CONTEXT-05

Doctoral Consortium

- Proceedings -

Chiara Ghidini
Bich-Liên Doan
Patrick Brézillon (eds.)

Paris, 5 July 2005

Preface

The papers published in these proceedings were presented at the 1st CONTEXT Doctoral Consortium. The Doctoral Consortium was held in Paris, France on July 6, 2005 during the Fifth International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-05).

The CONTEXT conference series is meant to provide an interdisciplinary forum where researchers can exchange ideas, methodologies, and results on context and has become a reference conference for all people doing research on context. Purpose of the 1st CONTEXT Doctoral Consortium was to bring together PhD students working on many different fields and aspects on the topic of context and to give them the opportunity to present and to discuss their research in a constructive, interdisciplinary, and international environment. The consortium was accompanied by prominent researchers on the topic of context to discuss the current research areas and to give helpful feedback to the students.

For this year's consortium, we received 27 submissions. Among them 15 have been accepted for presentation. The submissions demonstrate a variety of research topics and research approaches. Topics cover research on agent-based and multi-agent systems, cognitive science, context-aware applications and systems, heterogeneous information integration, knowledge representation, learning, linguistics, organizational sciences, semantic web systems, and ubiquitous computing. Each paper was presented and the discussed with prominent researchers and fellow students.

We thank the program committee members for all they did to ensure the high quality of accepted contributions and to stuck to an exacting reviewing schedule. We are extremely grateful to their efforts. We thank all the participants for their great interest and effort displayed both in preparation and presentation of their own work, as well as in the discussion of the contribution of others. Finally, we sincerely thank the organizing committee and local organizers of the CONTEXT-05 conference for their continuous support to preparing the Doctoral Consortium.

July 2005

Chiara Ghidini
Bich-Liên Doan
Patrick Brézillon

Organization

The CONTEXT'05 Doctoral COnsortium took place at the Pierre & Marie Curie University.

Organizing Committee

Doctoral Consortium Chair:	Chiara Ghidini (ITC-irst, Trento, Italy)
Local Chair:	Bich-Liên Doan (Supélec, France)

Program Committee

Varol Akman, Bilkent University, Turkey
John Barnden, University of Birmingham, UK
Carla Bazzanella, Universita' di Torino, Italy
Massimo Benerecetti, Universita' di Napoli, Italy
Paolo Bouquet, Universita' di Trento, Italy
Patrick Brézillon, University of Paris 6, France
Bich-Liên Doan, Supelec, France
Bruce Edmonds, Manchester Metropolitan University, UK
Anita Fetzer, University of Lueneburg, Germany
Alice G.B. Ter Meulen, University of Groningen, The Netherlands
Chiara Ghidini, ITC-irst, Italy
Rolf Nossum, Agder University College, Norway
Luciano Serafini, ITC-irst, Italy
Rich Thomason, University of Michigan, USA
Roy Turner, University of Maine, USA
Johan van Benthem, University of Amsterdam, The Netherlands
Terry Winograd, Stanford University, USA

Table of Contents

Speaker/Hearer Representation in DRT Model of Presupposition

Yafa Al-Raheb

University of East Anglia

1 Introduction

The paper presents a model of speaker and hearer mental spaces, in specific beliefs and acceptance spaces, and how they relate to presupposition and assertion. What speakers presuppose gives an indication as to what speakers believe or `accept` (weaker degree of belief) and what they believe hearers believe or accept (cf. AL-Raheb 2004). Presupposition is split into speaker presupposition and hearer presupposition and each agent's mental states as regards presupposition is represented in DRT (Discourse Representation Theory). Distinguishing speaker presupposition from hearer presupposition helps make the approach to presupposition within DRT more pragmatic than semantic, since we are not speaking of truth conditions. A series of checks has been devised for speakers and hearers to make presuppositions felicitously in a cooperative dialogue. The process of checks further enhances the differences between speaker and hearer presupposition.

2 Speaker and Hearer Presupposition

The relationship between presupposition and assertion is based on dividing presupposition into speaker presupposition and hearer presupposition. From the generational point of view (i.e. the speaker's), speakers make utterances to communicate new information, unless the intention is to change a hearer's beliefs. Generally speaking, to generate an utterance, there would be some discrepancy between the speaker's beliefs and the speaker's beliefs about the hearer's beliefs. The discrepancy motivates making an assertion,A, which may need presupposed arguments to be understandable. This is a top down approach, by first deciding on the assertion through checking belief discrepancies - building floor one - then finding the right presuppositions to make the assertion understood - laying the foundations for the ground floor for floor one which is already there.

Hearer presupposition differs in that being on the receiving end, utterances are split into presupposition and assertion, where possible, and presuppositions are first needed to establish links to objects in order for the new information to be understood by the hearer. For a hearer, assertions build on presupposition, i.e. a bottom up approach (assertion is supported by presupposition).

As for the A part of an utterance, the hearer can first `accept` (weakly believe) the new information and later on turn that assertion into a belief, by adding it to her belief set. However, it is worth mentioning that when making an utterance, both S and H focus their attention on A, which can get accepted by H. In such a case, H may adopt A as a belief and indicate so to S, making A a mutual belief, which may or may not later serve as a presupposition.

3 Checks on Presupposition

The treatment of presupposition presented here depends on the view that presupposition may introduce given or new information. Presupposition is also analyzed both from the speaker's and hearer's perspectives. Speaker presupposition differs from hearer presupposition in terms of three checks agents are hypothesized to perform when introducing or dealing with presupposition. The checks are (1) `clarification check` (Purver 2004), (2) `informativity check`, and (3) `consistency check` (van der Sandt 1992). The checks are similar in principle to Purver (2004) and van der Sandt (1992). However, they are developed here as a process which separates speaker generation from hearer recognition, allowing us to differentiate speaker presupposition from hearer presupposition. The three apply to both speaker and hearer and to presupposition and assertion. The checks are also applicable to assertions, but the discussion of checks and assertion is not covered in this paper.

3.1 Speaker Presupposition

To generate a presupposition, we are assuming that a speaker is bound by Grice's Cooperative Principle (1989). The `speaker` needs to have reason to believe that her presupposition is going to be both clear and consistent with the hearer's beliefs. The speaker can have previous context in memory that shows her presupposition to be consistent with her beliefs about the hearer's beliefs. However, when such evidence is lacking, the speaker can still introduce the presupposition and then make the judgment that the presupposition is consistent due to lack of negative feedback. For the informativity check, the speaker generally introduces a presupposition she believes to be known to the hearer. However, the speaker may wish to introduce a topic into the dialogue, knowing that the hearer has no previous knowledge of the topic. Such a move may indicate, for instance, that the topic is not the focus of her attention. According to this treatment of presupposition, whether the speaker believes the presupposition is new or old to the hearer, the speaker expects the hearer to 'go along' with and `accept` the presupposition since it is not the main focus (Simons 2003).

When the presupposition is known to the hearer, the speaker generally expects the hearer to bind it to information already accepted or believed. Of course, the hearer can experience some difficulty in understanding and ask for a 'clarification'.

However, the information presented as a presupposition may be new. The speaker may wish to introduce a topic into the dialogue, knowing full well that the hearer has no previous knowledge of the topic. Here, we follow Geurts's (1999) classification of remarkable and unremarkable information [1]. Unremarkable information is information that people can accept without too much questioning, such as having a brother or a sister. An example of remarkable information would be:

Example 1. My private jet arrives this afternoon.

Since not everyone has the means to have a private jet, such information will at least raise an eyebrow. Being cooperative, the speaker will assume, unless the hearer indicates otherwise, that the information she provides in the presupposition is unremarkable for the particular hearer in the particular context, and that the hearer will accommodate the information by either accepting it, or believing it. It has to be said that this is of course an idealized situation. The speaker does not always have beliefs concerning whether the hearer already believes the presupposed information or not.

Therefore, the speaker has to be prepared for cases when despite being cooperative, the hearer might perceive the information as unclear and/ or contradictory. What this means for this treatment of presupposition is that generally the speaker believes the new information presented in the presupposition is unremarkable; therefore, the speaker will expect the hearer to accommodate the new information. However, if the hearer finds the new information unusual or remarkable, the speaker can expect the hearer to check whether the presupposition is consistent with his beliefs or not. The speaker can also expect the hearer to ask for clarification if the presupposition is not clear.

If clear, the speaker can expect the hearer to accommodate the information and can assume that the information has been accepted, unless it is indicated through `strong positive feedback` that the information is actually believed. However, if the presupposition is not clear, the speaker can expect the hearer to ask for clarification and a clarification process takes place, where the hearer might ask for more clarification if the information is still not clear. When the information is clear, the speaker expects the hearer to provide feedback. However, the lack of feedback from the hearer about this information will be considered `weak positive feedback` that the hearer has accepted the information. Despite the speaker's best efforts to be cooperative, there are cases where the presupposition contradicts the hearer's previous beliefs.

In example, 2, in uttering S1, the speaker is being cooperative by following the presupposition checks, i.e. the utterance is clear, consistent and known to the hearer. The speaker believes that the hearer believes there is a new course book, drs5, i.e. that the speaker assumes this is known by the hearer (informativity check). The speaker assumes that by introducing the assertion about buying the course book, the speaker is being consistent. Additionally, being cooperative, the speaker believes the information communicated to be clear.

[1] What counts as remarkable/unremarkable in particular contexts is in itself an interesting research question that involves social and cultural boundaries. I do not pusue this here, since my focus is on the broader categories that need to be postulated in order to describe the operations on presupposition in dialogue.

Example 2. S1: I have to buy the new course book.
H1: Try the secondhand bookshop.

The model partially presented here is built on a reconciliation between two DRS variants provided by Kamp et al. (2005). The reconciled DRS is also extended to include acceptance space (for weaker beliefs) and the intention space is expanded to represent the linguistic content of the current utterance. The reconciled DRS allows the representation of speaker and hearer cognitive states separately before and after each utterance, following the stages of a dialogue from speaker generation to hearer recognition. Figure 1 shows speaker generation (uttering) S1. The speaker believes the hearer believes there is a new course book (presupposition). 'i' refers to the agent (here speaker), 'you', the other, 'p' for presupposition, 'a' assertion, 'c' for acceptance propositions, 'b' for belief, 'BEL' for believe, 'ACCEPT' for accept, and 'INT' for intention. DRSs are given labels so that they can be referred to, e.g. drs2. The attitude predicates refer to the `Attitude DRS`, BEL, ACCEPT, or INT. For example, attitude(i,'BEL', drs2) is equivalent to I believe the content of drs2.

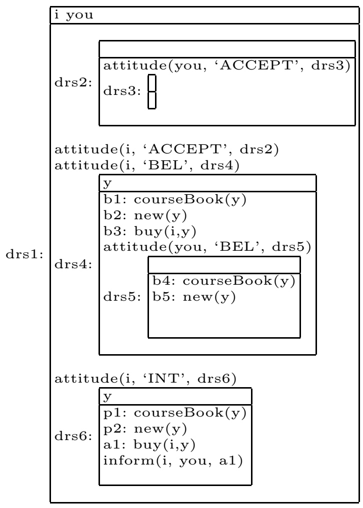

Fig. 1. Speaker Generation S1

As example 2 shows, the speaker's assumption about the hearer's beliefs turns out to be correct. The hearer does not seem to have problems with the presupposition, there is a new course book, and the hearer indeed makes a suggestion about where to buy the course book.

3.2 Hearer Presupposition

The first check to apply to `hearer` presupposition is the clarification check. When hearing P, the hearer first checks he has understood S correctly. If the presupposition is not clear, the hearer can ask the speaker to clarify her utterance. For example, consider a dialogue between a customer and customer service assistant about a gas heater:

Example 3. Customer: How long does it take to fix my gas heater?
Customer Service Assistant: Your what?
Customer: My gas heater needs fixing.

After checking whether the information is clear, the customer service assistant, asks the customer to clarify. In this particular case, the lack of clarity can be attributed to, for example, not hearing very well. Here, the clarification may have nothing to do with whether P is believed by H. The hearer expects the speaker to provide an explanation or clarification. The speaker needs to provide an explanation. When the information is cleared up, then the hearer can provide feedback that the information is clear. However, lack of feedback will be considered as `weak positive feedback`, i.e. the information is `accepted`. Generally, after providing an explanation, the speaker can conclude that the hearer now has no problems with the presupposition; in this example that the customer has a gas heater.

If we go back to example 2, and change it so that the hearer does not actually hold the belief that there was a new course book, example 4:

Example 4. S1: I have to buy the new course book.
H1: I didn't realize they changed it.

In this case, the representation of the hearer's recognition of S1 would reflect the discrepancy between the hearer's beliefs and what the speaker believes the hearer believes. Before performing the informativity check and the consistency check, the hearer first checks if the information is clear. The hearer encounters a problem with this belief discrepancy, necessitating a 'clarification process'. Until the information is clarified, the content of S1 is placed on hold, and is therefore not added to the hearer's acceptance or belief DRSs in figure 2. DRS6 shows that hearer has recognized the speaker's intentions and represents the linguistic content of the speaker's utterance, S1. Being cooperative, the hearer believes the speaker believes the content of her utterance, and adds to his beliefs about the speaker that the speaker believes the content of the presupposition and the assertion.

After receiving H1, the speaker recognizes that the hearer is having some problems with the presupposition and that a clarification process is needed. This means that unless the presupposition is clear (Clarification check), the hearer will not have accepted either the presupposition or the assertion. The speaker revises her beliefs about the hearer's beliefs in drs5 in figure 3 differs from drs5 in figure 1. The absence of more beliefs attributed to the hearer is

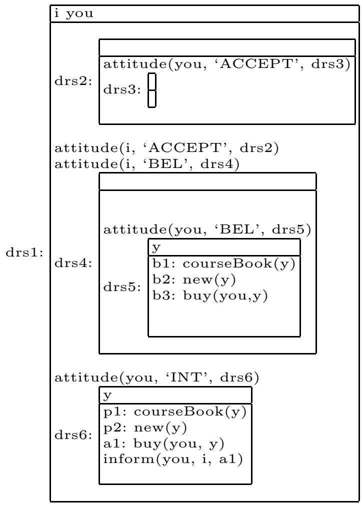

Fig. 2. Hearer Recognition of S1

quite significant. Where before the speaker expected the hearer to accept or even believe the presupposition and the assertion, the speaker now does not even add the content of S1 to the representation of hearer's cognitive state; the absence of propositions in drs5. Instead, the speaker's intention space and the speaker's hearer-intention recognition space are used to resolve problems with presuppositions and assertions before judgment about belief and acceptance is made. However, the speaker continues to hold the beliefs that she assumes that the hearer has formed about her beliefs (drs6, figure 3).

Having realized the hearer is having problems with the clarification check, the speaker can now address the problems. The speaker re-asserts that there is a new course book to the hearer. In this case, the speaker is attempting to change the hearer's belief or acceptance space concerning the book. The speaker provides an explanation to clarify the presupposition in S2, in example 5.

Example 5. S1: I have to buy the new course book.
H1: I didn't realize they changed it.
S2: Yes they did.

When the clarification process is resolved, the hearer can then check for informativity [2]. If the information the speaker presents as a presupposition is known to the hearer, i.e. part of acceptance or belief space, it is not necessary to add P as a new belief. Known presuppositions do not get accommodated, but get bound to previous referents in the context (Geurts 1999). We are assuming here

[2] It is worth noting that a clarification process can be initiated at any stage of the check process.

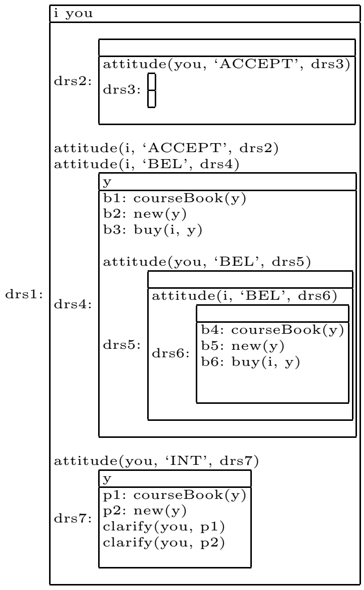

Fig. 3. Speaker Recognition of H1

that the hearer's knowledge of a presupposition means that this presupposition does not contradict with previous beliefs held by the hearer. The hearer may previously hold a belief about the presupposition, i.e. H may already bel P. In this case, it is not necessary to add a new belief that P. The same applies to acceptance. If the hearer's acceptance DRS or space already contains accept (P), nothing is added.

If the presupposition provides new information, the hearer then performs the consistency check. If the information provided contradicts previous beliefs, the hearer can reject the presupposition and an attempt would be made to remedy or fix the dialogue. For example,

Example 6. Speaker: Mia's husband is coming for dinner.
Hearer: This can't be! Mia is widowed!

When the information presented by the presupposition is consistent with the hearer's beliefs or acceptance space, the hearer makes a judgement about whether the information is remarkable or unremarkable. If the information is unremarkable, the hearer accommodates the new information by either accepting it or believing it.

If the presupposition in remarkable, the hearer can check again for clarity. This is a different type of clarity check from the one performed initially. This time the hearer requires an explanation for the oddness of the information in the presupposition. This is when the clarification process starts again. For example,

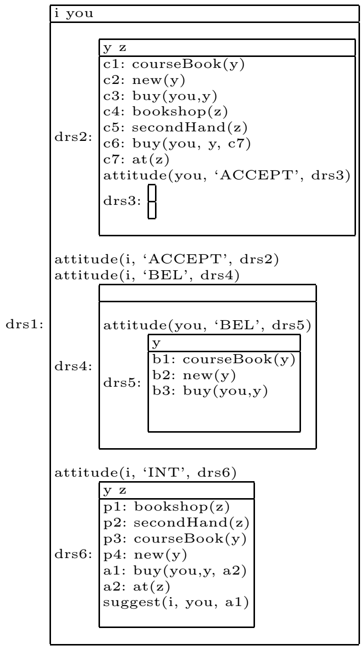

Fig. 4. Hearer Generation H2

Example 7. Speaker: My pet lion requires a lot of attention.
Hearer: Your pet what?
Speaker: Oh sorry, I mean one of those virtual pets you take care of.

Again, the hearer may provide feedback as to whether the explanation has been accepted or not. Unless negative feedback is provided, the hearer is expected to at least accept the presupposition. In addition, the speaker can assume that the hearer now has no problem with the presupposition. This is assuming that the hearer has indeed accepted the explanation provided by the hearer. If the hearer is not convinced by the speaker's explanation, an attempt at fixing or repairing the dialogue is needed.

Example 8. S1: I have to buy the new course book.
H1: I didn't realize they changed it.
S2: Yes they did
H2: You can buy it at a secondhand bookshop.

In example 8, being satisfied with the speaker's explanation and having performed the three check, the hearer finally `accepts` the problematic presupposition. Figure 4 shows the hearer accepting the speaker's explanation of the problematic presupposition in question by making a suggestion based on it. In this case, the speaker's reassertion of the presupposition serves as a confirmation that it is correct, and the hearer adds the presupposition to his acceptance space [3]. The hearer is also making a suggestion about where to get the course book. Again, assuming the hearer is being cooperative, the hearer expects the utterance to be clear and consistent. As this is a `suggestion`, the hearer's beliefs are weak beliefs and are, therefore, represented in his acceptance space, drs1.

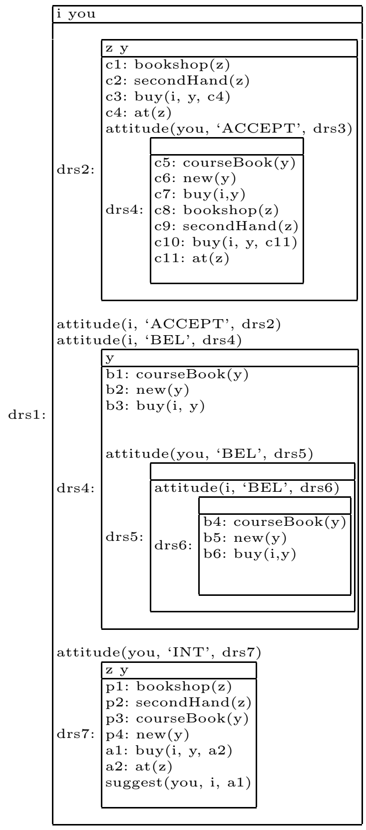

Fig. 5. Speaker Recognition of H2

Receiving a suggestion from the hearer about the book, where he used an assertion built on the presupposition in question, the speaker concludes that the hearer accepts (weakly believes) the presupposition as he has used it in

[3] Even though H2 is indirect, it acts as a confirmation of accepting the presupposition used originally by the speaker. The example raises interesting theoretical questions about how soon after questioning a presupposition the hearer will form strong beliefs about the presupposition. It is safer to presume that the hearer first accepts the presupposition, then eventually adopts it as a belief

his utterance and there are no longer any problems with the presupposition. This is the case even though the hearer has not provided explicit feedback or acknowledgment that the problem has been resolved. The suggestion in H2 end the clarification process (figure 5) . Having checked for clarity, consistency and informativity, the speaker also accepts the suggestion that she should try buying the book in a second hand bookshop.

4 Conclusion

What this paper has attempted to show is, firstly, a model of speaker and hearer presupposition in a representation of dialogue in DRT and, secondly, their relation to the participants' cognitive states. The presupposition checks ensure participants are being cooperative and details a strategy for dealing with clarifying problematic presuppositions and their effect on speaker and hearer beliefs. In general, the proposed modification to DRT introduce aspects of cognitive states involved in interaction; it is in this sense that the paper has sought to render DRT 'more pragmatic'.

References

[2004] Al-Raheb, Y. 2004. *Presupposition and Belief in DRT: Towards a New Implementation.* In: Ginzburg, J. and Vallduvi, E. Catalog 04: Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue, Barcelona, 144-145.

[1999] Geurts, B. 1999. Presuppositions and Pronouns: Current Research in the Semantics/ Pragmatics Interface Oxford: Elsevier

[1989] Grice, P. 1989. Studies in the Way of Words Cambridge, Massachusetts: Harvard University Press

[2005] Kamp, H., van Genabith, J. and Reyle, U. 2005 *The Handbook of Logic.* Gabbay, D. and Guenthner, F. http://www.ims.uni-stuttgart.de/ hans/.

[2004] Purver, M. 2004. Claire: the Clarification Engine. Ginzburg, J. and Vallduvi, E. Catalog Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue. Universitat Pompeu Fabra, Barcelona.

[1992] van der Sandt, R. 1992. *Presupposition Projection as Anaphora Resolution.* Journal of Semantics (9), pp 333-377

Problems with Multiple Focus and 'Only' in Hungarian

Kata Balogh

ILLC/Department of Philosophy
University of Amsterdam
k.balogh@uva.nl

Abstract. In my paper I discuss possible interpretations of Hungarian sentences containing multiple prosodic foci and 'only'. Hungarian as a discourse-configurational language has a special position for identificational/exhaustive focus, which makes the role of 'only' problematic. Next to this problem I focus on sentences with multiple foci with and without 'only' and discuss the possible interpretations: complex focus meaning (Krifka, 1991) and the scope-reading, where one focused constituent takes scope over the other one.

The main aim of my paper is to report the first observations of an ongoing research on the semantics and pragmatics of focus and 'only' in Hungarian. Consequently, I will not present a fully worked out theory, but rather a problem set with proposals for possible solutions.

I will point out two main problems with the interpretation of 'only' and focus in Hungarian. The first question is the interpretation of 'only' that goes along with exhaustive focus. Hungarian has a special structural position for (identificational) focus which is obligatorily pre-verbal, and the elements in this position get exhaustive interpretations. Hungarian 'only' (*csak*) always goes along with this identificational/exhaustive focus, which suggests that 'only' (*csak*) is not responsible for the exhaustive meaning. The second problem also concerns the function of 'only'. In the case of multiple focus, it is possible that two arguments are prosodically focused, modified by 'only' and the sentence has a *complex focus* (Krifka, 1991) interpretation (see (1)).

(1) Csak ZEPHIRT hívta fel csak CELESTE.
only Zephir.acc called VM only Celeste
'It is the Zephir, Celeste pair of whom the first was called by the second.'

In my paper I will discuss these problems, however, I will focus on example (1), since it seems to be very challenging for the semantic and syntactic analysis as well.

1 Focus and 'only' in Hungarian

Hungarian is a discourse-configurational language where the syntactic structure of a sentence is driven by discourse-semantic considerations (É. Kiss, 1995).

Hungarian has a special structural position for (identificational) focus[1] (see 2b) which is obligatorily pre-verbal, and the elements in this position are always interpreted exhaustively. See the neutral sentence in (2a) versus (2b) where the object is focused, it moves to the preverbal position – which was occupied by the verbal modifier (VM) in the neutral sentence – and gets exhaustive interpretation.

(2) a. Celeste felhívta Zephirt.
Celeste VM-called Zephir.acc

'Celeste called Zephir.'

b. Celeste ZEPHIRT hívta fel.
Celeste Zephir.acc called VM

'It is Zephir whom Celeste called.'

É. Kiss (1998a) distinguishes two types of focus in Hungarian: *identificational focus* (exhaustive) and *information focus* (non-exhaustive). Her main claims are that these two types are different both in syntax and semantics, and that identificational focus is not uniform across languages. Here I will discuss only the identificational focus, which has the following properties according to É. Kiss (1998a): it expresses exhaustive identification, certain constituents are out, it takes scope, it involves movement and it can be iterated.

1.1 Exhaustivity and 'only'

In this section we see how exhaustive focus can be treated in the classical theories of Structured Meanings (e.g. Krifka 1991, 2004) and the Partition Semantics (Groenendijk & Stokhof 1984, 1990). In these analyses 'only' and exhaustification are identified and we will see that this will not do for Hungarian.

Krifka (1991, 2004; among others) demonstrates the structured meaning framework for questions and the focusation of answers. Sentences with focus are expressed as a *focus–background pair* $\langle F, B\rangle$ so that if we apply the background to the focus $B(F)$ we get the ordinary interpretation.

(3) a. $[\![\text{CELESTE}_F \text{ came.}]\!] = \langle Celeste, \lambda x[came(x)]\rangle$

b. $B(F) = \lambda x[came(x)](Celeste) = came(Celeste)$

To get the intended meaning of Hungarian exhaustive focus we need an exhaustivity operator which also applies to the Focus-Background pair:

(4) $EXH(\langle F, B\rangle) = [B(F) \wedge \forall X \in Alt(F)[B(X) \rightarrow X = F]]$

With this exhaustivity operator we get the right interpretation for sentences containing identificational focus, but in this way sentences with identificational

[1] Here and later on I indicate the focused (accented) element with small capitals.

focus and sentences with 'only' will get the same interpretation, since the interpretation of 'only' (defined by Krifka, see below) and the exhaustive operator (above) are the same.

The focus sensitive particle 'only' is analysed as an operator which takes a focus-background structure:

(5) Meaning rule for 'only' (simple version):
$[\![\text{only}]\!](\langle F, B\rangle) = [B(F) \wedge \forall X \in Alt(F)[B(X) \rightarrow X = F]]$

The same holds for the question-answer analysis of Groenendijk and Stokhof (1984, 1990). For the semantics of linguistic answers they define an answer formation rule introducing an *exhaustivity operator*, which gives the minimal elements from a set of sets. This exhaustivity operator can be applied for Hungarian identificational focus:

(6) $EXH^n{=}\lambda\mathcal{R}^\backslash\lambda R^n[\mathcal{R}^\backslash(R^n) \wedge \neg\exists R'^n[\mathcal{R}^\backslash(R'^n)\wedge$
$R^n \neq R'^n \wedge \forall x_1...x_n[R'^n(x_1, ..., x_n) \rightarrow R^n(x_1, ..., x_n)]]]$

Also in this model exhaustification is analysed as 'only'. If we give the answer *'Babar and Celeste'* to the question *'Who came?'* in (7a), then it is interpreted as *'Only Babar and Celeste came'*. According to this, sentence (7b) and (7c) get the same interpretation as (7d).

(7) a. Ki jött el?
who came VM
'Who came?'

b. CELESTE jött el.
Celeste came VM
'It is Celeste who came.'

c. Csak CELESTE jött el.
only Celeste came VM
'Only Celeste came.'

d. $(EXH(\lambda P.P(c)))(\lambda x.came'(x)) =$
$\lambda P\forall x[P(x) \leftrightarrow x = c](\lambda x.came'(x)) =$
$\forall x(came'(x) \leftrightarrow x = c)$

1.2 'Only' pragmatics

For the above problem we proposed an analysis for 'only' and focus in the context of a question (Balogh, 2005). The facts above about Hungarian focus position suggest that 'only' is not responsible for the exhaustive meaning. On our view in Balogh (2005) 'only' has rather a pragmatic function to cancel the hearer's expectation.

For the question *Who came?* in (7a) sentence (7b) and (7c) are semantically equivalent answers, saying that it is Celeste and nobody else who came. Therefore it seems that the appearance of *csak* ‘only’ does not make any difference. But consider example (8) where we pose the same question in plural, so we make explicit an expectation of more persons who came.

(8) Kik jöttek el?
who.pl came VM
‘Who came?’

Question (8) cannot be answered with a simple identificational focus (7b), but (7c) – with ‘only’ – is felicitous. What ‘only’ does here is simply cancelling the expectation, therefore I claimed[2], that ‘only’ (in the answers) has a pragmatic rather than a semantic function.

To explain why we can answer the singular and plural questions above in a different way I use the partition semantics framework from Groenendijk and Stokhof (1984, 1990). They introduce a partition semantics account for questions and answers. The semantic interpretation of an interrogative is an equivalence relation over the set of possible worlds, thus an interrogative sentence denotes a partition of the logical space. Every block of the partition made by $?\phi$ contains the possible worlds where the extension of ϕ is the same.

(9) $[\![?\boldsymbol{x}\phi]\!] = \{(w,v) \in W^2 | [\![\lambda\boldsymbol{x}\phi]\!]^w = [\![\lambda\boldsymbol{x}\phi]\!]^v\}$

The focus expresses exhaustive identification, thus it contains an implicit exhaustivity (EXH) operator (along Groenendijk & Stokhof, 1984, 1990). Consequently, the proposition that a sentence with identificational focus denotes is one of the propositions in the partition made by the underlying question. Thus identificational focus selects one block from the partition, in case of (7b) the block containing the proposition *only Celeste came*.

From the questioner’s side (8) has an explicit expectation of more than one person (from the relevant domain) who came. This should be interpreted as a *restriction* on the partition, in this case we restrict us to the blocks where more persons came. For the identificational focus only the restricted area is accessible to select a block. Therefore we cannot reply to (8) with (7b) – the block where the proposition is *only Celeste came* is not among the available ones–, but we can reply with (10).

(10) Babar és Celeste jött el.
Babar and Celeste came VM
‘It is Babar and Celeste who came.’

[2] Following Zeevat (1994, 2002). In his examples ‘only’ seems to be superflous, and he concludes that the function of ‘only’ is less semantic and more pragmatic than it was discussed before. He suggests two possible ways to deal with this problem. The first one is that ‘only’ has a pragmatic function to cancel the expectation of the questioner, and the second one is that ‘only’ makes exhaustivity stronger in the way that it expands the extension of the restriction on the hidden wh-phrase in the topic. Considering the Hungarian data I prefer the first solution.

In fact, for question (8) it is not excluded to give an answer that expresses that Celeste and nobody else came, but in case of (8) we need ‘only’ to go explicitly against the expectation of the questioner. Thus ‘only’ deletes the restriction, whereby the blocks which were excluded before pop-up again, so they become accessible for the identificational focus to point out one of them.

At this point we can raise the question why we can answer (7a) with (7c) even if the questioner does not have any expectation about how many people came. I claim that in this case ‘only’ gives information about the answerer’s previous expectations, namely the answerer thought that there should be more people who came. But according to the questioner’s information state this additional information is irrelevant.

Summing up this section, I assume the following analysis. The interpretation of a question is an ordered pair. The first element is the semantic interpretation: the (unrestricted) set of propositions $\mathcal{P}$ (complete semantic answers). The second element is the set of propositions $\mathcal{P}'$ restricted by certain pragmatic effects.

(11) $[\![Q]\!] = \langle \mathcal{P}, \mathcal{P}' \rangle$, where $\mathcal{P}' \subseteq \mathcal{P}$

In a question-answer pair there is a felicity condition: the proposition denoted by the answer must be an element of the restricted set of propositions.

(12) In a question-answer pair Q-A where $[\![Q]\!] = \langle \mathcal{P}, \mathcal{P}' \rangle$ and $[\![A]\!] = p$,
A is felicitous iff $p \in \mathcal{P}'$.

‘Only’ has the pragmatic effect to cancel the expectation, it operates on $\mathcal{P}'$ such a way that it expands $\mathcal{P}'$ to include more possibilities.

2 Multiple focus constructions

The main idea above can be applied also to multiple foci. I will concentrate here on sentences that contain two prosodical foci. In this case there are two possible interpretations. First the *complex focus*[3] (e.g. (13a), (13b)), where we have semantically one focus: an ordered pair. Second the *real multiple foci* (e.g. (13c)), where the first focus takes scope over the second one.

(13) a. CELESTE hívta fel ZEPHIRT.
Celeste called VM Zephir.acc
‘It is the Celeste, Zephir pair of whom the first called the second.’

b. Csak CELESTE hívta fel ZEPHIRT.
only Celeste called VM Zephir.acc
‘It is the Celeste, Zephir pair of whom the first called the second.’

c. Csak CELESTE hívta fel csak ZEPHIRT.
only Celeste called VM only Zephir.acc
1. ‘Only Celeste called only Zephir.’ [the others nobody or more]
2. ‘It is the Celeste, Zephir pair of whom the first called the second.’

[3] The term is from Krifka (1991).

(13a) and (13b) get the intended interpretation according to the classical theories, the problem that arises here is the same as before, namely that 'only' and the exhaustivity operator are the same, so they do not distinguish between (13a) and (13b), but this can be solved by the above introduced pragmatic analysis of 'only' along Balogh (2005).

2.1 Exhaustification of multiple terms

In (13a) and (13b) exhaustivity applies to pairs, this is what Groenendijk and Stokhof's (1984, 1991) generalized definition of exhaustivity (6) does. In our examples there are two terms, so the interpretation goes as follows:

(14) $(EXH^2(\lambda R[R(c,z)]))(\lambda x\lambda y.called'(x,y)) =$
$\lambda R\forall x\forall y[R(x,y) \leftrightarrow [x = c \wedge y = z]](\lambda x\lambda y.called'(x,y)) =$
$\forall x\forall y[called'(x,y) \leftrightarrow [x = c \wedge y = z]]$

Krifka (1991) also gives an elegant analysis to deal with multiple focus in a compositional way. He gives a recursive definition of extended application for Focus-Background structures (15)[4] and defines the syntactic-semantic rules as follows (we give here only the ones relevant for our examples).

(15) $\alpha(\beta)$ functional application
$\langle\alpha,\beta\rangle(\gamma) = \langle\lambda X.[\alpha(X)(\gamma)],\beta\rangle$
$\gamma(\langle\alpha,\beta\rangle) = \langle\lambda X.\gamma(\alpha(X)),\beta\rangle$
$\langle\alpha,\beta\rangle(\langle\gamma,\delta\rangle) = \langle\lambda X \bullet Y.[\alpha(X)(\gamma(Y))],\beta \bullet \delta\rangle$

(16) S $\rightarrow$ NP VP; $[[_S$ NP VP$]]$ = [NP]([VP])
$\text{VP}_{tr} \rightarrow$ V NP; $[[_{VPtr}$ V NP$]] = \lambda S\lambda T\lambda x.T(\lambda y.S(x,y))([V])([NP])$
C $\rightarrow$ C_F; $[\text{C}_F] = \langle\lambda X.X,[C]\rangle$
C $\rightarrow$ FO C; $[[_C$ FO C$]] = \lambda\langle X,Y\rangle\lambda O[\lambda Z.O(\langle X,Z\rangle)(Y)]([C])([FO])$

According to this system the interpretation of (13b) is as follows:

(17) $Zephir_F$: $\langle\lambda T.T,z\rangle$
called $Zephir_F$: $\langle\lambda T\lambda x.T(\lambda y.called'(x,y)),z\rangle$
$Celeste_F$: $\langle\lambda T.T,c\rangle$
$Celeste_F$ *called* $Zephir_F$: $\langle\lambda X \bullet Y[X(\lambda x.Y(\lambda y.called'(x,y))],c \bullet z\rangle$
only $Celeste_F$ *called* $Zephir_F$:
$called'(c,z) \wedge \forall x \bullet y[x \bullet y \approx c \bullet z \wedge called'(x,y) \rightarrow x \bullet y = c \bullet z]$

This shows that both classical theories can easily deal with prosodically multiple foci that express semantically one focus, a pair. Both theories take an

[4] To make it simpler we give the rules without types in the main text, with types: $\alpha_{(\sigma)\tau}(\beta_\sigma)$ is type τ and interpreted as functional application
$\langle\alpha,\beta\rangle_{\langle(\sigma)(\tau)\mu,\sigma'\rangle}(\gamma_\tau) = \langle\lambda X_\sigma.[\alpha(X)(\gamma)],\beta\rangle_{\langle(\sigma)\mu,\sigma'\rangle}$
$\gamma_{(\sigma)\tau}(\langle\alpha,\beta\rangle_{\langle(\mu)\sigma,\mu'\rangle}) = \langle\lambda X_\mu.\gamma(\alpha(X)),\beta\rangle_{\langle(\mu)\tau,\mu'\rangle}$
$\langle\alpha,\beta\rangle_{\langle(\sigma)(\tau)\mu,\sigma'\rangle}(\langle\gamma,\delta\rangle_{\langle(\upsilon)\tau,\upsilon'\rangle}) = \langle\lambda X_\sigma \bullet Y_\upsilon.[\alpha(X)(\gamma(Y))],\beta \bullet \delta\rangle_{\langle(\sigma\bullet\upsilon)\mu,\sigma'\bullet\upsilon'\rangle}$

exhaustivity operator (practically the 'only') that applies to an orderd pair, so we get the right interpretation that it is the Celeste, Zephir pair of whom the first visited the second and there are no other pairs in the domain of which the call-relation holds.

2.2 Complex focus and two 'only's

Example (13c) has two meanings. On the first reading the first only-phrase takes scope over the second one, and on the other reading when we have again only a pair as focus. This latter case is problematic both for the Partition Semantics and for the Structured Meaning account. According to Groenendijk and Stokhof the two 'only's are the operators that exhaustify the phrases. Following this the interpretation is as follows:

(18) $(EXH(\lambda P.P(c)))((EXH(\lambda P.P(z)))(\lambda x\lambda y.called'(x,y)))=$
$(\lambda\forall y[P(y) \leftrightarrow y = c])((\lambda\forall x[P(x) \leftrightarrow x = z])(\lambda x\lambda y.called'(x,y)))=$
$\forall y[\forall x[\lambda y.called'(x,y) \leftrightarrow x = c] \leftrightarrow y = z]$

This result expresses the 'scope-reading' as it says only Mary is such that she called only Peter.

The same problem arises for the interpretation along Krifka's analysis, where the two 'only's are applied to the two focused constituents respectively. Here we also get the 'scope-reading' but not the complex focus reading.

(19) only Zephir$_F$: $\lambda P[P(z) \wedge \forall y[(y \approx z \wedge P(y)) \rightarrow y = z]$
called only Zephir$_F$: $\lambda x[called'(x,z) \wedge \forall y[y \approx z \wedge called'(x,y) \rightarrow y = z]]$
only Celeste$_F$: $\lambda P[P(c) \wedge \forall x[(x \approx c \wedge P(x)) \rightarrow x = c]]$
only Celeste$_F$ called only Zephir$_F$:
$\lambda P[P(c) \wedge \forall x[x \approx c \wedge P(x) \rightarrow x = c]](\lambda x[called'(x,z) \wedge$
$\forall y[y \approx z \wedge called'(x,y) \rightarrow y = z]])=$
$called'(c,z) \wedge \forall y[y \approx z \wedge called'(c,y) \rightarrow y = z] \wedge \forall x[x \approx c \wedge (called'(x,z) \wedge$
$\forall y[y \approx z \wedge called'(x,y) \rightarrow y = z]) \rightarrow x = c]$

A possible solution to solve the above problem is to suppose that in the case of the complex focus meaning of (13b) repeated here as (20a) semantically there is only one operator. This suggests that 'only' here is a resumptive operator, just like in the case of negative concord. This proposal is, however, not suitable because of the fact that dropping the second 'only' from the sentence does not lead to ungrammaticality but gives the same meaning (13c) repeated here as (20b).

(20) a. Csak CELESTE hívta fel csak ZEPHIRT.
'It is the Celeste, Zephir pair of whom the first called the second.'

b. Csak CELESTE hívta fel ZEPHIRT.
'It is the Celeste, Zephir pair of whom the first called the second.'

Instead, we suppose similarly to section 1.2 that ‘only’ and the exhaustivity operator are distinct. In this case there is one EXH operator that applies to the pair of the arguments, and the two ‘only’s work pragmatically saying that the fact that only Celeste called somebody and that only Zephir was called by somebody were both unlikely.

We also have to mention here the other meaning of (20a), namely the one where the first focus takes scope over the second one. In this case there are two exhaustivity operators. A relevant fact here is that this meaning differs in the information structure, because here the second focus is second occurrence, so the exhaustification of this constituent happened earlier in the discourse. At the point of the discourse where the sentence is uttered this expression comes as *old information* and happens to be in the scope of the first focus, which is the *new information*. In this way the two focused expressions are really apart, and there is no way for them to form a pair.

