

User Aspects of Explanation Aware CBR Systems

Jörg Cassens

Department of Computer and Information Science (IDI),
Norwegian University of Science and Technology (NTNU),
7491 Trondheim, Norway
jorg.cassens@idi.ntnu.no

Abstract. This paper addresses the problem of embedding explanation-aware intelligent systems into a workplace environment. We outline an approach with three different perspectives, focusing on the work process as a whole as well as user interaction from an interface and a system view. The theoretical background consists of Actor Network Theory, Semiotics, and Activity Theory. We further propose to integrate this workplace analysis into a design process for knowledge-intensive and explanation-aware Case-Based Reasoning systems.

1 Introduction

Explanations are an important vehicle to convey information in everyday human-human interaction. They help us to understand one another and enhance the knowledge of the communication partners in such a way that they accept certain statements. The partners understand more, allowing them to make informed decisions. The need for explanations provided by knowledge-based systems is well documented [1,2,3]. The adequacy of explanations is dependent on pragmatically given background knowledge. What counts as a good explanation in a certain situation is determined by context-dependent criteria [4].

Research on explanation is of interest today because it can be argued that the whole scenario on research on Knowledge-Based Systems (KBS) has changed: KBS are no longer considered as black boxes that provide a full solution to a problem. Instead, problem solving is seen as an interactive process (a socio-technical process). Problem descriptions as well as other input can be incomplete and changing. As a consequence, there has to be communication between human and software agents. Communication requires mutual understanding that can be essentially supported by explanations. Such explanations can improve the problem solving process to a large degree.

Case-Based Reasoning (CBR, [5]) is a research area in the field of AI. Its aim is to understand and build systems which are able to use previous experience in order to solve new problems. A CBR system is able to learn by storing experience in the form of so called cases, which describe problems and their solutions. When a new problem arises, a sufficiently similar previous problem has to be identified

and the former solution has to be adapted to the new problem. The new solution might also be based on more than one previous case.

We are suggesting a framework for the design of explanation-aware CBR systems which takes the usage of the system into account. We are therefore in need of methodologies which can describe the workplace environment in which the system is going to be used on the human-computer interaction level. This analysis is then integrated into the design process as a whole, as described in [6].

In order to understand how the system fits into a workplace situation, we propose a theoretical framework which is focusing on three different perspectives:

- Work process view: Actor Network Theory,
- HCI interface view: Semiotics, and
- HCI system view: Activity Theory.

2 Work Process View: Actor Network Theory

We model the context in which the system is implemented with the help of the Actor Network Theory, ANT [7,8]. The basic idea here is fairly simple: whenever you do something, many influences on *how* you do it exist. For instance, if you visit a conference, it is likely that you stay at a hotel. How you behave at the hotel is influenced by your own previous experience with hotels, regulations for check-in and check-out, the capabilities the hotel offers you (breakfast room, lifts), amongst others.

In effect, you are not performing from scratch, but are influenced by a wide range of factors. The aim of ANT is to provide an unified view on these factors and your own acting. According to Monteiro, an actor network in this notion is ‘the act linked together with all of its influencing factors (which again are linked), producing a network’ (see [8, p. 4]).

In this network, you find both technical and non-technical elements. By this, the ANT avoids the trap of either overstating the role of technological artifacts in a socio-technological system or underestimating their normative power by applying the same framework to both human actors and technological artifacts.

This makes it possible for us to understand how technological artifacts influence the doing of human actors in much the same way as other human actors.

3 HCI Interface View: Semiotics

When focusing on the interaction of a particular user with the system, we use the semiotics approach [9,10] to understand the peculiarities of interaction with intelligent systems. In the terminology of semiotics, human communication is a sign process. In contrast, conventional computer systems are only processing signals, lacking the necessary interpreting capabilities humans have.

We argue that in order to make intelligent systems work not merely as tools or media, but as actors to whose decision making abilities a human user can subscribe, the system must appear to the user as if it was capable of a meaningful

interaction. Since both processes, sign and signal, have to be coupled, the goal is to make an intelligent system behave in such a way that the user ascribes to the system the ability to participate in a sign process. The upper-level analysis of the work process helps in defining the aspects of user interaction where this ascription has to succeed in order to make the user believe in the system's capabilities.

One important challenge here is the ability of the system to show its capabilities. This can be described as a communication problem: the system has to interpret the actions of the user in a meaningful way and itself present results that make sense for the user. This process of sense-making is highly interactive: an intelligent partner in a communication process asks (meaningful) questions if an unclear situation occurs and is able to explain its own actions. The semi-otic approach is useful to analyse this sense-making process with the help of transferring knowledge about similar processes from other semiotic domains.

