Integrating Information for Organized Work

Geert-Jan Houben
Eindhoven University of Technology
Dept. of Math. and Comp. Sc.
P.O. Box 513
5600 MB Eindhoven
the Netherlands
houben@win.tue.nl

Frank Dignum
Eindhoven University of Technology
Dept. of Math. and Comp. Sc.
P.O. Box 513
5600 MB Eindhoven
the Netherlands
dignum@win.tue.nl

Abstract

In this paper we discuss our claim that the organization of information in databases, knowledge bases, hypertexts and other integrated structures depends on the processes for which they are used. These processes can be highly structured, completely ad hoc or anything in between. Whether a standard procedure can effectively and efficiently be followed in the process, however, depends on the quality of the information available. The human actors in the process need a mechanism to relate the quality of the information and the quality of the standard procedure to the goals of the business activity. To express this relationship between goals and information processing this paper proposes the use of agent technology. We argue that agent technology can be used to bridge the gap between highly structured situations with high quality data and ad hoc situations where little information is available or the information is not of the right quality.

1 Introduction

One might argue that ten years ago the most important electronically available information sources in companies were the (relational) databases. At that time all information relevant to the processes of the company would be modeled in these databases. Besides these large databases there also existed some small knowledge based systems that provided information for a small amount of specialized processes that needed more complex information.

Proceedings of the 4th KRDB Workshop
Athens, Greece, 30-August-1997

(F. Baader, M.A. Jeusfeld, W. Nutt, eds.)

http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-8/

This situation, however, has changed drastically over the past few years. It is not so much that relational databases are not the standard for information storage (they are still the most prominent means for information storage), but many more forms of storage, dissemination and processing of electronic information have become available and are widely used in practice. Not only are there other individual forms of information processing, we also see a more integrated approach to information system development. This trend towards integrated information systems is illustrated by the attention for concepts like hypertext systems, Intranets, document management systems (e.g. Saros, Documentum), groupware products (e.g. Notes, Exchange), and workflow management systems (e.g. Staffware, Cosa).

While the software tools may be still using traditional database technology, the users perceive that the tools enable them to use new ways of communication. For the users the data and information is stored and accessed in a completely different way as before. This has serious consequences for the way in which information systems are developed.

The traditional perspective of data and information processing does not cover the essential aspects of modern integrated information systems. It is not sufficient anymore to look only at a uniform language to represent information in the traditional database and knowledge base formats, but one should look at ways to access many different types of information intelligently. The key aspects are the integration of information from different sources, and the differences in disclosure principles for information accessed through different tools:

- integrating information from different sources implies mixing information in different technical formats and with different ways of handling by the users;

- accessing information that needs to be accessed through different tools requires a close cooperation between the tools and the need for a tool that allows the users a transparent access.

In this paper we propose to tackle this problem starting from the organizational point of view instead of starting from a technological point of view. That is, the starting point of describing the information should be the humans that need the information to perform some activity within an organization.

2 Quality Of Information

The human users of the information systems figure as the key actors in the processes of an organization. When they perform their activities, they process data and information retrieved from information systems. Traditionally, the approach has been that the architecture of the information system reflects the data processing aspects of some static business process: the designer of an information system has designed the system with a prescribed usage in mind. Therefore, the user is actually guided by the information system itself.

Nowadays much more unstructured information can be stored through hypertext, document information systems, etc. These systems are used as general knowledge repositories and are not focused on one process. This flexibility also has its consequences on the actual information processing. Besides this kind of flexibility information processing is also influenced by the flexibility that businesses want to attach to their business processes themselves: in this paper we will not elaborate on that issue, but assume that the reader is familiar with this desire to achieve increased flexibility.

The consequence of the desire for flexibility is that the use of information is much more controlled by people, as they are better in dynamically assessing the value of information for a business activity. In a large number of organizations knowledge workers need mechanisms to deduce which type of information can be used to perform a business activity. More than before, the knowledge worker must decide on the data to be used, the sources where the data can be retrieved, the format in which the data is retrieved, and the general quality of the information produced. Because information is stored in many different ways, an assessment of its quality is necessary to determine its value for a business activity.

The quality of information is determined by a number of properties. Some of these properties deal with the functionality: "how (with which semantics) is the information used?". Other important quality aspects include time (validity, availability), costs, and resources (human, information). In the end the quality is related to the effectiveness and the efficiency for the activity at hand.

Essentially, the users lack knowledge:

- *Data-knowledge.* Knowledge about the quality of information available within the organization: in order to be able to use a truly integrated information system, users must possess knowledge (meta-information) about the business information; without knowing what the information is worth, they cannot produce an optimal product or service.

