From Ad-hoc to Engineered Collaboration in Virtual Workspaces

Robert P. Biuk-Aghai
Department of Computer and Information Science
Faculty of Science and Technology
University of Macau
robertb@umac.mo

Simeon J. Simoff
Department of Computer Systems
Faculty of Information Technology
University of Technology, Sydney
simeon@it.uts.edu.au

John Debenham
Department of Software Engineering
Faculty of Information Technology
University of Technology, Sydney
debenham@it.uts.edu.au

ABSTRACT

Distributed collaboration over the Internet has become increasingly common in recent years, supported by various technologies such as virtual workspace systems. Often such collaboration is ad-hoc, and virtual workspaces are set up anew for each new instance of collaboration. We propose that much of the ad-hoc collaboration can be captured and transformed into patterns for reuse in future collaboration. This paper presents the results of the past five years of our work in this area. We introduce the notion of patterns of virtual collaboration; present a framework for extracting patterns of work in virtual workspace systems; and introduce an information model of virtual collaboration. We then present an overview of our data and process mining methods and reverse engineering techniques for discerning work processes carried out through virtual workspace systems. Finally we present our visual mining techniques that we use to discern aspects of work processes in virtual workspaces.

Keywords
Virtual collaboration, patterns, process mining, information visualization.

INTRODUCTION

Distributed collaboration, understood to mean the collaboration of teams across boundaries of space and time and aided by information and communication technology, has become increasingly common in recent years (Lipnack and Stamps, 1997). Many organizations are relying on the ability to bring together people for joint work in a virtual space, without having to bring the people involved together in a traditional face-to-face setting. The diverse technologies that support such collaboration – the collaboration platform – are ideally integrated as virtual workspace systems. The collaboration platform is usually set up using and integrating existing underlying workflow and groupware technologies (Bolcer and Taylor, 1998), and additional communication and collaboration tools, if necessary. The design of the actual virtual workspace can be viewed as a user-oriented semantic arrangement of the different interactive information structures – documents, processing tools, communication channels, etc. – in a way that supports the collaboration processes to be carried out. The aim of this arrangement is to meet some needs or requirements of the collaborators, whether this be an educational, research or business collaboration. Such requirements are usually expressed in terms of activities and their attributes, such as people who are executing those activities, objects/artefacts involved in the activities, etc. Examples of how the notion of activities is used in design, and how a design ontology can be refined are documented, respectively, in (Richards and Simoff, 2001; Simoff and Maher, 1998). The methods for composition of the collaboration platform can be roughly grouped in two categories. The first one, virtual architecture (Maher, Simoff and Cicognani, 2000) attempts to apply the principles and metaphors coming from architecture and building design. The second principle, the process engineering approach (de Vreede and Briggs, 2005), assumes that sufficient a priori knowledge about the collaboration process is available to make it possible to model it. The
Processes can be classified as either deterministic or non-deterministic. In a deterministic process, the steps within the process are well defined, thus the process can be modeled with workflow methodologies, and is referred to as a workflow process. In non-deterministic processes, not all steps can be planned ahead. While workflow processes have received much attention in the literature (Bolc and Taylor, 1998), and are supported by a number of modeling methods, few techniques exist for modeling partially planned or emergent collaboration processes. Such processes are common in information-rich, knowledge-intensive activities such as product innovation, electronic trading, collaborative design or online learning, to name just a few, and usually follow only general process structures, with details of the process emerging during execution. Processes of this type are not well supported by workflow technology, which requires entire processes to be defined in advance, and then enacted according to this definition. Instead, such collaboration processes need a greater degree of flexibility. Environments that are based on the notion of virtual workspaces, incorporating features of document management, interpersonal and group communication, notification, a configurable governance structure, and a set of activities that can be performed in these workspaces, provide a more adequate form of support (Biuk-Aghai, 2000).

Our research over the past five years has been concerned with capturing details of collaboration processes carried out through virtual workspace systems, in order to be used in new collaborative situations. Information on ad-hoc collaboration is obtained from a virtual workspace system, and transformed into a template of a collaboration process that can be reused. Our approach is to obtain patterns of virtual collaboration from a given virtual workspace system, and to transform these into a more abstract form more feasible for reuse. Our approach thus differs from the general patterns of collaboration described by (de Vreede, Fruhling and Chakrapani, 2005). Data mining is applied to basic patterns in order to obtain more abstract processes, and this is facilitated by information visualization. In the following we introduce the components of this work:

**SPACE-DATA-MEMORY FRAMEWORK**

Our goal is to derive actual collaboration processes from virtual workspace systems, and to transform these into templates for reuse. The principal components involved in this effort are: (1) virtual workspaces; (2) collaboration data; (3) data mining; and (4) a knowledge repository. These components are integrated in the 'Space-Data-Memory' framework, which has been presented in detail in (Biuk-Aghai and Simoff, 2001) and is shown in Figure 1.

