. Two types of knowledge are stored here: declarative knowledge – statements relative to the agent beliefs – and procedural knowledge – statements concerning the know-how of the agent, *e.g.* their plans and procedural abilities.

The importance of norms to action has determined the name we have chosen for the model component where the agent goals are represented. An agent goal may be simply understood as the *desire to perform an action* (which would motivate the designation of *conative*) but it can also be understood, especially in a social context, as the *result of the internalisation of a duty or social obligation* (which would motivate the designation of *deontic*). We have adopted the latter designation not only because we want to emphasise the importance of social obligations but also because personal desires can be seen as a form of ‘generalised’ obligation established by an agent for himself. This provides a *unification* of social and individual drives for action, which simplifies many aspects of the model.

The axiologic model component contains the value system of the agent, namely a partial order that defines the agent preferences with respect to norms. The importance of the agent value system is apparent in situations of conflict, when it is necessary to violate a norm. This preference ordering is dynamic, in the sense that it may change whenever the other internal components of the agent model change, reflecting different beliefs or different goals.

3 Intentions and Social Norms in the EDA Model

The multi-agent system metaphor that we have adopted for modelling organisations implies that organisations are seen as goal-governed collective agents, which are composed of individual agents.

In our model individual agents are autonomous, heterogeneous, rational, social agents. Therefore, they are compelled to make decisions and act in a way that, although not entirely deterministic, is constrained by their rationality. Actually their behaviour would be predictable if we knew all the details of their EDA model

components, the environment stimuli, their perception function and also the reasoning machine they use, because the ultimate goal of a rational agent is to maximize its utility.

Typically, artificial intelligence (AI) agent models consider intelligent agent decision processes as internal processes that occur in the mind and involve exclusively logical reasoning, external inputs being essentially data that are perceived directly by the agent. This perspective does not acknowledge any social environment whatsoever. In this paper we start from a totally different perspective, by emphasising the importance of social influences and a shared ontology on the agent decision processes, which then determines agents' activity. We shall henceforth refer to agent as 'it' although the EDA model also applies to human agents. In any case, we are particularly interested in the situations where information systems are formally described, thus making it possible for artificial agents to assist or replace human agents.

An important role of norms in agent decision processes is the so-called cognitive economy: by following norms the agent does not have to follow long reasoning chains to calculate utilities – it just needs to follow the norms.

However, instead of adopting a whole-hearted social sciences perspective, which is often concerned merely with a macro perspective and a statistical view of social activity, we have adopted an intermediate perspective, where social notions are introduced to complement the individualistic traditional AI decision models: a psycho-social perspective, whereby an agent is endowed with the capability of overriding social norms by intentionally deciding so.

Our model enables the relationship between socially shared beliefs with agent individual, private, beliefs; it also enables the analysis of the mutual relationships between moral values at the social level with ethical values at the individual level. However, we have found particularly interesting analogies in the deontic component, specifically in the nature of the entities and processes that are involved in agent goal-directed behaviour, by inspecting and comparing both the social processes and individual processes enacted in the deontic component of the EDA model.

This was motivated by the close relationship between deontic concepts and agency concepts, and represents a direction of research that studies agency in terms of normative social concepts: obligations, responsibilities, commitments and duties. These concepts, together with the concepts of power/influence, contribute to facilitate the creation of organisational models, and are compatible with a vision of organisations as normative information systems as well as with the notion of information field that underlies the organisational semiotics approach, on which the work presented in this paper is inspired.

As will be explained in more detail in the next section, an essential aspect of the EDA model is that the Deontic component is based on the notion of *generalised goal* as a kind of obligation, that encompasses both social goals (social obligations) and individual goals (self obligations). Following a traditional designation in DAI, we designate those individual *generalised goals* that are inserted in the agenda as *achievement goals*, as in [4]. Figure 3 describes the parallelism between mental and social constructs that lead to setting a goal in the agenda, and which justifies the adoption of the aforementioned generalised obligation. Here, p represents a proposition (world state). $B_\alpha(p)$ represents p as one of agent α 's beliefs. $O_\alpha^\beta(p)$

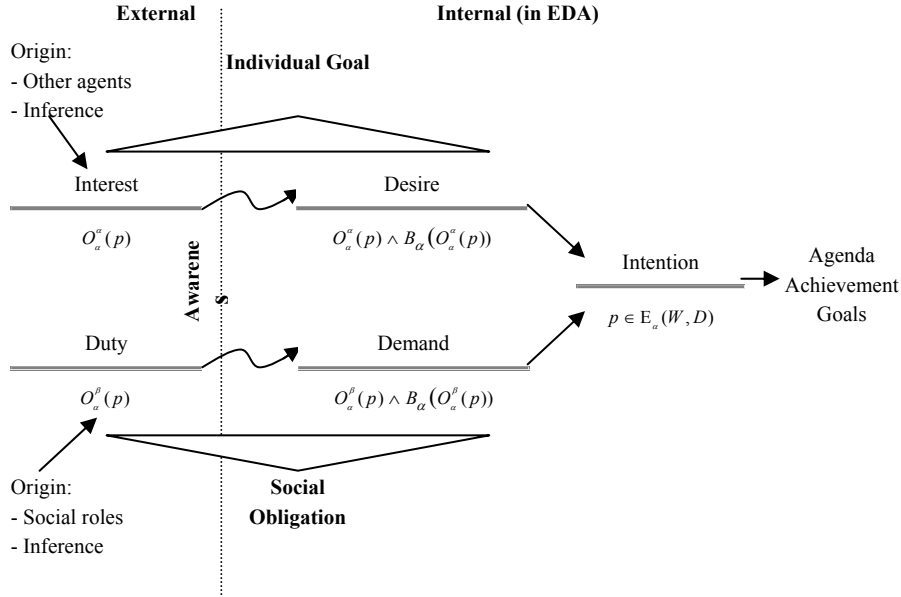


Fig. 3. Social and Individual goals parallelism in the EDA model.

represents the obligation that α must see to it that p is true for β . $O_\alpha^\alpha(p)$ represents the interest that α has on seeing to it that p is true for itself – a kind of self-imposed obligation. In this diagram $p \in E_\alpha(W, D)$ means, intuitively, that proposition p is one of the goals on α 's agenda.

Interest is one of the key notions that are represented in the EDA model, based on the combination of the deontic operator 'ought-to-be' [14] and the agentic 'see-to-it-that' *stit* operator [1]. *Interests* and *Desires* are manifestations of *Individual Goals*. The differences between them are the following:

- *Interests* are individual goals of which the agent is not necessarily aware, typically at a high abstraction level, which would contribute to improve its overall utility. Interests may be originated externally, by other agents' suggestions, or internally, by inference: deductively (means-end analysis), inductively or abductively. One of the most difficult tasks for an agent is to become aware of its interest areas because there are too many potentially advantageous world states, making the full utility evaluation of each potential interest impossible, given the limited reasoning capacity of any agent.
- *Desires* are interests that the agent is aware of. However, they may not be achievable and may even conflict with other agent goals; the logical translation indicated in the figure, $O_\alpha^\alpha(p) \wedge B_\alpha(O_\alpha^\alpha(p))$, means that desires are goals that agent α ought to pursue for itself and that it is aware of. However, the agent has not yet decided to commit to it, in a global perspective, *i.e.* considering all other possibilities. In other words, desires

become intentions only if they are part of the preferred extension of the normative agent EDA model [7].

It is important to point out the strong connection between these deontic concepts and the axiologic component. All notions indicated in the figure should be interpreted from the agent perspective, *i.e.* values assigned to *interests* are determined by the agent. Eventually, external agents may consider some goal (*interest*) as having a positive value for the agent and yet the agent himself may decide otherwise. That is why *interests* are considered here to be the set of all goals to which the agent would assign a positive utility, but which it may not be aware of. In that case the responsibility for the *interest* remains on the external agent.

Not all interests become desires but all desires are agent interests. This may seem contradictory with a situation commonly seen in human societies of agents acting in *others' best interests*, sometimes even against their desires: that's what parents do for their children. However, this does not mean that the agent desires are not seen as positive by the agent; it only shows that the agent may have a deficient axiologic system (by its information field standards) and in that case the social group may give other agents the right to override that agent. In the case of artificial agents such a discrepancy would typically cause the agent to be banned from the information field (no access to social resources) and eventually repaired or discontinued by human supervisors, due to social pressure (*e.g.* software viruses).

In parallel with *Interests* and *Desires*, there are also social driving forces converging to influence individual achievement goals, but through a different path, based on the general notion of social obligation. Social obligations are the goals that the social group where the agent is situated require the agent to attain. These can also have different flavours in parallel to what we have described for individual goals.

- *Duties* are social goals that are attached to the particular roles that the agent is assigned to, whether the agent is aware that they exist or not. The statement $O_{\alpha}^{\beta}(p)$ means that agent α ought to do p on behalf of another agent β . Agent β may be another individual agent or a collective agent, such as the society to which α belongs. Besides the obligations that are explicitly indicated in social roles, there are additional implicit obligations. These are inferred from conditional social norms and typically depend on circumstances. Additionally, all specific commitments that the agent may agree to enter also become duties; however, in this case, the agent is necessarily aware of them.
- *Demands* are duties that the agent is aware of¹. This notion is formalised by the following logical statement: $O_{\alpha}^{\beta}(p) \wedge B_{\alpha}(O_{\alpha}^{\beta}(p))$. Social demands motivate the agent to act but they may not be achievable and may even conflict with other agent duties; being autonomous, the agent may also decide that, according to circumstances, it is better not to fulfil a social demand and rather accept the corresponding sanction. Demands become intentions only if they are part of the preferred extension of the normative agent EDA model – see [7] section 5.7 for details.

¹ According to the Concise Oxford Dictionary, *demand* is “*an insistent and peremptory request, made as of right*”. We believe this is the English word with the closest semantics to what we need.

- *Intentions*: Whatever their origin (individual or social) intentions constitute a non-conflicting set of goals that are believed to offer the highest possible value for the concerned agent. Intentions are designated by some authors [11] as psychological commitments (to act). However, intentions may eventually (despite the agent sincerity) not actually be placed in the agenda, for several reasons:
 - They may be too abstract to become directly executed, thus requiring further means-end analysis and planning.
 - They may need to wait for their appropriate time of execution.
 - They may be overridden by higher priority intentions.
 - Required resources may not be ready.

The semantics of the prescriptive notions described above may be partially captured using set relationships as depicted in figure 4, below.

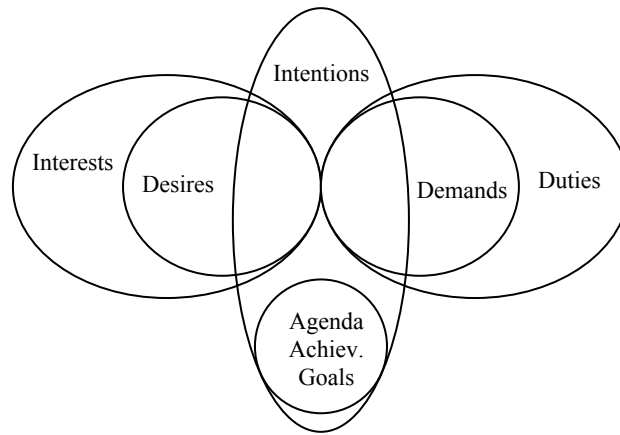


Fig. 4. Set-theoretic relationships among deontic prescriptive concepts.

When an agent decides to act in order to fulfil an intention, an agenda item is created – we adopt the designation of *achievement goal*. Achievement goals are defined as in [4] as goals that are shared by individuals participating in a team that has a *joint persistent goal*. Following the terminology of [4] agent α has a *weak achievement goal*, relative to its motivation (which in our case corresponds to the origin and perceived utility of that goal), to bring about the joint persistent goal γ if either of the following is true:

- α does not yet believe that γ is true and has γ being eventually true as a goal (*i.e.* α has a normal achievement goal to bring about γ)
- α believes that γ is true, will never be true or is irrelevant (utility below the motivation threshold), but has a goal that the status of γ be mutually believed by all team members.

However we do not adopt the notion of joint persistent goal for social coordination, as proposed by Cohen and Levesque [4] because their approach has a number of shortcomings, not only theoretical but also related to the practical feasibility of their model, which are well documented in [12].

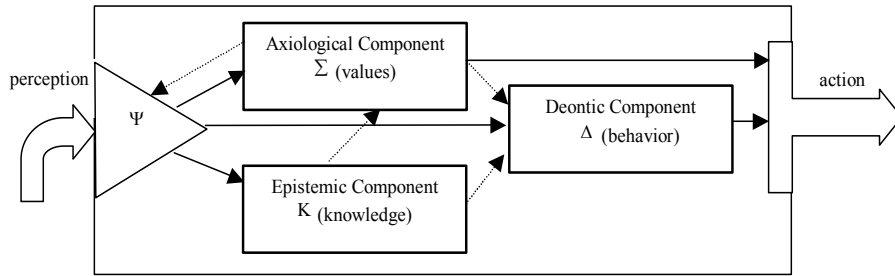


Fig. 5. The EDA model component relationships.

4 The EDA Model Internal Architecture

Using the social psychology taxonomy of norms, and based on the assumption that organisational agents' behaviour is determined by the evaluation of deontic norms, given the agent epistemic state, with axiological norms for solving eventual interest conflicts, we propose an intentional agent model, which is decomposed into three main components: the epistemic, the deontic and the axiologic. Additionally there are two external interfaces: an input (perceptual) interface, through which the agent receives and pragmatically interprets messages from the environment and an output (acting) external interface through which the agent acts upon the environment, namely sending messages to other agents².

A socially shared ontology is partially incorporated in an agent cognitive model whenever it is needed, *i.e.* when the agent needs to perform a particular role. In this case, beliefs are incorporated in the Epistemic component, obligations and responsibilities are incorporated in the Deontic component and values (using a partial order relation of importance) are incorporated in the Axiologic component – all indexed to the particular role that the agent is to play.

Figure 5 depicts the EDA model and its component relationships.

Ψ is a pragmatic function that filters perceptions, according to the agent ontology, using perceptual and axiologic norms, and updates one or more model components.

Σ is an axiologic function that is used mainly in two circumstances: to help decide which signs to perceive, and to help decide which goals to put in the agenda and execute.

K is a knowledge-based component, where the agent stores its beliefs both explicitly and implicitly, in the form of potential deductions based on logical reasoning.

Δ is a set of plans, either explicit or implicit, the agent is interested in and may choose to execute.

The detailed description of each component, including its internal structure, is provided in [7]. In this paper we focus on the system behaviour. The next sections

² In this paper we restrict our attention to the semiotic, symbolic, types of agent activity, ignoring substantive physical activities.

show how in EDA we specify ideal patterns of behaviour and also how we represent and deal with non-ideal behaviours.

5 Organisational Modelling with Multi-Agent Systems using the EDA Model

The EDA model may apply to both human and artificial agents, and is concerned with the social nature of organisational agents:

- Firstly, because it accounts for a particular mental structure (Epistemic-Deontic-Axiologic) that is better, for our purposes, than other agent mental structures proposed in the literature to model agent interaction. Specifically, we intend to use it for modelling information fields where social norms influence individual agents and are used by an agent to guide inter-subjective communication and achieve multi-agent co-ordination.
- Secondly, because the model is based on normative notions that are not only intended to guide the agent behaviour in a socially accepted way, but also to identify what sanctions to expect from norm violations, in order to let the agent take decisions about its goals and actions, especially when co-ordination is involved. The EDA model is based on the claim that multi-agent notions such as social commitment, joint intentions, teamwork, negotiation and social roles, would be merely metaphorical if their normative character were not accounted for.

Given its social-enabled nature, the EDA agent notion may be used to model and implement social activities, involving multi-agent co-ordination. However, although the agent paradigm described in this paper is suited to model team work and joint problem solving, the major novelty with respect to other current agent models is the normative flavour. EDA agents are able to co-ordinate on the basis of shared norms and social commitments. Shared norms are used both for planning and for reasoning about what is expected from other agents but, internally, EDA agents keep an intentional representation of their own goals, beliefs and values.

Co-ordination is based on commitments to execute requested services. Commitments are represented not as joint intentions based on mutual beliefs, as is the case of the Cohen-Levesque model, upon which the BDI paradigm is based, but as first-class social concepts, at an inter-subjective level. The organisational memory embedded in the representation of socially accepted best-practices or patterns of behaviour and the handling of sub-ideal situations is, in our opinion, one of the main contributions that a multi-agent system can bring about.

6 Representing Ideal and non-Ideal Patterns of Behaviour

The EDA model is a norm-based model. Norms are ultimately an external source of action control. This assumption is reflected specially in the Deontic component of the EDA model.

Standard Deontic Logic (SDL) represents and reasons about ideal situations only. However, although agent behaviour is guided by deontic guidelines, in reality an agent who always behaves in an ideal way is seldom seen. The need to overcome the limited expressiveness of SDL, and to provide a way to represent and manipulate sub-ideal states has been acknowledged and important work has been done in that direction, *e.g.* by Dignum *et al.* [6] and also by Carmo and Jones [2].

Contrary to SDL, the Deontic component of the EDA model is designed to handle sub-ideal situations. Even in non-ideal worlds, where conflicting interests and obligations co-exist, we wish to be able to reason about agent interests and desirable world states in such a way that the agent still is able to function coherently.

Behaviours may be represented as partial plans at different abstract levels. A goal is a very high abstract plan, whereas a sequence of elementary actions defines a plan at the instance level. The Deontic component is similar, in this sense, to what Werner [13] called the agent intentional state.

However, in our model, agent decisions depend both on the available plans and a preference relationship defined in the axiologic component. This value assignment, which is essential for determining agent intentions, *i.e.* its preferred actions, can change dynamically, either due to external events (perception) or to internal events (inference), thus dynamically modifying the agent's intentions.

Although our representation of ideal behaviours is based in deontic logic, we acknowledge the existence of problems with deontic logic, partially caused by the fact that the modal 'ought' operator actually collapses two operators with different meanings, namely 'ought-to-do' and 'ought-to-be'. Our solution, inspired in [8] and [1], is to use a combination of action logic and deontic logic for representing agentive 'ought-to-do' statements, leaving the standard deontic operator for propositional, declarative, statements. Agentive statements are represented as $[\alpha \textit{ stit} : Q]$ where α stands for an agent and Q stands for any kind of sentence (declarative or agentive). An 'ought-to-do' is represented using the conventional 'ought-to-be' modal operator combined with an agentive statement, yielding statements of the form $O[\alpha \textit{ stit} : Q]$ or, for short: $O_\alpha(Q)$.

The representation of behaviours using this kind of agentive statement has several attractive properties, including the fact that it is a declarative representation (with all the flexibility that it provides) and that there exists the possibility of nesting plans as nested *stit*s.

A plan is typically given to the agent at a very abstract level, by specifying the goal that it ought to achieve. The agent should then be able to decompose it into simple, executable, actions. This decomposition can be achieved by a means-ends process. By representing plans declaratively, as behavioural norms, the process becomes similar to backward chaining reasoning from abstract goals to more specific ones, until executable tasks are identified, the same way as in goal oriented reasoning in knowledge-based systems. This similarity enables the adoption of methods and tools from that area, being especially useful the inference engine concept.

However, the Deontic component does not control the Agenda directly, *i.e.* it is not responsible for setting the agent goals directly, because the prospective agent goals – similar to *desires*, in the BDI model – must be analysed by the Axiologic

component first, which computes their value accordingly to an internal preference relation, taking into account possible obligation violations.

7 Conclusions

The EDA model is a norm-based, theoretically sound, agent model that takes into account not only the intentional aspects of agency but also the social norms that prescribe and proscribe certain agent patterns of behaviour. The main components of this model (Epistemic, Deontic and Axiological) have a direct relationship with the types of norms that are proposed in the social psychology theory supporting the model. In this paper, however, we focused our attention essentially in the Deontic component, where the normative social aspects are more important, namely where ideal and sub-ideal behaviours are represented.

We consider agents to be goal-governed systems. All agent goals can be represented as obligations, encompassing both agent self-imposed obligations and social obligations – derived from moral obligations or commitments established in the course of their social activity.

Organizations can be seen as multi-agent systems with the EDA internal architecture based on deontic agency notions making it easier to understand some social aspects that in other agent models can only be modelled indirectly through joint-goals.

References

1. Belnap, N., 1991. Backwards and Forwards in the Modal Logic of Agency. *Philosophy and Phenomenological Research*, vol. 51.
2. Carmo, J. and A. Jones, 1998. Deontic Logic and Contrary to Duties. Personal copy of a pre-print manuscript.
3. Castelfranchi, C., 1993. Commitments: from Individual Intentions to Groups and Organisations. In *Working Notes of AAAI'93 Workshop on AI and Theory of Groups and Organisations: Conceptual and Empirical Research*, pp.35-41.
4. Cohen, P. and H. Levesque, 1990. Intention is Choice with Commitment. *Artificial Intelligence*, 42:213-261.
5. Conte, R., and C. Castelfranchi, 1995. *Cognitive and Social Action*, UCL Press, London.
6. Dignum, F., J-J Ch. Meyer and R. Wieringa, 1994. A Dynamic Logic for Reasoning About Sub-Ideal States. In J. Breuker (Ed.), *ECAI Workshop on Artificial Normative Reasoning*, pp.79-92, Amsterdam, The Netherlands.
7. Filipe, J., 2000. *Normative Organisational Modelling using Intelligent Multi-Agent Systems*. Ph.D. thesis, University of Staffordshire, UK.
8. Hilpinen, R., 1971. *Deontic Logic: Introductory and Systematic Readings*. Reidel Publishing Co., Dordrecht, The Netherlands.
9. Rao, A. and M. Georgeff, 1991. Modeling Rational Agents within a BDI architecture. *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (Nebel, Rich, and Swartout, Eds.)* pp.439-449. Morgan Kaufman, San Mateo, USA.

10. Stamper, R., 1996. Signs, Information, Norms and Systems. In Holmqvist et al. (Eds.), Signs of Work, Semiosis and Information Processing in Organizations, Walter de Gruyter, Berlin, New York.
11. Singh, M., 1990. Towards a Theory of Situated Know-How. 9th European Conference on Artificial Intelligence.
12. Singh, M., 1996. Multiagent Systems as Spheres of Commitment. Proceedings of the International Conference on Multiagent Systems (ICMAS) - Workshop on Norms, Obligations and Conventions. Kyoto, Japan.
13. Werner, E., 1989. Cooperating Agents: A Unified Theory of Communication and Social Structure. In Gasser and Huhns (Eds.), Distributed Artificial Intelligence, pp.3-36, Morgan Kaufman, San Mateo, USA.
14. von Wright, G., 1951. Deontic Logic. *Mind*, 60, pp.1-15.

PAPERS

NOMIS: A Human Centred Modelling Approach of Information Systems

José Cordeiro¹, Joaquim Filipe¹ and Kecheng Liu²

¹EST Setúbal/IPS, Rua do Vale de Chaves, Estefanilha, 2910-761 Setúbal, Portugal

²The University of Reading, Whiteknights, Reading, RG6 6AF, U.K.
j.cordeiro@computer.org, joaquim.filipe@estsetubal.ips.pt
k.liu@reading.ac.uk

Abstract. Current approaches to information systems (IS) still ignore its human nature resulting too often in technical IS project failures. One of the reasons seems to be their scientific and technical bias following a philosophical objectivist stance. One way to overcome this narrow view is to use a paradigm that takes properly into account the human element, his behaviour and his distinct characteristics within organisational and business domains. This is the purpose of Human Relativism (HR) a new philosophical stance centred in the human element and focused in the externalisation of his behaviour namely *human actions*. In order to introduce, describe and propose a suitable approach to IS development following HR principles this paper introduces NOMIS, a normative modelling approach for IS. NOMIS integrates the key notions and views of three IS socio-technical approaches, namely Organisational Semiotics, the Theory of Organised Activity and Enterprise Ontology in a coherent and consistent human-centred modelling approach. The new vision provided by NOMIS is intended to furnish a more realistic, comprehensive, and concise modelling of IS reality. In this paper some examples of this vision expressed by NOMIS views and some of its specific representation diagrams will be shown applied to an empirical case study of a library system.

Keywords. Information systems, Information systems development approaches, Information Systems modelling, Human-centred information systems, Human relativism, Organisational semiotics, Theory of organized activity, Enterprise ontology, NOMIS.

1 Introduction

Current approaches to information systems (IS) still ignore its human nature resulting too often in technical IS projects failure. One of the reasons seems to be their scientific and technical bias following a philosophical objectivist stance. To overcome the problems felt by these *hard approaches* a group of *soft approaches* have been proposed. These *soft approaches* usually follow a social constructivist philosophical stance that sees the world as socially constructed by individuals and groups where the social and human aspects take the leading role. Although promising to solve the issues of *hard approaches*, *soft approaches* did not succeeded as it can be deduced by its low adoption rate. Therefore, we think that a possible solution will be to have good

and accurate theoretical principles, a modelling representation expressing human and business reality according to those principles and a technical design and implementation produced from those models. First, and in order to provide the theoretical principles that take properly into account the human element and its behaviour, in [1] a new philosophical stance – Human Relativism (HR) – was proposed together with an analysis of *human action* seen as the kernel element of any approach following this stance. In this paper we use and extend the HR view by introducing a new modelling approach for IS – NOMIS – that integrates the key notions and views of three socio-technical approaches to IS, namely Organisational Semiotics [2], the Theory of Organised Activity [3] and Enterprise Ontology [4] in a coherent and consistent human-centred modelling approach built around the *human action* element. The new vision provided by NOMIS is intended to furnish a more realistic, comprehensive, and concise modelling of IS reality. Also, in this paper some examples of this vision expressed by NOMIS views and some of its specific representation diagrams will be shown applied to an empirical case study of a library system.

This paper is organised as follows: section 2 briefly presents the theoretical foundation of NOMIS, namely HR and NOMIS underlying theories. Section 3 makes a brief analysis and comparison of these theories and presents the integration principles adopted by NOMIS and its proposed modelling views. In section 4, a library system case study is used to illustrate the application of NOMIS, in particular some diagrams provided by each of its views will be presented. Section 5 presents some related work regarding other attempts to integrate the foundation theories of NOMIS and, finally, section 6 present the conclusions, work done and future work.

2 Theoretical Foundations of NOMIS

2.1 Human Relativism

Human Relativism (HR) [1] is a simple philosophical stance that recognises reality relative to humans without adhering to a subjectivist view of the world. This view is particularly important in businesses and organisations where the key elements are humans that have different understanding, experience, and knowledge of organisational reality. Also information expressed by language and used to represent and communicate this reality has a fundamental role in this context. Again, information is dependent on individuals and their perceptions, interpretations, knowledge, judgment, etc being impossible to formulate and state precisely. Thus, HR recognises and highlights the central role played by humans by acknowledging an *objective reality* as *human relative* and proposes two concepts – *observability* and *precision* – to deal with the unpredictability and inaccuracy factors introduced. In order to understand the first concept – *observability* – is necessary to distinguish between *perception*, the process of acknowledging the external reality through our senses, and *interpretation*, the meaning making process. Only information goes through the interpretation process, all other elements of human reality are just perceived. This restricts what is perceived by humans and, consequently, what is understood as *observable*. *Observable things* can be viewed as material or physical

individual things from the objectivist point of view. As an example a particular house is *observable*, but the concept of a house, an universal, is *not observable*. From this perspective HR makes the following assumption:

Anything that is observable will be more consensual, precise and, therefore more appropriate to be used by scientific methods.

In practice *observability* intends to remove ambiguities from human reality and to achieve the necessary precision needed to apply and use scientific methods. Besides using *observability* it is also possible to remove ambiguities by having a high degree of *precision* in any element of human reality. This second concept of *precision* in HR seeks to deal with this matter. To have a *high degree of precision* means to have a reduced level of ambiguity and different meanings in some term or element making it generally accepted, recognised and shared. One way of achieving precision, for example, is the use of physical measurement. It is simple to say, by using an appropriate instrument, if a specific string has or has not one meter of length, making it not so dependent on individuals.

From what have been said, an important Human Relativistic hypothesis is:

By adopting observable elements or high precision elements under a human relativistic view it is possible to derive a scientific and theoretical well founded approach to IS.

These simple ideas proposed by HR are, in fact, aligned with social constructivism and objectivism making a proper connection between them.

From an IS development viewpoint most of the problems felt with hard and soft approaches have its origins in the human element. This element introduces, in many different ways, an unpredictability factor that prevents the use of scientific and objective methods. HR identifies and highlights this point by recognising human behaviour, in general, as the source of IS development problems. This allows for using different approaches when human behaviour is or is not present. Therefore, another HR hypothesis and conclusion is:

We may freely apply technical approaches if there is no unpredictable behaviour present, specifically human behaviour.

2.2 Organisational Semiotics

One of theories supporting NOMIS is Organisational Semiotics (OS) [2] which was introduced by Ronald Stamper and his teams. OS perspective recognises and emphasises the social nature of IS and purposes a new philosophical stance – actualism – as a new way of looking into the world where reality and knowledge are constantly being constructed and altered by *human agents* through their *acts*. Human agents and their actions are, in fact, the essential elements of the two philosophical assumptions assumed in OS [5], namely:

- (1) there is no knowledge without a knower
- (2) his knowledge depend upon what he does

Or, for practical purposes:

- (1') there is no reality without an agent
 (2') the agent constructs reality through his actions

These assumptions were also the result of adopting and extending the Theory of Affordances of James Gibson (1904-1979) to the organisational domain. James Gibson was a psychologist who had formulated a radical 'ecological approach' to visual perception [6]. His theory was an answer to some ideas he felt to be wrong: 1) the notion that perception was a passive process with sensory information being just received and processed, and 2) the environment did not contain enough information to guide action. His answer was that perception is in fact an active process, a system that not only receives, but also search information and assures stability of its supply, acquiring own knowledge in this process. The perception system is able to detect invariants from the flux of information received from the environment. The invariants the agent recognises are those that matter for its survival or well being. This idea can be simply described by the concept of *affordance* which means what the environment can do for a creature, or what it affords the creature to do to. Affordances can be seen as repertoires of behaviour attached to each element that an agent identifies as invariants. OS extended this notion of physical affordance to the social world by introducing the concept of social affordance, which analogously, enables social action by an agent.

OS also establishes a particular dependency between affordances where an affordance cannot exist without the co-existence of another affordance. This kind of dependency known as *ontological dependency* is a key concept in OS. A selected set of affordances and their ontological dependencies, obtained from the different agents in the organisation, is used, in OS, to produce an ontological schema of the organisation known as an Ontology Chart (OC) where the key business terms are represented.

Following the social perspective adopted by OS, organisations are understood as social systems, acknowledging a social reality where people and their social relations play the main role. In these social systems people behave, judge, think and act according to social norms. Norms govern the attitudes of individuals and become a fundamental element in the living organisation. These norms are socially constructed being learned, sustained and improved by each generation directing people to behave in a predictable, civilised and organised way. OS defines the following general behavioural norm structure [7]:

IF condition **THEN** agent **ADOPTS** attitude **TOWARD** something

The human agent having the necessary information (condition) is expected to adopt an attitude that will trigger or influence his actions towards something.

Different groups of norms or norm systems act, in practice, as fields of forces binding people together and determining their expected behaviour. Each individual lives and shares different systems of norms such as those belonging to a nation, religion, tradition, family or a particular organisation, activity or business. Groups of people sharing a system of norms make up an information field. Stamper [8] provides the following definition of an Information field:

"A group of people who share a set of norms that enable them collaborate for same purpose, constitute an information field, where the norms serve to determine the information the subjects need to apply them."

Agents, affordances and its ontological dependencies, information fields and norms are the key notions applied by OS that provide a particular understanding of organisational reality.

2.3 The Theory of Organized Activity

The Theory of Organized Activity (TOA), thought and proposed by Anatol Holt [3] is the second foundation of NOMIS. This theory presents a new perspective of IS based on the concept of ‘Organized Activities’ or OA for short. This conceptual view of TOA and OAs can be described and it is confined in the following statement:

“I intend the expression ‘organized activity’ to mean a human universal. Like language, organized activity exists wherever and whenever people exist. It will be found in social groups of a dozen, or in social groups of millions - in the jungle and in New York City, in every culture, and at every stage of cultural/technological history. It is manifest in every form of enterprise, whether catching big game, coping with a fire, or running a modern corporation – even acquiring and communicating by language.” [4, pg.1].

OAs are intended to form the basis for a systematic analysis of human organisation(s) and TOA emphasises a group of aspects and components for every OA:

- A common communication language – expressed not only by words, but by actions and things as well, known as *units* and recognised by people sharing or involved in the same activity. Behind this idea there is an essential and associated metatheory called the Theory of Units (TU).
- Actions – which directly affect, involve or act on things or materials. Actions are related to a temporal dimension.
- Bodies – representing things or materials, related to a material dimension.
- Action Performers – *always* persons and/or *Organisational Entities*.

For planning OAs, TOA provides a diagrammatic language – Diplan [4] – where actions, action performers and bodies are shown together with the relationships between them.

TOA puts a special emphasis on actions that must be always *human actions*. According to Holt responsibility can only be attributed to humans and therefore computers and other tools cannot perform actions. Human actions, in TOA, are understood as motivated and driven by the interests of their performers.

Fig. 1 defines the OA kernel which relies in two dichotomies: persons/OE and actions/bodies. Referring to this kernel Holt states as a fundamental hypothesis of TOA that: “all organized activities, no matter how complex and subtle, can be usefully represented in these terms...” [4, pg. 56].

Besides TOA key concepts of actions and bodies, it also defines the concepts of *state* and *information*. A state in TOA only applies to bodies and is only understood within specific domains of action. This notion makes a TOA state different from the usual technical description of a state. Regarding information, in TOA it has the exclusive end use of making decisions, which determines the following course of actions. Information in TOA is carried in lumps by bodies, being those lumps

exclusive properties of those bodies. Information contents of a body depend on the context of its use and on the particular actors performing the actions. The same information can be used differently by different actors or in different contexts.

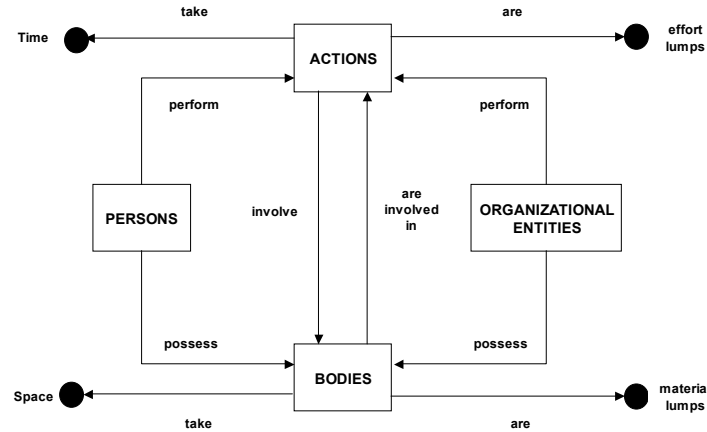


Fig. 1. Organised Activity kernel [3].

TOA view of information is considered ‘a conceptual climax of the theory’. Three statements are made [4, pg.173]: 1) it is the first and only concept of information that relates information to human decision, 2) shows promise for the definition of measures consistent with those of Claude Shannon, and 3) provides a basis for explicating all real-world operations performed on real-world information.

2.4 Enterprise Ontology

The third NOMIS foundational theory is Enterprise Ontology (EO). EO is a complete methodology comprising an ontological model, a systems theory, a model representation and a method for its application. The essential basis of EO is the Language Action Perspective (LAP) [9] where the essence of an organisation is seen as the intentions, responsibility and commitment of people made with the exchange of language acts. LAP sees language as effective action which drives and produces changes in the world. Following LAP, EO views organisations as networks of communicating people. In fact organisations seen at an *essential level*, where there is no material or technical support, are nothing else than a group of communicating people. At this level work is produced as a result of the exchange of language acts; social aspects take higher emphasis than production aspects. In this sense work production is the result of people intentions, commitments, obligations and responsibility. This high level view of an organisation is explored in EO by focusing in communication and coordination aspects modelled using a general communication pattern derived from the W&F basic conversation for action. This pattern known as *the basic transaction* and shown in Fig. 2, is used to model any system or organisation at the *essential level*. An extension of this pattern, which includes the exception paths is understood as a socioeconomic law. In this pattern there is a clear distinction between coordination acts (C-acts) used in conversations and production acts (P-acts) used for

objective action. C-acts are generally connected to commitments made by the speakers and P-acts to material or immaterial (such as decisions) actions made by the hearers.

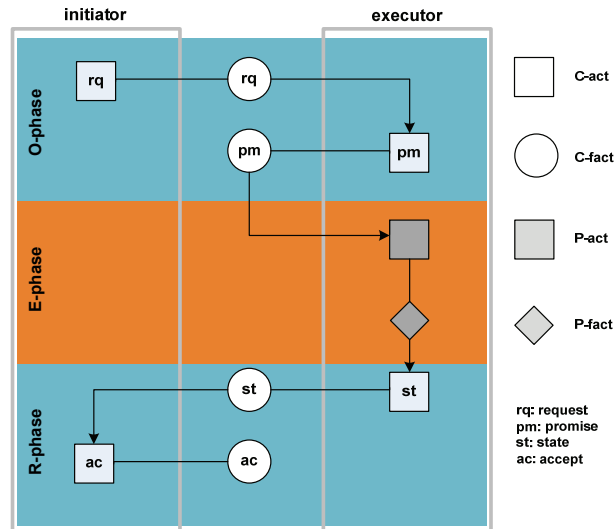


Fig. 2. The Basic Transaction Pattern [10].

Besides the philosophical foundations of EO based on LAP, there is a theory - the Ψ -theory - underlying EO that is defined in four axioms and one theorem. The first three axioms of this theory, namely the *operation*, *transaction* and *composition* axioms define respectively, the key elements modelled in EO, the transaction pattern and two particular compositions of these patterns. The fourth axiom – the *distinction axiom* – clearly establishes three different ontological levels for organisations emphasising the application of EO to the top level known as *the ontological level*. Finally, the organisation theorem completes the Ψ -theory. According to this theorem the organisation of an enterprise is a heterogeneous system composed by three homogeneous systems, namely the business, intellect and document organisation systems (B-organisation, I-Organisation and D-organisation for short).

3 Analysis of the Theoretical Foundations of NOMIS

Table 1. Comparison table of OS, TOA and EO key concepts.

	OS	TOA	EO
Focus	Organisations, agents and behaviour	Human organised activities	Organisations as networks of commitments
Kernel elements	Signs	Units	Business transactions
Context boundary	Information field	Activity	World ontology
Means of conducting business	Agent actions within a normative scope	Human action	Coordination acts and production acts
Business performers	Agents	Persons and OEs	Actor roles

3.1 Theoretical Core Concepts Analysis

Starting by comparing the basic assumptions of OS, TOA and EO presented in the previous section we found a similar perspective of the world and how it is experienced. In fact those theories share, as most soft approaches, the social constructivist stance view where reality is socially constructed. An IS according to these theories is understood as a network of people where connections are established by human actions, interactions and communication. A social dimension where individuals are related to each other and to their environment is always considered.

Regarding the instruments used to carry out the business and considered by each theory, we found linguistic and non linguistic actions performed by humans at the kernel of all of them. Whereas OS and TOA are based on general human actions, EO emphasises language acts, also human actions. One of the reasons for including only actions performed by humans has to do with the responsibility factor attached to those actions. Effectively, responsibility is a key concept in all these theories. To this concept, OS puts some emphasis also in authority, TOA in the interest of the human performers and EO in the intentions of those performers.

Besides those common assumptions of centred human based information systems, there is another common aspect concerning people, their actions and their acting environment: awareness and grounding of the context. Whereas TOA uses the activity as a defining context for any performed action, OS defines an information field as the context to which everyone is subjected in a world of norms and social behaviours. EO, on the other hand defines the context by acknowledging a world ontology defined by each system, usually an organisation.

Another two important and related concepts within IS are meaning and information. Bounded by context in the theories, 'meaning' in TOA is shared, dynamic and socially created, being established by the TOA *unit* concept. In OS 'meaning' is imported from semiotics through the sign concept. From the semiotic perspective a TOA unit is a socially established sign, whose meaning can be defined and socially validated using a criterion. Meaning in EO is established according to the ontological model represented by the EO ontological parallelogram. This model applies to a system or organisation as a world ontology defining the meaning context.

The second concept of 'information' is obtained by interpretation in EO and carried by language, in a broad sense, in LAP and by semiotic signs in OS. TOA has a different and particular view of information: "... it is always carried by bodies", and it "... is a kind of human resource, as essential as energy", which "end-use is in the making of decisions" [3, pg.130-131]. This means that information is useless, except for the making of decisions that can be a simple decision to act.

One last common and important aspect found in TOA, EO and OS is the role of technology as supportive to business systems or organisations, or else to systems of people as they are understood.

Concerning the relevant differences between the theories, they occur specially through the use/planning of human acting within any business or organisation. EO is based on language (as) actions for conducting business operation delegating other actions to a secondary level. On the other hand, TOA is strongly based on human actions of any kind involving always some material resource (a body). OS uses agents for doing the action according to a well formed formula where agents and actions are always present. In OS these actions are regulated by norms. An essential difference

regarding OS is that OS do not put the focus on how actions are conducted or on their effects but, emphasises the necessary conditions for actions through the affordance concept. A comparison table summarising the key ideas of this comparison is presented in Table 1.

3.2 A Path for a New Modelling Approach for ISD

The previous analysis of OS, TOA and EO show that there are some commonalities and some complementarities in them. Nevertheless, all of them provide innovative views of information systems supported by strong theoretical foundations. In spite their different views they share the same foundational basis: they all see the world under a constructivist philosophical stance, understanding information systems with a focus on individual and social aspects. This focus has also a common basis: the human action. TOA uses human actions as the drivers and kernel of activities. In EO there is a focus on communicative acts, which can be seen as a special case of human actions. On the other hand OS focus on affordances, but affordances are a kind of environmental states that afford human actions. Also in OS human action is a key term present in the universal well formed formula and in norms. The concept of human action is really a shared and connecting point between all these theories although they provide different views and use for human actions. Interestingly these views are in some way complementary. EO looks into communication aspects, TOA sees the effects of actions in bodies and as a consequence, action sequences and OS look into the necessary conditions for actions or what affords actions. A more comprehensive model of an IS could use all these views in a coherent way to provide an improved and comprehensive representation of the reality being modelled.

Getting back to the social constructivist view adopted by each theory there is another aspect related to this view that is common and adopted in TOA, EO and OS – the individual understanding of the basic activity terms within a social shared context. This is materialised by the unit concept within the activity context of TOA and the affordance terms related to the information fields in OS. EO uses a world ontology applied to systems to establish the social context. From this common perspective all terms should be understood relatively to its social context. In OS an information field defines the context for meaning, in TOA the context is provided by an activity and in EO it is established by an organisation or system. It is possible that the same terms will be used and understood differently in different contexts in these visions.

Another key concept adopted by OS is the norm concept. Norms are used by OS as a kind of force governing attitudes of people. In OS a group of norms define an information field. In particular, behavioural norms govern people actions. This is another transversal concept that can be used also in TOA and EO. However EO also defines a kind of norms that are expressed as action rules. These are in fact norms although they are only applied within the context of coordination acts.

One issue relating these theories is their particular understanding of information. OS relates information to the semiotic sign, whereas in TOA information is used for the making of decisions. EO adopts in some way the semiotic view of OS within its ‘ontological parallelogram’ [3].

From the perspective of action performers all the theories share a common understanding: OS uses the human agent, TOA the human performer and EO the actor

being all of them a representation of the same concept. In these theories an agent may be an individual or an organisation and responsibility is assigned to this agent accordingly. The other related concepts assigned to individuals and used in the different theories, namely interests, authorities, commitments, intentions can all be used concurrently without clashing to any fundamental aspect of them.

From a practical point of view for the structuring of information systems, TOA gives a good and practical perspective by focusing on activity decomposition of IS, which is also a natural way to understand and organise an IS.

In conclusion, a possible integration of the powerful and well grounded views of OS, TOA and EO promise to be a better representation of the reality that should be further explored.

4 The NOMIS Modelling Approach

4.1 Introduction

One of most important barriers to objective and successful design, development and implementation of technical IS has to do ultimately with the unpredictability factor introduced by the human element. HR showed a possible way to overcome unpredictability with the concepts of *observability* and *precision*. Effectively, unpredictability within IS is directly related to human behaviour where characteristics like intentions, meanings, responsibilities, feelings, decisions, judgments, interests, values and many others are difficult to describe, to predict or to control. Part of this behaviour is *observable*; it is possible to see or to hear what each individual does or talks. These are the *human actions*. From this perspective *human actions* is defined here as the expression or externalisation of human behaviour, or the *observable* part of human behaviour. According to the HR assumption that “anything that is observable will be more consensual, precise and, therefore more appropriate to be used by scientific methods” *human actions* should be an ideal element to be incorporated and used as the centre of analysis, design and modelling of an IS. This conclusion is fully aligned with the perspective of the theories analysed in the previous section. As was pointed before, each of them takes *human action* as a central element of their view.

This approach leads us also to think in the role of IT as a tool extending human capabilities by facilitating, improving, expanding and complementing human action.

Although human action appears to be an ideal basis for IS modelling, it also introduces a link to human behaviour which is difficult to deal with. A solution for this problem is already provided in OS through the use of norms. Norms, particularly behavioural norms, give the possibility to deal with human behaviour by describing ‘expected’ human behaviour. Norms act as fields of forces that make people tend to behave or think in a certain way. OS also uses the *Information Field* paradigm to describe different systems of norms that are shared and applied in different communities. These communities are composed by individuals belonging or acting within the same normative system. To identify and to take into account each normative field and the important norms regulating them and influencing human behaviour should be seen as a step further in the development of successful IS.