We saw above that there are several factors that give us the two readings: the complex focus reading and the scope-reading. In the following sections we discuss one more important factor, intonation, which seems to have a bigger role then it has been assumed before.

2.3 Intonation

It is obvious that for sentence (21) two different intonation patterns lead to two different meanings.[5]

(21) Csak CELESTE hívta fel csak ZEPHIRT. (=(13c))
only Celeste called VM only Zephir.acc

a. Csak $^{+}$Celeste $^{\sim}$hívta $^{\diagup}$fel ‖ csak $^{+}$Zephirt. ⟶ pair-focus
b. Csak $^{+}$Celeste $^{\sim}$hívta $^{\sim}$fel $^{\sim}$csak $^{+}$Zephirt. ⟶ scope-reading

This suggests that intonation has the role to yield the intended meaning, however, there is no one-to-one correspondence between the intonation patterns and the meanings. See (22) and (23):

(22) Csak CELESTE hívta fel ZEPHIRT. (=(13b))
only Celeste called VM Zephir.acc

a. Csak $^{+}$Celeste $^{\sim}$hívta $^{\diagup}$fel ‖ $^{+}$Zephirt. ⟶ pair-focus
b. Csak $^{+}$Celeste $^{\sim}$hívta $^{\sim}$fel $^{+}$Zephirt. ⟶ pair-focus (or *)

(23) CELESTE hívta fel ZEPHIRT. (=(13a))
Celeste called VM Zephir.acc

a. $^{+}$Celeste $^{\sim}$hívta $^{\diagup}$fel ‖ $^{+}$Zephirt. ⟶ pair-focus
b. $^{+}$Celeste $^{\sim}$hívta $^{\sim}$fel $^{+}$Zephirt. ⟶ pair-focus (or *)

[5] To simplify the indication of the intonation I use:
$^{+}$: pitch accent, $^{\sim}$: deaccented, $^{\diagup}$: rising intonation and ‖: little stop.

Interestingly only for the structure (21) we can get the scope-reading, for structures (22) and (23) the scope-reading is out. We cannot even ask *Who is that, who called only Zephir?* as (24a) but we can as (24b).

(24) a. *Ki hívta fel ZEPHIRT?
who called VM Zephir.acc

b. Ki hívta ZEPHIRT fel?
who called Zephir.acc VM

É. Kiss (1998a, 1998b) propose an elegant analysis of this type of constructions. She claims that F(ocus)P(hrase) (Bródy, 1990) iteration is possible, which has as a consequence that the first FP scopes over the second FP. According to this analysis, the second focused-phrase does not stay in situ, but also moves to an FP position. The verb moves to the first F-head, while it goes through the second one as well. In this way we get the following structure:

(25) a. Csak CELESTE hívta csak ZEPHIRT fel.
only Celeste called only Zephir.acc VM

'Only Celeste called only Zephir.'

b. $[_{FP}\ NP_i\ [_{F'}\ F + V_k\ [_{FP}\ NP_j\ [_{F'}\ F + t_k\ [_{VP}\ VM\ [_{V'}\ t_k\ t_i\ t_j]]]]]]$

But consider the following examples with the same structure, where we can get the complex focus reading as well with the corresponding intonation pattern.

(26) CELESTE hívta ZEPHIRT fel.
Celeste called Zephir.acc VM

a. $^{+}$Celeste $^{\diagup}$hívta $\|$ $^{+}$Zephirt $^{\sim}$fel. $\longrightarrow$ pair-focus

b. $^{+}$Celeste $^{\sim}$hívta $^{+}$Zephirt $^{\sim}$fel. $\longrightarrow$ scope

3 Conclusion and further work

In my paper I introduced the first observations of my ongoing research on focus and 'only' in Hungarian. The main aim is to point out several problems with the semantic analysis of Hungarian focus structure and 'only'. Here I concentrated on multiple foci and an interesting structure where the two prosodic foci go along with two 'only's respectively, but the sentence has also the complex focus meaning where semantically we have only one focus, an ordered pair. I discussed examples to show when the two prosodic foci take scope and when they form an ordered pair.

We need further research to give an appropriate semantic-pragmatic analysis of these examples taking into consideration the important role of syntax and intonation as well.

Further on we have to investigate the scalar meaning of 'only' (see e.g. van Rooij's (2002) analysis in terms of relevance). Scalar readings of 'only'-phrases is an important issue about the interpretation of 'only'. According to Hungarian data scalar 'only' and non-scalar 'only' behave differently in scope-relations. This suggests that 'only' has two lexical entries.

(27) Csak HÁROM FIÚ tud befogni csak ÖT CSIKÓT.
only three boys can hitch only five foals.acc
'Only three boys can hitch only five foals.'

In example (27) we can assume four possible readings: 1) the first 'only'-phrase (OP) is scalar and the second OP is exhaustive, 2) the first OP is scalar and the second OP scalar, 3) the first OP is exhaustive and the second OP is scalar, and 4) the first OP is exhaustive and the second OP is exhaustive. From these four possible readings the ones where the first 'only'-phrase gets scalar interpretation are ungrammatical. This suggests the generalization: if we have two only-phrases where the first scopes over the second, then the first one cannot be scalar, but has to be exhaustive (and distributive). A proper treatment of these facts must be left to another occasion.

References

1. Balogh, K. (2005): 'Only' and Exhaustivity in Hungarian. To appear in Proceedings of ESSLLI workshop on *Formal semantics and cross-linguistic data*, Edinburgh.
2. Bródy, M. (1990): Some remarks on the focus field in Hungarian. In *UCL Working Papers in Linguistics* 2: 201-225.
3. É. Kiss, K. (1995): Introduction. In É. Kiss, K. (ed.): *Discourse Configurational Languages*, Oxford Studies in Comparative Syntax, Oxford University Press, New York–Oxford, 3–27.
4. É. Kiss, K. (1998a): Identificational Focus Versus Information Focus. In *Language*, Vol. 74., No. 2., 245–273.
5. É. Kiss, K. (1998b): Multiple Topic, one Focus?. In *Acta Linguistica Hungarica* 45: 3-29.
6. Groenendijk, J. and M. Stokhof (1984): *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. Thesis, University of Amsterdam.
7. Groenendijk, J. and M. Stokhof (1990): Partitioning Logical Space. Annotated handout at the 2^{nd} European Summerschool on Logic, Language and Information, Leuven.
8. Krifka, M. (1991): A Compositional Semantics for Multiple Focus Constructions. In Jacobs, J. (ed): *Informationsstruktur und Grammatik, Sonderheft der Linguistischen Berichte*.
9. Krifka, M. (2004): Association with focus phrases. To appear in Molnar, V. and S. Winkler (eds.): *Architecture of Focus*, Mouton de Gruyter, Berlin.
10. Zeevat, H. (1994): Applying an Exhaustivity Operator in Update Semantics. In Kamp, H. (ed.): *Ellipsis, Tense and Questions.* DYANA-2 deliverable R2.2.b , ILLC, Amsterdam.
11. Zeevat, H. (2002): Only pragmatics. Position statement at the *One day "only"* workshop in Amsterdam.

Providing Context-Awareness Using Multi-Agent Systems in Ambient Intelligence

Oana Bucur, Philippe Beaune, Olivier Boissier

G2I/SMA Departement, Ecole N. S. des Mines de Saint Etienne
158, Cours Fauriel, 42023 Saint-Etienne, France
{bucur, beaune, boissier}@emse.fr

Abstract. Applications need facilities for recognizing, analyzing and adapting to context to provide dynamic, useful and user-centered results. In this paper, we present an ambient intelligence scenario based on the ability to adjust to context when making decisions. We propose a context-based multi-agent architecture consisting of context aware agents able to learn to discern relevant from non relevant context as well as to make appropriate decisions. This multi-agent system interacts with a context manager layer, which is able to answer context-related queries. This architecture has been tested in the development of a simple MAS for agenda management and as future work it will be enriched with more reasoning and learning capabilities and put into practice with the described ambient intelligence scenario.

1. Introduction and vision

The convergence of virtual environments, mobile communication and sensors, imposes a new paradigm: Ambient Intelligence, a vision of the future where we are surrounded by electronic devices, sensitive and reactive to people. Ambient intelligence appears poised to cause significant changes in people's everyday lives. With digital information, the ease of interaction between humans and computers can be greatly increased by broadening the interface media available and making mobile and portable communication free of wires and stationary units. The result of ambient intelligence is ultimately a more empowered computer with the benefits of added convenience, time and cost savings, and possibilities for increased safety, security, and entertainment. This technology has the potential to significantly impact business and government processes, as well as private life.

Ambient intelligence involves the convergence of several computing areas (as mentioned in [13]). The first is *ubiquitous* or *pervasive computing*. Its major contribution is the development of various ad hoc networking capabilities that exploit highly portable or else numerous, very-low-cost computing devices. The second key area is *intelligent systems* research, which provides learning algorithms and pattern matchers, speech recognition and language translators, and gesture classification and situation assessment. A third element is *context awareness*; research on this problem lets us track and position objects of all types and represent objects' interactions with their environments. Finally, an appreciation of the social interactions of objects in environments is essential.

Software Agent (SA) technology is promising in this field and thus, should have a major role in Ambient Intelligence development due to SA characteristics such as autonomy, adaptability, social behaviour and mobility. Agents should therefore provide to AmI systems the openness, the capacity to adapt and evolve trough learning and the context-awareness.

As context-aware applications are complex, recent work focuses on separating the development of such applications on different layers. In this way, the reusability of already existing layers simplifies the task for the developers and improves the modularity of the application. The typically used layers when developing a context-aware system are: sensing context, managing it and using contextual knowledge in a specific application. We will use these layers to define our objectives and approach.

2. Scenario - Leisure Activities in "Europe" Park

Marcus and Anna arrive in the "Europe" park. Their friends' child, Peter, who is in their charge for the day, accompanies them. Marcus, as well as Anna, has a PDA equipped with a personal assistant agent (*Perassis*).

The park has an informational system (*ParkInform*) able to provide a detailed map of the area (with all available services: fast-food, toilets, playgrounds, sport grounds, benches, etc.). The system can provide information about people located in these specific areas of the park: identity (if not prohibited by privacy matters), number of persons, as well as meteorological information: temperature, if it is raining/sunny/snowing, etc.

When entering the park, both agents (Marcus' and Anna's) display a list of tasks they are able to fulfill in that specific area: "surveillance", "guidance", "locate person", "search for groups". When entering a new system (area), *Perassis* agents are configured to search for other personal agents and ask what tasks they are able to solve. In this case, Anna and Marcus' agents know how to solve localization and guidance problems, using the services provided by *ParkInform*. The tasks "surveillance" and "search or groups" are unknown to them, but they have been proposed because other agents located in the park know how to fulfill them.

Anna selects: "guidance", "surveillance" and "locate persons". *Perassis* asks for details for each task, so Anna specifies that she needs: "guidance" for finding a playground for Peter and finding an available bench; "surveillance" for Peter (she introduces Peter's identification code) and "localization person": to be announced when friends are nearby. Marcus chooses "guidance" and "search groups". He needs to find groups willing to play football and then to be guided towards the field where the group is located.

Anna

Anna's *Perassis* agent is looking for playgrounds on the park's map. As it finds several, with different characteristics, it chooses the ones suitable for children under 10 years old (Peter is 9 years old) and it proposes them to Anna, by locating them on the map, specifying also the capacity of each playground and how many kids are already on them. Anna chooses one that is nearby, leaves Peter there and goes for the available bench that *Perassis* found for her.

Perassis has now to solve a problem it never encountered before: supervising somebody. By communicating with other personal assistant agents (that it locates in the park), it finds out that, from other's experiences, the agent should announce as a potentially dangerous situation when: (i) the child left the playground (*ParkInform* can provide this information, by sensing when a specified person leaves a specific area), (ii) there are less than 3 children in the playground, (iii) an unsupervised dog enters the playground, (iv) it starts raining, (v) the child stood in the sun more than one hour. In this way, Anna's *Perassis* knows when to alert Anna by using a specific sound and displaying an explanatory message on the PDA screen.

In the same time, *Perassis* detects a friend of Anna, Sarah, who enters the park area. It announces Anna and asks for permission to establish a connection with Sarah, in order for the women to meet. With Anna's permission, *Perassis* contacts Sarah (through her own personal agent, of course) and shares the location and the activity Anna is doing, so that Sarah can reach her easily. Sarah arrives near Anna and she has her 7 years old child with her. Anna proposes Sarah to let her child in the same playground where Peter is, so the two of them can spend some time together. After a while, Anna receives an alert message, that Peter is with only one other child on the playground (Sarah's child), so this situation is potentially dangerous. Anna reconfigures *Perassis* not to alert her when Peter is alone with a friend's child, and then continues her conversation with Sarah.

Marcus

In what concerns Marcus, his Perassis has also to deal with a novel problem: searching for groups that play (or desire to play) football in the park. By communicating with other personal assistant agents, it finds out that it should look for groups that are located on the football playfields, the number of persons in the groups should be an odd number and inferior to the playfield capacity and the match should have not started (or at least not longer than 5 min ago). Using these criteria, Perassis displays for Marcus a list of playgrounds where this kind of groups exists. Marcus asks also for information regarding the average age of the players and their physical condition. Perassis computes the information and memorizes the criteria as being also relevant when searching for a group. Marcus chooses one group that seems closer to his aptitudes and, using the guiding service provided by Perassis finds easily the playfield and integrates the team with fewer players.

After the match, Marcus asks Perassis to guide him towards a place in the park where he can eat. *Perassis* knows that, until now, it had to look for "fast-food" service on the map provided by *ParkInform*. The problem is that in this park, there is no "fast-food" service specified, so *Perassis* asks other agents what they look for when requesting to find an eating. One agent responds that it looks for "snack", and it suggests *Perassis* to locate a "snack" service on the map. Marcus' agent looks for it, finds several possibilities and proposes the closest one to Marcus. Problem solved!

These scenarios are based on the hypothesis that the agents agree to share at least a minimum of knowledge in order to be able to collaborate for finding solutions to different problems.

The system described in this scenario is an open system– agents come and go in/from the system at any time. It has a specific context: the park, the devices available, different areas and several services and contains heterogeneous agents (not necessarily developed by the same person and in the same manner; each agent is personalized by its user). Each agent has to solve several different problems, each problem depending heavily on contextual information

We described two types of situations:

a) An agent knowing how to solve a problem ("finding a place to eat" means "locate a fast-food service") in a context C1 ("South" park), finds itself in a new society where the context C2 ("Europe" park) is different. Other agents can help it find different solutions, adapted to the current context: in "Europe" park it has to locate a "snack" service.

b) An agent knowing how to solve a problem P1("locating a person"), needs to solve another problem P2("supervising a person"). It can, in order to acquire the knowledge needed to solve P2, ask for help from other agents in the system that have this kind of knowledge. This is what Perassis does when faced with the problem of supervising Peter or finding a group of football players for Marcus. User's feed back comes to enforce or change the knowledge the agent achieved through experience sharing with other agents.

We intend to develop this scenario by putting in practice:

- explicit context representation – we are interested in modeling this notion, so representing, managing and reasoning over/with context is the focus of our work
- a context management system able to provide: location, weather, number of persons in a place, etc.
- cooperation with other agents (sharing "context" – how to find the relevant context for a task – knowing that the number of kids in a playground is important when watching after one of them-, as well as how to use it – it is potentially dangerous to have only two children on a playground, so alert the user when this happens)
- active feed back from user (customizing the agent so that it shouldn't consider dangerous situation when the child is with a friend and not with an unknown person)
- agent's ability to adapt and to learn in order to evolve the mechanisms for adaptable behaviour

The main objective of our work is, therefore, to *propose a model for open Multi-Agent Systems in which agents adapt their behaviour to different contexts and cooperate to learn how to improve those reasoning abilities*. This model will be used for the extension of the abovementioned layered architecture with a new layer: reasoning over and with context. This layer wants to propose generic agents that are able (with a minimum of parameterization) to choose relevant context for a decision at hand and to use that context to make the appropriate decision. This will also simplify the task of context-aware applications developers, in that it will provide reasoning and learning mechanisms already included in context-based agents. In this way, developers will only need to 'personalize' the agent for the task at hand (with specific inputs and outputs, without changing the reasoning and learning methods).

3. Approach

Taking into account the architecture of a context-aware application presented in section 1., what we propose is to separate the "using context in applications" level in: level "application" and level "reasoning on/with context"(our research interest being this level of "reasoning on/with context"). The "reasoning" level will be represented by *agents* that *reason* on/with context for:

- solving the same problem in different contexts (finding a place to eat in different parks, for example)
- solving different problems in a defined context (for example, supervising a person and guiding another one to a specific location)

The major objective is therefore to develop agents with context-adaptable behavior, all this happening in an open, interoperable MAS. For this, agents will be able to reason on how to make decisions and to adapt their decision-making mechanism to context.

This work continues the Master project presented in [2] and [3]. In this project, a first proposal of an architecture of a context-aware application has been made. The idea is to separate the reasoning in the system in two levels: the "application" and the "reasoning on/with context" levels. Agents that reason on/with contexts are able to solve the same problem in different contexts or to solve different problems in the same context. In order to use context, agents should be able to represent and manage it. In order to have context-aware agents, the multi-agent system should provide them with a context management tool, generally called a Context Manager. The context manager is responsible of managing contextual information and providing agents with this information. Such a manager should be created and put at the disposal of agents, in order to enable them to be context-aware.

There are several research challenges in developing agents with context-adaptable behavior, especially in open multi-agent systems. Making intelligent context-dependent decisions means agents should be able to (1) decide what is the relevant context to take into consideration and (2) make intelligent decisions using that context. Both aspects are important and challenging: how should an agent know how to select the relevant context given a finality (e.g., a decision to make) and how should an agent know how to make decisions?

Answering these questions is particularly difficult in open multi-agent systems, where agents can enter at anytime. This work's objective is to design agents that are able to answer these questions even when entering in a new system, where they do not now how to achieve their objectives. This is the reason why the agents should be able to learn to improve their reasoning abilities. They should be able to learn either how to choose the relevant context or how to make the right decisions using that context.

In related work ([17]), the learning methods in multi-agent systems can generally be divided into: *individual* or *multi-agent learning*. In the first case an agent learns by using user's feedback and its past experience, while in the latter, it cooperates with other agents to share past experiences. The first method is important, especially in ambient intelligence scenarios, because an agent needs to be personalized, i.e., able to make decisions tailored to its user's preferences. The second method does not replace, but complements the first by allowing the agent to learn faster and even more, as it uses other agents' experiences too.

To summarize, the objective of this work is to design agents that adapt their behavior to context in open multi-agent systems used in ambient intelligence scenarios. An agent that adapts its behavior to the context should be able to *select* the relevant context for a decision and to *use* that context to make the decision. In order to use context, agents should be able to *represent* and *manage* it, by means of a Context Manager. Additionally, they should be able to improve their performances by *learning* how to select and use the context, especially in open multi-agent systems. The results of this work should be validated using one or several ambient intelligence scenarios. In what follows, we give a brief overview of the work we have already done in this field and how we plan to develop and use the obtained results in our future work.

4. Definitions

Agents are situated entities, meaning that they sense their environment and act accordingly. There are many definitions of context given in the literature ([1],[4], [8], [9], [14], [16], etc.). All these definitions have in common the fact that context may be further described as a set of attributes relevant for a finality. The *finality*, **f**, is the goal for which the context is used at a given moment (e.g. to decide whether a proposal for an appointment should be accepted or not, to determine whether the current situation is similar to another one or not, to understand a conversation, etc.). We can see the finality as being the information that is the most interesting for the agent at a given moment, the "fundamental fact, action, or belief" (cf. Webster's Dictionary) that determines the way it will act. Let's note $\boldsymbol{F}$ the set of finalities.

An entity is an instance of a "person, object or place" (as mentioned in [4]), but also an activity, an organizational concept (role, group), etc.

A *context attribute* (***a***) designates the information defining context, e.g. "Activity-Location", "NamePerson", "ActivityDuration". It is obvious that each context attribute has at least one value at a given moment, the value depending on several entities to which the attribute relates. For instance, the context attribute "NamePerson" defines the name of a person. When trying to instantiate this attribute, the needed parameter will be the specific person whose name interests us. The "PersonIsMemberOf" context attribute will also take a person as input, but will return (possibly) several groups in which that person takes part. Let's note V_a the definition domain of a, the set of possible values that a may take (example: V_{time} =[0,24[ ). We can therefore associate to each context attribute an instantiation function called ***valueOf***. We define *valueOf* for a context attribute as a function from A x P_a to $P(V_a)$, where **A** is the set of all attributes, $\boldsymbol{P(V_a)}$ is the power set of $\boldsymbol{V_a}$, and $\boldsymbol{P_a}$ is the set of parameters needed to compute the value of $\boldsymbol{a}$.

Not all attributes are relevant for a finality. We define ***is_relevant(a,f),*** a predicate stating that attribute $\boldsymbol{a}$ is relevant for the finality $\boldsymbol{f}$. Let's call ***RAS***(*f*) the *Relevant Attribute Set* for the finality $\boldsymbol{f}$: ***RAS***(*f*) = { $a \in \boldsymbol{A}$ | ***is_relevant***(*a,f*)=true }.

We will note an *instantiation of context attribute* $a \in \boldsymbol{A}$ as a pair *(a,v)* where *v* is the set of values $v \in \boldsymbol{P}(\boldsymbol{V_a})$ of *a* at a given moment. For instance, (Day, {14}), (roleOfPersonInGroup, {Team Manager}), (PersonIsMemberOf, {MAS Group, Center_X, University_Y}) are instantiation of respective context attributes *Day*, *roleOfPersonInGroup*, *PersonIsMemberOf.* Let's note $\boldsymbol{I}$ the set of instantiated context attributes as $\boldsymbol{I}$ = {(*a,v*) | $a \in \boldsymbol{A}$ ∧ ***valueOf***(*a*)=*v*}. We call *Instantiated Relevant Attribute Set* of a finality ***f,*** noted ***IRAS*(f)**, the set of instantiated context attributes relevant to finality *f*: ***IRAS***(f) = {(*a,v*) | $a \in$***RAS***(f) ∧ (*a,v*) ∈ $\boldsymbol{I}$}.

5. Proposed architecture for a context-based MAS

The proposed MAS architecture is composed of several learning agents (in our case study, we call them MySAM, because we validated our approach in a MAS for agenda management called MySAM - My Smart Agenda Manager), each of them being the personal assistant of an user. The agents can interact between them and can connect to a Context Manager that provides them with context knowledge by having access to the current state of the environment (cf. Figure 1).

To make contextualized decisions, every agent needs to know the values of the relevant context attributes in a situation. It acquires them through the Context Manager whose main functionalities are to let the agents know which are the context attributes that it can manage and to compute the values to instantiate the context attributes that are relevant for an agent at some point.

a. Context Manager (CM)

The main functionalities of CM are to let the agents know which is the context attributes set (defined in the ontology) that it manages and to compute IRAS corresponding to RAS given by the agents at some point of processing. When entering a society, an agent asks the corresponding CM to provide it with the context attributes that it manages. Acting as intermediary between agents and the environment, CM is able to answer requests regarding its managed context attributes.

The Context Knowledge Base contains the ontology of the domain and all context attributes that will be managed by the CM, on the other hand. The *instantiation* module computes the IRAS(f) for a given RAS(f). The *dependencies* module computes the values for derived attributes by considering possible relations between context attributes concerning their relevance: if one attribute is relevant for a situation and it has a certain value, then another attribute could also be relevant for that situation.

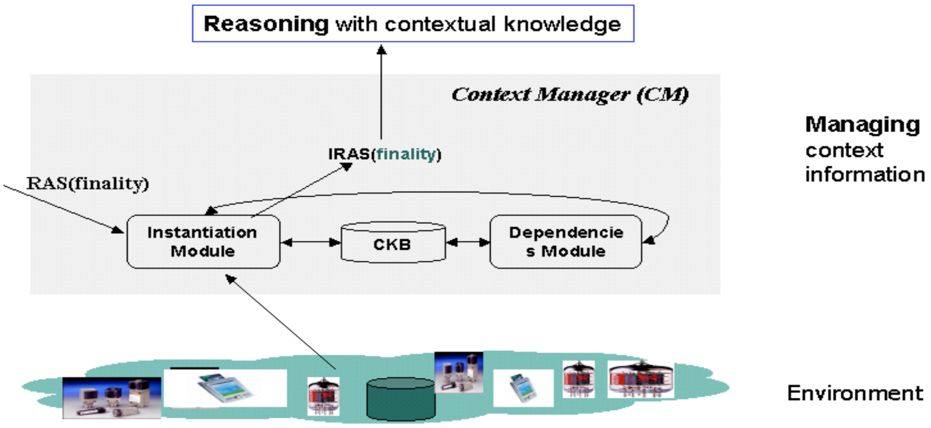

Fig. 1. Context manager architecture

b. Context-based learning agent

The context-based part of an agent is constituted of two main components: *selection* of relevant attributes of context for a certain purpose (the selection module) and *exploitation* of context knowledge to make decisions (the module of decision) (cf. Fig.2). Even if we named the agent mySAM in reference to the application which we used as a modeling scenario, this architecture of a context-based learning agent is general and it is not restrained to the frame of application considered to illustrate this approach. Although mySAM has some negotiation modules (in order to establish meetings), we present here only the part of the agent that is relevant for the management and reasoning on context.

In case of mySAM, the *selection* module provides the relevant attributes for a finality (RAS for this purpose). So, for instance, to decide if they should accept or not a 'two persons meeting', the selection module is going to construct an RAS made of {the things already planned, the role of the person with whom they negotiate}; for a finality of type 'seminar' meeting, the RAS will be {the participants, the seminar's theme, participants' research fields, …}.

The *decision* module decides whether to accept or not a meeting taking into account the values computed for relevant context attributes (the context knowledge). This module is therefore going to know how to accept a meeting if we have nothing planned for that period of time and if the person that demands this meeting is our chief, for instance. It will also know that it should not accept otherwise. The purpose, the situation that we want to solve is going to help the agent choose relevant context. It is therefore the context knowledge set (IRAS) which is helping the agent to decide how to act in this situation. In the case of MySAM, we implemented a multi-agent cooperation method (similar to what was proposed in [12] and [17]) so that agents can share their experiences in choosing the right context for a situation and then using the resulting IRAS to decide what to do (for details see [2], [3]).

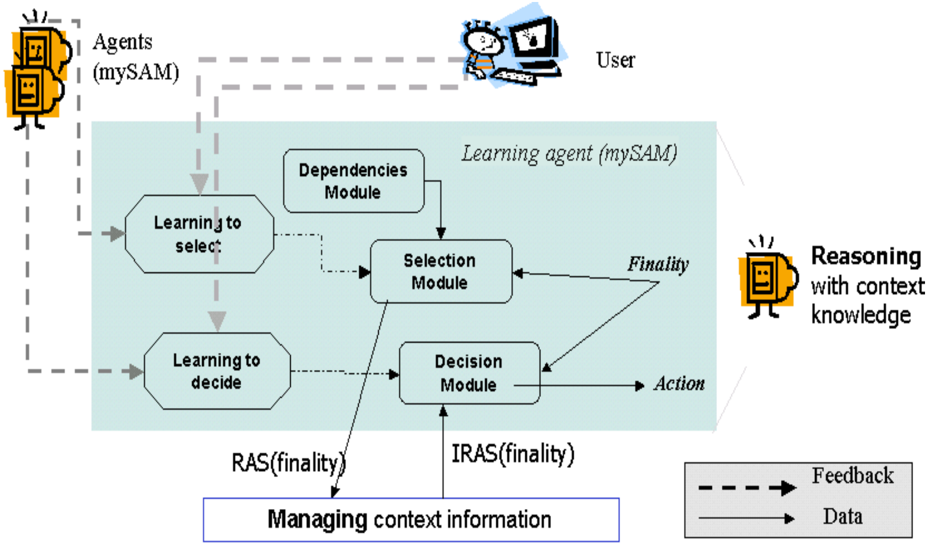

Fig. 2. Context-based agent architecture

6. Conclusions and future work

In this article, we have presented a scenario of ambient intelligence in which dealing with context (choosing the right context elements that are relevant for a situation and using context knowledge in making decisions) is the main issue. We proposed a definition of the notion of context, notion that is used in almost all systems without precisely and explicitly taking it into account. We have proposed a context-based architecture for a learning multi-agent system based on an ontology-based representation for context elements.

Our definition and architecture are not based on a specific application. The difference between our approach and the classical approach when using context in applications, is that we extended the typical architecture by adding a selection module and learning modules for the decision module (the usual approach when designing adaptable agents for decision-making- as in [5], [6], [7], [10], [11]) and for the selection module (so that the agent will improve its ability to select relevant context for a given finality). The selection module permits not to consider relevant context known in advance – which would make our system application-dependent.

Our future work we will focus on taking into account a mode complex Decision Support System, in order to highlight the use of context in each of the phases of a decision making process. Until now we focused on intelligence and choice phase to test how context can be used. Further more, we considered these phases as being atomic, and not decomposable into several sub-phases (sub-phases described in **[15]**). We plan to extend the use of the notion of context also for the design phase and to consider each step of the four phases of decision making. In what concerns learning, the framework will provide agents equipped with several individual learning algorithms and all needed modules to share contextual knowledge.

References

[1] Brezillon, P. – "Context Dynamic and Explanation in Contextual Graphs", In: Modeling and Using Context (CONTEXT-03), LNAI 2680, Springer Verlag p. 94-106, 2003
[2] Bucur O, Boissier O, Beaune P – "A Context-Based Agent Architecture for Learning How to Make Contextualized Decisions" (to appear in Proc. of MCMP 05), may 2005
[3] Bucur O – "Modeling context for a multi-agent learning of user preferences" – master dissertation – Ecole des Mines de Saint-Etienne, July 2004
[4] Dey, A., Abowd, G.– "Towards a better understanding of Context and Context-Awareness", GVU Technical Report GIT-GVU-00-18, GIT, 1999
[5] Lashkari Y., Metral M., Maes P – "Collaborative Interface Agents", Proc. of the Third International Conference on Information and Knowledge Management CIKM'94, ACM Press, 1994
[6] Lin S., J.Y.Hsu – "Learning User's Scheduling Criteria in a Personal Calendar Agent", Proc. of TAAI2000, Taipei.
[7] Mitchell T., Caruana R., Freitag D., McDermott J., Zabowski D.– "Experience with a learning personal assistant", Communications of the ACM, 1994.
[8] Pascoe J. – "Adding Generic Contextual Capabilities to Werable Computers". 2nd International Symposium on Wearable Computers (1998) 92-99.
[9] Persson P.– "Social Ubiquitous computing", Position paper to the workshop on 'Building the Ubiquitous Computing User Experience' at ACM/SIGCHI'01, Seattle.
[10] Scerri, P., Pynadath D., Tambe M.– "Why the elf acted autonomously: Towards a theory of adjustable autonomy " , First Autonomous Agents and Multi-agent Systems Conference (AAMAS), p. 857-964, 2002.
[11] Sen S., E.H. Durfee – "On the design of an adaptive meeting scheduler", in Proc. of the Tenth IEEE Conference on AI Applications, p. 40-46, 1994.
[12] Sian S. S. – "Adaptation Based on Cooperative Learning in Multi-Agent Systems", Descentralized AI, Yves Demazeau & J.P. Muller, p. 257-272, 1991.
[13] Shadbolt, N. – "Ambient Intelligence" – IEEE Intelligent Systems January 2003, p.2-3.
[14] Thevenin D., J. Coutaz. – "Plasticity of User Interfaces: Framework and Research Agenda". In Proceedings of INTERACT'99, 1999, pp. 110-117.
[15] Turban E., Aronson J.E. – "Decision Support Systems and Intelligent Systems", 6^{th} edition, Prentice Hall, 2001, pp. 33-121.
[16] Turner, R. – "Context-Mediated Behaviour for Intelligent Agents", International Journal of Human-Computer Studies, vol. 48 no.3, March 1998, p. 307-330.
[17] Weiss G., Dillenbourg P.– "What is "multi" in multi-agent learning?", P. Dillenbourg (Ed) Collaborative-learning: Cognitive, and computational approaches, p. 64-80, 1999.

Multi-context specification for Graded BDI-Agents

Ana Casali[1], Lluís Godo[2] and Carles Sierra[2]

[1] PhD Program in Information Technologies
EIA - University of Girona, Spain
UNR - Av Pellegrini 250, 2000 Rosario, Argentina.

[2] Institut d'Investigació en Intel·ligència Artificial (IIIA) - CSIC
08193 Bellaterra, Catalunya, España.

Abstract. In the recent past, an increasing number of multiagent systems have been designed to engineer complex distributed systems. Among the proposed architectures, one of the most widely used is the BDI agent presented by Rao and Georgeff. We consider that in order to apply agents in real domains, it is important for the formal models to incorporate uncertainty representation. With that aim we introduce in this work a general model for graded BDI agents, based on multi-context systems. The multicontext approach provides an adequate platform to specified agents, allowing a suitable representation of graded mental attitudes, and their interactions. The architecture proposed serves as a blueprint to design different kinds of particular agents.

1 Introduction

In the recent past, an increasing number of multiagent systems (MAS) have been designed and implemented to engineer complex distributed systems. Several theories and architectures have been proposed to give these systems a formal support. A well-known intentional system formal approach is the BDI architecture proposed by Rao and Georgeff [13]. This model is based on the explicit representation of the agent's beliefs (B), its desires (D), and its intentions (I). This architecture has evolved over time and it has been applied in several of the most significant multiagent applications developed up to now.

Modeling different intentional notions by means of several modalities (B, D, I) can be very complex if only one logical framework is used. In order to help in the design of complex logical systems Giunchiglia et.al. [6] introduced the notion of *multi-context system* (MCS). This framework allows the definition of different formal components and their interrelation. In our case, we propose to use separate contexts to represent each modality and formalize each context with the most appropriate logic apparatus. The interactions between the components are specified by using inter-unit rules, called *bridge rules*. Contexts have been used in diverse applications and particularly, have been used to model agent's architectures [11, 14], as a framework where the different components of the architecture and their interactions, can be neatly represented. Indeed one advantage of the

MCS logical approach to agency modeling is that it allows for rather affordable computational implementation. For instance, a portion of the framework described in [11] has been recently implemented using [5].

The agent architectures proposed so far mostly deal with two-valued information. Although the BDI model developed by Rao and Georgeff explicitly acknowledges that an agent's model of the world is incomplete, it makes no use of quantified information about how possible a particular world is to be the actual one. Neither does it allow desires and intentions to be quantified. There are a few works that partially address this issue and emphasize the importance of graded models. Notably, Parsons and Giorgini [12] consider the belief quantification by using Evidence Theory. They set out the importance of quantifying degrees in desires and intentions, but this is not covered by their work.

Our Thesis Project is about the development of a General Multi-context Model for Graded BDI Agents, specifying an architecture able to deal with the environment uncertainty and with graded mental attitudes. In this sense, belief degrees represent to what extent the agent believes a formula is true. Degrees of positive or negative desire allow the agent to set different levels of preference or rejection respectively. Intention degrees give also a preference measure but, in this case, modeling the cost/benefit trade off of reaching an agent's goal. Then, Agents having different kinds of behavior can be modeled on the basis of the representation and interaction of these three attitudes. We consider the individual aspects of agency, defining clear semantics for the different mental attitudes and suitable logics to represent each one. These are formalized in the respective contexts for beliefs (BC), desires (DC), and intentions (IC), and two functional contexts for planning (PC) and communication (CC). Furthermore, considering the social aspects of agency, a social context (SC) is introduced in our agent model, where the trust in other agents is modeled. These contexts are briefly schematized in the next Sections. We complete the agent architecture presenting in Section 8, a set of bridge rules. Preliminary results and an example of a Tourist Assistant Agent, illustrating the overall reasoning process of our graded model, can be found in [2]. Finally, in Section 9 we present some conclusions and future lines of work.

2 Graded BDI agent model

The architecture proposed is inspired by the work of Parsons et.al. [11] about multi-context BDI agents. Multi-context systems were introduced by Giunchiglia et.al. [6] to allow different formal components to be defined and interrelated. The MCS specification of an agent contains three basic components: units or contexts, logics, and bridge rules, which channel the propagation of consequences among theories. Thus, an agent is defined as a group of interconnected units: $\left\langle \{C_i\}_{i\in I}, \Delta_{br} \right\rangle$, where each context $C_i \in \{C_i\}_{i\in I}$ is the tuple $C_i = \langle L_i, A_i, \Delta_i \rangle$ where L_i, A_i and Δ_i are the language, axioms, and inference rules respectively. They define the logic for the context and its basic behavior is constrained by the axioms. When a theory $T_i \in L_i$ is associated with each unit, the implementation of a particular agent is complete. Δ_{br} can be understood as rules of inference with premises and conclusions in different contexts. The deduction mechanism of these systems is based on two kinds of inference rules, internal rules Δ_i, and bridge rules Δ_{br}. Internal rules allow to draw consequences within a theory, while bridge rules allow to embed results from a theory into another [4]. We have *mental* contexts to represent beliefs (BC), desires (DC), intentions (IC), and a social context (SC) which represents the trust in other agents. We also consider two *functional* contexts: for Planning (PC) and Communication (CC). In summary, the BDI agent model is defined as: $A_g = (\{BC, DC, IC, SC, PC, CC\}, \Delta_{br})$.

This multicontext specification of our model of agent allow us to set a clear and independent semantics for each mental attitude. We do not intent to define a global semantics for our multicontext graded BDI agent. This approach let us to represent separately the different graded mental attitudes, considering that they must be treated with a suitable logic in each case. The overall behavior of the system will be result of the logic representation of each intentional notion in the different contexts and the bridge rules. These rules, constitute an operational (hence extra-logical) part of the system behavior and permit us to modify the different theories.

In order to represent and reason about graded notions of beliefs, desires and intentions, we decide to use a modal many-valued approach. In particular, we shall follow the approach developed by Hájek et al. [8, 7] where uncertainty reasoning is dealt with by defining suitable modal theories over suitable many-valued logics. For instance, let us consider a Belief context where belief degrees are to be modeled as probabilities. Then, for each classical formula φ, we consider a modal formula $B\varphi$ which is interpreted as "φ is probable". This modal formula $B\varphi$ is then a *fuzzy* formula which may be more or less true, depending on the probability of φ. In particular, we can take as truth-value of $B\varphi$ precisely the probability of φ. Moreover, using a many-valued logic, we can express the governing axioms of probability theory as logical axioms involving modal formulae. Then, the many-valued logic machinery can be used to reason about the modal formulae $B\varphi$, which faithfully respect the uncertainty model chosen to represent the degrees of belief. In this proposal, we choose the Łukasiewicz infinite-valued logic but another selection of logics may be done for each unit.

3 Belief Context

The purpose of this context is to model the agent's beliefs about the environment. In order to represent beliefs, we use modal many-valued formulae, following the above mentioned logical framework. We consider the probability theory as the uncertainty model. Other models might be used as well by just modifying the corresponding axioms.

To reason about the credibility of crisp propositions, we define the BC language for belief representation, following Godo et al.'s [7], based on Łukasiewicz logic. In order to define the basic crisp language, we extend a propositional language L to represent actions, taking advantage of Dynamic logic [10]. These actions, the environment transformations they cause, and their associated cost must be part of any situated agent's beliefs set. The propositional language L is thus extended to L_D, by adding to it action modalities of the form $[\alpha]$ where α is an action. More concretely, given a set Π_0 of symbols representing elementary actions, it can be defined the set Π of plans (composite actions). Then L_D is defined in the usual way [2], if $\psi \in L$ then $\psi \in L_D$ and if $\alpha \in \Pi$ and $\varphi \in L_D$ then $[\alpha]\varphi \in L_D$. The interpretation of $[\alpha]\, A$ is *"after the execution of α, A is true"*. We define a modal language BC over the language L_D to reason about the belief on crisp propositions. To do so, we extend the crisp language L_D with a fuzzy unary modal operator B. If $\varphi \in L_D$, the intended meaning of $B\varphi$ is that "φ is believable". Formulae of BC are of two types:

- *Crisp (non B-modal):* they are the (crisp) formulae of L_D.
- *B-Modal:* they are built from elementary modal formulae $B\varphi$, and truth constants, using the connectives of Łukasiewicz many-valued logic:

• If $\varphi \in L_D$ then $B\varphi \in BC$

• If $r \in Q \cap [0,1]$ then $\overline{r} \in BC$

• If $\Phi, \Psi \in BC$ then $\Phi \rightarrow_L \Psi \in BC$ and $\Phi \& \Psi \in BC$ (where & and $\rightarrow_L$ correspond to the conjunction and implication of Łukasiewicz logic)

Other Łukasiewicz logic connectives can be defined from &, $\rightarrow_L$ and $\overline{0}$, for example $\neg_L \Phi$ is defined as $\Phi \rightarrow_L \overline{0}$. Since in Łukasiewicz logic a formula $\Phi \rightarrow_L \Psi$ is 1-true iff the truth value of Ψ is greater or equal to that of Φ, modal formulae of the type $\overline{r} \rightarrow_L B\varphi$ express that the probability of φ is at least r and will be denoted as $(B\varphi, r)$.