This framework suggests (1) how to obtain process knowledge from the logs of the virtual workspaces, and (2) how to feed discovered process knowledge back into the ongoing use of the virtual workspaces. In the centre of the framework are virtual workspaces. These are utilized in online collaboration, and through their use data about the collaboration carried out are collected (this is shown in the row labeled "collaboration level"). Collected data feeds into a knowledge discovery component where data mining methods and techniques are applied to discover patterns of collaboration processes. The output of this step is fed into a knowledge repository where it is available for feedback into the virtual workspace system. Thus collaboration processes that start out in an ad-hoc fashion and generate emergent sub-processes have the opportunity to become recurring processes through re-engineering and subsequent reuse. We use the term patterns of virtual collaboration to denote the chunks of processes that can be discovered using this framework. These chunks are discussed in the next section.

**PATTERNS OF VIRTUAL COLLABORATION**

Data collected from virtual workspaces contains patterns of the work carried out through them. These are patterns in a descriptive sense, i.e. structures in a pre-existing body of data that can be extracted using specialized data mining techniques, and that can be explored through information visualization. Descriptive patterns, once discovered and extracted from a body of data, have the potential to serve as sources for reuse, i.e. serving as prototypical prescriptions for use.

**Basic Notions**

A pattern in the descriptive sense is an abstract description of a structure existing in a body of data. For the patterns considered here, the body of data is related to virtual collaboration and originates from a virtual workspace system. It contains details of users, virtual workspaces, artefacts, communication channels, etc., as well as certain relationships among these. This data is interpreted to yield information, by applying pre-existing knowledge about the meaning of certain data items within the context of the given virtual workspace system.
This information is of two types: firstly, information relating to entities within the virtual workspace system, which we refer to as static-type information; secondly, information about the actions that take place within the virtual workspace system, which we refer to as dynamic-type information.

Static-type information represents structures of virtual space, whereas dynamic-type information represents behaviour associated with those structures. Static-type information consists of objects provided and maintained by the virtual workspace system. Examples of objects are virtual workspaces, documents, discussion forums, users, messages, etc. The specific set of objects depends on the given virtual workspace system, although in this regard there is usually a large degree of overlap among different virtual workspace systems.

Dynamic-type information consists of actions that occur within a virtual workspace system. Examples of actions are: creating a virtual workspace, opening a document for reading, posting a statement to a discussion forum, etc. There are also navigational actions, for example, switching from one workspace to another, navigating within a workspace, navigating the threads in a bulletin board, etc.

Collections of instances of similar actions can be generalized into a pattern of virtual collaboration. Below we present examples of such patterns. First, however, we introduce a graphical notation we use for representing patterns of virtual collaboration.

**EMOO Diagrams**

To facilitate communication about patterns of virtual collaboration, we use a graphical notation to represent essential aspects of such patterns. We have extended the MOO (multi-user object-oriented) notation, originally designed for the representation of cooperative business processes (Hawryszkiewycz, 2000), to create the EMOO (extended MOO) diagramming notation for representing patterns of virtual collaboration in virtual workspaces. The EMOO diagrammatic language represents the principal objects and actions involved in virtual collaboration: virtual workspaces, users, artefacts, and communication channels. In our work this formal language is used for knowledge representation of discovered patterns of virtual collaboration. An example of an EMOO diagram involving two users Bob and Alice involved in report preparation and discussion is given in Figure 2. In this diagram, users are represented by ovals, discussion forums are represented by hexagons, artefacts are represented by rectangles with rounded corners, actions are represented by arrows, and the virtual workspace is represented by the containing labeled rectangle.
Patterns of virtual collaboration, as presented above, may be of different granularity. A simple pattern may correspond to a simple action performed by a single user, such as a user posting a statement to a discussion forum; whereas a complex pattern may involve many individual actions performed by a group of users, such as the process of collaboratively preparing a research paper. Larger, more complex patterns are aggregations of a set of smaller, simpler patterns. Virtual workspace systems usually collect information of a fine-grained nature, such as records of individual actions performed by its users. Correspondences between patterns of different granularity can be defined, and collections of fine-grained patterns can thus be transformed into progressively more coarse-grained patterns. This makes it possible to derive patterns of larger-scale activity from the records collected by a given virtual workspace system. To be able to achieve this requires a consistent model that links actions at different levels of granularity, and allows inferences across these levels. The next section introduces a multi-layered model of information in virtual workspaces which forms the basis for performing this pattern derivation.

INFORMATION PYRAMID OF VIRTUAL COLLABORATION

Our model is the Information Pyramid of Virtual Collaboration (Biuk-Aghai, 2003). It consists of information about objects and actions, and their combination into specific patterns of virtual collaboration. This model consists of six levels, as depicted in Figure 3.

![Figure 3. The Information Pyramid of Virtual Collaboration](image-url)
infrastructure, such as web server access requests recorded in a web server log. Patterns of virtual collaboration at this level correspond to events in the software infrastructure.

**System level:** The level of the collaboration system itself, through which collaboration is carried out. Records of objects at this level are contained in the application data of the collaboration system, typically residing in files or database tables. Actions are the commands issued to the collaboration system, such as by a groupware client application. This information is of a larger scale than the corresponding information on the infrastructure level, so a single object or action on the system level usually corresponds to multiple objects or actions on the infrastructure level. Patterns of virtual collaboration at this level correspond to operations performed by the collaboration system.