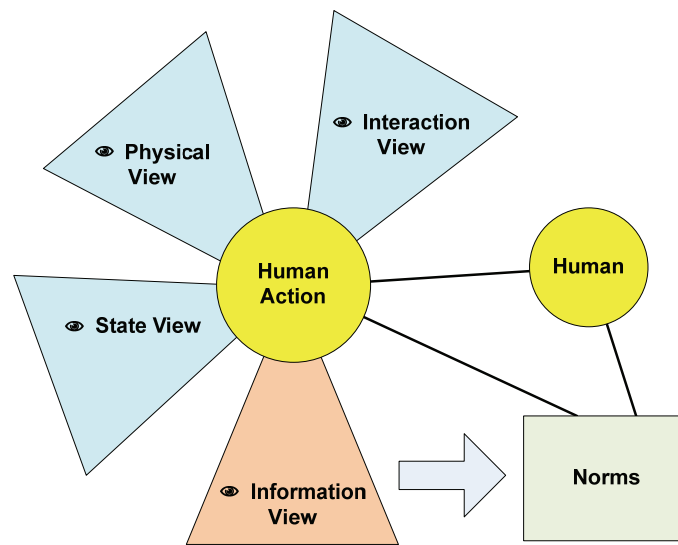


Fig. 3. The NOMIS modelling approach.

The previous view based on human actions, norms and information fields, which is guided by the HR philosophical stance is the basis of NOMIS, an acronym for **N**ormative **M**odelling of **I**nformation **S**ystems, introduced here as a new modelling approach for IS. Following this perspective NOMIS defines four separated views into the IS reality, namely the physical view, the state view, the interaction view and the information view. The first three views reflect the visions and perspectives of, respectively, TOA, OS and EO in which NOMIS is based. The last complimentary view relates to the important information dimension uncovered in the previous analysis. A simple picture expressing this structure is depicted in Fig. 3.

In the following sections each of the views proposed by NOMIS vision and its normative aspects will be briefly presented.

4.2 The Interaction View

The first view NOMIS proposes is the interaction view. The interaction view covers the communicational dimension of human action. All interactions involve communication and communication itself is a form of interaction. Any business or organisation is driven by a network of people performing actions coordinated by communication; interactions link people. From this perspective it is important to draw a special attention in how people interact and in particular communicate.

This view covers the language-action perspective of EO and LAP and extends it. By focusing on interaction aspects of human action, in particular human communication, this view is expected to capture the *essential* aspects of any business and organisations as EO sees and understood it. Furthermore, EO uses a single *business transaction* pattern to model organisations that translates to a coherent, consistent and sequenced group of specific communication acts and production acts. Although this pattern may be used and shown using this view, NOMIS perspective is

not restricted to a single pattern and other types of interaction patterns, possibly based in the richness of language acts types described in [9], are accepted to be used and represented.

In this view the different aspects involving interaction, such as who are the communicating actors, what interactions they perform, what communication links or channels connect them, and other *observable* aspects that may be addressed and represented.

4.3 The State View

The next NOMIS view is the state view. This view looks into environmental conditions and dependencies between them that enable an agent to act. This is the perspective of OS. The environment, including the appropriate elements, enables or affords the agent the ability to execute an action. These situations that NOMIS designates as environmental states (ES) and that the agent identifies as invariants, OS calls them affordances and shows them together with their dependencies in Ontology Charts.

A difference concerning OS is that *environmental states* defined in NOMIS and related to the *affordance* concept do not have to follow completely the rules established for Ontology Charts and, thus may avoid some of the problems related to its use. ESs, as defined in NOMIS, are composed by a body, or an information item, or a group of different bodies and information items in a particular state. The elements composing an ES have some observable form that may include information by using its representation. Nevertheless, ESs usually represents essential business states related by ontological dependency to other ESs and/or bodies and information items. The focus on *states* provided by this view surely is more *stable* than any focus in sequences of actions. A perspective also defended in OS related to OCs.

4.4 The Information View

The NOMIS Information view covers the information dimension of human action. The importance of information is recognised by all IS theories and its significance to human action should be emphasised. Most of human actions depend or rely on information in different ways. Some of them cannot even be performed without it. Therefore the identification of the important information required for each action must take special attention. There are some assumptions NOMIS makes in alignment with the theoretical background provided by its philosophical stance and foundations: (1) information does not exist without a material support, a body or a human actor and (2) information can only be created by humans or instruments and can only be consumed by humans. From a human action perspective there is a focus on what information is required or consumed by the human performer, what information he/her has access and what information he/her produces. From a design perspective it would be important to furnish all information that might be useful for the action. In this sense this view may be used to present the information needed and the underlying system responsible to furnish it to each or to a determined group of human actions. This would be the case of an *awareness* system.

Information is also used in Norms where it is related to agents and human actions. This is another aspect attributed to this information view - to identify and represent the information needed by norms.

Semiotics is a valuable help in this view for understanding information, how it is produced and how it is represented.

4.5 The Physical View

In NOMIS physical view, attention is focused on the material and *observable* aspects related to human action. This view covers the material dimension of human action expressed by TOA. Therefore, it addresses the relationships between bodies and actions: how bodies are affected by actions, how bodies are transported between actions. There is also a need to understand the role of bodies within each action. The same body may play different roles in respect to the action where it is used. A calculator may be used as a tool to help human make operations but can also be used as a simple material for a shop that sells it.

Physical context is another aspect of the material view that can be analysed from a physical perspective. Specific locations (space and time) are many times used for a group of actions; therefore it is useful to look for actions executed in a determined location. From a design viewpoint it will be possible to provide that location with the necessary tools, documents and instruments to help action execution.

A particular representation of this view could be an overall *single action view* where all the elements related to a particular action would be represented. This would allow thinking further into the human needs related to that specific action.

A last representation under this view that was not covered by the previous views is business processes representations showing action sequences. This type of models is most useful and common although under NOMIS approach shows different elements and follows different rules. As an example in NOMIS only *human actions* may be included, action sequence relates to expected behaviour regulated by norms and the initial situation before entering an action and the final situation next to leave it represent states of the environment.

4.6 Modelling Norms and Information Fields

The NOMIS views described previously shown a coherent and comprehensive view of IS centred in *human action* and *information*. Each of them offers a different perspective, however they are related in a consistent model of the IS. In fact, the elements shown and represented in each view must be the same. A coordination act is a human action and can be used in all the views, the same should happen for any other human action, body, human performer and information item. In spite this connection points NOMIS also uses the OS norm concept to regulate human actions and provide a way to model *expected* behaviour. In this case, only behavioural norms, which are related to human actions, are used. Cognitive, perceptual and evaluative norms related to people's beliefs, perceptions and judgements belonging to the intersubjective domain and difficult to use in practice are not included. Behavioural norms are represented analytically in a semi formal way as defined in OS.

With the purpose of helping to structure and to organise norms NOMIS proposes the use of a simple distinction between trigger norms, required norms and auxiliary norms related to the end use of the information from the norm *condition* part. Trigger norms are related to information that causes a human action to occur, required norms to information essential to the performance of an action and auxiliary norms to information auxiliary to an action execution.

Besides regulating human behaviour, groups of norms are used to establish information fields (IFs) where terms are understood by the community living under them. This notion imported from OS is used in NOMIS to define the boundaries of the terminology used in a particular IS. This idea is also in line with the theory of units from TOA, where each term is understood and defined by a criterion used and maintained by a community under a particular activity. In TOA the activity defines an information field. In other words, each IF defines and has a proper ontology.

5 Using NOMIS to Model a Library System

5.1 Introduction

The case study that will be used in this section is an empirical case study of a library system that is presented and described in [3]. In this library system the main emphasis goes to the registering process where anyone can become a member of the library. This membership state allows a member lending books, the second important library process. These two processes, their details and rules together with other information requirements are the essential elements of the library system case study. For the purpose of this section it is not necessary to look into the details of the library system although a complete description can be found in [3].

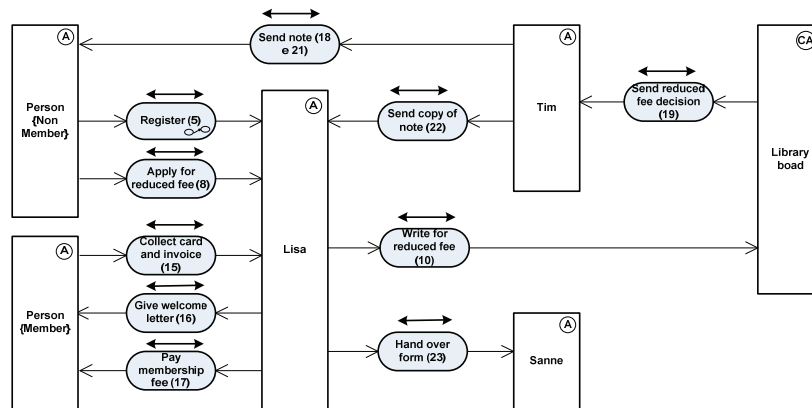


Fig. 4. The HID of the library registering process.

An additional note in this section regards the representation of NOMIS and its vision materialised by NOMIS views introduced in the previous section. In this sense NOMIS proposes a specific notation that defines the elements to be used for representing NOMIS views and suggests a set of diagrams for showing the different

perspectives according to each of them. In this introduction only a few examples of these diagrams will be shown for illustrative purposes. These diagrams will not use directly the NOMIS notation but a representation of some of its suggested diagrams using an extension of the Unified Modelling Language (UML) [11].

5.2 Interaction View

The NOMIS interaction view models the different interactions between individuals emphasising communicational aspects. A key diagram in this view is the Human Interaction Diagram (HID) that shows actor roles and the interactions linking them. Figure 4 presents a HID produced for the library case study. In this diagram we see the different interactions between actors and also interaction activities such as in the case of ‘register (5)’. In fact, it is possible to represent groups of related actions as interactions and to use also templates or patterns of those actions. The HID used here is ‘compatible’ with the Actor Transaction Diagram used by EO although it may describe other types of *business transactions* (interaction activities patterns). This view covers the *vision* of LAP into the business reality by addressing language acts and other forms of communication and interactions. The HID shown here is just one of the possible diagrams used to express this view. Another type is, for example, the Action Sequence Diagram showing action sequences, in this case mostly composed by *language actions*.

5.3 State View

NOMIS state view identifies key environmental states and the existential dependencies between them. Each environmental state (ES) enables a particular group of actions to occur. This acts like an OS *affordance* and, in fact, produces a similar business and organisational view to the one produced in OS using Ontology Charts. A key difference is that ES may include more than one element and the elements included should be part of the elements modelled by NOMIS namely *bodies*, *information items*, *actors* and other ES. For this view NOMIS uses the Body State Diagrams (BSD) to show the different states of a body and the Existential Dependency Diagram (EDD) showing ES and its existential dependencies.

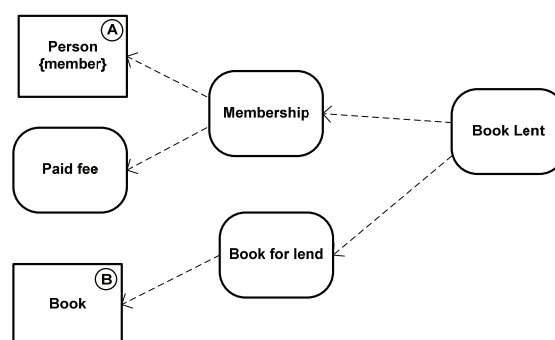


Fig. 5. The EDD of the library.

Figure 5 shows the EDD of the library system. It is interesting to note that the main activities of the library case are included in this diagram related to the key ESs. This is something that is defended in the OS view regarding OCs representing the important business concepts. Although similar, EDDs do not translate directly to OS OCs but the view they convey is in some way similar, capturing the essence of OS vision in this respect.

5.4 Information View

NOMIS information view identifies and models the information elements – the *information items* – used in a system. Besides this identification many other aspects of information are addressed here namely the information related to each human action, the production, access and consumption of information by each human actor and the supporting bodies of information. In the case of the library system most information is identified and represented in tables without a diagrammatic representation. However NOMIS suggest an Information Connection Diagram (ICD) to show key human actions where information is produced, accessed or needed.

5.5 Physical View

NOMIS physical view concerns the material aspects of human action. One possible model representation under this view is the life cycle of a particular body through the human action sequence in which it participates. In the case of the library, we took as an example the registration form and a Body Action Diagram (BAD) was created showing the registration form lifecycle (Figure 6). This diagram is a kind of a UML Activity Diagram showing actions, their sequence and bodies in particular states. This is the view produced by TOA using an action sequence perspective being this diagram compatible with TOA Diplans. BAD provides a typical representation of a business process although the main elements are human actions having *always* behind a human actor. These are the *real business processes* according to NOMIS Vision and their foundation theories. Besides BAD diagrams this view provides also ASD showing action sequences as mentioned before and Action View Diagrams (AVD) for detailing each action individually.

5.6 Information Fields and Norm Analysis

Besides the views produced by NOMIS also information fields (IF) and norms are addressed by NOMIS Vision. In the library case study the library system was considered as a single IF but in some cases is necessary to consider different IFs. In each IF there is a common form of expression and communication that may differ between them. NOMIS views can express IFs as areas in the diagrams and it is important to represent them in each system.

Regarding norms, they regulate human behaviour, in particular, human actions. One important use of norms in NOMIS vision is as action sequence determiners. Effectively, any action sequence is regulated by norms, humans may decide to not

follow these norms, and thus the sequence is not rigid. Another use of norms is to establish the dependencies of an action on their elements including environmental states. In general, there is a key principle that applies to all norms: ‘Any norm has a subject and an action element’, therefore they are just understood and applied in context of human actions. Norms represent the most extensive group of elements of an information system.

6 Related Work

The integration proposed in this work of TOA, OS and EO has never been proposed before although there are a few attempts regarding OS and EO or LAP as the EO broader field. TOA, as a *baby theory*, remain relatively unknown with no much research work produced. Thus, it was not integrated or even related to OS and EO/LAP in any research as far as the authors know. In the case of OS there are already a few links to LAP notions. Effectively, affordances depicted in OC may represent language-acts seen as semiotic signs standing for something else. In [12] Ronald Stamper addresses this fact by referring to *the semantics of communication acts* in the context of OS and OCs in particular. In this case it corresponds just as a perspective of LAP under an OS view and not a different use of both views as NOMIS proposes.

In [13] there is an acknowledgment of the power of the different views provided by OS and LAP and a proposal to integrate the *semantics* of OS with the *pragmatics* of LAP. However this integration is made in a different level without considering the different kernel elements used by NOMIS. Effectively, OS and LAP act together in the identification of requirements that are afterwards modelled by EO/DEMO business transactions resulting in a final LAP based only modelling.

[14] also proposes an integration of OS and LAP with the *RENISYS Method* but in this case the integration uses mainly OS NORMS and LAP conversational transactions lacking the broad integrating coverage of NOMIS.

In [15] there is another integration proposal that uses OS norms for extending the business process models defined with DEMO, and OS semantic analysis to help uncovering those norms.

Besides those integration proposes we found are some other minor contributions but all of them fail in the integration broadness, strength, theoretical foundation and extension with respect to NOMIS.

7 Conclusions and Future Work

This paper presented NOMIS, a normative approach to information systems modelling, as a new way to understand and model information systems based on Human Relativism. NOMIS approach joins and extends the views of Organisational Semiotics, the Theory of Organized Activity and Enterprise Ontology in a comprehensive and coherent way and provides a broader and accurate vision of organisational and business reality. These views describe NOMIS vision that is

represented by a proper notation not described in this paper. Also, a case study of a library system was presented and modelled using NOMIS that intended to illustrate NOMIS vision and give a simple idea of its application.

Regarding future work, the way NOMIS sees IS has never been applied before to the development of information systems although there are a few applications of its foundational theories. This opens a lot of possibilities for future work: the approach needs to be tested and developed further in real contexts. Many fields such as Human Computer Interaction, Ontology Engineering, Business Processes Management and other have to be reanalysed from a NOMIS perspective with possible contribution in both directions. A missing element in NOMIS is a methodology for information system development. NOMIS just proposes a modelling approach that enables the representation of information systems but the way from this representation to the development of concrete information systems is still needed. Besides methodologies NOMIS modelling requires tools as well. For the representation of NOMIS models it will be necessary to develop tools that will allow for verification and validation support of developed models.

In conclusion, much work can be done as future work within this new NOMIS approach that will be essential and necessary for its success.

References

1. Cordeiro, J., Filipe, J. and Liu, K.: Towards a Human Oriented Approach to Information Systems Development. In *Proceedings of the 3rd International Workshop on Enterprise Systems and Technology*. Sofia, Bulgaria (2009).
2. Liu, K.: *Semiotics in Information Systems Engineering*. Cambridge University Press, Cambridge, UK (2000)
3. Holt, A.: *Organized Activity and Its Support by Computer*. Kluwer Academic Publishers, Dordrecht, The Netherlands (1997)
4. Dietz, J.: *Enterprise Ontology, Theory and Methodology*. Springer-Verlag, Berlin Heidelberg, Germany (2006)
5. Stamper, R.: Signs, Norms, and Information Systems. In *Signs of Work*, Holmqvist B. et al (eds). Walter de Gruyter, Berlin, Germany (1996).
6. Gibson, J.: *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, (1979).
7. Stamper, R.: New Directions for Systems analysis and Design. In *Enterprise Information Systems*, Joaquim Filipe (ed). Kluwer Academic Publishers, Dordrecht, The Netherlands, (2000).
8. Stamper, R., Liu, K., Sun, L., Tan, S., Shah, H., Sharp, B. and Dong, D.: Semiotic Methods for Enterprise Design and IT Applications. In *Proceedings of the 7th International Workshop on Organisational Semiotics*, Setúbal, Portugal, (2004).
9. Winograd, T., Flores, F.: *Understanding Computers and Cognition*. Ablex Publishing Corporation, Norwood, NJ, USA (1986).
10. Dietz, J.: The Deep Structure of Business Processes. In *Communications of the ACM*, 49, 5, pages 59 -64, (2006).
11. Booch, G., Rumbaugh, J. and Jacobson, I.: *The Unified Modeling Language User Guide*. Addison-Wesley, (1999).
12. Stamper, R.: Exploring the Semantics of Communication Acts. In *Proceedings of 10th International Working Conference on the Language Action Perspective*. Kiruna, Lapland, Sweden, (2005).

13. Shishkov, D., Dietz, J. and Liu, K.: Bridging the Language-Action Perspective and Organisational Semiotics. In: *Proceedings of the Eighth International Conference on Enterprise Information Systems*, Paphos, Cyprus, (2006).
14. De Moor, A.: Language/Action Meets Organisational Semiotics: Situating Conversations with Norms. In *Information Systems Frontiers* 4:3, 257–272, (2002)
15. Barjis, J., Dietz, J. and Liu, K.: Combining DEMO Methodology with Semiotic Methods in Business Process Modeling. In *Information, Organisation and Technology*. K. Liu, R. J. Clarke, P. B. Andersen and R. K. Stamper (eds). Kluwer Academic Publishers, Boston, pages 213-246, (2001).

Methodological Support for the Design of Enterprise Information Systems with SDBS: Towards Distributed, Service-Oriented and Context-Aware Solutions

Boris Shishkov

IICREST, 53 Iv. Susanin Str., Sofia, Bulgaria
b.b.shishkov@iicrest.eu

Abstract. Taking in consideration the inability of traditional software development methods to meet to a full extent some new demands, such as ones related to service-orientation and context-awareness, we suggest in this paper directions to enrich software development methods, for (partially) resolving this drawback. We introduce and briefly discuss the SDBC approach – ‘SDBC – Software Derived from Business Components’, mentioning as well actual related work that is concerned with the alignment between business process modeling and software specification. We then discuss two of the most important challenges of current software development, namely adaptability (envisioning distribution and context-awareness as desired properties) and service-orientation. Based on this, we derive some actual solution directions that concern software development. We consider this as a contribution that relates to the further software development trends.

Keywords. SDBC, Components, Service-Orientation, Context-Awareness.

1 Introduction

Information and Communication Technology (ICT) has substantially influenced almost all spheres of human activity (including enterprise development), during the previous century, bringing on the stage information, as main asset, which in turn led to the problems of collecting, storing, processing, and communicating (enterprise-related) information [1]. Hence, Enterprise Information Systems (EIS) consist not only of software applications and software infrastructures to run them but also of people, hardware units, and so on. The EIS notion gets broader and more complex with the development of enterprises and the advances in enterprise technology. It is widely agreed nevertheless that ensuring the development and operation of enterprise software applications with predictable and improved cost, schedule, and quality, is of crucial importance for current EIS [2]. This inspires us to focus in particular on the development of ICT applications that are to support enterprises. Nevertheless, this was not a big challenge well until the late 1980s, in our view, not only because enterprise processes were not so complex (for example distributed and adaptable) as they currently are but also because the assumed support that ICT applications would deliver was mostly limited to the delivery of some calculation upon request. This changed rapidly in the following decade when more and more software development

tasks were dominated by the goal of (partially) automating human actions. This pushed, we believe, software engineering in a new ‘dimension’ where software systems were supposed to play complex role in the context of an enterprise. Hence, the alignment between business process modeling and software specification was becoming crucial.

This mentioned alignment used to be a weak point in most of the software development methods during the 1990s and even beyond [3]. What’s more, the software community used to consider the challenge of closing this gap as having a breakthrough importance [4]. Reporting actual (partial) research results in this direction, we relate in this paper the mentioned challenge to further and currently actual ones, such as adaptability and service-orientation, in order to establish a perspective on how ICT applications should be built in order to meet current demands, especially as far as EIS are concerned.

In particular, we introduce and briefly discuss one software development approach that reflects the traditional software development best practices, namely the SDBC approach – ‘SDBC – Software Derived from Business Components’ [5], mentioning as well actual related work that is concerned with the alignment between business process modeling and software specification. We then discuss two of the most important challenges of current software development, namely adaptability (envisioning distribution and context-awareness as desired properties) and service-orientation, concluding about the importance of enriching the well-established software development tools with adaptability and service –related features. We have a two-fold problem here, nevertheless: (i) The well-established software development methods are mostly envisioning the design of an ICT application which even though built in a component-based fashion, is self-contained – services performed by unknown components through the ‘Cloud’, being adapted dynamically to a need, are not considered. (ii) The current service-oriented and context-driven approaches are often too abstract with regard to the actual service realization that is to be anchored in particular ICT applications and some corresponding components.

Thus, the contribution of this paper relates to a proposed reinforcement of some well-established software development best practices through actual enrichments inspired by adaptability and service –related desired features.

The outline of the remaining of this paper is as follows: Section 2 introduces the SDBC approach and discusses other representative traditional software development approaches, for the sake of getting insight on some of the well-established software development best practices. Section 3 outlines two of the currently actual EIS-related challenges, namely distribution and adaptability, in order to inspire the further discussion on enriching traditional software development methods. Then Section 4 presents some proposed solution directions with regard to traditional software development methods and their desired enrichment. Section 5 exemplifies partially some of the presented views and proposed solution directions. Section 6 briefly outlines related work. Finally, Section 7 presents our conclusions.

2 SDBC and Traditional Software Development Methods

In this section we outline briefly the SDBC approach as an approach that is based on the traditional software development principles and we discuss subsequently some other traditional approaches.

SDBC. In summarizing the approach *Software Derived from Business Components – SDBC* [3], we use the following abbreviations as applied in Figure 1: **bc** – *Business Component* (a business sub-system that comprises exactly one business process); **bk** – *Business CoMponent* (a model of a Business Component, which model is adequately elaborated in terms of statics and dynamics); **glbk** – *general Business CoMponent* (which is re-usable by extension); **gcbk** – *generic Business CoMponent* (which is re-usable by parameterization); **ssm** – *software specification model*; **sc** – *software Component* (an implemented piece of software representing a part of an application); **sk** – *Software CoMponent* (a conceptual specification model of a Software Component).

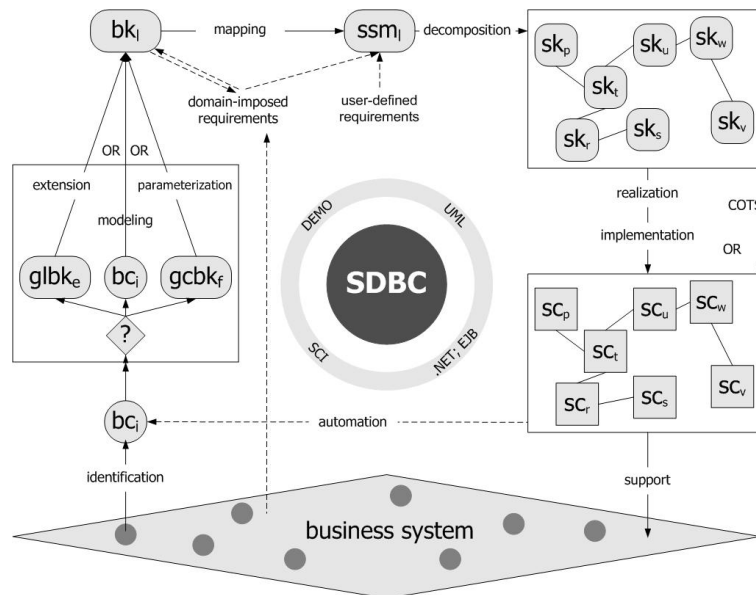


Fig. 1. Outline of the SDBC approach [5].

The Figure shows that SDBC is about a component-based business-process-modeling-driven specification and realization of software. The starting point is the consideration of a business system. Business Components are identified from it. This can be done through the *SCI* technique – *Structuring Customers' Information* [5]. The Business Components should then be reflected in corresponding Business CoMponents, in supplying an adequate modeling foundation for the further software specification activities. Another way of arriving at a Business CoMponent is by applying re-use: either extending a general Business CoMponent or parameterizing a generic Business CoMponent. DEMO and other Language-Action-Perspective -driven

modeling tools [6] are relevant as far as Business CoMponents' specification is concerned. Each Business CoMponent should be then elaborated with the domain-imposed requirements, for the purpose of adding elicitation on the particular context in which its corresponding Business Component exists within the business system. Then, a mapping towards a software specification model should take place, possibly driven by the DEMO-UML transformation mechanism [5]. The domain-imposed requirements as well as the user-defined requirements are to be considered here, since the derived software model should reflect not only the original business features but also the particular user demands towards the system-to-be. The (UML-based) software specification model would need then a precise elaboration so that it provides sufficient elicitation in terms of structure, dynamics, data, and coordination [3]. The model needs also to be decomposed into a number of Software CoMponents reflecting functionality pieces. Then these Software CoMponents are to undergo realization and implementation, being reflected (in this way) in Software Components. This final set of components might consist of such components which are implemented (using software component technologies, such as .NET or EJB, for instance) based on corresponding Software CoMponents and such components which are purchased. Finally, the (resulting) component-based application would support the target business system, by automating anything that concerns the initially identified Business Component(s) identified from the mentioned system.

Other Traditional Methods. The challenge of capturing the essential aspects about business processes for the purpose of further software specification, has been addressed not only by the SDBC approach but also by methods such as Catalysis [7] and Tropos [8] as well as by the Model Driven Architecture – MDA [9].

The Catalysis method provides a coherent set of techniques for business analysis and system development as well as well-defined consistency rules across models. However, these techniques concern the software design perspective and have no theoretical roots relevant to the modeling of business processes. Hence, the business process modeling as conducted in Catalysis would inevitably be superficial and therefore the method cannot guarantee an adequate capturing of all related real-life aspects, including semantic and pragmatic ones. In addition to this, Catalysis does not have mechanisms for a mapping between business process models and software specification models. Therefore, a definite strength (in this regard) of SDBC is that, relying on the LAP-OS 'combination', the approach supports adequately the business process modeling task and the software design activities in SDBC stem from a pure business process model, guaranteeing that the application-to-be would function adequately in the business environment in which it would have to be integrated.

The strengths of the method Tropos relate to its capability of conducting a sound requirements analysis, considering the business processes which are to be supported by the application-to-be. From such a business process modeling point of view, the method addresses the software design. The mentioned requirements analysis includes elicitation not only of the 'early requirements' that concern the original business reality but also of the 'late requirements' which are about a corresponding updated (desired) business reality. The analysis is driven by a thorough consideration of the intentions of stakeholders, modeled as goals which are then reflected in the system's global architecture. Its definition is in terms of sub-systems interconnected through data, control, and other dependencies. Then a detailed design follows. Therefore, all

this features Tropos as a powerful method for designing software, which appropriately refers to the task of capturing essential real-life aspects that concern the modeling of business processes. However, the method is incomplete with regard to some of these aspects – it is not exhaustive in handling semantics and is insufficiently concerned with essential pragmatic issues, such as communicative actions, negotiations, coordination, and so on [5]. Further, the method lacks (just like Catalysis) clear and complete guidelines (and elaboration) on how to reflect the business process modeling output in the specification of the application-to-be. Such a specification would therefore inadequately reflect the original business model.

MDA prescribes three viewpoints from which models of the application-to-be (our target software system) should be defined: Computational Independent Models (CIMs) should focus on the environment and requirements of the system, abstracting from the system's construction; Platform-Independent Models (PIMs) should focus on the functionality of the system without revealing details on the specific technological platform on which the system is built, and Platform-Specific Models (PSMs) should define how a PIM is built using some specific platform. Therefore, the Computational Independent Modeling as well as the CIM-PIM transformation relate to the problem addressed in the current paper, namely the achievement of an adequate business-software alignment which is concerned with all relevant real-life aspects. However, bridging business process models and application design by using Computational Independent Modeling and realizing a CIM-PIM mapping, are issues not enough explored, as it is well known. The MDA Community still misses sound guidelines and procedures on how to discover Computational Independent Models and how to reflect them in Platform Independent Models.

What we conclude is that traditional software development approaches fundamentally address the link between the construction of ICT components (done by developers) and the delivered functionality (mainly consumed by users). Hence, traditional software development is not considering the possibility to directly 'consume' services from the 'Cloud', adjusting this in an ad-hoc fashion, upon the appearance of a necessity.

3 Towards More Distributed and Adaptable Solutions

In this section we outline, as mentioned in the Introduction, two of the currently actual EIS-related challenges, namely distribution and adaptability, in order to inspire the further discussion on enriching traditional software development methods.

Distributed Web Services. We start discussing the actual (EIS-related) challenges (which challenges are in our view insufficiently reflected in traditional software development approaches), by firstly addressing the service concept not only because service-orientation usually demands a heavy distribution (that is not fully in line with most software development approaches) but also because the current significant changes in software technology are centered around this concept [10].

From an abstract point of view, a service represents a piece of well-defined functionality that is available at some network endpoint and is accessible via various transport protocols and specialization formats. The functionalities provided by

services cover a vast spectrum reaching from low level features like offering storage capabilities, over simple application functions like changing a customer address, to complex business processes like hiring a new employee.

To usefully utilize the service perspective in developing ICT applications would mean according to Shishkov & Van Sinderen [11] the ability to create new applications from existing services, independently of who provides these services, where they are provided, and how they are implemented. It should be nevertheless mentioned that by ‘creating’ an application, it is not meant constructing the application from the scratch; what is ‘created’ is the end result that represents a functionality consumed by users. This is the significant change in our vision concerning software technology – developers would no longer possess full control over all software components that realize services. What is more, the development task may split:

- some developers would just focus on the development of small software modules delivering generic adjustable services to whoever might be interested in using them;
- other developers would not develop software components any longer, focusing instead on the composition of complex functionalities (for this, they would be using available generic services).

According to some recent views of Frank Leymann [12], a new kind of middleware is currently evolving for the partial support directed to the dealing with services in such a way, with the idea that based on the specification of the functionality needed, the middleware determines (automatically or through a developer’s intervention) a (composite) service that would deliver the required functionality. This is not only affecting the application functionality creation but it would concern performance, taking into account that services would require in most cases processing power of back-end server systems. Thus both application creation and run time processing would substantially rely on support from the ‘Cloud’ and this should certainly point in particular to web services that represent services which are created and executed through the Web. Hence, in order to be of actual use, such services would demand enabling technology standards and the web service technology stack as according to Papazoglou [10] outlines some actual web service technologies and standards – Figure 2.

As it is seen from the Figure, web services’ relying on a transportation protocol is crucial. Although not tied to any specific transport protocol, web services build on ubiquitous Internet connectivity and infrastructure to ensure nearly universal reach and support. Hence, their mostly relying on HTTP (the connection protocol used by web servers and browsers) and XML (a widely accepted format for all exchanging data and its corresponding semantics) looks logical.

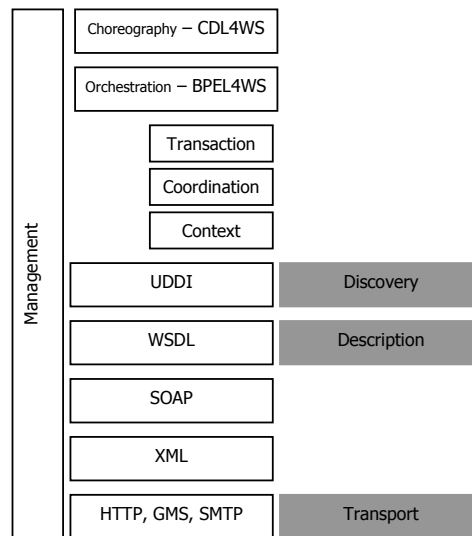


Fig. 2. Web services technology stack [10].

Having this as a ‘foundation’, we are to mention further the core web service standards, namely SOAP, WSDL, and UDDI:

- SOAP (Simple Object Access Protocol) is a simple XML-based messaging protocol on which web services rely in exchanging among themselves information. SOAP implements a request/response model for communication between interacting web services.
- WSDL (Web Service Description Language) is a language that specifies the interface of a web service, providing to the requestors a description of the service in this way.
- UDDI (Universal Description, Discovery, and Integration) represents a public directory that not only provides the publication of online services but also facilitates their eventual discovery.

When we then have to compose web services, we need to introduce some orchestration, defining their control flows [13], such as sequential, parallel, conditional, and so on, with this ending up with the determination of complex processes that may span many parties. BPEL4WS (Business Process Execution Language for Web Services) can support usefully such composition activities.

As the collaboration among many parties (through their web services) is concerned, a common observable behavior (choreography) would often need to be defined. CDL4WS (Choreography Description Language for Web Services) can usefully support such collaboration descriptions.

Adaptable Systems with Context-Aware Behavior. The utilization of a generic service for a specific user-related situation logically relates to acquiring knowledge on the context of the user and also exploiting this knowledge to provide the best possible service, which is labeled as context-awareness by Shishkov & Van Sinderen [14].

We hence claim that taking the end-user context into account is important in adequately delivering a service. Examples of end-user context are the location of the user, the user's activity, the availability of the user, and so on. We do assume that the end-user is in different contexts over time, and as a consequence (s)he has changing preferences or needs with respect to services.

Usefully enriching in this perspective traditional application development that may be SDBC-driven for example, means that the application under consideration or the component under consideration need some 'sensitivity' with regard to the changes in the end-user's context.

A schematic set-up for a context-aware application is depicted in Figure 3. Here, the application is informed by sensors of the context (or of context changes), where the sensing is done as unobtrusively (and invisibly) for the end-user as possible. Sensors sample the user's environment and produce (primitive) context information, which is an approximation of the actual context, suitable for computer interpretation and processing. Higher level context information may be derived through inference and aggregation (using input from multiple sensors) before it is presented to applications which in turn can decide on the current context of the end-user and the corresponding service(s) that must be offered.

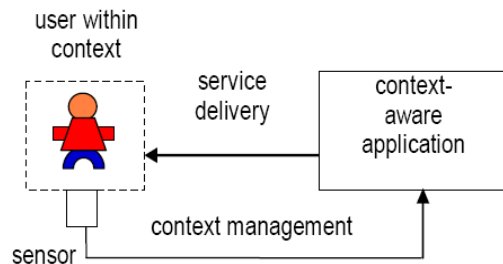


Fig. 3. Schematic representation of a context-aware application [14].

Based on previous research, the authors claim that the establishment of what desirable behavior corresponds to a (captured) change in end-user's context, is a challenge of great importance. For this reason, we would like to especially discuss this challenge in the current paper.

This challenge points to several interrelated sub-challenges, discussed briefly below:

- The Capturing: it is not trivial to establish through sensors what change in end-user's situation has actually occurred not only because the raw sensor data is very often at high risk of being misinterpreted but also because establishing adequate Quality-of-Context levels that are reliable is rarely realistic in our view.
- The Delivery: although a deterministic approach for defining the delivered behaviors, when it is precisely known still in advance what is the service that the system should deliver in one or another particular end-user context situation, appears to be a reliable solution, it is questionable how realistic is it to adequately foresee all possible end-user context situations at design time; as for non-deterministic solutions, we doubt how much reliable they would be taking into account the observed by us failure (in general) in much artificial-intelligence -related projects.

- The Response Capacity: even though it should be in most of the cases possible to predict most of the situations the end-user may appear to be in, there would always be such situations that are completely out of expectation and if they would demand service(s) that in turn need(s) unavailable resources, it would be impossible to deliver timely the proper service(s) to the end-user.

What we conclude is that traditional software development approaches can have chances to stay adequate if they allow for a paradigm shift according to which:

- Components and not applications are developed, with no restriction on how and with what other components the component under development will interact;
- The service delivered by a component is separated conceptually from the implementation of the component, keeping at the same time correspondence and traceability that would allow for different interactions of the service delivered by the component with other services, with these interactions realizable just through the service descriptions;
- A component is to allow for delivering different versions of its service putting this in dependence from the end-user situation.

A schematic representation of this vision is exhibited in Figure 4:

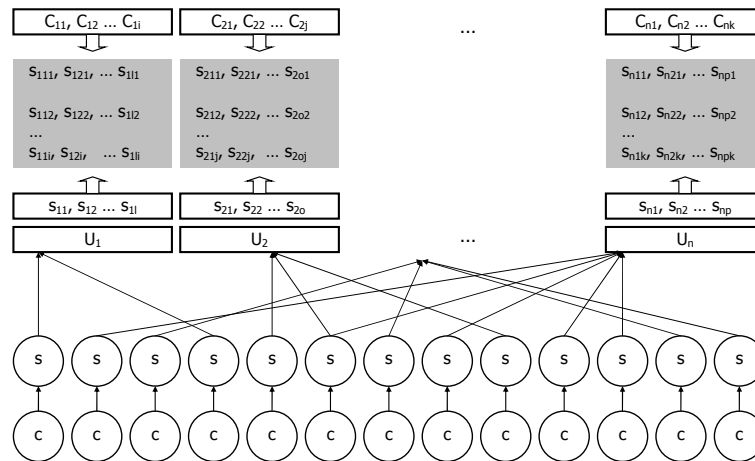


Fig.4. Actual view on the software development challenge.

As the Figure suggests, software components (C) appear to be the units to be developed, intended to deliver functionality pieces to a broad non pre-defined public. The manifestation of these functionality pieces are the services (S). It is to be noted that the corresponding underlying components remain 'hidden' from the broad public. Being utilizable in unlimited ways, a service may appear in tens and hundreds of utilization schemes (U), in each of which schemes it is combined in different ways with other different services. Hence, for a particular utilization scheme, there should be services derived from the service level (S) which are appropriately parameterized, composed and combined. The aggregate of this represents the desired functionality with regard to a utilization scheme. What should also be taken into account however is the context corresponding to each of the utilization schemes – this would point to a

number of different possible context states (C_i); for each of these context states, the service output should appear in different versions. Therefore, each matrix (colored in grey in the Figure) corresponds to the service support that relates to a particular utilization scheme – there are different versions of combinations of services for each possible context state.

The visions that were presented in this section inspire us to propose some solution directions with regard to traditional software development methods and their desired enrichment – this will be done in Section 4.

4 Solution Directions

As already mentioned, we will propose some solution directions with regard to traditional software development methods and their desired enrichment. We will do this by firstly further analyzing several actual demands towards the development of software and secondly – proposing some partial solution directions on this basis.

Analysis. Inspired by the demands identified and also discussed in the previous section, which can be summarized as: (i) floating components; (ii) utilizable services; (iii) sensitivity to end-user's context, we add as well the following issues that are to be taken into consideration:

- the *run-time control* will be slipping away from the hands of the user since different services are delivered by components which are implemented/run on remote servers where the user has limited or no control;
- the inevitable *infrastructure dependency* regarding service utilization will push much more attention to the service environments development, which together with the development of generic components, will be comprising the main part of the software development efforts.

These changes put in our view less technical and more business oriented people as the actual producers of the solutions that are finally delivered to the user; it becomes more necessary to be able to properly combine functionalities in the context of a business process and different requirements rather than being able to implement components that deliver functionality pieces (such tasks are already becoming trivial in many cases).

Proposal for Reinforcing Traditional Software Development Practices. Considering (i) the traditional application development ways, and particularly the perspective of the SDBC approach, (ii) the current needs for distribution, service-orientation, and context-awareness, and also the conclusions (derived) on their applicability within software development, (iii) the brief analysis (see the first part of the current section) on the impact of the change in the way we look upon software development, we propose a software development vision (outlined in Figure 5) that we expect would better fit the upcoming demands:

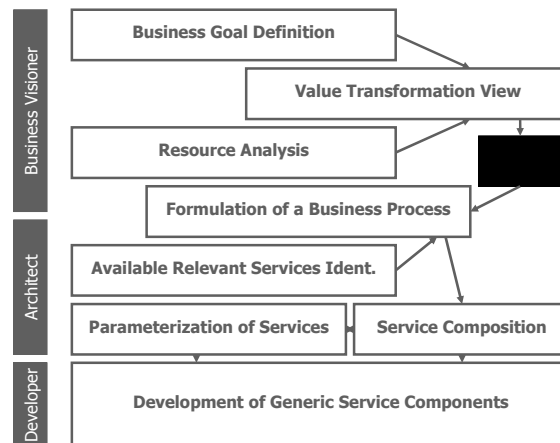


Fig. 5. Vision on some actual software development tasks and roles.

As it can be seen from the Figure, the role of a Business Visioner is becoming more important and far reaching - Business Visioners are defining a business goal and this goes often irregardless of the technology to be used, it is just an abstract goal statement. There is another activity that in our opinion is to go in parallel with the mentioned one: resource analysis (it is impossible to realize even a very small business goal without resources - resources are to be analyzed in order to justify the adequacy of the business goal). What is nevertheless of significant importance in our view is the definition of the desired value transformations: how we transform our available resources in order to achieve the business goal. This gives already a rough view on the business process to be addressed but how to move from the value transformation view to a specification of a business process, is a challenge and we still have not seen a definitive transformation vision. It is to be mentioned that technology is applied not for the sake of applying technology but as a way to improve effectiveness, efficiency, and quality. We need a business-level motivation or justification that a web service is a better 'fit' in a particular business model than a human service, for example. For this reason, the decision to use web service(s) has business-level roots. Services should be identified for knowing how they can be used with regard to the business process. For this reason, the Business Process formulation relies not only on Business Visioners' value transformation view but also on Architect's service-related view. Moreover, after there is a business process that is (partially) implementable through web services that are basically known as type, the particular service selection may start (what is the best service solution or the cheapest, and so on). Once the services are selected for being used, they would need to be parameterized and possibly composed (service composition is a topic that will not be discussed in this paper). These all are responsibilities of the Architect. Finally, the Developer must have implemented the components which realize the actual services.

Considering distributed systems and typical distributed architectures and inspired by previous research, we propose a layered design architecture illustrated in Fig. 6:

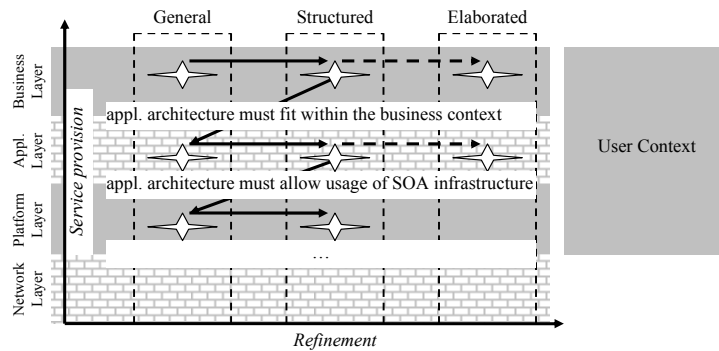


Fig. 6. Proposed layered architecture.

As it can be seen from the Figure, we distinguish 4 service provisioning layers, namely Network Layer (concerned with networking protocols), Platform Layer (concerned with (IT) infrastructures), Application Layer (concerned with the application logic), and Business Layer (concerned with the business logic that is not delegated to the application layer).

Further, we consider 3 ‘degrees’ of refinement, namely General (‘black-box’) view, Structured (high-level ‘white-box’) view, and Elaborated (detailed ‘white-box’) view.

Next to that, we enforce two relevant desired properties (requirements):

- The service(s) provided by the Application Layer must fit within the business context;
- The architecture of the Application Layer must be SOA compliant.

Furthermore, we consider end-user’s context as crosscutting at least with regard to the Platform, Application, and Business layers since in different context situations, different business steps and solutions are to be considered which leads in turn to different services and they in turn are projected on platforms. Nevertheless, for the sake of brevity, we will not discuss in more detail the proposed architecture.