- *Process-knowledge.* Knowledge of the activities and processes in which they use the information: in the traditional dedicated information systems the users/system interaction reflects the intended processes; in most of the modern practical cases the processes have a goal of their own, not reflected in the specific information systems: in the abstract integrated information systems it is left to the knowledge worker to use the specific information systems in order to fulfill the process goals.

Traditionally, information system development has focused on the data aspects and the quality of the data for a prescribed purpose. The changes in the way businesses are run imply that this does not suffice anymore.

3 Process Information

An approach that in the last years tries to support knowledge workers in their grip on the processes that play a role in the integrated information systems, is the use of process models. Concepts like workflow management, document management or product data management, start from a view on the information that focuses more on the process, and less on the data.

Considering workflow management (see [Geo95] for an elegant introduction), we see that the information processing in an organization is perceived as a logistical process. In this process the data or information is simply considered as material that must be distributed among the right users. Moreover, the humans involved are seen only as resources. This paradigm is generally valid if the work processes are highly structured, and produce end-products in high volumes. More and more researchers and practitioners find that only in a small number of cases these conditions hold, and that the average process requires a different approach.

There is another drawback. While the design of an optimal logistical process contributes heavily to the performance of the information processing, this approach alone does not quite acknowledge the problem of using information from different sources. The information used in the workflow originates in different databases and knowledge bases, but also comes from the less formal knowledge of the workers involved.

The knowledge of the workers includes information on:

- the data and its accessibility
- the data and its role in the business process
- the process as it is being executed (the status of the process)
- the applicability of a predefined process (procedure) to the specific case on hand (how do the data and process match with this specific case).

This last item, the applicability of the generic processes to a specific case, is in our opinion essential in the development of an adequate business process together with its supporting information system. As in most situations things are not as structured as we have been used to as information systems developers, we need a different approach. In ad hoc or less structured situations knowledge workers feel a need to customize the procedure to the actual needs of the specific case (see for example [Kir94]). We argue that this type of knowledge and insight is not acknowledged in workflow applications.

4 Customization Of Processes

The most characteristic aspect of the attention for less structured processes is dealing with deviations from the predefined standard process. In situations where human office workers remain a vital part of the process, aspects like exceptions, inconsistencies and uncertainties must be taken into account.

These deviations make that processes are customized, which means that they are adapted to satisfy the specific needs of the actual situation. In any process that involves intelligent people making decisions about the way in which work is done, this kind of customization is vital.

The need for customization becomes even more prominent, when the process involves people that communicate. Intelligent office workers that communicate with each other in order to produce a common decision, requires that the processes acknowledge the special aspects of communicating people. Modern information systems need to pay more attention to facilitate this communication: groupware, CSCW, etc. offer ways to implement efficient and effective exchange of information among workers, but these facilities are often not embedded elegantly in the integrated information processing.

Information sharing and work activity coordination should acknowledge the "office memory": the implicit and often informal mechanisms that are used by the employees to communicate and coordinate. There are virtually no office processes that can work without it. As in only a fraction of the cases it is feasible to make the memory itself explicit, in the general case only the use of the memory should be taken into account in the model.

5 Integration Of Process And Data Aspects

The integration of all relevant aspects can be tackled bottom up from the databases and knowledge bases. In the first two sections we have argued that there are a number of difficulties in the bottom up use of information. Especially, it is rather difficult to assess the quality of the information used, with respect to the activities on hand: "what does this information contribute to the product or service that I am producing?".

In the previous section we have argued that the top down approach suggested by the workflow management paradigm, is not always the route to follow. It does not acknowledge that the use of information requires knowledge (meta-information) on the effectiveness and efficiency of using that information.

We feel that a combined approach is necessary (in line with conclusions from [Abb94] and [She96]). Specifically in cases where the processes are less structured, may be even ad hoc, and where the quality of the information is an issue, we argue that an adequate support of the knowledge worker is necessary. They should have the knowledge to assess the combination of procedure and information that is available and relate that to the goal of the activities to be performed. By this we mean that the human actors have the means to decide how an activity is actually performed, while using their professional knowledge of the available data and of the prescribed procedure as a suggestion for an ideal process.

This approach uses a different attitude towards the role that people play in business process and their interaction with information systems. Instead of trying to model and facilitate all exceptions, inconsistencies etc. as being part of the explicit model of the information system, we feel that these deviations must be handled differently: the model must support these deviations by accepting that the explicit model only captures part of the story. The remainder is left implicit, and the information system must be instructed to facilitate people in doing things implicitly within the bounds of the explicit process. This applies especially to the use of heterogeneous information sources.

We aim at using agent technology to assist the users in navigating the different information sources rather than aiming at building a uniform interface to all sources.

6 Agent Technology

In the context of this paper we do not elaborate further on the motivation for studying integrated information systems. This paper contains the characteristics of our proposal for an approach to model and develop information systems.