Belief Semantics and Axiomatization

The semantics for the language BC is defined, as usual in modal logics, using a Kripke structure. We have added to such structure a ρ function in order to represent the world transitions caused by actions, and a probability measure μ over worlds. Thus, we define the Kripke structure $M_B = \langle W, e, \mu, \rho \rangle$ where:

- W is a non-empty set of possible worlds.
- $e : V \times W \rightarrow \{0,1\}$ provides for each world $w \in W$ a Boolean evaluation of each propositional variable $p \in V$, that is, $e(p,w) \in \{0,1\}$.
- $\mu : 2^W \rightarrow [0,1]$ is a finitely additive probability measure such that for each crisp φ, the set $\{w \mid e(\varphi, w) = 1\}$ is measurable [8].
- $\rho : \Pi_0 \rightarrow 2^{W \times W}$ assigns to each elementary action a set of pairs of worlds denoting world transitions.

Extension of e to L_D formulae: e is extended to L using classical connectives and to formulae with action modalities –as $[\alpha]\, A$, setting: $e([\alpha]\, A, w) = min\,\{e(A, w_i) \mid (w, w_i) \in \rho(\alpha)\}$.

Extension of e to B-modal formulae: e is extended by means of Łukasiewicz logic truth-functions and the probabilistic interpretation of belief as follows:

- $e(B\varphi, w) = \mu(\{w' \in W \mid e(\varphi, w') = 1\})$, for each crisp φ
- $e(\overline{r}, w) = r$, for all $r \in Q \cap [0,1]$
- $e(\Phi \& \Psi, w) = \max(e(\Phi) + e(\Psi) - 1, 0)$
- $e(\Phi \rightarrow_L \Psi, w) = \min(1 - e(\Phi) + e(\Psi), 1)$

Finally, the truth degree of a formula Φ in a Kripke structure $M = \langle W, e, \mu, \rho \rangle$ is defined as $\|\Phi\|^M = \inf_{w \in W} e(\Phi, w)$. Notice, that a theory T in this context may include formulae like $(B[\alpha]\varphi, r)$. According to the belief semantics, this formula means that the probability that the actual world is w' where $[\alpha]\varphi$ is true, is at least r. This probability captures in some sense, the probability of failure of plan α. As mentioned in Section 2, to set up an adequate axiomatization for our belief context logic we need to combine axioms for the crisp formulae, axioms of Łukasiewicz logic for modal formulae, and additional axioms for B-modal formulae according to the probabilistic semantics of the B operator. Hence, axioms and rules for the Belief context logic BC are as follows:

1. Axioms of propositional Dynamic logic for L_D formulae (see e.g. [10]).
2. Axioms of Łukasiewicz logic for modal formulae (see e.g. [8]).
3. Probabilistic axioms
   $B(\varphi \to \psi) \to_L (B\varphi \to B\psi)$
   $B\varphi \equiv \neg_L B(\varphi \wedge \neg\psi) \to_L B(\varphi \wedge \psi)$
   $\neg_L B\varphi \equiv B\neg\varphi$
4. Deduction rules for BC are: modus ponens, necessitation for $[\alpha]$ for each $\alpha \in \Pi$ (from φ derive $[\alpha]\varphi$), and necessitation for B (from φ derive $B\varphi$).

Actually, one can show that the above axiomatics is sound and complete with respect to the intended semantics described in the previous subsection (cf. [8]).

4 Desire Context

In this context, we represent the agent's desires. Desires represent the *ideal* agent's preferences regardless of the agent's current perception of the environment and regardless of the cost involved in actually achieving them. We deem important to distinguish what is positively desired from what is not rejected. According to the works on bipolarity representation of preferences by Benferhat et.al. [1], positive and negative information may be modeled in the framework of possibilistic logic. Inspired by this work, we suggest to formalize agent's desires also as positive and negative. Positive desires represent what the agent would like to be the case. Negative desires correspond to what the agent rejects or does not want to occur. Both, positive and negative desires can be graded. As for the BC language, the language DC is defined as an extension of a propositional language L by introducing two (fuzzy) modal operators D^+ and D^-. $D^+\varphi$ reads as "φ is positively desired" and its truth degree represents the agent's level of satisfaction would φ become true. $D^-\varphi$ reads as "φ is negatively desired" and its truth degree represents the agent's measure of disgust on φ becoming true. As in BC logic, we will use a modal many-valued logic to formalise graded desires with Łukasiewicz logic as the base many-valued logic. In this context the agent's preferences will be expressed by a theory T containing quantitative expressions about positive and negative preferences, like $(D^+\varphi, \alpha)$ or $(D^-\psi, \beta)$, as well as qualitative expressions like $D^+\psi \to_L D^+\varphi$ (resp. $D^-\psi \to_L D^-\varphi$), expressing that φ is at least as preferred (resp. rejected) as ψ.

A complete formalization of semantics, in terms of generalized possibilistic Kripke structures, and a correct axiomatics can be found in [2].

5 Intention Context

In this context, we represent the agent's intentions. They, as well as desires, represent the agent's preferences. However, we consider that intentions cannot depend just on the benefit of reaching a goal φ, but also on the world's state w and the cost of transforming it, into a world w' where the formula φ is true. By allowing degrees in intentions we represent a measure of the cost/benefit relation involved in the agent's actions towards the goal. The positive and negative desires are used as pro-active and restrictive tools respectively, in order to set intentions. Note that intentions depend on the agent's knowledge about the world, which may allow –or not– the agent to set a plan to change the world into a desired one. We present two kinds of graded intentions, intention of a formula φ considering the execution of a particularly plan α, noted $I_\alpha\varphi$, and the final intention to φ, noted $I\varphi$, which take into account the best path to reach φ. Thus, if the agent's IC theory T contains the formula $I_\alpha\varphi \rightarrow_L I_\beta\varphi$ then the agent will try φ executing the plan β before than executing plan α. On the other hand, if T has the formula $I\psi \rightarrow_L I\varphi$ then the agent will try φ before ψ

We define the IC Language in the similar way as we did with BC. Let L denote the basic propositional language and Π the set of actions corresponding to the dynamic propositional language L_D. The intention to make φ true must be the consequence of finding a feasible plan α, that permits to achieve a state of the world where φ holds. Then, for each $\alpha \in \Pi$ we introduce a modal operator I_α, and a modal operator I, in the same way as we did in DC. $I\varphi$ will represent that the agent intends φ by means of the "best plan" known. We use Łukasiewicz multivalued logic to represent the degree of the intentions. As in the other contexts, if the degree of $I_\alpha\varphi$ is δ, it may be considered that the truth degree of the expression "the goal φ is intended by means of action α" is δ. If the degree of $I\varphi$ is γ, it may be considered that the truth degree of the expression "the goal φ is intended" is γ. The computation of the degree of $I_\alpha\varphi$ for each feasible plan α is left to a suitable bridge rule (see (2) in next Section 8).

Semantics and axiomatization for IC

The semantics defined in this context shows that the value of the intentions depends on the formula intended to bring about and on the benefit the agent gets with it. It also depends on the agent's knowledge on possible plans that may change the world into one where the goal is true, and their associated cost. The models for IC are Kripke structures $M_I = \langle W, e, \rho, u\rangle$ where W, e and ρ are defined as in BC, and $u : \Pi \rightarrow [0,1]^{W\times W}$ assigns to each action α a utility distribution $u_\alpha : W \times W \rightarrow [0,1]$ such that: $u_\alpha(w', w) = 0$, if $(w', w) \notin \rho(\alpha)$ and if $\alpha' = \langle\beta;\alpha;\gamma\rangle$ then $u_{\alpha'} \leq u_\alpha$. $u_\alpha(w', w)$ can be understood as the utility of reaching state w from state w', by means of action α. Then, the upper and lower bounds for the range of utilities of reaching φ from w' by means of α are:

- $U^+(w', I_\alpha\varphi) = \sup\{u_\alpha(w', w) : e(w, \varphi) = 1\}$
- $U^-(w', I_\alpha\varphi) = \inf\{u_\alpha(w', w) : e(w, \varphi) = 1\}$

Then, the truth evaluation of the $I_\alpha\varphi$ at w' can be in principle any value in the interval $[U^-(w', I_\alpha\varphi), U^+(w', I_\alpha\varphi)]$. There are in such a case extreme alternatives: a pessimistic attitude would correspond to take $e(w', I_\alpha\varphi) = U^-(w', I_\alpha\varphi)$, an optimistic attitude would be to take $e(w', I_\alpha\varphi) = U^+(w', I_\alpha\varphi)$. Notice that $U^+(w', .)$ and $U^-(w', .)$ are a possibility and a guaranteed possibility measures on formulas, respectively, so they can be axiomatized in Lukasiewicz logic [8].

6 Social Context

We introduce a Social Context (SC) to deal with the social aspects of agency. This context has the purpose of filtering the agent's information interchange. The incoming information must be analyzed and filtered depending on the trust that the agent has in its source. To equip an agent with the social aspects, it is important to model and support the agent's trust. In an agent community different kinds of trust are needed and should be modeled. In a first stage, the purpose of the SC in our model of agent, is to filter all the information coming from other agents. We have inspired our work in the Belief, Inform and Trust (BIT) logic presented by Liau [9]. One of the central ideas formalized in BIT logic is that if $agent_i$ is informed by $agent_j$ about φ, the $agent_i$'s beliefs about φ depends on the trust the $agent_i$ has in $agent_j$ with respect to φ. In the framework of this logic all the formulae are crisp.

Assuming we have a multiagent system scenario with a finite set of agents: $\{agent_i\}$, $i \in A$, the language for this context is a basic language L extended by a family of modal operators T_{ij}, where $i, j \in A$. We consider the trust of an $agent_i$ towards an $agent_j$ about a formula φ, $T_{ij}\varphi$, may be graded taking values in [0,1], to express different levels of trust. A belief-based degree of trust has been discussed in [3]. As in the other contexts, we used a many-valued treatment for the trust of an agent towards others. Then, if the degree of $T_{ij}\varphi$ is τ, we shall consider that the truth degree of the sentence "$agent_i$ trusts in $agent_j$ about φ" is τ. We again chose the Łukasiewicz logic as the underlying many-valued logic.

The models for SC are defined in a similar way as we did in the other contexts using a Kripke structure. As for the modal formulae, we follow the intuition that the trust of $\varphi \wedge \psi$ may be taken as the minimum of the trusts in φ and in ψ, hence we interpret the trust operator T_{ij} as a necessity measure on non-modal formulas. Then, the corresponding axiomatics is set in a similar way than in IC for the pessimistic attitude. In a multiagent system scenario, if the $agent_i$ is informed by $agent_j$ that φ is true, then the statement $N_{ij}\varphi$ will be a crisp formula in the Communication context (outlined in next Section 7). One of the central axioms for trust in the BIT logic [9] is: $(B_i N_{ij}\varphi \wedge T_{ij}\varphi) \rightarrow B_i\varphi$, where $i, j \in A$. We present a multi-context version of this axiom. As belief, inform and trust formulae are represented in different contexts, we use a bridge rule (see (5) in section 8) to formalize it, and we extend this rule to a many-valued framework.

7 Planner and Communication Contexts

The nature of these contexts is functional and they are needed components of our model. In this work we only draft their funtionalities in relation with the mental contexts presented. There is much work to do with respect Planner and Communication contexts, but is out of the scope of this work and they are left as black boxes with some imput-output functionalities. The Planner Context (PC) has to build plans which allow the agent to move from its current world to another, where a given formula is satisfied. This change will indeed have an associated cost according to the actions involved. Within this context, we propose to use a first order language restricted to Horn clauses (PL), where a theory of planning includes at least the following special predicates:

- *action(*α*, P, A,* c_α*)* where $\alpha \in \Pi_0$ is an elementary action, $P \subset PL$ is the set of preconditions; $A \subset PL$ are the postconditions and $c_\alpha \in [0, 1]$ is the normalised cost of the action.

- *plan(*φ*,* α*, P, A,* c_α*)* where $\alpha \in \Pi$ is a composite action representing the plan to achieve φ, P are the pre-conditions of α, A are the post-conditions $\varphi \in A$, c_α is the normalized cost of α.

- *bestplan(*φ*,* α*, P, A,* c_α*)* similar to the previous one, but only one instance with the best plan is generated.

Each plan must be feasible, that is, the current state of the world must satisfy the preconditions, the plan must make true the positive desire the plan is built for, and cannot have any negative desire as post-condition. These feasible plans are deduced by a bridge rule (see (1) in the next Section 8).

The Communication unit (CC) makes possible to encapsulate the agent's internal structure by having a unique and well-defined interface with the environment. This unit has a propositional language with the modality N_{ij}, where $N_{ij}\varphi$ represents "$agent_i$ is informed by $agent_j$ about φ". The theory inside this context will take care of the sending and receiving of messages to and from other agents in the Multi Aagent society where our graded BDI agent lives.

8 Bridge Rules

For our BDI agent model, we define a collection of basic bridge rules to set the interrelations between contexts. In this Section we comment the most relevant rules of an $agent_i$[3].

From the positive and negative desires –represented by $\nabla(D^+\varphi)$ and $(D^-\psi, \mathit{threshold})$ respectively, the beliefs of the agent, and the possible transformations using actions, the Planner can build feasible plans. Furthermore, a filter is used to select the plans with a belief degree of achieving the goal after its execution, greater than some b-threshold. The following bridge rule does this:

$$\frac{\begin{array}{c} DC : \nabla(D^+\varphi), DC : (D^-\psi, \mathit{threshold}), PC : \mathit{action}(\alpha, P, A, c), \\ BC : (B([\alpha]\varphi), \mathit{bthreshold}), BC : B(A \to \neg\psi) \end{array}}{PC : \mathit{plan}(\varphi, \alpha, P, A, c)} \quad (1)$$

[3] We only explicit the i subscript in bridge rule (5) where another agent takes place

The intention degree trades off the benefit and the cost of reaching a goal. There is a bridge rule that infers the degree of $I_\alpha\varphi$ for each feasible plan α that allows to achieve the goal. This value is deduced from the degree of $D^+\varphi$ and the cost of the plan α. This degree is calculated by function f as follows:

$$\frac{DC : (D^+\varphi, d), PC : plan(\varphi, \alpha, P, A, c)}{IC : (I_\alpha\varphi, f(d, c))} \tag{2}$$

Different functions model different individual behaviors. For example, the function might be defined as $f(d, c) = (d+(1-c))/2$, if we consider an *equilibrated agent*. Then, $(I\varphi, i)$ will be computed where i is the maximum degree of all the $I_\alpha\varphi$, where α is a feasible plan for φ. The plan α_b that allows to get the maximum intention degree to φ –i–, will be set as the *best plan* in the Planner.

We also need bridge rules to establish the agent's interactions with the environment, meaning that if the agent intends φ at degree i_{max}, the maximum degree of all the intentions, then the agent will focus on the plan α_b –*bestplan*– that allows the agent to reach the most intended goal φ:

$$\frac{IC : (I\varphi, i_{max}), PC : bestplan(\varphi, \alpha_b, P, A, c_\alpha)}{CC : C(does(\alpha_b))} \tag{3}$$

Through the communication unit the agent perceives all the changes in the environment and particularly receives the information from other agents. This information is introduced in the belief context, using a suitable order preserving transformation h, by the following bridge rule :

$$\frac{CC_i : N_{ij}\varphi, SC_i : (T_{ij}\varphi, \tau)}{BC_i : (B\varphi, h(\tau))} \tag{4}$$

Figure 1 shows the graded BDI agent proposed with the different contexts and some of the bridge rules relating them.

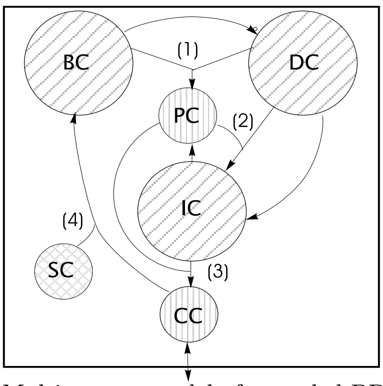

Fig. 1. Multicontext model of a graded BDI agent

9 Conclusions and Future Work

We have presented a BDI agent model that allows to explicitly represent the uncertainty of beliefs, graded desires and intentions. This architecture is specified using multi-context systems and is general enough to be able to model different types of agents. The agent's behavior is then determined by the different uncertainty measures used, the specific theories established for each unit, and the bridge rules. As for future work, we are considering two directions. On the one hand we want to continue with the extension of our multi-context agent model to a multiagent scenario, including other kind of relations with other agents. On the other hand, from an computational point of view, our idea is to implement each unit as a prolog thread, equipped with its own meta-interpreter, in charge of manage inter-thread (inter-context) communication. This implementation will support both, the generic definition of graded BDI agent architectures and the specific instances for particular types of agents. The implementation will also allow us to experiment and validate the formal model presented.

References

1. Benferhat S., Dubois D., Kaci S. and Prade, H. Bipolar representation and fusion of preferences in the possilistic Logic framework. In *KR-2002*, 421-448, 2002.
2. Casali A., Godo Ll. and Sierra C. Graded Models for BDI Agents. In Proceedings of *CLIMA V*, Leite J. and Torroni P. eds., 18-33. Lisboa, Portugal, 2004. Also in Proceedings of *EUMAS'04*, Barcelona, Spain, 2004.
3. Castelfranchi C. and Falcone R. Social Trust: A Cognitive Approach, in *Trust and Deception in Virtual Societies*, Kluwer Academic Publishers, 55-90, 2001.
4. Ghidini C. and Giunchiglia F. Local Model Semantics, or Contextual Reasoning = Locality + Compatibility *Artificial Intelligence*,127(2):221-259, 2001.
5. Giovannucci A. Towards Multi-Context based Agents Implementation. IIIA-CSIC Research Report, 2004.
6. Giunchiglia F. and Serafini L. Multilanguage Hierarchical Logics (or: How we can do without modal logics) *Journal of Artificial Intelligence*, vol.65, 29-70, 1994.
7. Godo, L., Esteva, F. and Hajek, P. Reasoning about probabilities using fuzzy logic. *Neural Network World*, 10:811–824, 2000.
8. Hájek, P. *Metamathematics of Fuzzy Logic*, vol.4 of Trends in Logic. Kluwer, 1998.
9. Liau C. J., Belief, Information Acquisition, and Trust in Multiagent Systems - a modal formulation. *Artificial Intelligence* 149, 31-60, 2003.
10. Meyer J. J. Dynamic Logic for Reasoning about Actions and Agents. *Workshop on Logic-Based Artificial Intelligence*, Washington, DC, 1999.
11. Parsons, S., Sierra, C. and Jennings N.R. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3): 261-292, 1998.
12. Parsons, S. And Giorgini P. On using degrees of belief in BDI agents. In Proceedings of *IPMU-1998*, Paris, 1998.
13. Rao, A. And Georgeff M. Modeling Rational Agents within a BDI-Architecture. In *KR-92*, 473-484 (ed R. Fikes and E. Sandewall), Morgan Kaufmann, 1991.
14. Sabater,J., Sierra, C., Parsons,S. and Jennings N. R. Engineering executable agents using multi-context systems. *Journal of Logic and Computation*12(3): 413-442, 2002.

Grounding is not Shared Understanding: Distinguishing Grounding at an Utterance and Knowledge Level*

Mauro Cherubini[1], Jakko van der Pol[2], Pierre Dillenbourg[1]

[1] CRAFT, Swiss Federal Institute of Technology, Lausanne (EPFL),
Bât. CE1.631, Station 1
CH-1015 Lausanne, Switzerland
{Mauro.Cherubini, Pierre.Dillenbourg}@epfl.ch
http://people.epfl.ch/Mauro.Cherubini
[2] IVLOS, University of Utrecht, Heidelberglaan 8
kamer 384 (Bestuurgebouw)
NL-3508 Utrecht TC, The Netherlaands
j.vanderpol@ivlos.uu.nl

Abstract. This paper argues that for the study and facilitation of collaborative learning, existing theories of grounding such as that of Clark and Shaefer [5] cannot be applied without adjustments. When comparing collaborative learning and conversation, four dimensions can be identified where grounding at a knowledge level differs from the grounding at an utterance level. Firstly, the indirect access and the existence of a range of manifest meanings, poses the need for a notion of 'groundedness'. Secondly, we propose providing evidence in 'co-referenced actions' to be an important process as well as an additional marker to assess grounding. Thirdly, instead of simply repairing misunderstandings after they arise, 'perspective taking' becomes a more prominent mechanism. Fourthly, effort into grounding is turned from needing to be minimised, into needing to be 'optimised'.

1 Introduction

Many studies of Computer Supported Collaborative Learning (CSCL) that identify grounding as an important process, analyse it using the theory of (or models based on) Clark and Shaefer [5]. However, the application of their theory within the field of CSCL holds some problems. As a linguistic theory, it analyses conversation on a micro or 'utterance' level and is not developed to describe the macro or 'knowledge' level, which is associated with learning. While the micro level focuses on the dialogue interchange occurring between two or more interlocutors, the macro level refers to the

* Collaboration for this paper was partly made possible by funding from the Dutch Organisation for Scientific Research (NWO).

shared understanding that is constructed as a consequence of that exchange [10]. We argue that the observable presentation and acceptation of utterances, as described in Clark and Shaefer's contribution theory, cannot automatically be translated into the sharing of knowledge. As Koschmann's [14] example of a learning conversation between surgeon and student in an operation room shows, even repeated presentation and acceptation phases of a concrete referent in a shared environment, can result in different personal representations at a knowledge level.

Since language is not a direct translation of a speaker or writer's knowledge, the interaction between knowledge and language that we find within CSCL, is a complex one [1]. While everyday human interaction has developed to be very efficient in the recognition of mutual intentions, communicating about knowledge (or 'semantic grounding' [2]) cannot automatically rely on the same unproblematic and self-regulating character of 'grounding-for-conversation'. Our reason for stressing this, is while we believe in the great potential of communication to produce learning, we want to caution that not all communication will automatically do so. When analysing or designing for collaborative learning, we need to take into account the idea that successful conversation is not necessarily the same as successful knowledge sharing.

To explore the complex interaction between conversation and learning, this article will investigate the (subtle) differences of the characteristics, evidence, principles and mechanisms of grounding at the micro and macro level. To give this a practical context, we will present two examples: the use of mobile messaging for a spatial collaboration task 1.1, and the use of asynchronous electronic discussion for collaborative text processing 1.2.

1.1 Example A

In this first example[1], we will illustrate the limited information that acknowledgements give us about grounding at a knowledge level with an instance of human-to-human IT mediated communication, where two agents are coordinating for a meeting in an urban environment. The two peers exchange messages using an SMS (Short Messaging System, a system used to exchange short text content on mobile phones) chat, with the aim to guiding themselves in the actual space, towards the goal to reach a physical co-presence. Below, we will report the exchange transcript (Table 1) and the reference to the city map (Fig. 1).

---

[1] We recognise that this first example does not pertain to formal learning situations. However, we believe that informal learning as the coordination for a meeting can be assimilated to the mechanisms of knowledge construction related to conventional learning

Table 1: Transcript of the example conversation. In the third column we coded the transcript using the formalisation proposed by Traum [24]

Agent	Msg. #	Contrib./Act	Message Content	[Actual Action]	Map.#
A	1	initiate / initiate^{I}(1)	Can we meet at St. Francis church at 9?	[Standing in "x"]	1
B	2	ReqRepair^{R}(1)	Ok. Where is it? I am at the St. Paul 's station.	[Standing in "y"]	1
A	3	repair^{I}(1)	Go to the central plaza. Take left and the first right. Then the first left. See ya	[Standing in "x"]	1
B	4	ack^{R}(1)	Ok. I am on my way.	[Walking towards "z"]	2
B	5	ReqRepair^{R}(2)	I am lost. No way on left. I took right at the first junction but there were two streets. I took right again.	[Walking back towards "j"]	3
A	6	repair^{I}(2)ReqAck^{I}(2)	No, sorry. There you must stay on the main road. You should see me.	[Walking towards "k"]	3
B	7	ack^{R}(2)	Ok.	[Walking towards "k"]	4
A	8	initiate^{I}(3)	I am waiting at the red cross on the left hand side of the st.	[Standing in "k"]	4
B	9	ack^{R}(3)ReqAck^{R}(3)	Found the red cross office. Where are you?	[Standing in "w"]	5
A	10	cancel^{I}(3)initiate^{I}(4)	No, sorry it was another cross :-) Keep going for another two blocks.	[Standing in "k"]	5
B	11	ack^{R}(4)	Ok.	[Walking towards "k"]	5

Fig. 1: Map references used in the transcript at Table 1

If we try to model the described situation using Clark and Shaefer's Contribution Model$_{1989}$ [5], or Traum's Grounding Acts Model$_{1999}$ [24], we reach the conclusion that A and B have grounded their conversation at each acknowledgment. More precisely: once both presentation and acceptance phases have been completed, the peers will have grounded a certain contribution (at utterance n. 4, 7, 9, 11). There is the tendency in CSCL to correlate the rate of acknowledgments with the level of shared understanding on the assumption that the provision of evidences of reception is enough to infer the understanding of the signal and the corresponding incorporation in the contributor's beliefs. Additionally, when using these models, it is difficult to operationalise a lack of understanding, as in the example provided when B leads to point "z", because B provided clearly evidence of acceptance as per message 4, on Table 1.

Our claim is that in order to take into account the complexity of this kind of interaction we need to look at the situation from a knowledge construction point of view. From there, we argue, new descriptors of grounding are needed. Therefore, to stay with our example, we can say that the respondent B had an illusion of grounding between point "y" and point "j", until s/he realised that multiple solutions were possible and s/he did not have enough information to solve the ambiguity.

1.2 Example B

The second example was collected in a study of using asynchronous discussion groups for the collaborative processing of academic texts [25]. This example serves mainly to show the fuzzy, ambiguous nature of communication when talking about abstract concepts, and the more nuanced levels of grounding that have to be distinguished. To fully illustrate this point, we would have to give an account of a whole discussion thread with 10 or 20 messages where students try to make sense of their subject matter. Then we would see a conversation that 'in form' would behave like the contribution theory describes, while it still would be very hard to determine how the different messages relate to one another on a knowledge level. Often, while all messages roughly concern the same topic and all consist of slightly different points of view, it will remain unclear whether anyone understood someone else's message, or whether any new knowledge was 'build' (see Fig. 2).

Discussion statement: "According to Laurillard, phenomenography is a research method that focuses on task specific characteristics"

Date: January 12, 2000 05:06 PM
Subject: wrestling with the statement

I find the start statement of this discussion not as clear than it is at first sight. If I read Laurillard I can only partly confirm the statement. De question is, however, whether phenomenography is a suitable research method for task specific characteristics.

Date: January 19, 2000 02:49 PM
Subject: some more clarity

In our class of 12/1/2000 my group has posed a 'reading question' about phenomenography. In the discussion that followed a few things obviously came forward, after which a joint
conclusion was drawn, which is the very thing that: Laurillard prefers the phenomenographical approach (p.35). The drawbacks that she mentions after that are aimed against old forms like observation of behaviour and in favour of phenomenography. Only by asking about it and not by observing, can one know how student experience a certain concept.

Fig. 2: Sample from an electronic discussion between university students. Grammatical errors were copied, to give a correct impression of the original ambiguities

The second message, which aims to provide ‘some more clarity’ on the abstract subject of Laurillard’s account of 'phenomenography', can be considered as an attempt for grounding at a knowledge level. The most important aspect of this follow up message, is that it present a certain gradual ‘topic drift’: the question “what is phenomenography” is answered by a message about the question “how does Laurillard evaluate phenomenography”. This topic drift results in the fact that existing views are being exchanged, but because of a lack of relevance of one message to another, little new knowledge can be build.

2 Four Dimensions of Grounding at a Micro and Macro Levels

Using the presented examples, we will now elaborate on the difference between the micro and macro level, in four interrelated dimensions (see Fig. 3). Firstly, our examples show that the broad range of possible meanings on a knowledge level makes grounding more difficult, and is more likely to result in partial understanding then at a conversation level. Secondly, when it comes to measuring successful grounding, we propose to look at levels of commitment and co-referenced action, which might demonstrate (degrees of) shared knowledge better then acknowledgements. Thirdly, we will look at the underlying principles and see that because grounding is essentially efficiency-driven, the notion of ‘effort’ plays a central, but different, role at both levels. Finally, we will investigate where this effort is or should be directed and identify of perspective taking [13] as a primal grounding mechanisms on the knowledge level.

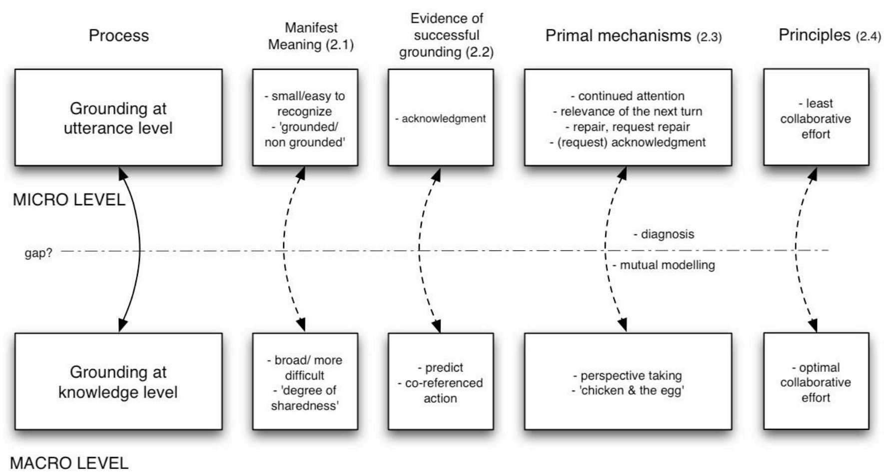

Fig. 3: A four-component model of grounding at utterance and knowledge levels

2.1 Manifest Meaning

Knowledge can never be accessed directly. As Laurillard [16] states, we have to *infer* conceptual information ('descriptions of the world') from the physical or communicative interactions we make in this world, thus making abstract learning, or communicating about knowledge, an essentially *mediated* phenomenon. Since this mediation is never perfect, and 'common ground' can never be reached completely ([11] referring to Wittgenstein), we will use the notion of 'mutual cognitive environment' instead [26]. Sperber and Wilson define a cognitive environment as *the set of facts that are 'manifest' at a certain moment to a person: the facts that he or she is capable of representing and accepting as true or probably true*. Or, in our words, what is manifest for a certain person is the range of possible meaning that is evoked or triggered by the presented evidence, in a certain context. This collection of meanings that are associated to a certain action, concept or statement can even be so broad that it includes contradictory points of view [3]. The difference with Clarks description of common ground, is that to say two people share a cognitive environment does not imply they make the same assumptions; merely that they are *capable* of doing so.

While Clark's experiments started from the idea that a piece information x is either known or unknown to person A or B, the notion of 'manifestness' shows that there are also many degrees in between, and many different ways of 'knowing piece of information x'. We can say that the bigger the overlap is between the manifest meanings of different conversation partners, the more successful their grounding. When looking at the two levels we distinguished, we can state that the need for a notion of 'groundedness', which can account for subtle differences in interpretation, is even greater at a knowledge level than it is at an utterance level. Or, as Andriessen and Alargamot ([1], p. 8) put it: "*semantic understanding is something gradual*". Also, the smaller and more focused a range of manifest meanings is, the better the chances for successful grounding. This depends on what one is grounding: an intention or speech act, a literal meaning, a statement, or a certain point of view. The more elaborate and complex the grounding object, the more difficult grounding. Because the range of possible interpretations will usually be broader at the knowledge level than at the utterance level, grounding will also be more difficult at that level, and "a communicative intention can be fulfilled without the corresponding informative intention being fulfilled" [26]. We would like to see the distinction between the micro and the macro level not strictly as a dichotomy, but rather as a range, for instance going from recognizing simple intentions, to recognizing literal meanings, more elaborated statements and finally complex points of view.

2.2 Evidence of Successful Grounding

In concordance with Sperber and Wilson's account of the *evidence* that messages provide to guide their interpretation, the same can be said about analysing grounding. The more evidence we have, the more we know about the levels of shared understanding (though it may never be conclusive). As we have stated in the introduction, we do not think acknowledgements are always a valid measure of shared understanding. Apart from different goals at the two levels (see 2.4), Ross, Green and House [21] have shown that an (partial) 'illusion of shared knowledge' is not only possible, but also even likely to occur (called the *false consensus effect*). Therefore we propose to look at verbal and physical *actions* as well. Bereiter's term 'knowledgeability' [3], or 'being able to take intelligent action', indicates that (verbal or physical) actions intrinsically contain knowledge. In our first Example 1.1 the 'information bearing actions' one can identify are the coordination of tuning attempts with the agreed plan. If the pair agree to a certain strategy and then implement it coherently, we can infer that the pair successfully grounded to a high degree. Or, to state it more generally, if someone '*commits to a previous statement, and subsequently does something directly related to it in the forthcoming action or statement*' (we use this notion of commitment in accord with [8]). Since this relatedness between communicative actions requires a large overlap in the cognitive context and shared referents, we will label them as *co-referenced actions*.

In the asynchronous discussion groups (Example 1.2) we can look at the alignment of questions and answers. An answer that follows a question seems like a legitimate and useful speech act (utterance level), but only if the relevance of the content is established, we can deduce if it is also a successful knowledge-building act. On a knowledge level, for an action to be 'co-referenced', it is required that it refers to a shared piece of knowledge and needs to be relevant from someone else's view. According to Sperber & Wilson: "*something is relevant to an individual when it connects with background information he has available to yield conclusions that matter to him: say, by answering a question he had in mind, improving his knowledge on a certain topic, settling a doubt, confirming a suspicion, or correcting a mistaken impression.*" ([27], p. 608). Our examples show that, while at an utterance level, both recognising a certain speech act, (such as identify a question by its question mark) and providing a relevant response (giving an answer) is pretty straightforward, on knowledge level the requirements for action to be relevant or co-referenced are much higher.

2.3 Grounding Mechanisms

At an utterance level, human communication is very efficient by investing minimal effort in elaborate message design or conscientious interpretation, but rather by jumping to (subjective) conclusions and repairing a possible misunderstanding *after* it arises. At a knowledge level however, we have seen that because of the mediated nature of grounding and the more complex collections of associated (manifest) meanings, this presents more problems. Miscommunication can be both harder to detect (thus cannot be relied upon to reveal itself) and to repair. Therefore, the grounding mechanisms at the knowledge level might present the most important shift from the utterance level. To understand what nuanced meaning other people attribute to certain statements, one must 'put oneself in the other's shoes' and try to identify which meaning will be relevant for that person [26]. In order to infer someone else's cognitive environment or 'frame of reference', both for reading and writing messages (audience design), we rely on strategies like *perspective taking* [13] and *mutual modeling* (for a definition see [19]).

While at an utterance level repair mechanisms are know to be self-regulating (the less shared understanding, the more grounding will take place, see for example [10]), this is less evident for knowledge level perspective taking. It seems that at this level, the 'chicken & the egg' relation between grounding and common ground ("*It is hard to find some if you don't have some already and you don't have any unless you find it*", [15], p. 4) is even more prevalent than it is at the utterance level. This shows that at a macro level, knowledge of other's perspectives plays a role as a prerequisite as well as an outcome and the same goes for one's knowledge of the subject matter. Because identifying another's frame of reference is easier if one has knowledge of the different possible frames of reference that exist, perspective taking is also tied to existing knowledge. This underlines the reciprocal relationship between individual and collective processes in collaborative learning [23]: it is not only so that individual learning results from collaborative processes, but individual knowledge also influences the success of collaboration.

2.4 Grounding Principles

First of all, grounding is functional and driven by mechanisms of efficiency, as Clark & Wilkes-Gibbs [6] demonstrate with their 'principle of least collaborative effort' and Wilson and Sperber [26] in their 'relevance-theoretic comprehension procedure'. The fact that in grounding: no more effort will be invested that what is 'sufficient', can explain the lack of co-referenced actions in our examples. For students the costs (relative to the goals) may simply be too high, especially because high-level learning goals are usually translated into practical tasks, with which students deal in a pragmatic way. Taking the perspective of someone else may take more effort than staying within one's own perspective, and what is 'sufficient to continue the conversation' might not be 'sufficient for learning' [2]. That is why, for learning, instead of trying to 'minimize the collaborative effort', we strive for an: *optimal collaborative effort* [9].

The effect of effort into perspective taking and co-referenced actions is twofold: not only does relevant feedback enhance collaborative knowledge building, but the effort after shared meaning itself is also strongly associated with learning [22], especially if the effort is directed at the knowledge level (or 'semantic grounding', see [2]). Spending effort into trying to understand another perspective *is* learning: it is leaving one's preconceptions and trying to new information and insights. The is also true for reading, since *perspective taking* for comprehending messages is closely related to the comprehending process when studying scientific texts.

3 Conclusions

Context is inextricably present when we grasp meanings and when we infer knowledge. The pragmatic tradition of relevance highlights the action-oriented nature of intelligence, where the term 'action' is to be understood in a broad sense that includes reasoning behaviour, or communicative acts [7].

When looking at the relevance of communicative actions in collaborative learning, we have described that providing evidence and acting in a co-referenced way is crucial for developing a shared understanding. The more evidence is presented, the easier it becomes to take another's perspective, act in a co-referenced way and enhance the degree of shared understanding.

We suggest the implications for design and research on grounding in collaborative learning might involve an effort to facilitate grounding at a knowledge level. For instance, communication tools could be developed that provide more (focused and detailed) contextual information which serves to limit the range of manifest meanings of the concepts that are being used, and thus to increase the chances of shared understanding. Also, since the use of acknowledgements as markers of shared understanding is problematic, we propose to create markers that can give an account of the relevance of communicative actions in regards to the reasoning process. As an example of this, we think 'operationalising' the 'co-referencedness' of actions on a knowledge level, as measure of shared understanding, would be a valuable effort. As a final remark, we want to conclude with a question for discussion:

- How can we 'operationalise' the concept of Perspective Taking?

The concept of co-referenced action, as defined in this paper, presents several limitations. It is sometimes difficult to have a multi-modality of communication that may make visible incongruences between intentions and actuations. We need other markers. Relevance theory states that an input is relevant when, and only when, its processing yields a positive cognitive effect [26], meaning that the receiver of the input will generate and act an expectation of a particular cognitive effect to be achieved by the incoming input [18]. These 'acts' of expectation may be observed and accounted as yielding back to the reasoning and inference process of the learner and on the particular perspective s/he is taking in relation to the processed input.

References

1. Alamargot, D. Andriessen, J.: The "power" of text production activity in collaborative modelling: nine recommendations to make a computer supported situation work. In M. Baker, P. Brna, K. Stenning, A. Tiberghien (eds): The Role of Communication in Learning to Model. Lawrence Erlbaum Associates. NJ (in press)

2. Baker, M., Hansen, T., Joiner, R., Traum, D.: Collaborative Learning: Cognitive and Computational Approaches, chapter The role of grounding in collaborative learning tasks, Pergamon / Elsevier Science, Amsterdam (1999) 31–63
3. Bereiter, C.: Education and mind in the Knowledge Age. Lawrence Erlbaum Associates, Inc. (2002)
4. Billig, M.: Ideological dillema's; a social psychology of everyday thinking. SAGE publications, London (1988)
5. Clark H. H., Shaeffer, E. F.: Contributing to discourse. Cognitive Science 13 (1989) 259-294
6. Clark, H.H., Wilkes Gibbs, D.: Referring as a Collaborative Process, Cognition, 22 (1986) 1-39
7. Ekbia, H. R., Maguitman, A. G.: Modeling and Using Context (Proceedings of the conference CONTEXT 2001, Dundee, UK, July), volume 2116 of Lecture Notes in Artificial Intelligence, chapter Context and Relevance: A Pragmatic Approach. Springer, Berlin, (2001) 156–169
8. DiEugenio, B., Jordan, P. W., Thomason, R. H., Moore, J. D.: The agreement process: an empirical investigation of human-human computer-mediated collaborative dialogues. International Journal of Human Computer Studies. (2000) 1–57
9. Dillenbourg, P., Traum, D., and Schneider, D.: Grounding in multi-modal task-oriented collaboration. In Proceedings of the European Conference on AI in Education, pages, Portugal (2000) 415–425
10. Dillenbourg, P. and Traum, D.: The complementarity of a whiteboard and chat in building a shared solution. Journal of the Learning Sciences, 26 (in press)
11. Draper, S. W., Anderson, A.: The significance of dialogue in learning and observing learning. Computers Educ., 17(1) (1991) 93-107
12. Gergle, D., Kraut, R. E., Fussel, S. R.: Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. Journal of Language and Social Psychology, 23(4) (2004) 491-517.
13. Järvelä, S., Häkkinen, P.: Web based cases in teaching and learning - reciprocal understanding and perspective taking in conversation. CAL99 (Computer Assisted Learning), London (1999)
14. Koschmann, T.: Reconsidering Common Ground: Examining Clark's Contribution Theory in the OR. ECSCW 2003: the Eighth European Conference on Computer-Supported Cooperative Work, Kluwer Academic Publishing (2003)
15. Krauss, R. M., Fussell, S. R.: Mutual knowledge and communicative effectiveness. In J. Galegher, R. E. Kraut, C. Egido (eds.): Intellectual teamwork: Social and technological foundations of cooperative work. Hillsdale, NJ: Erlbaum (1990)
16. Laurillard. D.: Rethinking university teaching. London, Routeledge (1993)
17. Mäkitalo, K., Häkkinen, P.: Building and maintaining common ground in web-based interaction. In Proceedings of the 9th Conference for Research on Learning and Instruction, Fribourg, Switzerland (2001)
18. Matsui, T.: Modeling and Using Context (Proceedings of the conference CONTEXT 2001, Dundee, UK, July), volume 2116 of Lecture Notes in Artificial Intelligence, chapter Experimental Pragmatics: Towards Testing Relevance-Based Predictions about Anaphoric Bridging Inferences, Springer, Berlin (2001) 248–258
19. Nova, N., Wehrle, T., Goslin, J., Bourquin, Y., Dillenbourg, P.: The impacts of awareness tools on mutual modelling in a collaborative video-game. In Favela, J. and De-

couchant, D., editors, Proceedings of the 9th International Workshop on Groupware, Autrans, France (2003) 99–108
20. Pickering, M. J., Garrod S.: Toward a mechanistic psychology of dialogue. BBS (under review)
21. Ross, L., Greene, D., House, P.: The false consensus phenomenon: an attributional bias in self-perception and social perception processes. Journal of Experimental Social Psychology, 13 (1977) 279-301
22. Schwartz, D. L., & Lin, X. D.: Computers, productive agency, and the effort after shared meaning. Journal of Computing in Higher Education, 12 (2) (2000) 3-33
23. Stahl, G.: A model of collaborative knowledge-building. In Proceedings of the Fourth International Conferences of the Learning Sciences (ICLS), MI, USA. Ann Arbor (2000)
24. Traum, D. R.: Computational models of grounding in collaborative systems. In working notes of AAAI Fall Symposium on Psychological Models of Communication (1999) 124–131
25. Van der Pol, J.: Identifying and modelling variables in complex CSCL-situations. Case study: the use of asynchronous electronic discussions. CSCL2002 workshop Designing Computational Models of Collaborative Learning Interaction, Boulder, Colorado (2001)
26. Wilson, D., Sperber, D.: Relevance: communication & cognition. Blackwell publishing, Oxford (1995)
27. Wilson, D., Sperber, D.: Relevance Theory. In G. Ward and L. Horn (eds.) Handbook of Pragmatics. Oxford: Blackwell (2004) 607-632

Fragile Contextual Effects in Judgment of Length.