In the following Section, we partially illustrate our presented views.

5 Illustrating Example

To illustrate partially our views, we consider the ‘derivation’ of an Education Mediator (EM) that would support customers in a number of ways, in an e-learning context. By ‘customers’, we mean the users of EM’s services; those could be students and teachers (in the simplest case). Furthermore, we address (for the sake of brevity) only EM’s advice provisioning service: *a customer can receive from EM advice which of the Student/Teacher entities (registered in the system) best satisfy a need* (for example, which is the best teacher with respect to a particular student demand). To receive advice from EM, the customer approaches EM’s *ADVISOR* (an entity that is internal with regard to EM, which is responsible for handling the advice provisioning). It should be nevertheless noted that the Advisor may be shielded from the customer by the EM and in such a case the customer would be ‘talking’ to the EM

and the EM would in turn route requests to (and results from) the Advisor. Approaching the Advisor, the customer should specify a request: course type (e.g. lecturing course or experimental course), preferences (e.g. closest to a particular subject), and so on. Based on this (and acting ‘through’ the Match-maker, to be introduced further on in this paragraph), EM’s *REQUEST HANDLER* (an entity that is internal with regard to EM also; this entity processes requests) generates a standardized request specification, appropriately synthesizing some of the information provided by the customer. This is delivered then to EM’s *MATCH-MAKER* (an entity that is also internal with regard to EM; this entity is responsible for finding a match using the standardized request and considering what is currently available); the Match-maker realizes matches driven by particular criteria, chosen by the customer (and represented in the standardized request), for instance: a preference for a teacher from a particular country or institution or the earliest available teacher. In order to realize a criterion-driven match, the Match-maker applies relevant rules and procedures, nevertheless needing input from EM’s *DATA SEARCHER* (an entity that is also internal with regard to EM; this entity is responsible for searching). The Data searcher searches through the information concerning the available (Student/Teacher) entities and also applies procedures to it. This hence supports the identification of candidate matches relevant to the particular customer’s request. The Match-maker applies its rules and procedures to realize a final match, passing this information to the EM’s Advisor.

Considering the above-presented briefing, a business entity model is built (Figure 7), with a notation that is inspired by DEMO [6].

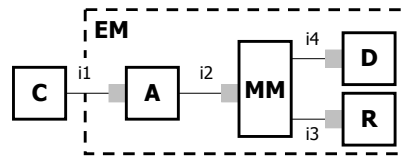


Fig. 7. Business entity model for the EM case.

The identified entities are presented in named boxes – these are Customer (C), Advisor (A), Match-maker (MM), Request handler (R), and Data searcher (D). Interactions i1 – i4 are identified as follows: between C and A (i1), between A and MM (i2), between MM and R (i3) and between MM and D (i4). As for the delimitation, C is positioned in the environment of the education mediation system EM, and A, MM, R and D together form the EM system.

We model then interactions using the notations of UML Activity Diagram [16]: i3 and i4 are to be progressing in parallel and only after they have been exhausted (the standardized requests and candidate matches have been delivered) the match-making can be done (i2) followed by the advice (i1) – this is illustrated in Figure 8 (upper part). This is the *business process level*, as labeled in the Figure, and it is assumed that human-driven roles (and responsibilities) stay behind each of the interactions and as it is about human activities, much is driven by complex organizational (and societal) norms, much is actually done using best practices, and much is done in an intuitive way. IT services nevertheless require many explicit definitions. That is why the IT services that correspond to the business-process-level interactions, are considered

together with other related issues, as it is shown in Figure 8 (lower part), depicting the *IT service level*, as labeled in the Figure.

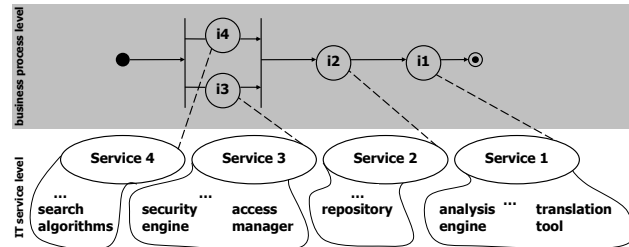


Fig. 8. Service derivation for the EM case.

As it is seen from the Figure, searching requires *search algorithms*, request processing requires an adequate supportive *security engine* and *access control facilities*, match-making needs *repositories* with candidate matches and match criteria, the delivery of an advice requires an *analysis engine* and sometimes, a *translation facility*, just to name a few.

We need to further extend this model, particularly with respect to ‘IT Services level’, by considering an adopted service pattern proposed in [11] according to which a coordination service (supported by an information service) orchestrates the work of the other services. Hence, the final EM service model is presented in Figure 9:

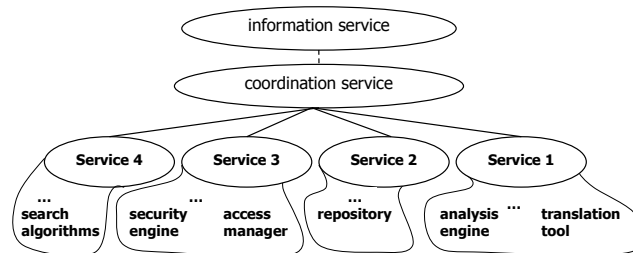


Fig. 9. Service model for the EM case.

As it is suggested by the Figure, the coordination service orchestrates the work of the other services, namely Service 1, Service 2, Service 3, and Service 4. Taking into account the case information and the considered domain, we label these 4 services in the following way:

- Service 1: **Educational Mediation Service;**
- Service 2: **Educational Broker Service;**
- Service 3: **Educational User Agent Service;**
- Service 4: **Educational Resource Discovery Service.**

This example, although simplified, partial and incomplete, illustrates the view on the paradigm shift, as discussed in Section 3, implicitly suggesting that business entities are not to be considered any more exclusively as source for identification of software

components but it could often be that the models of such entities are just valuable to actually give the right restrictions with regard to the services to be composed.

6 Related Work

In this paper, we have presented a service-orientation and context-awareness –rooted vision on software development, as a proposal for a desired reinforcement of traditional software development approaches and methodologies.

With regard to considering related work, among the work that is focusing on core service-orientation and partially on context awareness concerns are [15,16,17], driven mainly by consideration of particular key problems, such as service composition and web service technologies, overlooking nevertheless the issue of requirements-functionality alignment. On the other hand, there is reported research concerning the requirements-functionality alignment, mainly related to relevant methods, such as Catalysis [7], Tropos [8], and SDBC [5] which however lack the proper service-orientation focus.

7 Conclusions

In this paper, have discussed traditional software development approaches, putting the discussion in the perspective of some of the important current challenges related to software development, such as service-orientation and context awareness. Also, based on an analysis of the change brought to the desired vision on software development by these actual challenges, we have considered the impact of this change influencing already the ways software is developed. Finally, we propose a partial vision on how these new influences can be incorporated in software development, through some new software development tasks and roles. We have partially exemplified our vision.

Hence, the contribution of this paper represents an explicit proposal concerning the development of software and limited to proposed enrichments inspired by the consideration of the mentioned challenges.

As further research, we plan partial re-work of the SDBC approach, in the direction of service-orientation and adaptability.

References

1. Encyclopedia Britannica (Home), <http://www.britannica.com>
2. Software Engineering Institute (Home), <http://www.sei.cmu.edu>
3. Shishkov, B., Dietz, J. L. G., Liu, K.: Bridging the Language-Action Perspective and Organizational Semiotics in SDBC. In ICEIS'06, 8th Int. Conf. on Enterpr. Inf. Systems (2006)
4. Liu, K.: Semiotics in Information Systems Engineering. Cambridge University Press, Cambridge (2000)
5. Shishkov, B.: Software Specification Based on Re-usable Business Components. Delft, The Netherlands: Sieca Repro (2005)

6. Dietz, J.: Enterprise Ontology, Theory and Methodology. Springer-Verlag, Berlin Heidelberg (2006)
7. D'Souza, D. F. and A.C. Willis: Object, Components, and Frameworks with UML, the Catalysis Approach. Addison-Wesley, USA (1998)
8. Tropos Project (Home). <http://www.troposproject.com>
9. Object Management Group (OMG): MDA – Guide, V1.0.1, omg/03-06-01 (2003)
10. Papazoglou, M. P.: Web Services: Principles and Technology. Pearson Pr. Hall (2008)
11. Shishkov. B. and Van Sinderen, M. J.: Service-Oriented Coordination Platform for Technology-Enhanced Learning. In: I-WEST'09, 3rd International Workshop on Enterprise Systems and Technology. INSTICC Press (2009)
12. Leymann, F.: Combining Web Services and the Grid: Towards Adaptive Enterprise Applications. CAiSE Workshops (2005)
13. Van Sinderen, M. J.: From Service-Oriented Architecture to Service-Oriented Enterprise. In: I-WEST'09, 3rd International Workshop on Enterprise Systems and Technology (2009)
14. Shishkov. B. and Van Sinderen, M. J.: On the Design of Context-Aware Applications. In: I-WEST'08, 2nd Int. Workshop on Enterprise Systems and Technology. INSTICC Press (2008)
15. Alonso, G., Casati, F., Kuno, H., Machiraju, V.: Web Services, Concepts, Architectures and Applications. Springer-Verlag, Berlin-Heidelberg (2004)
16. Bosworth, A.: Developing Web Services. In: 17th Int. Conference on Data Engineering (2001)
17. Pasley, J.: How BPEL and SOA are Changing Web Services Development. Internet Computing, IEEE (2005)

Service-Oriented Architecture for Household Energy Management

Marten van Sinderen¹, Martijn Warnier² and Boris Shishkov³

¹University of Twente, Enschede, The Netherlands

²Delft University of Technology, Delft, The Netherlands

³IICREST, 53 Iv. Susanin Str., Sofia, Bulgaria

m.j.vansinderen@utwente.nl, M.E.Warnier@tudelft.nl
b.b.shishkov@iicrest.eu

Abstract. Recent advances in pervasive computing foster automated systems to support household energy management. Smart devices can be used to monitor both the consumption of household appliances and the presence and activity of people in the household. Based on monitoring results, intelligent feedback to residents and intelligent control of appliances is possible. In this paper we present a service-oriented architecture for household energy management in order to cope with interoperability and flexibility issues that exist in the home environment. We also propose a home service bus that realizes core properties of the service-oriented architecture, and thus facilitates integration of existing solutions and development of new applications.

Keywords. Household energy management, Energy consumption monitoring, Context monitoring, Home service bus, Service-oriented architecture (SOA).

1 Introduction

Technological developments such as home networking, sensor networks and pervasive computing [45, 57, 21, 44] have provided many possibilities to apply ICT in the home environment to support needs and desires of the residents. In this paper we focus on ICT support for household energy management, however we explicitly consider extension of proposed solutions to other application areas.

This paper builds on the work presented in [48] and is a companion paper of [52]. In [48] the authors present a new approach for household energy management. The approach consists of an explanation of how results from autonomic and context-aware computing can be applied to contribute to the realization of certain defined objectives for energy management. This leads to the identification of behavioral patterns and key components, constituting a basic architecture for household energy management.

The energy management system has to base its management decisions on information gathered from controlled appliances and from context sources. The representation of this information as knowledge with which the system can adaptively reason is outside the scope of [48] but is further explored in [52].

This paper extends [48] in yet another direction. It reconsiders the basic architecture and extends this architecture based on requirements such as

interoperability and flexibility. The resulting architecture is service-oriented, and features programmable process and decision services for monitoring and controlling appliances. We claim that the architecture is generic enough to be specialized and extended to support other home/household application areas such as home automation and homecare. For this purpose we propose a home service bus as a common service-oriented backbone.

The contribution of the paper lies in the exploration of service-oriented architecture solutions for smart home applications in general and household energy management in particular. We discuss several architectural patterns that exploit the service-oriented paradigm and we show how heterogeneous technology solutions can be integrated.

The remainder of this paper is structured as follows: Section 2 present the motivation for working on household energy management and presents related work. Section 3 summarizes the basic system architecture introduced in [48]. Section 4 discusses a basic service-oriented architecture for household energy management system. This architecture is derived from the basic system architecture, considering interoperability and flexibility requirements. This section also briefly explores possibilities to extend the architecture for multi-level energy management, which will be relevant if household energy management systems are being federated. Section 5 considers high level goals that drive the decision-making behavior of the household energy management system. Section 6 discusses the possibility of employing a home service bus to realize key properties of the service-oriented architecture. Finally, Section 7 presents our main conclusions and outlook.

2 Motivation and Related Work

The motivation to work on household energy management emerges from the general concerns about environmental load of consumption, and resulting policies to achieve energy preservation, i.e. using energy efficiently and reducing energy consumption [24]. We briefly discuss efforts and methods to achieve energy preservation in the home and then turn to technology-based solutions that can lead to energy efficient smart homes.

2.1 Energy Conservation

Household energy use accounts for a significant portion of the total energy footprint of OECD countries, and reducing the energy consumption and emissions from homes and communities is therefore considered a cost effective way to contribute to a sustainable society [16]. Three general routes have been identified for reducing energy consumption in the residential sector [55]:

1. Promote low-energy buildings;
2. Promote energy-efficient domestic equipment;
3. Promote energy conscious behavior of end-users.

We still know little about the behavior of end-users: how they manage their energy consumption, and how they can be motivated to change to a more energy conscious

behavior [7]. A review on the effectiveness of interventions showed that rewards are most effective in achieving energy conservation, although the effect may be short-lived, followed by frequent feedback [2]. A study on feedback revealed the key features for being successful: feedback should be frequent, persistent and appliance-specific, in addition to being presented in a clear and appealing way [23].

Encouraging energy conservation in the home is not easy, primarily because the impact and amount of energy use are not readily apparent. The following measures are considered in order to improve on this situation [5]:

- **Energy labeling:** Providing energy labels for white goods and appliances is well-established, and is an important point-of-sale source of consumer information on comparative lifecycle energy use.
- **Immediate feedback:** Energy monitoring using smart meters and advanced user interfaces is an obvious point-of-use source of consumer information on daily energy use. Some authors have claimed up to 20% reduction in total domestic energy use, the precise amount depending on the content and method of the feedback [55, 11, 23]. Recent work [12] indicates that actual implementation of smart metering at the household level is still in its infancy and hard evidence on what can be achieved by it is scarce.
- **Carbon labeling:** Instead of providing information on the lifecycle energy use of equipment, it would be more appropriate to give information on the total embodied energy and carbon content of a product. This so-called carbon labeling would entail consideration of energy required and emissions induced for production, packaging, transportation, storage as well as lifecycle use. Although attractive from an energy conservation point of view, mandatory carbon labeling is probably hard to achieve because of opposition from industry groups [5].

Household energy consumption varies dramatically with time of day and time of year [55]. The electricity system would be able to maximize its efficiency if smoother or more predictable electrical demand profiles are available. For this reason, interest areas also include:

- **Energy use components and patterns:** Domestic electricity use is dominated by heating our house and our water [16]. A report of the US Department of Energy [19] gave the following contributions to household energy use: appliances 30%, air condition, heating and ventilation 31%, and water heating 9%. Components of energy use may be roughly classified as predictable, moderately predictable and unpredictable [55]. Furthermore, energy use patterns are different for different types of equipment (e.g., for lighting, cooking, heating), and overall energy use patterns are different for different household compositions (e.g., single, couple, family with children) [30]. Despite these differences, clear peak times of energy use exist, which place high demands on the energy grid. The possibilities to use or change energy use patterns in order to smoothen energy use are limited. Most promising are multi-user systems [30] which can rely on the more predictable behavior of a community, compensate for irregular behavior of individuals, and time shift individual demands in order to smoothen the overall consumption peaks of the community.
- **Green micro-generation:** Domestic generation of energy with renewable sources, such as sun, wind and thermal heat, has many advantages. Heat and electricity can be produced locally, which may be used for immediate daily energy needs or locally

captured and used during (peak) energy demand periods. Alternatively, surplus power from micro-generation may be sold to the local power company and injected in the grid [15]. Use of fossil fuels for centralized energy production is thus decreased, losses due to centralized production and transportation over the grid are reduced, and strain on the energy grid during peak periods is lowered. Besides the environmental benefits, consumers also have the financial gain of saving on energy costs or selling to the power company. However, despite the potential benefits, there are still substantial barriers to the large-scale use of micro-generation [4].

2.1 Pervasive Computing and Smart Homes

Household energy management has been addressed by several researchers in the smart homes domain. The smart homes concept, also called home automation or domotics, is about the integrated application of technologies that help to improve the quality of life of the residents of the home [31, 40]. Energy management is just one of the potential application areas of smart homes, next to controlling household appliances and multi-media equipment, providing assistance to elderly and disabled, and increasing safety and security. The technologies considered here have evolved from advances in pervasive computing, also known as ubiquitous computing or ambient intelligence [53, 44, 1]. Important topics in this domain, which also have high relevance to household energy management, include:

- **Sensors and actuators:** RFID and other pervasive technologies [21, 3] enabled the development of miniature sensors and actuators with communication capabilities. These devices can be embedded in our environment, and integrated in equipment, daily objects, cloths or even our body. Thanks to their communication capabilities, they can be identified, traced and sometimes remotely controlled using other computers.
- **Context-awareness:** Sensors that measure physical properties of an environment can communicate sensor data across the home network. Based on aggregation of such data and logical inference, so-called context information can be derived which is relevant to end-user applications [13]. The latter may adapt their behavior to be more efficient or effective, given the perceived or predicted context status of their environment. In particular, applications can be more useful to end-users if provided services can be personalized to the changing situation and corresponding needs of individual users [39, 17]. Personalization can be in terms of functionality or content provided by a service, or in terms of time, form or location of its delivery. Use of automatically derived context information, instead of explicit user input, ensures that the personalization is non-intrusive to the end-user. User patterns or habits may be learned by analyzing historical context information obtained over longer time periods, and matching real-time context information with these patterns allows pro-active context-aware behavior [49].
- **Network infrastructure and interoperability:** A smart home will comprise various components, which may have been acquired at different times, supplied by different vendors, and developed by different manufacturers. This potential heterogeneity of components, combined with the diversity of physical media solutions for home environments, poses a serious interconnection and interoperability problem [18, 28].

Most existing houses were built without communication infrastructure, so they have to be retrofitted with new network technology. Powerline solutions and wireless networks present retrofits which are easy to install, flexible and relatively cheap, compared to dedicated wiring, although they are limited with respect to bandwidth and reliability. On top of low-level protocols for network connectivity, several protocols exist to communicate content and control devices. These protocols can be grouped in application areas, such as entertainment, lighting, appliances, and climate control, effectively forming different home networks [50]. In order to achieve integration of these different home networks and associated applications, middleware and service-oriented approaches have been proposed [50, 56].

- **Smart applications:** The previous topics are all necessary to enable the *raison d'être* of smart homes, namely to provide smart applications. Smart applications can increase comfort, interest or fun for users, without requiring explicit management or control from users. This means that a smart home system should be able to monitor itself or its environment, interpret monitoring results with respect to user goals, and decide on actions that would contribute to attaining the user goals. Several solutions exist for learning, knowledge representation, knowledge reasoning, and decision making, which have been applied in various smart applications, including inhabitant-aware home automation [49, 8], load balancing on the electric grid [26], reactive and proactive household energy control [34], and location aware resource management [43].

Despite the considerable progress made in pervasive computing, smart home solutions have not yet been massively deployed. Main reasons are complex installation and interfaces paired with lack of interoperability [28]. Moreover, many proposed smart home systems enable control over devices and artifacts, rather than promoting control over the lives of people. The latter has been identified as a key factor of success [14]. Such control over lives may be used to pursue ‘calm living’, although [42] claims that success of pervasive computing is more likely if it is used to engage people, to provoke them to be creative and active.

An important trigger for starting a smart home revolution may come from government policies to install smart energy meters into the home nationwide [16]. From this situation, energy conservation in the home and energy selling to the grid may evolve as potential killer applications.

3 Basic System Architecture

Figure 1 illustrates the basic architecture for household energy management, as introduced in [48]. It shows the controlled domain with household appliances which are being monitored and/or which can be controlled concerning energy consumption. For this purpose, each appliance is instrumented with sensors and/or actuators. The controlled domain also contains context sensors embedded in the home environment. One type of context sensors is used to measure physical properties of the environment such as temperature, humidity and air quality. Another type of sensors can be used to determine the presence, location, and activity of the residents, and be used in

combination with software applications such as an electronic agenda to provide supplementary information for reasoning.

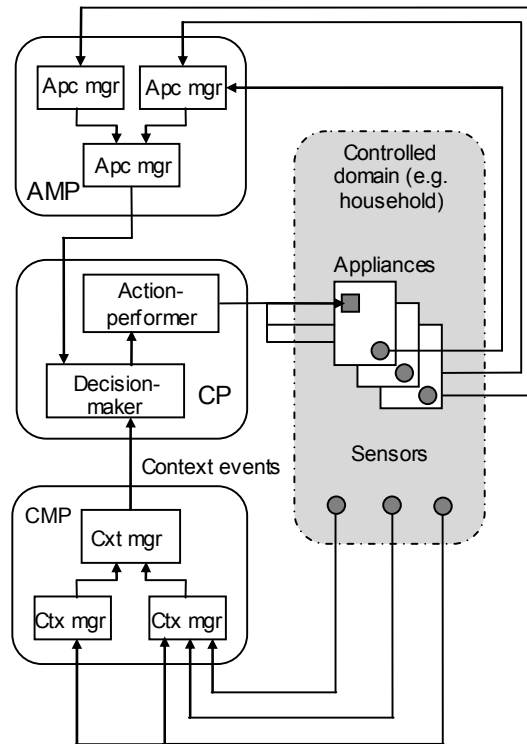


Fig. 1. Basic architecture for household energy management (adapted from [48]).

The monitored energy consumption of several appliances may be collected by an appliance manager (*Apc mgr* in Fig. 1) in order to get useful insights and derive consumption aggregates for certain categories of appliances. Multiple appliance managers may be organized in a tree structure with a root/top appliance manager that produces the right-level consumption information on basis of which control decisions can be made. The coordination of tasks of appliance managers is called the Appliances Management Process (*AMP*).

Similarly, raw context data of one or more context sensors may be collected by a context manager (*Cxt mgr* in Fig. 1) in order to derive more reliable or higher-level context information. Through a process of context reasoning involving one or more context managers the right-level context information is derived which can be used for taking control decisions. The coordination of tasks of context managers is called the Context Management Process (*CMP*).

Both the results of the *AMP* and *CMP* are used to take control decisions. For example, if the *AMP* indicates that the consumption is approaching a previously set maximum level, certain appliances may be temporarily switched off. However, if the *CMP* indicates that residents are in the house then no actions should be taken that would go against a previously set comfort level. Control decisions are based on a set

of rules that aim at realizing end-user goals by proposing actions based on a view of the currently evolving situation. End-user goals are for example reducing energy consumption and maintaining a certain comfort level. The view of the situation is based on information received from the *AMP* and *CMP*. Possibly also learned patterns can be considered. If an evolving situation matches such a pattern, the actions defined for that pattern will be proposed. Taking control decisions is the responsibility of a component called *Decision-maker*. Another component, called *Action-performer*, translates the proposed actions into operations to be performed on selected appliances. For example, the *Action-performer* may send a request to the actuator of the freezer to temporarily power down.

4 Basic Service-oriented Architecture

The basic architecture of the household energy management system identifies components and their interconnection structure but it does not consider interoperability and flexibility issues. One can expect that devices for monitoring and control will be offered by many different vendors. These devices will have widely varying capabilities and use different communication solutions, depending on the type of appliance or environment in which they will be embedded. A similar reasoning applies to software components, which may employ many different algorithms for reasoning and decision making. Developing a dedicated interoperability solution for a given set of heterogeneous components is hard, and most likely results in a system with no or limited possibilities to change the interconnection structure, replace or add components, and use components' functions in various combinations and orders. The latter may be an important drawback if the system is to be deployed in different types of homes and households and has to be maintained under evolving user requirements and technology developments.

We adopt a service-oriented architecture in order to cope with the problems of heterogeneity and rigidity [20, 37]. This means, among others, that we assume that components expose their functionality through service interfaces, and can be accessed accordingly using applicable Web service standards [38]. Proprietary solutions should then be wrapped such that functionality is appropriately translated between standard-based public service interfaces and internal technology-specific interfaces.

4.1 Overview of Services

We initially identify the following services in our basic service-oriented architecture for a household energy management system:

- Consumption (*Csn*) *sensor* service: This is a service provided by a (wrapped) sensor attached to or embedded in a household appliance for measuring energy consumption. It supports either a notification or a request-response message exchange. In the former case a notification is sent to interested clients at regular intervals or in case of relevant events. The notification contains measurement values. Relevant events may be raised on crossing consumption thresholds or starting/ending thermostatically triggered activity cycles. The request-response message exchange is

initiated by interested clients, which send a request to the service and ask measurement values to be returned in a response.

- Power actuator (*Pwr actr*) service: This is a service provided by a (wrapped) actuator attached to or embedded in a household appliance for power management. This means that the actuator can be instructed to perform a power on/off operation or switch to another power mode. The service supports a one-way message exchange with which a client can invoke one of the defined operations on the service.
- Context (*Cxt*) *sensor* service: This is a service provided by a (wrapped) sensor embedded in the home environment for measuring a context attribute. It supports either a notification or a request-response message exchange. In the former case a notification is sent to interested clients at regular intervals or in case of relevant events. The notification contains measurement values or a value identifying the event type. Relevant events may be raised on crossing context attribute value thresholds such as a critical temperature, or a step change in the context attribute value such as a person entering a room. The request-response message exchange is initiated by interested clients, which send a request to the service and ask measurement values or event indications to be returned in a response.
- Appliance Management Process (*AMP*) service: This service provides consumption information to the *CP* service (see below). For this purpose it supports either a notification or a request-response message exchange, to notify the *AP* service or to respond to a request from the *AP* service, respectively. The consumption information passed to the *AP* service has the right scope and content to take control decisions. The *AMP* service is also a client of one or more *Csn sensor* services. Using the latter services, it collects measurements values concerning energy consumption of controlled appliances, which it uses to derive the higher-level consumption information aggregates.
- Context Management Process (*CMP*) service: This service provides context information to the *CP* service (see below). For this purpose it supports either a notification or a request-response message exchange, to notify the *AP* service or to respond to a request from the *AP* service, respectively. The context information has the right semantic level to take control actions. The *CMP* service is also a client of one or more *Cxt sensor* services. Using the latter services, it collects measurement values or event notifications concerning the context of the residents, which it uses to derive the higher-level context information aggregates.
- Control Process (*CP*) service: This service provides an interface to the end-users for programming the household energy management system. Programming may consist of specifying simple workflows or rules, or choosing between alternative pre-defined workflows or rules and possibly supplying values for parameters. The *CP* service is a client of both the *AMP* service and the *CMP* service, which provide the necessary information to take control decisions. In addition, it is a client of one or more *Pwr actr* services, which can be instructed to perform certain operations in accordance to the decisions.

Fig. 2 illustrates the basic service-oriented architecture, using an ad-hoc notation to distinguish between client ports (to invoke operations on an external service) and service ports (to offer operations which can be invoked by external clients).

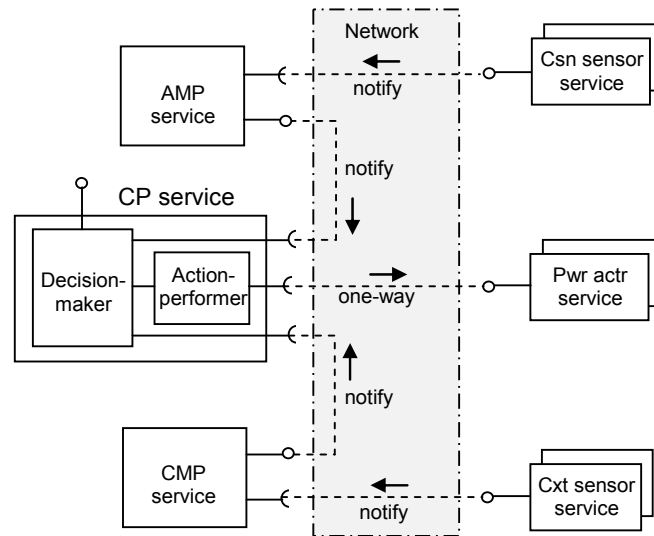


Fig. 2. Basic service-oriented architecture.

4.2 Service Granularity and Service Hierarchies

According to the basic system architecture described in the previous section, *AMP*, *CMP* and *CP* are processes each involving (potentially) multiple components. This brings us to the question whether we want to consider the services corresponding to these processes as individual services or as composite services. Each ‘process’ service may either be provided by an integrated implementation of the components identified in the process or it may be provided by an orchestration of ‘component’ services. To answer this question one has to consider what is the right granularity of services [27].

We foresee that in case the controlled domain is a single household, there is no need to have the processes implemented as an orchestration of separate services corresponding to the identified components. The reason for this is that the processes are probably not physically distributed and there is no opportunity to reuse the functionality of the components as services. On the other hand, in case the controlled domain covers a larger geographical area such as an apartment building or a city block, which contains a large number of controlled appliances, the use of composite services might be a good idea.

Several alternatives exist to structure the household energy management system in terms of services when we consider multiple levels or hierarchical domains of energy management. The motivation for considering this is twofold. First of all, by having a system that can coordinate individual household energy management systems, more opportunities exist to avoid peaks and to balance demand and supply (in case of micro-generation). This is particularly interesting to the local power company. Secondly, by organizing households into a collective, leverage is created for individual households to negotiate better prices.

We briefly explore two alternatives for coordinating individual household management systems. These alternatives are not necessarily optimal or practical, but

illustrate the spectrum of possibilities for coordination. We defer a more thorough treatment of this topic as future work. Let us assume we have two levels of energy management, say an apartment level and an apartment building level. The first alternative assigns to the highest level considerable responsibility and authority for direct control based on the overall situation (see Fig. 3):

- Each apartment has a single appliance manager and a single context manager, which report to an appliance manager and a context manager at the apartment building level, respectively. The top-level managers inform the *CP* at the apartment building level. This *CP* has a decision-maker that analyzes consumption and context, and makes decision proposals. The decision proposals are forwarded to the apartment building's action-performer, which propagates decision proposals in an appropriate way to action-performers at the apartment level.

The second alternative assigns to the highest level no responsibility and authority for direct control, but instead allows changing the policy for control based on the overall situation (see Fig. 4):

- Each apartment has its own *AMP* and *CMP*, which report to the apartment's *CP*. Each apartment's *CP* makes its own analysis with the decision-maker and controls its own appliances accordingly with the action-performer. However, each apartment's *CP* also forwards the notifications of the *AMP* and *CMP* (or less frequent summary reports) to the *CP* at the apartment building level. The latter *CP* has another task and structure than the *CP*s at the apartment level. It analyzes the consumption and context information with the objective to decide whether criteria for decision rules at the apartment level should be modified. If so, it informs selected *CP*s at the apartment level of the new criteria.

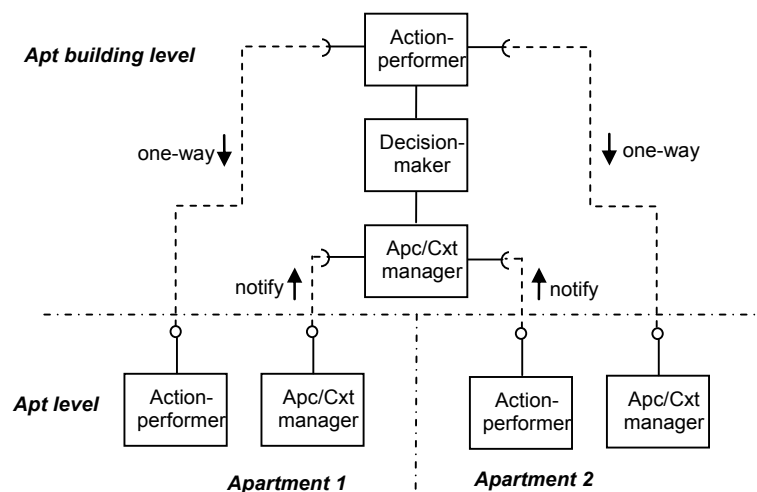


Fig. 3. Multi-level energy management – alternative 1.

The second alternative resembles a reflective architecture as has been proposed for some middleware systems [9]. A reflective system offers a separation of concerns and provides inspection and adaptation of its own behavior [32]. These properties would

facilitate dynamic adaptation [6] and self-healing [35], as is generally required in wireless settings with mobile and context-aware applications.

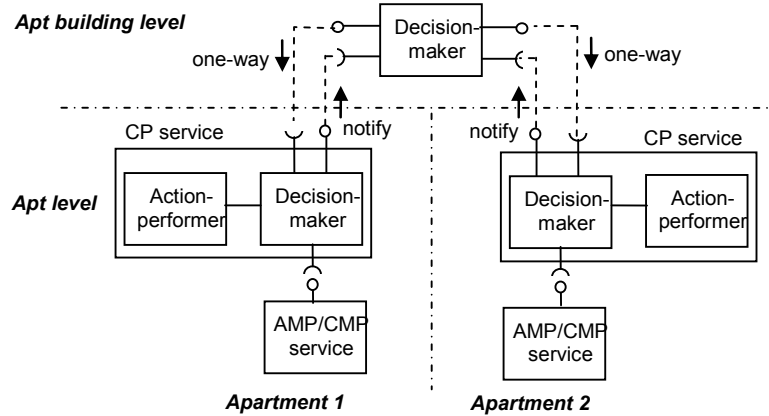


Fig. 4. Multi-level energy management – alternative 2.

In our case of household energy management, two meta-levels of reflection can be recognized: (i) at the first level, appliances are monitored, and decisions are made to change the consumption behavior of appliances based on some management policy model; and (ii) at the second level, households are monitored, and decisions are made to change the energy management policy of households based on some performance goal model. Fig.5 shows these levels. In this paper, we will not further explore the exploitation of reflection for our service-oriented architecture, but leave this for future work.

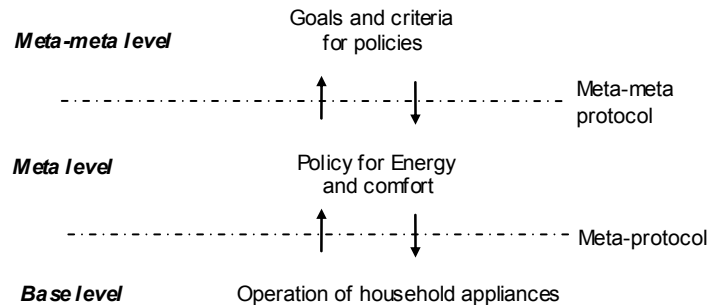


Fig. 5. Reflective levels in energy management.

5 Goal-driven Control Process

The control process that provides the *CP* service and comprises the *Decision-maker* and *Action-performer* (see Fig. 2) is driven by end-user goals and realized by the execution of a rule-based policy that fulfils these goals. The overall goal that we project is a combination of energy and comfort conservation. Since the components in

this goal are to some extent conflicting, and their weights may be case-specific as well as dependent on the stakeholders, balancing these components is very challenging. We can distinguish between two stakeholder types: the resident or house owner, typically the consumer of energy (if we disregard micro-generation), and the power company, the main producer of energy. Their interests can be characterized as follows (see for example also [54]):

- Producer: (i) keep peaks below what can be supported, as power outages and power quality fluctuations lead to consumer complaints; (ii) minimize overcapacity and avoid large variations in time, as this would otherwise require expensive dimensioning of the grid; and (iii) minimize supply and waste, for environmental reasons.
- Consumer: (i) minimize local consumption, for economic and environmental reasons; and (ii) optimize local comfort, for consumer-selfish reasons.

Underlying the rule-based policy is a knowledge representation that allows the selection and evaluation of rules. We do not discuss here the requirements that hold for this knowledge representation and how the knowledge base can be up to date. For this the interested reader is referred to [52]. Instead, we end this section with some (loosely formulated) examples of rules that could be used by a control process:

1. If a room is not used, then any device in the room from the list (light, heating, air conditioning, entertainment equipment, etc.) can be switched off, unless somebody has indicated that it should stay on.
2. During low consumption periods, battery-powered devices (such as electric cars, household robots, battery-powered appliances and tools) can be charged as far as they need charging.
3. During peak consumption periods, thermostatically controlled devices in the house (such as room heating, air conditioning, refrigerator, freezer) can be operated in interleave or low-power mode.

6 Home Service Bus

The service-oriented architecture depicted in Fig. 2 comprises a network component that provides the connection infrastructure for the services. We assumed that this component is capable of realizing the message exchange between service requestors (clients) and service providers (services), using available network technologies. However, we may attribute additional generic functionality to this infrastructure, thus relieving individual service developers from the task of implementing such functionality over and over again. This idea is similar to that of enterprise service bus [33, 46], and we therefore call this infrastructure the home service bus. Presumably, a home service bus will have similar functionality as an enterprise service bus, but would favor different implementations because of the specific requirements of home applications and the characteristics of home networks.

The home service bus plays a key role in realizing the service-oriented architecture. First of all, it can provide the register-find-bind functionality which is one of the central promises of service-oriented architecture [10, 20]. And secondly, it

can provide a virtual operational environment to service requestors. We will briefly expand on these functions.

The home service bus manages a service directory which contains relevant information on services. The information on a service is added to the directory if the service is registered. This can be done automatically, when the service is deployed, or after deployment when it is discovered by the bus. The information covers both application aspects of the service as well as network aspects of the service endpoint. A service requestor can find a service it is interested in by enquiring the directory. The directory uses the application related information to match registered services with capabilities required by the service requestor, and selects the best service from the qualified ones by using run-time operational information. The service requestor can then bind to the selected service using the network information on the service endpoint.

Effectively, a service requestor can find and subsequently use a service without needing to know the technical details of this service and that the service may be just one of many qualified services. The set of qualified services represents the virtual service in which the service requestor is interested. The home service bus hides internal implementation details of the services, such as the used programming language, runtime environment and hardware platform, but also allows the service requestor to interact with the selected service without knowledge of the network address or the communication protocol. Several kinds of virtualization or mediation functions may be provided by the service bus in order to cope with the heterogeneity of service implementations. For example, wrapping and transformation functions are needed if peer components do not share a common communication protocol or message format, as can be expected in the heterogeneous home environment. Some services may reside on a home gateway or server system, where Web services for communication (SOAP) and description (WSDL) can be supported. However, resource-limited devices, such as sensors and actuators, have to use product type specific industry standards. Therefore, the home service bus has to support such standards, and internally apply proper interface wrapping and/or protocol translation in order to make these devices appear as web-enabled devices to components that natively support web services. Additional functions are also possible, including but not limited to:

- **Routing:** Messages may be routed to a particular system in case of error situations, or to a preferred service or set of services based on workload or availability. For example, service selection for the virtual service may be done on a per-request basis, using request-time operational and possibly context information.
- **Publish/subscribe:** Service requestors may subscribe to certain responses or events with a single subscribe request to the service bus. The service bus then takes care of notifying the requestor of responses or events according to the specification in the subscribe request, until the subscription is withdrawn. For example, energy consumption measurement values or events from consumption sensors can be delivered this way to appliance management processes.
- **Booking:** Service requestors may book certain services which are not immediately available or not immediately needed. The service bus schedules provisioning of booked services to requestors based on priorities or overall performance criteria. For example, energy supply can be considered as a service and smart appliances may

request certain energy-hungry tasks to be performed, but leave the timing to the bus.

Fig. 6 shows the discussed functionality of the home service bus as infrastructure services.

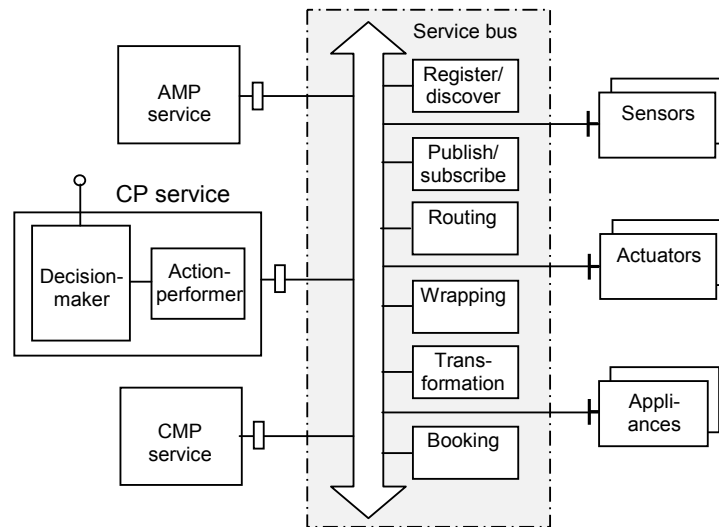


Fig. 6. Home service bus as mediator between services relevant for household energy management.

It is interesting to consider how the home service bus can be realized using various existing industry standards, and what alternatives exist for connection infrastructures that support service-oriented architecture. Relevant standards include Open Service Gateway initiative (OSGi), European Installation Bus (EHS), Home Audio Video interoperability (HAVi), Universal Plug and Play (UPnP), and Jini. In particular, several projects have explored the use of OSGi in combination with other standards for the smart home [50, 40, 56, 25].

Another interesting question concerns the right functionality of the home service bus, considering the different application areas that exist for the smart home domain. Many research projects have focused on the development of a network infrastructure or middleware for the smart home with one specific application area in mind, such as infotainment and home automation [41], elderly care and independent living [51, 36, 22], and energy management [29, 47]. This led to different proposed solutions. However, it is unlikely that multiple home service buses will co-exist in the same home. The problem is that some functionality may be generic in one application area, but may not be used in another area. Another problem is that different application areas will have their own industry specific standards. Finding a cost-effective compromise on which functions to include and which standards to support is therefore an important challenge

7 Conclusions

Household energy consumption accounts for a significant portion of most nations' overall consumption. It is therefore worthwhile to investigate possibilities of reducing energy consumption of households through effective energy management. To be effective, household energy management systems should promote energy conscious behavior of end-users. This is not easy to achieve, since we still know little about the behavior of end-users. Some studies have shown that providing immediate feedback to residents on consumption and related costs can have a positive impact on user behavior and lead to considerable energy savings. Advances in pervasive computing have enabled smart home applications, with energy management as one of the targeted application areas. For example, this development led to affordable technical solutions for providing immediate feedback through smart metering and advanced user interfaces. Smart home applications for energy management can also be used to automate part of the energy conscious behavior of people, namely by making such applications responsible for realizing comfort and consumption (c.q. reduction) targets that have been formulated by the residents.

Automated systems to support household energy management consist of heterogeneous components, are based on various industry standards, and are subject to evolving user requirements and technology developments. We claim that adoption of a service-oriented architecture for household energy management systems is particularly useful to cope with interoperability and flexibility issues. We illustrate the service-oriented approach for a system that is able to monitor both the consumption of household appliances and the presence and activity of residents, and that accordingly takes 'smart' decisions on when and where to intervene in the consumption of appliances. We further discuss the role of a home service bus to realize some of the key properties of a service-oriented architecture.

Several challenges and topics for future work have been identified. One major challenge is the federation of multiple household energy management systems in order to exploit economies of scale. The federated system would have two or more hierarchical levels of management, and it is as yet unclear which architecture would best support multi-home energy conservation opportunities while sufficiently ensuring autonomy of individual households. We also like to extend our household energy management system with micro-generation. Especially, it is interesting to consider the possibility of dynamically matching demand and supply of energy in combination with buying energy from and selling energy to the electricity grid. Another major challenge is finding an acceptable architecture and technical configuration of the home service bus such that it can act as the service-oriented backbone for all application areas in the smart home.