4 HCI System View: Activity Theory

In our framework, we use Activity Theory (AT, [11]) to analyse the use of artifacts as instruments for achieving a predefined goal in the work process and especially to understand the transformation of the artifact itself and the individual and collective work practice during this process.

Since an AI system is more a partner in a work process than a tool, its role in the user interaction changes. Whereas a classical informatics system is a passive translator and memory of praxis, the intelligent system is constantly re-shaping praxis through its use. Looking at a decision support system, the decision making process itself is transformed by the ability of the system to react differently, e.g. through accumulated experience and usage context.

But since AT itself models artifacts as being preformed as socio-cultural entities, we can describe the artifacts in a way which takes this modification into account. Again, our upper-level model helps us to identify the mediation process and the role of both human and non-human actors in the usage process.

The ability of an intelligent system to adapt to the user is very important. In the process of re-shaping praxis, a user expects from an (as-if) intelligent system that it adapts to the changed situation. In the beginning of the usage of a Case-Based diagnostic system, it will be important to explain to the user in detail why a particular case was matched to a new problem, but the user expects from an intelligent partner that the same match will be explained in less detail when occurring very frequently (since the artifact should be changed by the changed praxis, that is here the accumulated knowledge on both parts). On the other hand, in the event of a breakdown situation, the level of detail in explanations given by the system should be increased again.

In addition, the notion of action cycles [12] is helpful for mapping the CBR system model to existing work processes. For example, identifying those situations where feedback to the system is both required and fits into the existing work process helps in avoiding obtrusive system behaviour.

5 Conclusions

The three perspectives presented allow system designers to analyse different socio-technical aspects of the targeted workplace environment. They can be used together to get a more complete model, but this is not always possible or even necessary. For example, in recent work we have investigated how Activity Theory alone can be used to model the knowledge needed in context-aware systems [13].

On the other hand, combining these three approaches allows modelling different aspects of human-computer interaction ranging from the socio-technical network to the design of the user interface itself on the knowledge level. Our goal is therefore to integrate these perspectives further and combine them with other steps of the CBR system lifecycle [6].

References

1. Swartout, W.: What Kind of Expert Should a System be? XPLAIN: A System for Creating and Explaining Expert Consulting Programs. *Artificial Intelligence* **21** (1983) 285–325
2. Buchanan, B.G., Shortliffe, E.H.: *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison Wesley, Reading (1984)
3. Swartout, W., Smoliar, S.: On Making Expert Systems More Like Experts. *Expert Systems* **4** (1987) 196–207
4. Leake, D.B.: Goal-Based Explanation Evaluation. In Ram, A., Leake, D.B., eds.: *Goal-Driven Learning*. MIT Press/Bradford Books, Cambridge (1995) 251–285
5. Aamodt, A., Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications* **7** (1994) 39–59
6. Roth-Berghofer, T.R., Cassens, J.: Mapping Goals and Kinds of Explanations to the Knowledge Containers of Case-Based Reasoning Systems. In Muñoz-Avila, H., Ricci, F., eds.: *Proceedings ICCBR 2005* (to appear). (2005)
7. Latour, B.: Technology is Society made Durable. In Law, J., ed.: *A Sociology of Monsters*. Routledge (1991) 103–131
8. Monteiro, E.: Actor-Network Theory. In Ciborra, C., ed.: *From Control to Drift*. Oxford University Press (2000) 71–83
9. Nake, F.: Human-Computer Interaction – Signs and Signals Interfacing. *Languages of Design* **2** (1994) 193–205
10. Andersen, P.B.: What Semiotics Can and Cannot do for HCI. *Knowledge-Based Systems* **14** (2001) 419–424
11. Bødker, S.: Activity Theory as a Challenge to Systems Design. In Nissen, H.E., Klein, H., Hirschheim, R., eds.: *Information Systems Research: Contemporary Approaches and Emergent Traditions*. North Holland (1991) 551–564
12. Fjeld, M., Lauche, K., Bichsel, M., Voorhoorst, F., Krueger, H., Rauterberg, M.: *Physcial and Virtual Tools: Activity Theory Applied to the Design of Groupware*. *CSCW* **11** (2002) 153–180
13. Kofod-Petersen, A., Cassens, J.: Activity Theory and Context-Awareness. In Schulz, S., Leake, D.B., Roth-Berghofer, T.R., eds.: *IJCAI Workshop Proceedings: MRC 2005* (to appear). (2005)