Penka Hristova

Central and East European Center for Cognitive Science,
Department of Cognitive Science and Psychology,
New Bulgarian University, 21 Montevideo Street, Sofia 1635, Bulgaria
(phristova@cogs.nbu.bg)

Two experiments test the possibility of enlarging contrast effect due to properties of the target stimulus that are irrelevant to the task. Both experiments require judgment of lengths of lines, which might be differentiated depending on their color. According to the JUDGEMAP model irrelevant characteristics take part in judgment by directing the activation. In this manner the irrelevant information modulates dynamically the content of WM and thus influences judgments of the stimulus in different contexts. Experiment1 shows significant effects of irrelevant characteristics upon judgments of middle lines under limited presentation time. Experiment2 explores the possibility of enlarging the effect by making use of two irrelevant dimensions, which correlate. It fails to demonstrate any effect of irrelevant dimensions, probably because of opposing sources of assimilation. Finally several possible explanations and some potential ideas for overcoming the brittle contextual effects in judgment are discussed.

1. Introduction

The goal of this paper is to give an idea about the way people decide that something is good or bad, beautiful or ugly, big or small, etc? How do they manage to assess something on a given scale or to specify that it is several times better than another in any respect? To be more precise, it will try to shed some light on the mechanisms that might underlie contextual influence on judgment. The implicit assumption behind this goal is that judgment *could not* be differentiated from the current context.

Existing theories that try to explain the relativism in human judgment stress on different characteristic of context. Some of them treat context as part of the surrounding environment –contextual stimulus presented at the time of judgment [2] or a set of stimuli that a person has already been judged [12,14,15]. Others are more interested in internal context – the most active information [4], the most accessible and applicable information during judgment [17,18,21]. The starting point taken at this paper is the definition of context proposed by Dynamic Theory of Context [9]. It treats context both as comprising of some elements of the surrounding environment and some internal states or associative relevant elements retrieved from memory. This definition resembles quite well the idea for working memory (WM) adopted in DUAL cognitive architecture [6,7,8]. The content of WM might be considered as a dynamic entity embracing different information for the same task in the course of time. Thus, WM seems to be a good substitute of the notion of "*context*".

Another potential problem of the current theories is that they rarely explain judgment by the means of concrete mechanism. Empirical data, however, could hardly be differentiated into groups depending on contextual forces that change judgment in a systematic manner. This might be a good reason for looking a level below, on the mechanism that might underlie contextual influence in judgment. Different researches demonstrate different factors that induce judgment. Sometimes manipulations of the same factor might lead either to contrast or assimilation. For example, presenting an anchor, which value is within the values of stimulus range conduct assimilation, while presentation of an anchor that do not belong to the values of the stimulus set shifts judgments of the same stimuli in the opposite direction, e.g. contrast [13,16]. It could however be that different factors influence judgment differently - presenting a moderate in attractiveness woman face within a set of other more attractive faces lead to contrast (lower rating), but the same face presented simultaneously with another more attractive face would raise its rating (assimilation)[19].

One of the goals of JUDGEMAP project was to propose concrete mechanism for explaining the way context influences judgment [10]. JUDGEMAP is a computational model of judgment based on the DUAL cognitive architecture. DUAL takes the advantage of decentralized representations of concepts, objects and events, and parallel emergent computations. It is explicitly designed for modeling context-sensitivity of human cognitive system. Since DUAL is a cognitive architecture it provides an opportunity for modeling different cognitive processes with the same mechanisms. JUDGEMAP models judgment as a process of *mapping* a set of retrieved and perceived stimuli onto the scale - it ensures that smaller magnitudes would receive smaller ratings and vise versa. It is integrated with AMBR model [6,11] for analogy making. JUDGEMAP and AMBR both emphasize the role of associative memory. They retrieve a set of associative relevant memory elements and process them together with the perceived one in the scope of WM. JUDGEMAP strongly relies on memory as a potential starting point for the subsequent judgment. In this respect it is similar to the idea of constructed norm proposed by Norm Theory [5].

2. Predictions of JUDGEMAP

JUDGEMAP points to several potential mechanisms that might underlie contextual forces in judgment. Some of them concern the retrieval of elements in the WM, others the mapping between the set elicited in the WM and the required scale. This paper presents two experiments that test the model's prediction about mechanisms *responsible for construction* of the content of the WM. In JUDGEMAP spreading activation mechanism carries over the retrieval of information from the LTM. The activation comes from INPUT (target and potential contextual stimuli from the environment) and GOAL (the task to judge the stimulus on the required scale) nodes. It spreads through the network of micro-agents that represent concepts, exemplars and episodes. Thus, concepts related to the target one, or possessing the same features as the target stimulus becomes also active. The features representing the target stimulus activate some of the related concepts and their exemplars. The activation spreads through both relevant and irrelevant features of the target. This lead to an interesting *prediction* that *irrelevant dimension* also *takes part in the judgment process*. For example, if a person has to judge the height of a particular target person, his/hers judgment will be affected from the extra knowledge about the target. If the target person is a musician, the stimuli, which would be elicited in WM would be prevailing musicians. Moreover, if the target person is also a blond it is quite likely that the exemplars elicited in the WM would be predominantly of other blonds.

This specific prediction of JUDGEMAP received some empirical support [10]. Ss were asked to judge the length of lines, which were either green or red. Green lines form a positively skewed set, while red ones – a negatively skewed set. Both set posses the same range of stimulus lengths but different density. Green and red lines have the same end points but green lines were relatively short, while red lines were relatively long. The assumption behind this experimental procedure lies in the Range-Frequency Model [12,14,15]. It predicts a contrastive shift in rating of a stimulus that is included in uneven distribution.

In this experiment, lines were evenly distributed with respect to their length but unevenly with respect to their color. Thus if Ss *pay any attention* to irrelevant information they would judge the same line differently depending on its color. More precisely, rating of the green lines would be higher than ratings of the red ones despite they were equal in length. This was exactly what we obtain -the difference in judgments of green and red lines was small those significant (0.046). This result was interpreted as being in favor of our hypothesis, namely, that people elicited different sets of stimuli when judged green and red lines. Green lines retrieved more green exemplars than red lines with the same physical length and vise versa. Moreover, surprisingly simulations with JUDGEMAP model demonstrated similar in size effect of irrelevant dimension upon judgment of the length of lines (0.053).

However, several questions remained:

1. *Whether the obtained difference was accidental one.*

This question could not be answered easily without a series of experiments that display the same effect of the irrelevant dimension in a consistent manner. In this respect, our expectations are the following ones:

- To receive a raise in judgments of positively distributed stimuli in comparison with negatively skewed ones depending on their irrelevant characteristic;
- This difference to be higher for middle stimuli than for the rest. This specific hypothesis is a sequence of Range-Frequency model.

2. *Can we enlarge the effect of the irrelevant dimension?*

This question concerns the size of contextual effect. Its answer should specify such testable conditions that might enlarge the differences in ratings of the same line depending on its irrelevant dimension. One possible way of studying the impact of context on any cognitive process is to put it under "stress". This suggestion was explored in experiment1. The time for stimulus presentation was limited to 100ms. However, Ss were not urged to judge the stimuli faster, they simply were limited in their perception of each line segment. This seems a reasonable way to put judgment under stress, because it restricts the influence of some possible environment anchors like borders of the screen and predispose to more mistakes.

The goal of the second experiment was also to enlarge the impact of the context on judgment. It closely resembles another prediction of JUDGEMAP model, namely that activation that spreads to several dimensions that feed the same information raise the possibility this particular information to be elicited in comparison with a situation in which only one dimension activate it. The most active element in WM is the element that shares nearly all elements of the target stimulus. Thus, if several irrelevant dimensions correlate the irrelevant information would have better chances to direct memory retrieval. Hence the context, understood like the content of WM would include more elements that share the same irrelevant to the task information.

3.Experiment1: Judgment of the length of lines under "stress"

Stimulus lines were projected for a very *short time* (100ms) that allows Ss only to detect stimulus but not to compare it unrestrictedly to the rest of the surrounding sources of useful information. This peculiarity of the experimental procedure seemed to us enough for causing some potential troubles for the judgment process. We hope that by making judgment harder will obtain more mistakes and smaller precision in judgment. Restricted external sources of relational information and the opportunity for more mistakes seem to enable the possible influence of irrelevant information upon retrieval of similar elements in the WM.

Ss were asked to judge the length of lines that appear always horizontally, but in random positions on the screen. Each line was projected for a very short time on the computer screen- for only 100ms. The subsequent answer did not require a prompt, rather the computer "wait" for the Ss's answers. The Ss were instructed to press the button when they are sure what rating the target line deserves.

The irrelevant dimension was embedded in these so simple stimuli through the color of the lines. The lines were either green or red. Depending on their single irrelevant dimension, the lines took part in a positive or a negative distribution. All lines form a uniform distribution of lines. Thus if subjects consider only the relevant dimension of stimuli, they would pay attention only to the length of the lines and would judge lines independently of their color. It could however be that Ss might not disregard the irrelevant to task color of the target lines. Than they would probably shift their ratings of the same line depending on its color – lines from positive distribution would be judged higher than lines from negative distribution despite they are equal in length.

Method

*Design***:** The experiment has a 14x2 within-subject factorial design. The independent variables are length (varying at 14 levels) and color (varying at 2 levels) of the lines. The experimental design was counterbalanced in order the positively and the negatively skewed stimuli to be presented either in green or in red. In the first experimental condition, the green lines were positively skewed, while red lines form negatively skewed distribution. In the second experimental condition, red lines were positively skewed, while green lines were negatively skewed. The dependent variable is the mean rating of the length of lines on a 7-point-scale.

Stimulus Material: A set of 14 lengths was designed. The smallest one was 180 pixels, the longest one was 505 pixels, and the increment was 25 pixels. Each length was presented 24 times forming a set of 336 lines. Each line was presented either in red or green. The presentation of the lines in the first experimental group is placed in table 1. In second experimental group the presentation of lines was just on the opposite, e.g. red lines formed positively skewed distribution (include relatively short lines) and green lines formed negatively skewed one (include relatively long lines).

Table 1. Presentation schedule of lines in group 1, where green lines were positively skewed and red lines were negatively skewed one.

Length	Number of the green lines (Positively skewed distribution)	Number of the red lines (Negatively skewed distribution)
1 and 2	21	3
3 and 4	18	6
5 and 6	15	9
7 and 8	12	12
9 and 10	9	15
11 and 12	6	18
13 and 14	3	21

Procedure: Each line was presented horizontally against a grey background in a *random position* on the screen for *100 ms.* Ss were instructed to press a button from 1 to 7 on the keyboard whenever they are sure on their rating. They were also asked to leave their hand on the desk after each pressing of the key in order to be sure that the time for reaching each button on a keyboard is relatively equal. When subject press the button corresponding to his/hers answer the next line appears on the screen (randomly drawn from a set of 336 lines). The experiment was conducted in sound-attenuated room and lasts around 15 minutes.

Participants: 31 students from New Bulgarian University. Part of them participates in order to satisfy a course requirement, others were paid 0.5 EURO.

Results and Discussion

The judgments of each line were averaged depending on their color. Thus each participant has 28 mean judgments (14 lines*2 colors). Color and length was analyzed as a within subject factor, while group was a between subject factor. The group did not differentiate results (F (1.29) = 0.215, p=0.646), thus for plainness, we would use the skew of the distribution as an indicator for contextual dimension instead of using its color (red or green) in the rest of the text. Moreover, the effect of irrelevant dimension that we are studying in the experiment, was formalized trough the skewedness of the lines - we expect positively skewed lines to receive greater ratings than negatively skewed ones with the same physical length. Unfortunately, this was not happen. The difference between mean judgments of lines with different color did not reach significance: F (1, 29) =0.070, p=0.793. The mean judgment for each line presented in different color is presented in fig. 1.

It could however be that judgments of middle length were more sensitive to our experimental manipulation. In line with our prior expectations - difference in judgments of positively and negatively skewed stimuli that belong to the middle range should be more expressed. The repeated measurement analysis performed on length 7 and 8 reveal a significant effect of color upon judgments of the length of lines: F (1, 29) =4.400, p=0.045. The difference between mean judgment of positively skewed lines (5.005) and mean judgment of negatively skewed lines (4.918) was 0.087.

As we expected, judgments of middle lines were influenced in higher extent from their color than the rest of the lines. We found small those significant effects of irrelevant dimension (0.087). Timing limitations for the lines presentation did not increase the impact of irrelevant dimension. The effect of color did not enlarge under these specific conditions. A potential explanation for this result might be that contextual effects in judgments are *quite fragile*. Moreover, it could be that people are quite precise in measuring physical characteristics of surrounding world. The correlation between judgment and length was found to be 0.81. Thus contextual impact has to be somewhere within this quite small interval between 0.81 and 1. In this respect difference in the decimals of contextual effect between the reported experiment and previous one seem to be offhand.

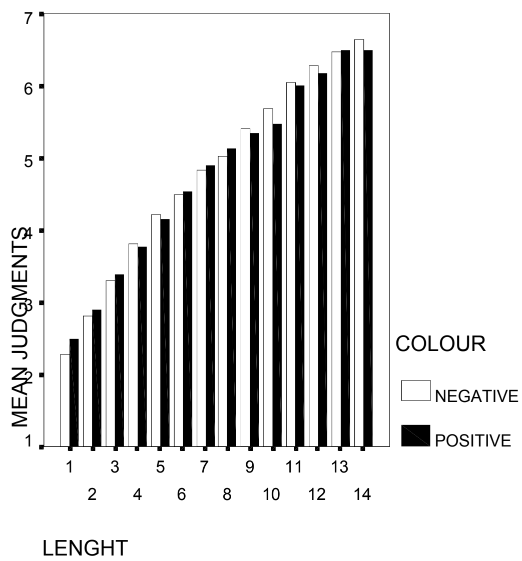

Fig. 1. Mean judgments of all lines depending on their color. Black bars stand for judgments of positively skewed stimuli, white bars – for the negatively skewed ones.

4.Experiment2: two correlating irrelevant characteristics

The goal of this experiment was to explore the situation in which stimuli could be characterized as sharing one relevant to the task characteristics and several irrelevant ones. Stimulus material was the same like in the experiment1. The difference is that experiment2 manipulates two irrelevant characteristics, which correlate. In DUAL, the activation of particular agents in the LTM should become greater if they receive activation simultaneously from several places. Thus positively skewed lines are joined with a particular sound, while negatively skewed ones are - with other sound. Then positively skewed exemplars should receive greater activation when a line from the same set is judged in comparison with a situation in which the lines differ only by their color. This manipulation might help as to elicit with higher probability the exemplars that belong to the target category. Then if the content of WM depends on irrelevant information and retrieved exemplars are used for the consequent judgment, the effect of irrelevant dimensions should be greater.

The following experiment tries to get the advantage of this DUAL prediction. The color of the lines was coupled with a sound, in order to obtain a correlation of two irrelevant to the task dimensions. Green lines always appeared with the sound "laser", while red ones with the sound "whoosh".

Method

Design: The irrelevant dimensions were an independent variable that was varied in two levels. The lines could be red or green, coupled with the sounds "whoosh" or "laser" respectively. Color of the positively skewed line was green in the first group and red in the second group. The color of the negatively skewed stimuli was red in the first group and green in the second group. The length of the lines was also an independent variable that varies in 14 levels.

The dependent variable was the mean judgments of each line on a 7-point scale.

Stimulus Material: The same 14 lines used in the previous experiment were presented with the same frequency depending on their irrelevant characteristics – lines with lengths 1 and 2 were presented 21 times in green to the accompaniment of the "laser" sound, lines with lengths 3 and 4 - 18 times and so on. The experiment was counterbalanced with respect to the irrelevant characteristics.

Procedure: The same as in the first experiment, but without any time limitations for presentation of lines. Each line appears simultaneously with a particular sound - green line with "laser" sound, red lines with "whoosh" sound.

Participants: 29 students from New Bulgarian University. Part of them participates in order to satisfy a course requirement, others were paid 0.5 EURO.

Results and Discussion

Group did not differentiate the effect of the irrelevant dimension upon judgment of the length of 14 target lines ($F(1, 27) = 1.363$, $p = 0.253$). Repeated Measurement Analysis did not reveal significant influence of color upon judgment of the length of all lines ($F(1, 27) = 2.448$, $p = 0.129$). Color also did not differentiate judgments of the middle lengths ($F(1, 27) = 1.929$, $p = 0.176$).

The lack of statistically significant effect might due to the second irrelevant dimension added in this experiment in comparison to the first one. The sounds could be perceived as quite different from participants and thus giving the opportunity for opposing shifts in their judgments. As Goldstone [1] shown in his experiments, people incline to assimilate their judgments toward a category if they perceive the target as similar to it. Moreover, judgments might be assimilated toward categories that are constructed on-line based only on some sort of contextually sensitive similarities. It is possible, experiment2 to oppose two contextual forces – contrast that due to the skew of lines against assimilation toward categories that represent lines, which appear along with "laser" or "whoosh" sound. "Whoosh" and "laser" sound are different enough to provoke such effect.

5.Conclusion

The paper points out the need of explicit starting definition of context in exploring contextual effects in judgment. Two experiments try to reveal the mechanisms of contextual influence upon judgment. The starting point of these experiments was JUDGEMAP model, where spreading of activation is the mechanism responsible for locating of relevant to the task sources of information. The activation spreads from Input and Goal nodes to the elements in LTM, which posses similar characteristics as the target ones. The crucial point here is that activation spreads not only through relevant, primary to the task dimensions but also through irrelevant ones. Thus any information that characterizes target object might be considered during judgment if it enters WM. We try to formulate some special situations, for which the irrelevant information might possibly change the judgment in a systematic fashion.

Both experiments study the effect of the irrelevant information upon judgment of length of lines. The irrelevant information was embedded in these so simple stimuli through their color or concomitant sound. Experiment 1 demonstrates small significant differences in ratings of the length of the middle lines depending on their color. The lines distributed positively depending on their color were judged as longer than lines distributed negatively although they were equal in length.

The theoretical maximum for the expected impact of the irrelevant information could be calculated by the equations of the Range-Frequency Model if the R and Fr variables are calculated for all green and red lines separately. Then two sets of theoretically expected judgments might be acquired. Calculated in this manner mean of the positively skewed lines was 4,4364 and for the negatively skewed one - 3,5636. This means that the absolute maximum of the difference between the ratings of the positively and negatively skewed lines 0,87. The prediction that JUDGEMAP makes did not argue for such size of the contextual effect. JUDGEMAP predicts that the comparison set will be formed dynamically for each stimulus separately and the color of the lines will take part in this process of comparison set formation as well as many other factors. Moreover, target lines retrieve in memory mixed set of lines with respect to their color. Green line retrieved more green lines than red lines and on the opposite, red lines call in memory more red lines. Therefore, we cannot even imagine such a strong difference when colors are irrelevant to the task characteristic. In this case, it seems more realistically to anticipate a smaller difference, e.g. somewhere on the half of the theoretical maximum. The results, however, did not reach such values. The size of contextual effect obtained in experiment1 (0.087) did not come closer to the theoretical maximum (0.87) than the effect registered in the experiment from 2004 (0.046) [10].

It could however be that people are quite precise in judgment of physical length. They might possess the capabilities for judging consistently physical dimensions of the surrounding world. The correlation between judgment and length of the lines found in the reported experiments is 0.81. This fact works against our hypothesis. The impact of the irrelevant to the task dimension seems to be convicted preliminary. Contextual effects in judgment of the length of lines should be held within this quite narrow interval between 0.81 and 1. Such situation seems to be quite restrictive for studying of any contextual effects in judgment. One way to overcome this unfavorable situation is to change the target dimension from judgment of the length of lines to something else that people are not so accurate. Finding a way to reduce the judgment precision of the target dimension should open up place for larger effects. Another way to explicate the effect is considered to be the scale [20]. If a scale with smaller number of intervals is used this might raise the effect of the uneven distribution.

Surprisingly, however, in the second experiment we didn't obtain any significant effect of the irrelevant dimension, even the effect that we have already demonstrated in the experiment from 2004 [10] and experiment 1 reported here in more details. This might be an indication that our starting model is wrong. Context, probably, could not be considered as equivalent to the WM or probably the spreading activation is not so powerful mechanism in the case of judgment. It could however be that the lack of any effect in the second experiment is a sigh for switching on the opposite assimilation effect. As Goldstone [1] reported, contrast and assimilation depend on perceptual similarity between the target stimulus and the rest of the stimuli judged during the experiment. In the case of judgment of green and red lines that appear always in the company of a particular sound it seems reasonable to suppose that people were provoked to construct two separate categories – one for the green lines that appear along with the "laser" sound and one for the red lines that appear along with the "whoosh" sound. We could imagine that people disregard the color of the target lines better than the color along with a particular sound. Moreover, sounds were short. They were heard till the lines appeared and hence, probably before Ss began to judge. They also imply some sort of reasonable meaning that also might push the Ss toward some classifying of the stimuli. Such reflections and the previous two experiments were the reason of interpreting the results as not contradicting to our main hypothesis that irrelevant information takes part in the judgment process. Thus we were more disposed to interpret the result from the second experiment as showing a conflict of two different mechanisms – on one hand, the expected contrast effect depending on the irrelevant dimension, on the other, the assimilation effect toward a category reported by Goldstone [1]. One additional contextual feature - two different sounds join to our stimulus material made the observed effect to disappear.

One way to overcome this unfavorable situation is to change the target dimension from judgment of length of lines to something else that people are not so accurate. Finding a way to reduce the judgment precision of the target dimension should open up place for larger effects. Another way to explicate the effect is considered to be the scale [20]. If a scale with smaller number of intervals is used this might raise the effect of the uneven distribution.

References:

1. Goldstone, R. (1995). Effects of Categorization on Color Perception. *Psychological Science*, vol.6 (5), 298-304.
2. Helson,H. (1964), Adaptation Level Theory: An experimental and systematic approach to behavior, *New York: Harper&Row*.
3. Herr, Sherman and Fazio (1982) On the Consequence of Priming: Assimilation and Contrast Effects. *Journal of Experimental Social Psychology*,vol.19,323-340.
4. Higgins, E.T., W. S. Rholes, and C.R. Jones (1977), Category accessibility and impression formation, *Journal of Experimental Social Psychology,* 13, 141-154.
5. Kahneman, D., and Miller, D. (1986), Norm Theory: Comparing Reality to Its Alternatives, Psychological Review, Vol. 93, No 2, 136-153.
6. Kokinov, B. (1994a). A hybrid model of reasoning by analogy. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory: Vol. 2. Analogical connections* (pp. 247-318). Norwood, NJ: Ablex

7. Kokinov, B. (1994b). The DUAL cognitive architecture: A hybrid multi-agent approach. *Proceedings of the Eleventh European Conference of Artificial Intelligence*. London: John Wiley & Sons, Ltd.
8. Kokinov, B. (1994c). The context-sensitive cognitive architecture DUAL. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.
9. Kokinov, B. (1995), A Dynamic Approach to Context Modeling. In P. Brezillon & S. Abu-Hakima (Eds.*), Proceedings of the IJCAI-95 Workshop on Modeling Context in Knowledge Representation and Reasoning.* LAFORIA 95/11.
10. Kokinov, B., Hristova, P., Petkov, G. (2004). Does Irrelevant Information Play a Role in Judgment? *In: Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ.
11. Kokinov, B. & Petrov, A. (2001). Integration of Memory and Reasoning in Analogy-Making: The AMBR Model. In: Gentner, D., Holyoak, K., Kokinov, B. (eds.) *The Analogical Mind: Perspectives from Cognitive Science*, Cambridge, MA: MIT Press.
12. Pardicci, Allen (1965), Category Judgment: A Range-Frequency model, *Psychological Review*, 72(6), 407-418.
13. Parducci, A.and L. Marshall (1962), Assimilation vs. contrast in the anchoring of perceptual judgments of weight, *Journal of Experimental Psychology*, Vol.63, No. 5, 426-437.
14. Pardicci, Allen and Linda F.Perrett (1971), Category Rating Scales: Effects of relative spacing and Frequency of Stimulus Values, *Journal of Experimental Psychology*, 89(August), 427-453.
15. Parducci, Allen (1974), Contextual Effects: A Range-Frequency Analysis, *Handbook of Perception*, vol.2, NY: Academic Press, 127-141.
16. Sherif, M., Taub, D., and Hovland, C. (1958), Assimilation and contrast effects of anchoring stimuli on judgments, *Journal of Experimental Psychology*, Vol.55, No.2, 150-155.
17. Srull, Th. and Wyer, J.(1979), The role of category accessibility in the interpretation of information about persons: Some determinants and implications, *Journal of Personality and Social Psychology*, 37,1660-1672.
18. Srull, Th., Wyer, J. (1980) Category Accessibility and Social Perception: Some Implications for the Study of Person Memory and Interpersonal Judgment. *Journal of Personality and Social Psychology*, vol. 38 (6), pp 841-856.
19. Wedell, Douglas H., Allen Parducci, and R. Edward Geiselman (1987), A Formal Analysis of Ratings of Physical Attractiveness: Successive Contrast and Simultaneous Assimilation, *Journal of Experimental Social Psychology*, 23(May), 230-249.
20. Wedell D.H., Parducci A., and Lane M.(1990), Reducing Contextual Effects for Ratings of Clinical Case Histories, *Journal of personality and Social Psychology* ,Vol.58, N 2, 319-329.
21. Wyer, R.S. (1974), Changes in meaning and halo effects in personality impression formation, *Journal of personality and Social Psychology*, Vol.29, No6, 829-835.

Context as a Key Concept in Information Demand Analysis

Magnus Lundqvist

School of Engineering, Jönköping University, P.O. Box 1026,
SE-551 11 Jönköping, Sweden
magnus.lundqvist@ing.hj.se

Abstract. It has been suggested that problems in information searching, storing, and processing related to information overflow could be reduced by introducing information systems based on the Information Logistics approach providing the right information to the right place at the right time. However, in order to do so an understanding of what the right information, time, and place are is necessary, i.e. the information demands motivating the system must be known. In this paper a definition of the term information demand as well as different aspects thereof are presented together with ideas on how they can be analyzed. It is advocated that the key to understanding such demands are to understand the context in which they exist and that the large part of such contexts can be derived from different types of Enterprise Models.

1 Introduction

According to the Delphi group some 70 percent of "business professionals" today feel that they spend too much time on finding and processing information and in many cases without ever succeeding in doing this, time that could be better spent on performing work. On top on this they perceive the software support for information retrieval as insufficient [1]. As work contexts tend to become more and more complex, document archives, guidelines, and process descriptions grow in size the perceived problem will probably increase. Not only is the right information needed, it is also needed to be found at a particular point in time and at the location and technical platform, the user happens to be at, at that time [2].

As a consequence of this the research field Information Logistics (ILOG) was established by various research groups, including the Swedish Center for Information Logistics (CIL)[1], Technical University of Berlin and the Institute for Software- and Systems Engineering (ISST)[2] within the German Fraunhofer Group[3].

---

[1] For more information please visit http://www.cil.se (in Swedish).
[2] For more information please visit http://www.isst.fraunhofer.de.
[3] For more information please visit http://www.fraunhofer.de.

1.1 Information Logistics as a Concept

The main idea of Information Logistics is to provide the right information at the right place at the right time. This results in three key aspects to be taken into account in Information Logistics [3]:

- *Content* – the users want just the information relevant for their specific situation. This means that an ILOG-system must be able to decide what is considered to be relevant content, i.e. select, aggregate and provide only the right content.

- *Time* – in order to provide content in a way that does not contribute to information overflow the system needs to provide the information just in time, here simply being defined as the point in time when the user wants the information.

- *Location* – an ILOG-system has to take into account the location of the user at the moment some particular information is needed. Depending on this the system needs to consider how information should be formatted and distributed to the user.

Furthermore users' demand for information may require several sources of content to be used in order to meet that demand and the users location also may dictate information demands (ID), e.g. information that normally would be considered relevant for a particular user may not be relevant if the user is not at a specific location. There's also the task of deciding what "the right time" is. It would be of little use to the user if the information were to be provided at any point in time. Even though the information might be both relevant and useful in itself this would still require the user to store the information for retrieval at a later time when the information actually is needed. Information provided too late is obviously useless from any point of view.

So far the main work in the area has been focusing on defining the area of information logistics as such and relevant components. This includes the design and development of a software framework [2] and the evaluation of this framework in real-life application cases such as Weather Information on Demand (WIND) [4] and Smart-Wear® [5]. Furthermore a lot of effort has been put on better representing different aspects of the above mentioned characteristics resulting in the introduction of the concepts Context [6] and Situation [7] to the area. These concepts embody different ways of capturing, representing and deciding such aspects as location, "right information" and "right time" etc.

Not only does all this work provide the necessary tools to build ILOG-applications but also it clearly states the users' and their information demands central position in such applications. The aim of the research done within the Information Engineering group at Jönköpings University is to develop a general method for analyzing the different aspects of users' information demand that cannot be automatically derived from different sensors and sources. It is expected that such a method would contribute to and improve the quality of Information Logistic systems as well as simplify the process of developing them.

1.2 Information Demand: the Main Idea

A general method for analyzing and modeling information demand needs not only to be general enough to work in an arbitrary application context. It also has to consider the highly dynamic nature of information demand as well as the fact that a lot of information about the users of an ILOG-system is needed in order to decide what information the system should provide and to whom.

Starting point for development of such a method should be a clear and unambiguous definition of the term information demand. In this paper such a definition is presented together with a motivation to why the concept of context is so important in this setting. This will support evaluation of the definition with respect to concepts present and situations constantly arising within small-scale business applications.

2 Information Demand Analysis

Even though quite some work has been done concerning information demand within the ILOG-area it has, up to this point in time, mostly been from an application oriented perspective rather than focused on general analysis and modeling methodologies. There exist some concepts within different areas, such as Information Retrieval, Data Mining, and Information Extraction that are similar to the concept of information demand and definition thereof presented here. Examples of such concepts are information requirements (analysis) [8], information need [9] [10]. Furthermore similar concepts are also used in areas such as business informatics [11] and social sciences [12]. Unfortunately none of these concepts is suitable for the purpose intended here since they only partly consider essential aspects of information demand like location and time. Common information retrieval tools, such as Internet search engines, can be taken as example to illustrate this. In those the time perspective usually is the time it takes to answer a request for information and then such a demand can be forgotten until the next time a user has it again. There of course exists exceptions, like search engines based on contextual information retrieval but even in those the time perspective does not exceed the user's current task, though such tasks might be stored in profiles for later use [13]. Furthermore, since the user actively visits and uses the search engine no more attention to the location aspect than the development of interfaces for different platforms is needed. In ILOG-systems the information need for the whole life time of the system has to be known as well as the user's location at any given time. To this the different types of demands for information is added. Yet another comparison between ILOG-systems and search engines can be used to illustrate the differences: while a search engine usually only needs to provide information that according to some method is deemed relevant based on a user's information request that always has to be possible to formalize in some query language. ILOG-systems need to handle demands for a wide range of situations rather than just one type or form of queries. Thus the need for a generic definition of information demand well adopted for information logistics solutions is quite clear.

2.1 Definition of Information Demand

As a general definition of information demand from information logistic perspective the following definition is presented:

Information Demand is the constantly changing need for current, accurate, and integrated information to support (business) activities, when ever and where ever it is needed.

This definition implies a number of aspects that must be considered when analyzing information demands:

- *Changing* means that the resulting models need to be able to capture the dynamics of information demand in order to reflect changes over time.
- *Current* and *accurate* requires some form of quality and relevance measurement.
- *Integrated* as well as *(business) activities* implies a need for awareness of the context in which the demand exists as well as some mechanism for understanding when a switch in context takes place.
- *When ever* and *where ever* states that the timing and location aspects of the information demand are important to analyze. Although not obvious, it also states that the demands as well as the possibilities to fulfill them may vary depending on time and location.

In the light of these implications it is assumed that to capture and model information demands it is needed to consider different dimensions of the demand and the reality in which it exists.

2.2 Context as a Dimension of Information Demand

The above clearly identifies the complexity of information demand as a concept. It has been proposed that this complexity can be handled by breaking down the concept into several different but interconnected dimensions [14] illustrated in figure 1.

To be able to support (business) activities and provide integrated information it's necessary to capture and evaluate information about these activities. Thus the concept of context shown in figure 1 is not only the largest, but also considered to be the most important aspect of ID since this defines the settings in which the users' information demands exist. That is to say that providing the "right information at the right time and place" entails that it should be right given the demanding party's context.

It is also important to notice the use of the term business activities above because even though it certainly might be relevant to speak about information demand from the perspective of individuals in such contexts as family life, various spare time activities, etc. it is reasonable to assume that for the most part these contexts are too informal and random in nature and the information demand less critical and frequent for any information demand analysis based on the ideas presented in this paper to be meaningful or useful. As a consequence of this the focus in this paper is on information demands within a business application context.

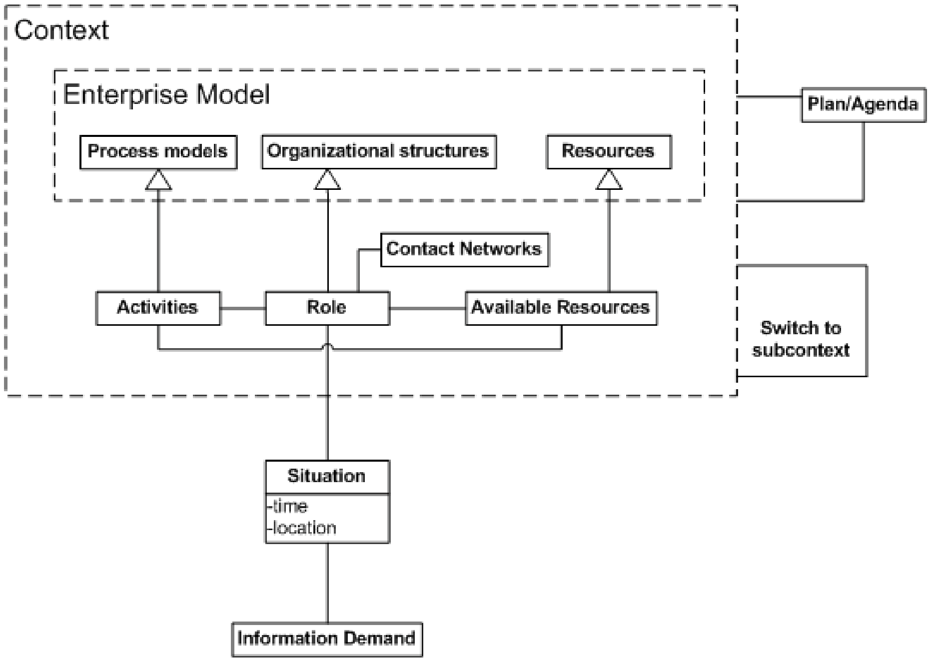

Fig. 1. Information model describing dimensions of Information Demand

There exist many different definitions of context in areas such as ubiquitous computing, contextual information retrieval, data warehousing, decision support systems etc. but for the purpose of information demand analysis context is here simply defined as:

An Information Demand Context is the formalized representation of information about the setting in which information demands exist and is comprised of the organizational role of the party having the demand, work activities related, and any resources and informal information exchange channels available, to that role.

From this definition several important concepts can be identified and the most central of them all from an information demand perspective is that of role, thus when context is mentioned here it is considered to be the context of a particular role. It could be argued that it is equally relevant to speak of an information demand as related to a specific activity and that some resources are necessary to perform some activities no matter who performs them but role is nevertheless the one concept that, as it will be shown here, in a natural way interconnects the others.

One can of course imagine situations where there exist overlap between roles, when one individual has several different roles, or even when several different individuals share one role. However, in its simplest form role can be described as a part of a larger organization structure clearly defined by the responsibility it has within that structure. Associated with that role are a number of activities that fall within, as well as define, the responsibility of that role, i.e. a role is also defined by the activities it performs within an organization. Furthermore there are a number of different resources available to a specific role which can be utilized to perform activities. Such resources might be anything from supporting information systems to a particular device used for some reason in such a performance and are dependent of both the role as well as the activity to be performed, i.e. not all resources are available to all roles within an organization and not all resources are suitable to use for all activities.

Finally it is proposed that an information demand context incorporates yet another important concept, not acknowledged enough in today available analysis methodologies. In the model depicted in figure 1 above this concept is named contact networks and describes the informal information exchange channels that always, at least to some degree, seem to exist between peers despite not being based on, or formally represented in, any organizational structures, process descriptions, or flow charts. Such informal structures might be based on anything from personal networks, the comfort of turning to other individuals with whom a common interests or the same educational or demographical background is shared. Even seemingly futile things as the layout of an office landscape might be a reason to the existence of such structures. These structures are in their very nature tacit and personal and are brought into an organization by individuals but since the model above does not include humans but rather the formalized view on humans as roles within an organization such structures are in the model connected to a role.

It is also important to emphasize that there is not always exactly one and only one context for every user. In fact every user may, as already mentioned, have a number of different contexts, to some degree unique to that user, between which he is constantly switching. Contexts may at times even overlap with each other. Examples of such contexts are different roles within an enterprise, associations, or even family, all with their different and unique demands for information. Moreover, within a context there may be sub-contexts that change everything from the role of a user to the activities to be performed [14]. Also worth noticing is the granularity of an information demand analysis. So far in this paper the perspective has been the one of an individual role but might just as well be that of an organizational unit, a division or an enterprise since they too have information demands related to some role within a larger organizational structure, perform activities, and have a variety of resources to their disposal. Even though the nature of the information demand of an organization might be inherently different than that of a single individual due to the differences in role and its responsibilities within a larger organizational structure the context definition is still assumed to be valid. As a consequence of this it can be stated that an individual role's context is just a sub-context to the context of the organization that role belongs to.

2.3 Deriving Time and Locations from Context

Once the context has been used to determine the actual subject or type of information demanded there are still two other important dimensions to consider, time and location, i.e. to what place and/or platform, and at which point in time, the information is needed. In order to do this the concept of situation is utilized, according to the definition that was introduced to the ILOG-area by Meissen et al. [15]. For the purposes presented in this paper a situation is simply put just a way to describe or represent what the user is or should be doing, when he is doing it and where he is doing it. Situations are derived from and belong to a specific context, even though the same situation can occur in several contexts. The different situations together with the contexts determine a user's information demand.

3 Deriving Context from an Organization

Deriving contextual information can, as already stated, be done in many different ways and from many different sources. One could imagine interviews with different roles within an organization, work- or information flow analysis, various kinds of process modeling methodologies and so on. Based on the background given in this paper it is here proposed that one such suitable source for deriving the contextual information necessary for information demand analysis is some form of Enterprise Models.

3.1 Enterprise Models as the Source of Information Demand Context

Enterprise Modeling (EM) has been described as the art of externalizing enterprise knowledge. This is usually done with the intention to either add some value to an enterprise or share some need by making models of the structure, behavior and organization of that enterprise [16]. The motivation often given to why enterprises should be analyzed and modeled is that it helps, and to some degree is a prerequisite for, better management, coordination and integration of such diverse things as markets, processes, different development and manufacturing sites, components, applications/systems, and so on as well as contributes to an increased flexibility, cleaner and more efficient manufacturing etc. Such models usually include [16]:

- Business processes
- Technical resources
- Information flow
- Organizational structures
- Human resources

There might be other aspects like financial flows, decisions, goals, product information etc. but those listed above are the essential ones. If business processes are considered to be sequences of activities, technical resources and information to be resources, organization to be the structure in which roles (humans) can be identified it also corresponds well to different aspects of context presented in section 2.2.