References

1. Aarts, E., Wichert, R.: Ambient Intelligence. In: Technology Guide, pp. 244-249. Springer (2009)
2. Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T.: A Review of Intervention Studies Aimed at Household Energy Conservation. *Journal of Environmental Psychology* 25: 273-

- 291 (2005)
3. Ahuja, S., Potti, P.: An Introduction to RFID Technology. *Communications and Network* 2 (3): 183-186 (2010)
 4. Allen, S. R., Hammond, G. P., McManus, M. C.: Prospects for and Barriers to Domestic Micro-generation: A United Kingdom Perspective. *Applied Energy* 85 (6): 528--544 (2007)
 5. Burgess, J., Nye, M.: Re-materialising Energy Use through Transparent Monitoring Systems. *Energy Policy* 36: 4454-4459 (2008)
 6. Capra, L., Emmerich, W., Mascolo, C.: CARISMA: Context-aware Reflective Middleware for Mobile Applications. *IEEE Transactions on Software Engineering* 29 (10): 929--945 (2003)
 7. Chetty, M, Tran, D., Grinter, R.E.: Getting to Green: Understanding Resource Consumption in the Home. In: 10th International Conference on Ubiquitous Computing, UbiComp, pp. 242-251. ACM Press (2008)
 8. Cook, D.J., Youngblood, M., Das, S.K.: A Multi-agent Approach to Controlling a Smart Environment. In: *Designing Smart Homes – The Role of Artificial Intelligence*, pp. 165-182. LNCS 4008, Springer-Verlag (2006)
 9. Costa, F. M., et al.: The Role of Reflective Middleware in Supporting the Engineering of Dynamic Applications. In: *Reflection and Software Engineering*, pp. 79-98. LNCS 1826, Springer-Verlag (2000)
 10. Curbera, F. et al.: Unraveling the Web Services Web - An Introduction to SOAP, WSDL, and UDDI. *IEEE Internet Computing* 6 (2): 86-93 (2002)
 11. Darby, S.: The Effectiveness of Feedback on Energy Consumption – A Review for DEFRA of the Literature on Metering, Billing and Direct Displays. Environmental Change Institute, University of Oxford (2006)
 12. Darby, S.: Smart Metering: What Potential for Householder Engagement? *Building Research & Information* 38 (5): 442-457 (2010)
 13. Dey, A. K.: Understanding and Using Context. *Personal and Ubiquitous Computing Journal* 5 (1): 4-7 (2001)
 14. Davidoff, S., et al.: Principles of Smart Home Control. In: 8th International Conference on Ubiquitous Computing, UbiComp, pp. 19-34. LNCS 4206, Springer-Verlag (2006)
 15. De Silva, L. C., Methew, S.: Energy Efficient Smart Homes. In: 1st International Conference of the IET Brunei Darussalam Network, pp. P82: 1-4 (2008)
 16. DEFRA – Department for Environment, Food and Rural Affairs: Climate Change: Taking Action (2010), <http://www.defra.gov.uk/environment/climate/documents/taking-action.pdf>
 17. Dockhorn Costa, P., Ferreira Pires, L., Sinderen, M. van: Designing a Configurable Platform for Mobile Context-aware Applications. *International Journal of Pervasive Computing and Communications* 1 (1): 13-25 (2005)
 18. Edwards, W.K., Grinter, R.: At Home with Ubiquitous Computing: Seven Challenges. In: *International Conference on Ubiquitous Computing, UbiComp*, pp. 256--272. LNCS 2201, Springer-Verlag (2001)
 19. Energy Information Administration: Reginal Energy Profile - U.S. Household Electricity Report (2005), http://www.eia.doe.gov/emeu/reps/enduse/er01_us.html
 20. Erl, T.: *Service-oriented Architecture: Concepts, Technology, and Design*. Prentice Hall (2005)
 21. Estrin, D., Culler, D., Pister, K., Sukhatme, G.: Connecting the Physical World with Pervasive Networks. *Pervasive Computing* 1 (1): 59--69 (2002)
 22. Farella, E., Falavigna, M., Ricco, B.: Aware and Smart Environments: The Casattenta Project. *Microelectronics Journal* 41 (11): 697--702 (2010)
 23. Fischer, C.: Feedback on Household Electricity Consumption: A Tool for Saving Energy? *Energy Efficiency* 1: 79-104 (2008)
 24. Geller, H. et al.: Policies for Increasing Energy Efficiency: Thirty Years of Experience in OECD Countris. *Energy Policy* 34 (5): 556--573 (2006)

25. Gouin-Vallerand, C., Giroux, S.: Management and Deployment of Applications with OSGi in the Context of Smart Homes. In: 3rd International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob. IEEE Computer Society (2007)
26. Gustavsson, R.: Agents with Power. *Communications of the ACM* 42 (3): 41--47 (1999)
27. Haesen, R., Snoeck, M., Lemahieu, W., Poelmans, S.: On the Definition of Service Granularity and Its Architectural Impact. In: *Advanced Information Systems Engineering*, pp. 375--389. LNCS 5074, Springer-Verlag (2008)
28. Holroyd, P., Watten, P., Newbury, P.: Why Is My Home Not Smart? In: 8th International Conference on Smart Homes and Health Telematics, ICOST, pp. 53--59. LNCS 6159, Springer-Verlag (2010)
29. Inoue, M., Higuma, T., Ito, Y., Kushiro, N., Kubota, H.: Network Architecture for Home Energy Management System. *IEEE Transactions on Consumer Electronics* 49 (3): 606--613 (2003)
30. Jenny, A., Díaz López, J. R., Mosler, H.-J.: Household Energy Use Patterns and Social Organisation for Optimal Energy Management in a Multi-user Solar Energy System. *Progress in Photovoltaics: Research and Applications* 14: 352--362 (2006)
31. Jiang, L., Liu, D.-Y., Yang, B.: Smart Home Research. In: 3rd International Conference on Machine Learning and Cybernetics, pp. 659--663 (2004)
32. Kon, F., Costa, F., Blair, G., Campbell, R.H.: The Case for Reflective Middleware. *Communications of the ACM* 45 (6): 33--38 (2002)
33. Leymann, F.: Web Services: Distributed Applications without Limits. In: *Database Systems for Business, Technology and Web*, BTW (2003)
34. Long Ha, D., Ploix, S., Zamai, E., Jocomino, M.: A Home Automation System to Improve Household Energy Control. In: 12th IFAC Symposium on Information Control Problems in Manufacturing. International Federation on Automatic Control (2006)
35. Mei, H., Huang, G., Tsai, W.-T.: Towards Self-healing Systems via Dependable Architecture and Reflective Middleware. In: 10th IEEE International Workshop on Object-oriented Real-time Dependable Systems, WORDS, pp. 337--344 (2005)
36. Mikalsen, M., et al.: Interoperability Services in the MPOWER Ambient Assisted Living Platform. In: 22nd International Congress of the European Federation for Medical Informatics - Medical Informatics in a United and Healthy Europe, pp. 366--370. IOS Press (2009)
37. O'Brien, L., Merson, P., Bass, L.: Quality Attributes for Service-oriented Architectures. In: International Workshop on Systems Development in SOA Environments, SDSOA (2007)
38. Papazoglou, M. P.: *Web Services: Principles and Technology*. Pearson Prentice Hall (2008)
39. Prekop, P., Burnett, M.: Activities, Context and Ubiquitous Computing. *Computer Communications* 26 (11): 1168--1176 (2003)
40. Ricquebourg, V. et al.: The Smart Home Concept: Our Immediate Future. In: 2006 1st IEEE International Conference on e-Learning in Industrial Electronics, pp. 23-28. IEEE Press (2006)
41. Röcker, C., Janse, M. D., Portolan, N., Streitz, N.: User Requirements for Intelligent Home Environments: A Scenario-driven Approach and Empirical Cross-cultural Study. In: 2005 Joint Conference on Smart Objects and Ambient Intelligence - Innovative Context-aware Services: Usages and Technologies, pp. 111-116. ACM (2005)
42. Rogers, Y.: Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences. In: 8th International Conference on Ubiquitous Computing, UbiComp, pp. 404-421. LNCS 4206, Springer-Verlag (2006)
43. Roy, A., Das, S. K., Basu, K.: A Predictive Framework for Location-aware Resource Management in Smart Homes. *IEEE Transactions on Mobile Computing* 6 (11): 1270-1283 (2007)
44. Saha, D., Mukherjee, A.: Pervasive computing: A paradigm for the 21st century. *IEEE Computer*, 36 (3): 25-31 (2003)

45. Satyanarayanan, M.: Pervasive Computing: Vision and Challenges. *IEEE Personal Communications* 8 (4): 10--17 (2001)
46. Schmidt, M.-T., Hutchison, B., Lambros, P., Phippen, R.: The Enterprise Service Bus: Making Service-oriented Architecture Real. *IBM Systems Journal* 44 (4): 781--797 (2005)
47. Shah, N., Tsai, C.-F., Chao, K.-M.: Monitoring Appliances Sensor Data in Home Environment: Issues and Challenges. In: 11th IEEE Conference on Commerce and Enterprise Computing, CEC, pp. 439--444. IEEE Computer Society (2009)
48. Shishkov, B., Warnier, M., Sinderen, M. van: On the Application of Autonomic and Context-aware Computing to Support Home Energy Management. In: 12th International Conference on Enterprise Information Systems, ICEIS. INSTICC (2010)
49. Si, H., Kawahara, Y., Morikawa, H., Aoyama, T.: A Stochastic Approach for Creating Context-aware Services Based on Context Histories in Smart Home. In: 3rd International Workshop on Exploiting Context Histories in Smart Environments (ECHISE) (2005)
50. Valtchev, D., Frankov, I.: Service Gateway Architecture for a Smart Home. *IEEE Communications Magazine* 40 (4): 126-132 (2002)
51. Wang, F., Turner, K.J.: Towards Personalised Home Care Systems. In: 1st ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA. ACM (2008)
52. Warnier, M., Sinderen, M. van, Shishkov, B.: Adaptive Knowledge Representation for a Self-managing Home Energy Usage System. In: 4th International Workshop on Enterprise Systems and Technology, I-WEST. SciTePress (2010)
53. Weiss, R. J., Craiger, J. P.: Ubiquitous Computing. *The Industrial-organizational Psychologist* 39 (4): 44-52 (2002)
54. Willis, K. G., Garrod, G. D.: Energy Efficiency: Economic Constraints to Environmental Quality. *International Journal of Environmental Technology and Management* 1 (3): 334-337 (2001)
55. Wood, G., Newborough, M.: Dynamic Energy-consumption Indicators for Domestic Appliances: Environment, Behaviour and Design. *Energy and Buildings* 35: 821-841 (2003)
56. Wu, C.-L., Liao, C.-F., Fu, L.-C.: Service-oriented Smart-home Architecture Based on OSGi and Mobile-agent Technology. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Review* 37 (2): 193-205 (2007)
57. Xu, N.: A Survey of Sensor Network Applications. *IEEE Communications Magazine* 40 (8): 102-114 (2002)

Hovering Information: A Self-Organising Infrastructure-Free Information Storage and Retrieval Service

Alfredo A. Villalba Castro¹, Giovanna Di Marzo Serugendo² and Dimitri Konstantas¹

¹Department of Information Systems, University of Geneva
CUI-Bat A, 7 rt de Drize, 1227 Carouge, Switzerland

²Computer Science and Information Systems, Birkbeck (University of London)
Malet Street, London, WC1E 7HX, U.K.

{Alfredo.Villalba, Dimitri.Konstantas}@unige.ch
dimarzo@dcs.bbk.ac.uk

Abstract. This paper introduces the Spatial Memory concept which defines any geographical location as a memory where information can be stored and retrieval by user-applications. The Hovering Information Middleware is also introduced as a constrained implementation of the Spatial Memory concept. This middleware is based on results presented on previous works related with Hovering Information. Two main modules of the middleware are discussed and evaluated, the Storage and the Retrieval Modules. The evaluation is done by simulations using both random-way-point and real mobility patterns. Real mobility patterns are generated with the SLAW real mobility model. The simulations evaluate current storage and retrieval algorithms and outlines the major drawbacks and potential improvements for next milestones.

Keywords. Mobile services, Geolocalisation, Ad-hoc networking.

1 Introduction

Nowadays, we are witnessing the extraordinary progress of technology in terms of miniaturization and communication. Fixed and mainly mobile devices are much smaller and much more powerful than ten years ago. It has become normal for someone walking in the street to carry with him a mobile phone having as much *memory and computing power* as a personal computer of the nineties. Besides this fact, current devices (both fixed and mobile) possess *powerful wireless communication interfaces* which make them able to communicate with distant servers via the Internet or even to communicate between them in an ad hoc manner.

In the last decade, mobile devices as they are becoming more powerful have started playing an important role enabling people to be connected not only between them (phone calls and text messages) but to the Internet and then to all the ad hoc services. In addition, the increasing development of geo-localized services leads us to the design of even more powerful services in mobile devices.

Besides the technological improvements and the undisputed success of the Internet, the scheme of how information is produced, stored and retrieved has been

also evolving from centralized schemas to *community and collective* ones. We are not any more limited to applications working with only one or a group of some powerful servers on which information (databases, videos, profiles,...) is stored. We are now facing services such as peer-to-peer sharing files (kaaza, bittorrent,...), user-produced streaming video (YouTube), and social networking (MySpace, Facebook, LinkedIn,...). These applications are different from classic ones in two points: either they are peer-to-peer and there is no more a centralized server, or the information they manage is produced in a collaborative and continuous way by the users.

However, mobile devices are still playing a passive role since most of them are consumers of information but not producers. It is becoming usual for us to send and receive emails, to browse and to navigate in a city thanks to the geo-localization in-built services, all this in our mobile phone. But still, these applications are only consumers of information and they have to get connected to a centralized server via the Internet. What is still missing is applications benefiting from the potential of collective work arising among mobile devices, sharing both information and computing. Sensor networks, for instance, follow this paradigm but we want to extend this vision to any set of mobile and stationary devices.

In addition, applications following the centralized scheme may face problems of network saturation or robustness against failures. Moreover, stigmergy-based, context-awareness or spontaneous social networking applications need to use information being meaningful only locally, for a limited period of time, thus making the use of (centralized) servers unnecessary. There is no need for the information to travel a large geographical region when both the producer and consumer of that information are next to each other.

In this paper, we introduce the Spatial Memory vision in which any region of the earth may be defined as a memory able to store and retrieve any kind of information. Spatial memory as a general concept may rely on any kind of technology from centralized servers to mobile devices, from infrastructure to mobile ad hoc networks. However, we propose a concrete implementation of spatial memory called Hovering Information which has been already presented in previous works [12,13,14]. In other words, we have extended the Hovering Information concept to the more general concept of Spatial Memory.

In addition to the conceptual definitions we present in this paper, we also present a middleware architecture for the Hovering Information implementation. We discuss about its requirements and properties. We also present the storage and retrieval algorithms which are the core of the middleware. Storage and retrieval have been evaluated through simulations using both random-way-point mobility pattern and real mobility patterns.

1.1 Related Work

The Virtual Infrastructure project [4567] defines virtual (fixed) nodes implemented on top of a MANET. This project proposes first the notion of an atomic memory, implemented on top of a MANET, using the notion of quorums or focal points where a reasonable amount of mobile nodes intersect. The motivation behind this project is the development of a virtual infrastructure on top of which it will be easier to define or adapt distributed algorithms such as routing, leader election, atomic memory,

motion coordination, etc. Hovering information shares similar characteristics, it tries to benefit from the mobility of the underlying nodes, but the goal is different. We intend to provide a hovering information service on top of which applications using self-organising user-defined pieces of information can be built.

GeOpps 9 proposes a geographical opportunistic routing algorithm over VANETs (Vehicular Ad Hoc Networks). This work focuses on routing information to some geographical location, it does not consider the issue of keeping this information alive at the destination, while this is the main characteristics of hovering information.

The work proposed by 10 aims to disseminate traffic information in a network composed by infostations and cars. While the idea is quite similar to that of hovering information, keeping information alive in its relevant area, this study does not consider the problem of having a limited amount of memory to be shared by many pieces of information or the problem of fragmentation of information. It also takes the view of the cars as the main active entities, and not the opposite view, where it is the information that decides where to go.

The Ad-Loc project 1 proposes an annotation location-aware infrastructure-free system. Notes stick to an area of relevance which can grow depending on the location of interested nodes. Notes are kept in their relevance area is by periodically broadcasting location-aware information to neighbouring nodes. This work also proposes to use this annotation system as a cache for Internet files in order to spare bandwidth. In this case, URLs are used as note identifiers. Similarly to the previous work, nodes are the active entities. In addition, in this case the size of the area of relevance grows as necessary in order to accommodate the needs of users potentially far from the central location. The information then becomes eventually available everywhere.

The ColBack system 28 is part of the MoSAIC project and intends to set up a collaborative backup system for mobile devices. This system does not focus on geolocalized information but replication strategies and replica scheduling and dissemination techniques could be used as inspiration for hovering information replication algorithms.

PeopleNet 11 describes a mobile wireless virtual social network which mimics the way people seek information via social networking. It uses the infrastructure to propagate queries of a given type to users in specific geographical locations called bazaars. Within each bazaar the query is further propagated between neighbouring nodes via peer-to-peer connectivity until it finds a matching query. The proposed queries propagation inside bazaar techniques could be a source of inspiration when we will develop query to retrieve specific hovering information.

2 Spatial Memory

The main idea is that of making a geographical region to behave as a memory where information of any kind could be stored, modified and read by user applications. This information will be stored in devices (mobile and fixed) being inside the region. The environment is highly dynamic as devices can leave the region, fail, run out of memory or just be switched off; or new devices can enter into the region.

Let us develop a more global vision of the concept. We consider the surface of the earth as a whole geographical region in which plenty of computers, servers, mobile phones, PDAs, RFIDs, smartphones, iPhones, sensors, etc are storing and gathering information; processing it and producing still new information. The exchange of information between all of them is via the Internet or local networks using wireless or fixed technologies. Let us call all these devices as nodes and let us sub-divide the whole region in arbitrary sub-regions that we will also call them regions by simplicity.

Each region has potentially some storage capacity where information can reside; this capacity depends on the number, characteristics and state of devices being inside the region at some point in time. We can imagine that information could be of any type and size, from text to video and from some bytes to several megabytes (or even gigabytes).

Once a piece of information created, it will have a self-organizing nature. The information will be responsible of staying at the region with which it is associated, or to migrate from one region to another if requested. It will be responsible of being accessible by user applications that will be able to read and modify it. To achieve all this goals, information will use the storage, computing and communication resources of the nodes. Regions are passive elements of spatial memory and pieces of information are the active ones.

It is important to notice that one major difference between Spatial Memory and current information management systems (databases, servers, p2p) is the way of addressing information. Traditional systems associate an address to a server or a peer storing information through which we can reach the information on request. However, in the Spatial Memory concept the way of addressing information is by geographical regions. We ask for information having some type and content, but being stored at some particular region. However, it is also possible to ask for information having some characteristic without specifying its location, in this case it is up to Spatial Memory to figure out where the information is stored. In both cases, there is a strong relation between geographical location and information because the latest is anyway stored at some region.

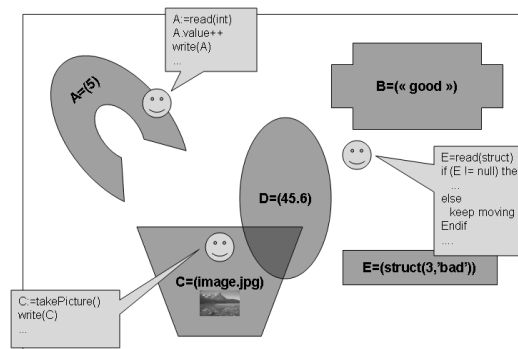


Fig. 1. Spatial Memory Vision.

User applications will only have to make use of simple primitives to basically read or write information from/to some region. Spontaneous social networking, disaster

areas, stigmergy-based and context-aware applications are the most appropriate for using this paradigm. Figure 1 depicts a geographical region where several pieces of information co-exist and each is associated with a specific region. We can also observe the presence of user applications that access the information of their interest in order to perform some tasks.

Extrapolating the vision of Spatial Memory to current information management systems (i.e. web, video/audio streaming, databases, etc.), we can consider this information as information attached to the regions of these servers. Whenever specific information is required by a client, the data will then migrate from the region of the server to the region of the client, where it is processed by user-applications. Another view is of considering the whole earth as the region of information stored in servers.

2.1 Requirements and Issues

Setting up a system such as Spatial Memory is undoubtedly not trivial since the environment is highly dynamic and unreliable. In the paragraphs below, we enumerate and discuss the most relevant requirements and issues.

Persistency. This is a major issue because of the dynamism and unreliability of the environment. In such conditions, an existing piece of information is likely to disappear when a node switch off or leave some region; a node runs out of memory; or some network errors happen. It is then extremely important that information find by itself the way to stay alive, to survive. In the other hand, time life of information depends on its nature. There might be some information being relevant only for some few minutes or hours; and another for some weeks or months. It is evident that short time life information fits better with this environment.

Reliability. While persistency deals with keeping information alive, reliability consists on keeping information accessible for nodes willing to read or modify it. Again, this is a very challenging issue to address. An existing piece of information may leave its meaningful region and be the last one piece on this area. Therefore, any node willing to access it will not be able to do it even though the information still exist. Although a fully reliability is the ideal situation, several applications might not need such a quality to still work well. For instance, instantaneous social networking applications may be more flexible in terms of reliability compared to disaster areas applications.

Consistency. Being Spatial Memory a highly distributed information storage system, inconsistency problems will follow. We can identify two sources of inconsistency. First, the fact that a piece of information will certainly exist in several nodes (distributed), a consistency issue arises when some node modifies its content and this modification must take some while to be known by other nodes hosting the same piece of information. Another situation is concurrent modification of a piece of information, two different version of the same information could exist at the same time. Last case, when the same information exist in two different and not overlapping regions, in this case each region could have a different version. In all these cases, ensuring consistency is a challenge. However, inconsistency in information could also be an advantage in some cases since human social information is most of the time contradictory (there is no an absolute knowledge) and society still works.

Atomicity. When writing, reading or modifying a piece of information, Spatial Memory must ensure the atomicity of these operations. Some information could be split in several pieces of information and spread in several nodes depending on the storage capacity of nodes. Later, this pieces will be merged into one piece of information to be used by applications. This kind of behaviour is definitely a challenge to overcome when dealing with atomicity. Another source of problems is communication failures but we could delegate this responsibility to the overlay network.

Trust. The nature of information is dynamic and user defined. These aspects are both an advantage and at the same time a source of trust challenges. We have to address challenges such as how to trust to some information published by some one else, or how to be sure that some information is not a trap for us. More an application relays on the reliability of the source of information, more it will sensible to the trust of information. As a first approach, we can image a kind of reputation systems in order to solve this problem.

Integrity. An application should be able to modify only its related information. In this way, integrity of information has to be conserved. It is in this context that attributes such as owner and access rights are aimed to exist. At the same time, it arises a problem of supplanting an identity and then modifying an information that normally we could not do it. Again, it is another challenging issue to solve when designing a concrete implementation of Spatial Memory.

Cooperation and Privacy. Spatial Memory requires a high degree of participation from nodes. But, this means that nodes have to be willing of sharing theirs storage, computing and networking resources. Otherwise, such a system will not work because it needs a critical mass of participants. Besides this problem of cooperation, it arises a problem of security and privacy. Participant nodes need enough guarantees that they will not be attacked because of being more open to cooperate, or that their private data would be accessible by anyone. A possible solution could be setting up a virtual machine on nodes to create a fully protected space where spatial memory could work.

3 Hovering Information Middleware

This section outlines a particular design of the Spatial Memory concept described above. Leveraging on our previous work on hovering information [1314], we show here how to build such a service through a middleware called Hovering Information Middleware (HIM).

3.1 Global Overview

The middleware assumes the presence of mobile (fixed) nodes moving around a geographical 2D area. Each node is able to store and process information, that we call it hovering information, as well as sent it to and receive it from neighbouring nodes in an ad hoc manner. User-applications are able to create information to be stored at a specific geographical region chosen by the application itself or by the middleware.

The creation process of information could be both triggered by a human user or by the application itself in an automatic way.

Figure 2 depicts several applications running on different groups of nodes; each application interacts with the hovering information middleware in order to perform its task.

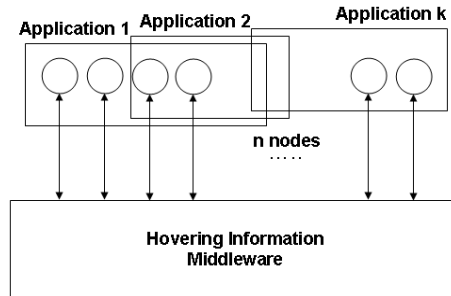


Fig. 2. Applications Deployment.

A piece of hovering information can be of any type from simple raw text to videos. Once created, a piece of information might be attached to a precise geographical region defined by the user-application. If not, it is up to the middleware to find a region where to store it. In both cases, information can be retrieved from anywhere unless strict region-bounded constraint is specified.

From an application perspective, hovering information is a concept characterising self-organising information responsible to find its own storage on top of a highly dynamic set of mobile devices. A piece of hovering information is attached to a specified location, which we call the anchor location. The main requirement of a piece of hovering information is to keep itself stored in the vicinity of the anchor location, which we call the anchor area, despite the unreliability of the device on which it is stored. Whenever the mobile device, on which the hovering information is currently stored, leaves the area around the specified storage location, the information has to hop - hover - to another device.

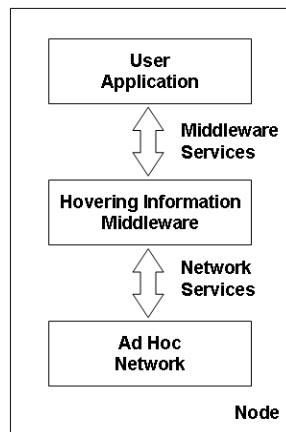


Fig.3. Node Architecture.

Pieces of hovering information have to be seen as information particles "suspended" into the given geographical region, being *attracted* by their anchor location. A piece of hovering information comes with a unique identifier, the actual data value to store and additional policies for spreading, self-destruction, etc. By means of simplicity, we will use the term information or hovering information in an exchangeable way.

From a conceptual point of view, the middleware itself is responsible of two main tasks: storing information (persistence) and retrieving it. We assume that the middleware use network primitives such as sending and receiving data, and services such as discovering neighbouring devices. On top of this, applications make use of the middleware services. Figure 3 illustrates these three layers in a node and their relation.

The current middleware we propose is designed to work in mobile ad hoc networks. In previous works [1314], we studied the behaviour of some simple but fundamental algorithms being responsible of the basic mechanism of keeping information persistent in a highly dynamic environment. Sub-section 4 discuss more in detail these algorithms.

3.2 User-application Primitives

The way how an application interacts with the middleware is through an Application Protocol Interface (API). Besides basic primitives for writing and reading information, the middleware proposes primitives for subscribing an application to some information-related events; and for defining regions, filters and policies. Let us explore more in detail each of these primitives.

1. $write(h, a, p)$ – This primitive allows an application to write a piece of information h into a region a having the p policies.
2. $write(h, p)$ – This primitive allows an application to write a piece of information h having the p policies. It is to the middleware to define the most appropriate region where h will be written.
3. $read(q, a)::h$ – This primitive allows an application to retrieve a piece of information h matching the query q and being stored into the region a . If no information matches the query, a null value is returned.
4. $read(q)::h$ – This primitive allows an application to retrieve a piece of information h matching the query q . It is up to the middleware to decide where to look up for this information. If no information matches the query, a null value is returned.
5. $subscribe(h)$ – This primitive allows an application to automatically be notified when information h has been modified or removed.
6. $notify()$ – This primitive is called by the middleware to inform to an application that a subscribed information has been modified or removed.
7. $region(w)::a$ – This primitive allows an application to define a geographical region by its vertices w . A priori, the geometrical form of this region should be any. A region will be the semantic scope of a piece of information. The definition could be also done by passing the center and the radius of the region, in which case the region will a circle.
8. $applyRegionOperator(operator, a, b)::c$ – This primitive allows an application to

combine two regions a and b and produce a new one c which will be the results of an *operator* such as union, intersection or subtraction among others. This kind of operators may be useful when extending the semantic scope of some information for instance.

9. *defineFilter(expression)::f* – Existing several pieces of information stored in the global region and many of them being stored in the same region or in overlapped ones, it should be possible to define a filter f when reading some information. This filter would define by an *expression* the region or regions, the type of information, the user id application, the owner id, the content, the author id, etc.

3.3 Middleware Architecture

The hovering information service is designed as a middleware service. It is a distributed service and each node must implement it. The Figure 4 outlines the proposed architecture.

The core of the proposed middleware are the Storage and Retrieval modules, both coordinate their behaviour according to the network properties (i.e. network saturation), local device properties (i.e. node routines), environment properties (i.e. neighbours state) and application requirement (i.e. low delay response when reading an application related information).

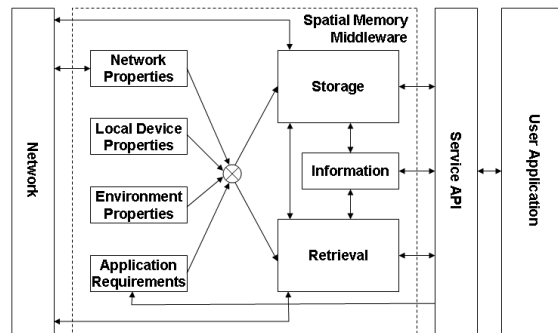


Fig. 4. Hovering Information Middleware Architecture.

Storage. This module is responsible of storing information in a region. It has as inputs data such as the network state, the local node state and the neighbours state. This data is provided by the respective modules. Some previous research has been done for this module and two algorithms have been proposed: attractor point and location-based caching 14.

Retrieval. This module is responsible for looking for an information requested by an user application. The information is searched in a particular region and filtered after the defined filter. The simple case is when information is stored at the current region, otherwise a process of propagation of the query will be triggered and the requested information, if found, will migrate to the querying area of the user application.

Network Properties. This module provides a transparent access to the network properties and current state such as traffic congestion, bandwidth, delay and current

neighbours among others. These network characteristics are used by the Storage and Retrieval modules in order to adapt their behaviour.

Local Device Properties. Local information of the device such as memory size, current position, current speed and prediction of location is made available by this module. This information is part of the set of parameters that Storage and Retrieval modules take in consideration to adapt their behaviour.

Environment Properties. This module makes available some other environment parameters such as neighbours nodes and their characteristics such as their current position and velocity vector.

Application Requirements. This module provides to the Storage and Retrieval modules parameters defining the expected performances an user applications requires to work in the right manner. Modules should have to adapt their efforts in order to reach these requirements as best as possible.

Information. This module is responsible of managing and processing the hovering information that is stored or retrieved from a region. It deals with attributes such as the age of the information, generation, priority, type, owner, etc. Meta-data such as replication time, version, creation time and generation are also processed in this module in cooperation with Storage and Retrieval modules (cf. §4).

3.4 Middleware Properties

The proposed middleware's properties come out from the design of the Storage and Retrieval modules and from the nature of the context which is highly dynamic and unreliable.

Infrastructure-Free. There is no need of infrastructure for this service. It is a kind of opportunistic service that profits from the presence of devices being around to do its job. It does not rely on any centralized server.

Self-Organisation. Once created and introduced in a region a piece of information, this last is completely autonomous. It looks by itself to survive and be accessible for user applications. The kind of operations that information could be able to do for this goal include replication, propagation, splitting, merging, swarm behaviour, etc.

Best-Effort Service. Being the environment highly dynamic, the service provided by spatial memory is not fully reliable and user applications should be designed having in mind this characteristic. However, user applications are able to specify their minimal requirements in order to achieve their goals.

Migration. We emphasize the fact that semantic scope of information is bounded to some region. However, it might be possible to replicate this scope some where else or to be expanded. In this case, semantic scope should be extended and modified and then a migration of information should be done.

Garbage Collector. Pieces of hovering information come with policies. Those policies prescribed a time-to-live limit after which the information self-destroys. This starts counting from the last time the information has been accessed or modified. This however means that the data is in fact modified to remember the last time of access.

Load-Balancing. The place where a piece of hovering information is stored is decided by either the application or the middleware (if not specified an specific region). In both cases, information replicates itself in quest of survive and competes with other information. Therefore, several different pieces of information may co-exist and previous results 14 have shown that geographical region is fairly distributed among different pieces of information.

4 Storage and Retrieval Algorithms

In this section we describe the current design of the storage and retrieval modules. We discuss about the main issues when designing them, the current implementation and its drawbacks, as well as the next milestone improvements in order to design a fully working middleware.

4.1 Storage Module

The information storage module is responsible of storing pieces of information, created by user-applications, in the most appropriate geographical region. As we mentioned in §2.1, keeping information alive and available are major challenges of such a system because of the highly dynamism of the environment.

Current Storage Algorithms. In previous work we have already presented two storage algorithms, called as replication algorithms: attractor point (AP) and broadcast-based (BB). We present the main concepts related to these algorithms.

Once a user-application creates a piece of information, the hovering information middleware will store it in the most appropriate geographical area called the *anchor area*. This piece of information can be tightly related to a geographical area (e.g. a context-aware information) or not, in which case the middleware can store information anywhere. In this paper, an *anchor area* is defined as a circle having a center, the *anchor location*, and a radius, the *anchor radius*, both defined by the user application to an specific piece of information. For simplicity we assume in this paper that an *anchor area* is always defined as a circle, but we could imagine that it could have any shape than a circle.

The AP and BB algorithms are based on self-replication. When a replica is in risk of disappearance it starts replication itself. To this aim, an *anchor area* and its surroundings are divided in four continuous areas, all of them having as center the *anchor location*. The *secure area* which is a disc having a radius smaller than the *anchor radius*, therefore this area is inside the *anchor area*. The *risk area*, which is the prolongation of the *safe area* beyond the *anchor area*, is a ring having a minimal radius equal to the radius of the *safe area* and a maximal radius greater than the *anchor radius*. The *meaningful area* is also a ring having a minimal radius equal to the maximal radius of the *risk area* and a maximal radius being normally quite far away from the *anchor location*. And finally the *irrelevant area* is the rest of the rest of the area where information is likely not relevant.

When a replica is in the *risk area*, it starts replication itself constantly in the quest of surviving. While the BB algorithms replicates a replica into all the neighbours of the respective host, the AP algorithms tries do it better by targeting only those neighbours being the k -closest ones (replication factor) to anchor location. We define the following attributes: replication factor (k_R), replication timer (T_R) and cleaning timer (T_C). A replication being in risk replication itself each T_R seconds and a node removes irrelevant replicas (those being in the *irrelevant area*) each T_C seconds.

Besides the two replication algorithms, we have also introduced two caching policies: the location-based caching (LBC) and the generation-based caching (GBC). In fact, in a context where several applications co-exist and each of them has several pieces of information (several context variables), it might happen that replicas compete for storage resources. That is why caching policies have been defined in order to choose the most appropriate replicas when storage space is not enough in a hosting node. The LBC policy is the most interesting and the most appropriated after our previous results 14. It is based on the location of the competing replicas and the hosting node. The most relevant replica to the host's location will have the priority.

Drawbacks of the Current Storage Algorithms. Previous replication algorithms and caching policies are not good enough yet. Although, they achieve their goals and their performances are good, the mayor drawback resides on the network overhead. The following paragraphs discuss these issues.

Once a replica is in its risk area, the replication frequency depends on the network saturation, the current state of host node's neighbours, the life cycle of the replica, and the kind and priority of the information. This means that the replication mechanism adapts its behaviour to the saturation of the network in order to reduce the bandwidth consumption. Concerning the state of neighbours, a host node is constantly receiving messages from its neighbours, some of them are related to hovering information. In this way, by receiving some replication messages from away, a host node can estimate which replicas are already in its direct neighborhood, and a replication reduces its replication if there are several copies of it around. Concerning the lifecycle, a replica would have the tendency to replicate itself very often at the beginning of its existence, and then will gradually reduce this frequency assuming that there are enough replicas out there. Finally, a higher priority information will invest more effort on replication itself.

The current design of the replication algorithms is thought limited. A replica replicates itself when being in its risk area and keeps doing while being there with a well defined replication frequency. Therefore, we can straight notice the main drawback: network overhead. The current host of a replica may slowly pass through the risk area, or event stay there, and the replica would keep replicating itself permanently. An stop mechanism is needed. Moreover, the replication frequency is constant but we could imagine that this may evolve along the time as well. A compulsive replication might be suitable at the beginning of a prior of risk but not after a while of estimating that survival is quite guaranteed.

The current AP replication algorithm replicates a replica potentially in risk to a set of its neighborhood nodes being the closest ones to the anchor area of the replica. However, this implies that a multicast or several unicast messages are sent. The question that arises from this implementation is that broadcasting the replica to all neighbours might be more efficient and more or less equally expensive in terms of

power.

Previous results shows that the current replication algorithms get good performances in terms of keeping hoverinfos alive and available. However, the major drawback is the network overhead even though the AP does much better than a flooding scheme. From all this, we conclude that, and as stated and thought before, a replication decision based on the nodes behaviours may significantly improve and reduce the network overhead. The main idea resides in classifying nodes in two categories. Those who are most suitable to stay for a long time in a geographical area and those that not. The former would be the most suitable for storing replicas related to the respective location.

4.2 Retrieval Module

Once a mobile application or another system has created a hovering information and stored it in a geographical area, we must be able to retrieve this hoverinfo when required. We distinguish two kinds of retrieval mechanisms: a passive one and an active one. While in the *active retrieval* the user application asks for some information, in the passive *retrieval* the information looks for matching user applications (publish/subscribe scheme). The answer message, if the query matches, propagates following a unicast scheme. In this paper, we only cope with the *active retrieval*.

We describe the active query propagation. Once a user application has asked for some information, the respective node generates a query that will propagate in quest of matching some replica stored in some node's buffer. The approach we present here deploys a mitigated and geographically bounded flooding when the query is in the anchor area of the respective information. Otherwise, the query is propagated following a geographical routing until it gets to the anchor area of the information and then the previous method is triggered.

Each time a query arrives to a node, the node verifies whether the query has been already received in which case the node drops the query. Otherwise, the node verifies if the query matches any of the hoverinfos stored in its buffer to later send this hovering info back to the demanding node. If the query does not match any hoverinfo, it propagates itself towards neighboring nodes. The decision of propagation is taken following an uniform distribution and a threshold previously defined.

Once a query has found a node containing a replica matching the query, the node sends back to the node source of the query an answer message with the respective hoverinfo. To route the message back, the node sends an unicast message to the node from where it received the query previously. Each node receiving this message behaves in the same way until the message gets to the destination node. Another way of sending back an answer without using the historic information is by only performing a geographical routing.

5 Evaluation

We evaluated the behavior of the above the current storage and retrieval modules

under different scenarios by varying the number of nodes, the number of hovering informations and the number of queries.

We performed simulations using the OMNet++ network simulator (distribution 3.3) and its Mobility Framework 2.0p2 (mobility module) to simulate nodes having a simplified WiFi-enabled communication interfaces (not dealing with channel interferences) with a communication two different communication ranges of 30m and 120m. Concerning the mobility module, an update interval of 1s and a window time of 10s has been used.

5.1 Simulation Settings and Scenarios

The generic scenario consists of a surface of 500m x 500m with mobile nodes moving around following a Random Way Point mobility model with a speed varying from 1m/s to 10m/s without pause time, or following a Real Mobility Pattern after the SLAW model 15.

In the generic scenario, pieces of hovering information have an anchor radius (r) of 50m, a safe radius (r_{safe}) of 30m, a risk radius (r_{risk}) of 70m, a relevance (meaningful) radius (r_{rel}) of 200m, and a replication factor of 4 (k_R).

Each node triggers the replication algorithm every 10 seconds (T_R) and the cleaning algorithm every 60 seconds (T_C). Each node has a buffer capacity of up to 20 different replicas. The caching algorithm is constantly listening for the arrival of new replicas.

Based on this generic scenario, we defined specific scenarios with varying number of nodes: from 20 to 200 nodes, increasing the number of nodes by 20; varying number of different pieces of hovering information existing in the system: from 20 to 200 hoverinfos, increasing the number of pieces by 20; and by producing 5 queries per hoverinfo created. Each of this scenarios has been investigated with different replication algorithms and the LBC caching policy.

We have performed 10 runs for each of the above scenarios. One run lasts 3'600 simulated seconds. All the results presented here are the average of the 10 runs for each scenario, and the errors bars represent a 95% confidence interval. All the simulations ran on a Linux cluster of 32 computation nodes (Sun V60x dual Intel Xeon 2.8GHz, 2Gb RAM).

5.2 Critical Mass

Figure 5 outlines the average availability of the AP and BB replication algorithms combined with the LBC caching policy. An average availability value has been computed for several scenarios containing each a different number of nodes.

We can observe that as expected the BB algorithm has in all the cases an average availability bigger than that of the AP algorithm. However, for this latest, we note an strange behaviour in comparison to previous results 1314, the availability does not keep growing as the number of nodes increases but it starts decreasing after 80 nodes. There are two issues for these results: either they are incorrect and the whole simulations should be verified, or any some missing detail is missing and they are correct. In the latest case, the critical mass which is defined as the minimal number of

nodes per area unit so that some acceptable level of availability can be reached (this threshold depends on the user-application but for the moment we assume that an acceptable one should be bigger than 90%) would be around 100 nodes for the BB algorithm, and there would not exist for the AP algorithm since average availability is lower than 80%.

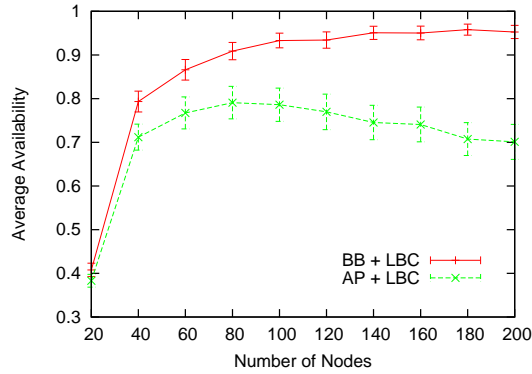


Fig. 5. Availability.

5.3 Network Overhead

Figure 6 depicts the number of sent messages for both the AP and the BB replications algorithms under several scenarios, each having a different number of nodes.

Again, in comparison to the results of the critical mass, we observe an strange behaviour for the AP algorithm. The number of sent messages is bigger than that of the BB for the range 20 to 140 nodes, and it starts decreasing after 80 nodes. These results are also incompatible with those of previous works [13, 14].

We define the network overhead as the number of messages exchanged between nodes per unit of time. One goal of storage and retrieval algorithms is to minimize this variable while maximizing the storage/retrieval performances. However, since our results seem to be biased, we can not conclude anything for the moment than verifying current results.

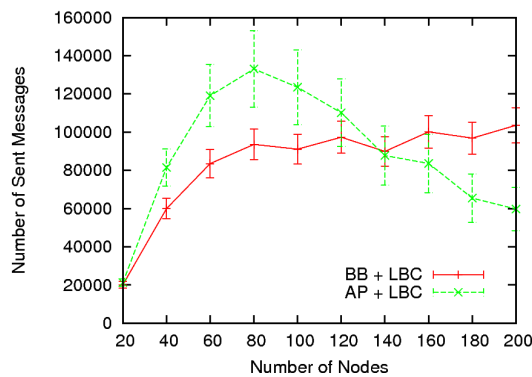


Fig. 6. Sent Messages.

5.4 Absorption Limits

Figure 7 shows the average availability for both the AP and the BB replications algorithms under several scenarios each having a different number of hovering informations but always either 100 or 200 nodes.

A hovering information system has limits. We can not store information as much as we want because of some inherent limitations: the whole storage capacity which is defined as a function of the number of nodes and the buffer size at each node. However, this absorption limit is below of this theoretical limit because of the redundancy of replicas.

The results of this paragraph shows that and absorption limit is reached for the BB because we can observe that the number of hoverinfos created in the system is not bigger than 120 hoverinfos. We can explain this results by the fact that nodes' buffer have a limit of 20 hoverinfos and that at some point they get full and no more hoverinfos are accepted. In the other hand, for the AP algorithm, we observe that a bigger amount of hoverinfos is absorbed by the system. However, these results should be verified because of the strange behaviour of the AP algorithms noticed in previous sub-sections.

5.5 Communication Range

Figure 8 depicts the average availability for both the AP and the BB replications algorithms under several scenarios having as main difference the communication range: 30m and 120m.

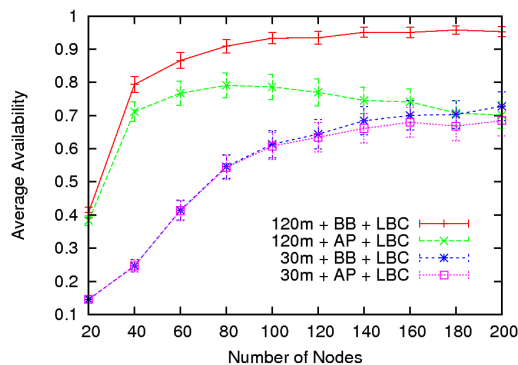


Fig. 7. Availability and Communication Range.

We observe that as expected the availability is lower for a communication range of 30m. However, and in opposition to previous sub-sections, we notice that the behaviour of the AP in a configuration of 30m is more reliable and consistent with what has been expected and what has been found in previous works [1314]. This fact leads us to think about the correctness of the results for 120m of communication range.

5.6 Real Mobility Patterns Scenarios

The previous results have studied the behaviour of hovering information middleware under an hypothetical scenarios on which every node moves randomly. This is quite rare to happen in reality, unless we consider a market or festival where we could observe a more chaotic behaviour in the movement of people. Therefore a set of simulations has been performed using real mobility traces or more precisely pseudo real mobility patterns, which have been generated by the SLAW mobility model 15.

Several scenarios have been defined each having different density of people moving around. Figure 9 presents a mobility trace of 200 nodes moving after the SLAW model, and an histogram of the locations by where nodes pass through. We can clearly observe, from both the trace and the histogram, that there are between 4 and 6 hotspots which could be considered as points of concentration of people. For instance in a real situation, these focal points might be considered as the main building of a university campus.

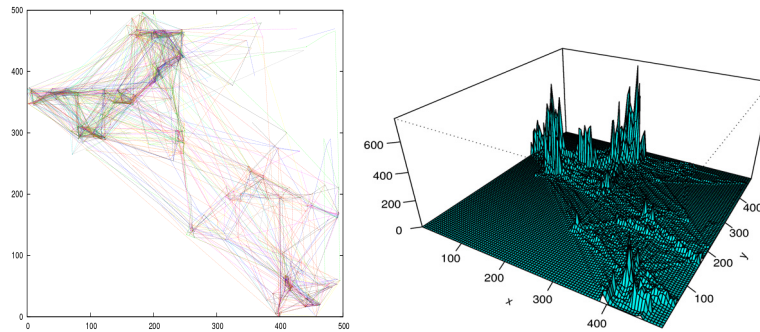


Fig. 8. Trace and histogram for SLAW real mobility model.

Figure 10 depicts the trace and histogram of 200 nodes moving after a random-way-point mobility pattern which has been extensively used during this technical report and previous works. We can notice that nodes have the tendency to pass more time in the middle of the area rather than in the sides. We can conclude from this that hoverinfos being in the middle will have more chances of surviving than those being created in the sides of the area. In the case of the real mobility patterns, the similar reasoning could be applied to the focal points.