We suggest an agent-based approach as a basis to support the work of persons in an organization. It takes the persons as central entities in the workflow and models the workflow through commitments (and conversations) between the persons involved. Our approach involves the use of Action Workflow [Med92], where each task can be modeled through the following cycle[1]:

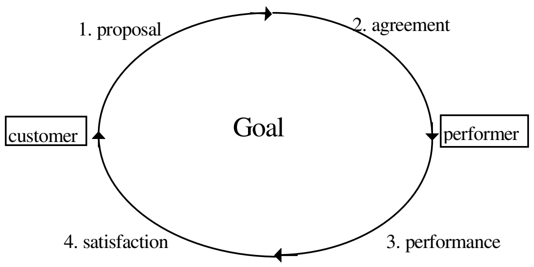

Figure 1: The model for a task (Action Workflow)

In [Hou97] we described how the business processes can be modeled using this approach. There are two important aspects in this approach. First we have given the persons who perform the work a more important role than only that of resource of the process. Using the above cycles makes clear in which context certain tasks are performed. It indicates who is responsible in case of exceptions to the process and also which person has the initiative at every point in time. This is important information in case additional information has to be gathered for the execution of the task.

The second important aspect in this modeling approach is the explicit indication of the goal of the task. Although there usually is some type of standard procedure to reach the goal of the task, most important is that the goal of the task is reached. This makes it possible to create alternative plans when the standard procedure fails to reach the goal.

The third arc in this cycle indicates the step where the actual work is performed. In principle the performer has to perform the following tasks to reach the goal:

- First he assesses whether the standard procedure is appropriate for the present task based on: the goal, a (formal) description of the standard procedure and the quality of the available information. This task acknowledges the fact that in the average business process procedures are only applicable up to a certain level. Moreover, it assures that the goal is considered the essential element, not the activity itself.

- The second task of the performer is to collect the necessary information from the heterogeneous information sources. The information that is needed depends on the procedure that is followed: the standard procedure or an ad hoc solution that is created for the task at hand. This second task involves a global perspective on the common knowledge.

- Finally the performer has to execute some actions that will achieve the goal of the task at hand. Once this is done he will inform the customer, who will accept or reject the result.

In order to support the above steps we propose to use a cooperative information agent as described in [Ver97]. The idea to use agents to support the workers in the performance of their tasks is motivated by the fact that agents are a good platform to describe the goal oriented performance of tasks. The cooperative information agents that we propose to use have the architecture of Figure 2.

The contracts form the basis of the tasks that the agent places on its agenda. The contracts are formed through requests and orders from other agents. These can be seen to take place in the first two arcs from

[1] Note that in the figure we speak about performer and customer. These names refer to the role that persons play in the execution of an action. These names do *not* refer to the role that an external client plays in a business process: the name customer has nothing to do with the traditional role of customer in a business process.

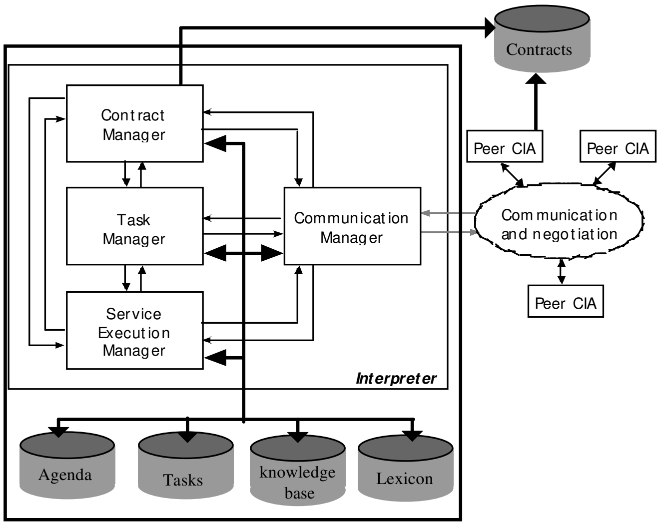

Figure 2: The architecture for cooperative agents

the Action Workflow cycle. The contract contains the goal of the task that is requested. The tasks are assessed by the task manager and the service execution manager. The latter keeps track which tasks can be performed by the agent itself and which tasks have to be requested from other agents. The contract manager keeps track of the status of the contract, i.e. whether the goal is already reached or it can still be reached, whether exceptions occurred and which actions have to be taken in that case. A more thorough description of the agents can be found in [Ver97].

We distinguish two types of agents in the context of work support. For each information resource type there is an agent that can assist the user to access that type of information. The agent that supports the task of the user is called the user-agent, the other agents are called resource-agents.