A wide range of different methods or architectures for enterprise modeling exists today ranging from the IDEF-family [17] of methods to such complex frameworks as GERAM (Generalized Enterprise Reference Architecture and Methodology) [18] and CIMOSA (CIM Open System Architecture) [19]. This has introduced the problem of many different and interoperable tools on the market and thus resulted in the development of unified modeling language like UEML (Unified Enterprise Modeling Language) [20]. However, for the purpose of deriving information demand contexts the choice of languages is not important as long as the models can be understood and analyzed.

3.2 Context Derivation: an Example

As a very small example on how information demand context may be derived from an EM a simple but classic example from the UEML-initiative, the PC-installation scenario represented in EEML (Extended Enterprise Model Language), has been borrowed [20]. The process model can then repeatedly be broken down into smaller and smaller tasks until a suitable level of analysis is received. For the purpose of context derivation a suitable level has already been suggested in the concept of role. In figure 2 a process has been broken down to a sub-process comprised of all those tasks, or activities in terms of context, that may be expected to be performed by one single role, say a computer technician, within the organization.

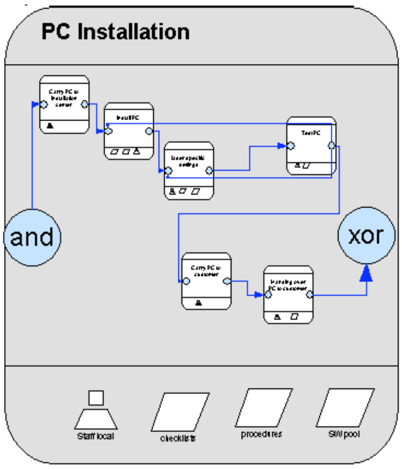

Fig. 2. An EEML task comprised of several sub-tasks

For each of those activities the different types of resources available and/or necessary for performing that particular task and some of the information needed to do so can be easily identified from the model. Of course a lot of additional information like goals, conditions for evaluation of the fulfillment of such goals is captured in these models. In some cases it might even be possible to derive timing aspects from them but since all this is outside the scope of information demand contexts no further attention is given to these aspects here.

However, since such models contain processes, resources as well as organization structures and thus almost all of the aspects of an information demand context presented in this paper obviously can be derived from such models it can be argued that they are nothing less than formalized information demand context descriptions. These models are however not complete context descriptions due to their failure to capture the informal structures discussed in section 2.2. One could easily imagine the computer technician used in the example above asking his or hers colleague sitting in the next office down the hall for information on some subject despite the fact that the model does not express such an information flow between those two roles in the organization. In order to capture such informal information exchange channels or information of a more tacit nature that not easily are represented in information systems either extensions to EM or the use of other methods is necessary.

While the EM approach certainly seems to be a suitable for deriving information demand contexts this requires enterprises structured enough to actually have models. For semi- or non-structured enterprises it is reasonable to believe that such models are less likely to exist and hence make this approach unsuitable in these cases.

4 Summary and Outlook

Problems in information searching, storing, and processing related to information overflow are today well recognized within a wide range of different areas. It is suggested that such problems could be reduced by introducing information systems based on the Information Logistical ideas of providing the right information to the right place at the right time. However, in order to do so an understanding of what the right information, time, and place are is necessary, i.e. the information demands motivating the system must be known. It is advocated that the key to analyze such demands is to show ample consideration to the setting in which they exist. This could, as it has been shown, to a large part be done with Enterprise Models as a starting-point.

This approach is however not completely without problems since far from all enterprises have, or even can have, such models describing their build up and activities. Especially many Small- and Medium Sized Enterprises (SME) are known to ignore or be unaware of EM as a useful tool [16] which may limit the applicability of the approach since SME make up for a large part of all enterprises having information overflow related problems. Furthermore some additional shortcomings in this approach are due to the fact that EM does not take any informal information exchange channels that might be present within an enterprise into consideration.

In order to solve problems like this and to evaluate ideas presented here a series of empirical studies have to be, and are being, performed focusing specifically on information use within SME and informal information exchange channels. Furthermore, work done on the subject presented in this paper has so far focused more on information use, information demand, and ILOG rather than context as a research area and the interrelationships between the different concepts presented in this paper. This is of course something that has to be attended to in the near future.

To define an analysis methodology is only considered to be the first step towards sufficient and efficient information demand analysis. The main objective of the research ideas presented here is to utilize such a methodology for the collection of enough knowledge about these demands to be able to derive information demand patterns, which are believed to be the best way to shorten and simplify the analysis necessary. Future work will therefore also include the application of the methodology on several enterprises as well as the evaluation of any defined patterns.

References

1. Delphi Group: Perspectives on Information Retrieval. Delphi Group, Boston, US, October 18, 2002. Available via http://www.delphigroup.com.
2. Deiters, W., Löfeler, T., Pfenningschmidt, S.: The information logistics approach toward a user demand-driven information supply. Proceedings CMSD 2003.
3. Mosnik, G.: Introduction to information logistics. Växjö University, School of Mathematics and System Engineering, Växjö (2002)
4. Jaksch, S., Pfenningschmidt, S., Sandkuhl, K., Theil, C.: Information Logistic Applications for Information-on-Demand Secenarios: Concepts and Experiences from WIND Project. Proceeding Euromicro conference, Antalya, Turkey, Septemper 2003, IEEE Computer Society Press (2003)
5. Deiters, W., Heuwinkel, K.: SmartWear: A personalized wearable information broker. Proceedings ICEWES Conference 2002, BTU Cottbus, Germany (2002)
6. Haseloff, S.: Context Gathering – an Enabler for Information Logistics. In Deiters W., Kutsche R., Sandkuhl K. (Eds.): Knowledge Supply and Information Logistics in Enterprises and Networked Organisations. Proceedings Intl. Workshop, associated with Multikonferenz on Wirtschaftsinformatiok, Essen (2004)
7. Meissen, U., Pfennigschmidt, S., Sandkuhl, K., Wahnfried, T.: Situation-based Message Rating in Information Logistics and its Applicability in Collaboration Scenarios. Accepted Paper, Euromicro 2004 Special Session on "Advances in Web Computing", August 31-September 3, 2004.
8. Winter, R., Strauch, B.: Demand-driven Information Requirements Analysis in Data Warehousing. Proceedings of the 36th annual Hawaii International Conference on System Sciences (2003) 231–240
9. Liu, R., Lin, W.: Mining for interactive identification of users' information needs. Information Systems, Vol. 28 (2003) 815–833
10. Cole, C.: Intelligent information retrieval – diagnosing information need. Part 1. Information Processing & Management, Vol. 34, No. 6 (1998) 709–720
11. Fuellhart, K., Glasmeier, A.: Acquisition, assessment and use of business information by small- and medium-sized business: a demand perspective. Entrepreneurship and Regional Development, Vol. 15, No. 3 (2003) 229–252
12. Moore, N.: A model of social information need. Journal of Information Science, Vol. 28, No. 4 (2002) 297–303
13. Challam, V., Gauch, S.: Contextual Information Retrieval Using Ontology-Based User Profiles. http://www.ittc.ku.edu/keyconcept/publications/vishnu.pdf
14. Lundqvist, M., Sandkuhl, K.: Modeling Information Demand for Collaborative Engineering. Post-conference proceedings of 2nd Intl. Workshop on Collaborative Engineering, Stara Lesna (2004)
15. Meissen, U. et al.: Context- and Situation-Awareness in Information Logistics. Post proceedings of Workshops of the International Conference on Extending Database Technology (EDBT), LNCS, Springer-Verlag, Berlin Heidelberg (2004)

16. Vernadat, F., B.: Enterprise Modeling and Intergration (EMI): Current Status and Research Perspectives. Annual Reviews in Control 26 (2002) 15–25
17. IDEF-I-CAM Definition. Knowledge Based Systems Inc (2000)
18. General Enterprise Reference Architecture and Methodology, version 1.6.3. IFIP-IFAC Task Force on Architectures for Enterprise Integration (1999)
19. CIMOSA Technical Baseline – Part 1 Version 3.2. CIMOSA Association. Private publication (1996)
20. Berio, G. et al.: Deliverable 3.1 – Requirement analysis: initial core constructs and architecture, Annex 2. Project UEML, University of Torino, Torino (2003)

Context Aware Mobile Devices and Reconfigurable Computing

Timothy O' Sullivan

Computer Science Department,
University College Cork,
Ireland
t.osullivan@cs.ucc.ie

Abstract. The ubiquitous computing vision of "everywhere, anytime" introduces performance and versatility issues for future personal mobile devices. These handheld computers will be required to be multi-functional and will be expected to execute a broad range of compute intensive applications. These devices are currently physically constraint by their size, their computational power and their networking ability. These constraints will challenge future personal mobile devices in achieving their multi-functional, high-performance requirements and will jeopardise their overall role within the ubiquitous computing vision. Our work combines the research paradigms of reconfigurable computing and context aware agent technology to improve the versatility and performance of future personal mobile devices. Reconfigurable logic allows a handheld computer provide application diversity and maintain performance levels whilst remaining portable. Context aware agents provide an intelligent middleware enabling mobile devices to fully derive the benefits of reconfigurable resources. An architectural framework currently being deployed and tested as well as overall future objectives are outlined.

1 Introduction

Future next-generation mobile devices are expected to deliver a broad range of multi-functional, compute-intensive applications [1]. The ubiquitous computing vision of "everywhere, anytime" is driving this expectation. Researchers need to respond to these market pressures through the creation of more proficient design and deployment methodologies for personal mobile devices. Their approaches will however have to account for the traditional limiting constraints of mobile devices such as physical size, computational power and networking ability.

This position paper outlines our research focus and proposed solutions to enabling future personal mobile devices meet their performance and versatility requirements. Our proposal incorporates reconfigurable computing and agent technology into the environment of a mobile device. Reconfigurable logic enables a resource constrained mobile device to dynamically adapt its hardware. This empowers a device to deliver a range of computationally-intensive applications whilst still maintaining a level of high performance. This reconfigurable hardware can be incorporated both as an integral component of a portable computer's architecture and as an external resource (e.g. an adaptive server) that is accessible over a network.

Agent technology provides a sophisticated middleware framework that can support mobile devices in utilising surrounding reconfigurable resources. Context awareness is a key component in enabling agents to effectively exploit these reconfigurable resources. A context aware agent has enhanced decision-making ability allowing for intelligent and proactive utilisation of resources.

Agents are particularly well-suited to wireless networks as they are efficient in their use of bandwidth and they can deal with intermittent network connections. Additionally, an agent can effectively represent, communicate and work towards a user's preferences and interests.

In the rest of this paper, an examination of related work is presented in section two. The third section presents an overview of an architectural framework for context aware reconfigurable mobile devices. The goal of this framework is to enable future mobile devices meet their performance and versatility requirements. The section highlights proposed context-aware deployment strategies incorporated within the architecture. The technologies utilised to realise the framework of the overall system are also detailed. Section four depicts an experimental prototype. Finally, section five concludes with an outline of future research.

2 Related Work

The issue of mobile devices meeting their future performance and versatility requirements is increasing in importance within the embedded systems co-design community. Reconfigurable computing is seen by many as a realistic solution to this growing problem. Field Programmable Gate Arrays (FPGAs) are an enabling technology of reconfigurable hardware.

There have been a number of research efforts investigating the potential of integrating reconfigurable hardware into a physical mobile device. Reconfigurable technology is recognised as a viable implementation platform for portable multimedia-based embedded devices [2]. This work presents an embedded system called Cam-E-Leon which can dynamically modify its software and reconfigurable hardware functionality through network downloads. Configuration of the appliance is achieved through a traditional client-server and message-based TCP/IP architecture.

A flexible server management architecture is also identified as a key objective for networked embedded devices that contain reconfigurable hardware [3]. The proposed framework is implemented using Enterprise Java Bean (EJB) technology and has a system management API constructed with XML. This approach offers an application download on demand service and an on-the-fly update of hardware acceleration components.

There have also been a number of research efforts investigating the potential of integrating reconfigurable hardware into the environment of a client system (i.e. as an adaptive server). These research efforts focus on developing middleware solutions to support client systems in utilising adaptive servers to improve their system performance. An attempt to establish ubiquitous access to remote reconfigurable hardware has been previously outlined [4]. The objective of this work is to allow a network of reconfigurable hardware modules be transparently accessible by client applications through a JINI-based infrastructure.

Middleware capable of discovering under-utilised computing nodes containing reconfigurable FPGA-based accelerator boards has also been developed [5]. The proposed strategy for sharing remote reconfigurable resources extends an off-the-shelf job management system. This framework enables effective scheduling of client requests for access to remote reconfigurable hardware.

The field of medical research also recognises the potential benefits of adaptive server technology [6]. This work proposes agents as a middleware solution and details their ability to efficiently exploit networked adaptive servers.

The issues of system performance and versatility are two related issues that are especially prevalent for future personal mobile devices. This position paper asserts context aware agents and reconfigurable hardware as being key components to help address these issues. Reconfigurable logic can be utilised both as an integral component of a device's architecture or as an external resource within its distributed environment. Context aware agent technology provides mobile devices with a sophisticated middleware framework enabling them to effectively utilise reconfigurable resources within their environment.

3 Architectural Framework

Our research combines the paradigms of reconfigurable logic and context aware agent technology to improve the performance and versatility of future personal mobile devices. A focus of this work has been the development of an architectural framework. This framework currently encompasses two deployment strategies to help a future mobile device provide application diversity and maintain high-performance whilst respecting their resource constraints.

3.1 Context-Aware Deployment Strategies

An overview of the system architecture and the associated deployment strategies are presented in Figure 1. Each deployment strategy can be briefly described as follows:

1. *Context Aware Agent Offloading Computation to an Adaptive Server using a Location influenced Negotiation and Bidding Technique*

This deployment strategy enables a mobile device to offload performance intensive computations to neighbouring adaptive servers. Adaptive servers are valuable distributed resources that can improve the system performance and versatility of portable devices. The deployment strategy recognises that reconfigurable resources within adaptive servers are costly commodities that can be effectively utilised with agents to ensure satisfactory return on investment.

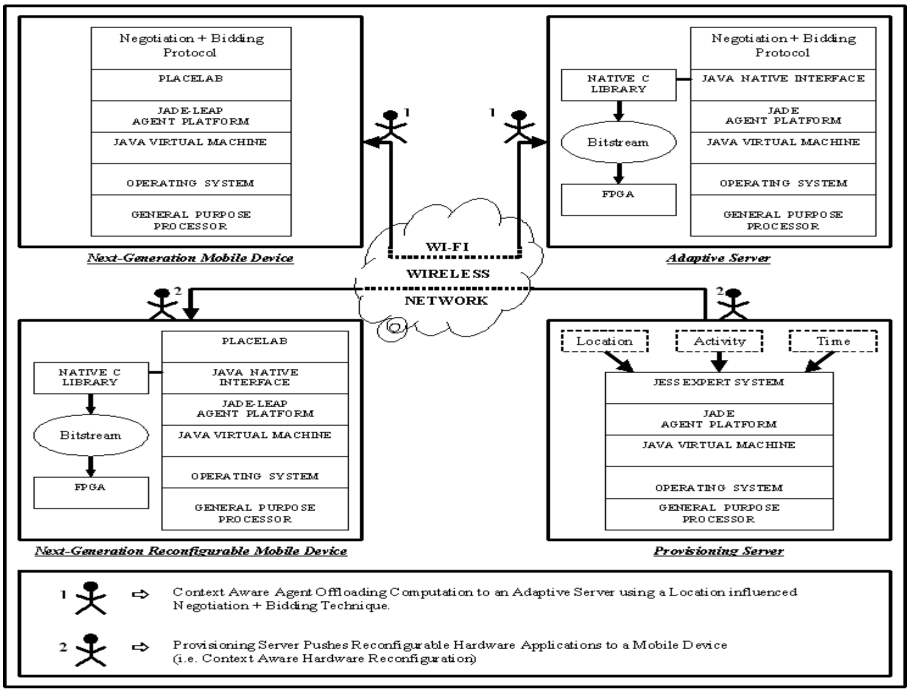

Figure 1: System Architectural Framework of Deployment Strategies

The negotiation strategy employs concepts which are based upon the contract-net protocol [7]. The proposed offloading protocol builds upon previous work by introducing the location of the portable device as an influencing factor in the adaptive server bid calculation [8]. This contextual information helps an adaptive server to identify the urgency of a computational request enabling the server to weight its bid calculation accordingly.

An example scenario of the potential benefit of this context-based offloading strategy would be within a distributed telemedicine environment. The location of a medical practitioner dictates the urgency of a computational request e.g. a request from a physician located in the emergency room of a hospital should have a higher priority weighting than a request from a physician in a ward.

The negotiation process is initiated by a mobile device through the communication of a call-for-proposals request to all adaptive servers within its environment. Each adaptive server submits a bid to execute the mobile device computational request. The bid is determined by examining their current queue of jobs and estimating the total service time. This examination of the queue takes into account the location of the current mobile device request. The priority level associated with the location of the incoming request dictates the placement of the potential computation within the adaptive server's queue of jobs. The result of the evaluation combined with an estimate of the time required to service the current computation request determines the adaptive server bid. The mobile device assesses all adaptive server bids to determine the best offload option.

The location aware aspect of the decision making process identifies the urgency of the user request. This helps to optimise the quality of service experienced by a mobile device user. The negotiation protocol also inherently encapsulates a load-balancing mechanism. This maintains a fair workload distribution amongst adaptive servers enabling the distributed system avoid both bottlenecks and under-utilisation of resources.

2. *Provisioning Server Pushes Reconfigurable Hardware Applications to a Mobile Device (i.e. Context Aware Hardware Reconfiguration)*

This deployment strategy seeks to enhance utilisation of reconfigurable resources contained within a mobile device. The strategy proposes using a context aware agent-based framework to push reconfigurable hardware applications to a mobile device. This configuration management strategy builds upon previous work by employing the concept of context in the decision making process [9]. Previous research focused on the dynamic provisioning of reconfigurable hardware based product updates to mobile devices.

The proposed deployment strategy increases the frequency of configuration by seeking to push reconfigurable hardware applications to a mobile device depending on the contextual environment of a user. This enables a portable device to have its hardware configuration dynamically tailored to the usage (i.e. application) requirements of a user.

The contextual elements required to enable effective configuration management are the location of the mobile device, the time of day, and the activity of the user. These elements are interpreted by a rule-based expert system operating on a provisioning server to determine if a reconfigurable hardware application should be pushed to a mobile device. The user activity is derived from a predetermined schedule of user appointments and their associated application preferences. An example scenario of this deployment strategy within a distributed telemedicine environment for a medical practitioner is shown in Table 1.

3.2 Agent Development Environment

The proposed architectural framework utilises JADE as the active agent platform on all provisioning and adaptive servers [10]. JADE-LEAP is the active agent platform on all mobile devices [11]. JADE (Java Agent Development Environment) is a Java-based open source development framework aimed at developing multi-agent systems and applications. JADE-LEAP (JADE-Lightweight Extensible Agent Platform) is an agent based runtime environment that is targeted towards resource constrained mobile devices. Both JADE and JADE-LEAP conform to FIPA (Foundation for Intelligent Physical Agents) standards for intelligent agents. FIPA is a standards organization established to promote the development of agent technology [12].

Table 1. Example Medical Scenario of Deployment Strategy for Context Aware Hardware Reconfiguration of a Mobile Device

Reconfigurable Hardware Application Pushed to Mobile Device	**Mobile Device Location**	**Time**	**User Activity**	**Reason**
Image Processing Application	Hospital Ward	9 AM	Patient Check-up	Practitioner analyses patient scans using an image processing application.
Speech Recognition Application	Office	1 PM	Taking Patient Assessment Notes	Practitioner takes notes on patients using a voice recognition system.
Video Processing Application	Home	7 PM	Personal Time	Practitioner likes to watch films on mobile device whilst relaxing at home.

3.3 Reconfigurable Technology

The proposed approach to executing a reconfigurable hardware based computation involves an agent interfacing with native C-libraries to dynamically manipulate FPGA circuitry. The native C-Library enables the agent to communicate and configure the hardware portion of its code to the reconfigurable logic. An agent can interact with the FPGA by controlling execution, obtaining feedback, and performing dynamic and partial reconfiguration. The device library provides an interface between an FPGA and the agent platform.

Hardware modules to be configured onto the FPGA are defined in Handel-C. This is a programming language built upon the syntax of conventional C that has additional parallel constructs to gain maximum benefit in performance from the target hardware [13]. Handel-C is used to produce an intermediate hardware format definition of the hardware algorithms (e.g. EDIF, VHDL). These are synthesised to a bitstream configuration for the target FPGA using Xilinx place and route tools [14].

3.4 Location Tool

A contextual element required for successful deployment of the proposed architectural framework is knowledge of the location of the mobile device. This is facilitated within the framework through the incorporation of Place Lab technology. This is an open source development project that uses a radio beacon-based approach to location [15]. An agent executing on a portable device can use the Place Lab component to estimate its geographic position. This is achieved by listening for unique identifiers (i.e. MAC addresses) of Wi-Fi routers. These identifiers are then cross-referenced against a cached database of beacon positions to achieve a location estimate.

3.5 Rule Based Expert System

The deployment strategy to push reconfigurable hardware applications to a portable computer is reliant on a rule based expert system to inform the decision-making process of agents on the provisioning server. Jess is the rule engine and scripting language employed within the framework [16]. This is a Java-based expert system that can be utilised to interpret and evaluate the contextual elements of a portable device to recommend hardware configurations.

4 Experimental Prototype

An experimental prototype is currently being developed to establish the viability of our approach in meeting the future performance and versatility requirements of mobile devices. The deployment strategy of enabling a portable device to offload computation to a neighbouring adaptive server using a location influenced negotiation and bidding technique has been implemented.

The prototype system is targeted towards telemedicine imaging applications which are computationally intensive for mobile devices. This prototype enables a physician to retrieve scanned patient images as shown in Figure 2. The left screenshot displays the options available to a physician in terms of their patients, associated scanned images and imaging algorithms that can be applied. The right screenshot shows a patient's original brain scan and a filtered edge-detected image created in real-time by an adaptive server.

Edge detection algorithms are used widely in medical practise to aid physicians in their patient analysis. The performance benefit of implementing an edge detection algorithm with reconfigurable hardware has observed an increase in speed of a factor of twenty in comparison with an implementation of the algorithm in software [17].

Figure 2: Medical Prototype Screenshots

The prototype environment consists of a Dell Axim PDA running a Pocket PC 2003 operating system and executing the JADE-LEAP agent platform using a Personal Java implementation of a Java Virtual Machine called Jeode.

The PlaceLab software plug-in resides on the mobile device enabling an accurate location estimate to be communicated to adaptive servers within a call-for-proposals computation request.

Four adaptive servers execute within agent containers on a high-end Pentium PC executing the JADE agent platform. They are connected to a Celoxica RC200 FPGA development board, enabling the execution of a reconfigurable hardware code portion of an offloaded computational request. Agents communicate between the mobile device and the adaptive servers over a Wi-Fi network.

5 Conclusions & Future Work

This position paper outlines our strategy of enabling future personal mobile devices meet their performance and versatility requirements. Our solution motivates the use of both reconfigurable computing and context aware agent technology.

This will provide handheld computers with a dynamic physical architecture (i.e. FPGA), a sophisticated middleware framework (i.e. context aware agents), and a high-performance support infrastructure (i.e. adaptive servers). This combination places future personal mobile devices in a favourable environment to help them meet their system performance and versatility goals. Two deployment strategies highlighting the potential of our work are briefly outlined.

Our future objectives include fully realising and evaluating the experimental prototype system. Additionally, there will be a further examination of additional services and benefits offered to future personal mobile devices through reconfigurable computing and context aware agent technology.

References

1. Fleischmann, J., Buchenrieder, K., and Kress, R.: Java Driven Codesign and Prototyping of Networked Embedded Systems. In Design Automation Conference (DAC), (1999)
2. Guccione, S., Verkest, D., and Bolsens, I.: Design Technology for Networked Reconfigurable FPGA Platforms. In Proceedings of Design, Automation and Test in Europe Conference (DATE), pp. 994-997, (2002)
3. Nitsch, C., Lara, C., and Kebschull, U.: A Novel Design Technology for Next Generation Ubiquitous Computing Architectures. In Proceedings of Reconfigurable Architectures Workshop, (2003)
4. Indrusiak, S., L., et al.: Ubiquitous Access to Reconfigurable Hardware: Application Scenarios and Implementation Issues. In Proceedings of Design, Automation and Test in Europe Conference (DATE), (2003)
5. Gaj, K., at al.: Effective Utilisation and Reconfiguration of Distributed Hardware Resources using Job Management Systems. In Proceedings of the Parallel and Distributed Processing Symposium, (2003)
6. Mapen, E., B., and Rosiene, J.: Integrating Adaptive Computing with Distributed Services Medicine. In Proceedings of 12th Annual IEEE Symposium on Computer-Based Medical Systems, (1999)
7. Smith, R.: The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. In IEEE Transactions on Computers, vol. 29, pp. 1104-1113, (1980)
8. O' Sullivan, T., and Studdert. R.: Handheld Medical Devices Negotiating for Reconfigurable Resources using Agents. In Proceedings of 18th IEEE International Symposium on Computer-Based Medical Systems, Dublin, Ireland, (2005)
9. O' Sullivan, T., and Studdert, R.: Configuration Management for Networked Reconfigurable Embedded Devices. In Proceedings of IEEE/IFIP Mobility Aware Technologies & Applications, Florianopolis, Brazil, (2004)
10. Bellifemine, F., Poggi, A., and Rimassa, G.: JADE – FIPA Compliant Agent Framework. In Proceedings of PAAM, (1999)
11. JADE-LEAP: http://sharon.cslet.it/project/jade
12. Foundation for Intelligent Physical Agents (FIPA).: FIPA Agent Management Specification, available at http://www.fipa.org
13. Handel-C Language Reference Manual, version 2.1, Celoxica Limited, (2001)
14. Xilinx Corporation: http://www.xilinx.com
15. LaMarca, A., et al.: Place Lab: Device Positioning Using Radio Beacons in the Wild. In Proceedings of Pervasive 2005, Munich, Germany, (2005)
16. Friedman-Hill, E.: Jess in Action, Rule-Based Systems in Java. Manning Publications, Greenwich, (2003)
17. Daggu, V., R., Muthukumar, V.: An Efficient Reconfigurable Architecture and Implementation of Edge Detection Algorithm using Handel-C. In Proceedings of International Conference on Information Technology: Coding and Computing, (2004)

Context-enhanced Ontology Reuse

Elena Paslaru Bontas

Free University of Berlin
Institute of Computer Science
Takustr. 9 14195 Berlin, Germany
paslaru@inf.fu-berlin.de

Abstract. Though the importance of knowledge sharing and reuse is well recognized by the Semantic Web community, ontology reuse is currently associated with major overheads, mainly because of the lack of methods and support tools to cope with the heterogeneity of the existing ontological sources. In this paper we present a thesis proposal which aims at developing methods to increase the productivity of ontology reuse. We introduce "ontology context" and a model for representing it, and briefly describe how contextual information may be used to evaluate source ontologies for reuse purposes.

1 Motivation

The Semantic Web[2] initiative envisions a new generation of the World Wide Web as a machine-processable network of semantically annotated documents and semantics-aware Web Services. For this purpose, it proposes standards for representation languages, which adapt well-established languages from the Knowledge Representation field to the requirements of the World Wide Web, and promotes the deployment of *ontologies*, regarded as a means for a shared knowledge understanding and a way to formally represent some domain of the world.
Ontologies are recognized as a key technology in various application domains on the emerging Semantic Web[6]. Though the definition of ontologies as "formal specification of shared conceptualization"[7] is generally accepted in the community, the current state of the art of the Semantic Web mirrors a different and sometimes contradictory understanding of the ontology concept. Existing ontologies are at most *formal specifications of conceptualizations*, but they are rarely built to be shared or reused.
An increasing number of academic and industrial initiatives formalize application knowledge using OWL and RDFS ontologies. Nevertheless the emerging ontologies seldom reflect a consensual or at least application-independent view of the modelled domain. These ontologies, though encompassing valuable amounts of knowledge, cannot be easily evaluated for and reused in new usage contexts. However, for an extensive dissemination of ontologies as envisioned by the Semantic Web community, both ontology designers and ontology users need a means to *understand, evaluate and assert existing ontologies*. This issue is vital especially for large-sized ontologies, describing knowledge-intensive domains (e.g. medicine,

biology, legislation). In this case the evaluation can not be performed completely manually due to the complexity and the dimensions of the modelled field.[1]
The general aim of the thesis is to analyze ontology reuse in order to identify critical factors that complicate, amend the quality or even guarantee the success of this process. The objective of the work is to develop a framework, which offers methods to increase the reusability of *existing ontologies*, by simplifying the user interface to these sources, and provides means to develop *new, more reusable ontologies*. The remaining of this paper is organized as follows: Section 2 gives a overview of our approach and introduces the concept of *ontology context* – a means to improve the quality of current reuse processes. Section 3 presents related work, which is relevant to the notion of context in computer science and the reuse issue in the Semantic Web field. The preliminary results of our work are subject of Section 4, while future work is discussed in Section 5.

2 General Approach: Improving Reusability through Context

This thesis proposal analyzes the difficulties related to *current* ontology reuse processes with the aim of providing methods to improve the reusability of available ontologies in user-defined application contexts.
For the achievement of these purposes we examined several use cases of reuse-oriented ontology engineering and identified critical factors which influence the ontology reusability[15].
The analysis of the use cases revealed that for the evaluation of an ontology w.r.t. its practical significance in a specific setting both ontology engineers and ontology users need not only experience in the application domain, but also information *about the ontology* itself: about its intrinsic properties (e.g. its vocabulary, the underlying graph topology and the representation language) and, with the same importance, about external features like its scope, history, purpose, authors or best practices of its usage in existing systems. Since several ontologies may describe the same domain in different ways, a successful evaluation and thus a successful reuse of an existing model depends also on the user-perceived capability to comprehend critical modelling decisions made during the conceptualization process or ontological commitments and assumptions.

We believe that the absence of such “contextual” information – which we herewith call *ontology context* – is a major obstacle to a wide-spread dissemination of ontologies. In a context-enhanced ontology reuse scenario (Figure 1), the ontology engineer has access to a repository of potentially relevant ontologies, which is additionally linked to a pool of *ontology context* information items. Every ontology corresponds to a set of context objects that, among other things, describe the roles and tasks related to previous ontology utilizations. Several types of

[1] In the same time the re-usage of existing sources in these domains is inevitable since a new implementation is associated with significant costs.

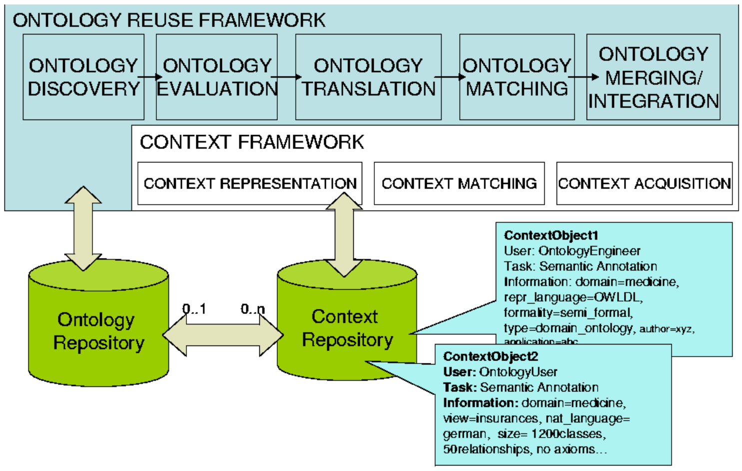

Fig. 1. Context-enhanced Ontology Reuse

contexts are defined for different groups of users (e.g. for ontology engineers, for domain experts) and Semantic Web resources (e.g. ontologies, parts of ontologies) , while parts of the information comprised by a context object are intended for automatic, for human processing, or both. Ontology context is represented in a pre-defined, machine-readable form: this defines a framework for the generation of context information for new ontologies, thus simplifying this task, and enables the refinement of specific *automatic* ontology engineering tasks such as *ontology discovery* and *ontology matching*. A valuable ontology repository is inconceivable without a usage-related classification of the sources and the associated search features. Both of these features can be realized if ontology context information is taken into account in the functionality of the repository. W.r.t. ontology matching, the ontology context is applied for example in a pre-matching phase in order to automatically select matching strategies fitting to the corresponding context information. A possible usage of context for evaluation purposes is briefly described in Section 4.3.

3 Related Work

3.1 Previous definitions of context

Although the importance of context and contextual reasoning has been often emphasized in different research communities, there is no common definition for this concept. Previous definitions of "context" have emerged in cognitive psychology, philosophy and also in particular areas of computer science, especially when dealing with natural language processing or automatic reasoning. Artificial Intelligence research adopted three main directions when dealing with context:

it can be either a situation in the general sense of the term, a part of knowledge or both of them[1, 13, 9]. However, since these definitions have proved to be hardly applicable in other computer science domains, numerous context-aware applications define context by enumeration of potentially application relevant information sources or by choosing synonyms for context (as in the area of mobile computing[19]). A primary concern in mobile computing is context-awareness, i.e. the awareness of the physical environment surrounding users and their mobile devices. Both in real time systems and mobile applications the focus is placed on the acquisition and processing of contextual information, while a clear definition of context is of secondary importance[5]. In the Semantic Web community the most frequent definitions regard this concept as a specific ontology standing for a domain of discourse, as a view upon a domain or as a mediation layer between ontologies – a position related to previous approaches of contextual/local reasoning. We conclude this section with an (incomplete) definition of context provided by the Free On-line Dictionary of Computing[2]: context is "that which surrounds and gives meaning to something else". This definition, without claiming to tap the full potential of the underlying concept, represents quite well the way humans think about "context" apart from any scientific considerations. Besides it outlines two aspects that are significant to classify information as "contextual".[3] The term *surrounds* reflects the relative nature of context. Context is information/knowledge which can be used *additionally* to perform some task. The second part of the definition mentions the term *meaning*, as a synonym for *relevance*. Since every available piece of information could be considered contextual[4], additional parameters are required to specify the *relevant* fragment. The current task and the relevance criteria imposed by the user of the context operate as filters to distinguish between *related* information and *relevant related* information about the context target. Some of the mentioned context definitions are used in our work in order to declaratively represent contextual information about ontologies and theoretically ground the underlying conceptual model.

3.2 Ontology Reuse

The difficulties related to ontology reuse are reflected by the relatively manageable number of research papers addressing this issue from a theoretical point of view, as well as from the absence of significant best practices or case studies. An analysis of the reuse process and the definition of its main stages are provided in [17, 25]. Case studies on this topic are described for example in [23, 18, 25]. Though most of the available ontology engineering methodologies mention the possibility of *reusing* existing knowledge sources in building new ontologies, the reuse and re-engineering of ontologies has been rather poor explored in this field. For example, Uschold and King[26] describe in detail how to build an ontology

[2] FOLDOC: http://foldoc.doc.ic.ac.uk/foldoc/index.html.

[3] These two parameters are not the only choice in the attempt of defining context, but they are significant in order to identify contextual information and to differentiate among types of contexts.

from scratch, but on the matter of reusing existing ontologies they only give a general explanation of the approach. Several methodologies address this issue only in the context of ontology customization: extracting relevant fragments of very comprehensive ontologies[22, 16].
Several research initiatives address subjects related to the reuse of ontologies such as *ontology evaluation*, *ontology comparison* and *metadata for ontologies*. Ontology evaluation is usually considered in conjunction with a specific application system[3]. Ontology comparison approaches rely mainly on basic syntactic features of ontologies e.g. number of classes or on matching algorithms taking into account the vocabulary and/or the taxonomical structure of an ontology[12]. Though the importance of metadata has been often emphasized by the Semantic Web community, a metadata scheme for ontologies is still subject of current work[10]. Ontology context as proposed in this paper and ontology metadata as understood by recent approaches do share a common goal, that is to offer additional information about ontologies. In our opinion the main difference between context and metadata consists however in the underlying information model. While metadata is a plain collection of useful information items about ontologies, the context model distinguishes between several types of information with different processing scopes and usages, with a particular focus on the task and on the user the corresponding contextual information is relevant for.

4 Preliminary Results

As mentioned before, the general approach of this thesis proposal starts with the analysis of typical reuse case studies, in order to identify requirements for a successful reuse process. This examination suggests that additional, usage-related information – contextual information – about ontologies can improve several stages of the reuse process. On this basis we develop a model to represent ontology contexts and analyze its usage for ontology evaluation purposes.

4.1 Analysis of the Use Cases

The goal of a first use case was the development of an ontology for the domain of Recruitment. The ontology was intended to be used in a Semantic Web job portal by allowing a uniform representation of job postings and job seeker profiles and semantic matching in job seeking and procurement tasks. Several taxonomies for the description of skills, classification of job profiles and industrial sectors have been developed by major organizations in these fields (e.g. the German Federal Agency of Employment) and thus are expected to be used at a large scale. Therefore using these standards was an essential requirement for interoperability purposes. The second use case aimed at developing an medical ontology for the domain of pathology. The ontology was used in a retrieval system to semantically annotate patient reports. The complexity of the application domain made the building of a pathology ontology from scratch extremely costly. Given the wide scale of medical ontologies already available – UMLS, Gene Ontology,

MeSH to name only a few – the focus of the ontology engineering process in this second use case was on reusing these ontologies in the new application setting.
Typically ontology reuse starts with the identification of knowledge sources useful for the application domain, which differ both in the represented content, and in the formalization (thesauri, XML-Schemes and DTDs, ER models, informal taxonomies, textual descriptions, OWL etc.), followed by their evaluation and integration[17]. An automatic integration of the source ontologies means not only the translation of the representation languages to a common format, but also the matching of the schemes[17]. According to the experiments the success of reusing ontologies is currently influenced by two factors: the means to evaluate and compare the candidate ontologies – a task which is semi-automatically, at most – and the methods to integrate the selected sources to the target ontology – a task which is envisioned to be performed automatically. Further on, both tasks are significantly influenced by the *complexity and understandability* of the ontology and by the *quality of the integration strategy.*
We identified features affecting the complexity or the user-perceived understandability of an ontology and classified them according to Stamper's semiotic framework in syntactical, semantic and pragmatic features [20].
Syntactical features offer quantitative and qualitative information about the ontology and its underlying (graph) topology. Examples of syntactical features include the number of concepts and properties for each class, the depth of an inheritance tree, the number of incoming properties, the number of concept instances, the average path length, the number of connected components. Since ontologies are published in an open network like the Semantic Web, it is also important to consider the links a particular ontology has to other networked information sources [14]. Finally, there is qualitative, representation language-dependent information like the representation language itself, the number of syntax constructs used and syntactical correctness (validity).
Semantical features are related to the formal semantics of the representation language and the meaning of the ontology content: i). consistency (as measured by a reasoner), ii). correctness (i.e. whether the asserted information is true), iii). readability (i.e. the non-ambiguous interpretation of the meaning of the concept names w.r.t. a lexicon, the usage of human-readable concept names), iv). level of formality (e.g. highly informal, semi-informal, semi-formal, rigorously formal[24]), v). type of model (upper-level, domain ontology, thesaurus etc.), vi). ontology domain (i.e. the modelled domain e.g. medicine), vii). representation paradigm (i.e. the class of representation languages w.r.t. expressivity such as a specific Description Logic), and viii). natural language (e.g. English).
Heuristic and pragmatic features refer to information about the history of the ontology, for example when, by whom and to which purpose it was developed, whether multiple versions are available or about the engineering process the ontology originally resulted from. The latter topic is relevant for ontology engineers intending to (partially) re-use the ontology within an information system: the original engineering methodology, tools used during the development process and the input information sources. As input information sources one can

mention extern ontologies which are, partially or in a modified form, included to the current ontology. Another example is a domain-relevant document corpus used by ontology learning programs.

4.2 A Model for Contextual Information

In order for the contextual information to be automatically processed, shared and exchanged on the Semantic Web one needs an explicitly representation of context and its information sources. On the basis of our considerations from the previous section we developed a core *context model* which can be used to specify and process the context information in a transparent manner.

The core concept of the context model – formalized in OWL – is the class Context referencing its target, an instance of the class SemanticWebResource and its user (instances of OntologyEngineer, DocumentAuthor etc.). Subclasses of SemanticWebResources are Ontology, WebDocument or even OntologicalPrimitive and GroupOfOntologicalPrimitives. For each SemanticWebResource one can define the relevant context information by means of OWL properties. For example, for the class Ontology we formalized the features mentioned in Section 4.1. Syntactic features are conceptualized as DatatypeProperties with an integer or a string range for the numerical and the qualitative information respectively. Semantic features are modelled as OWL classes or individuals. For interoperability purposes every instance of OntologyDomain references a topic in the Open Directory taxonomy[4], which was translated to OWL for this purpose. The class DomainType is used to define the generality levels of a conceptualization as in [8, 27]. By means of the class RepresentationParadigm and its individuals (e.g. some Description Logic) one can define the supported ontological primitives (e.g. supports existential constraints). General-purpose FormalityLevels are defined as in [24, 27]. The properties of different subclasses of Ontology such as thesauri, taxonomies or Semantic Web ontologies can be declared by means of OWL constraints. For example a taxonomy is formalized using a representation paradigm supporting concepts and is-a relationships and is usually semi-formal.