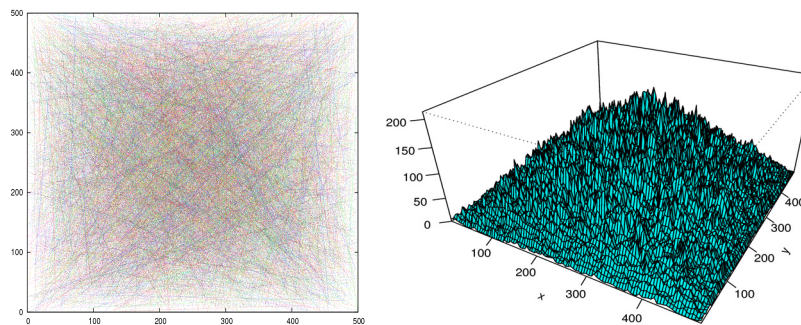


Fig. 9. Trace and histogram for random-way-point mobility model.

Figure 11 outlines the average availability for the AP and the BB replication algorithms under several scenarios having different number of nodes, which move after the SLAW real mobility model. We observe that for more than 80 nodes, the average availability for both algorithms is bigger than 60%, and for the BB algorithms is bigger than 80% which is not probably an acceptable value but it is not bad. For less than 80 nodes results are bad or strange, again further verifications and improvements have to be done.

6 Conclusions

In this technical report, we have described the Spatial Memory vision which defines any geographical region as a memory being able to store and retrieve any kind of information. We have also presented the Hovering Information concept re-engineered as a middleware implementing a constrained version of the spatial memory concept.

We have defined the API and the architecture of the Hovering Information Middleware, and described its properties. We have notably discussed about the current state and the major drawbacks of the Storage Module's replications algorithms, the Attractor Point (AP) and the Broadcast-based (BB), which were introduced in previous works. We have also introduced the Retrieval Module which for the moment implements a basic version of queries propagation, matching, and responses propagation back in a very simple way.

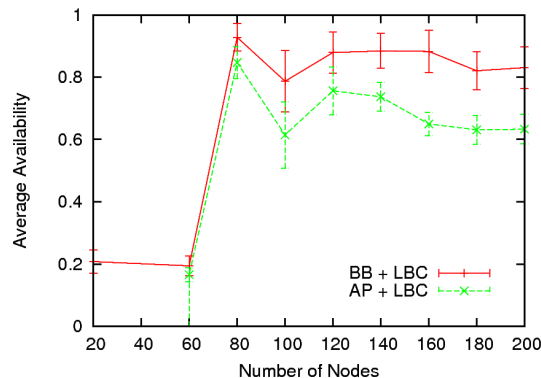


Fig. 10. Real Mobility Patterns – Availability.

We have run a set of several simulations using random-way-point mobility model and SLAW real mobility model. However, some results are strange and further verification is required, mainly for the AP algorithm. In what concerns real mobility patterns, results are encouraging because of the quite high performances for the BB replication algorithm.

It is clear that some improvements are needed in the storage and retrieval algorithms. One of them, and most probably the most interesting one, goes by adapting the behaviour of the algorithms to the daily behaviour of nodes, for instance to the nodes' mobility routines.

References

1. D. J. Corbet and D. Cutting. Ad loc: Location-based infrastructure-free annotation. In ICMU 2006, London, England, Oct. 2006.
2. L. Courts, M.-O. Killijian, D. Powell, and M. Roy. Sauvegarde cooperative entre pairs pour dispositifs mobiles. In UbiMob '05: Proceedings of the 2nd French-speaking conference on Mobility and ubiquity computing, pages 97–104, New York, NY, USA, 2005. ACM Press.
3. G. Di Marzo Serugendo, A. Villalba, and D. Konstantas. Dependable requirements for hovering information. In Supplemental Volume - The 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07), pages 36–39, 2007.
4. S. Dolev, S. Gilbert, L. Lahiani, N. A. Lynch, and T. Nolte. Timed virtual stationary automata for mobile networks. In OPODIS, pages 130–145, 2005.
5. S. Dolev, S. Gilbert, N. A. Lynch, E. Schiller, A. A. Shvart sman, and J. L. Welch. Virtual mobile nodes for mobile ad hoc networks. In DISC, 2004.
6. S. Dolev, S. Gilbert, N. A. Lynch, A. A. Shvartsman, and J. Welch. Geoquorums: Implementing atomic memory in mobile ad hoc networks. In DISC, 2003. .
7. S. Dolev, S. Gilbert, E. Schiller, A. A. Shvartsman, and J. Welch. Autonomous virtual mobile nodes. In DIALM-POMC '05: Proceedings of the 2005 joint workshop on Foundations of mobile computing, pages 62–69, New York, NY, USA, 2005. ACM Press.
8. M.-O. Killijian, D. Powell, M. Banâtre, P. Couderc, and Y. Roudier. Collaborative backup for dependable mobile applications. In MPAC '04: Proceedings of the 2nd workshop on Middleware for pervasive and ad-hoc computing, pages 146–149, New York, NY, USA, 2004. ACM Press.
9. I. Leontiadis and C. Mascolo. Geopps: Opportunistic geographical routing for vehicular networks. In Proceedings of the IEEE Workshop on Autonomic and Opportunistic Communications. (Colocated with WOWMOM07), Helsinki, Finland, June 2007. IEEE Press.
10. I. Leontiadis and C. Mascolo. Opportunistic spatio-temporal dissemination system for vehicular networks. In MobiOpp '07: Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking, pages 39–46, New York, NY, USA, 2007. ACM Press.
11. M. Motani, V. Srinivasan, and P. S. Nuggehalli. Peoplenet: engineering a wireless virtual social network. In MobiCom '05: Proceedings of the 11th annual international conference on Mobile computing and networking, pages 243–257, New York, NY, USA, 2005. ACM Press.
12. A. Villalba and D. Konstantas. Towards hovering information. In Proceedings of the First European Conference on Smart Sensing and Context (EuroSSC 2006), pages 161–166, 2006.
13. A. Villalba Castro, G. Di Marzo Serugendo, and D. Konstantas. Hovering information - self-organising information that finds its own storage. In IEEE International Conference on Sensors, Ubiquitous and Trust Computing (SUTC'08), 2008
14. A. Villalba Castro, G. Di Marzo Serugendo, D. Konstantas, Hovering Information - Infrastructure-Free Self-Organising Location-Aware Information Dissemination Service, 2nd ERCIM Workshop on eMobility in conjunction with WWIC 2008, May 2008.
15. Kyunghan Lee, Seongik Hong, Aeong J. Kim, Injong Rhee, Song Chong, SLAW: A Mobility Model for Human Walks, In Proceedings of the 28th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), April 2009.

Business Modeling and Value Network Design Case Study for a Tele-Rehabilitation Service

Lambert J. M. Nieuwenhuis

University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
L.J.M.Nieuwenhuis@utwente.nl

Abstract. During the last decade, many useful telemedicine pilots have been conducted, however, only a few telemedicine services did actually reach the market and were successfully being deployed. One of the reasons is that costs and benefits are not equally distributed across the organizations of a value network. Value network analysis combined with business modeling right from the beginning of a project may improve the success rate of telemedicine services development and deployment. This paper presents results of this approach used in a case study for Myotel, a wireless rehabilitation service for treatment of chronic work related neck shoulder problems.

1 Introduction

Telemedicine may improve healthcare, both in a qualitative and in a quantitative way. During the last decade many useful technologies have been developed but fail to actually reach the market [4, 17]. Often, telemedicine services are provided by multiple organizations being part of a large value network. In practice, the services innovation is hampered by the fact that costs and benefits of a service innovation are not equally distributed across the network. An organization with only costs and hardly any benefits will not contribute to the innovation. A second roadblock in bringing telemedicine services to the market is the existence of regulations limiting the full power of the telemedicine services [4]. Business modeling is seen as a solution to bring (technological) innovations to successful deployment and several determinants have been identified for success [4, 2, 5]. By using business modeling right from the beginning of a project, the failure rate of telemedicine service innovations may be lowered. This is because such business model designs are expected to be more viable as a result of better alignment with available resources and capabilities as well with their external environments [17]. In this paper, we present a business modeling approach for Myotel, a wireless rehabilitation service for treatment of chronic work related neck shoulder problems. This approach allows for both a qualitative and quantitative analysis needed to engineer a viable value network by allocating activities to organizations such that costs and benefit are sufficiently matched.

2 Myotel Case Description

The case consists of a so-called myofeedback tele-rehabilitation system that can be used for patients with work related chronic neck shoulder pain. The system monitors muscle relaxation during daily activities via sensors and actuators implemented in a wearable garment, which is connected to a PDA. The system provides continuous feedback when there is too little muscle relaxation. The monitoring data is sent wirelessly – e.g., via a GPRS, UMTS or HSDPA connection – to a back end system, which can be accessed by health care professionals. These health care professionals can use the system for optimizing treatment working more efficiently by saving on face-to-face contact hours with their patients and giving them more personalized feedback as well (Figure 1 gives a high level architectural overview of the system).

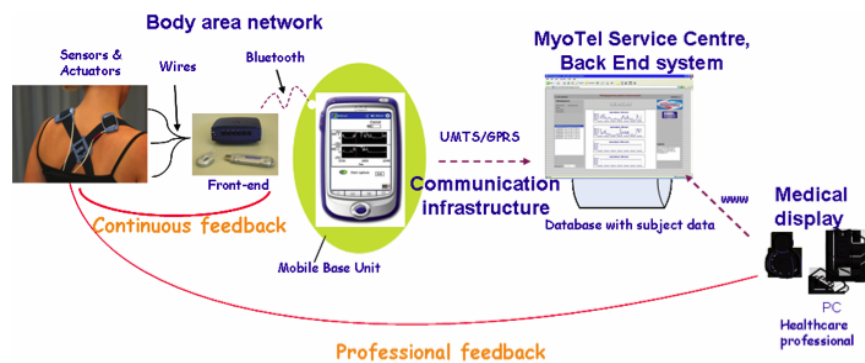


Fig. 1. Myotel system overview.

3 Business Modeling and Telemedicine Services

During the 1970s the business model concept was used for describing IT-related business processes [12, 18]. More recently, the business model concept has been used for analyzing market structures as well as strategic choices related to positioning of organizations within these market structures [16, 8]. A widely used business model definition within this context is that of Chesbrough and Rosenbloom [5] who concisely define a business model as “a blueprint for how a network of organizations co-operates in creating and capturing value from technological innovation”. In our view, it is important to explicitly distinguish the two main types of value to be created: customer value (value delivered from a customer perspective) as well as monetary value (value delivered from a provider perspective). So we define, in similar words, a business model as “a description of the way a company or a network of companies aims to make money and create customer value” [7, 11].

In literature, initially attention has been paid to empirically define business model typologies [8]. In recent years business model research started focusing on exploring business model components and developing descriptive models [11, 15].

Afuah and Tucci [1] see businesses as systems consisting of components (value, revenue sources, price, related activities, implementation, capabilities and

sustainability), relationships and interrelated technology. Osterwalder and Pigneur [14] more systematically define four business model components, i.e., product innovation, customer management, infrastructure management and financial aspects. Broens et al. [4] identified five determinant categories that influence implementation of telemedicine interventions: (1) technology, (2) acceptance, (3) financing, (4) organization and (5) policy and legislation. For our case, the technology has been addressed by medical trials proving the medical effectiveness of the Myotel treatment [9]. A recent project focused on the acceptance by conducting Myotel trials in four European countries. Section 4 proposes a business engineering approach addressing the financial in particular the revenue streams and organizational aspects, i.e., value networks and related roles. Section 5 deals with the policy and legislation.

4 Value Chain Engineering using Business Models

We engineered the value chain by addressing the differences between the current traditional treatment and the future treatment of work related chronic neck shoulder pain using the tele-rehabilitation service, using the following three-step approach:

- The first step identifies the main activities needed to provide the service, by defining the role that needs to be played by an actor and a related short description. A certain type of organization or a specific professional with a given set of competences can play a role.
- The second step determines costs related to each role for the current and future situation. This step clarifies the economic impact of tele-rehabilitation services by determining the cost increase or decrease for each activity identified in the first step. From an economic point of view, the implementation of the tele-rehabilitation service makes sense only when the overall cost level of the future situation decreases.
- The third and last step is the “engineering of the business” by allocating activities to existing or to be created organizations. The objective is to create a value network where each organization benefits from the introduction of the tele-rehabilitation, i.e., additional costs must be compensated by additional benefits.

The above-mentioned process requires insights from business model surveys, literature studies, as well as expert interviews. Several cross-organizational workshops are needed to design business models, parameterize quantitative analysis and provide feedback to all parties. The next sections describe the above-mentioned steps needed for the engineering the business needed to viably provide the tele-rehabilitation service.

Step 1: Identification of Main Activities. In order to develop an initial qualitative business model design for the myofeedback tele-rehabilitation service, we organized a half-day workshop for twelve experts within the field of myofeedback and tele-treatments from four European countries in which the service could be offered – The Netherlands, Belgium, Sweden and Germany. Table 1 presents the activities of the main value network roles that will be affected by introducing the tele-rehabilitation service.

Step 2: Quantitative Model. The next step the costs for the activities of per role are

determined (see Table 2). We only look at the activities that change when the traditional treatment of neck shoulder is replaced by the new tele-treatment. These activities are presented in italics.

Table 1. Main roles and activities.

<i>Roles</i>	<i>Activities</i>
Patient	Gets treated by the myofeedback system
Health care professional	Offers the professional care and guides tele-rehabilitation activities
Tele-rehabilitation service provider	Offers the actual tele-rehabilitation service, including helpdesk, training & certification, public relations,
Hardware provider	Offers hardware, e.g. for communication devices and sensors
Network Provider	Offers mobile communications services
Software developer	Develops the tele-rehabilitation system software
Software platform provider	Offers the tele-rehabilitation software platform
Insurance company	Offers health insurance to patient and absence insurance to employer
Employer	Employs persons who are patient treated by tele-rehabilitation system
Medical research & development organization	Does R&D related to developing training material and certification.

Traditional activity based costing approaches are used to determine costs of the activities. For each activity we determine for each activity the number of times N the activity is carried out and the cost price P per activity. With respect to cost price, we distinguish between investments and yearly costs. The investments are onetime costs for training and education (needed when new employees get involved) and investments for equipment (needed when more devices are needed or old devices are worn out). We consider a period of ten years. The volume and cost tables are part of an Excel spreadsheet that simulates the provisioning of the tele-rehabilitation service in year i from 2008 to 2018. Multiplying N_i and P_i gives the overall costs for year i for each activity. The values for N_i are based on an S shaped technology adoption curve. The values for P_i are based on today's market prices that develop over time, i.e., technology prices decline (deflating prices), whereas, e.g., salary costs for professionals increase (inflating prices). This enables Net Present Value calculations over the ten years period as well.

In our study, experts in each country have determined the values for numbers and cost prices for each of the four countries under study. The Excel spreadsheet presents the set of parameters to the experts and provides for instantaneous feedback for what-if analysis. In this way, we obtained a useful tool for evaluating costs and benefits in direct interaction with the experts.

Results of the Quantitative Analysis. The results of the simulation for the Netherlands are presented Fig. 2. For each year of the ten-year period, the diagram

shows how investments and operational costs change due to replacing the traditional treatment by the tele-rehabilitation treatment. The tele-rehabilitation treatment leads to additional investments and operational costs changes for the following activities:

The Service Provider activities require investments in equipment, e.g., garments, sensors and devices for the patient as well as back office equipment. The estimated life time of the back office is three years; the life time for other equipment is set to two years.

- The Service Provider activities require also yearly operational costs to maintain, operate and use software licenses for the above-mentioned ICT.
- The Service Provider activities include investments to train health care professionals in using the equipment. This is a onetime impact, disappearing once all professionals are familiar with using the system. Also the Health Care Professional invests time (equals money) when he or she is being trained.
- The Health Care Professional saves operational costs per patient, because fewer treatments are needed (from nine to four treatments).
- Finally the employer of the patients saves costs due to absence reduction (fewer treatments) and lower productivity losses (because of the treatment).

Table 2. Activities per role.

<p>myofeedback service provider</p> <ul style="list-style-type: none"> • <i>manage tele-rehabilitation service (overhead)</i> • <i>develop tele-rehabilitation market (marketing)</i> • <i>acquire tele-rehabilitation customers</i> • <i>build back office</i> • <i>manage back office</i> • <i>build device service</i> • <i>manage devices needed for treatment</i> • <i>train myofeedback service delivery personnel</i> • <i>deliver myofeedback service</i> • <i>request reimbursement myofeedback treatment</i> • <i>receive payment for myofeedback service</i>
<p>health care provider</p> <ul style="list-style-type: none"> • <i>develop tele-rehabilitation treatment</i> • <i>train personnel tele-rehabilitation treatment</i> • <i>diagnose patient</i> • <i>consult patient with traditional treatment</i> • <i>consult patient with tele-rehabilitation treatment</i> • <i>request reimbursement treatment</i> • <i>receive payment for treatment</i>
<p>employer</p> <ul style="list-style-type: none"> • <i>employ traditionally treated employee</i> • <i>employ tele-rehabilitation treated employee</i>
<p>patient</p> <ul style="list-style-type: none"> • <i>undergo traditional treatment</i> • <i>undergo tele-rehabilitation treatment</i>

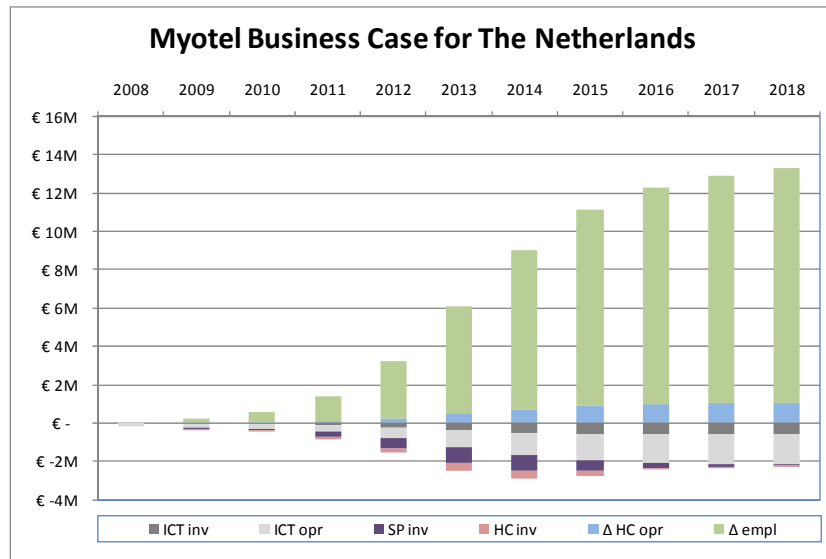


Fig. 2. Results of Myotel in The Netherlands.

Conclusions of the Quantitative Analysis. The cost benefit calculations revealed three critical insights that would be unknown without quantitative business model analysis (the related figures show the calculations for the Dutch market; for the other three countries, similar conclusions can be made):

- The tele-rehabilitation treatment is more expensive than the traditional treatment – mainly due to ICT related investments and operational costs.
- The IT investments are not likely to be compensated by the related labor savings for the Health Care Professional.
- However, the expected absence reduction and productivity increase of employers treated with the myofeedback systems compensates for the above mentioned additional cost.

The results shows that for example in The Netherlands, we may expect that employers obtain the most benefits related to implementing the myofeedback service. Therefore these organizations should be seen as the main potential revenue source when deploying the myofeedback teletreatment service.

Step 3: Value Network Engineering. The final step comprises the allocation of activities to existing or new organizations, part of the entire value network delivering the service to the patients and professionals. These organizations are part of a value network delivering products and services to each other on a commercial basis, i.e., one company pays for the product or service delivered to the other. The roles must be mapped such that each organization is able to cover the costs by payments originating from benefits. The experts mapped most of the roles easily to existing organizations, except for the role of tele-rehabilitation service provider. This role relates to costs (operations and investments) for IT devices and infrastructure that must be financed through the benefits of the employers. From a range of alternatives, the value network presented in was selected as most viable implementation. The employer (or the related

insurance company) of employees suffering from chronic neck shoulder problems has an incentive to use the tele-rehabilitation service.

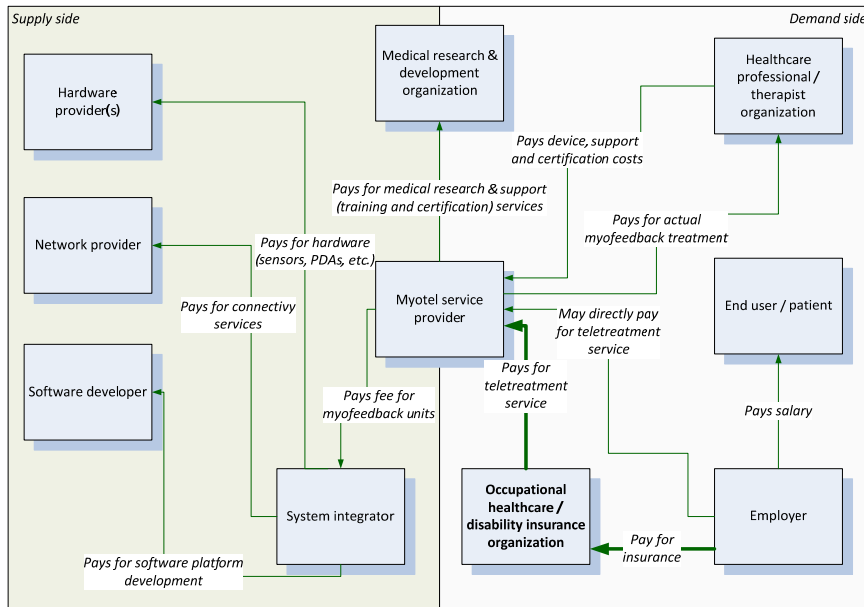


Fig. 3. Value network for Myotel.

5 Regulation

After deriving a viable value network design, we focused on performing a regulatory validation of the design for each of the four countries participating in the research. The main goal was to get insights with respect to the influence of the national healthcare regulation in each of the four participating countries[4]. According to Saltman[3], the influence of national healthcare regulation on the healthcare sector in a country can be divided into two different aspects – policy objectives and managerial mechanisms.

Policy objectives include regulation that is concerned with specific policy goals that influence a broad public, focusing on e.g. providing a healthcare system that is accessible for the whole population or educating citizens about clinical services, pharmaceuticals and a healthy lifestyle.

Managerial mechanisms have a more practical and operational character and target specific regulations that are needed to reach the goals as described in the policy objectives. Saltman et al [3] recognize a number of components that affect healthcare management capabilities and are associated with greater operating efficiency and effectiveness of both human and material resources. The following three main components of managerial mechanisms can be identified:

- State influence – does the government control the healthcare market, is it tax financed or is it dominated by private for profit organizations?

- Licensure and liability – how can be ensured that healthcare professionals meet competence standards and that malpractice will be prevented as well?
- Financial regulation – how does regulation influence the financial structure of the healthcare system?

Because of the practical and operational character of these components and their influence on the operating efficiency and effectiveness of both human and material resources, it is expected that these components have the largest influence on e-health service value networks. Therefore these three components formed the focus of our regulatory validation. Via a workshop with the experts from the four countries as mentioned in Step 1, we discussed the impact of national healthcare regulation in the Netherlands, Germany, Sweden en Belgium on the value network structure that resulted from the previous step.

When analyzing the results from the expert workshop and related expert interviews, we conclude that from the three regulatory elements as identified, the financial element has the most impact on the value network structure. Because reimbursements for e-health services – e.g. teleconsults – are not fully implemented yet, the potential efficiency increases related to the teletreatment service cannot be fully capitalized yet. Only in the Netherlands and Sweden, recognizes that the work related chronic pain is a work related injury. In both Belgium and Germany only severe injuries are officially recognizes as work related. Because of this, the occupational healthcare can be regarded as a potential revenue source and important value network role only in the Swedish and Dutch value network structures of the teletreatment service. Another critical aspect appeared to be the reimbursement factors: because a Dutch healthcare professional treating neck shoulder problems gets paid for the entire treatment, regardless of the specific treatment, there is an incentive to work more efficiently. In the other three countries, these healthcare professionals are paid per hour. As a result, for them there is no strong financial incentive to work more efficiently.

State influence also impacts the value network structure. Because of the healthcare market in Sweden is highly regulated with relatively independent and autonomous regional healthcare institutes, Swedish government agencies are expected to play an important role in the Swedish teletreatment value network. However, in the other three countries, where state influence is lower, insurance organizations are expected to be important actors in the teletreatment value network structures. The impact of the final component, licensure and liability, was expected to be of less influence on the national value networks of the teletreatment service.

Based on the results related to the regulatory validation, the Dutch and Swedish healthcare markets appeared to be currently the most viable markets for deploying the teletreatment service. In both the Netherlands and Sweden it was possible to include the occupational healthcare role in the value network, which turned out to be a potential revenue source critically important for viably deploying the teletreatment service.

6 Conclusions

In this paper, we described our business model engineering approach to early stage business model and value network development for a tele-rehabilitation service in the R&D deployment phase. We use a three-step approach. First, the main activities that are affected by the introduction of the new treatment are identified. Second, the investments and operational costs for each activity are determined. Finally, the activities are allocated to organizational units such that costs and benefits of the treatment can be matched.

Step 2 and 3 of our method led to critical deployment insights that would otherwise be unknown or learned at a much later phase of the development process. Improving the viability and feasibility of business model and value network designs in an early deployment stage may lead to substantial savings in costs and resources.

The analysis can be augmented by analysis considering the environmental factors like market, technology and regulatory environments.

Although the first results are encouraging, the method and empirical results need to be further validated and the relationships between the qualitative and quantitative analysis as part of the business_model_action design cycle should be further integrated as well.

References

1. A. Afuah and C. Tucci, *Internet business models and strategies: Text and cases*, McGraw-Hill Higher Education, 2000.
2. R. Amit and C. Zott, "Value creation in e-business", *Strategic Management Journal*, 22 (2001), pp. 493-520.
3. R. B. Saltman Regulating incentives: the past and present role of the state in health care systems. *Social Science and Medicine*. 2002;54(11):1677-84.
4. T. H. F. Broens, R. M. H. A. H. in't Veld, M. M. R. Vollenbroek-Hutten, H. J. Hermens, A. T. van Halteren, L. J. M. Nieuwenhuis, "Determinants of successful telemedicine implementations: a literature study", *Journal of Telemedicine and Telecare*, 13 (2007), pp. 303-309.
5. H. Chesbrough and R. S. Rosenbloom, "The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-off companies", *Industrial and Corporate Change*, 11 (2002), pp. 529-555.
6. R. Cole, S. Purao, M. Rossi and M. Sein, *Being proactive: where action research meets design research, ICIS 2005*, Las Vegas, USA, 2005, pp. 325-336.
7. T. Haaker, E. Faber and H. Bouwman, "Balancing customer and network value in business models for mobile services", *International Journal of Mobile Communications*, 4 (2006), pp. 645-661.
8. J. Hedman and T. Kalling, "The business model concept: theoretical underpinnings and empirical illustrations", *European Journal of Information Systems*, 12 (2003), pp. 49-59.
9. H. J. Hermens and M. M. R. Vollenbroek-Hutten, "Towards remote monitoring and remotely supervised training", *Journal of Electromyography and Kinesiology*, 18 (2008), pp. 908-919.
10. P. Jarvinen, "Action research is similar to design science", *Quality & Quantity*, 41 (2007), pp. 37-54.

11. B. Kijl, H. Bouwman, T. Haaker and E. Faber, *Developing a dynamic business model framework for emerging mobile services*, ITS 16th European Regional Conference, Porto, Portugal, 2005.
12. E. F. Konczal, "Models Are for Managers, Not Mathematicians", *Journal of Systems Management*, 26 (1975), pp. 12-15.
13. H. Mason and T. Rohner, *The venture imperative: a new model for corporate innovation*, Harvard Business School Press, 2002.
14. A. Osterwalder and Y. Pigneur, *An e-business model ontology for modeling e-business*, 15th Bled Electronic Commerce Conference, Bled, Slovenia, 2002, pp. 17-19.
15. A. G. Pateli and G. M. Giaglis, "A research framework for analysing eBusiness models", *European Journal of Information Systems*, 13 (2004), pp. 302-314.
16. M. E. Porter, "Strategy and the Internet", *Harvard Business Review*, 79 (2001), pp. 63-76.
17. T. Spil and B. Kijl, "E-health Business Models: From pilot project to successful deployment", *IBIMA Business Review*, 1 (2009), pp. 55-66.
18. P. Stähler, *Geschäftsmodelle in der digitalen Ökonomie. Merkmale, Strategien und Auswirkungen*, Patrick Stähler, 2002.
19. J. Tennent and G. Friend, *Guide to business modelling*, Profile Books, 2005.

Design of a Service-Oriented e-Learning Environment

Dumitru Dan Burdescu¹, Marian Cristian Mihăescu¹
Costel Marian Ionașcu² and Bogdan Logofătu³

¹University of Craiova, Software Engineering Department, Craiova, Romania

²University of Craiova, Analysis, Statistics and Mathematics Department, Craiova, Romania

³University of Bucharest, CREDIS Department, Bucharest, Romania

{burdescu, mihaescu}@software.ucv.ro, icostelm@yahoo.com
logofatu@credis.ro

Abstract. Within e-learning domain there is a tremendous effort trying to loosely couple the data management and data deployment in e-Learning environments. This paper presents a design of such a architecture which consists of three main pieces: (1) Global Repository, (2) Repository Manager and (3) Web Deployer. The architecture ensures scalability, interoperability and service composition for an enterprise e-Learning environment. The Global Repository consists of a suite of services that actually manage all the assets of the e-Learning environment. The Repository Manager is a data management tool which allows the maintenance of all assets from a desktop application. The Web Deployer tool actually creates e-Learning environments by using assets from the Global Repository. The overall architecture is Service Oriented such that each service publishes its functionality in a standardized way letting other services to access and use the functionality in a flexible manner.

1 Introduction

Enhancing on-line environments become more and more evident in almost all existing systems. One important approach is to use a recommender system for students. This is one important way of upgrading a traditional e-Learning platform. In this way, e-learning becomes the way to empower a workforce with the skills and knowledge it needs. So far, traditional e-learning systems were developed in the means of static software applications, lacking in educational multimedia environments and personalized capabilities and without any interest given to the real users input and feedback [1, 2].

Large amount of hypermedia may be accessed using an e-Learning platform thus leading to large amounts of data regarding user behavior. This data may be used for intelligent user profiling, content reaching and classification, personalized intelligent interface.

This paper presents a process and the outcome of our effort in developing a recommendation mechanism for students that run along an e-Learning environment. The content of the e-Learning platform is centrally administered and is supposed to be custom structured. An e-Learning infrastructure needs to be set up such that disciplines and associated materials are presented in a structured way. The platform is supposed to have the capability of capturing and saving user's performed activities.

Second section presents related work regarding intelligent recommender systems. The third section presents the e-Learning content representation and fourth section presents the logging mechanism of student's interactions. These two sections represents a prerequisite for the e-Learning environments that want to obtain recommendations from IRS. Section five presents the methodology of building learner's model, section six presents the procedure for computing the knowledge level for a student and section seven presents the procedure for obtaining recommendations. These sections represent the core business logic of IRS. Section eight presents the software architecture of IRS. Main software components with functionalities and interactions are presented. Section nine presents how IRS and client e-Learning platform are put together. Finally, conclusions are presented.

2 Related Work

Substantial research in the design of technology enhanced learning has been conducted within programs such as the Web-based Inquiry Science Environment [3], BioKids [4], Thinkertools [5], Modeling Across the Curriculum [6], and other research programs.

These projects have investigated the use of models and visualizations, idea maps and modeling engines, graphs and data probes, collaboration, and inquiry learning. They have pushed the boundaries of what technology can do, providing new functionality for curriculum designs and scaffolding students, teachers and classrooms. Such research promises to transform learning from its traditional form of knowledge dissemination to new forms of knowledge community and inquiry learning [7]. This represents the next-generation of e-Learning systems that use innovative methods and materials and finally promote an intelligent community curriculum developers.

Another research direction courses web-usage-based mining [15]. In this direction there are major contribution regarding preprocessing and preparing of data [16], recommendation of actions in e-Learning according performed actions [17].

There were proposed models for assisting evaluation of learner's in e-Learning systems [18]. Implementing many of these research directions has been done using data mining algorithms [19]. There were employed clustering algorithms [20], algorithms for obtaining sequential models [21], algorithms for association rule creation [22, 23].

These research directions concretized into non invasive recommendation systems for learners [24]. Such system is also employed for obtaining recommendations regarding the materials that need to be studied by learners [25].

Automatic recommendation implies that the user profiles are created and eventually maintained dynamically by the system without explicit user information. Examples include amazon.com's personalized recommendations and music recommenders like Mystrand.com in commercial systems [26], and smart recommenders in e-Learning [28], etc. In general, such systems differ in the input data, in user profiling strategies, and in prediction techniques.

Several approaches for automatic personalization have been reported in the literature, such as content-based or item-based filtering, collaborative filtering, rule-based filtering, and techniques relying on Web usage mining, etc [27]. In the e-learning

area, one of the new forms of personalization is to give recommendations to learners in order to support and help them through the e-Learning process.

To deliver personalized content to users with diverse backgrounds, data mining techniques have been used in e-learning systems in recent years [29, 30]. The data mining approach uses all the available information about existing users, such as system logs, to learn user models and then use these models for personalization..

3 e-Learning Content Representation

Concept mapping may be used as a tool for understanding, collaborating, validating, and integrating curriculum content that is designed to develop specific competencies. Concept mapping, a tool originally developed to facilitate student learning by organizing key and supporting concepts into visual frameworks, can also facilitate communication among faculty and administrators about curricular structures, complex cognitive frameworks, and competency-based learning outcomes.

To validate the relationships among the competencies articulated by specialized accrediting agencies, certification boards, and professional associations, faculty may find the concept mapping tool beneficial in illustrating relationships among, approaches to, and compliance with competencies [6].

Figure 2 is a typical example of a concept map [14]. It addresses a simple question, "What is a plant?" and illustrates how cross-links can be made and concepts organized. This is just one way the concepts listed on the left of the figure can be positioned in a hierarchical fashion to show relationships. The list of concepts and the relationship between them depend entirely upon professor because different professors create different concept maps on the same subject. This is the reason why the concept map associated with a subject is a point of continuous improvement.

Recent decades have seen an increasing awareness that the adoption of refined procedures of evaluation contributes to the enhancement of the teaching/learning process. In the past, the teacher's evaluation of the pupil was expressed in the form of a final mark given on the basis of a scale of values determined both by the culture of the institution and by the subjective opinion of the examiner. This practice was rationalized by the idea that the principal function of school was selection - i.e., only the most fully equipped (outstanding) pupils were worthy of continuing their studies and going on to occupy the most important positions in society.

The usage of concept maps has a proper motivation. Using this approach, the responsibility for failure at school was to be attributed exclusively to the innate (and, therefore, unalterable) intellectual capacities of the pupil. The learning/ teaching process was, then, looked upon in a simplistic, linear way: the teacher transmits (and is the repository of) knowledge, while the learner is required to comply with the teacher and store the ideas being imparted [7].

Usage of concept maps may be very useful for students when starting to learn about a subject. The concept map may bring valuable general overlook of the subject for the whole period of study.

It may be advisable that a concept map should be presented to the students at the very first meeting. This will help them to have a good overview regarding what they will study.

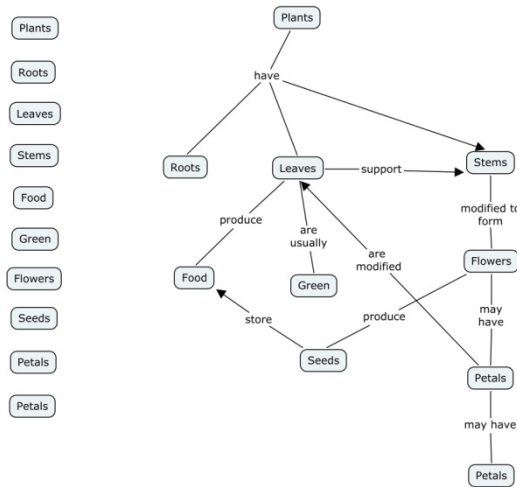


Fig. 1. An Example of a Concept Map (Novak, The Institute for Human and Machine Cognition) [14].

The structuring of the teaching course material is performed according to the instructional objectives defined by course manager. This structuring will allow dynamically build custom e-Learning course materials in accordance with learner’s previous activity, current knowledge model and goals that were set up by administrators and learner himself.

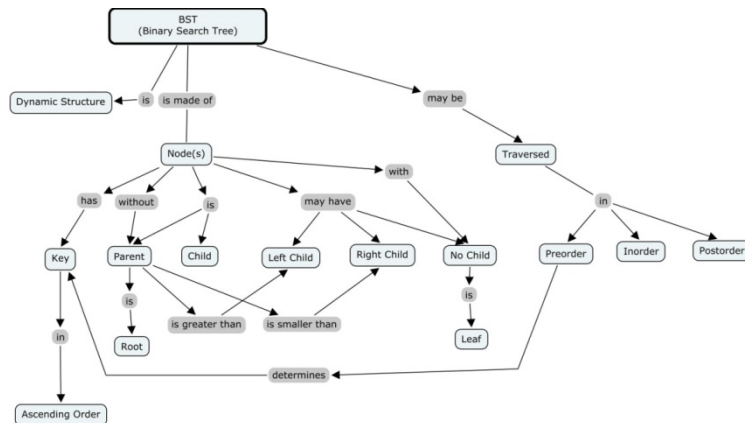


Fig. 2. The structure of a course.

The knowledge representation model has a direct influence regarding the outcome of the construction procedure of the learning material. Under these circumstances the knowledge model becomes the central part of the building mechanism. It may be regarded as a reference point for all learners. It may be assimilated to a representative virtual learner for all learners that contributed to experience repository data.

(wbe 2010)Once a course manager has been assigned a discipline he has to set up its chapters by specifying their names and associated document. For each chapter course managers have the possibility of setting up two pools of questions: one for

testing and one for final examination. The e-Assessment tool offers also the possibility the structure of testing and examination activities.

The course manager specifies the number of questions that will be randomly extracted for creating a test or an exam. Let us suppose that for a chapter the professor created 50 test questions and 60 exam questions and he has set to 5 the number of test questions and to 10 the number of exam questions that are to be randomly withdrawn. It means that when a student takes a test from this chapter 5 questions from the pool of test question are randomly withdrawn. When the student takes the final examination at the discipline from which the chapter is part, 15 questions are randomly withdrawn: 5 from the pool of test question and 10 from the pool of exam question. This manner of creating tests and exams is intended to be flexible enough for the professor. In Figure 1 it is presented the logical structure of the Database discipline.

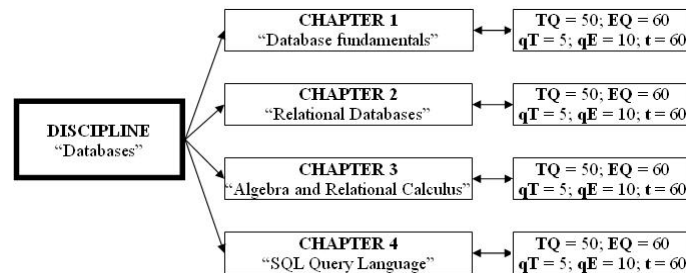


Fig. 3. Structure of Database discipline.

All tests and exams are taken under time constraints. For each chapter the professor sets up a number of seconds necessary to answer questions that chapter. When a test or exam is taken all the seconds are summed thus obtaining a maximal interval of time in which the student has to finish the test. The elapsed and remaining time are managed on server side and presented to the student after each answered question. In Figure 1 the time parameter is denoted by t and its value is 60 seconds for each question. Figure 2 presents how the management of the Database discipline looks in the e-Assessment tool.

The professor has also enough flexibility for creating and editing questions. A question may contain pictures, thus equations, formulas or other graphics may be imbedded into it. For each question the professor sets up the visible answers and the correct answers. For example, if a question has four possible answers he will have to check the checkboxes called A, B, C and D stating that the student will have four choices. Error checking is enforced such that when a question has three visible answers and the correct answer is D an error is shown to professor and the question cannot be saved.

4 Logging Student Interactions

Tracking of actions is required to facilitate the data collection for the needed performed actions. The design of actions tracking module needs to select proper features that describe the performed actions of the learned. Among these features (or

attributes) there are: user id, the date and time the action has been performed, details about the performed action, and so on. From the design phase of the platform, there were adopted two methodologies for monitoring actions.

Since the business logic of the platform is Java based, *log4j* utility package was employed as a logging facility and is called whenever needed within the logic of the application. The utility package is easy to use; *log4j.properties* properties file manages the logging process. The setup process states the logs are saved in *idd.log* file. The main drawback of this technique is that the data from the file is in a semi-structured form. This makes the information retrieval to be not so easy task to accomplish. On the advantages, logging activity may be very helpful in auditing the platform or even finding security breaches. This logging facility is also very helpful when debugging during development or when analyzing peculiar behavior during deployment.

To overcome the semi-structured shape of logged activity a structured way of gathering activity information was enforced. The activity table was added in the database and all actions were recorded in the manner of one record per action. In the table 1 it is presented the structure of activity table.

Table 1. Structure of activity table.

Field	Description
id	primary key
userid	identifies the user who performed the action
date	stores the date when the action was performed
action	stores a tag that identifies the action
details	stores details about performed action
level	specifies the importance of the action

In Table 1 the action field is represented by a tag. The detailed explanation of what the tag means is set in a properties file. For each language a separate properties file is created, each file containing the same tags but with description in a different language.

The details field stores specific information regarding the action that was executed. For example, if a secretary modifies the profile of a student in the details field there will be stored information about what fields were updated.

The level field specifies the importance of the executed action. There are defined three level of importance: 0, 1 and 2 where level 0 specifies the critical actions.

After five months of deployment, the activity table contains more than 50,000 records and we suppose that until the end of the learning cycle there will be close to 100,000 records. All this logged activity may also be very helpful in an audit process of the platform. The records from the activity table represent the raw data of our analyzing process.

Figure 4 presents a sample of the log with the performed activities by students.

```

$URL = http://apps.software.ucv.ro/tesys/servlet/tesys?actionCommand=Student
2: $URL&actEvent=login&username="student"&passwd="student"
4: $URL&actEvent=listCourses
5: $URL&actEvent=listChapters&materieid=1
15: $URL&actEvent=viewTests&materieid=1
10: $URL&actEvent=viewTest&testid=3
75: $URL&actEvent=listCourses
500: $URL&actEvent=takeTest&materieid=1&capid1=2&capid2=3
55: $URL&actEvent=readMessage&messageld=3
20: $URL&actEvent=sendMessage&receiverId=5&senderId=2&text="test message"
3: $URL&actEvent=logout

```

Fig. 4. Sample scenario for a virtual learner.

5 Building Learner's Model and Predicting Needed Resources

Clustering is one of the most useful tasks in data mining process for discovering groups and identifying interesting distributions and patterns in the underlying data. Clustering problem is about partitioning a given data set into groups (clusters) such that the data points in a cluster are more similar to each other than points in different clusters [8].

In the clustering process, there are no predefined classes and no examples that would show what kind of desirable relations should be valid among the data that is why it is perceived as an unsupervised process [9]. On the other hand, classification is a procedure of assigning a data item to a predefined set of categories [10].

Clustering produces initial categories, in which values of a data set are classified during the classification process. From all clustering algorithms categories we chose to have a closer look on those that use partitioning methods. k-Means algorithm [11] is taken since it is simple and straightforward. The steps of clustering process are presented in Figure 3. The procedure follows the standard knowledge discovery [10] but is accustomed for our specific situation.

In the k-Means algorithm, given a database of n objects and k , the number of clusters to form, a partitioning algorithm organizes the objects into k partitions ($k \leq n$), where each partition represents a cluster. The clusters are formed to optimize an objective partitioning criterion, often called similarity function, such as distance, so that objects within a cluster are "similar", whereas the objects of different clusters are "dissimilar" in terms of database attributes. So, the first step is to define a list of attributes that may be representative for modeling and characterizing student's activity.

The classic k-means algorithm is a very simple method of creating clusters. Firstly, it is specified how many clusters are being thought: this is the parameter k . Then k points are chosen at random as cluster centers. Instances are assigned to their closest cluster centre according to the ordinary Euclidean function. Next the centroid, or the mean, of all instances in each cluster is calculated – this is the "means" part. These centroids are taken to be the new centre values for their respective clusters. Finally, the whole process is repeated with the new cluster centers. Iteration continues until the same points are assigned to each cluster in consecutive rounds, at each point the cluster centers have stabilized and will remain the same thereafter [12].

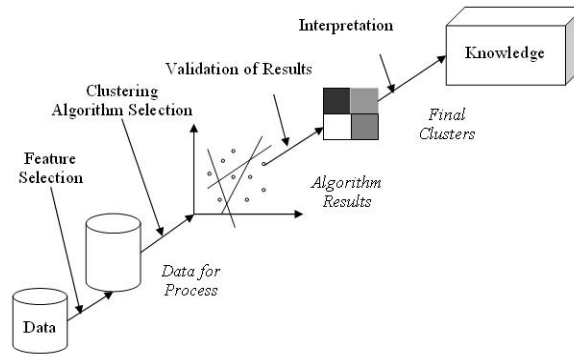


Fig. 5. Steps of clustering process.