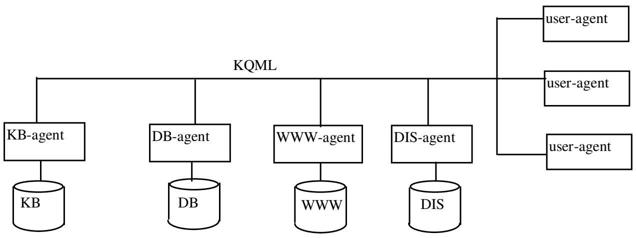

Figure 3: User-agents and resource-agents

The architecture of the two types of agents is identical, but the user-agent contains knowledge about the task it supports and the location and types of information needed for that task. More specifically the agent should have the following elements in its knowledge base[2] to support the task of the user:

1. Formal description of the standard procedure: the starting point,
2. Type and location of necessary information (including persons),
3. Formal description of alternative steps to be taken: the deviations,
4. Goal formation rules to plan the ad hoc procedure, including an assessment of the different possible procedures or alternatives.

The resource-agents contain knowledge about the type of resource they maintain. Each agent contains a meta-data model of the information. Also they can translate the agent communication messages into the language for that particular resource. Using an agent communication language means that each agent only needs one set of translation rules instead of rules between every pair of information sources. As agent communication language we will use KQML. The same approach has been taken in the Infomaster project [Inf97], although the architecture of the agent is different internally.

Of course the translation rules of the agents are quite complicated to develop. However, we expect to build upon the work that has already been done in the area of interoperable and cooperative information systems (see e.g. [Pap92], [Bri92] and [Jam94]).

In practice the approach can imply that the agents help the humans to decide on the procedure to be used. While in some cases the agents can be developed to automatically assess the properties of the procedure and information, in most cases the humans will act as the users of the knowledge delivered by the agents. This means that the humans are in control and they combine their own professional expertise with the knowledge from the agents to decide on the procedure to follow in order to achieve the goal at hand.

7 Conclusion

The modern integrated information systems ask for a new approach to develop information systems. By acknowledging the role that human office workers intelligently play in communication and coordination, the

[2] In this paper we cannot go into more details about the actual knowledge that is needed. For the purpose of this paper it suffices to know that the agents must be able to gather all information concerning the decision which procedure is to be executed.

interplay between activities and goals can effectively be handled. In this paper we have argued that there is a need for a meta-model that facilitates the design of integrated information systems by focusing on the role that humans play within a business. The humans make that the standard procedures are customized to be valid for a specific case. As the quality of this customization appears to be a measure for the success of a business, we have studied the aspects involved in the customization. We have found that the human actors need to have knowledge about a number of aspects, in order to be able to effectively choose the right procedure for a case.

We have sketched here that agent technology can support the humans in gathering the information that they need to assess the procedure, the data involved, the alternatives available and the rules to compose a new procedure. With the use of agents we can model the implicit parts of the processes effectively, thus producing a more robust model of the entire business process and its supporting information system.

References

[Abb94] K. Abbott, S. Sarin. Experiences with Workflow Management: Issues for the Next Generation. *ACM Conference on Computer-Supported Cooperative Work*, 113-120, 1994.

[Bri92] M. Bright, A. Hurson, S. Pakzad. A Taxonomy and Current Issues in Multidatabase Systems. *IEEE Computer*, 25(3):50-60, 1992.

[Geo95] D. Georgakopoulos, M. Hornick, A. Sheth. An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure. *Distributed and Parallel Databases*, 3:119-153, 1995.

[Hou97] G.J. Houben, F. Dignum. Combining activity-based and goal-based models to realize effective and flexible support of group work. submitted for publication.

[Inf97] http://infomaster.stanford.edu/, July 1997

[Jam94] H. Jamil, P. Johannesson. Semantic interoperability - context, issues and research directions. *Proceedings of the second conference on Cooperative Information Systems* (Toronto, Canada), 180-191, 1994.

[Kir94] St. Kirn, R. Unland, U. Wanka. Mamba: Automatic Customization of Computerized Business Processes. *Information Systems*, 19(8):661-682, 1994.

[Med92] R. Medina-Mora, T. Winograd, R. Flores, F. Flores. The Action Workflow Approach to Workflow Management Technology. *Computer-Supported Cooperative Work 92*, 281-288, 1992.

[Pap92] M. Papazoglou, S. Laufman, T. Sellis. An organizational framework for cooperating intelligent information systems. *Journal of Intelligent and Cooperative Information Systems*, 1(1):169-202, 1992.

[She96] A. Sheth e.a. Report from NSF Workshop on Workflow and Process Automation in Information Systems. *SIGMOD Record*, 25(4):55-67, 1996.

[Ver97] E. Verharen, F. Dignum. Cooperative Information Agents and Communication. *Cooperative Information Agents 1st Int. workshop, (Springer LNAI 1202)*, 195-209, 1997.