The pragmatic category contains classes such as OntologyTask, OntologyApplication, IndustrialSector, OntologyAuthor, EngineeringMethodology. For the categorization of ontology applications we use a modified version of the ACM classification.[5], while the industrial sectors relate to the NAICS taxonomy.[6]A separate ontology describes typical roles ontologies may have and tasks they may be involved in. We analyzed the state of the art of ontology-based information systems according to recent surveys on this topic [21, 11] in order to establish the range of purposes ontologies are currently used for and the roles ontologies play within these tasks.

The formal definition of the context information as an OWL ontology offers significant advantages: first it allows the automatic processing of ontology contexts

[4] http://www.dmoz.org
[5] http://www.acm.org/class/1998/
[6] North American Industry Classification System: http://www.census.gov/epcd/www/naics.html

in various stages of the ontology reuse process (see the next section for details about the usage of the model). Second it guides ontology engineers in annotating existing or newly built ontologies with this kind of useful information.
Certain fragments of the context model, especially those reflecting intrinsic properties of an ontology, e.g. representation language or number of classes, can be acquired (or computed) automatically. The remaining parts are supposed to be provided and maintained by humans. Nevertheless, the usage of pre-defined ontologies such as the Open Directory aims at assisting the manual acquisition process and in the same time assure a common basis for context matching operations.

4.3 Context Usage

As underlined before, the discovery, the evaluation and the semantic matching of mostly heterogeneous ontological sources can draw profit from the usage of ontology context. In the following we will focus on its usage for evaluation purposes. The remaining usage scenarios are discussed elsewhere.
Both ontology engineers and domain experts need context information to estimate the relevance of the source ontologies involved in the reuse process. Information about the original purpose and the projects/systems using a given ontology can give hints about its fitness of use in new application settings. For example a thesaurus is likely to have a different formality level as that *required* by a semantic retrieval application. An ontology whose concept names are English can hardly be used to automatically annotate French documents, while a textual classification can not be offhand translated into a formal Semantic Web taxonomy because of its modelling ambiguities. Further on an ontology engineer may trust ontologies originated from known authors or organizations or ontologies whose quality has been proven or certified in previous application settings before investing resources in trying to reuse it for her current application goals. In the examined case studies we have been directly confronted with such issues in our attempt to evaluate a large set of very different ontologies.[7] The evaluation framework using ontology context was thus derived from the empirical findings of our studies and has already proved its usability in these settings – it can provide a quick assessment of the relevance of the source ontologies to the application context. The evaluation procedure is adapted for two groups of users: ontology engineers and domain experts and consists of a weighted set of questions. A first category of questions aims to specify the application setting the final ontology is to be integrated in, while the ontology-related questions address mainly the ontology engineer to specify in detail the desired (technical) properties of the target ontology (a small excerpt of the questionnaire is depicted in Figure 2). The result of the selection is translated to a semantic query which is automatically matched against the context information readily available in OWL in the context repository. A list of relevant ontologies and the corresponding context

[7] We evaluated over 20 ontologies in the HR case study. Over 100 ontologies were taken into consideration in the medical example.

objects is presented for a final manual assessment. The usage of ontology context does not change the human-driven nature of ontology evaluation. However the features contained by the proposed context model have proven to be useful for humans to assist this decision making process. Furthermore, the availability of this information becomes fundamental for situations in which the ontology engineer is not familiar with the ontology to be evaluated or when this ontology is utterly too complex to be "read through" by humans.

Application setting	
Question	Alternative answers with ratings
Kind of system are you building?	Concepts of the ACM-Ontology
Industrial sector of your application?	Concepts of the NAICS Ontology
The task the target ontology will be used in?	Concepts of the TaskOntology
The role of the ontology in this task?	Concepts of the TaskOntology
Ontology	
The domain/view of the ontology?	Concepts of the OpenDirectoryOntology
The size of the ontology?	Integer
The natural language of the ontological primitives?	List of common languages

Fig. 2. Excerpt of the evaluation questionnaire

5 Conclusions and Future work

This paper presents first ideas towards a more efficient ontology reuse. The major part of the paper describes the analysis of two case studies on ontology reuse and their results towards a formal model to represent ontology context. Further on we briefly describe current work concerning the usage of ontology contexts in ontology evaluation tasks. Besides the work presented in this paper, we are currently investigating the usage of the presented context model in matching operations and are implementing a prototypical rule-based generic matching framework to demonstrate the usability of context for such purposes.
Future work includes the issue of context acquisition and an overall evaluation of the approach. Some ontology features can be extracted or computed directly, while features such as ontology domain, natural language, ontology type, sub-ontologies can be derived heuristically using additional information sources. A first step towards the evaluation of the results is intended to be the empirical validation and refinement of the context model in collaboration with a team of ontology engineers and domain experts with experience in the ontology engineering field. The evaluation of overall approach against other reuse approaches will be performed by comparing the costs arisen in the reuse process (i.e. person months efforts) and the quality of the outcomes (i.e. the target ontology).

References

1. J. Barwise and J. Perry. *Situations and Attitudes.* MIT Press, 1983.

2. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.
3. C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks. Data Driven Ontology Evaluation. In *Int. Conference on Language Resources and Evaluation*, 2004.
4. P. Brezillon. Context in Artificial Intelligence: I. A Survey of the literature. *Computer and Artificial Intelligence*, 18(18):321 – 340, 1999.
5. G. Chen and D. Kotz. Survey of Context-Aware Mobile Computing Research. Technical Report TR2000-381, Dartmouth College, 2000.
6. D. Fensel. *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. Springer Verlag, 2001.
7. T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43(5-6):907–928, 1995.
8. N. Guarino. Formal Ontology and Information Systems. In *Proc. of the FOIS'98*.
9. R.V. Guha. Context dependence of representations in CYC. Technical report, Microelectronics and Computer Technology Corporation, 1993.
10. KnowledgeWeb European Project. Identification of standards on metadata for ontologies (Deliverable D1.3.2 KnoweldgeWeb FP6-507482), 2004.
11. KnowledgeWeb European Project. Typology of ontology-based processing tasks (Deliverable D1.1.3 KnoweldgeWeb FP6-507482), 2004.
12. A. Maedche and S. Staab. Measuring Similarity between Ontologies. In *Proc. of the European Conf. on Knowledge Acquisition and Management EKAW02*, 2002.
13. J. McCarthy. Notes on formalizing context. In *Proc. of the IJCAI'93*, 1993.
14. Mark Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
15. E. Paslaru Bontas, M. Mochol, and R. Tolksdorf. Case Studies in Ontology Reuse. In *I-KNOW2005, 5th International Conference on Knowledge Management*, 2005.
16. B. J. Peterson, W. A. Andersen, and J. Engel. Knowledge Bus: Generating Application-focused Databases from Large Ontologies. In *Proc. of the KRDB*, 1998.
17. H. S. Pinto and J. P. Martins. A methodology for ontology integration. In *K-CAP 2001: Proc. of the International Conference on Knowledge Capture*, 2001.
18. T. Russ, A. Valente, R. MacGregor, and W. Swartout. Practical Experiences in Trading Off Ontology Usability and Reusability. In *Proc. of the KAW99*, 1999.
19. B. Schillit, N. Adams, and R. Want. Context-aware computing applications. In *Proc. of the Int. Workshop on Mobile Computing Systems and Applications*, 1994.
20. R. Stamper. The Semiotic Framework for Information Systems Research. *Inf. Systems Research: Contemporary Approaches and Emergent Traditions*, 1991.
21. SWAD European Project. Semantic Web applications - analysis and selection (Deliverable Project SWAD IST-2001-34732), 2001.
22. B. Swartout et al. Toward Distributed Use of Large-Scale Ontologies. In *Proc. of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop*, 1996.
23. M. Uschold, P. Clark, M. Healy, K. Williamson, and S. Woods. An Experiment in Ontology Reuse. In *Proceedings of the KAW98*, 1998.
24. M. Uschold and M. Grüninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), 1996.
25. M. Uschold, M. Healy, K. Williamson, P. Clark, and S. Woods. Ontology Reuse and Application. In *Proc. of the FOIS'98*, 1998.
26. M. Uschold and M. King. Towards a methodology for building ontologies. In *Proc. of the IJCAI'95, Workshop on Basic Ont. Issues in Knowledge Sharing*, 1995.
27. Y. Wand and R. Weber. Information Systems and Conceptual Modelling: A Research Agenda. *Information Systems Research*, 13(4), 2002.

Judgment as Mapping (JUDGEMAP2)

Georgi Petkov

Central and East European Center for Cognitive Science,
Department of Cognitive Science and Psychology,
New Bulgarian University, 21 Montevideo Street, Sofia 1635, Bulgaria
e-mail: gpetkov@cogs.nbu.bg

Abstract. This paper presents one model of judgment, based on the cognitive architecture DUAL. Its main assumption is that judgment is not a separate module, but is integrated with the other cognitive abilities. The model tries to explain some of the known phenomena in this field – both assimilation and contrast effects, the influence of the irrelevant features of the stimuli, the importance of the context. The model does not put any limitations over the type of the scale or the complexity of the stimuli. Several simulations and their results are briefly presented. Three of the simulations require judgment of length of lines on seven-point scale – lines that participate in uniform distribution, lines that belong to a skewed set, and lines that share different irrelevant characteristic. Two – dimensional stimuli was judged in the fourth simulation.

1. Introduction

"Please, judge the length of the line that you see on a monitor on ascending scale from one to seven." Such instruction illustrates in its explicit form the judgment task. However, people do not make judgments only in the artificial psychological laboratories. On the contrary, they judge every day – *'I like this man, I do not like that one'*, *'I prefer this cheese'*, etc. Something more, judgments cannot be separate from the overall cognition. They might take part in choice, in decision-making, in analogy-making, etc. This fact is important constrain for each theory of judgment. The phenomena have to be explained with abilities that are more fundamental for the human cognition. We began our work over the model from this starting point.

One disadvantage of most of the theories is that they do not focus on the process of judgment, but only describe the result – usually with mathematical equations. This is another reason to look for computational models based on larger architectures, and to understand better the relationship between judgment and the other abilities.

Another problem of the existing theories is that they usually separate the stimuli into simple and complex ones. Psychophysics analyze judgments of unambiguous stimuli like lines, sounds, etc, whereas the researchers from the field of the social psychology work on a very complex ones, like traits or emotions. Our assumption is that it is whole continuum between them and that the same basic mechanisms might be involved in both types of judgment.

The predecessor of JUDGEMAP2 – JUDGEMAP [11] – illustrated the role of the irrelevant dimensions of the stimuli in judgment. It, however, did not give satisfactory explanation of the assimilation effect, and did not work correctly with complex stimuli. Now, JUDGEMAP2 demonstrates, together the role of context and sources of the range and frequency effects [15], the sequential assimilation effect [13], and the ability to judge stimuli using unspecified number of dimensions.

2. Theories for judgment

The main theories of judgment can be roughly separated into three main groups.

The first point of view is that judgment is a process of *measuring the similarity/dissimilarity to a standard*. According to the classical ideal point approach [3,18], people have their "ideal points" and when judge they measure the distance between it and the stimulus. The Adaptation Level Theory [6] proposes that the standard (AL) dynamically changes, depending on the recent experiences. The Norm Theory [7] follows a similar approach, however, the standard here is called "norm" that is constructed on the spot rather than retrieved from long-term memory.

A second group of approaches looks at the judgment as a process of *classification*. Each rating forms a subcategory and the target stimulus has to be classified into one of those subcategories. The Range-Frequency Theory [14,15] postulates two independent constraints that such category subdivision should satisfy: the range of value variation within all subcategories should be about the same, and the number of examples in all subcategories should be about the same. ANCOR [16] is a computational model within this approach, based on ACT-R – it describes the process of learning of prototypes of these subcategories, their dynamically updates, and the process of classification of the target stimulus into one of them.

JUDGEMAP project assumes third point of view. The target stimulus is not compared to the comparison set, but is rather included in it and then a mapping between the elements of the comparison set and the set of scale elements is established. This mapping should satisfy as much as possible the specified in the instruction structural constraints – higher stimulus magnitudes should receive higher ratings, and almost equal differences between magnitudes should correspond to almost equal differences between the corresponding ratings.

3. The main goals and assumptions of the model

JUDGEMAP2 is modeled under the assumptions that judgment is not a separate process and cannot be distinguished from the other cognitive abilities. The mechanisms, underlying the model, seem to be fundamental for the cognitive system. The demonstration of the concrete data, received by various psychological experiments, should be not starting point, but rather should serve for verification of the model. Our goals were to define clearly some of the principles that we think are most important for the human cognition at all; to find out the most successful computational mechanisms that can model them; and to describe the experimental data at the level of those mechanisms.

Our main assumptions were that the human memory is associative and constructivists; that mappings between structures are in the core of the cognition; and that context is a key factor in all cognitive processes.

The model is based on *associative memory*. Such memory works flexible and fast in a natural environment [1]. When the system works on a certain item, it keeps the close associations of this item more active, because it is probable to face them soon. For example, if the task of a subject is to judge lengths of lines, it is useful to keep ready in the memory other similar lines, other lengths etc. Very important condition such type of memory to work is activation to *decay*, if no future irritations appear.

The belief that the *structural mapping* (and analogy-making) is not an isolated human faculty, but is fundamental for cognition [4,12], is another assumption of JUDGEMAP2. By mappings, the system dynamically integrates the new information with the old one, manipulates and adjusts either of them until they fit consistently. JUDGEMAP2 is able to capture small similarities in the structures and to map them. It is a pressure those mappings to grow up and to involve other agents. Just like in the process of crystallization, the system strives to a stable equilibrium, changing quantitatively itself.

Some theories understand judgment as precise measurement and look at the contextual influences as adding a noise. Rather, we assume that *context* is a necessary condition for a flexible and effective cognition. In order one system to be flexible, the set of all possible alternatives should be as large, as possible. In order to be effective, the set of the actually considered alternatives should be quite small. The context determinates the *relevant* paths for searching and solves this obvious contradiction.

4. Description of the model JUDGEMAP2

DUAL architecture [9,10] and AMBR model [12] combine all mentioned above features, thus are an appropriate base for building JUDGEMAP.

DUAL is a multi-agent hybrid architecture that is designed explicitly to model context sensitivity. The context is assumed as the state of the system in a given moment, i.e., the whole pattern of activities of the agents. There is no clear boundary between task and context; the behavior of the model changes continuously in response to the environment and to the given goals.

DUAL does not have any central executor and its global behavior emerge from the local interactions of a huge number relatively small micro – agents. AMBR1 [8] and AMBR2 [9] are models for analogy-making that combine the ideas for associative distributed memory and structural mapping.

Each DUAL agent has connectionists and symbolic part. From the connectionists point of view the system works like a localist neural network. Only a small number of agents, which activation level exceeds a certain threshold, form the Working Memory. The agents also have a symbolic part – each agent 'stands' for something – object, property, relation, hypothesis, etc. It also can perform very simple symbolic operations – can send or receive short messages to its neighbors, to modify its framework or to create new agents. The two parts interact in a very important way. The symbolic operations have a 'price', which is paid with activation. Consequently, the most active agents work faster; the less active ones work slower; and the inactive ones do not work at all. In the opposite, depending of the symbolic operations, new agents born and this changes the overall pattern of the activation too.

There are two special sources of activation in JUDGEMAP2 – INPUT and GOAL, representing respectively the perceptions and the goals of the system. The stimulus to be judged, together with the scale and possibly contextual elements are attached to INPUT node, whereas the relations, responsible for the target – the knowledge that stimuli with larger magnitude correspond to higher ratings – to the GOAL node. The comparison set is formed due to the spreading activation. In the WM enter other similar stimuli; relevant concepts, together with their prototypes (if such exist); both relevant to the task and irrelevant features of the stimuli. Recently judged stimuli also stay in the WM. The activation spread through concepts that are more general and back to their specific exemplars. However, there are only a few links from the concepts to their instances. The links to the recently used ones are created when the instances enter in the WM and their weights slowly decrease.

The other mechanisms, however, do not wait until the comparison set is formed. DUAL architecture assumes that cognition is continuous process. It is possible to work on several tasks simultaneously without any specified order. The same principles are used for the perceptions – each element can be attached to INPUT in any moment and can stay there unspecified time.

The first work of each instance-agent, when enter in WM, is to emit *marker*. The marker passing mechanism, like in the classical semantic network tradition, serves for searching paths between agents. Marker passing is symbolic operations and due to the main DUAL principles, its speed depends on the activation level – the most active agents work faster, the less active ones – slower, and the inactive ones do not work.

4.1 Comparison-relations

When receive marker from instances, the concepts create opposite links to them and transmit the marker to their superclasses and to the relations, which interest them. The *comparison-relations* form such type of relations (Fig.1).

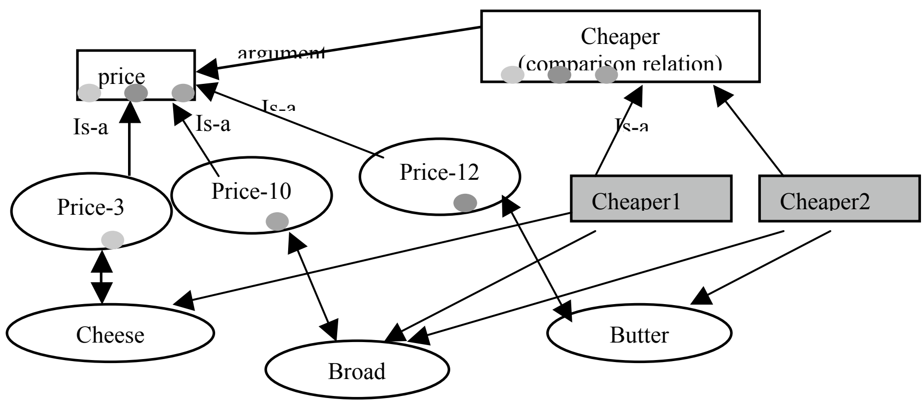

Fig.1 The work of the comparison relations. Comparison relations collect the received from their arguments markers, compare their origins and create new instances of the relation that they are responsible. The speed of this work depends on the activation of the participating agents. White colored ones are the permanent agents, grey colored – the temporal ones, created as a result of the work of the comparison relation ‘cheaper’. The markers and their ways are represented with small grey circles.

They are equipped with a special symbolic routine. They can compare two magnitudes and create new agents, responsible for the specific relation between magnitudes. For example, the comparison relation 'cheaper' wants from the concept 'price' to transmit to it all markers that receive. Suppose that the instances 'price-of-chese-1' and 'price-of-broad-2' enter in the WM, emit markers and those markers one by one come in 'cheaper'. The latter compares the magnitudes of both prices (if they are mentioned in their frame), and creates a new agent – for example 'broad-2-is-cheaper-than-cheese-1'. The new agents are incorporated in the main network and have full rights with the other agents (the only exception is that they die when depart from WM). Such temporal agents are constructed on the spot and they represent the add-hoc, contextual dependent knowledge.

4.2 Correspond-relations

Correspond-relations form another type of agents, responsible for creation of *hypotheses* for mapping. The correspond-relations are able to check whether some very simple rules are satisfied by the agents, for which the relation is aware. Those relations usually are goal – depended and enter in the WM through the GOAL node. For example, if the task is to judge lengths of a line on a numerical scale, one such correspond-relation agent represent the information that longer lines have to be judged with higher ratings. Now suppose that this agent know that 'line-1' was judged with 'rating-3' and that 'line-2' is longer than 'line-1'. This information makes a pressure 'line-2' to be judged with a higher rating. In such a case the correspond-relation agent chooses the most active one from the available ratings (for example 'rating-4') and creates a hypothesis that 'line-2' correspond to 'rating-3' (Fig2).

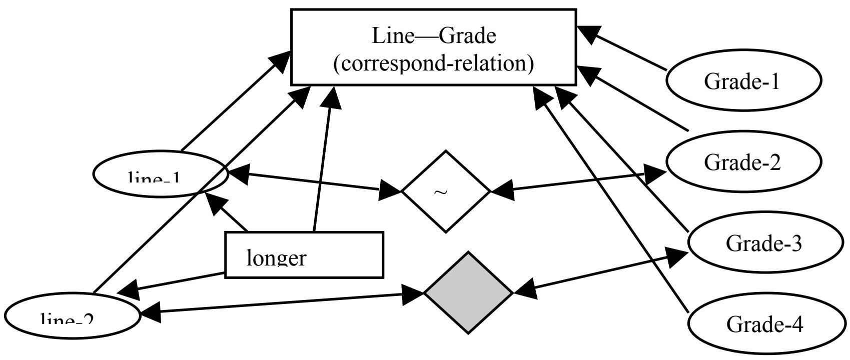

Fig.2. The work of the correspond-relations. Correspond-relations receive information from the active stimuli, ratings, and comparisons (via markers). They combine this information and create new hypotheses for correspondences (in gray). For example, if 'line-2' is longer then 'line-1' and 'line-1' is rate with 2, this is a reason to rate 'line-2' with the most active higher rating. Correspond-relations that are more complicated work with more than two stimuli (see text for details). The speed of the creation of hypotheses depends on the activation levels, and only a few of them are created during the time before the response.

In addition, there are also more complicated (and hence slower working) correspond-relations. For example, for judging lines, it is important to keep in mind also the knowledge that almost equal differences in the magnitude have to correspond to almost equal differences in the scale.

For different reasons and in different moments, number of new hypotheses for correspondence between the target stimulus and the ratings emerge. Each hypothesis receives activation from the two elements that connect, and from its justifications, i.e. the reason to be created. Some of these hypotheses duplicate each other and in this case, they combine their justifications; the controversial ones create inhibitory links between them. In this way, in parallel with the other processes, a constrain satisfaction network is formed. It is interconnected with the main one and their members participate in the global activation spreading process. The system gives response when a certain hypothesis about the target wins its competitors. Then it receives the next stimulus without any reloads and continues with its judgment.

5. Verification of the effects, observed in the experiments

Numbers of experiments have demonstrated that the contextual influence in judgment cannot be viewed just as a stochastic noise. In a certain circumstances, deviations of the ratings from their initial base level are systematic and robust. One such effect is the so-called sequential assimilation – the tendency the ratings to shift toward the ratings of the previous judged stimuli. The effect has been observed form various authors and has been demonstrated with extremely different types of stimuli – from loudness of sounds [13], through lengths of lines [16], to attractiveness of woman faces [17]. The source of this effect is not clearly understood, but there are evidences that assimilation occurs at all levels of the information processing. Goldstone [5] demonstrates it at the level of perception, Lokhead [13] shows with series of experiments that it occurs also at higher levels – reasoning and response.

Unfortunately, JUDGEMAP2 does not deal with the level of lower perceptions. It, however, assumes that assimilation is a size effect of the natural and fundamental process of spreading activation. The scale is represented with number of ratings, interconnected each other. At each fixed moment of time, a small part of the scale is most active – usually around the ratings of the recently judged stimuli.

The final decision depends on the activation level of the competing hypotheses. The latter depends on their activation. Each hypothesis receives activation from two main sources – its justification and its arguments. Whereas it is clear that the final decision should depend on the justifications (reasons for its creation), it is not so obvious that it may depend also on the activation of its arguments. This dependence, however, can be better illustrated in analogy-making. As more salient (active) is a given item, as salient are its corresponding analogies.

The opposite contrast effect also is robust. Parducci [15] demonstrated that when a skewed set of stimuli is given for judgment, there is a tendency the ratings to shift in direction, opposite to the skew. For example, if in the set of lines dominate the short ones, than all lengths would be overestimated.

According to JUDGEMAP2, this tendency is a consequence of the properties of the mappings. Structural mapping involves pressure for one to one mapping. This is the reason inhibitory links to be created between hypotheses, which involve the same agent. This induces the tendency ratings to be used almost the same times.

Another interesting phenomenon is the influence of the irrelevant to the task properties of stimuli upon judgment. Experimental results shows that if the whole set of stimuli is uniform, according to the judged property, but is separated into two skewed subsets, according to a certain irrelevant dimension, the ratings depend on the value of that irrelevant dimension [11].

Comparison set is formed due associations from relevant, but also from irrelevant properties. Consequently, the memories, which share more characteristics with the target, would be more active in the WM. If happens such stimuli to be concentrated around one of the endpoints of the judged set, then the inhibitory links between competing hypotheses would cause a shift of the ratings toward to the other endpoint.

6. Simulations

Four simulations are presented to illustrate some of the abilities of the model.

The first one requires judgment of the length of line segments on a seven-point scale. The set of stimuli was uniformly distributed and the results illustrate the sequential assimilation effect.

In the second simulation, the set of lines was skewed and the results were compared with those from the first simulation. With synchrony with our expectations, contrast effect according to the direction of skew, was demonstrated.

The third simulation repeated the result, received by psychological experiments [11, 2]. Uniformly distributed set of length of lines was judged, but half of the stimuli were red and positive skewed, whereas the other half - green and negative skewed. As a result, it was a small difference between the ratings, given to lines with the same length, but different color.

In the fourth simulation, the model judged two-dimensional stimuli. It has to rate the preference of different cheeses, depending on their price and quality. Again the sequential assimilation effect appeared, as like the experimental results show.

6.1 Simulation 1

The main goal of this simulation was to illustrate the sequential assimilation effect, and to serve as a base for comparison for the second simulation.

We used a set of 112 lines. Each line was represented with two agents – one stands for the line itself, the other – for its property ‘length’. There was 14 different sizes – from 100 to 1400 with increment 100. Eight lines participate in each size group. The stimuli were given to the model in a random order, each one - immediately after judgment of the previous one. The scale was represented with seven agents standing for the ratings. The neighbor ratings were interconnected. They were instances of the concept ‘seven-rate-scale’, attached to INPUT. The correspond relation ‘longer<->higher rating’ was attached to GOAL. The other relations and concepts were in the long-term memory and were activated due the spreading activation. This procedure was repeated 15 times, simulating 15 individuals.

Results: The averaged relation between the lengths and ratings is illustrated with the dashed line on Fig.3, left panel, together with the results of the second simulation for a comparison. The overall correlation between length and rating was 0.549 (p<

0.01). The sequential assimilation effect was 0.292 (p<0.01), measured as a correlation between the given rating and the previous one.

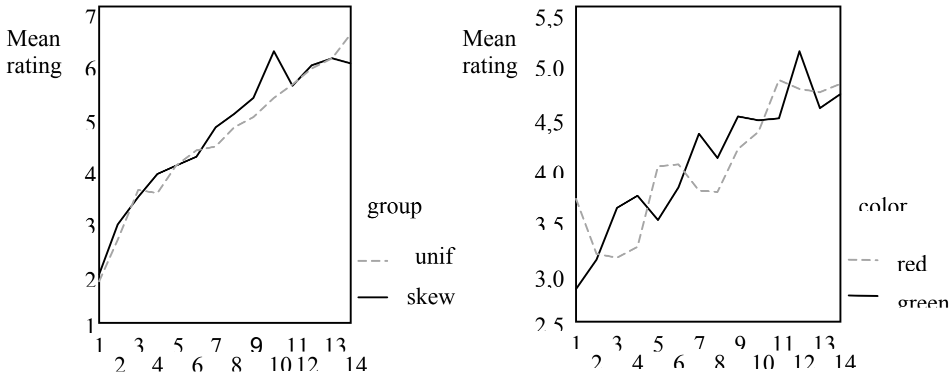

Fig.3. On the left panel: The averaged results from simulations 1 and 2. The lengths of the lines are on abscise, the averaged ratings – on ordinate. The dashed line shows the results from the uniform set (sim. 1), the solid one – from the positively skewed set (sim. 2).
On the right panel: The results from Simulation 3. The solid lines present the main rating of the green lines (positively skewed), the dashed one – of the negatively skewed red lines.

6.2 Simulation 2

The main goal of this simulation was to demonstrate the frequency effects [14,15]. When people judge skewed set of stimuli, they tend to shift their ratings in direction, opposite to the skew.

This simulation repeated the procedure from the first one, but with positively skewed set. The overall set consisted of 56 stimuli, distributed into 14 size groups. In the first two groups (lengths 100 & 200), there were seven lines in each, in the next two (lengths 300 & 400) – six lines, etc, in the last two (1300 & 1400) – one line.

Results: The averaged relation between lengths and ratings is illustrated with the solid line on Fig.3, left panel, together with the results of the first simulation. The ratings, given by each 'subject' for each length in each simulation were averaged and ANOVA analysis showed that the skew of the set is significant factor (p=0.047).

The overall correlation between length and rating was 0.557 (p< 0.01). The sequential assimilation effect was 0.337 (p<0.01).

6.3 Simulation 3

Third simulation replicates the experimental design, performed by Kokinov, Hristova, Petkov [11]. The lines differ in both length and color. The overall set of length of lines is uniformly distributed. Half of the stimuli were red and positive skewed, the other half - green and negative skewed. The experimental results showed shift of the ratings in direction, opposite to the skew.

Each line was represented with three agents – one for the agent itself, one for its length, and one for its color. All set consists of 112 stimuli, distributed uniformly by length – exactly like in the first simulation. Half of the stimuli, however, were green,

and the other half – red. The green subset was positively skewed – it consisted of the 56 lines, distributed by length exactly like in simulation 2. The rest of the lines were green and they formed negatively skewed subset. The procedure remained the same.
Results: The dependency between the lengths and averaged ratings for both groups is presented on Fig.3, right panel. The mean of the mean ratings of all red categories was 3.996, while the mean of the mean ratings of all green categories was 4.022. The difference of 0.026 was significant tested by ANOVA analysis ($p=0.033$). The results were consistent with the data, received by the experiment. The size of the color main effect is very small, but significant. The sequential assimilation was 0.248 ($p<0.01$).

6.4 Simulation 4.
The fourth simulation verified the model with stimuli with two relevant dimensions, and checked whether the assimilation effect still exists. The set consisted of 112 cheeses; each one had a certain quality and certain price. They correlated, just like in natural economical situation. The task was to judge the *rating* of each cheese. The concept 'better' was represented in the long-term memory as a superclass of two concepts – cheaper and more tasty (more quality). There was comparison- and correspond-relations for both of them.
Results: The correlation between the current and the previous stimuli (sequential assimilation effect) was 0.123 ($p<0.01$). It was no correlation nor between the prices and ratings, nor between qualities and ratings, because of their correlation.

7. Conclusions

The model for judgment JUDGEMAP2, based on the architecture DUAL, was presented. It is integrated with the AMBR model for analogy-making and inherits their basic assumptions – context-sensitiveness is a necessary condition for human cognition; human memory is associative and constructive; analogy-making is not a isolated human ability, but is fundamental for the cognition.
JUDGEMAP2 consider judgment is based on mapping between the sets of stimuli and ratings. The target stimulus is included in the comparison set and the response is a result of structural mapping between them. The main pressure of this mapping is to preserve the order relations.
Four simulations with the model demonstrate respectively the sequential assimilation effect; the range and frequency effects, the contrast effect, produced by the influence of irrelevant to the task properties of the stimuli; and the capability to deal with more complex stimuli, when the task require to consider two relevant characteristics.
The model, however, is not without shortcomings. It is just a beginning of a hard work and trace out some long term goals – to integrate perceptions in the task; to involve much more complex stimuli for judgment; to build bridges to more cognitive abilities, like choice and decision-making.

Acknowledgments
I would like to thank prof. B. Kokinov and the AMBR research team for their continuous support and stimulating environment.

References:

1. Anderson, J.: The Newell Test for a Theory of Cognition. Behavioral and brain science, 2003, 26
2. Arieh, Y. & Marks, L.: Context Effects in Visual Length Perception: Role of Ocular, Retinal and Spatial Location. Perception and Psychophysics, 2002, vol. 64 (3), 478-492.
3. Coombs, C. (1964). A Theory of Data. NY: Wiley.
4. Gentner, D., Holyoak, K. & Kokinov, B.: The Analogical Mind. Cambridge, 2001, MA: MIT Press.
5. Goldstone, R.: Effects of Categorization on Color Perception. Psychological Science, 1995, vol.6 (5), 298-304.
6. Helson, H.: Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior. NY, 1964: Harper and Row.
7. Kahneman, D., Miller, D.: Norm Theory: Comparing Reality to Its Alternatives. Psychological Review, 1986, vol. 93 (2),136-153.
8. Kokinov, B.: Associative memory-based reasoning: How to represent and retrieve cases. In T. O'Shea and V. Sgurev (Eds.), 1988, Artificial intelligence III: Methodology, systems, applications. Amsterdam: Elsevier.
9. Kokinov, B.: The DUAL cognitive architecture: A hybrid multi-agent approach. Proceedings of the Eleventh European Conference of Artificial Intelligence, 1994. London: John Wiley & Sons, Ltd.
10. Kokinov, B.: The context-sensitive cognitive architecture DUAL. Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society, 1994. Hillsdale, NJ: Lawrence Erlbaum Associates.
11. Kokinov, B., Hristova, P., Petkov, G.: Does Irrelevant Information Play a Role in Judgment? In: Proceedings of the 26th Annual Conference of the Cognitive Science Society, 2004. Erlbaum, Hillsdale, NJ.
12. Kokinov, B. & Petrov, A.: Integration of Memory and Reasoning in Analogy-Making: The AMBR Model. In: Gentner, D., Holyoak, K., Kokinov, B. (eds.) The Analogical Mind: Perspectives from Cognitive Science, Cambridge, 2001, MA: MIT Press.
13. Lockhead, G.: Psychophisical Scaling: Judgments of Attributes or Objects? Behavioral and Brain Science, 1992, vol. 15, 543-601.
14. Parducci, A.: Category Judgment: A Range-Frequency Model. Psychological Review, 1965, vol. 72 (6), 407-418.
15. Parducci, A.: Contextual Effects: A Range-Frequency Analysis. In: Carterette, E., Friedman, M. (eds) Handbook of Perception. 1974, vol. II. Psychophysical Judgment and Measurement. NY:Academic Press, 127-141.
16. Petrov, A. & Anderson, J.: ANCHOR: A Memory-Based Model of Category Rating. In: L. Gleitman & A. Joshi (eds.) Proceedings of the 22nd Annual Conference of the Cognitive Science Society, 2000. Hillsdale, NJ: Erlbaum.
17. Wedell, D., Parducci, A. & Geiselman, R: A Formal Analysis of Ratings of Physical Atractivness: Successive Contrast and Simultaneous Assimilation. Journal of Experimental Social Psychology, 1986, vol 23, pp 230-249.
18. Wedell, D., Pettibone, J.: Preference and the Contextual Basis of Ideals in Judgment and Choice. Journal of Experimental Psychology: General, 1999, vol. 128, pp 346-361.

Context-based Community Support

Mario Pichler

Software Competence Center Hagenberg (SCCH)
Hauptstrasse 99, A-4232 Hagenberg, Austria
mario.pichler@scch.at,
http://www.scch.at/people/mario.pichler

Abstract. The aim of the paper is to outline a doctoral thesis investigating the question about how members of outdoor communities can be provided with community-specific services while carrying out community activities. Strong emphasis is laid on describing the background and motivation to justify the selection of the research problem. The research problem, derived subproblems, the applied methodology, and the expected contributions along with organization of work are presented.

1 Background and Motivation

The starting point for the work presented in this paper was the question about what kind of future mobile and ubiquitous applications (MUA) can provide added-value to their users. Mobile operators are concerned with this question, and this is especially true for Europe. Many consortia address this problem and aim at developing scenarios of innovative mobile applications or to lay the foundations for building them [1–5].

As a first step to contribute to this issue, we tried to identify the needs of mobile people through desk (literature) studies, field studies, development of prototypes, analysis of current available mobile devices with respect to their satisfaction of mobile people's needs, and an analysis of artifacts people use and interact with when being mobile [6]. The next step then was to develop a number of innovative scenarios of future mobile and ubiquitous applications that satisfy the discovered needs.

1.1 Experiences Gathered Through a Field Study

For one scenario, we experimented with some of the location-based services (LBS), offered by an Austrian operator, at the Linz Intl. Airport in December 2002. In fact, we did some research into so-called search services, and chose to search for restaurants and taxis in our vicinity. The study was repeated in October 2003. Details about the study can be read in [7].

Two main issues among the experiences we gathered along with our conclusions are:

1. The provided contents of the investigated location-based services were not comprehensive or correct, respectively. Hence, it seems to be challenging keeping information up to date or comprehensive if a central authority manages these data on a "global" scale.
2. The navigation paths on the used mobile device to get the information needed were quite long—we felt too long for mobile people. These navigation problems are caused by insufficient user interface design and/or the oversupply of offered services, some (or even most) of them do not make sense in a number of situations.

Fact was that we did not perceive real added value for mobile people by using the investigated location-based services. Based on the investigated shortcomings and our conclusions, the first of two major concerns of our initial research proposal, introduced in the following section, has been identified.

1.2 Initial Research Proposal

The first of two major concerns of our initial proposal for the dissertation, introduced in this paper, is the following:

Research concern 1 (RC1): *"Our first concern for the next generation of value added mobile applications is to provide mobile people with services that are very specific for the context area—an area representing a specific context—these people are in at a time. We refer to those services as **contextual services**."*

Context Areas and Wireless Context Area Networks Therefore, we introduced the notion of *wireless context area networks (WCANs)*. The idea of WCANs is to subdivide the environment into areas that represent a specific context and to provide only such information and services that are relevant for the respective context area—besides time and location-independent information and services which are usable everywhere. A possible set of context areas a human might meet during a workday is illustrated in Figure 1. The difference of our proposed contextual services to location-based services is that the scope of a context area not necessarily has to be constrained by physical borders. Instead, a *context area* can also be defined by logical boundaries. Context areas that are constrained logically can be defined through activities or tasks people are engaged in, like waiting on the tram station, driving on the highway, waiting in front of the concert hall and the like. More details about the WCAN idea together with further motivating facts can be read in [8].

At this point, we want to agree on the widely accepted importance of context awareness for the success of mobile and ubiquitous applications.

Support for Infrastructure-enhanced Communities Based on our investigations, we came up with a catalog of around twenty mobile application scenarios. As a next step, we analyzed the scenarios developed at that time. For

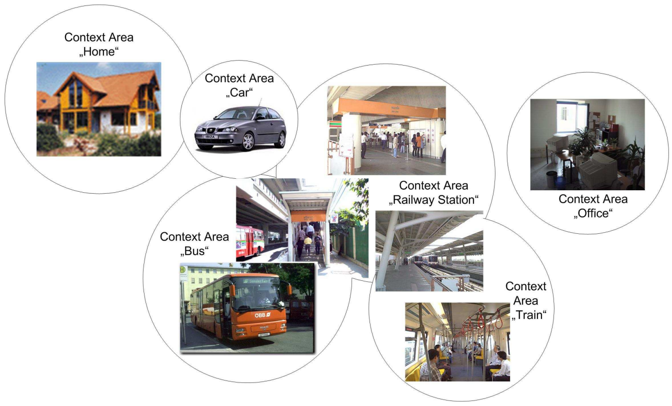

Fig. 1. Relevant Context Areas During a Workday

a number of scenarios, we identified that there is not only an information or service access need for one person. We also identified a need to get in contact or to stay in touch with people having similar interests or are carrying out similar activities (like-minded people) [9].

The following enumeration represents an excerpt of domains, for which application scenarios have been developed. Domains in which this need is of relevance are highlighted: *automobile association*, *traffic broadcasts*, carrying business, building companies, architecture, banking, *fair and event organizers*, *stadium operators*, mobile network operators, *tourism*, city council and local government, *public utilities* and e-Government, mobile professionals and *mobile teamwork*, supermarkets and building centers, *race track operators*, office and work place, service industry, and *public transport operators*.

The afore mentioned need goes beyond the possibilities offered by mobile telephony—the most ubiquitous mobile application—which primarily represents a one-to-one communication paradigm. As opposed to this traditional one-to-one communication paradigm, the communication requirements here reach from *one-to-many* to *many-to-many*, like it is known from chat-rooms and groupware systems [10].

Thereof, we derived our second major concern of our initial research proposal for this work:

Research concern 2 (RC2): *"Our second concern for the next generation of value added mobile applications is to support the networking of individuals to form communities. We envision communities being built of humans, who come together in physical proximity, reside in equal or similar situations (e.g. people waiting on train- or tram stations), or do have the same interests. Spoken more general these communities will be built of humans, who come together in the same context area. Further on, we will refer to those communities as* ***context-based communities****."*

In line with our first major concern to offer contextual services in context areas, offering context-based community services requires an understanding of the context of the users of those services. We assume that in context areas, as shown in Figure 1, basic assumptions of the context of people in the respective context area can be made. We assume that such environments can be made intelligent (or smart) in the way that detection of users' contexts and the provision of contextual services and context-based community services, respectively, is possible. The reason is that basic infrastructure for the realization can more or less easily be installed in those areas. At the time of writing, several projects are dealing with these issues (cf. [11, 12]).

Therefore, to go a step ahead of our initial proposal of primarily supporting communities in more or less infrastructure-enhanced areas, we see another promising application domain for context-based community support.