From a different perspective, for a cluster, there may be computed the following two parameters:

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (\text{the means})$$

$$\sigma = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \dots + (x_n - \mu)^2}{n - 1}$$

(the standard deviation)

The sum of all probabilities for all clusters is 1. If we know which of the distributions each instance came from, finding the parameters is easy. On the other hand, if the parameters are known finding the probabilities that a given instance comes from each distribution is easy. The general clustering algorithms result in crisp clusters, meaning that a data point either belongs to a cluster or not. The clusters are non-overlapping and this kind of partitioning is further called crisp clustering.

Cluster validity assessment is the final step in the process and practically validates the results. There are two criteria proposed for clustering evaluation and selection of an optimal clustering scheme [9]: compactness and separation.

Also, the influence of the maximum number of clusters C_{max} , related to the weighting factor, in the selection of the optimal clustering scheme is discussed in [12].

6 Computing the Knowledge Level

The main input data for computing the knowledge level is represented by the context data from the platform and the activity performed by the learner taken into consideration. Next lines present a sample XML file with data representing the context of the platform.

```
<module>
  <id>1</id>
  <name>Computer Science</name>
```

```

<discipline>
  <id>1</id>
  <name>Algorithms and Data Structures</name>
  <chapter>
    <id>1</id>
    <name>Binary Search Trees</name>
    <quiz>
      <id>1</id>
      <text>text quiz 1</text>
      <visibleAns>abcd</visibleAns>
      <cotectAns>a</cotectAns >
    </quiz>
    .....
  </chapter>
</chapter>
  <id>2</id>
  <name>Height Balanced Trees</name>
</chapter>
.....
</discipline>
</module>

```

One of the critical issues regarding knowledge measurement is represented by weight assignment procedure for quizzes. Weight assignment is automatically performed according with data recorded within Experience repository module. The weight of a quiz is computed according with the number of learners that answered incorrectly to that quiz. If n is the total number of learners that answered a question q_i and n_{wrong} is the number of students that answered incorrectly the question than the formula for computing the weight of the quiz is:

$$\text{weight}(q_i) = \frac{n_{\text{wrong}}}{n}$$

This formula will assign values between 0 and 1. Values closer to 1 mean high weight and are obtained for high number students that answered incorrectly to questions. For example, if 25 students answered to a quiz and 20 of them answered incorrectly this means the weight of the quiz is 0.8.

7 Procedure for Obtaining Recommendations

For Algorithms and Data Structures disciplines there was created an infrastructure for Binary Search Trees chapter. For this chapter there were defined 16 concepts as in figure 4. For each concept there are created 5 to 10 quizzes. Each instance is represented by a set of 3 features: nL -number of loggings, nT -number of taken tests, $avgW$ -average of weight for taken test.

- K-means clustering algorithm is performed and 4 clusters are obtained. The distribution into clusters is:

- Cluster 0 – 25 instances – 12.5 %
- Cluster 1 – 63 instances – 31.5 %
- Cluster 2 – 80 instances – 40 %
- Cluster 3 – 32 instances – 16 %

The virtual representatives have the characteristics presented in table 1.

Table 2. Cluster characteristics.

Cluster ID	nL	nT	avgW
0	<10	<5	<0.3
1	10 50	5 ... 10	0.3 ...0.5
2	50 ... 100	10 ... 20	0.5 ... 0.7
3	>100	>20	>0.7

When a learner with 15 loggings and 5 taken tests enters the e-Learning platform he is placed as belonging to cluster 1. For him, the target cluster is cluster 2.

All concepts are taken into consideration for the virtual representative of cluster 1 and for analyzed user. Table 2 presents the value of weights for some concepts.

Table 3. Weights of concepts for a cluster representative and analyzed learner.

Concept	C1	C2	C3
Representative of cluster 1	0.2	0.4	0.5
Analyzed learner	0.1	0.1	0.2

For a threshold value of 0.2 there are obtained concepts C2 and C2. These concepts need more study and thus their associated document is presented as recommendation. More than this, the associated quizzes are recommended for testing purposes.

After the learning session all performer activities are saved into Experience learner repository. Each time the learner will log in the e-Learning environment his portfolio of actions is rebuild and the cluster to which the learner belongs is determined.

Within the knowledge repository model the model is rebuild at certain intervals, for example 2 days. The newly obtained model represents a challenger model.

The quality of the challenger model is determined by computing the likelihood of a test dataset. Measure of goodness-of-fit is determined by the logarithm of the likelihood, or log-likelihood: and the larger this quantity, the better the model fits the data. Instead of using a single test set, it is also possible to compute a cross validation estimate of the log-likelihood. For our instances the value of the log-likelihood is -2.4 which represent a promising result in the sense that instances (in our case students) may be clustered in four disjoint clusters based on their used features.

Validation of results produces the final clusters that implement the model. The validation procedure has two main outcomes: firstly it proves the correctness of results for current dataset and gives an idea of how the model will perform on new data.

Classification using Bayes technique was also used for obtaining recommendations

regarding assets (e.g. documents, quizzes) that need more attention. For each resource there was computed a context in which a label is defined. The features that describe

an instance are nT – number of tests, $avgW$ – average weight of taken tests and fG – final grade. Table 3 presents few sample contexts.

Table 4. Sample resource features and tag.

id	nT	avgW	fG	tag
0	5	0.3	4	yes
1	10	0.5	5	yes
2	60	0.7	8	no
3	65	0.76	7	no

In table 4 may be seen that for resource 1 there were taken 10 tests with an average weight of 0.5. The final grade was five and thus the recommendation is *yes*. All these data represent the training data for the Naive Bayes classifier that will be used for determining what resourced may be recommended.

The performance of the model was evaluated by using 10-fold cross validation technique. The results are presented as percentage of correctly classified instances (78.71%) and incorrectly classified instances (21.19%) and confusion matrix.

Whenever needed, the learner may issue a request for advice. All his performed activities regarding activities are taken into consideration and the classifier is run such that a label is created for each resource. Resources that are labelled with *yes* are recommended to the learner as needing further study.

When a learner enters the platform, he will automatically be clustered and, thus, the next clustered with higher value of average coverage is determined. If there is no better cluster, than the clustering process will be performed for the students that belong to the very same cluster, thus obtaining new clusters.

The procedure that determines the advice is:

```

procedure DetermineAdvice(Learner l, Collection collection)
{
  K = classifyLearner (l, collection);
  T = determineTargetCluster (K, collection);
  for (int s=0; s < NC; s++) {
    switch distance( coverage(cs, K), coverage(cs, T) )
      case "veryLow": return NO_ADVICE;
      case "low" : return NEED_EXAMPLE;
      case "avg" : return NEED_DETAILED_DESCRIPTION;
      case "high": return NEED_OVERVIEW;
    }//end for
  }//end procedure

```

The above procedure determines the advice for a learner l having already computed a collection of clusters named $collection$. Firstly, it is determined the cluster K to which the learner l belongs. Cluster K is one of the clusters produced by the clustering process. Then it is determined the target cluster T . This is the cluster with higher weight coverage than K . Once clusters K and T are determined each concept is taken into consideration. NC represents the number of concepts and s is iterator. For each concept it is computed the distance between the values of coverage for concept cs in both K and T clusters. If the distance is very low than any advice is issued. If the distance is low than it seems the learner needs to make small progress in understanding that

concept. That is why, he just needs to access the examples document that is assigned to that concept. If the distance is average than it seems the learner needs to make more progress in understanding that concept. In this situation the learner is advised to study the detailed description document that is assigned to that concept. If the distance is high we may say that learner has no knowledge regarding the discussed concept. In this situation the advice is to start studying the concept overview.

The distance function makes intensive usage of concept maps. The concept maps are used for computing the quantity of knowledge a learner has accumulated at a certain moment in time. The concept map is a graph where nodes are represented by concepts. Once a user has activity and he has answered tests questions there may be computed the weight of the graph which represents the quantity of accumulated knowledge. Thus, for a cluster of learners there may be determined a representative learner (which is a virtual learner) whose level of accumulated knowledge represents the average of accumulated knowledge for all learners that belong to that cluster. In this manner, there might be determined a representative learner for all cluster and thus, an Euclidian distance may be defined.

Once the advising process is finished the user is invited to have spent some time for actually following the received advice. Once this is accomplished, the testing activities may resume. All performed activities are again monitored and saved along with older ones. Whenever the learner seems its appropriate he may ask again for advice. At this point, the learner is clustered again. This means he is placed in one of the existing clusters. Determining the advice will follow the same procedure. Hopefully, due to previous advice, the learner is in a “better” cluster and thus the advices are more oriented towards examples. If this is not the situation it means that the learner has not progressed in improving his knowledge coverage of the concepts and thus advices regarding a better reading of the overview document are issued.

The analysis process has the goal of determining clusters of users/students based on their activity. Each user represents an instance for the clustering process and is represented by a set of parameters. The obtained clusters are used for classifying students. A student may be “moved” from one cluster to a “better” one by determining the concepts that are not covered well. This is done by a weight function that computes a generic weight, which is representative for that cluster. In this way, clusters may be “ordered” and a Euclidian function may be defined.

The clustering process is used for putting together students with the similar activity. The activity is represented by performed actions of learners within the e-Learning environment. Once clusters are created there is used a weight function for computing the representative learner for each cluster. The procedure will recommend concepts based on the distance function presented further in this section.

Firstly, the parameters that represent a learner and their possible values must be defined. For this study the parameters are: *nLogins* – the number of entries on the e-Learning platform; *nTests* – the number of tests taken by the learner; *noOfSentMessages* – the number of sent messages to professors; *chapterCoverage* – the weighted chapter coverage from the testing activities. Their computed values are scaled to one of the following possibilities: VF – very few, F – few, A – average, M – many, VM – very many. The number of attributes and their meaning has a great importance for the whole process since irrelevant attributes may degrade classification performance in sense of relevance. On the other hand, the more attributes we have the more time the

algorithm will take to produce a result. Domain knowledge and of course common sense are crucial assets for obtaining relevant results.

After the parameters have been named they are computed for each learner that participates in the study. In Figure 3 this step is named Feature Selection and produces the Data for process. Data is represented by the whole history of all users that may be found in relations of the database (e.g., activity, exam results, test results, messages, etc.) and in semi structured log files. The Feature Selection will produce the set of instances (sometimes called points) that will represent the input for Clustering Algorithm Selection. Depending on algorithm a number of clusters is obtained each instance being assigned to one or more clusters.

In the next paragraphs there will be described in detail the whole process of knowledge discovery. Everything starts with the data from the database of the e-Learning platform.

The database of the platform contains 21 relations. Among the most important ones are: user, role, userrole, usersections, sections, questions, testquestions, examquestions, testresults, examresults, messages and activity.

The preparation gets data from the database and puts it into a form ready for processing of the model. Since the processing is done using custom implementation, the output of preparation step is in the form of an arff file. Under these circumstances, we have developed an offline Java application that queries the platform's database and crates the input data file called activity.arff. This process is automated and is driven by a property file where there is specified what data/attributes will lay in activity.arff file.

For a student in our platform we may have a very large number of attributes. Still, in our procedure we use only four: the number of logings, the number of taken tests, the number of sent messages and the weighted chapter coverage from the testing activities. Here is how the arff file looks like:

```
- @relation activity
- @attribute nLogings {VF, F, A, M, VM}
- @attribute nTests {VF, F, A, M, VM}
- @attribute noOfSentMessages {VF, F, A, M, VM}
- @attribute chapterCoverage {VF, F, A, M, VM}
- @data
- VF, F, A, A,
- F, A, M, VM,
- A, M, VM, A, V,
- VM, A, VM, M,
```

As it can be seen from the definition of the attributes, each of them has a set of five nominal values from which only one may be assigned. The values of the attributes are computed for each student that participates in the study and are set in the @data section of the file. For example, the first line says that the student logged in very few times, took few tests, sent an average number of messages to professors and had average chapter coverage.

In order to obtain relevant results, we pruned noisy data. We considered that students for which the number of logings, the number of taken tests or the number of

sent messages is zero are not interesting for our study and degrade performance; this is the reason why all such records were deleted.

With this step, we actually accomplished the Feature Selection step from the clustering process. Now we are ready for running the k-Means clustering algorithm.

Running the EM algorithm from Weka [14] package created three clusters. The procedure clustered 91 instances (34%) in cluster A, 42 instances (16%) in cluster B and 135 instances (50%) in cluster C. The following table shows in which cluster the instances belong after running the EM algorithm.

Table 6. Distribution of instances after EM algorithm.

Instance	Cluster A	Cluster B	Cluster C
1	1	0	0
2	1	0	0
3	0	1	0
...
268	0	0	1

Figure 6 presents the distribution of instances after running EM algorithm, with centroids marked. The figure is generated by plotting each student as a black dot.

The algorithm has $O(n)$ time complexity, where n is the number of instances in the dataset. In the performed experiments, we observed that the number of clusters has no importance regarding algorithm's complexity.

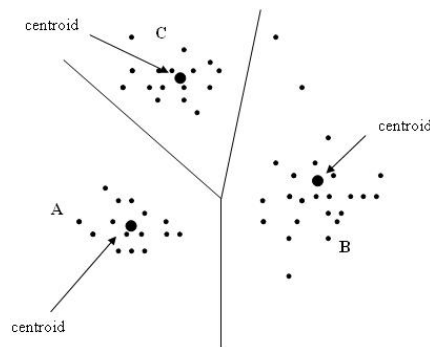


Fig. 6. Distribution of instances after k-Means algorithm with centroids.

The model validation is accomplished by computing the likelihood of a set of test data given the model. Weka measures goodness-of-fit by the logarithm of the likelihood, or log-likelihood: and the larger this quantity, the better the model fits the data. Instead of using a single test set, it is also possible to compute a cross validation estimate of the log-likelihood. For our instances the value of the log-likelihood is -2.61092, which represents a promising result in the sense that instances (in our case students) may be classified in three disjoint clusters based on their activity.

When a learner enters the platform, he will automatically be clustered and, thus, the next clustered with higher value of average coverage is determined. If there is no better cluster, than the clustering process will be performed for the students that belong to the very same cluster, thus obtaining new clusters.

8 Software Architecture of Intelligent Recommender System (IRS)

One of the key issues in making the above presented shift a real success resides in the ability to design a custom knowledge representation model. Therefore, new knowledge management technologies are needed to prove the effectiveness of modern e-Learning environments.

Custom knowledge representation will enable designing context-aware environments and therefore create the premises for shifting towards intelligent e-Learning environments.

The benefits of such approach regard helping professors to prepare high quality e-Learning content. Learners will benefit by being able to access needed learning material such that their knowledge level will optimally increase.

This section presents the architecture of the system that will manage and use the custom knowledge representation model.

The main components that make up the system are:

Central Business Logic Module – this module contains the logic for accessing the e-Learning infrastructure and for sending queries and receiving responses from the Knowledge Miner Module. It also represents the main entry point into the system. Here is gathered the general operation logic.

Experience Repository Module – this module gathers in a structured format all the data regarding the actions that were performed by learners.

Knowledge Model Repository – this module manages the current knowledge model representation. This model is used whenever the intelligent character of an action is needed.

Knowledge Miner Module - this module gathers the business logic for querying the knowledge model repository.

e-Learning Infrastructure Module – this module represents the classical view of an e-Learning environment.

The proper operation of central business logic module and experience repository module is driven by an experience properties file. This file contains the definitions of the actions that are to be logged as experience during the operation of the system.

Knowledge model repository functionality is managed by a properties file which specifies the employed technique for building the model. This properties file has as input the properties file that sets up the experience repository module.

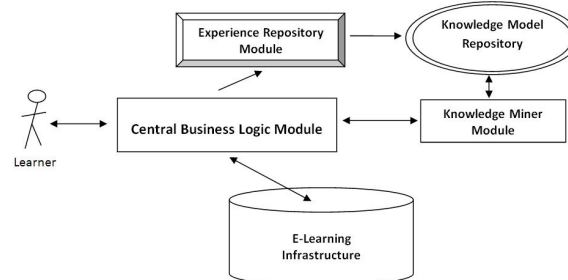


Fig. 7. System's architecture.

Knowledge miner module runs according with the specifications set up by the data analyst. The specifications regard the specific educational goals required by the administrators of the e-Learning environment.

E-learning infrastructure represents the classical view of an e-Learning system. It gathers all the assets managed by e-learning environment: users (e.g. learners, professors, and administrators), disciplines, chapters, course documents, quizzes. It also embeds the needed functionalities for proper running, like security, course downloading, communication, testing or examination.

Central business logic module along with e-Learning infrastructure represents the classical structure of an e-Learning environment. Experience repository module, knowledge model repository and knowledge miner module may be regarded as an intelligent component that runs along the e-Learning environment in order to enhance it.

9 Conclusions and Future Work

We have created a procedure of data analysis, which may be used for improving the level of acquired knowledge of a student.

An e-Learning platform has been set up such that a set of students may study a discipline that is well structured. This means that the discipline is divided into chapters and that each chapter has assigned a concept map. More than this, each concept defined in the concept map has assigned a set of documents and a set of quizzes. The documents that are assigned are: overview, detailed description and examples.

The experiment is performed on 268 students that used Tesys e-Learning platform.

The platform on which the study may be performed needs built-in capabilities of monitoring activities performed by learners. The business goal of the platform is to give students the possibility to download course materials, take tests or sustain final examinations and communicate with all involved parties. To accomplish this, four different roles were defined for the platform: sysadmin, secretary, professor and student.

The process's goal is improve the level of acquired level by custom advising regarding the concepts that need further attention. An off-line application was developed for creating the input data files that are analyzed. Data analysis is done using k-Means clustering algorithm. The main goal of the procedure is to create clusters of students based on their recorded activity.

For k-Means clustering we have computed determined three clusters. The obtained validation values of log-likelihood show a success in clustering users.

Cross-validation technique proved that obtained models may perform in the same manner on new data. The obtained results show that using k-Means algorithm provides sound results.

The experiment proves that machine learning algorithms and concept maps may be successfully used for improving quality of learning. Still, further refines of the procedure may be needed. Improvement in structuring the discipline regards obtaining more adequate concept maps.

Improvement regarding the analysis process may be done using different strategies. Different data source may be considered. Different clustering procedures or even

different machine learning algorithms may also be considered. Having this in mind, different goals may be taken into consideration.

Learner's characterization may have a predictive value in the sense that, from the activities a student has made there, conclusions about his learning proficiency might be pulled out. On the other hand, platform's characterization may have as the result an estimation of the capability of an e-learning system to grade and order students according to their accumulated knowledge. This analysis is critical for having as a conclusion that a system can support generalized tests.

Further experiments need to be run with data obtained from the same platform setup (same disciplines, test and exam questions, etc.) but with other students. In future we will try to follow the same analysis process with data from other platforms.

This business logic of generating advice considers that target cluster the cluster with next immediate higher concept coverage. A good study may be done by always considering as target cluster the cluster with best coverage results. This means that the learner is in a position to have a more rapid learning curve.

The main drawback of the approach proposed in the study regards the complexity of the procedure itself. The main steps (1) data collection, (2) data filtering (3) feature selection (4) running the algorithms and (5) having and interpreting results represent a distinct process that needs custom setting. Automation of the process and integration within an e-Learning platform is the final goal.

Acknowledgements. This work was supported by the strategic grant POSDRU/89/1.5/S/61968, Project ID61968 (2009), co-financed by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007 – 2013.

References

1. M. Sullivan, "Collaboration: teacher pedagogy, student practice", National Teaching and Learning Forum, Vol. 7 No. 5, pp. 4–6, 1988.
2. H. Varenne, "On Pedagogical Authority: One Teacher's Choice", (Online). Available: http://varenne.tc.columbia.edu/hv/pol/pedagog_author.html (last access: December 12, 2009).
3. WISE, (Online), Available: <http://wise.berkeley.edu> (last access: December 12, 2009)
4. BOKIDS, (Online), Available: <http://www.biokids.umich.edu> (last access: December 12, 2009]
5. THINKERTOOLS, (Online), Available: <http://thinkertools.org/> (last access: December 12, 2009]
6. MAC, (Online), Available: <http://mac.concord.org> (last access: December 3, 2009)
7. J. L. Kolodner, P. J. Camp, D. Crismond, B. Fasse, J. Gray, J. Holbrook, S. Puntambekar and M. Ryan, "Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice", *The Journal of the Learning Sciences*, 12 (4), pp. 495-547, 2003.
8. S. Guha, R. Rastogi and K. Shim, "ROCK: A Robust Clustering Algorithm for Categorical Attributes", In *Proceedings of the IEEE Conference on Data Engineering*, Sydney, Australia, pp. 512-521, 1999.

9. M. J. A Berry and G. Linoff, *Data Mining Techniques For Marketing, Sales and Customer Support*. John Wiley & Sons, Inc., USA, 1996.
10. M. U. Fayyad, G. Piatetsky-Shapiro, P. Smuth and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*. AAAI Press, 1996.
11. I. H. Witten and E. Frank, *Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, 2000.
12. M. Halkidi, M. Vazirgiannis and I. Batistakis, “Quality Scheme Assessment in the Clustering Process”, *Proceedings of PKDD*, Lyon, France, 2000.
13. Weka, Available: www.cs.waikato.ac.nz/ml/weka
14. J. D. Novak, “Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations”, Mahwah, NJ: Lawrence Erlbaum Associates, 1998.
15. J. Srivastava, R. Cooley, M. Desphande and P.N. Tan, “Web usage mining: Discovery and applications of usage patterns from web data”. *SIGKDD Explorations*, 2000.
16. R. Cooley, B. Mobasher and J. Srivastava, “Data preparation for mining World Wide Web browsing patterns”, *Knowledge and Information Systems*, 1999.
17. O. R. Zaiane, “Building a recommender agent for e-Learning systems”, *Proc. of 7th Int. Conf. On Computers in Education*, Auckland, New Zealand, 2002.
18. L. Guo, X. Xiang and Y. Shi, “Use web usage mining to assist background online e-Learning assessment”, *4th IEEE ICALT*, 2004.
19. W. Abramovicz, T. Kaczmarek and M. Kowalkiewicz, “Supporting topic map creation using data mining techniques”, *Australian Journal of Information Systems*, 2004.
20. J. Han and M. Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufman, San Francisco, 2001.
21. R. Srikant and R. Agrawal, “Mining sequential patterns: Generalizations and performance improvements”, *EDBT: Proc. of teh 5th Int. Conf. On Extending Database Technologies*, 1995.
22. P. N. Tan, V. Kumar and J. Srivastava, “Selecting the right objective measure for association analysis”, *Information Systems*, 2004.
23. R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules”, *Proc. of the 20th Int. Conf. on Very Large Databases*, Santiago, Chile, 1994.
24. W. Lin, S.A. Alvarez and C. Ruiz, “Efficient adaptive-support association rule mining for recommender systems”, *Data Mining and Knowledge Discovery*, 2002.
25. E. Spertus and L. Stein, “A hyperlink-based recommender system written in squeal”, *Proc. ACM CIKM’98 Workshop on Web Information and Data Management*, 1998.
26. B. Mobasher, “Data Mining for Web Personalization”, *The Adaptive Web: Methods and Strategies of Web Personalization*, Lecture Notes in Computer Science, New York, 2006, Vol. Springer-Verlag, Berlin-Heidelberg.
27. O. Nasraoui, “World Wide Web Personalization”, Invited chapter in “*Encyclopedia of Data Mining and Data Warehousing*”, 2005, J. Wang, Ed, Idea Group.
28. O. R. Zaiane, “Building a Recommender Agent for e-Learning Systems”, in *Proc. of the 7th International Conference on Computers in Education*, Auckland, New Zealand, December, 2002, 3 – 6, pp 55-59.
29. C. Romero, S. Ventura, “Educational Data Mining: a Survey from 1995 to 2005,” *Expert Systems with Applications*. Elsevier 1:33 (2007) 135-146.
30. C. Romero, S. Ventura, J. A. Delgado and P. D. Bra, “Personalized Links Recommendation-Based on Data Mining in Adaptive Educational Hypermedia Systems,” in *Creating New Learning Experiences on a Global Scale*, Springer Berlin /Heidelberg(2007), pp.292-306.

Model Driven Service-Oriented Approach for Smart Building Energy Management

¹Ding-Yuan Cheng, ²Nasaraf Shah, ³Chen-Fang Tsai
²Kuo-Ming Chao and ¹Chi-Chun Lo

¹Institute of Information Management, National Chiao-Tung University
300, Hsin-Chu, Taiwan

²DSM Research Group, Department Computing and the Digital Environment
Coventry University Coventry, CV1 5FB, U.K.

³Department of Industrial Management, Aletheia University, Taiwan, ROC
cclo@faculty.nctu.edu.tw, kewas@iim.nctu.edu.tw
n.shah@coventry.ac.uk, k.chao@coventry.ac.uk
au1204@mail.au.edu.tw

Abstract. This paper describes a new framework which is featured by integrating two well-known system modeling methods namely model driven approach and service-oriented architecture to facilitate web service development and to effectively satisfy a group of service consumers' subjective requirements and dynamic preferences. These models can be combined at different levels and configured into heterogeneous structures to form various building blocks for development of service-oriented applications. A case study is used to illustrate how to systematically use our proposed framework to construct a smart service-oriented environment system for effective energy management.

1 Introduction

Service-oriented computing has become an important trend in IT development. Especially, the recent convergence of ubiquitous computing and context-aware computing in an attempt to integrate numerous types of sensors, heterogeneous communication protocols, and programming languages to produce an effective and efficient solution to the design of a distributed smart environment has seen as a considerable challenge. A number of methods, frameworks and tools to design a smart environment using service oriented approach have been proposed in the past [1][2][3]. In general, the aim of these systems is to maximize users' comfort level, and minimize the cost of software design, code implementation, application installation, and system maintenance. The communications and interactions among sensors, objects, and human participants are inherently complex, as they involve different protocols and languages. It can be viewed as configuration problem that requires various components to work together in cooperative and coordinated manner in order to produce an optimised environment to meet the system requirements. Service-Oriented Architecture (SOA), which offers a way to take sensors, objects, participants as services, can provide an effective approach to improve their communication and coordination, so the human

participants can effectively interact with the system and offer their opinions and preferences through a coherent technology.

A web service can be invoked and executed by a number of consumers simultaneously. Most service discovery and recommending systems recommend a list of services according to the functional or non-functional requirements given by service consumer. Thus, the service consumer can choose the best fit based on a set of criteria. This kind of cases are based on an assumption that is the cardinality between an instance of service and a service consumer is built upon one to one relationship. However, in some cases such as smart environment, an instance of a service needs to serve a number of consumers. In other words, an instance of a service needs to meet various requirements specified by different service consumers. It does not only satisfy one consumer's needs, but also of most consumers. It, however, can be very difficult when conflicting requirements occur. In order to address the above issue, services should be able to conduct self-adjusting and self-organizing to meet the requirements or preferences given by most consumers along with changing environments.

Hence it is important that the design of service should consider the user preferences. The configuration or composition of services should also take the environmental factors into account. So the system needs to adapt to environment changes by adjusting or configuring its services dynamically in accordance with the data sent by sensors. For example, the staffs in one office use a collection of the same devices or functions provided by the system. Each function is intended as a service such as light service, air condition service, heating service, and ventilation service, etc. Service consumers could have different preferences or requirements on the services such as room temperature. Some users may subjectively feel hot for the current room temperature, but some may feel cool or comfortable. How to set the temperature for air condition/heating service according to these consumers' preferences can be a challenging issue. Existing research works in the context-aware system do not sufficiently address this issue. To increase interoperability among these autonomous services is required, so these services need to have well-defined interfaces and functions. The context-aware system would encompass the capacities such as group consensus reaching mechanism and service self-organizing mechanism to improve system adaptability.

In this study, we design an intelligent context-aware service system based on Service Component Architecture (SCA) and Model Driven Approach (MDA). It combines these two methods to build a ubiquitous computing environment. The goal of this framework or system is not only to effectively integrate sensors and services but it also provides a systematic and adaptive approach to construct services in a dynamic environment.

The remainder of the paper is organized as follows. In Section 2 the awareness management process is outlined along with the basic notions required to model the problem domain. Section 3 discusses a SOA platform. We identify the basic services and briefly discuss the basic components of the platform. Then we present an example to illustrate how smart service in an environment involving human users can be used to capture and convey information. Finally we conclude the paper.

2 Related Works

In this section, we review some existing studies and research work in relation to SOA standards and middleware.

Although smart home and smart office concept have recently attracted a great attention from research community and industry in Europe and United States, a number of Smart Home projects within the past few years [4], [5], [6] have been carried out. Also, AT&T Laboratories at Cambridge [7] built a dense network of location sensors to maintain a location model shared between users and computing entities. Microsoft's EasyLiving [8] focuses on a smart space that is aware of users' presence and automatically adjusts environment settings to suit their needs. More recently, Hewlett Packard's CoolTown [9] provides physical entities with "Web presence" and allows the users to navigate from the physical world to the Web virtual environment by picking up links to Web resources using various sensing technologies. These projects have provided great insight into smart building and environment, but we are particularly interested in three different issues of these technologies, namely service selection, consensus formulation method and service modeling.

2.1 Web Services

Many researchers have focused on approaches or methods on improving service discovery protocols [10][11][12][13][14] to increase accuracy in service discovery and selection. One of the most widely used protocols for publishing service is Universal Description, Discovery and Integration (UDDI) that includes service registry with explicit specifications to enable service advertisement. In our previous studies on the UDDI specification [15], we use the tModel to represent the QoS for web services composition. After that, several researchers have designed the semantic query mechanism into UDDI registry, and mapping RDFS upper concepts to UDDI data model to increase the precision in service selection such as [16] [17] [18].

Services can be classified into atomic or composite [19] services. An atomic service is a basic unit which cannot be decomposed further. A composite service is made of a collection of existing services which can be atomic or composite. Many research works on service composition have used workflow-based approach, Artificial Intelligence (AI) planning, and other modeling methodologies to compose services. In the workflow-based composition methods, [20] proposed composite service definition language (CSDL) to reform the flow. They used a static work flow generation in their proposed platform. Some researchers presented Semantic Web Service composition method based on Model-Driven Architecture (MDA) [20] [22], [23] [25], and UML [21]. These composite services are specified using standard UML model to generate specifications and to produce applications using MDA concepts.

2.2 Consensus Forming Methods

Decision making is one of the most complicated administrative processes in management. In a decision-making process, decision makers may encounter multiple criteria

for evaluation. Therefore, Multiple Criteria Decision Making (MCDM) is one of the most well known branches in decision making. MCDM can be divided into two categories: MODM and MADM. A further discussion about MODM and MADM can be found in Yoon and Hwang [26][27]. MADM has been widely used by decision makers in management processes to evaluate and rank possible alternatives.

In a decision making processes, a group of decision makers could be involved and it is called Group Decision Making (GDM), so the all members' options or preferences have to be considered. Most of the GDM problems have strategic dimensions and can be complicated due to their multi-criteria framework involving many subjective and quantitative factors. Optimal utilization of the time and resources is a key element sought by many GDM methods. Various researchers have focused their efforts on increasing the ability of making quality group decisions[28, 29, 30, 31, 32, 33, 34, 35, 36, 37].

An effective web service discovery mechanism should be able to search and assess services based on their QoS and service contents as well as users' requirements. The service assessment or selection often involves multi-criteria decision-making process [38]. So, the GDM is applicable to service selection when the service consumers have inconsistent or conflicting requirements, as it can be considered as a reasoning process for reaching group consensus on their requirements for web service selection.

2.3 Web Service Modeling Methodology

There are several ways to compose services at the design time. Model-driven Architecture (MDA) [39] is a software architecture framework proposed by the Object Management Group (OMG). MDA consists of a set of standards that assist the system in creation, implementation, evolution and deployment. The key technologies of MDA are Unified Modelling Language (UML), Meta-Object Facility (MOF), XML Meta-Data Interchange (XMI) and Common Warehouse Metamodel (CWM).

MDA emphasized the importance of modeling for the software architecture design. MDA suggests a three-layered approach. The Computation Independent Model (CIM) describes a system from the computation-independent point of view to address structural aspects of the system. The Platform Independent Model (PIM) defines a system in terms of a technology-neutral virtual machine or a computational abstraction. The Platform Specific Model (PSM) consists of a platform model that captures the technical platform concepts and a model geared towards the implementation technique. The lifecycle of MDA development is shown as Fig. 1.

In [40], context-awareness is an essential aspect for service utilization, especially when frequent interactions take place between users and environments. In this paper, a solution for developing context-aware web services applications is proposed. The methodology includes a model driven approach to separate the web application functionality development from the context adaptation at the development phases (analysis, design, and implementation). In essence, context adaptation takes place on top of the web application business functionality to facilitate system evolution.

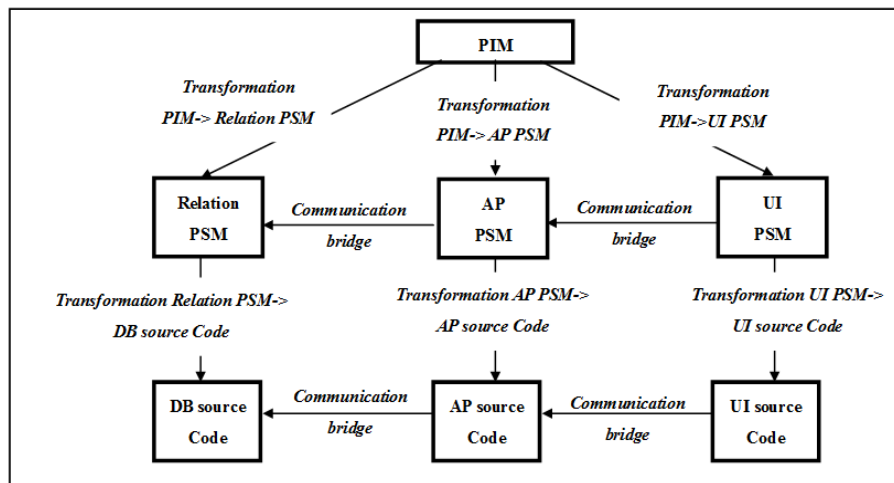


Fig. 1. Lifecycle of MDA Development.

3 Design Framework

Model-Driven Architecture (MDA) is a well-developed concept for software design modeling and implementation. The meta-model plays an important role in MDA. Also, Meta Object Facility (MOF) is the kernel of transformation between different MOF layers. The Unified Modeling Language (UML) is the most widely used in software engineering. The key task of UML modeling is to identify the class, attribute, operation, and their relations. A class might inherit from another class. Moreover, a class can also have many functions which are called operation or method. Therefore, the design concept is not fully applicable to Service-Oriented applications because services are loosely coupled operations or functions and there is no inheritance property and internal facility to store states in services.

Hence, in this work, we proposed an agile SOA modeling methodology which combines MDA and Service-Oriented modeling methods. The main goal is not only to speed up the intelligent system development according to the SOA principles, but also be able to identify group consensus on service's QoS requirement and contents in order to maximize their users' satisfaction. In SOA, a required function can possibly be satisfied by multiple services. Therefore, the process of service selection and discovery needs to consider functional as well as non-functional (i.e. QoS) requirements. The characteristic of our proposed methodology allows the services to have ability of self-adjustment in the process of composition in line with emerging requirements or environment states.

3.1 Overview

There are four kinds of models in MDA. During initial phase, the business analyst analyzes the user requirement in CIM. Next, the PIM based on the outcomes derived

from the previous phase to define the functionalities, the structure, and the behavior of a system. The PSM focuses more on implementation and execution platform which could be a specific programming language or database.

In our propose framework, we distinguish two types of requirements which are functional requirement and comfort requirement in initial phase. After analysis phase, the users can use any tools or modeling language which they are familiar with to model the system. Here, we adopt the service-oriented modeling framework (SMOF) as a modeling framework. In PIM to PSM phases, we use Service Component Architecture (SCA) as transforming methodology. Also in PSM phase, we use the Service Data Object (SDO) to manipulate the connection between application and the database. Our proposed framework is based on MDA that includes SPEF and MOF. The overall architecture is shown in Fig. 2.

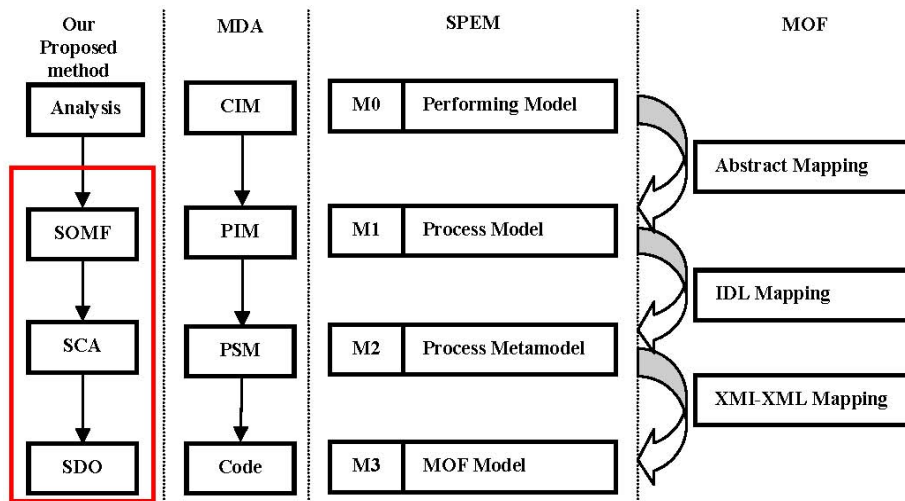


Fig. 2. Life-cycle of our proposed MDA service-oriented approach.

Adopting our proposed methodology to design context-aware systems has the following advantages:

(1) Services-Oriented Modeling: It can reduce the deficiency of object-oriented modeling in service-oriented applications, as UML meta-model cannot provide the necessary support. In service-oriented environment, software and hardware can be represented as services. Services are more transparent and loose coupled, which contain a collection of independent functions or operations as compared to objects. Objects heavily rely on their interdependencies and their internal states to operate.

(2) Autonomous Behavior: A context-aware system based on SOA possesses abilities such as autonomous adjustment, autonomous management, and autonomous deployment to satisfy diverse requirements from multi-user. The group consensus approach collects the preferences from the users and reason over them to provide a basis for system self-adjustment in order to meet the majority of users' requirements.

(3) Annotating sensor data with semantics: Sensor data could be value of temperature, humidity or an expression representing other conditions, but this data could imply a

condition such as light brightness or weather. The sensor devices and their sensed data can be grouped together to become services and annotated with semantics for reasoning.

(4) Information Streams Fusion and resource description: The resource including data, services, computation resource, and device profile will be described explicitly with their location and characteristics. This can benefit locating, allocating and re-deploying resources.

3.2 Top-down Modeling Analysis

Our proposed Model Driven Service-Oriented Approach (MDSOA) provides a top-down modeling analysis method. As mentioned above, we combine MDA concepts in the software design and system implementation with web services. The requirements can be separated into two types: functional requirement, and comfort requirement. Functional requirement defines scale, quantity, and function of all hardware and software. For example, the number of lights, air-condition units, dehumidifier or heater, etc. The comfort requirement is related to users' preferences which are about QoS.

The requirements collected from the previous CIM phase lay the foundation for modeling the required services. An analyst can use any modeling language such as UML to model the requirements. Because some of the service-oriented features cannot be satisfied with UML, it requires another modeling language to specify service flows, service relations, and service capabilities. Hence, we adopt Service-Oriented Modeling Framework (SOMF) [41] in the PIM phase. SOMF is a discipline of modeling business functions and system behaviors based on services.

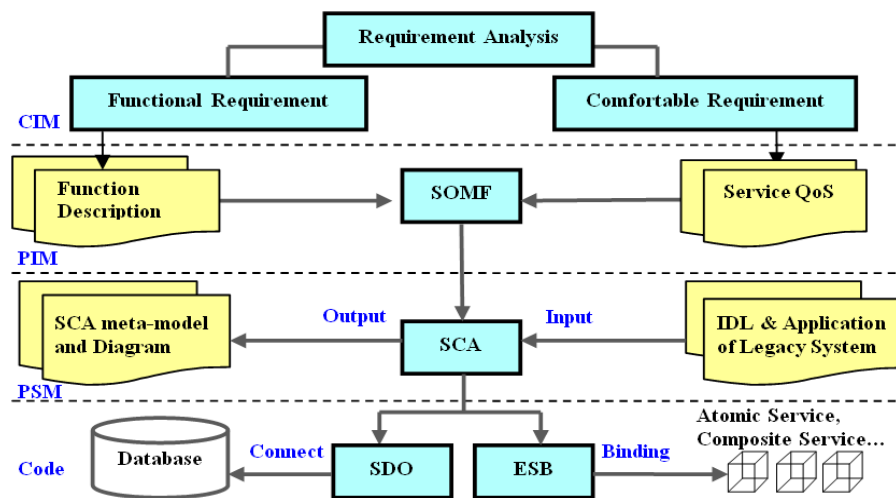


Fig. 3. MDA Mapping Architecture of Top-Down Modeling Analysis.

In PSM phase, the main task is to draw SCA diagram and obtain a system meta-model. After that, the SDO (Service Data Object) and ESB (Enterprise Service Bus)

can connect to database and bind services together. SDO aims to provide a consistent means of handling data within applications, regardless of its source and format. It provides a unified way of handling data of databases and services. ESB is used to integrate applications, coordinate resources, and manipulate information. The proposed architecture is depicted in Fig. 3.

3.3 Button-up Assembling Analysis

From the top-down modeling analysis, we can analyze the system development life cycle from abstraction stage to implementation. However, the system should be adaptable and self-manageable according to changes occurring in the environment, so it is able to take the users' feedbacks and environment changes for system adaption. Hence, we proposed a bottom-up analysis to increase the system's capability in adaption to the environment.

We adopt the SCA assembly model which deals with the aggregation of components and their linkages. The assembly model is independent of implementation language. SCA is a set of specifications which is used to build applications and systems by deploying new service and composing existing components. SCA does not only extend and complement prior approaches to implementing services, but also provide a programming model for building applications and systems based on a SOA.

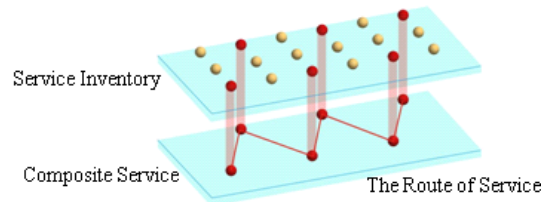


Fig. 4. Bottom-Up Assembling.

The business functions are supported by a series of services which are assembled together to support a particular operational need. The assembly model deals with the aggregation of components and the linking of the components. These services run on the server containers. A service inventory includes a number of server containers.

4 Examples

In this section, we will demonstrate how MDA and SCA approach can be used to develop smart building services for building energy management system. Furthermore, we adopted TOPSIS to resolve conflicts in group decisions.

4.1 Architecture

In the proposed smart office system, each device has a unique ID and it provides spe-

cific service. For example, the motion sensor detects object movement within its designated range. The temperature sensor measures temperature and humidity. These sensors can be considered as an atomic service. All the service descriptions in smart office can be stored in a registry. The overall architecture is shown as Fig 5.

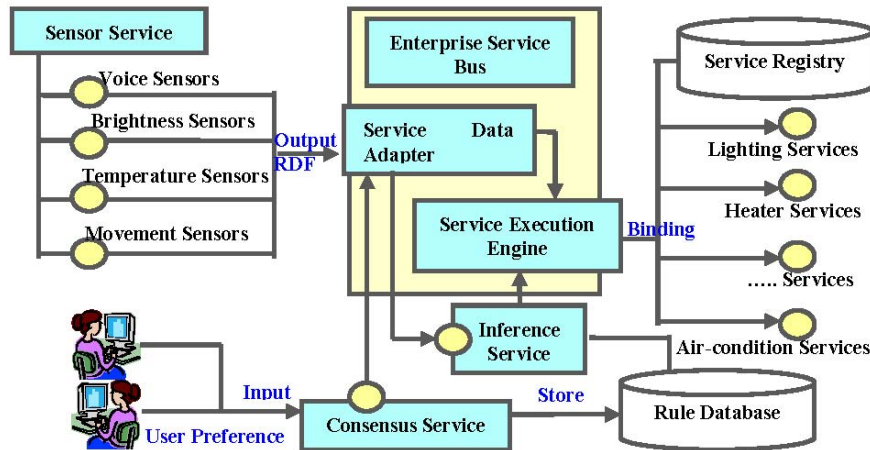


Fig. 5. System architecture.

(1) Services Repository. The aim of semantic web service is to locate services automatically based on the functionalities provided by the web services. These services are registered in a service repository which is UDDI. Therefore, we use the JUDDI to build UDDI environment which provides Business Entities, Service Entities, Binding Templates, and tModels to represent the business details and its services. Services registered with in JUDDI can be searched by name, location, business, bindings or tModels.

(2) Consensus Service. We design a consensus service which is based on the TOPSIS method to reason over a group of users' preferences to identify their potential agreements. These preferences can be very subjective and inconsistent and they could be represented in different ways. The preferences can be associated with uncertainty, fuzziness, and incompleteness. The consensus service is able to identify common requests from the majority of users and made recommendations to the device service to adjust the devices such as the light or air-condition services.