1.3 Support for Communities in the Wild

People are not only traveling between context areas as shown in Figure 1. People have special interests and are carrying out specific activities (hobbies etc.). Often these interests are satisfied and the activities are carried out in places (in the fields, woods, mountains) where infrastructure support as mentioned above is not available at all or only to a limited degree.

Thereby, we are primarily thinking of communities like hikers, alpine skiers, mountain bikers, cyclists, runners, horsemen, and the like (see Figure 2). They all have in common that the activities are performed outdoors (in the wild). Also, there are services for those communities available on the Internet (forums, chat rooms, newsgroups, portals for specific interests and activities etc.) which members of the respective communities use to share their experiences or route descriptions (e.g. GPS route descriptions), as is the case for hikers, mountain bikers, and horsemen for instance.

Fig. 2. A Sample of "Communities in the Wild"

These services are easily usable when sitting in front of a desktop PC. However, those services cannot be used when people are carrying out their activities.

Implications We notice that members of the above-mentioned communities are using technologies to stay in touch and to share artifacts with other community members. They are not able to do this when performing their "community" activities. Thereof, we derive potential for investigating *communities in the wild* as potential application domain for future mobile and ubiquitous applications. According to our first concern of providing contextual services to mobile people, we also here see a need for situation-specific (community) services.

However, detecting the context of mobile people in the wild seems to be harder than is the case with infrastructure-enhanced areas. In the former case, context detection and detecting the presence of other community members have to be done without infrastructure support. No single access point to community services—as is the case with the Internet in front of the desktop PC—is available here. Additionally, infrastructural services are available for community members in exceptional cases only.

Thereof, with the aim of contributing to the question stated at the very beginning of this paper, we derive the research problem that will be tackled by this thesis.

1.4 Problem Statement

Based on the motivations derived from the background presented above, the following problem is studied in this thesis.

Research problem (RP): *"How can members of outdoor communities be provided with community-specific services when carrying out their community activities?"*

This general question includes a number of subquestions that have to be answered. Therefore, this question is divided into the following subproblems.

Subproblems (SP):

(SP1): *"What kind of outdoor communities can benefit from the availability of community-specific services when community members are carrying out their activities?"*

This question deals with the identification of relevant communities and their (information and communication) needs.

(SP2): *"How can the presence (or availability) of other community members in the wild be detected?"*

Mobile people often are members of communities without explicitly noticing this fact, or, are not noticing the availability of communication means to get in touch with like-minded people. Because of that, they are not able to communicate with other members of the community or exchange information or other (soft) goods.

(SP2.1): *"How can the context of a person in the wild be detected?"*

The question here is, how the situation of mobile person (i.e. the situational context [13]) can be detected.

(SP2.2): *"How can shared contexts (i.e. being members of the same community) of several mobile people in the wild be detected (context matching)?"*

Other than using for instance the same chat room to talk to like-minded people on the desktop, the question is how to identify like-minded people when being mobile. Having detected the presence of other community members, community-specific information and communication services can be provided to the members.

2 Methodology

The following methodology is being applied in order to answer the research questions:

Identification of mobile people's basic needs. An understanding of the needs of mobile people will be necessary in order to provide them with adequate supporting technology.

Scenario development. Application scenarios of context-based community applications based on their needs are the next work step. Some of the so far developed scenarios are presented above. Further work is planned in order to identify aspects that are relevant in various application domains.

Field studies. Field studies are performed to gather first-hand experiences to refine scenarios. This is done through participation of the author in a number of envisioned outdoor community activities.

Analysis of field studies and scenarios. Analysis of field studies and scenarios are done to identify common characteristics as well as differences to "desktop-supported" communities.

Concept development. A concept for the realization of the motivating scenarios of context-based community applications will be developed.

Proof of concept. Proof of concept will be carried out through discussions with community members (users), academics, and industry. Furthermore, prototypical scenario implementations of context-based community applications shall be used to proof the developed concepts and drawn conclusions.

Survey on Related Work. Parallel to all other activities, survey on related work is done through literature studies and participation in scientific conferences and other relevant events.

3 Contributions

The presented work will make the following contributions:

1. Investigation and identification of outdoor communities as potential application fields for future mobile and ubiquitous applications.

2. Concept and framework for supporting context-based (outdoor) communities with MUA, including concepts for:
   (a) Detecting the context (situation) of a mobile person.
   (b) Detecting similar (matching) context of several mobile people as basis for community building.
3. Experiences about the application of the developed concepts to recognize similar contexts of several mobile people as basis for community applications.
4. Experiences about the potential of applying mobile and ubiquitous technologies to support outdoor communities.

4 Organization of Work

In this section we briefly describe the next work steps.

4.1 Context-based Communities

This part builds the basis for the work, as the area of study—context-based communities—is motivated and justified. For this reason, we investigate community support in general and its several forms in particular. Related notions of virtual-, online-, blended-, mobile-, wearable communities, and others, as partially presented in the Section 5, are introduced. Based on the results and predictions in these related areas, future potential of context-based community applications is pointed out.

At the time of writing, we see two groups of people, or even communities, in consideration of supporting their activities through mobile and ubiquitous technologies and applications:

1. Outdoor communities, who are already using communication technology (web portals etc.) to stay in touch with other community members. Up to now, support for staying in touch with others while carrying out community activities is limited.
2. Members of communities, who are using technology or other artifacts to support their community activities (e.g. bike computer, GPS device, avalanche beacon). We see such communities as early adopters of new supporting technology.

4.2 Analysis and Requirements

Within this part of the thesis, we analyze the motivating community scenarios to identify their specialties and common characteristics. Furthermore, requirements for a concept to support community members by the application of mobile and ubiquitous technologies and applications are derived.

4.3 Context Detection and Activity Recognition

Based on identified requirements of mobile community applications and analysis of work already done in related areas, a concept to detect contexts of mobile people will be developed. Among the questions to answer will be the following:

- Which information (environment of the person, activities of the person etc.) is relevant to detect the context/situation of a person?
- Which kinds of context data can be detected by technology?

4.4 Context Matching

Based on the mechanism of detecting the context of a single person, an approach to detect shared contexts of several people (in the wild) shall be developed.

Basis for community applications are shared situational contexts (cf. [13]) of several people. Positional context—similar to [13]—is not of primary importance in our work, as shared situational contexts may occur anywhere.

5 Excerpt of Related Work

In this section, we present an excerpt of related work from an application perspective—support for communities.

Referring to the initial question at the very beginning of this paper, in [14], four potential segments of *mobile community* support from a mobile network operator's perspective are described:

1. Support of mobile community services for free peer-to-peer communities
2. Implementation of own customer-based community platforms and services
3. Support for provider-independent, nationwide community-platforms
4. Support for "mobilizing" existing virtual communities

They conclude that successful mobile community services in segments 1 and 2 might be rare, while segment 3 and 4 might be promising for future community services. With respect to this segmentation, we see our intended community support primarily within segment 1, as we intend to support "communities in the wild". However, a combination of segment 4 and 1 also seems valuable, as there exists a number of (virtual) community services, which could be "mobilized" to support community members when they are "in the wild".

Kortuem introduces the notion of *wearable communities* as a social network that emerges when enough people use their wearable computers to form webs of personal relationships [15]. Similar to *online communities*, they are organized around affinities and shared interests, bringing together people who do not necessarily know each other personally (cf. [9]). In his PhD-work on *wearable computers and contextual awareness* [16], Starner also stresses the use of wearable computers, combining messaging tools, wireless connectivity, and a head-up display to support informal networking among colleagues to form a type of *intellectual collective*. These notions are very similar to our notion of context-based communities, bringing together people who have similar interests or are carrying out similar activities. However, as we intend to support communities in the wild (including e.g. sports activities), we are not sure if the use of "heavy" wearable technologies is suitable in this case.

Following our aim of finding a way to support outdoor communities (communities in the wild) during community activities, closely related work from an application domain point of view is Hocman [17]. It is an application to add value to traffic encounters among motorcyclists. The application is designed for handheld devices equipped with wireless ad-hoc networking interfaces, which are used by the bikers. Every time, two or more bikers come within close proximity of each other, HTML documents are shared among them, documenting who has been met during the trip. In this way, social interaction among bikers is enabled, as personal information is stored in the shared HTML documents.

Another related work is the Hummingbird [18, 19]. The Hummingbird is an inter-personal awareness device (IPAD) that enhances the awareness of presence between people in a group. It is based on a Nintendo GameBoy handheld video game. If two or more Hummingbirds belonging to the same group are within physical proximity (100m or less), it produces a sound and displays the identity of the other Hummingbird users in the vicinity. In the cited articles, the usage of Hummingbird devices by six ski instructors during a five-day trip was studied. The study showed that the Hummingbirds were used primarily as a support for informal social interaction among the ski instructors.

In line with a paper title *Mobile Communities - Extending Online Communities into the Real World* and the concerns presented therein [20], one of our concerns is to make online community services available to community members, when they are carrying out their activities outside (in the wild). Examples for such communities are users of devices supporting outdoor activities (e.g. bicycle computers, GPS devices). Using those devices, users are recording tour descriptions (hiking tours, bike tours etc.), which are later uploaded to a Web portal, from where other users can download the routes to their computers, and further, to their mobile devices. One of our aims is to support sharing of such artifacts among community members not only when sitting in front of desktop PC, but rather during ad-hoc encounters in the wild (*ad-hoc communities*).

6 Conclusion

In this paper, support for outdoor communities (communities in the wild) as an answer to the question about how to provide added-value to mobile people through mobile applications has been presented. Along with this suggestion comes the question about how members of outdoor communities can be provided with community-specific services while carrying out community activities. This question represents the research problem of an ongoing doctoral thesis. Background and motivation, the research problem, derived subproblems, applied methodology, and the expected contributions along with organization of work have been presented.

References

1. MB-net: Network of excellence in mobile business applications and services. http://www.mobiforum.org/mbnet/ (2002) Last visited: May 2005.
2. MobiForum: The m-commerce portal. http://www.mobiforum.org/ (2000) Last visited: May 2005.
3. UMTS: UMTS Forum. http://www.umts-forum.org/ (1996) Last visited: May 2005.
4. WWRF: Wireless World Research Forum. The book of visions 2001 — visions of the wireless world (2001)
5. WWI: Wireless World Iinitiative. http://www.wireless-world-initiative.org/ (2004) Last visited: May 2005.
6. Pichler, M.: Identifying mobile people's information needs. SCCH-TR-0511 (2005)
7. Pichler, M.: Enabling contextual services in wireless context area networks. In Ferscha, A., Hörtner, H., Kotsis, G., eds.: Advances in Pervasive Computing. Band 176, Austrian Computer Society (OCG) (2004) 111–117
8. Pichler, M.: Enabling communities in physical and logical context areas as added value of mobile and ubiquitous applications. In Crestani, F., Dunlop, M.D., Mizzaro, S., eds.: Proc. Mobile HCI 2003 Workshop on Mobile and Ubiquitous Information Access. Volume 2954 of LNCS., Udine, Italy, Springer (2004)
9. Pichler, M., Hofer, T., Leonhartsberger, G.: Bringing together people: Social benefit of mobile ad hoc applications. In: Proc. CHI 2002 Workshop on Mobile Ad Hoc Collaboration, Minneapolis, Minnesota, USA (2002)
10. Johansen, R.: Groupware: Computer Support for Business Teams. Free Press, New York (1988)
11. Cortese, G., M.Lunghi, Davide, F.: Context-awareness for physical service environments. In Riva, G., Vatalaro, F., Davide, F., Alcañiz, M., eds.: Ambient Intelligence, http://www.ambientintelligence.org, IOS Press (2005)
12. AN: Ambient networks project. http://www.ambient-networks.org/ (2004) Last visited: April 2005.
13. Gellersen, H., Schmidt, A., Beigl, M.: Multi-sensor context-awareness in mobile devices and smart artefacts. Mobile Networks and Applications **7** (2002) 341–351
14. Fremuth, N., Tasch, A., Fränkle, M.: Mobile communitiesnew business opportunities for mobile network operators? In: Proc. of the 2nd Interdisciplinary World Congress on Mass Customization and Personalization (MCPC), Munich, Germany (2003)
15. Kortuem, G.: A Methodology and Software Platform for Building Wearable Communities. PhD thesis, Department of Computer and Information Science, University of Oregon (2002)
16. Starner, T.E.: Wearable Computing and Context-Awareness. PhD thesis, MIT Media Laboratory, Cambridge, MA (1999)
17. Esbjörnsson, M., Juhlin, O., Östergren, M.: The hocman prototype — fast motor bikers and ad hoc networking. In: Proc. of MUM 2002, Oulu, Finland (2002)
18. Weilenmann, A., Holmquist, L.E.: Hummingbirds go skiing: Using wearable computers to support social interaction. In: ISWC '99: Proceedings of the 3rd IEEE International Symposium on Wearable Computers, Washington, DC, USA, IEEE Computer Society (1999) 191–192
19. Weilenmann, A.: Negotiating use: Making sense of mobile technology. Personal and Ubiquitous Computing **5** (2001) 108–116
20. Hillebrand, C., Groh, G., Koch, M.: Mobile communities - extending online communities into the real world. In: Mobile and Collaborative Business. (2002) 7–17

Electronic Operational Documentation Use in Civil Aviation

Jean-Philippe Ramu

European Institute of Cognitive Sciences and Engineering (EURISCO International)
4, Avenue Edouard Belin
31400 Toulouse, France
+33.5.62.17.38.38
jean-philippe.ramu@onecert.fr

Abstract This paper discusses the development of a new generation of electronic operational documentation in the aviation domain. A matrix of flight crew activities describing the use of operational documentation is introduced in order to take into account both, flight crew encountered situations and operational documentation use. I discuss the use of "context" as a mean to develop an electronic operational documentation tool. The objective of a contextual operational documentation tool is to develop operational documentation so that its use is no longer an interruption of the operational work, but is a naturally integrated part of flight crew's work management.

Introduction

In the aeronautical domain, operational manuals are the means to support the storage and use of useful information, as well as technical facts describing or limiting operations. The developments of operational manuals are a continuous process integrating lessons learned from on line operations or training practices and is the technical reference for system behavior descriptions. Lessons learned are related to the iteration of best practices and manufacturers and airline training organizations widely communicate them.

Despite significant advances in aircraft design, the present operational manuals format and medium have remained basically unchanged. These manuals remain "classic" paper documents, and are organized accordingly. However, progress has been made in systems to store documentation electronically. Electronic documentation introduces the concept of structured modular documentation. The emerging computer standards such as the eXtensible Mark-up Language (XML) will enable industrial application of object-oriented databases. In the near future, operational manuals will migrate from static paper output to dynamic electronic format.

Operational manuals have always been used as a reference publication and this will always be maintained. New information technologies will enable the continuous consultation of operational manuals enhancing its impact on the user learning process. The availability of multiple devices and networking capabilities will enable large amounts of information to be taken into account, as well as the integration within the

working environment of electronic operational manuals as a performance support tool [1].
The migration of operational manuals from the status of reference to the status of an integrated performance support tool will not be trivial. The operational environment in aviation, especially in the aircraft cockpit, is already sustained by information. The role of a new performance support tool will have to be identified with respect to the existing environment. Its conception will go beyond the content of the documentation; properties for the right exploitation of the content will have to be specified. The changes will impact both end-user's daily operations and training, as well as organization operational manuals production and culture.
In this paper, first I describe the transition of operational documentation into electronic format and discuss its integration into today's aviation environment. Then I describe the use of operational documentation by flight crews in the training and operational environment. Next, I introduce a matrix of flight crew activities using operational documentation information. Finally, I introduce the notion of context as a framework for the management of this matrix during flight crew's work practices. Electronic

Operational Documentation in Aviation

The International Civil Aeronautic Organization (ICAO) states that the definition of an operating manual for flight crews is [2]: "Manual where are indicated all procedures, instructions and indications for the operating personnel in the execution of their tasks." Hereafter, I will define operational documentation as the overall content of these manuals.
Multiple organizations produce operational documentation. The manufacturer provides technical documentation in order to deal with the aircraft functioning. National authorities provide information with respect to airports and airspace infrastructures. Specialized organizations provide meteorological bulletins. Airline assembles existing information and customizes the content with respect to their particular practices.
The transition from paper to electronic format will enable the integration of documentation into one single database, eliminating today's documentation segmentation into manuals. It will be up to the documentation interface, the so-called Electronic Flight Bags (EFB) [12], to filter the relevant content with respect to its particular use.
The aviation domain has already experienced a transition of its information sources into the electronic world. In the past, classic cockpit generation instruments were made of dials and needles. Each of the information pointed by cockpit needles was relevant for the flight management, and was accurate to the actual state of the aircraft. At this time, paper operational documentation was the sum of all other available information enabling the flight crews to foresee the flight navigation and understand the systems functioning conditions and respective possible consequences.
Today, the transition of the instrumentation information into electronic format produced the so-called "glass cockpit" generation, which is the norm for most of

commercial aircraft. Information that was in the past presented separately with needles are now merged and integrated in electronic displays. This integration was not limited to the existing classic cockpit instrumentation, but the space and arrangement possibilities of the displays enabled to assemble correlated information, and even explicitly add new information. Most added information was information previously only reachable through an operational documentation tool. Examples are the Airbus' Electronic Centralized Aircraft Monitoring Systems (ECAM) and its system displays, the Navigation Display (ND) and the chart representation with waypoint and routing presentation [8].

With the transition of operational documentation into electronic format (Electronic Operational Documentation (EOD)), all information will have the possibility to be merged and filtered. Operational documentation might be connected to specific sensors just like instrumentation is. In the general aviation, the development of e-books coupled with Ground Positioning System (GPS) will replace static paper charts(usually part of operational documentation [10]). Its use becomes similar as a ND in commercial aviation. The more EFB will be sophisticated, the less the physical distinction between instrumentation and operational documentation will be obvious.

While operational documentation was only available in paper (static) format, the user himself always triggered its use. Now that electronic (dynamic) operational documentation become available, the possibility exist to develop an EFB that provide the user with the right amount of information, at the right time, for its use in training or in operations. In the next section I will describe the management of operational documentation in civil aviation, with the aim of developing an EFB that is a naturally integrated part of flight crew's work management.

Operational Documentation Management in Aviation

It is recognized that an error or a lack of information may have a performance impact. As a consequence, organizations produce and dispatch operational documentation and associated revisions and bulletins as necessary. Weather information is a typical example, but the evolutionary property is applicable to all operational documentation sources. Flight crews may also forget things. As a consequence, in order to ensure an adequate level of qualification, flight crews need to acquire and refresh their individual knowledge and competences recurrently.

In the aviation domain, several processes exist for the flight crew to acquire and maintain sufficient and up-to date competences. Flight Training Organizations (FTO) dispatch ab-initio training, special devices such as Computer Based Training (CBT) enable individual learning. Present technologies enable to simulate operations, and give the opportunity to improve and test flight crew's competences. Full Flight Simulators (FFS) are able to train the flight crew in a realistic manner under critical and improbable circumstances. Training environments are also a possibility to provide information feedback. All those training and simulation environments, together with operational documentation use, aim at giving the flight crew means to obtain and maintain its qualifications.

Finally, the role of operational documentation is to help the flight crew to access the right information at the right time, in order to enable efficient operations. It is a mean to support situational awareness, which is defined as "...the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." [6]. During operations, learning on the job gives the flight crew the possibility to maintain its basic competences. As for training environments, operational environments are a possibility to provide information feedback.
The management of an integrated EFB will have to take into account all the above processes. It will have to be a mean for the flight crew to autonomously optimize his individual qualification; it will have to assist him in simulation and operational environment; and also be a mean for organizational feedback. Each of these activities will need the manipulation of operational documentation. Hereafter, I call "documentation activities" the activities involving electronic operational documentation, and I further categorize the properties that guide the flight crew documentation activities.

Categories of Electronic Operational Documentation Activities

I will categorize the activities involving flight crew information resources such as operational documentation around two axes, the "time to task" axe and the "documentation activity workload" axe. The "time to task" axe refers to the dynamical property of the domain, and takes into account the elapsed time between the documentation activity done to perceive information and the task accomplishment for which this information is foreseen. I will divide it into three categories:

- "Documentation activity at hand" is the time to task for which the documentation activity matches the actual situation need. For instance, the take-off phase in the aviation domain needs to monitor carefully the speed in order to accomplish the take-off tasks. This category is heavily constraint by the dynamical property of aviation and needs a constant iteration of the perception-action loop matching the situation evolution. The flight crew competences will influence the perception action iteration efficiency to reach the task goals (see also [9]);
- "Anticipation documentation activities" is the time to task for which the documentation activity matches an expected modeled situation need. Monitoring meteorological forecast is a typical anticipation documentation activity. This activity is less constraint by the dynamical property of aviation domain as an expected situation refers to an assumed situation to be encountered in the "near" future. The available elapsed time between the anticipation documentation activity and the expected situation will influence the strategy of documentation activity used. The flight crew competences will influence the perception anticipation iteration efficiency to choose expected tasks matching the expected modeled situation;
- "Training documentation activity" is the time to task for which the documentation activity matches a chosen modeled situation need. A fully chosen modeled situation requires to disconnect from real operations and to enter a simulated

environment. The dynamical property of aviation is not anymore a constraint, but a model that may be frozen. Documentation activities such as demanding trainer advices, use of other information sources, may be done without restrictions.

The “documentation activity workload” axe represent the flight crew effort needed to acquire information contained within the operational documentation through the EFB interface. In order to categorize this axe I will study four possibilities. Table 1 summarizes the possible cases. I will consider that a group of information that I call g(i) might be either: (1) the flight crew knows g(i) exist and knows how to interpret g(i); (2) the flight crew knows how to get g(i); (3) g(i) is available in the EFB database; and (4) g(i) is displayed in the flight crew’s working environment. Each of those functions (exist g(i), get g(i), available g(i) and displayed g(i)) can take two values: “YES” or “NO”. I will hereafter discuss each case, using the example of an airspeed indicator in order to describe the situation (see Fig. 1):

Fig. 1. Typical general aviation airspeed indicator

Table 1. Operational documentation group of information g(i) possible states

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)	(P)
EXIST g(i) ?	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES	NO	NO
GET g(i) ?	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
AVAILABLE g(i) ?	YES	YES	YES	YES	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO
DISPLAYED g(i) ?	YES	YES	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES
	NEAR SUPPLY	REFRESHER ASSISTANCE	EXPLANATION ASSISTANCE	ATTENTION ASSISTANCE	FAR SUPPLY	TARGET HELP	NOT CONSIDERED	DISCOVERING HELP	FEEDBACK	FEEDBACK	NOT CONSIDERED	USELESS HELP	NOT CONSIDERED	NOT CONSIDERED	NOT CONSIDERED	NOT CONSIDERED

(A) The flight crew knows the group of information exist and knows how to get it. The group of information is available and displayed: The flight crew knows that the airspeed is indicated through the airspeed indicator, he knows where the airspeed indicator is, and knows that the needle present an actual aircraft speed. I will call this category of activity the “near supply” category;

(B) The flight crew knows the group of information exist but does not know how to get it. The group of information is available and displayed: The flight crew does not remember where he can read airspeed information. He will need to scan his environment with the hope of re-discovering the information, or use a dedicated database output interface in order to refresh his memory. I will call this category the "refresher assistance" category;
(C) The flight crew does not know the group of information exist but knows how to get it. The group of information is available and displayed: This time the flight crew sees the airspeed indicator in his working environment but is not able to interpret it. He will need an assistant in order to understand the airspeed indicator. Answering the questions: what is it? what does it serve? are typically documentation activities of that category. For instance, the "AIRSPEED" notice in the centre of the indicator is helpful for this category. I call this category the "explanation assistance" category;
(D) The flight crew does not know the group of information exist and does not know how to get it. The group of information is available and displayed: I have to be a little bit imaginative to encounter this situation with our classic airspeed indicator. This category is interesting if we imagine that the speed increases, and goes above 140 knots. If this situation is submitted to restrictions, we could imagine that the indicator displays a message relating to that restriction. In that case, it could happen that the flight crew doesn't see the information as it just dynamically appeared, but the information is available and displayed. If this information is judged critical, we could imagine an attention getter in order to ask the flight crew to notice the information. I call this category the "attention assistance" category. Once he identifies the restriction (and not the attention getter), he enters again the previous category description;
(E) The flight crew knows the group of information exist and knows how to get it. The group of information is available but not displayed: Due to the anemometer conception, the airspeed tool measures a dynamical pressure. This measure is proportional to the aircraft speed, and is a convenient indication in order to translate aircraft aero-dynamical performances, but does not indicate the True Air Speed (TAS). The TAS is the real aircraft speed in the mass of air and is a speed used for navigation calculations. In order to read the TAS out of the airspeed indicator, it is necessary to use the rotating knob at the bottom left of the instrument, and set a value on the upper window function of the outside pressure and temperature. Only then, the needle will show on the outer scale the TAS. I will call this category of documentation workload the "far supply" category. Once the documentation "far supply" activity is completed, we fall into the "near supply" category;
(F) The flight crew knows the group of information exist but does not know how to get it. The group of information is available but not displayed: Following the description of (E), it may be the case that the flight crew does not know the correct pressure and temperature at that time. In that case, the flight crew will have to ask for this information, or use a dedicated database output interface. The flight crew may know that the meteorological chart provides this information, but does not find it directly. I will call this category the "target help" category;
(H) The flight crew does not know the group of information exist and does not know how to get it. The group of information is available but not displayed: This could be the case during a training session, if the trainer asks the flight crew to explain the significations of the colored banner of the airspeed scale. The initiation of the

documentation activity does not present directly (i),but refers to (i). I will call this documentation activity the “discovering help” activity;
(I) The flight crew knows the group of information exist and knows how to get it. The group of information is not available and not displayed: This case does not refer to a documentation activity using operational documentation directly. In this case it is explicit that the flight crew does not get (i) in the database as (i) is not available. However, it may be the case that the flight crew has annotated his documentation. It could also emphasize a lack of information provided by the database repository. This case might need a “feedback” documentation activity;
(J) The flight crew knows the group of information exist but does not know how to get it. The group of information is not available and not displayed: This case is similar as (I), but the flight crew discovers (i) without prior knowledge of (i). This case might as well need a “feedback” documentation activity;
(L) The flight crew does not know the group of information exist and does not know how to get it. The group of information is not available and not displayed: Following the case (H), it could be that the database repository is not able to provide helpful information to the flight crew, the flight crew not knowing that a “discovering help” will be useless. In this case, the flight crew might loose confidence if he does not find helpful information. I call this documentation activity the “useless help” activity;
(G) (K) (M) (N) (O) (P) are cases not considered, as they can logically be rejected or be similar to another case already described.
In order to have a better overview of the documentation activity workload categories, Fig. 2 illustrates the paths through the categories if step-by-step one state at a time would change:

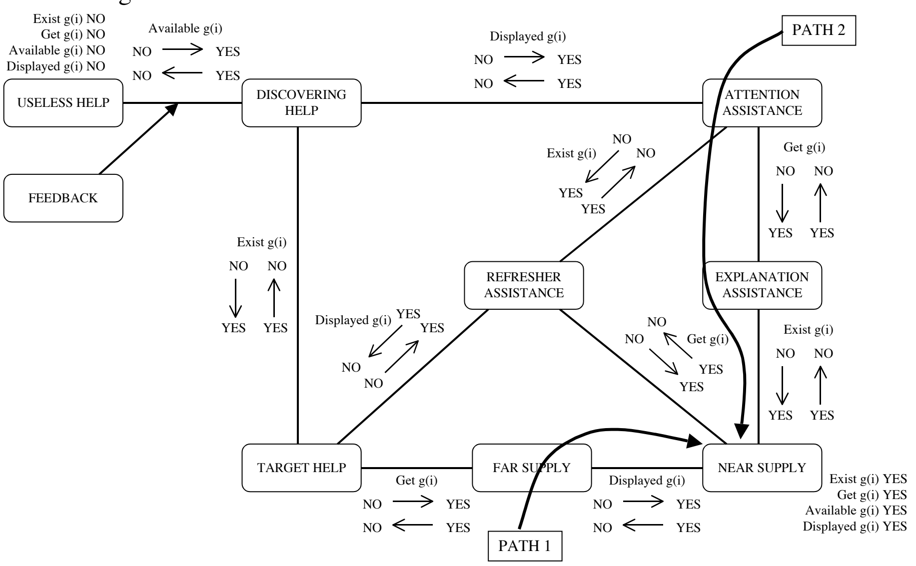

Fig. 2. Documentation activities workload categories paths

It illustrates those paths going from a “useless help” category (flight crew’s worst case) to the “near supply” category (flight crew’s best case). Those paths represent

possible uses of operational documentation. However, the flight crew might enter those paths through each of the separate categories, and leave it whenever he decides it. For instance, the flight crew can enter the path in the “far supply” category, than reach a group of information thus entering the “near supply” category (Fig 2, path 1). He might also enter the path in the “attention assistance” category, with an alarm noise, than need “explanation assistance” in order to interpret the “near supply” failure description (Fig 2, path 2).
In order to correlate the documentation activity workload axe categories with the time to task axe categories, and to maintain at a suitable level of detail, we will group the documentation activities workload categories into three categories of higher level. I will call (see also [3]):

- “Supply documentation activities” category if {available g(i) YES and exist g(i) YES and get g(i) YES }. This category correspond to any documentation activity when the flight crew already knows what group of information he needs and knows where to acquire it;
- “Assistance documentation activities” category if {available g(i) YES and displayed g(i) YES}. This category relates to activity involving information that the system displays, and remains under this category until the flight crew know how to interpret this information and/or knows how to get this information. It is either a lack of knowledge of the flight crew that must be recovered, or an automation of the information system that provides relevant information assuming it may assist the user;
- “Help documentation activities” category if {get g(i) NO and displayed g(i) NO}. This category relates to documentation activities when there is a communication interaction between the flight crew and the information system, as the flight crew does not know how to gather the group of information and it is not yet displayed.

Fig. 3 summarizes the two axes categories definitions and describes a documentation activity matrix.

Discussion

Although I have merged the concepts of instrumentation and operational documentation information, the scope of the study is not to take into account all information produced under those two products. Our goal is to describe a new approach in the use of EOD in the form of an intelligent assistance system [4], for which the production and use is under transition.
In order to characterize our EFB interaction, I need a concept that enables to model (1) the situation in which the user is evolving (time to task axe), (2) the documentation tool (documentation activity workload axe), and (3) the user itself which will evolve in the matrix space in respect to his specific knowledge. The notion of “context” has this property. It can be defined as [5]: “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves.” In order to formalize the context, I intend to use the ontology technology [7]. I call contextualized

operational documentation the correlation of a domain specific ontology modeling the context with the electronic operational documentation [11].

The research question of our study is: How it is possible to develop electronic operational documentation that supports the operational activities as an integrated information system in the flight crew's work management. The study will focus on two phases. Phase 1 considers the EFB as a stand-alone tool and will demonstrate the interaction in the "training documentation activity" and "anticipation documentation activity" part of Fig. 3. In this phase, the "anticipation documentation activity" can be seen as a flight crew preparation duty. Phase 2 will integrate the "documentation activity at hand" part and will demonstrate the interaction of the EFB as an integrated tool within the flight crew's operational environment, with the possibility to be correlated with instrumentation.

A further investigation of the study is to describe strategies of use of an integrated EFB in order to: (1) reduce the need of EFB use as help device in a task at hand use; and (2) maximize the learning effect of an EFB use as training supply. The first of which may improve efficiency of flight operations, and the second reduce the use of expensive full flight simulator devices.

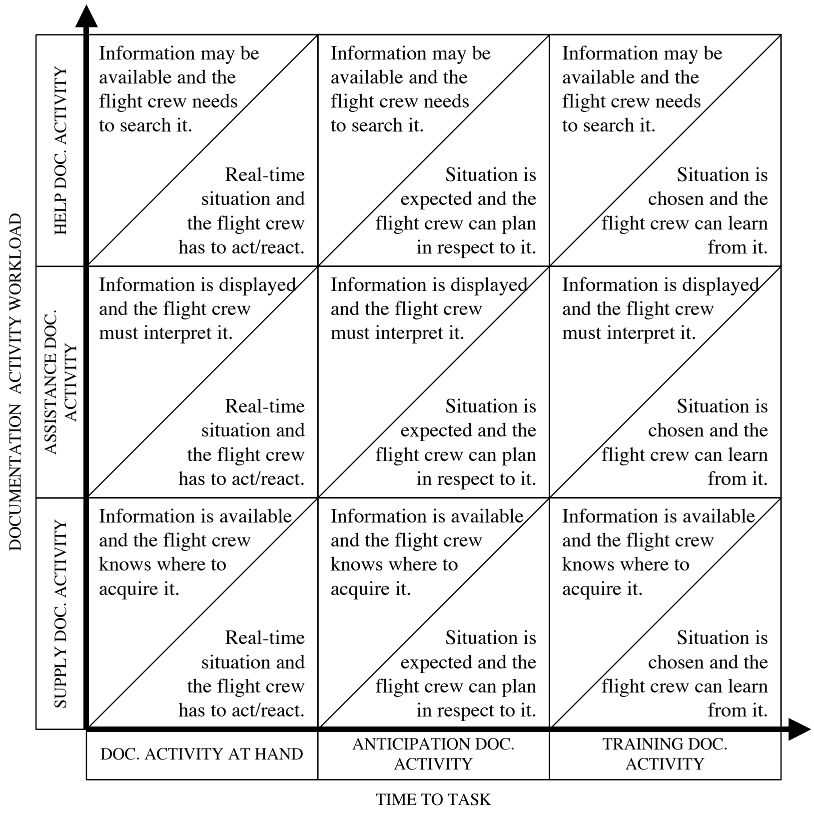

Fig. 3. Documentation activities workload and time to task documentation activities matrix

Acknowledgment

Thank you to Yvonne Barnard, my project manager; and Guy Boy & Guy Gouardères, my thesis directors.

References

1. Barrera Esquinas, J.; Durstewitz, M. (2002). Evolution of Human-Device in the Field of Technical Documentation. In S. Chatty, J. Hansman, & G. Boy, Proceedings of the International Conference on Human-Computer Interaction in Aeronautics, p.80-86. Menlo Park, California: AAAI Press
2. Beillard, A. (2002). Procédures opérationnelles – Généralités. Institut Aéronautique Jean Mermoz.
3. Boulier, D. (2005). Aide-toi, l'aide t'aidera. Prise et emprise dans les aides home-machine. A paraître dans "systèmes d'aides" Gapenne et Boullier, eds, 2005.
4. Boy, G. (1991). Intelligent Assistant System. Academic Press Inc., San Diego.
5. Dey, A. (2001). Understanding and Using Context. Personal and Ubiquitous Computing (pp. 4-7).
6. Endsley, M.R. (1988). Design and Evaluation for Situation Awareness Enhancement: Proceedings of the Human Factors Society 32nd Annual Meeting. Santa Monica, CA: Human Factors Society. 97-101.
7. Gruber, T.R. (1993) A translation approach to portable ontology specification, Knowledge Acquisition 5 (1993) 199-220.
8. Harris, D. (2004). Head-down Flight deck Display Design. In Human Factors for Civil Flight Deck Design. Ashgate publishing Limited. England.
9. Hollnagel, E. (2004). Time to think and time to do? I can fail, and so can you! Human Factors and Aerospace Safety 4(3) 169-179, Ashgate Publishing.
10. JEPPESEN (2004). Electronic Flight Bag. http://www.jeppesen.com
11. Ramu J.-P. (2004). Contextualized Operational Documentation: an Application Study in Aviation. In S. Schultz & T. Roth-Berghofer, Proceedings of the Workshop on Modeling and Retrieval of Context, p. 109-120, 27th German Conference on Artificial Intelligence, Ulm, Germany
12. Shamo, M. (2000). What is an Electronic Flight Bag, and what is it doing in my cockpit? Proceedings of the International Conference on Human-Computer Interaction in Aeronautics,p.60-70. 27-29 September, Toulouse, France.

An Approach for Context-based Schema Integration in Virtual Information Environments*

Heiko Stoermer[1] and Rodolfo Stecher[2]

[1] University of Trento,
Dept. of Information and Communication Tech.,
Via Sommarive 10, 38050 Trento, Italy
stoermer@dit.unitn.it
[2] Fraunhofer IPSI
Integrated Publication and Information Systems Institute
Dolivostrasse 15, 64293 Darmstadt, Germany
stecher@ipsi.fhg.de

Abstract. In so-called Virtual Information Environments (VIEs), which consist of meta-information about vast document collections, the exchange and integration of hierarchical classification schemas are important issues in terms of semantic interoperability. We see schema exchange between agents and the resulting integration process as a context switch. Thus we present an approach for semantic schema integration based on formal logics of context, by exploiting the contextualization features of an existing schema matching algorithm and storing the elicited semantics of the schema in a VIE for later use by semantic-based community processes.

1 Introduction

With the advent of the Semantic Web [4], Knowledge Management faces new challenges in the form of massively distributed information systems. There exists an large amount of documents in the WWW, containing a host of information which is currently not managed with respect to semantics but only existing as raw data. Recent research projects such as the Vike Framework (see www.vikef.net) try to attack the challenge of managing the contained information by building virtual information environments (VIEs) which extract information from distributed document repositories into a local information system to provide semantic-enabled services to a community of users. This information does not only consist of the documents themselves, but also of meta-information and structures.

* The work described in this paper has been partly funded by the European Commission through grant to the project Vikef under the number IST-507173. We would like to thank the creators of the CtxMatch algorithm as well as the members of the Vikef project team for the numerous discussions and contributions to the idea of the Core Vikef Infrastructure and semantic-enabled service development.

One of the crucial tasks in this respect is the management of hierarchical schemas. It is a well know fact that schemas are widely used in information systems for organizing large numbers of documents [12]; examples include web directories, computer file systems or product catalogues. In a VIE, we envision at least the following use cases involving schemas:

- During *creation* of the VIE, the knowledge engineer wants to provide the user with one or more default taxonomies for hierarchical browsing of documents.
- There will be *internal* uses of schemas, e.g. for normalizing metadata with schemas such as the Dublin Core (see http://dublincore.org/).
- During *evolution* of the VIE, schemas from other VIEs have to be integrated for compatibility reasons.
- VIEs have to be *merged*, or parts of one VIE has to be *imported* into another, so schema integration or schema matching will take place.

We believe that in massively distributed information systems an a-priori agreement on knowledge, knowledge exchange and centralization is very hard to achieve [9, 5], and is also inconsistent with numerous managerial aspects of Knowledge Management [6]. For this reason the common paradigm of shared knowledge in combination with local schemas does not seem applicable, but rather converts into a problem of local knowledge and shared schemas, which we see as a matter of context.

2 Towards Schemas in Context

When speaking about semantic information environments, we think of a large number of agents in the form of *semantic peers*[3], which are autonomously working systems operating on their own local knowledge bases. The reason for this view is that in our opinion operation on shared knowledge is very hard to achieve and is only applicable in small and controlled environments, such as company intranets or intra-community information systems. But if we try to achieve integration or interoperation of information systems, the current standard approach of creating large-scale shared knowledge will hardly scale up to something the size of the (Semantic) Web, and is also conceptually problematic because in our opinion knowledge is never context-free [26, 28] and can thus never be perfectly shared. It is much more likely that we will face more complex scenarios which basically can be described as interoperation between semantic peers, achieved rather through shared schemas than through shared knowledge.

Taking into account the formal theory of context in [2], for our setting we believe it to be sensible to define that a semantic peer actually represents a context, because it can operate in isolation and contains a local knowledge base and its own logics and reasoning. As a logical consequence, exchanging a schema between semantic peers can be defined as a context switch; from an interpretation

[3] For details on this notion refer to [11].

point of view, the schema will be detached from its original context and the new interpretation will be based on the knowledge base of the new context.

In Fig. 1 we contrast the notions of knowledge exchange versus schema exchange: on the left-hand side of the figure, we illustrate the aforementioned scenario of shared knowledge, which will lead to *identical interpretations* of any schema between agents. The right part highlights the more likely scenario of a schema switching context, which leads to *different interpretations* of the same schema between agents.

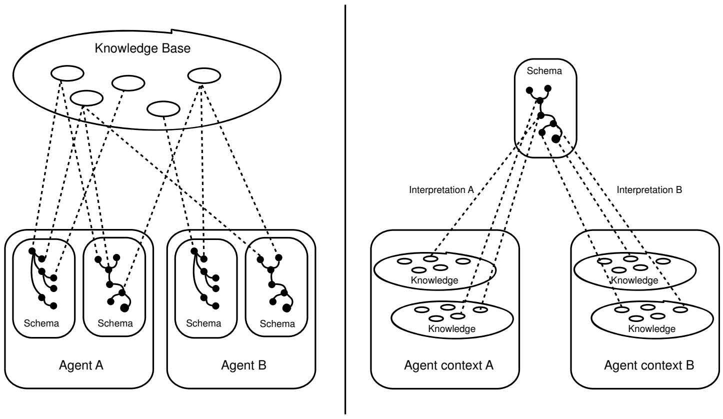

Fig. 1. Shared knowledge and local schemas vs. local knowledge and shared schemas.

The resulting interpretations will differ between the agents, which makes it legitimate to question the usefulness of sharing a schema for interoperability if the interpretations of the schema are not identical. But in fact, even communication between human agents is submitted to this problem[4] and a rational analysis of *realistic* scenarios can lead to the conclusion that this problem can hardly be eliminated. Yet it makes semantic interoperability harder – but not impossible – to achieve: in our case the basic challenge is a) to provide intelligent analysis of the knowledge implicit in schemas and b) manage a persistent and machine-processable representation of this knowledge.