(3) Service Execution Engine. Service Execution Engine (SEE) provides a run time environment for service binding and execution. We use an open source system, Tuscany, which is an Enterprise Service Bus (ESB) for our SEE. SEE is not only a web server container but also can forward the message to other web services accordingly which are deployed on the other server. For achieving goal of service execution, we use the Synapse, which is an open source project as a routing server on Apache web server. Synapse provides a simple, lightweight and fully open source SOA infrastructure to assemble and manage composite applications as well as route messages. Synapse supports HTTP, SOAP, SMTP, JMS, FTP and file system transport for message exchange using XSLS, XPath and XQuery to bind the web services. Fig 6 shows the service execution engine architecture.

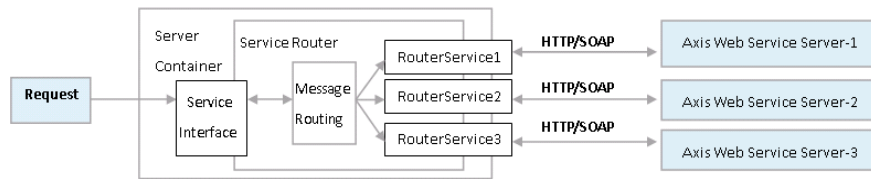


Fig. 6. Service Execution Engine architecture.

(4) Rule Database. The rule database is used to store the expert rules and user preferences. Moreover, we used a tool kit, FuzzyJ, as the fuzzy inference engine in our prototype system. The inference engine contains essential pre-defined knowledge for interpreting and classifying the information e.g., Very Cold, Normal, Warm, Hot, Very Hot. It consists of primitive and composite fuzzy terms, modifier and quantification of fuzzy terms, and fuzzy rules. Primitive terms are a set of atomic terms that represent a collection of raw data collecting by service broker.

(5) Sensor Services. The sensor device is associated with a sensor service which collects and senses the environmental parameters. Sensor service transmits the sensor data to service execution engine. The description of service is represented in RDF. The sensor data in stored in XML format can be parsed consistently by other services.

4.2 Module Design

The prototype system is a smart office system and it is designed by following SCA principles. We designed several composite services which are called SensorComposite, InferenceComposite, UserOpinionComposite, ConsensusComposite and DeviceComposite. The SensorComposite is composed of several components which include FireSensorComponent, SmogSensorComponent, VoiceSensorComponent, MovementSensorComponent, TemperatureSensorComponent and BrightnessSensorComponent. The SensorComposite is responsible for all kinds of sensor services. All sensor data is transmit to InferenceComposite to evaluate whether the change of the device parameters such as brightness of light is necessary or not. Another important mechanism is related to handling the group preferences. The UserOpinionComposite is used to collect users' preferences. The feedbacks and preferences from users are sent to ConsensusComposite to calculate and obtain their consensus if there is any. The result of group consensus is then sent to the DeviceComposite.

The DeviceComposite would also receive the sensor data and preferences from sensors and users. All the data and signals are stored in XML-format file. We model these composite services by using SCA diagram which is shown in Fig. 7.

The system gathers environment related data to generate appropriate information for use. For example, a user may leave his seat, and move to conference room. The motion detection sensor would sense the situation, and the light will be turn off. The patterns of interactions among these appliances and their controls can be driven by analyzing sensor data. Thus, the detailed workflow design or relationship between user and appliance are not necessary to be prescribed at design time. Hence, the proposed framework includes a SCA approach which consists of service-oriented modeling and assembling mechanism. Sensor and intelligent mechanisms can be considered as ser-

vices. It can not only activate component automatically, but also improve the service capability on semantic interpretation and to reuse the existing service to build new composite services to meet new requests.

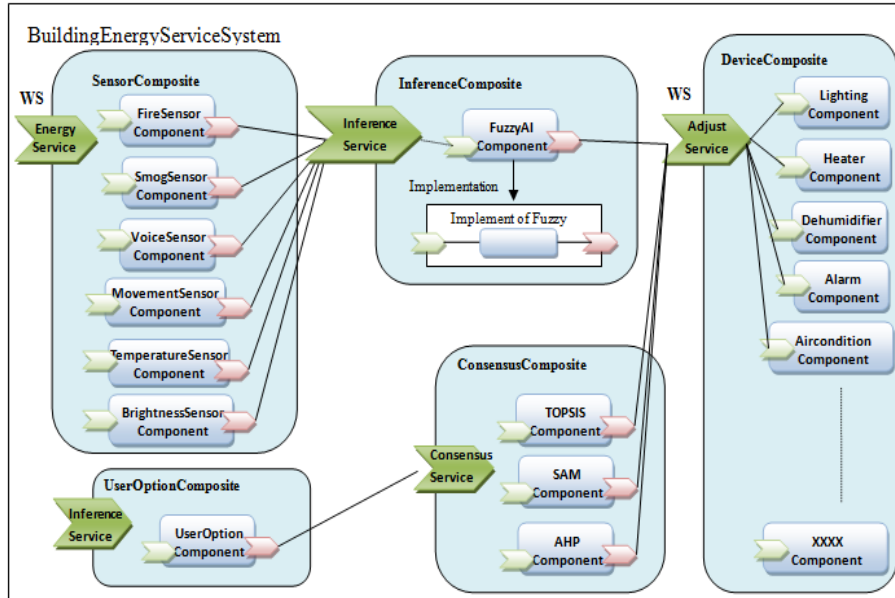


Fig. 7. Service Composite based on SCA.

4.3 Implementation Details

We use a sequence diagram to help readers understand how services are activated and adjusted to meet users' requirements. The system has sensors to detect brightness, voice and temperature and the corresponding devices to control them. The sensor service transmits the sensor data to ESB. The ESB binds the inference service with rules to reason over data and pass the outcomes to the device services. The device services control and adjust the devices by giving appropriate commands and parameters (such as dimmable light and air-condition).

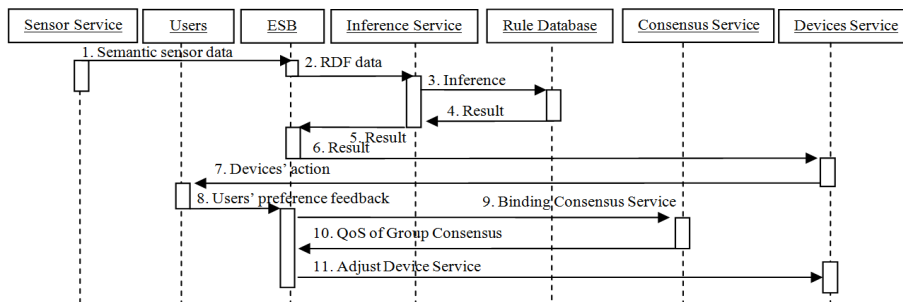


Fig. 8. Sequence Diagram of service support the self-adjust between ESB and users.

If the users feel too hot or gloomy, they can make requests to adjust the light or air-condition. These requests are collected and submitted to the consensus service. Since not every user wants to change the lighting or temperature and they may have conflicting requests or preferences, the consensus service will reason over these data to produce the recommendations which satisfy most users' preferences or requirements. If there is any change required in the environment, the device services will set new parameters and send commands to the device such as light and air-condition according to the recommendations.

5 Discussion

A number of research projects have been established in an attempt to solve the issues associated with context-aware systems. Although sensing and perceptual technologies have been increasingly recognized as key methods to develop smart environment, the main problem is that a group of users might have different requirements in the same space. Achieving context-awareness needs to take into account the variations present in the environment and users' opinions. The system does not only analyze the system functional requirements, but also need to consider the users' preferences.

Combining SCA and MDA could provide an adaptive platform to develop an autonomous adaptation system. The main characteristic of SCA is that it supports declarative foundation which enables it to access and compose services of diverse appliances. Thus a context-aware system can use semantic annotations to locate and bind services dynamically.

6 Conclusions

In this paper, we have demonstrated MDA approach along with SCA concept to develop Service-Oriented applications. The appliances in the framework modeled as services are annotated with semantics to alleviate the difficulty in development. The composition mechanisms, message routing and data driven functions support the intelligent control between appliances and sensors.

Our main contribution is a system modeling methodology which is based on MDA and SCA to facilitate the development of smart building energy management system which often involves complex activities and interactions among sensors, devices and human users. Although sensors can control the devices, it needs to work cooperatively with other devices and human users. In a context-aware system with multiple users involved, it is important to develop a synchronous information stream and fusion. Our proposed Model Driven Service-Oriented approach provides a solution for developers to extend service features to accommodate existing devices or applications without rewriting them.

This study presents an overview of MDA with emphasis on the application of smart office based on SCA framework which provides a mechanism to support necessary steps required for service composition. The developers can take advantage of these existing services and supporting functions to produce composite service for intelligent

control. A prototype of the proposed framework based on a number of existing hardware such as Arduino and software such as Synapse, and Tuscany has been developed to test the feasibility of the proposed approach. Further development of the system by introducing more intelligent rules and repositories is needed.

Acknowledgements. The research is supported by the National Science Council of Taiwan under the grant No. NSC 96-2416-H-009-008-MY3. This research is also partially supported by EU FP7 program under grant agreement no 224609.

References

1. Kapitsaki, G., Kateros, D. A., et al.: Service Composition: State of the art and future challenges. In: Mobile and Wireless Communications Summit, Budapest, 2007
2. Chakraborty, D., Yesha, Y., Joshi, A.: A distributed service composition protocol for pervasive environments. In: Wireless Communications and Networking Conference, Atlanta, 2004
3. Josef, N., Alam, S., Mohammad, C.: Integrating Mobile Devices into Semantic Services Environments. In: The 4th International Conference on Wireless and Mobile Communications, Athens, 2008
4. Georgia Institute of Technology–Georgia Tech Aware Home- available at <http://www.cc.gatech.edu/fce/ahri>, 2002
5. Massachusetts Institute of Technology–House, available at http://architecture.mit.edu/house_n, 1999
6. University of Rochester–Smart Medical Home, available at http://www.futurehealth.rochester.edu/Smart_Home, 2004
7. A. Harter et al., The Anatomy of a Context-Aware Application, Proc. 5th Ann. ACM/IEEE Int'l Conf. Mobile Computing and Networking (Mobi-Com 99), ACM Press, p.p. 59–68, 1999
8. B. Brumitt et al., “Easy Living: Technologies for Intelligent Environments,” Proc. 2nd Int'l Symp. Handheld and Ubiquitous Computing (HUC 2000), LNCS 1927, Springer-Verlag Publication Inc., 2000, p.p. 12–29.
9. T. Kindberg et al., “People, Places, Things: Web Presence for the Real World,” Proc. 3rd IEEE Workshop Mobile Computing Systems and Applications (WMCSA 2000), IEEE CS Press, p. 19, 2000
10. UPnP Device Architecture 1.0, UPnP Forum, Dec. 2003
11. JINI Technology Core Platform Specification, v. 2.0, Sun Microsystems, June 2003
12. Service Location Protocol, v. 2, IETF RFC 2608, June 1999
13. Salutation Architecture Specification, Version 2.0c, Salutation Consortium, June 1, 1999
14. Tom Bellwood, IBM. Luc Clement, Microsoft, Claw von Riegen, SAP, UDDI Version 3.0.2, UDDI Spec Technical Committee Specification, Dec. 2004
15. Chi-Chun Lo, Ding-Yuan Cheng, Ping-Chi Lin, Kuo-Ming Chao, A Study on Representation of QoS in UDDI for Web Services Composition, Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, 2008, p.p. 423-428, ISBN: 978-0-7695-3109-0
16. Anton Naumenko, Sergiy Nikitin, Vagan Terziyan, Jari Veijalainen, Using UDDI for publishing metadata of the semantic web, Springer Boston Publication Inc., Vol 188, p.p. 141-159, 2005

17. Jim Luo, Bruce Montrose and Myong Kang, An Approach for Semantic Query Processing with UDDI, SPRINGER-VERLAG, Lecture notes in computer science, 2005, NUMB 3762, p.p. 89-98, ISSN 0302-9743.
18. Asuman Dogac, Ibrahim Cingil, Gokce Laleci, and Yildiray Kabak, Improving the Functionality of UDDI Registries through Web Service Semantics, Lecture Notes in Computer Science (2002), Volume 2444, Springer Berlin, p.p. 9-18, ISBN 978-3-540-44110-6.
19. M. Papazoglou and D. Georgakopoulos. Guest editor introduction: Service oriented computing. ACM SIGSOFT Software Engineering Notes, 46(10):24–28, 2003
20. F. Casati, S. Ilnicki, and L. Jin. Adaptive and dynamic service composition in EFlow. In Proceedings of 12th International Conference on Advanced Information In Proceedings of 13th International Conference on Advanced Information Systems Engineering (CAiSE), Interlaken, Switzerland, Springer Verlag Publications Inc., June 2001
21. H. Schuster, D. Georgakopoulos, A. Cichocki, and D. Baker. Modeling and composing service-based and reference process-based multi-enterprise processes. In Proceeding of 12th International Conference on Advanced Information Systems Engineering (CAiSE), Stockholm, Sweden, Springer Verlag Publications Inc., June 2000
22. Merabti, M., Fergus, P., Abuelma'atti, O., Yu, H., Judice, C.: Managing Distributed Networked Appliances in Home Networks. Proceeding of the IEEE 96(1), p.p. 166–185 , 2008.
23. Fatima-Zahra Belouadha, Hajar Omrana and Ounsa Roudiès, A model-driven approach for composing SAWSDL semantic Web services, International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010
24. Bernhard Bauer, Jörg P. Müller, MDA Applied: From Sequence Diagrams to Web Service Choreography, ICWE 2004, LNCS 3140, pp. 132–136, 2004
25. Roy Gronmo, Michael C. Jaeger, Model-Driven Semantic Web Service Composition, Proceedings of the 12th Asia-Pacific Software Engineering Conference table of contents, Pages: 79 - 86, 2005, ISBN ~ ISSN:1530-1362 , 0-7695-2465-6
26. Zhengdong Zhu, Yanping Chen, Ronggui Lan and Zengzhi Li, Study of MDA Based Semantic Web Service Composition, Information Technology Journal, Volume 8, Issue 6, 2009
27. B. Bauer and J. P. Muller. Mda applied: From sequence diagrams to web service choreography. In ICWE'04 (LNCS 3140), p.p. 132-136, 2004
28. Simona Motogna, I. Lazăr, B. Pârvi, and I. Czibula, An Agile MDA Approach for Service-Oriented Components, Electronic Notes in Theoretical Computer Science, Volume 253, Issue 1, p.p. 95-110, Oct. 2009
29. Hwang, C. L. and Yoon, K. Multiple Attribute Decision Making: Methods and Applications. New York: Spriner-verlag, 1981
30. Yoon, K. and Hwang, C. L., Multiple Attribute Decision Making: An Introduction. Thousand Oaks: Sage Publications Inc., 1985.
31. G. Bordogna, M. Fedrizzi, G. Pasi, A linguistic modeling of consensus in group decision-making based on OWA operators, IEEE Transactions on Systems, Man and Cybernetics: Part A. Systems and Humans 27, p.p. 126– 132, 1997
32. N. Bryson, Group decision-making and the analytic hierarchy process: exploring the consensus-relevant information content, Computers & Operations Research (23) (1996) 27– 35.
33. C. Carlsson, D. Ehrenburg, P. Eklund, M. Fedrizzi, P. Gustafsson, P. Lindholm, G. Merkurjeva, T. Riissanen, A.G.S. Ventre, Consensus in distributed soft environments, European Journal of Operational Research 61 (1992) 165–185.
34. A. Efstathiadsea, S. Tassoub, A. Antoniouc, Strategic planning, transfer and implementation of advanced manufacturing technologies (AMT), Technovation 22 (4). p.p. 210– 212, 2002.
35. F. Herrera, E. Herrera-Viedma, J.L. Verdegray, A model of consensus in group decision-making under linguistic assessments, Fuzzy Sets and Systems 78 p.p. 73– 87, 1996

36. J. Kacprzyk, M. Fedrizzi, *Multiperson Decision-Making Models Using Fuzzy Sets and Possibility Theory*, Kluwer Academic Publishers Inc., Dordrecht, 1993.
37. J. Kacprzyk, M. Fedrizzi, H. Nurmi, *Consensus Under Fuzziness*, Kluwer, Norwell, MA, 1997
38. E.E. Karsak, Distance-based fuzzy MCDM approach for evaluating flexible manufacturing system alternatives, *International Journal of Production Research* 40, 2002, p.p. 3167–3181.
39. Object Management Group. *MDA Guide V1.0.1*. OMG, 2003
40. Marco Brambilla, Stefano Ceri, Federico Michele Facca Politecnico di Milano and Irene Celino, Dario Cerizza and Emanuele Della Valle Ceferiel, *Model-Driven Design and Development of Semantic Web Service Applications*, *ACM Transactions on Internet Technology (TOIT) archive* Volume 8, Issue 1 No.3 , Nov. 2007, ISSN:1533-5399
41. Bell, Michael, *Introduction to Service-Oriented Modeling. Service-Oriented Modeling: Service Analysis, Design, and Architecture*, Wiley & Sons Publishers Inc., 2008, ISBN 978-0-470-14111-3.

Adaptive Knowledge Representation for a Self-Managing Home Energy Usage System

Martijn Warnier¹, Marten van Sinderen² and Frances M.T. Brazier¹

¹Faculty Technology, Policy and Management
Delft University of Technology, The Netherlands

²Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, The Netherlands

M.E.Warnier@tudelft.nl,
m.j.vansinderen@ewi.utwente.nl, F.M.Brazier@tudelft.nl

Abstract. Automated and efficient energy management has many potential benefits for producers and consumers of energy, and the environment. Focusing on energy management on the consumer side, this paper considers two forms of energy management: minimizing energy usage in single households and avoiding peaks in energy consumption in a larger area.

A combination of context aware and autonomic computing is used to describe an automated and self-managing system that, by analyzing context information and adapting to its environment, can learn the behavior of household occupants. Based on this information, together with user defined policies, energy usage is lowered by selectively powering down devices. By powering specific thermostatically controlled devices on or off energy can also be redistributed over time. This is utilized to avoid global peaks in energy usage.

The self-managing system reasons about context and other information and acts when required. This information is the knowledge with which it can adaptively reason, about to take to ensure efficient energy usage. This paper explores the requirements that hold for representing this knowledge and how the knowledge base can continuously and adaptively be updated: to be self-managing.

1 Introduction

Efficient energy management forms an important challenge in today's society as conventional energy sources become more and more scarce and more eco-friendly alternatives are not yet evaluated at a large scale. Automated approaches to efficient energy management are currently still limited and mostly used by large power consumers such as factories [10, 16, 17], but with the advancement of the Smart Grid [2] and other recent advancements in sensor networks [1, 19], such as cognitive networks [18], this changes rapidly.

This paper addresses this challenge at two different, and possibly conflicting, levels: (i) at the level of a single household, where the goal is to lower energy consumption and (ii) at the level of group of households (a neighbourhood), where the goal is to lower peak usage in energy consumption. In [15] a new approach based on autonomic and context aware computing is introduced. This paper extends that approach. It explores

the requirements for an adaptive knowledge representation in the context of a home energy management system. The system exploits sensor information to monitor electric appliances and their surroundings. Based on this environmental information the status of appliances is updated, influencing energy consumption.

The core of the home energy management system is formed by a rule based system [5]. Such systems are typically very deterministic in nature, always producing the same outcomes for the same input: a more dynamic approach is required for self-management. Therefore, the main challenge addressed in this paper is the question how, in the context of a home energy management system, a knowledge base can continuously and adaptively be updated: to become self-managing.

The remainder of this paper is organized as follows: in the next section the approach introduced in [15] is summarized. Section 3 discusses requirements for the self-managing home energy system and Section 4 outlines an approach to adaptive and self-managing knowledge representation. The paper ends with a brief discussion and conclusions.

2 An Autonomic and Context Aware Home Energy Management System

A new green computing approach on the consumer side (demand side management) is proposed in [15]. The proposed system considers two forms of energy management: minimizing energy usage in single household and avoiding peaks in energy usage for a larger residential area. A basic architecture is proposed to achieve this goal. A service oriented framework, based on the complementary approaches of autonomic computing and context aware computing, is introduced. Context information is continuously used to analyze the energy requirements of a household. Based on this information, the home energy management system determines whether and how to influence the energy consumption of individual devices. By selectively powering thermostatically controlled devices –such as fridges, air conditioning units, electrical heating– on or off, energy consumption is redistributed over time [12, 15] avoiding peaks in energy usage. Such devices make up around 25% of the total energy consumption in the USA [9].

Residences of the household can add their own preferred energy usage profile. The resulting service oriented framework reduces energy usage in households in an intelligent and user friendly manner. Actual consumption can be analyzed with respect to consumption constraints set by the system. If either consumption constraints are (close to being) violated or user needs are unnecessarily high, then energy consumption should be decreased and corresponding control actions need to be exercised on the appliances of interest.

Fig. 1 illustrates the proposed architecture of the home energy management system.

The MAPE (monitor, analyze, plan, and execute) control loop from autonomic computing [6] is adapted to be used to analyzing energy consumption of electrical appliances. Considering a pool of electrical appliances that are instrumented to allow monitoring and control. The monitoring consists of measuring the energy consumption of these appliances. The measurements are fed to a control process (the Appliances Management Process or AMP in Fig. 1), which interprets these as the actual consumption,

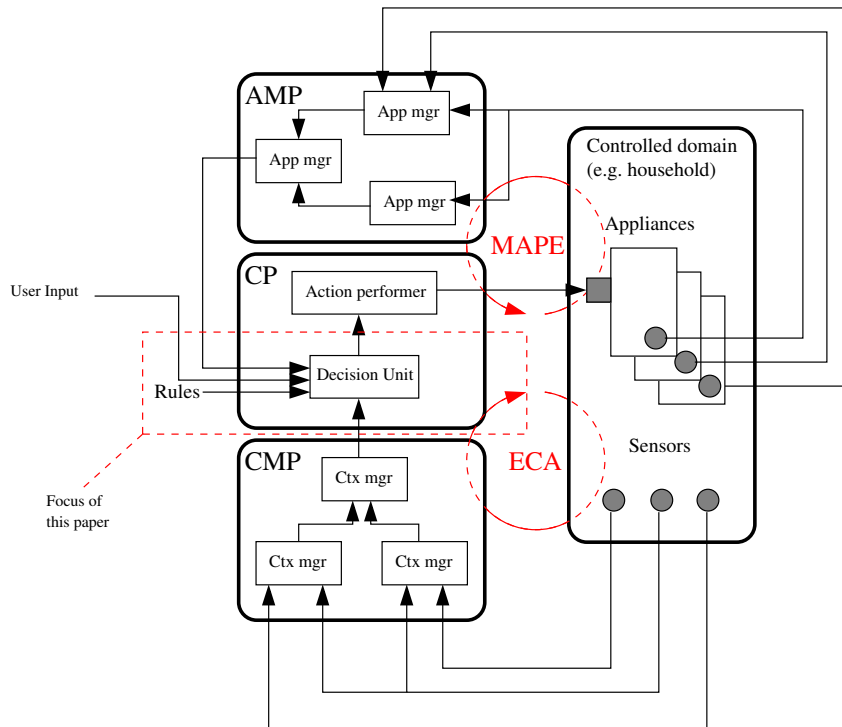


Fig. 1. The basic architecture of the home energy management system. Adapted from [15].

and compares the consumption with the consumption constraints (in the Decision Unit in Fig. 1). If, as a result of this analysis, it is decided that control actions are needed, an action plan is produced. The action plan is derived with an algorithm that considers time-shifting of the active state of appliances. Subsequently, the plan is executed (in the Action Performer in Fig. 1) by performing the indicated control actions on the selected appliances. Figure 1 illustrates the application of the MAPE control loop in the top of the figure. The Decision Unit, which forms the focus of this paper (see Fig. 1) is discussed in detail in Section 4.

With regard to analyzing consumer needs, the event-control-action (ECA) pattern from context-aware computing [4] is used. Figure 1 illustrates the application of the ECA pattern at the bottom of the figure. The environment of the appliances is considered. It is assumed that this environment is instrumented with sensors that are able to measure relevant conditions. For example, measurements may be used to determine context changes or situations, such as the presence of one or more persons in the house or in a particular room, the activity mode (sitting, walking, sleeping) of a person, or a person entering or leaving the house. Context situations and changes can generally not be directly or reliably measured by a single sensor. A context management process (the CMP in Fig. 1) is responsible for producing events that indicate the occurrence of a context situation or change, based on reasoning which potentially involves sensor data

from several sources. Events are fed to a control process, which applies them in rules to determine actions related to perceived needs. For example, if nobody is in the house, a rule may establish the action to set the preferred value of the heating at 15 degrees Celsius. Whether the actions are really required depends on the supported needs. For example, if the preferred value of the heating is already set to 15 degrees Celsius, no action is needed. The comparison of the perceived and supported needs leads to an action plan, which, if not empty, is subsequently executed by performing the indicated control actions on the selected appliances.

A service-orientated architecture (SOA) is used to implement the approaches outlined above, see [15] for more details.

3 Requirements for Knowledge Representation

The high level goal of the energy management system is twofold: On the one hand, on the household level, the system should *minimize the total energy consumption* of the household, within fixed boundaries set by the household owner. On the other hand, on the neighborhood level, the system should minimize fluctuations in energy demand, i.e., keep the energy consumption of the whole neighborhood as constant as possible. Note that these goals can be conflicting. Energy providers can prioritise one over the other by providing (monetary) incentives. The energy management system's high level goals provide the first system's requirement:

1. the system should be flexible enough to optimize either of the two high-level goals of the system: minimize local energy consumption or minimize global fluctuations in energy consumption

Moreover, the system needs to be highly adaptive, in particular:

2. the system should be able to adapt its behavior at runtime and change the high-level goal, depending on input from the environment

Finally, users should be able to customize the systems to their specific needs, setting limits to the adaptive behavior of the system:

3. users, i.e., home owners, should be able to customize the system to their specific needs.

Note that the last requirement is the most important one. This can potentially limit the adaptive behavior of the systems. However, it is crucial that users should be able to override the energy management system, even if this means that, for example, the air condition is set to maximum in each room. No users will allow a fully autonomous system to manage their energy usage. This issue is discussed further in Section 5.

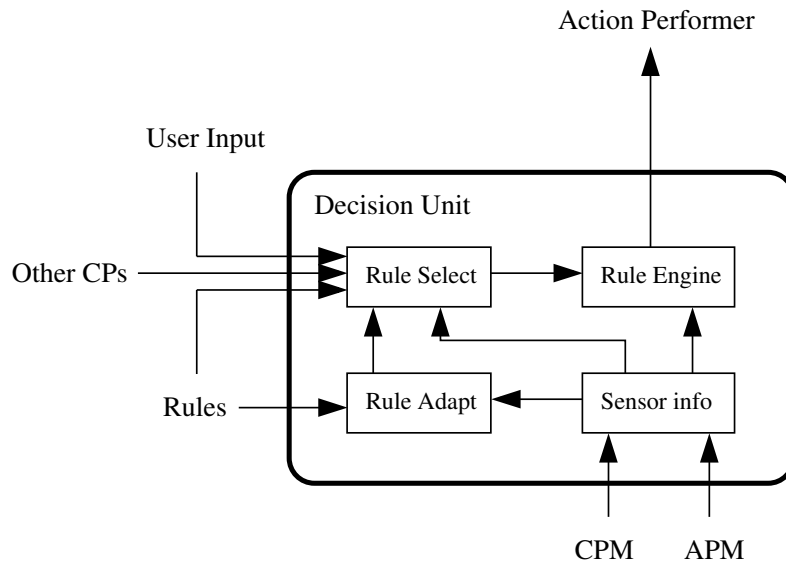


Fig. 2. The decision unit from Fig. 1.

4 Towards Adaptive and Self-Managing Knowledge Representation

The main self-managing component of the energy management system is formed by the *Decision unit* shown in Fig. 1. This unit has to (autonomously) decide how appliances are adjusted to meet the goals of the system. Fig. 2 shows the decision unit in detail.

The decision unit as a whole takes input from four different sources, namely (i) sensor input, (ii) user input, (iii) other control process units (CPs) at other households and (iv) rules. The sensor input comes from the Context Management Process (CMP) and the Appliances Management Process (AMP). These units process the sensor information and provide (aggregated) context and appliances status information which forms the basis for the adaption process. The user input gives the home owner the opportunity to override the (autonomous) decision unit, for example, indicating that the air conditioning unit in a room cannot be put off.

To meet the goal of the system, global (at a neighborhood level) minimization of the fluctuations in energy consumption, the control processes (CPs) in individual households need to communicate with each other. The CPs can be organized in a virtual tree overlay [13] and work together to meet this goal, for example using the approach outlined in [12, 14].

The heart of the Decision Unit is formed by a rule based system which consist of Rule Select, Rule Adapt and Rule Engine components, see Fig. 2. It is assumed, since

the application domain is known and unlikely to change rapidly, that the knowledge base of the home energy system uses a fixed ontology, i.e., all rules are formulated in the same (fixed) language. Rule based systems are traditionally very deterministic systems. They consist of rules of the form shown in Example 1 below:

Example 1 (*Rule base system*)

matching condition 1 \Rightarrow effect 1
 matching condition 2 \Rightarrow effect 2

 matching condition n \Rightarrow effect n

△

By default the ordering of the rules defines the (operational) semantics of the system. Rules are evaluated in order, and the first rule with a matching condition is executed, i.e., later rules that might match are discarded. Execution of a rule leads to an effect. In this case something like altering the status of an appliance, i.e., lowering the temperature of a fridge. Since the environment will change continuously different rules will be executed over time. However, given the same conditions the same rules will be executed, making the system completely deterministic (static). To make the system more adaptive to its environment it should evolve over time, to meet the demands of a specific household. There are several options that can be used to make the rule based system adaptive, these include: (i) **load specific set of rules based on the environment**, for example in the *rule select* unit a specific set of rules can be loaded that has as goal to minimize global fluctuations instead of minimizing the local energy use of a household. (ii) **evolving rules**, the rules can be changed based on genetic algorithms [3] or a neural network [7]. (iii) **weighted rules**, in this approach all rules are weighted. Instead of executing the first matching rule all rules are selected and the rule with the highest weight is executed. Fuzzy logic [8] like approaches can be used for this. (iv) **hybrid approach**, combine some of the options above.

Evolving rules (item 2 above) effectively is typically difficult. Neural networks and genetic algorithms try to merge and combine existing rules to produce new (better) ones. However, a fitness function is required to determine if a newly generated rule is better than existing ones. Finding a suitable fitness function is typically very hard. Therefore this alternative is not further studied in detail here. Instead, a hybrid approach that combines item 1 and 3 above is explored. The proposed adaptive rule based system has the following properties:

- rules are weighted
- rules are bundled in a set, called the *device set*, per device
- rules can be added and removed per set
- sets can be added and removed to the *rule base*, the active rule sets that are used by the rule engine

For each device there is an associated device set, consisting of weighted rules, that determines the (adaptive) behavior of the device. Consider, for example, the following three rules, shown in Example 2 below, that are part of the device set for controlling the ac-unit in the master bedroom:

Example 2 (*Weighted Rules controlling an AC unit*)

```

.....
(0.3) #people in room ≥ 1 & t > tmax & AC = off ⇒ AC = on & cooling=10
(0.7) #people in room ≥ 1 & t > tmax & AC = off ⇒ AC = on & cooling=2
.....
(0.5) #people in room = 0 & AC = on           ⇒ AC = off
.....

```

△

All three rules have a weight (0.3, 0.7 and 0.5 respectively). The first two rules share the same pre-conditioning, that evaluates to the value true if there is more than one person in the room, the ac-unit is turned off and the current temperature (t) is higher than some predetermined temperature (t_{\max} , for example 25 degrees Celsius). However, the result of executing the rules is different: executing the first rule results in turning the AC unit at setting 10, in the second case the AC unit is turned on at setting 2. A higher setting leads to a faster cooling of the room, but also a higher (at least initial) energy consumption. Since both these rules share the same precondition, the one with the highest weight (the second rule with weight 0.7) is executed.

By adapting the weights of rules, different energy consumption patterns emerge. Again looking at the two rules mentioned above, in the current situation energy consumption is low over a longer period. If the weights of the two rules are swapped, the result would be a higher energy consumption, but for a longer period. In effect, the first mechanism that is used to adapt the system is this changing (adapting) of the weights of the rules. This makes it possible to adapt the reaction of the system to a specific situation.

Note that the third rule (with weight 0.5) has a different pre-condition. It evaluates to true if there is no one in the room and the AC unit is on. The effect of executing the rule is that the AC unit is turned off.

The *Rule Select* unit (from Fig. 2) loads selected device sets into the rule engine. By periodically loading different rules into the rule engine, i.e., by changing the *rule base*, the system adapts to its environment. Based on input from the environment different (possibly conflicting) high level goals can be met by different rule-bases. This loading of different device sets provides the second adaptation mechanism of the system.

User input, i.e., from home owners, can be mapped easily to rules that deal with a specific appliance, for example, the air conditioning unit in the bedroom. The rule is seen in the Example 3:

Example 3 (*User generated Rule controlling an AC unit*)

$$(1.0) \text{ true} \Rightarrow \text{AC} = \text{on}$$

△

User generated rules should always evaluate to true (hence the pre-condition in the rule above). Also note that such rules should always be loaded (unless specifically cancelled by the user) and should have high weights (in this example, the maximum value of 1.0) to ensure that user generated rules are executed. Finally, note that the weight of such rules should, in principal, not be adapted by the system.

The system provides two adaptation mechanisms: weight adaptation which is handled by the adapt rule component and selective device set loading (and unloading), which is handled by the rule select unit. Separating these mechanisms has the advantage that its easier to reason about adaptation policies at a higher (strategic) level. This is left for future work, as are specific rule adaptation policies.

In summary, the Decision Unit takes input from the user and other CPs. Based on input from the environment (sensor info), weighted rules are adapted and selected. The rule engine selects all matching rules and chooses the ones that have the highest weight per appliance. These are then send to the action performer which adapts the status of the appliances. This whole process is repeated periodically.

5 Discussion and Conclusions

This paper discusses an approach and architecture for a home energy system based on an adaptive and self-managing knowledge representation. The system is based on a weighted rule based system that adapts continuously to its environment. One of the main challenges of this system is to meet its different, possibly conflicting, goals. And while the current architecture should make this possible it remains to be seen if these goals are not too conflicting to be unifiable in practice. It might be necessary to drop the goal of lowering the global fluctuations in energy consumption to meet the user's preferences and minimize the local energy consumption of the household. Simulations and/or experiments should provide more insight on this issue. This is left for future work.

Another issue is how to scale this up to collections of households. A hierarchical structure could be used in which the architecture can be repeated at different levels of granularity. For example, a household has appliances as units that are being controlled; an apartment has living units as the units that are being controlled; a city block has apartments as units being controlled; etc. Finding the correct clustering of households [11] that are controlled by one processing unit forms another challenge.

From a technical perspective it is not very difficult, with the proposed architecture, to force control processes in different households to cooperate to reduce peaks in (global) energy usage. However, this might lead to some considerable discomfort with home owners, for example if they can use their air conditioning unit at the maximum

setting, because a global reduction in energy consumption is required. Monetary incentives, provided by energy producers who benefit from reduced peak usage, might help lessen the discomfort of the home owner, as would specific policies set by local governments. However, if this will be an acceptable solution remains to be seen. This issue is further outside the scope of this paper.

A related issue is if an (semi) autonomous home energy management system will be accepted by users. However, since there is both a monetary incentive (energy usage is lowered which in turn leads to a lower energy bill) and since users can override the behavior of the system this is probably less of an issue than the one discussed above.

Acknowledgements

The authors like to thank Boris Shishkov who was jointly (together with the first two authors) responsible for the main idea behind the proposed energy management system. The research presented in this paper has partially been funded by the Next Generation Infrastructures project ‘Self-Managed Dynamic Institutions in Power Grids: Sharing the Cost of Reliability’.

References

1. I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002.
2. P. Clark W. Gellings. *The Smart Grid, enabling energy efficiency and demand response*. The Fairmont press, 2009.
3. A. Corcoran and S. Sen. Using real-valued genetic algorithms to evolve rule sets for classification. *ICEC*, 94(2405):120–124.
4. P. Dockhorn Costa, L. Ferreira Pires, and M. van Sinderen. Concepts and architectures for mobile context-aware applications. In *Handbook of research on mobile multimedia*, Information Science Reference. Hershey, New York, 2008.
5. F. Hayes-Roth. Rule-based systems. *Communications of the ACM*, 28(9):932, 1985.
6. IBM Corporation. An architectural blueprint for Autonomic Computing. White paper, 2005.
7. E. Keedwell, A. Narayanan, and D. Savic. Evolving rules from neural networks trained on continuous data. In *The 2000 Congress on Evolutionary Computation CEC 00*, pages 639–645, 2000.
8. G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic: theory and applications*. Prentice Hall Upper Saddle River, NJ, 1995.
9. P. Mazza. The Smart Energy Network: Electrical Power for the 21st Century. *Climate Solutions*, 2002.
10. A. Middelberg, J. Zhang, and X. Xia. An optimal control model for load shifting - With application in the energy management of a colliery. *Applied Energy*, 86(7-8):1266 – 1273, 2009.
11. E. Ogston and F. M. T. Brazier. Apportionment of control in virtual power stations. In *In the proceedings of the international conference on infrastructure systems and services 2009: Developing 21st Century Infrastructure Networks*, 2009.
12. E. Pournaras, M. Warnier, and F. M. Brazier. Local agent-based self-stabilisation in global resource utilisation. *International Journal of Autonomic Computing (IJAC)*, 2010.

13. E. Pournaras, M. Warnier, and F. M. T. Brazier. Adaptive Agent-based Self-organization for Robust Hierarchical Topologies. In *ICAIS '09: Proceedings of the International Conference on Adaptive and Intelligent Systems*. IEEE, September 2009. (to appear).
14. E. Pournaras, M. Warnier, and F. M. T. Brazier. A distributed agent-based approach to stabilization of global resource utilization. In *Complex, Intelligent and Software Intensive Systems, International Conference*, pages 185–192, Los Alamitos, CA, USA, 2009. IEEE Computer Society.
15. B. Shishkov, M. Warnier, and M. van Sinderen. On the Application of Autonomic and Context-Aware Computing to Support Home Energy Management. In *The proceedings of the 12th International Conference on Enterprise Information Systems (ICEIS 2010)*, 2010.
16. M. Stadler, W. Krause, M. Sonnenschein, and U. Vogel. Modelling and evaluation of control schemes for enhancing load shift of electricity demand for cooling devices. *Environmental Modelling & Software*, 24(2):285 – 295, 2009.
17. G. Strbac. Demand side management: Benefits and challenges. *Energy Policy*, 36(12):4419 – 4426, 2008. Foresight Sustainable Energy Management and the Built Environment Project.
18. R. Thomas, D. Friend, L. DaSilva, and A. MacKenzie. Cognitive networks: adaptation and learning to achieve end-to-end performance objectives. *IEEE Communications Magazine*, 44(12):51–57, 2006.
19. F. Zhao and L. Guibas. Wireless sensor networks. *Communications Engineering Desk Reference*, page 247, 2009.

A Multimedia Database Server for Content-Based Retrieval and Image Mining

Cosmin Stoica Spahiu, Liana Stanescu and Dumitru Dan Burdescu

University of Craiova, Faculty of Automation, Computers and Electronics, Craiova, Romania

Abstract. This article presents a possible extension of a software tool implemented in C++ that manages multimedia data collections from medical domain. An element of originality for this database management system is that it includes a series of algorithms used for extracting visual information from images (texture and color characteristics) along with classical operations needed for databases servers. It is also presented a data mining algorithm adapted to the database system that will be included in a future version.

1 Introduction

The images are an important class of multimedia data. The WWW is one of the biggest multimedia repositories, including text data, images, video and audio data. Most of the data type exchanged in real world is images. Although there have been made efforts for developing search engines, content-based retrieval is rarely implemented.

The raw data is not always useful. The real advantage is when data mining techniques can be applied and obtained knowledge. That is why first step is to adapt the techniques used for images processing.

In order to make the mining technology to be successful, it should be developed for other types of data, especially for images. The image mining should consider automatic classification, knowledge extraction, connections between images and other new patterns. The extension of the data mining in the imagistic field is a natural extension. It is an interdisciplinary domain that includes artificial vision, patterns recognition, data mining, automatic learning, databases and artificial intelligence.

More than that, image mining must consider spatial information. The same pattern might have several interpretations. That is why the mining algorithms for images are different than the classical ones. An image pattern must be represented in a suggestive form to the users using some characteristics of images. The information can be presented at different levels of details: pixel, object, semantic concept, or pattern.

Most of the data mining activities have been made based on the similarity analysis between a query image and the images from database. There are two categories of image retrieval techniques: systems that use a text descriptor of the image and systems that use visual content.

In the first category the images are described based of a text defined by the user. They are indexed and retrieved based on basic descriptors such as: image size tags, image type, acquisition date, owner id, keywords, etc. The types of queries that can be

executed are: find images from database that match the following criteria: date of capture (before 2010), size > 150 KB, and tag “clouds”.

The text descriptors are usually added by a human operator since the automatic generation is hard to be done without incorporating visual information. It is a process that is hard to be applied nowadays since the volume of information is high. More than that, these descriptors are subjective and depend on the users’ perspectives.

When using the second category, the queries that can be executed follow the next pattern: “find all the images that are most similar with the query image”.

The paper is structured as follows: Section 2 presents content-based retrieval systems. Section 3 presents an overview of the server, Section 4 presents the image data mining functionality and Section 5 presents the conclusions.

2 Content-based Retrieval

The Most important aspect in content based retrieval is to find a method to measure the similitude of the images. The properties needed for methods used to compute the similitude, are:

- Easy computing
- Correspondence with human reasoning

There are two ways to implement content based retrieval: content-based retrieval for k-nearest vicinity (retrieves the most k similar images) and domain query (retrieves all the images that have the similarity between specified ranges). In order to make fast comparisons of the images, the system has to process off-line the images and to extract and store the characteristics of the image. For example, the color histograms describe the colors distribution and they are extracted before executing any content based query.

The dissimilitude of two images must be a metric that generate small values for similar images and large value for images that have only few in common. The performance of such a system is limited by the quality of the images characteristics [4][8].

The architecture of a content based retrieval system is presented in Figure 1.

a) The content-based region query: this type of query compares the image based on their color regions. In the first step of the query, the images’ regions are being queried instead of whole images. The total similitude of two images is computed based of the distances computed for each region in part.

The content-based retrieval can be improved in quality by adding spatial information to the query. In this case it is considered both the similitude distances of the characteristics (texture, color) and the spatial values of the regions.

b) Spatial query of the images. In the last years there have been developed techniques for spatial indexing that permits to retrieve objects based on the objects positioning. These researches compare images where there have been already defined regions or objects, as in Figure 2 [7].

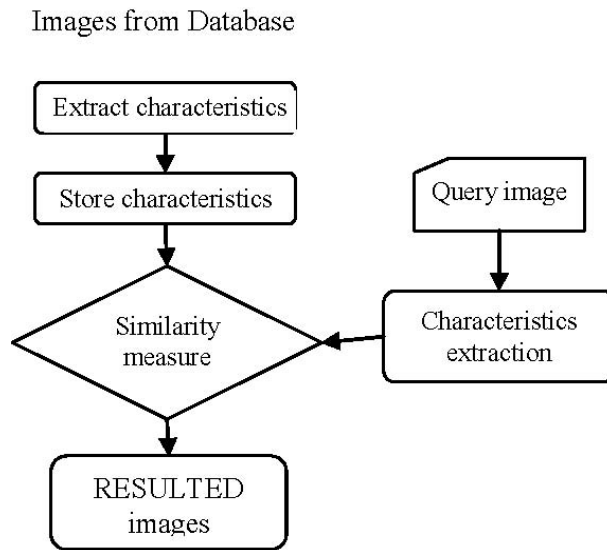


Fig. 1. Architecture of the Content-based retrieval system.

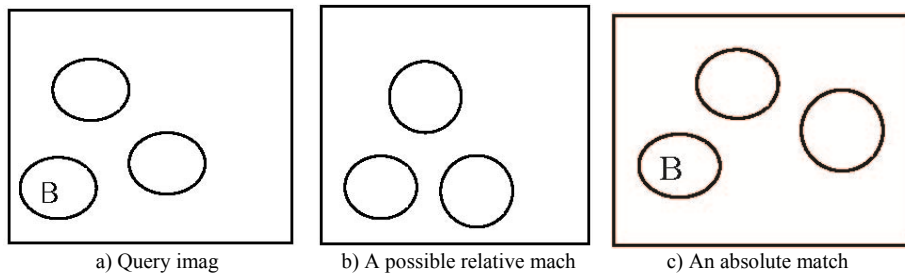


Fig. 2. Spatial query of the images.

In order to extract the color regions of the images there have been implemented the Apriori algorithm proposed by Smith and Chang. For each color region that have been detected it should be stored the following information:

- Image id
- Region id
- Color set of the image
- Coordinates of the regions

All these information will be used in the content-based image query process.

The spatial information is represented by the minimum rectangle that bounds the region.

The characteristics that have been used in the proposed system are: color histogram and texture (extracted using the Gabor method) [9][10].

The similitude between images is computed using histogram intersection and quadratic distance.

3 MMDBMS Overview

The MMDBMS that have been implemented allows database creation, table and constraints adding (primary key, foreign keys), inserting images and alphanumeric information, simple text based query and content-based query using color and texture characteristics. The software tool is easy to be used because it respects the SQL standard. It does not need advanced informatics knowledge and has the advantage of low cost. It is a good alternative for a classical database management system (MS Access, MS SQL Server, Oracle10g Server and Intermedia), which would need higher costs for database server and for designing applications for content-based retrieval.

Figure 3 presents the general architecture of the MMDBMS [1][2][3].

In the first step any application that uses the server must connect to the database. This way it will be created a communication channel between them. All commands and responses will use this channel to send queries requests and receive answers.

The server has two main modules: kernel engine and database files manager. The kernel engine includes all functions implemented in the server. It is composed from several sub-modules each of them with specific tasks[1][2]:

The Main Module. It is the module, which manages all communications with the client. It is the one that receives all queries requests, check what is the type of query requested, extracts the parameters of the query and calls the specific module to execute it.

Queries Response Module. After the query is executed, the results will be sent to the Queries Response Module. It will compact the result using a standard format and then return it to the client. The client will receive it on the same communication channel used to send the request.

Select Processing Module. If the main module concludes that is a SELECT SQL command, it will call the Select Processing module. This module extracts the parameters from the query and then search in the database files for specific information. If the query is a SELECT IMAGE query, it will use for comparison the similitude of characteristics instead equality of parameters.

Characteristics Extraction Module. When the main module receives a SELECT IMAGE or a UPDATE query which uses an image that is not already in the database it is needed first to process it.

This module is called to extract the color and texture characteristics of the image. The data of the results will be used to initialize a variable of IMAGE data type.

Update Processing Module. When the query received from the user is an UPDATE command, it will be called to execute it.

Delete Processing Module. It is called when the user executes a DELETE command. The kernel executes only logic deletes. It never executes physical deletes. The physical deletes are executed only when a "Compact Database" command is sent by the user.

The second main module is the Database Files Manager. It is the only module that has access for reads and writes to the files in the database. It is his job to search for information in the files, to read and write into files and to manage locks over databases. When a client module request a read form a file it is enabled a read lock for

the specific file (that represents a table in the database). All other read requests will be permitted but no writes will be allowed. If the client module request a write to file, it will be enabled a write lock. No other requests will be allowed until the lock is canceled.

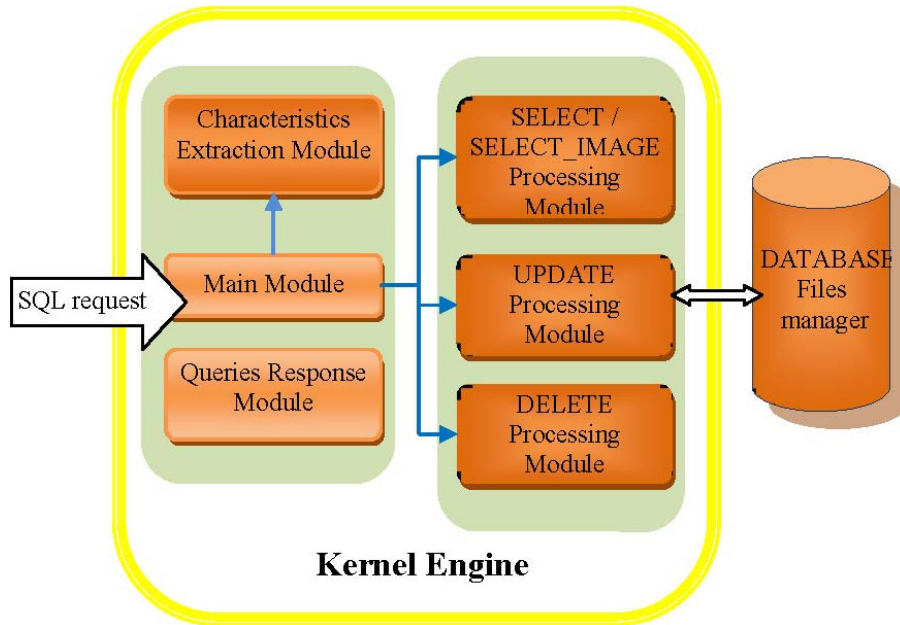


Fig. 3 General architecture of the system.

4 Image Data Mining

The information in raw form is not always useful. The real benefit comes when “interesting” patterns can be obtained based on the association rules. An association rule tells us about connections between two or several objects. It is a rule of type $A \Rightarrow B$, where A and B are objects satisfying the condition $A \cap B = \Phi$.

In order to find specific combinations for objects that appear together in different associations we have studied the Apriori algorithm. It will be implemented in a future version of the system.

This algorithm will select the most “interesting” rule, based on two parameters called support and confidence. The rule $A \Rightarrow B$ shows that anytime A appears in a transaction, it is very possible to appear B also.

The support parameter shows the statistic meaning of a rule:

$$supp(A \Rightarrow B) = \frac{|\{T \in D | A \subseteq T \wedge B \subseteq T\}|}{|D|}$$

The confidence parameter shows the strength of a rule:

$$conf(A \Rightarrow B) = \frac{|\{T \in D | A \subseteq T \wedge B \subseteq T\}|}{|\{T \in D | A \subseteq T\}|}$$

The rule probability confidence is defined as conditional probability:

$$p(B \subseteq T \mid A \subseteq T).$$

The association rule can be also between two or several objects ($A, B \Rightarrow C$) where $A, B, C \subseteq U$.

The association rule is stronger as the confidence parameter is higher. This last parameter specifies the minimum support for frequent objects. All the subsets of frequent objects are also frequent. An object can be frequent only if it is found to be frequent in one of the algorithm's steps.

The Apriori algorithm is presented next.

Algorithm 1 Apriori Algorithm

```

1. Find frequent item sets;
 $F_1 = \{u_i \mid \|u_i\| > \text{minimum support}\}$ 
for ( $K = 2; F_{K-1} \neq \emptyset; K++$ ) do
   $C_K = \{c_k \mid c^{(a)} \wedge c^{(b)} \in F_{K-1}\}$ , where:
   $c_k = (u_{i_1}, \dots, u_{i_{k-2}}, u_{i_{k-1}}, u_{i_k})$ 
   $c^{(a)} = (u_{i_1}, \dots, u_{i_{k-2}}, u_{i_{k-1}})$ 
   $c^{(b)} = (u_{i_1}, \dots, u_{i_{k-2}}, u_{i_k})$ 
   $\|c_k\| = 0$ ;
  for ( $\forall T, T \subseteq D$ ) and ( $\forall c_k, c_k \in C_K$ ) do
    if ( $c_k \in T$ ) then
       $\|c_k\| = \|c_k\| + 1$ ;
    end if
  end for
   $F_K = \{c_k \mid \|c_k\| > \text{minimum support}\}$ 
end for
 $F = \bigcup_K F_K$ 
2. Use the frequent itemsets to generate
strong association rules.

```

This algorithm will be implemented in order to find different patterns used for automate classification of the images, based on the diagnosis. The system will be able to find which part of the extracted features is characteristic for each disease. It will also be able to say for example which characteristics are connected (used in early diagnosis for some diseases).

Because the data volume is higher and higher in the last years it is important to find efficient algorithms for data mining. The presented algorithm scans the data for few times depending to the biggest most frequent object. New enhancements can be added by reducing the number of database parsing and the number of candidates that were generated.

The version of Apriori algorithm that is based on partitions needs only two parsing of the database. The database is divided in disjoint partitions, each of them small enough to fit the memory.

During the first scan, the algorithm reads each partition and finds the most frequent local objects. During the second parsing the algorithm computes the support for each frequent local object, from the entire database. If one object is frequent in the database it must be frequent at least in one of the partitions. That is why to the second partition there are found the supersets with all the potential frequent sets of objects.

5 Conclusions

The paper presents a possible extension of a software tool implemented in C++ that manages multimedia data collections from medical domain. An element of originality for this database management system is that along with classical operations for databases, it includes a series of algorithms used for extracting visual information from images (texture and color characteristics). It is also presented a data mining algorithm adapted to the database system that will be included in a future version.

It is created for managing and querying medium sized personal digital collections that contain both alphanumerical information and digital images (for examples the ones used in private medical consulting rooms). The software tool allows creating and deleting databases, creating and deleting tables in databases, updating data in tables and querying. The user can use several types of data as integer, char, double and image. There are also implemented the two constraints used in relational model: primary key and referential integrity.

The advantages of using this intelligent content-based query visual interface are that the specialist can see images from the medical database that are similar with the query image taking into consideration the color and texture characteristics. In this way the specialist can establish a correct medical diagnosis based on imagistic investigation frequently used nowadays.

The system will include a data mining module that will be used for automate classification of images and finding “interesting” patterns between characteristics of the images.

This software can be extended in the following directions:

Adding new types of traditional and multimedia data types (for example video type or DICOM type - because the main area where this multimedia DBMS is used it is the medical domain and the DICOM type of data is for storing alphanumerical information and images existing in a standard DICOM file provided by a medical device)

Studying and implementing new algorithms for data mining that performs faster on large image collections.

References

1. Stoica Spahiu C.: A Multimedia Database Server for information storage and querying. In: Proceedings of 2nd International Symposium on Multimedia – Applications and Processing (MMAP'09), Vol. 4, (2009), pp. 517 – 522.
2. Stoica Spahiu C., Stanescu L., Burdescu D. D., Brezovan M.: File Storage for a Multimedia Database Server for Image Retrieval. In: Proceedings of The Fourth International Multi-Conference on Computing in the Global Information Technology, (2009) pp.35-40
3. Stoica Spahiu C., Mihaescu C., Stanescu L., Burdescu D. D., Brezovan M.: Database Kernel for Image Retrieval. In: Proceedings of The First International Conference on Advances in Multimedia, (2009), pp. 169-173
4. Kratochvil M.: The Move to Store Images In the Database (2005). http://www.oracle.com/technology/products/intermedia/pdf/why_images_in_database.pdf

5. Del Bimbo A.: Visual Information Retrieval, Morgan Kaufmann Publishers. San Francisco USA (2001)
6. Smith J. R.: Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis. PhD. thesis, Graduate School of Arts and Sciences. Columbia University (1997)
7. Gevers T.: Image Search Engines: An Overview. Emerging Topics in Computer Vision. Prentice Hall (2004)
8. Stanescu L., Burdescu D. D., Brezovan M., Stoica Spahiu C., Ion A.: A New Software Tool For Managing and Querying the Personal Medical Digital Imagery. In: Proceedings of the International Conference on Health Informatics, Porto – Portugal: (2009) 199-204
9. Samuel J. Query by example (QBE). <http://pages.cs.wisc.edu/~dbbook/openAccess/thirdEdition/qbe.pdf>
10. Query by Example, http://en.wikipedia.org/wiki/Query_by_Example
11. Muller H., Michoux N., Bandon D., Geissbuhler A.: A Review of Content based Image Retrieval Systems in Medical Application – Clinical Benefits and Future Directions. Int J Med Inform 73 (2004)
12. Khoshafian S., Baker A. B.: Multimedia and Imaging Databases. Morgan Kaufmann Publishers, Inc. San Francisco California (1996)

Contextualising Information Quality: A Method-Based Approach

Markus Helfert¹ and Owen Foley²

¹School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland

²School of Business, Galway Mayo Institute of Technology, Dublin Road, Galway, Ireland
markus.helfert@computing.dcu.ie, owen.foley@gmit.ie

Abstract. Information Quality is an ever increasing problem. Despite the advancements in technology and information quality investment the problem continues to grow. The context of the deployment of an information system in a complex environment and its associated information quality problems has yet to be fully examined by researchers. Our research endeavours to address this shortfall by specifying a method for context related information quality. A method engineering approach is employed to specify such a method for context related information quality dimension selection. Furthermore, the research presented in this paper examined different information systems' context factors and their effect upon information quality dimensions both objectively and subjectively. Our contribution is a novel information quality method that is context related; that is it takes the user, task and environment into account. Results of an experiment indicate as well as feedback from practitioners confirm the application of our method and indeed that context affects the perception of information quality.

Keywords. Information quality, Method, Context, Method engineering.

1 Introduction

With the increasing importance of Data and Information Quality (DIQ), much research in recent years has been focused on developing DIQ frameworks and dimensions as well as assessment approaches. Researchers have developed a plethora of frameworks, criteria lists and approaches for assessing and measuring DIQ. Despite the vast amount of DIQ research, discussions with experts and practitioners as well as recent studies indicate that assessing and managing DIQ in organizations is still challenging and current frameworks offer only limited benefit.

Several researchers have addressed the question on how to define DIQ and many have confirmed that DIQ is a multi-dimensional concept [1, 2]. Following prominent definitions of "quality" as "fitness for use", most researchers acknowledge the subjective nature of data and information quality. Aiming to assess DIQ, many objective and subjective assessment techniques have been proposed. Mostly, these are developed for one specific context or domain, with limited general applicability. Furthermore, despite the inherent subjective character of quality, foremost frameworks and assessment methodologies are limited to consider the subjective character in which the assessment is performed.

Objective DIQ assessment uses software to automatically measure the data in database by a set of quality rules whereas subjective DIQ assessment uses a survey or interview approach to measure the contextual information by data consumers. A single assessment result can be obtained from objective DIQ assessment. However, we may obtain different assessment results from different information consumers. With the development of both objective and subjective DIQ assessment, researchers suggested to combine these two assessment methodologies. For example, Pipino et al. [3] provide a framework to combine objective and subjective DIQ assessments. Kahn et al [4] propose the PSP/IQ model, in which they assign two views of quality: conforming to specifications (objective) and meeting or exceeding consumer expectations (subjective).

Recognising the problem of limited context in DIQ frameworks, the aim of this paper is to present an approach, which assist in adapting DIQ frameworks to various contexts. In our work we follow the recent observation by researchers, to adapt research results to specific application by providing an approach of contextualising models [5]. In contrast to contribute yet another DIQ framework, in this article, we describe a method to contextualise DIQ frameworks. In this sense, “context” relates the content of the DIQ framework to the IS environments [6]. Context itself is described by various contextual factors characterizing the IS environment. Following design science research, the method is being developed and refined using a method engineering (ME) approach. The proposed method was developed based on experiences from the airline industry. A number of experiments are conducted to test the proposed method. In addition, IS professionals are interviewed to further verify the method and to study the impact of IS context on DIQ.

The paper is organised as follows. The research is introduced in general terms followed by an outline of the problem statement and objective. A review of related IQ work as a field within IS follows. Subsequently, details of the contextualisation method together with an initial application are presented. Contribution, limitations and future work conclude the paper.

2 Problem Statement and Research Objective

As a direct result of user dissatisfaction with the quality of the information produced by IS [7], practitioners and researchers have been concerned for several years about the quality of information and data. The problem becomes increasingly important with the rapid growth in the amount of data that enterprises store and access [8-10]. The information –often of poor IQ- is being used ever increasingly for critical decision making at all levels within the organisation, resulting in significant IQ related problems.

Some examples of these problems are summarized for instance in Al Hakim [11] who provides examples from many areas, outlining the reason along with the particular IQ dimension affected. The examples indicate how the generation of information from disparate sources can impact on many aspects of an enterprise, often not even considered when the IS was initially designed. The impact of these IQ problems has prompted researchers to examine such aspects as IQ dimensions, IQ assessment and IQ management. The impact of the various dimensions of IQ requires

measurement and examination. Furthermore, there has been a huge financial impact associated with the lack of IQ. For instance, it is estimated that poor quality information costs American business some \$611 billion a year [12]. Addressing this problem, several enterprises have invested considerably in efforts to “clean up” their information, to improve IQ and to define rules and routines to assess and manage IQ.

Research has addressed this challenge from various directions. Management process and guidelines have been proposed in order to manage IQ. The database community has contributed several approaches for data cleansing and assessing data quality in databases. In addition, software engineering has focused on improving the quality of software. As a result, research related to IQ has evolved significantly over the last two decades. Numerous frameworks, dimensions and metrics have been put forward [1, 2]. Approaches have been developed to measure the impact of these dimensions on IQ and to improve IQ. However, the benefits of these approaches such as better quality software, easier to use systems, readily acceptable software and increased IQ all have the potential to be foregone by changing contexts or situations.

The evolving nature of IS context presents new and very distinct challenges to IQ research. Primarily among these is the dynamic nature of IS context. IS designers no longer have the luxury of complete control over the nature of the IS context post deployment of the IS. Usually software systems are developed with tried and tested methodologies for a particular context with certain requirements. IQ measurements and management approaches can be defined and deployed for this particular context. Once implemented, contexts evolve and requirements may change. However with such fundamental changes in context, a question about its impact on IS become significant [13, 14].

IQ measurements and management approaches are usually not evolving along the various contexts, and indeed changes in context do usually not undergo a systematic consideration. Systems and IQ measurement approaches are developed for a particular context. Nonetheless the importance of a high level of IQ remains a requirement. As a result often the perception of IQ via subjective survey instruments is progressively getting poorer [15]. Indeed, this observation may explain a frequent criticism, that despite large investments in IQ and software systems, end users are still not satisfied with IQ and the usability of systems. Variation in the context, require adaptation of the IQ measurement approach to cater at least for different requirements and changes in perception. Presently the approach to a changing context and IQ is ad-hoc and not systematic.

The problem can be illustrated by an example from the airline sector, which has been examined by us. The organisation has an airline maintenance and information system in use for many years. Different types of users using the IS on a regular basis for the performance of their duties. Also, for several years the organisation has IQ routines and assessment approaches in place. Pilots, engineers, administrators and technicians are required to rate the IQ of the system on a regular basis for quality control and legal obligations. Surprisingly, over the last two years, the MIS (Management Information Systems) department has experienced a dramatic increase for requests to verify IQ of the IS. This has become a very resource intensive exercise with many of the requests requiring no alteration to data values. On closer examination, we observed that in tandem with the increase in IQ requests the IS context has changed. For instance, the access modes to the system have changed over the years. The single point of access via data entry personnel has evolved over time to

the point where many users interface with the IS from mobile devices over wireless networks. As a consequence, the procedures and IQ assessment approaches in place did not reflect the current situation with various access modes and changing contexts. The organisation did not have any systematic approach to cater for these changes.

The brief example from the airline industry illustrates the requirement for a systematic approach to adapt IQ frameworks for various contexts. Our approach presented in this article, assists to contextualise IQ frameworks and thus cater for various and changing contexts. The necessity for this research arises from the ever increasing dynamics that exists with respect to IS deployment. As our observation from the case study shows, perceptions of IQ may alter as the result of evolving contexts. Research related to IQ has not or only in a limited manner addressed and recognised this problem. In contrast to define yet another IQ framework, we believe that the application of existing IQ frameworks requires a clear, concise and systematic approach to cater for dynamic and evolving contexts. The traditional static deployment of IQ frameworks do not consider adequately the changing factors of IS context. The challenge of this research is to specify a systematic approach in the form of a method that considers the IS context, allowing for the evaluation of IQ in various contexts.

3 Related Work

Our work builds on and contributes to the research related to IQ, which has developed a large number of frameworks, assessment approaches and criteria list. An overview of research related to IQ is provided for instance in Ge and Helfert [16]. Ge and Helfert have examined the definition of IQ and suggest that it can be defined from a consumer perspective and a data perspective. The concept of fitness for use [2] is widely regarded in the literature as a definition for IQ from the consumers view point.

Table 1. Selected Information quality frameworks and its application context.

Author and year of Publication	Application context	Author and year of Publication	Application context
Morris et al. 1996 [17]	Management	Huang et al. 1999 [18]	Knowledge Management
Redmann 1996 [7]	Data Bases	Chengalur-Smith et al. 1999 [19]	Decision-Making
Miller 1996 [20]	Information Systems	Berndt et al. 2001 [21]	Healthcare
Wang/Strong 1996 [2]	Data Bases	Xu et al. 2002 [22]	Enterprise Resource Planning
Davenport 1997 [23]	Information Management	Helfert/Heinrich 2003 [24]	Customer Relationship Management
Ballou et al.1998 [25]	Data Warehousing	Amicis/ Batini 2004 [26]	Finance
Kahn/Strong 1998 [27]	Information Systems	Xu/Koronios 2004 [28]	E-business
Rittberger 1999 [29]	Information Service Providers	Knight/Burn 2005 [15]	World Wide Web
English 1999 [30]	Data Warehousing	Li/Lin 2006 [31]	Supply Chain Management
Ballou/Tayi 1999 [32]	Data Warehousing		

The definition of IQ from a data perspective examines if the information meets the specifications or requirements as laid down in IS design. IQ from a consumer perspective led to the development of subjective IQ measures, whereas IQ from a database perspective resulted in objective IQ measures.

Many IQ frameworks have been developed in order to classify dimensions that will allow for IQ assessment. As shown in table 1, we reviewed prominent frameworks and analyzed these according to the application context for which they were proposed. Although claims are sometimes made to provide a generic criteria lists, on closer examination most research has been focused on investigating IQ within a specific context. The frameworks differ in selected IQ criteria as well as assessment techniques.

The *complexity of the information system architecture* is just one of several contextual factors to characterise IS environments. In literature there is strong support that types of users and types of IS result in different requirements and therefore perceptions of DIQ [33-35]. Empirical research concluded that user evaluation of IS is directly influenced by system, task and individual characteristics. Besides, several examples in DIQ literature illustrate that the departmental (organisational) role plays an important factor in user's opinion and perspectives [11, 33, 36]. Although recognising that there are several contextual factors, in our research we limit the set of contextual factors to 4:

- (1) User role
- (2) Organisational department,
- (3) IS Architecture
- (4) Task complexity.

The initial deployment of an IS generally caters adequately for these context factors. However over time these evolve and require a fresh analysis in order to accurately represent the true nature of the context of the IS. This can only be done if examinations of the factors within a context are continually updated and revised. These factors can be further classified with appropriate components. The updating of the factors and reclassification of components should be carried out in an iterative and systematic fashion similar to TDQM [37].

Table 2. Common IS Context Factors.

Factor	Component
User Role	Manager / Specialist
Organisational Department	IT Department / Non IT Department
IS Architecture	Workstation, Service Oriented, Mobile
Task Complexity	Operational, Strategic

The significant change and increase in complexity of IS context has in many cases occurred independently of the underlying databases that are accessed [13]. An application may have been designed, built and tested with a mature software development method for a particular context. Yet within a very short period of time it may be accessed and predominantly employed from a different context [11]. Data models have also evolved [9]. However a considerable number of IS in use today have data models designed prior to the contexts that are employed to access them. There are as a result multiple accesses from diverse and complex contexts.

4 A Method for Contextualising IQ Frameworks

In order to develop an approach to contextualise IQ frameworks, we follow design science and apply a method engineering (ME) approach. ME as a discipline has been recognised over the last decade. It is concerned with the *process* of designing, constructing and adapting generic artefacts such as models, methods, techniques and tools aimed at the development of IS [38]. Punter [39] describes the discipline from a process perspective where methods are comprised of phases, phases are comprised of design steps, and design steps are comprised of design sub-steps. He states that to every design step or sub-step, certain product-oriented method constituents (e.g. techniques, procedures) can be assigned.

In order to describe methods, Gutzwiller [40] proposes a meta-model for methods that includes activities, roles, specifications, documents and techniques. Figure 1 below illustrates the relationship between these elements. The meta-model facilitates a consistent and concise method, which in turn allows for their application in a goal oriented, systematic and repeatable fashion. According to Gutzwiller [40] activities are the construction of tasks which create certain results. These activities are assigned to roles and the results are recorded in previously defined and structured specification documents. The techniques comprise of the detailed instructions for the production of the specification documents. Tools can be associated with this process. The meta-model describes the information model of the results.

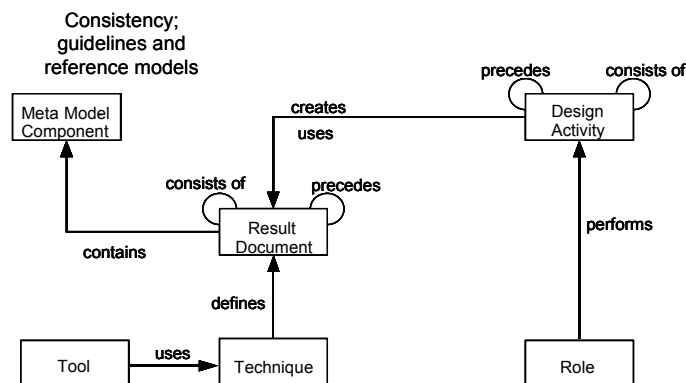


Fig. 1. Method Engineering Approach.

Applying a method based perspective on contextualising IQ frameworks, we identified four main activities that describe the contextualisation process: (1) Identify and prioritising contextual factors, (2) Selecting and prioritise IQ dimensions with associated IQ measures, (3) Implement IQ measures, (4) review and improve.

Besides the contextualisation process (activities), the second main element is the description of result documents. We provide a consistent result document for contextualised IQ frameworks. The meta-model is illustrated in Figure 2 that outlines the relationship between the different components of context factors and IQ measurements.

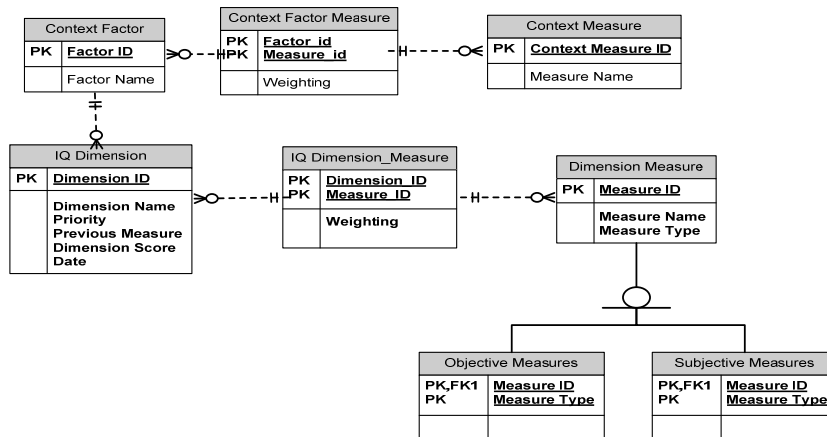


Fig. 2. Meta Model for Contextualising IQ Frameworks.

The initial application of the ME approach to our problem was examined with respect to a library IS. A diverse user population accesses this IS from three different contexts. Figure 3 specifies the general contextualisation processes that allow us to conduct an experiment to validate our approach. We consider 4 main activities (1) Identify contextual factors, (2) Quantify and prioritise IQ requirements, (3) Implement selected IQ Measures, (4) Improve. Activity (1) is usually carried out by the IS Manager, involving interviews and surveys with domain experts. This activity completes and measures context factors. Activity (2) identifies and prioritises IQ metrics and requirements. Carried out by Business Analysts, for this activity and to priorities IQ dimensions we selected a specific technique: Leung’s metric ranking. Subsequently the Information Technology Manager and Software developers implement the IQ measures, in the form of Service Analysers, Integrity Checker and IQ surveys. Finally, IS manager, Information Technology managers together with the Business Analyst review the context factors as well as IQ measures, and thus initiate a continuous improvement process.

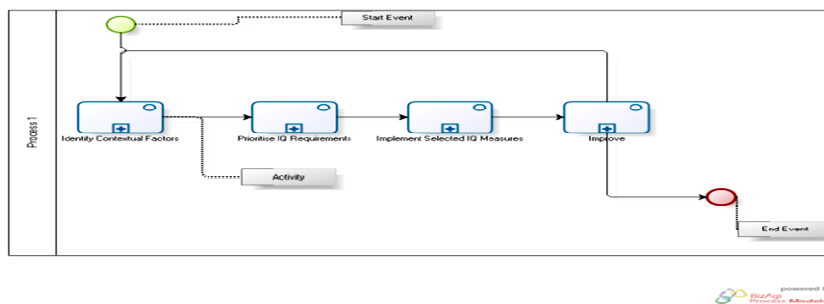


Fig. 3. High Level View of IQ Dimension Selection Method.

This general contextualisation process was detailed in sub-processes. Figures 4 to 7 outline in detail the process description that is required in order to apply our method. In conjunction with the meta-model in figure 2, the result of the subsequent

processes, provides us with a detailed set of activities and tools that allow for context factor and dimension selection.

In order to select appropriate context factors as described in figure 4 the role groups of the IS, tasks, associated IS service and access devices are identified and assessed. This assessment is done in conjunction with the domain users and IS experts. Upon identification of these a classification order is assigned to each of the context factors. The context factor and context factor measure tables in figure 2 are propagated with the appropriate values.

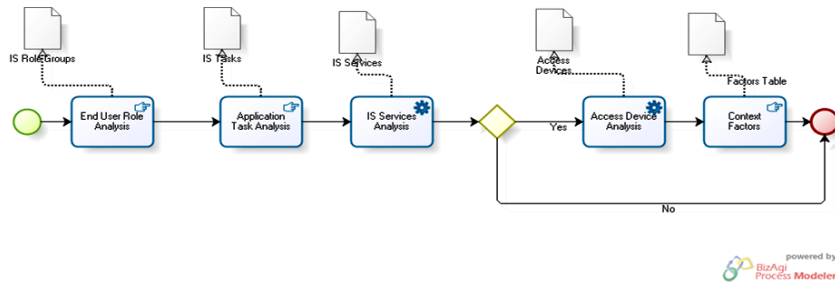


Fig. 4. Identify Contextual Factors.

Upon identification of the context factors there is a requirement to identify and prioritise dimension selection appropriate to a particular context. This requires domain experts, IS and IT managers to rank dimensions in order of priority. This may involve the application of domain metrics or survey instruments to ascertain the most important dimensions. Once this process is completed the appropriate dimension tables outlined in figure 2 are updated.

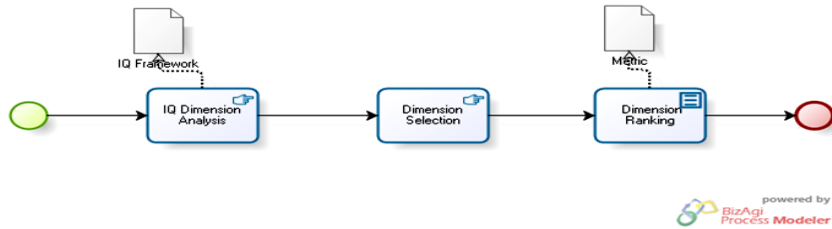


Fig. 5. Prioritise IQ Dimension Requirement.

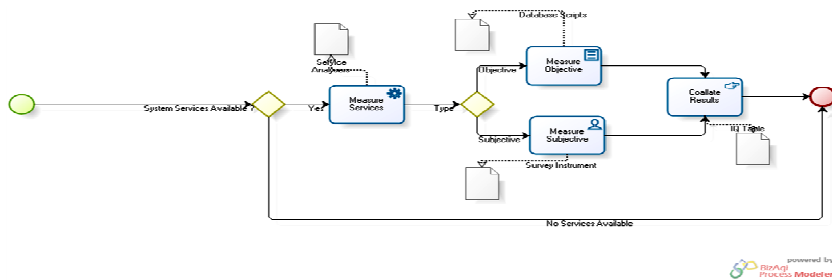


Fig. 6. IQ Measurement.

The selection of appropriate dimensions requires them to be measured. It is important that the sequence of measures is followed correctly as outlined in figure 6. The measures as outlined in table 3 require the availability of services to be checked initially followed by the objective database and subjective measures.

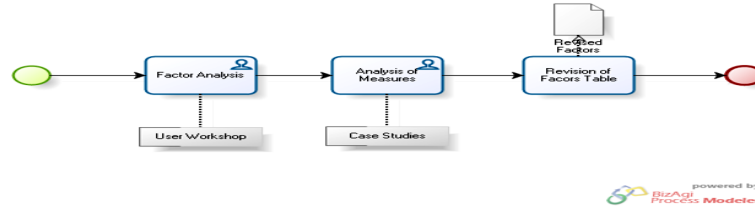


Fig. 7. Improve.

The final process of improvement involves revised context factor analysis by means of user work and measures. The revised factors are updated in the context factors table.

The contextualisation process and the result meta-model are further extended by a set of objective and subjective IQ measures. These are part of the meta-model in Figure 2. The objective measures are further subdivided into two categories data base integrity measure and software service measure. A summary of the objective and subjective IQ measures is available in Table 3. Subjective IQ Measures follow a common questionnaire approach, using questionnaire construct and a 5-point Likert scale.

Now we have described the process steps (activities) for contextualising IQ frameworks. Further we have specified objective and subjective measures for IQ. Reviewing the meta model in Figure 2, we also need to specify how various IQ measures from different IQ dimensions are aggregated.

Many researchers have proposed ways to aggregate single measures of IQ dimensions, often underlying a weighted aggregate of single values for IQ dimensions [2]. Although, recently some researchers have attempted to propose IQ value curves and trade-offs by analysing the potential impacts of IQ, many researchers propose to measure the overall impact of IQ as weighted aggregate. A principle measure of the weighed sum of all the criteria (IQ_i) is illustrated below

$$IQ = \sum_{i=1}^N \alpha_i IQC_i \quad \text{where} \quad \begin{array}{l} \forall \alpha_i : 0 \leq \alpha_i \leq 1 \\ \sum_{i=1}^N \alpha_i = 1 \end{array}$$

Equation 1. Aggregate measure of information quality.

We follow this prominent aggregation of IQ measures by weighted sums. This is reflected in our method and meta-model by specifying priorities in forms of weights. The aggregated value should define the quality level that characterizes information source. The approach to use the average as an aggregation functions may not be suitable among heterogeneous dimensions since dependencies introduces bias that negatively affect the reliability of the assessment procedure. This might be problematic, as changes in the context will have an impact on other dimensions and as a consequence the aggregate score.

Table 3. Information Quality Measures.

	Measure Name	Description	Measure
Data base Integrity	Free-of-Error	The dimension that represents whether data are correct.	Free-of-Error Rating = $1 - \left\langle \frac{N}{T} \right\rangle$ Where N = Number of data units in error and T = Total number of data units.
	Completeness	Schema, Column and Population	Completeness Rating = $1 - \left\langle \frac{C}{T} \right\rangle$ Where C = Number of incomplete items and T = Total number of items.
	Consistency	Referential Integrity, Format	Consistency Rating = $1 - \left\langle \frac{C}{T} \right\rangle$ Where C = Number of instances violating specific consistency type and T = Total number of consistency checks performed.
	Timeliness	The delay in change of real world state compared to the modification of the ISs state. The difference between the times when the process is supposed to have created a value and when it actually has.	Timeliness Rating = $R - I$ Where R = IS State Time I = Real World Time
Software Service	Database Listener	DB process	Binary Measure
	Web Service	Web Service Process	Binary Measure
	Mobile Access	Security Access	User Profile Dependent
Subjective Information Quality Measures	Timeliness	This information is sufficiently current for our work. This information is not sufficiently current for our work. This information is sufficiently timely. This information is sufficiently up-to-date for our work.	
	Free of Error	This information is correct. This information is incorrect. This information is accurate. This information is reliable.	
	Completeness	This information includes all necessary values. This information is complete. This information is sufficiently complete for our needs. This information covers the needs of our tasks. This information has sufficient breadth depth for our task.	
	Consistency	This information is consistently presented in same format This information is presented consistently. This information is represented in a consistent format.	

5 Initial Application and Analysis

We applied the developed Method to contextualise IQ Frameworks in an experiment involving 48 users. By considering various contexts, this allows us to measure the impact of the context on the perception of IQ. The **USER ANALYSIS** in Figure 4 identified 3 groups broken down between librarians, library users and technicians. This involved the users completing a number of **TASKS** with respect to information retrieval as they pertained to each particular group. The selected dimensions of the framework for the particular **ENVIRONMENT** are then broken down into objective and subjective measures. The results from application of the measures are compared upon completion.

The tasks for each user group were specific to that group. Members of the groups were randomly allocated to each of the access devices. The experiment involved control of one independent variable namely the access device. Each of the groups were assigned tasks particular to their profile. The tasks within each group were conducted from three different contexts namely workstation, web, and mobile. The first requirement of the method is to analyse the software services identified and selected. This is a binary test and examines the availability of the service. The application of the method first checks the availability of the three software services identified. In the event of service availability the objective integrity analysers are initiated. The results of this analysis are stored in an IQ table-space similar to an audit table-space [41]. The subjective survey instrument is run in conjunction with the objective integrity analysers.

The analysis of results indicate that a relationship exists between the level of IQ and the context of IS access. This confirms the requirement to select IQ dimensions appropriate to individual contexts. The implementation of our method and its validation by means of an experiment demonstrate the significance of context. A uniform application of dimensions without consideration for context we contend will not accurately reflect the true state of IQ for an IS.

We describe the process of data collection and analysis. The initial step in the analysis is the binary test of services. This important step in our method as outlined in Figure 6 it allows for the identification of dimensions associated with various software services. Subjective analysis only takes place when this analysis is complete. All the services were present therefore the subjective and objective tests were applied to all participants. In order to test the level of significance of the remainder of the results it is necessary to apply an appropriate statistical technique to the data gathered. We need to ascertain if there is a relationship between the context of the IS (Web, Workstation or Mobile) and the level of IQ. As IQ is a multidimensional concept it is necessary to do this at a dimension level.

There is a clear indication from this initial analysis that the context of the IS is significant. However in order to strengthen and build on this finding Field [42], suggests that appropriate inferential statistical techniques should be applied. A review of this literature indicates that One Way Independent Analysis of Variance (ANOVA) is appropriate. Field [42] also suggests that this technique be used when three or more statistical groups and different participants in each group will be used.

Based on confidence interval of 95%, $v_1=2$ and $v_2=27$, the critical F statistic, $F_{.05,2,15}= 3.354$ are within the “Reject H_0 ” which leads us to the conclusion that the population means are not equal.

The ANOVA test statistically indicates that the population means are not even. We have rejected H_0 . Caulcutt [43] indicates that it is possible to determine which of the sample means is statistically significant using the Scheffé Test. According to our results statistically users rate the IQ dimension of free-of-error best from the workstation context in comparison with both web and mobile. They also rate the web context statically more significant or satisfied than the mobile context.

6 Summary, Future Research and Limitations

This research contributes to the analysis of IS context and IQ. Although frequently mentioned, foremost research lacks in explaining adequately the impact that IS context has on IQ [15]. In recent times companies have invested heavily in IQ programmes in an effort to improve IQ [11]. Our research demonstrates that a relationship exists between IS context and other dimensions in an IQ framework. We have specified a method that allows context to be considered when selecting dimension. The traditional techniques of measuring IQ dimensions will also require examination as relationships between the context and other dimensions have also been established. This research contributes to the field of IQ research by providing a method and test environment that can be employed in a context related manner. It has the potential to allow organisations to measure the impact of introducing new contexts post the development of an IS.

Although the research revealed interesting results, our research currently concentrates only on a subset of dimensions. The application of further experiments addresses this limitation. Correlation has been used to examine the relationship between IQ dimensions; Analysis of Variance will be completed for all sections of the research. This will allow for a full examination of experimental data. Furthermore it is intended to improve and extend the prototypical implementation of the tool as software application.

References

1. Wand, Y. and R. Wang, *Data Quality Dimensions in Ontological Foundations*. Communications of the ACM, 1996. 39: p. 184-192.
2. Wang, R. and D. Strong, *Beyond Accuracy: What Data Quality Means to Data Consumers*. Journal of Management Information Systems, 1996. 12: p. 5-34.
3. Pipino, L., Y. Lee, and R. Wang, *Data Quality Assessment*. Communications of the ACM, 2002: p. 211-218.
4. Kahn, B. and D. Strong, *Information Quality Benchmarks: Product and Service*. Communications of the ACM, 2002: p. 184-192.
5. Recker, J., et al. *Model-Driven Enterprise Systems Configuration*. in *18th Conference on Advanced Information Systems Engineering*. 2006. Luxembourg: Springer.
6. Lee, Y., *Crafting Rules: Context-Reflective Data Quality Problem Solving*. Journal of

- Management Information Systems, 2004. 20(3): p. 93-119.
7. Redman, T., *Data Quality for the Information Age*. 1996: Artech House Publishers.
 8. AlHakim, L., *Challenges of Managing Information Quality in Service Organizations*. 2007: Idea Group Pub.
 9. Date, C., *Databases in Depth*. 2005: O Reilly.
 10. Loshin, D., *Enterprise Knowledge Management - The Data Quality Approach*. 2001: Morgan Kaufmann.
 11. AlHakim, L., *Information Quality Management Theory and Applications* 2007.
 12. Olson, J., *Data Quality: The Accuracy Dimension*. 2003: Morgan Kaufmann.
 13. Adelman, S., L. Moss, and M. Abai, *Data Strategy*. 2005: Addison-Wesley.
 14. Sangwan, R., et al., *Global Software Development Handbook*. 2007: Auerbach Publications.
 15. Knight, S. and J. Burn, *Developing a Framework for Assessing Information Quality on the World Wide Web*. *Informing Science*, 2005. 8: p. 159-172.
 16. Helfert, M. and M. Ge. *A Review of Information Quality Research*. in *11th International Conference on Information Quality*. 2006.
 17. Morris, S., J. Meed, and N. Svensen, *The Intelligent Manager*. 1996, London: Pitman Publishing.
 18. Huang, K., Y. Lee, and R. Wang, *Quality Information and Knowledge Management*. 1999.
 19. Chengalur-Smith, I. and P. Duchessi, *The Initiation and Adoption of Client Server Technology in Organizations*. *Information and Management*, 1999. 35(7): p. 77-88.
 20. Miller, *An expanded instrument for evaluating information system success*. *Information and Management*, 1996. 31(2): p. 103-118.
 21. Berndt, E.R. and D.M. Cutler, *Medical Care Output and Productivity*. 2001: National Bureau of Economic Research.
 22. Xu, H., et al., *Data Quality Issues in implementing an ERP*. *Industrial Management & Data Systems*, 2002. 102(1): p. 47-59.
 23. Davenport, T., *Information Ecology: Mastering the Information and Knowledge Environment*. 1997, Oxford: Oxford University Press.
 24. Heinrich, B. and M. Helfert. *Analyzing Data Quality Investments in CRM: a Model Based Approach*. in *8th International Conference on Information Quality*. 2003. Boston, Massachusetts, USA.
 25. Ballou, D.P., et al., *Modelling Information Manufacturing Systems to Determine Information Product Quality*. *Management Science*, 1998. 44(4): p. 462-484.
 26. Amicis, F.D. and C. Batini, *A Methodology for Data Quality Assessment on Financial Data*. *Studies in Communication Sciences*, 2004. 4(2): p. 115-137.
 27. Kahn and Strong. *Product and Service Performance Model for Information Quality: an update*. in *4th International Conference on Information Quality*. 1998. Boston, Massachusetts, USA.
 28. Xu, H. and A. Koronios, *Understanding Information Quality in E-business*. *Journal of Computer Information Systems*, 2004. 45(2): p. 73-82.
 29. Rittberger, M. *Certification of Information Services*. in *Proceedings of the 1999 Conference on Information Quality*. 1999. Boston, Massachusetts, USA.
 30. English, L., *Improving Data Warehouse and Business Information Quality*. 1999, New York: Wiley & Sons.
 31. Li, S. and B. Lin, *Assessing Information Sharing and Information Quality in Supply Chain Management*. *Decision Support Systems*, 2006. 42(3): p. 1641-1656.
 32. Ballou, D.P. and G.K. Tayi, *Enhancing Data Quality in Data Warehouse Environments*. *Communications of the ACM*, 1999. 42(1): p. 73-78.
 33. Fisher, C., et al., *Introduction to Information Quality*. 2006: MIT.
 34. Goodhue, D.L., *Understanding user evaluations of Information Systems*. *Management Science*, 1995. 41(12): p. 1827-1842.
 35. Misser, P., et al., *Improving data quality in practice: a case study in the Italian public*

- administration*. Distributed and Parallel Databases, 2003. 13(2): p. 125-169.
36. Ballou, D.P. and H. Pazher, *Modeling completeness versus consistency tradeoffs in information decision contexts*. IEEE Transactions in Knowledge and Data Engineering, 2003. 15(1): p. 240-243.
 37. Wang, R.Y., *A Product Perspective on Total Quality Management*. Communications of the ACM, 1998. 41(2): p. 58-65.
 38. Brinkkemper, S., K. Lyytinen, and R.J. Welke, *Method Engineering Principles of method construction and tool support*. 1996: Chapman & Hall.
 39. Punter, T. and K. Lemmen, *The META-model: towards a new approach for Method Engineering*. Information and Software Technology, 1996. 38(4): p. 295-305.
 40. Gutzwiller, T., *Das CC fur den Entwurf von betrieblichen, transaktionsrieentierten Informationssystemen*, Physica. 1994: Heidelberg.
 41. Greenwald, R., R. Stackowiak, and R. Stern, *Oracle Essentials: Oracle Database 11g*. 4th ed. 2004: O Reilly.
 42. Field, A. and G. Hole, *How to Design and Report Experiments*. 2006.
 43. Caulcutt, R., *Statistics in research and development*. 1991: CRC Press.

Author Index

Brazier, F.	132
Burdescu, D.	99, 142
Castro, A.	70
Chao, K.-M.	117
Cheng, D.-Y.	117
Cordeiro, J.	17
Filipe, J.	3, 17
Foley, O.	150
Helfert, M.	150
Ionascu, C.	99
Konstantas, D.	70
Liu, K.	17
Lo, C.-C.	117
Logofatu, B.	99
Mihaescu, M.	99
Nieuwenhuis, L.	89
Serugendo, G.	70
Shah, N.	117
Shishkov, B.	36, 52
Spahiu, C.	142
Stanescu, L.	142
Tsai, C.-F.	117
van Sinderen, M.	52, 132
Warnier, M.	52, 132

Proceedings of the 4th International Workshop on
Enterprise Systems and Technology - I-WEST 2010
ISBN: 978-989-8425-44-7