3 Challenges of Schema Interpretation

Given the fact that when exchanging schemas we can not rely on shared knowledge and thus have to re-interpret the exchanged schema based on local knowledge, let us illustrate the challenges that arise from schema interpretation.

[4] What one human says and what another human understands can differ considerably. For further reading on this topic, refer to various works from the area of Philosophy, such as [13, 27, 22]

Classification schemas usually follow the pattern of a concept hierarchy where nodes have labels that are meaningful for humans (see Fig. 2).

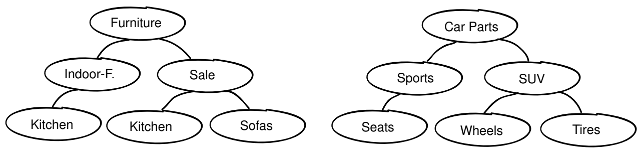

Fig. 2. Structurally identical but semantically independent schemas, with free-form labels and inhomogeneous relation types.

While the human agent is intuitively able to deduce the meaning (semantics) described by a schema, this task is a complex problem of Artificial Intelligence that has been approached in a number of ways [25, 24, 23, 3].

In our opinion the reason for this amount of different approaches are the problems arising out of semantic heterogeneity. For an illustration of these problems we adapt the classification of semantic heterogeneity in ontologies, suggested in [8]:

The syntactic and capability level. On this level, problems relate to the choice of a schema representation language and its technological capabilities.

- The schema to be integrated is technically not composed in a language that allows for usage with knowledge in an ontology. (I.e. the schema might be serialized in plain XML as a bare conversion of a file system hierarchy structure.)
- The schema to be integrated is technically composed in a language that does provide for semantic facilities, but they have not been exploited. The semantics of the schema do only exist in the creator's mind.
- The schema to be integrated is technically composed in a language that does provide for semantic facilities, and they have been made use of, but the referred knowledge base has not been provided or is not accessible.

The terminological level. Node labels are usually formulated in free text [3], and have to be analyzed with methods of Natural Language Processing (NLP). Due to this free-form notation, manifold label representations can be thought of, which lead to hard parsing problems such as the string *Indoor-F.* in Fig. 2, or more complex labels that formulate whole phrases but violate language syntax by leaving out spaces, e.g. *ImagesOfTrentino*. Another issue is general ambiguity present in natural languages, e.g. in the form that there might exist a number of different words for the same object (synonymy), or that one word can refer to a number of different objects (polysemy).

The semiotic/pragmatic level. This level refers to heterogeneity issues arising from the fact that there may exist circumstances leading to different interpretation of schemas in different contexts.

– Simple graph analysis of the hierarchical structure is obviously prone to error, as illustrated in Fig. 2, which shows two structurally identical schemas with completely different semantics.
– The kind of relation[5] between hierarchically connected nodes is not explicit and can also vary between parent/child pairs of the same schema. In Fig. 2 we see that *Furniture → Indoor Furniture → Kitchen* could be intuitively analyzed by a human as a specialization, while in *Furniture → Sale → Kitchen* this formalization is not applicable.

4 Solution Approach: Semantic Elicitation

To overcome some of the various challenges in schema analysis and management presented to far, we present the idea of embedding the process of *semantic elicitation* described in [10] into VIEs, to provide a persistent representation of a contextualized schema as a base for intelligent services and reasoning.[6]

The algorithm as a whole, called CtxMatch [10], aims at *semantic coordination*[7] between hierarchical classifications. For the reason that schema coordination might only be one of the possible application areas, we will only exploit the first step of the algorithm which will produce a contextualized representation of a schema based on several types knowledge. This representation, in Description Logics (DL), will then be stored within the VIE to provide higher levels of the VIE architecture with information about the schema that can be processed by automated reasoners.

CtxMatch uses three types of knowledge to extract information from a schema: lexical knowledge, by means of NLP and a knowledge base to identify and disambiguate node labels, domain knowledge in the form of a domain ontology, and structural knowledge derived from the arrangement within the schema itself. Based on these knowledge sources, the approximated meaning of every node in a concept hierarchy is made explicit in a DL formula. In its original form, CtxMatch will then use this internal DL representation of *two* contextualized schemas to calculate semantic mappings between nodes.

In the first phase of the algorithm, node labels are being parsed and mapped to all possible senses in the lexical knowledge base (WordNet [14]). The result is a list of so-called *atomic concepts*. As a second, more complex step, the implicit structural knowledge of the schema is analyzed, irrelevant atomic concepts are eliminated and the remaining variations (called *complex concepts*) of the complete paths in the hierarchy (e.g. *Furniture → Indoor Furniture → Kitchen* in Fig. 2) are encoded in logical formulae, which can then be stored in the VIE for further processing.

[5] For details on relation types in conceptual modeling, such as composition, specialization, instantiation, etc. see for example [20].
[6] The well-researched, formal theory of context is described in several papers [2, 18, 17, 16] and seems applicable to this kind of problem. This theory serves as a foundation for our solution approach to the problem of intelligent schema integration in VIEs.
[7] For details on *semantic coordination* please refer to the introduction of [7].

With this approach, we hope to attack the following challenges illustrated in Sect. 3:

Syntactic level. We do not require the exchanged schemas to have any reference to a knowledge base. The node labels and the implicit structural information suffice for analysis and integration of the schema.

Terminological level. The CtxMatch algorithm implements several heuristics and methods of NLP for label parsing, even in more complex forms.

Semiotic/pragmatic level. We do not rely only on mere graph analysis, but also take into account the semantics of the labels by locating their respective concepts in a local knowledge base, as well as structural aspects of the hierarchical classification by trying to guess relations between the discovered concepts.

There are three major benefits of this approach: for one, we perform *semantic analysis* of the schemas based on a combination of three knowledge sources *in context*, as opposed to statistical or mere structural analysis which lack reference to the semantic level. Secondly, we provide for *machine processability* of the implicit knowledge of a schema, in that every node of the schema is represented by a complete DL description of its (approximated) meaning. Finally, the *abstraction level* of the representation is believed to be high enough to allow for various uses effected by reasoning processes, such as schema coordination, mapping of documents to schemas or mapping of schema nodes to ontology concepts.

5 The Vikef project

Vikef (Virtual Information and Knowledge Environment Framework) is an application-oriented Integrated Project within the IST Sixth Framework Programme of the European Community, which started in April 2004. Its main aim is to bridge the gap between the (partly) *implicit* knowledge and information conveyed in scientific and business information, content and knowledge (ICK) resources and the *explicit* representation of knowledge required for a targeted and effective access, dissemination, sharing, and use of such ICK resources by scientific and business communities. Vikef will provide a framework for building in a parameterizable and re-usable way virtual information and knowledge environments, based on a community specific combination of semantic-enabled services tailored to the requirements of the specific domain, for different communities.

The semantic-enabled services use semantic representations of the domain knowledge, called *Semantic Resource Networks* (SRNs). SRNs consist of Resource Description Framework (RDF) [21] statements about relevant instances in the target domain, based on underlying domain ontologies [19] represented in Ontology Web Language (OWL) [1]. The semantic-enabled services use selected parts of such SRNs in combination with domain-specific inferencing rules to filter out relevant knowledge and to infer new explicit knowledge relevant for a specific community task (e.g. find trends in a research area for a conference organizer or visitor). We will consider now the SRN building and consuming process, where integration of data corresponding to different schemata, semi-automatic discovery and definition of inter-schema mappings and of similar elements in metadata schemata and ontologies play a fundamental role.

In the VIKEF project, metadata for building SRNs can be extracted from underlying document collections through analysis, or imported from already existing metadata sources. In both cases, the schemata from the extraction process or from the already existing metadata sources have to be transformed into an *Annotation Template* (AT). An AT is a general XML schema that will be mapped to the domain ontology of interest. ATs can be regarded as a *middle* representation that is used for *normalizing* the, possibly many, data sources into a common representation, containing the elements that are of interest for the target application (domain). Since there could be several different ontologies (e.g. for slightly different applications in one domain or in overlapping domains), it could be necessary to map an AT to different ontologies. The opposite scenario will also be considered, where there are different ATs that need to be mapped to one ontology (integration of several different metadata extraction processes and reuse of results of past extraction processes).

The different schema levels that can be distinguished in VIKEF for the representation of semantic information are (see also figure 3):

Instance Level: The instance level contains the acquired and harmonized content objects;

Metadata Level: The metadata level contains a) Annotated documents that contain meta information as markup within the objects; b) Separate metadata;

Pragmatic or Schema Level: The schema level may contain schemata for data structuring, like for example DBLP Schema or Dublin Core, as well as classification schemata; the ATs are also at this level.

Semantic Level: The semantic level contains the domain ontology or ontologies;

The main distinction between the Schema level and the Semantic level is not on the type of information structures located on each of these levels. Indeed both of the levels may contain e.g. RDF Schemata. The difference is on the locality of the information: The pragmatic level contains local schemata as they are used for applications, reflecting individual domain understanding. In contrast to this, the semantic level contains conceptualizations shared by a certain community, i.e. mainly ontologies.

Semantic Elicitation application areas are

- between the schema of the metadata extraction components and a VIKEF AT,
- between the schema of already existing metadata (from external sources, for example DBLP schema) and an AT,
- between different ATs, in order to re-use previously acquired metadata, and
- between different classification schemata or structures.

Another application area of Semantic Elicitation would be in the area of relating values (not any more on the schema, but on the instance level) of metadata instances.

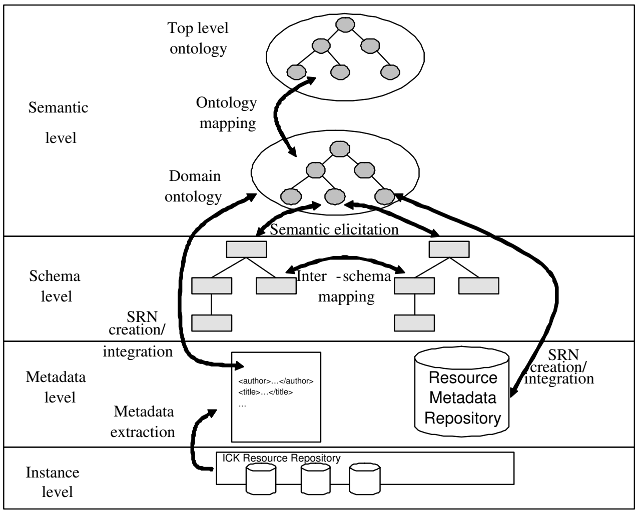

Fig. 3. Schema Levels

Mappings between objects of these levels are required for different purposes. The Semantic Resource Network is itself a result of a mapping between the Schema Level and the Semantic Level that is systematically applied to a set of metadata from the Metadata Level. In every step where data conveying to different schemata have to be related or integrated, Semantic Elicitation will contribute to semi-automate the mapping definition task, levering in high degree the remaining manual work needed. The combination of metadata coming from different sources raises the issue of data cleansing in order to avoid redundant entries in the SRN but this issue is not addressed in this paper.

6 Summary and Further Works

We have presented an approach for schema integration in Virtual Information Environments: contextualizing schemas based on local, structural and lexical knowledge and storing the resulting DL representation persistently within the VIE to allow higher levels of the architecture for intelligent, semantic-based services. This approach has been illustrated to be applicable in a real framework that is currently under development by the Vikef research project.

The idea presented is in an early stadium and has to be researched and verified more in depth. In the course of further works we expect various challenges, some of which are connected to the current implementation of the CtxMatch algorithm, whereas others deal with the usage of the DL representation within a VIE.

Although being self-contained and functional for its original purpose of semantic schema matching, usage of the current implementation of CtxMatch within a VIE has the following drawbacks which will have to be addressed by our research group:

- The input format is restricted to XML Schema.
- The output format of the contextualization functionality is produced for a specific DL reasoner.
- WordNet is used as a provider of both lexical and domain knowledge.
- No secondary ontologies can be added to the contextualization process.

We will also have to take a closer look at the further usage of the DL representation of the contextualized schema. In VIEs built upon standard Semantic Web technologies such as RDF and OWL, it is not yet fully explored how a DL representation will fit into the big picture and be able to provide a sensible base for the connection of documents, local knowledge, schemas and reasoning services. We expect having to define a more generic or intermediate representation of a contextualized schema to be independent of specific reasoning components.

References

[1] Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. *OWL Web Ontology Language Reference - W3C Recommendation*, February 2004. http://www.w3.org/TR/rdf-concepts/.

[2] Massimo Benerecetti, Paolo Bouquet, and Chiara Ghidini. Contextual reasoning distilled. *J. Exp. Theor. Artif. Intell.*, 12(3):279–305, 2000.

[3] Sonia Bergamaschi, Silvana Castano, Maurizio Vincini, and Domenico Beneventano. Semantic integration of heterogeneous information sources. *Data Knowl. Eng.*, 36(3):215–249, 2001.

[4] Tim Berners-Lee, James A. Hendler, and Ora Lassila. The semantic web. *Scientific American*, May, 2001. http://www.sciam.com/2001/0501issue/0501berners-lee.html.

[5] Matteo Bonifacio, Paolo Bouquet, Gianluca Mameli, and Michele Nori. Kex: A peer-to-peer solution for distributed knowledge management. In Dimitris Karagiannis and Ulrich Reimer, editors, *PAKM*, volume 2569 of *Lecture Notes in Computer Science*, pages 490–500. Springer, 2002.

[6] Matteo Bonifacio, Paolo Bouquet, and Paolo Traverso. Enabling distributed knowledge management: Managerial and technological implications. *Informatik - Zeitschrift der schweizerischen Informatikorganisationen*, 1:23–29, 2002.

[7] Paolo Bouquet, editor. *AAAI-02 Workshop on Meaning Negotiation*. American Association for Artificial Intelligence (AAAI), AAAI Press, July 2002.

[8] Paolo Bouquet, Jerome Euzenat, Enrico Franconi, Luciano Serafini, Giorgos Stamou, and Sergio Tessaris. Knowledgeweb d2.2.1 specification of a common framework for characterizing alignment. Technical Report DIT-04-090, University of Trento, Dept. of Information and Communication Tech., http://eprints.biblio.unitn.it/archive/00000653/01/090.pdf, September 2004.

[9] Paolo Bouquet, Fausto Giunchiglia, Frank van Harmelen, Luciano Serafini, and Heiner Stuckenschmidt. C-owl: Contextualizing ontologies. In Fensel et al. [15], pages 164–179.

[10] Paolo Bouquet, Luciano Serafini, and Stefano Zanobini. Semantic coordination: A new approach and an application. In Fensel et al. [15], pages 130–145.

[11] Paolo Bouquet, Luciano Serafini, and Stefano Zanobini. Coordinating semantic peers. In Christoph Bussler and Dieter Fensel, editors, *AIMSA*, volume 3192 of *Lecture Notes in Computer Science*, pages 75–84. Springer, 2004.

[12] G. C. Bowker and S. L. Star. *Sorting things out: classification and its consequences.* MIT Press, 1999.

[13] Donald Davidson. A nice derangement of epitaphs. In LePore [22].

[14] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, May 1998.

[15] Dieter Fensel, Katia P. Sycara, and John Mylopoulos, editors. *The Semantic Web - ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings*, volume 2870 of *Lecture Notes in Computer Science*. Springer, 2003.

[16] Chiara Ghidini and Fausto Giunchiglia. Local models semantics, or contextual reasoning=locality+compatibility. *Artif. Intell.*, 127(2):221–259, 2001.

[17] F. Giunchiglia and P. Bouquet. Introduction to contextual reasoning, an artificial intelligence perspective. In *B. Kokinov: Perspectives on Cognitive Science*, pages 138–159. New Bulgarian University, 1997.

[18] Fausto Giunchiglia. Contextual reasoning. *Epistemologia - Special Issue on I Linguaggi e le Macchine*, XVI:345–364, 1993.

[19] Tom R. Gruber. A translation approach to portable ontology specifications. Technical Report KSL 92-71, Computer Science Department, Stanford University, California, 1993.

[20] Giancarlo Guizzardi, Heinrich Herre, and Gerd Wagner. On the general ontological foundations of conceptual modeling. In Stefano Spaccapietra, Salvatore T. March, and Yahiko Kambayashi, editors, *ER*, volume 2503 of *Lecture Notes in Computer Science*, pages 65–78. Springer, 2002.

[21] Graham Klyne and Jeremy J. Carroll. *Resource Description Framework (RDF): Concepts and Abstract Syntax W3C Recommendation*, February 2004. http://www.w3.org/TR/rdf-concepts/.

[22] Ernest LePore, editor. *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Basil Blackwell, Oxford, 1986.

[23] Jayant Madhavan, Philip A. Bernstein, AnHai Doan, and Alon Y. Halevy. Corpus-based schema matching. In *ICDE*, 2005.

[24] Jayant Madhavan, Philip A. Bernstein, and Erhard Rahm. Generic schema matching with cupid. In Peter M. G. Apers, Paolo Atzeni, Stefano Ceri, Stefano Paraboschi, Kotagiri Ramamohanarao, and Richard T. Snodgrass, editors, *VLDB*, pages 49–58. Morgan Kaufmann, 2001.

[25] Bernardo Magnini, Luciano Serafini, and Manuela Speranza. Making explicit the hidden semantics of hierarchical classifications. In Amedeo Cappelli and Franco Turini, editors, *AI*IA*, volume 2829 of *Lecture Notes in Computer Science*, pages 436–448. Springer, 2003.

[26] John L. McCarthy. Notes on formalizing context. In *IJCAI*, pages 555–562, 1993.

[27] John Searle. *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge University Press, 1979.

[28] Luciano Serafini and Paolo Bouquet. Comparing formal theories of context in ai. *Artif. Intell.*, 155(1-2):41–67, 2004.

A Context-Management Infrastructure for Ubiquitous Computing Environments

Andreas Zimmermann

Fraunhofer Institute for Applied Information Technology
Schloss Birlinghoven
53754 Sankt Augustin, Germany
Andreas.Zimmermann@fit.fraunhofer.de

Abstract. De-contextualization is death for any information. Context-awareness is the key asset in future information services for Ubiquitous and Pervasive Computing. Indeed, contextualization currently lacks a few properties making it transparent and applicable for everyone. The thesis introduces Context-Management as a novel and unified framework in order to hide complex technical details from developers and end users while developing context-aware systems. This framework provides software packages to support developers in integrating context-awareness in ubiquitous computing environments and tools for the management of content combined with context information.

1 Problem Statement

The central issue of research in Ubiquitous Computing is the creation of applications that deliver the right content in the right form to the right person at the right time. The implementation of systems, which are aware of their computing context and the context of the user, demands the consideration of three main functional areas: context acquisition, context synthesis, and context use. Today's terminals provide a range of input and output devices, such as a microphone, a pen that can be used for pointing, handwriting, drawing or sketching, a medium resolution video camera, a medium resolution screen to display video, graphics and text, and a loudspeaker. It is also sure to assume that the next generation terminals will come with wireless interfaces that allow one to connect additional input and output devices, for example GPS and motion sensors, sensors for ambient temperature or humidity, and non-invasive electrophysiological measures of the user's state.

Context-aware computing is a mobile computing paradigm in which *applications can discover and take advantage of contextual information* (such as user location, time of day, nearby people and devices, and user activity). Through the course of a day, users interact with a number of stationary and mobile systems. Two technologies allow users to move about with computing power and network resources at hand: portable computers and wireless communications. The relationship between systems and devices is constantly changing due to user mobility. The application of context-awareness in ubiquitous computing environments suffers from some main problems:

Complexity. The management of technologically highly different sensors and other information sources in combination with different devices for information presentation introduces a high level of complexity. In addition, the technology and algorithms for the interpretation, comparison and analysis of context data is difficult to access, as well as the manipulation of historical context data.

Inexperienced Developers. Product managers, application designers and system integrators know best about demands for context-sensitive functionalities in their application. In parallel, they have to react to a shortened delivery time to be competitive in a dynamic market. As inexperienced developers, they face the crucial challenge of realizing such systems.

Weak Reusibility and Dependency on the Domain. Many context-aware systems are specialized applications, tailored to one domain or environment and rarely reusable in other applications. Common problems that emerge during the development and the reuse of components are the strong dependency on the domain, no open and standardized interfaces, no uniform representation of context models.

Lacking Standards for Exchanging Context Information. Currently, context-aware applications lack standards for exchanging context information as well as a generic architecture and processing pipeline. The absence of a uniform structure of the model which ensures a common understanding of the syntax and semantics at all parts of the application complicates the sharing of context information with other system parts.

Distributed Environment. In a ubiquitous computing environment context is inherently distributed, since entities in the real world are physically distributed. Distribution occurs on different levels: on a conceptual level where information is distributed and on an implementation level where system components are distributed [1].

The key idea of the thesis is to ease the design and development of context aware systems in Ubiquitous Computing by making it transparent and applicable for everyone. By shifting the focus from the developer to a more inexperienced group of users, a *Context-Management System* will provide both software packages for developers and design tools for inexperienced developers or non-experts. A unified framework hides complex technical details and includes the entire design process with reusable components for context acquisition, context modeling, controlling the application behavior and displaying information in the right form.

2 Related Works

In Ubiquitous Computing context-aware systems exploit their surrounding context to increase the fit of a service or an application to the user's needs beyond the user centered evaluation. The term context-aware computing was first introduced by Schilit [2] who describe context as location, nearby people and objects and changes to those objects over time. A general and widely applicable definition of the term context was proposed by Dey [3], who defined context as "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves*". In recent research in context-adaptive systems, numberless approaches for the identification of different context types exist, such as user context, computing context and physical context [4] or identity, location, time, and environment [5]. Other researchers further distinguish between static and dynamic context parts [6]. A distinction of different context types is important, since each type obtains a specific relevance and demands a special treatment. Additionally, when the context-information is recorded across a time span, we can obtain histories like interaction, movement or event histories, which are also useful for certain applications [7].

Context-awareness has been recognized as a vital feature for modern computing applications. Consequently, various context-aware infrastructures and prototype applications have already been developed. Many of these systems have been built application-driven without paying much attention to the reusability of components within other domains. As a very fist step, Schilit [4] presented a system architecture supporting context-aware mobile computing, which made it possible to easily build such applications. The *Context-Toolkit* [8, 9] aims at making the development of context-aware applications easier to perform. In the style of GUI-Widgets, they defined Context-Widgets mainly for supporting the acquisition and delivery of context. In contrary, Brown [10] has focused on how to support application designers with a general architecture supporting non-programmers in building context-aware applications. In recent publications, Schmidt [1] presents a concept for distributing context in a ubiquitous computing environment.

The Context-Toolkit developed by the Future Computing Environments Group at the Georgia Institute of Technology targets to provide a reusable solution for context handling that facilitates the implementation and deployment of interactive context-aware applications. The toolkit incorporates various services related to the gathering and supply of context, including an encapsulation of context sensors, access to context data, context storage, and a distributed infrastructure. Furthermore, the toolkit comprises a design process for building context-aware applications.

The Context-Toolkit addresses many issues related to context gathering and supply, but its main flaw is the absence of a formal context model and means for controlling an application or change variable parameters inside an application. In addition, the functionality of interpreters for context derivation is very limited as they are usually employed for simple data type conversions only. As a result, the support of context comparisons is limited as well.

A recent and ongoing research activity is the approach towards a context-management infrastructure for ubiquitous computing environments pursued by Henricksen [11] at the University of Queensland and the Distributed Systems Technology Centre. This infrastructure facilitates the development of context-aware applications through the provision of generally required functionalities such as context gathering, context-management and context dissemination. These functionalities are based on a context model that addresses the diversity of contextual information, its quality and complex relationships among context data and temporal aspects. The context-management infrastructure is organized into a three-layered hierarchy of interacting components: Context-Aware Application layer, Context-Management layer and Awareness Module layer. The communication among these layers is carried out by a notification service.

The context model provided by this approach contains some simplifications that make a comprehensive representation of complex contextual information difficult, since the question remains how entities' attributes and their values are modeled. In conjunction with the context gathering no augmentation processes have been considered so far. In addition, context data filtering can only be carried out if the data that is to be filtered refers to the same context attribute.

Within the scope of the Context Service project of the IBM Research Center a middleware infrastructure aimed at gathering and disseminating context in ubiquitous computing environments has been developed [12]. This service-based approach provides applications with contextual information at a high level of abstraction. This provision covers the gathering of entities' context from various sensors, maintaining context and responding to queries by clients. Further issues like context quality, context storage and extensibility are also regarded. In contrast to other approaches this project explicitly addresses privacy and security concerns. The used context model bases on a form metaphor which describes a particular type of contextual information and is composed of a set of attributes [13].

The concepts proposed by the Context Service so far lack maturity, since the form abstraction used to represent context evidently does not distinguish between different context attributes. This fact makes it difficult to handle the potential complexity of context. Furthermore, the architecture of the Context Service considers processes of context data augmentation only marginally, since neither context data derivation nor a filtering of context data are carried out.

3 Research Activities

The focus of the thesis was the development of an infrastructure for Context-Management that will allow for the flexible enhancement of existing services and applications lacking context-aware functionality. The new approach of *Context-Management* will allow end-users to easily integrate such functionality in their domain-specific applications. Based on the concrete problem definition and a broad analysis of the state-of-the-art the classical methodological approach continued with the definition of requirements, the conceptualization and design, prototyping, and testing and evaluation.

3.1 Requirements Definition

A brief section of the key requirements when designing a Context-Management platform to ease the development of context-aware applications are presented here.

End-User Involvement. Domain experts and product managers know best about the required system behavior and have been identified as the key people to adapt context-aware behavior. Their involvement in the application design process postulates the provision of an easy management of contextual information.

Reusability. The design of many models has been geared to particular application domains or environments, thus failing to be universally applicable. Reusable standard components need to be solidly founded upon a fundamental understanding of capabilities and limitations of context-modeling and context-management.

Formal Processing Pipeline. In order to overcome the lacking standards for exchanging information, a generic architecture and processing pipeline has to be investigated. The easy sharing of context information with other parts of the system is only achieved by the offering of a straightforward API.

Transparent Context Model Generation. Processing contextual information in a consistent manner can only be facilitated by a formal model and thus by a universal structure for context. In addition, by formal models more flexibility is achieved regarding extensibility and adaptability to changing requirements.

From Sensors to Actuators. Even if there is sensor data and a context model available, there is still a need for control components deciding what consequences must be taken, if the model fulfills certain conditions. Flexible actuator components need to reflect these consequences by mapping them into real world actions.

3.2 Conceptualization and Prototyping

The examination of existing approaches has shown that so far no solutions exist that satisfactorily meet all the requirements. As a consequence, the main research activity of the thesis has been the derivation of novel concepts and unified framework in order to hide complex technical details from developers and end users realizing context-aware systems in ubiquitous computing environments. For this reason, this work introduces the term *Context-Management* as a new design approach:

Context-Management allows the creation, integration and administration of context-aware applications. Context-Management considers the definition of relevant context-parameters, the link between these parameters and information sources, their utilization for and the definition of the targeted adaptive behavior.

The Context-Management System (CXMS) consists of two main parts as shown in the center of Figure 1: The Context-Modeling Tool on the left and the Content-Management System (CMS) on the right hand side. In the following we describe the four main components of the CXMS, the *Context-Toolkit*, the *Design-Tool*, the *Mobile Collector* and the *Content-Player* in more detail.

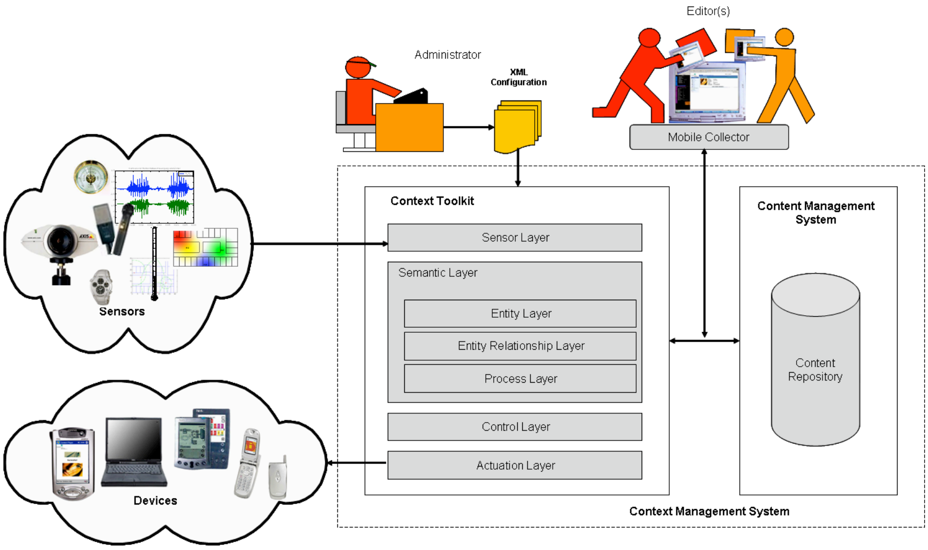

Fig.1 Architecture of the Context-Management System

3.2.1 The Context-Toolkit

In order to achieve reusability and domain-independency in the developing process, the contextualization process has to be generalized. As a result this process is mapped onto a generic four-layer architecture: recognition of relevant parameters (Sensing), model-based interpretation of information (Modeling), interpretation of information and decision making (Controlling), and generating appropriate output (Actuation). Hereby, a component-based approach facilitates the addition, removal and replacement of components on each layer independently.

The developed Context Toolkit (cf. Figure 1 in the center) provides software developers with an extension to the JAVA programming language for the implementation of context-aware applications. The libraries consist of ready-to-use software components that hide complex technical details from the developer. The implementation of the Context-Toolkit complies with the presented layered approach.

The sensor layer consists of software objects that receive incoming data, perform a cleaning and a fusion of sensor values and fire sensor change events notifying objects of the semantic layer. Already existing sensors for position tracking, noise or motion detection can be extended by complying with the sensor interface and offer methods to

access the sensor value and the value quality. In its current implementation the Context-Toolkit does not support automatic detection or recognition of sensors.
Each context attribute on the semantic layer is connected with zero or more sensors and in turn sensors deliver information to one or more attributes, thus creating a cross-linked network between sensors as information sources and attributes as information interpreters. Context attributes receive sensor values, perform a model-based interpretation and map them to semantically more enriched values or derive their values based on algorithmic data fusion from more than one sensor.
The modeling approach chosen in this thesis is the use of attribute-value pairs as a simple but flexible context representation. Each context is an enumeration of one or more context attributes, and each entity owns one static context by default and several types of dynamic contexts in addition. For every type of context, the implementation allows for the configuration of a persistent storage of changes of this type of context as a history. Contexts may be memorized triggered by three types of events: on every change event, based on a fix time interval or explicitly. As a part of the process layer this history function supports the profiling task of personalization engines and enables the derivation of preferences, interests, and so forth.
The Context-Toolkit supports three types of matching procedures for entities: First, two entities can be compared, to see if their contexts are similar to each other (e.g. if users in the same room) and determine certain relationships between entities. Second, a matching can be performed to see, if the entity's context fulfills certain criteria (e.g. for the determination of stereotypes). A third matching procedure enables the filtering of entities from a list (e.g. for a database query). Therefore, the toolkit offers the configuration of qualifiers (results are true or false) and of more complex similarity measures (results lie in the interval [1..0]).
The components described so far form the basis for a rule system that controls the desired behavior of the target application. This rule system consists of a set of hierarchically ordered rules, whereby each rule is composed of a precondition and action part. If the precondition of a rule is fulfilled the rule fires and all associated actions are executed. Preconditions consist of Boolean expressions build up of qualifiers. Currently available actions perform changes to context attribute values and of relations between entities, the selection of content and the presentation of content on several devices. Correspondingly, the indicator/actuator components focus on displaying content on several mobile devices.

3.2.2 The Design-Tool

Due to its compliance to this new standard architecture, its platform independency, and the transparency of the component based system design, the Context-Management System offers different abstraction levels for users on several programming skill levels. Inexperienced users have access to a graphical user interface providing a look-and-feel for a straight forward realization of context-aware systems with predefined components. Experienced developers directly access configuration files, reuse existing or add new software components that are conform to the layered approach for context-aware systems as proposed in the work.

The mode of operation of the software components provided by the Context-Toolkit and the entire design of the targeted context-aware application are to a certain extent configurable using the XML language. Thus, we designed an appropriate design and authoring interface, which is a regular editor for XML files furnished with some intelligence, in order to prevent users from doing mistakes (cf. Figure 2). This interface compounds sub-panels that are necessary for the application design: Sensors, attributes, modeling, control, queries and actuators.
The graphical user interface allows for the addition and removal of pre-implemented sensors and actuators, as well as the creation of the cross-linked network between sensors and attributes mentioned above. Figure 2 shows an example for a context model consisting of attributes with associated sensors. Additionally, the interface allows a preparation of qualifiers and preconditions in a treelike representation and the definition of rules for controlling the application behavior.

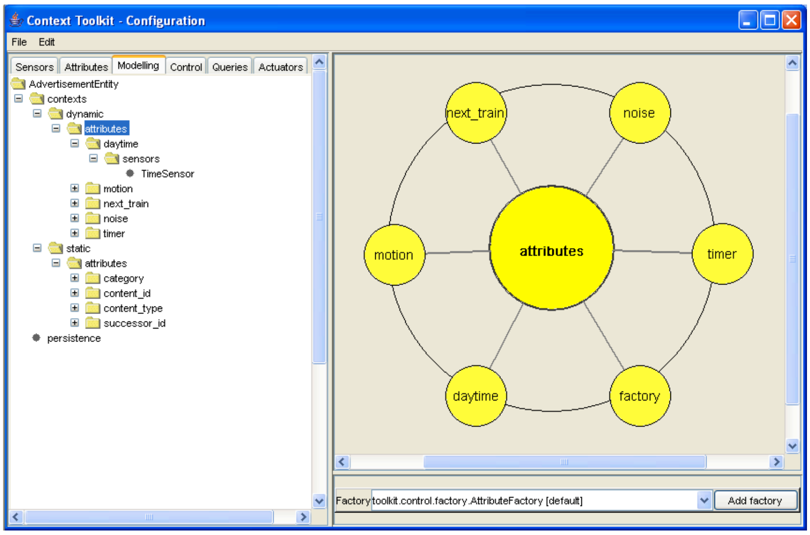

Fig.2 Screenshot of the Design-Tool of the CXMS

3.2.3 The Mobile Collector

The annotation of content with context data is a special task which can most effectively be done on the move or directly in the context of use. Mainly our approach supports users with a tool for “recording” or capturing context data together with content from a Content-Management System. The Mobile Collector offers content providers with an efficient tool for the production of contextualized content.
Figure 4 illustrates the Mobile Collector running on a Tablet PC. The right hand side shows the web front-end of the Content-Management System. This screen provides the author with functionality for adding, removing, searching and browsing content such as images, sounds, videos or even entire HTML pages. The left hand side of the Figure depicts the current context of the device in the lower panel and the current

sensor values in the upper panel. If there are any changes to the values of the sensors or context attributes, both panels are immediately updated. Since the left panel is a browser plug-in, it does not affect the user while browsing the net. It reflects the context model defined with the Context-Toolkit and allows the user to capture the current context (i.e. to freeze its values), to edit context attribute values and to (un)select attributes that are considered (ir)relevant in this specific situation.
If the correct content is selected and the context is adjusted appropriately, the author can easily create the link between those two by clicking the snapshot-button and store this link to the persistent memory of the Mobile Collector. By using a context-snapshot approach authors of content or annotators can capture context meta-data for content. One context-snapshot consists of current values for each context-attribute in the specific moment the snapshot-button has been pressed. Since the Mobile Collector is a tool for collecting context snapshots and linking them to appropriate contents, it is not used for administrative purposes. The removal of the context annotation for a specific content requires the use of an administrator tool at a desktop PC.

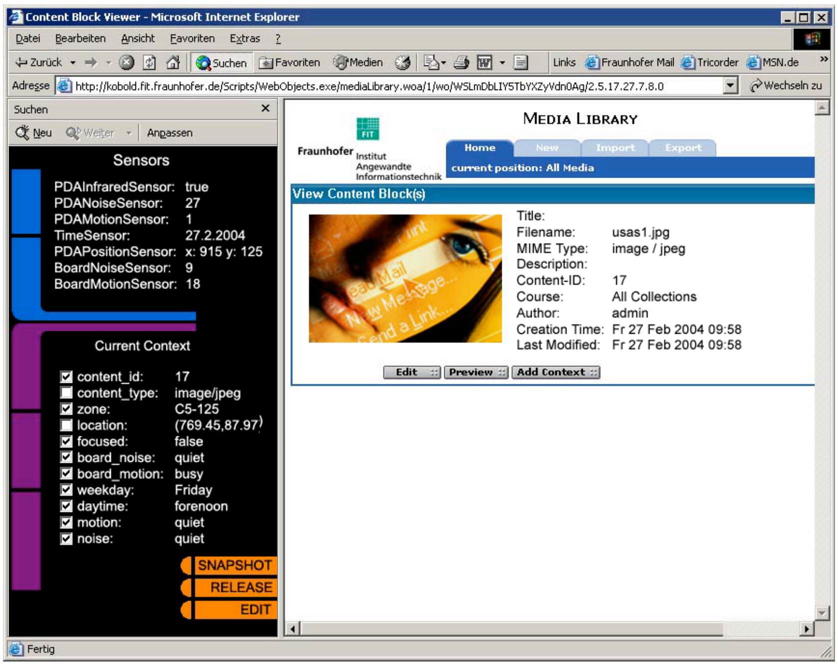

Fig.3 Screenshot of the Mobile Collector

3.2.4 The Content-Player

The Content Player is an adapted browser that runs on a mobile device like a Personal Digital Assistant (PDA). Almost like the Mobile Collector, the sensors connected to the mobile device are readout and their values are sent to the server. The server interprets the sensor values and selects the content from the content-management system that fits best the current situation of the Content Player. The context value vector stored with each content block (cf. Section 3.2.3) is the basis for a filtering process that retrieves content in a specific situation. The retrieval procedure compares the stored context-snapshot with the values of the context the user of the Content Player currently is in and returns the content block associated with the best match. Then the Content Player refreshes the browser and displays the page with the new content.

3.4 Testing and Evaluation

Theory and architecture have been implemented in the form of an operational Context-Management System. An initial set of software components is available to enable the development of context-adaptive services in ubiquitous computing environments. As a proof of concept three exemplary use cases depict the advantages and drawbacks of the approach and serve for evaluation purposes: an intelligent advertisement board, a mobile museums guide and a treasure-hunting game. At the CeBIT 2004 the Fraunhofer Institute FIT presented an intelligent advertisement board as an illustrative description of how the abstract models of the described infrastructure defined earlier are instantiated. The museums guide was part of the developments within the LISTEN project [14], which has been a 3-year research project founded by the European Commission. The treasure hunting game has been developed as an example for a mobile gaming application played by a group of 20 young scientists in June 2004.

References

1. Schmidt, A.: Ubiquitous Computing Computing in Context. Lancaster University, U.K. (2002)
2. Schilit, B. N., Adams, N. I., Want, R.: Context-Aware Computing Applications. In: Proc. Proceedings of the Workshop on Mobile Computing Systems and Applications (1994)
3. Dey, A. K., Abowd, G. D.: Towards a Better Understanding of Context and Context-Awareness. College of Computing, Georgia Institute of Technology, (1999)
4. Schilit, B. N.: System architecture for context-aware mobile computing. Columbia University (1995)
5. Gross, T., Specht, M.: Awareness in Context-Aware Information Systems. In: Proc. Mensch und Computer - 1. Fachuebergreifende Konferenz (2001) 173-182
6. Zimmermann, A., Lorenz, A., Specht, M.: Reasoning From Contexts. In: Personalization for the Mobile World: Workshop Proceedings on Adaptivity and User Modeling in Interactive Systems (2002) 114-120
7. Chen, G., Kotz, D.: A survey of context-aware mobile computing research. Computer Science Department, Dartmouth College, Hanover, New Hampshire (2000)
8. Salber, D., Dey, A. K., Abowd, G. D.: The Context Toolkit: Aiding the Development of Context-Enabled Applications. In: Proc. Proceedings of the 1999 Conference on Human Factors in Computing Systems (1999) 434-441

9. Dey, A. K., Abowd, G.D. and Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Human-Computer Interaction Journal 16 (2001) 97-166
10. Brown, P. J.: The Stick-e Document: a framework for creating context-aware applications. In: Proc. Proceedings of the Electronic Publishing (1996) 259-272
11. Henricksen, K., Indulska, J., Rakotonirainy, A.: Generating Context Management Infrastructure from Context Models. In: Proc. 4th International Conference on Mobile Data Management (MDM) - Industrial Track (2003)
12. Ebling, M. R., Hunt, G.D.H., Lei, H.: Issues for Context Services for Pervasive Computing. In: Proc. Workshop on Middleware for Mobile Computing (2001)
13. Lei, H., Sow, D.M., Davis, J.S., Banavar, G., Ebling, M.R.: The Design and Applications of a Context Service. ACM SIGMOBILE Mobile Computing & Communications (2002) 45-55
14. Eckel, G.: LISTEN - Augmenting Everyday Environments with Interactive Soundscapes. In: Proc. Proceedings of the I3 Spring Days Workshop "Moving between the physical and the digital: exploring and developing new forms of mixed reality user experience" (2001)