**User level:** The level on which individual users operate, performing actions on objects residing in virtual workspaces. Objects at this level are virtual workspaces and other objects contained in them, whereas actions are the operations performed by users, such as for instance opening a document for reading. Objects and actions at this level are often abstractions or aggregations of corresponding objects and actions at the system level. Patterns of virtual collaboration at this level correspond to operations performed by a single user.

**Collaboration level:** The level on which multiple users work in collaboration with each other. Objects at this level are virtual workspaces and other objects contained in them, whereas actions at this level are the operations performed by multiple users. Objects at this level mostly correspond closely to those at the user level. However, actions at this level are abstractions of multiple user-level actions. Patterns of virtual collaboration at this level correspond to operations performed by groups of users.

**Task level:** The level where larger-scale activity involving several lower-level actions takes place. Objects at this level are groupings of multiple lower-level objects, whereas actions at this level are the tasks performed by multiple users, consisting of certain combinations of actions and objects from lower levels. Patterns of virtual collaboration at this level correspond to tasks performed by groups of users.

**Process level:** The highest level of the Information Pyramid, here collections of related tasks, constituting work processes, are performed by groups of users. Objects at this level are combinations of multiple lower-level objects involved in the process, and actions are collections of task-level actions. Patterns of virtual collaboration at this level correspond to processes performed by groups of users.

These levels are broadly categorized, relative to the user level at which actual actions are performed by users, as micro, meso and macro level (shown on the left of Figure 3).

**REVERSE ENGINEERING OF COLLABORATION PROCESSES**

The reverse engineering of collaboration processes includes two complementary methods: an automated analytical method for reverse engineering, and a visualization method which is focused on identifying the role of individual components in the collaboration processes.

**Automated Analytical Process Derivation**

The automated analytical method employs the 'Space-Data-Memory' framework and the Information Pyramid, aiming to recover, or discover, the structure of a collaboration process. It expresses it primarily using the EMOO modeling notation introduced above, supplemented with rich pictures and transition diagrams, modeling notations taken from (Hawryszkiewycz, 2000). A rich picture is a diagram that presents a functional perspective of a collaboration process, capturing relationships and connections between people, tasks, and artefacts; whereas a transition diagram expresses the sequencing of tasks in a collaboration process. As the discovered knowledge is also presented back to humans, rich pictures are used for representing entire processes, transition diagrams for showing task sequences, and EMOO diagrams for showing individual task detail. The method proceeds in the reverse order of the methodology presented in (Hawryszkiewycz, 2000): first individual task models are obtained, then these are combined to a process model, and finally a model of task sequences is obtained, as illustrated in Figure 4.

**Task analysis:** We have found that users often create separate virtual workspaces for different tasks. Thus there is a correspondence between space and task such that individual virtual workspaces can be seen as being equivalent to individual tasks. Analyzing a task aims at producing a task model, represented in the form of an EMOO diagram. Depending on the virtual workspace system in which the collaboration was carried out, this may be a straightforward mapping through the Information Pyramid that can be fully automated, or it may require a manual process of identifying and mapping...
modeling elements. EMOO diagrams contain mainly three modeling elements, namely roles, artefacts, and discussion forums, which may be related through certain defined types of relationships.

**Process analysis:** Once task models have been produced, relationships between tasks need to be analyzed in order to discover which tasks belong to the same process. A number of methods are available to aid in this analysis. One method is to analyze shared task elements, such as artefacts, discussion forums, roles, users, etc. The higher the proportion of shared elements between a pair of virtual workspaces, the greater the likelihood that the tasks in the two spaces are related and are part of the same collaboration process. Another method of analysis is to mine traversal patterns between virtual workspaces. This can reveal a network of spaces among which their users traverse back and forth. Such networks are a good indication of related tasks that are part of the same process. A further method is to look for so-called “handover points”, where objects are passed from one virtual workspace to another. Such handovers occur when an object, such as an artefact, is produced by one task as its output, and is received by another task as its input. A handover point usually is a good indicator that two tasks are part of the same process. The output of this step is a process model, expressed as a rich picture.

![Figure 4. Method for reverse engineering of processes](image)

**Temporal analysis:** Once a process model has been obtained, further analysis can be performed to derive a task sequence model. This analysis takes the temporal relationship of actions in different virtual workspaces into account. To determine task sequences, an analysis of temporal action relationships is performed on a pair of tasks taken from the process model. A temporal sequencing of actions in virtual workspaces, and thus of corresponding tasks, is derived, and parallel or interleaved tasks, as well as task loops, are identified. Once all task sequences have been identified, a task sequence model can be produced, represented in the form of a transition diagram.

The Information Pyramid assists with the above described analyses by providing specifications of transformations of information to higher levels. Information on the levels above the base level of the Information Pyramid (i.e. the level at which data logged by the virtual workspace system is obtained, which is usually the system or user level) can be derived from those at lower levels through transformations such as aggregations. This is achieved through initial specification of a model of each level of the Information Pyramid, in terms of its constituent information items (objects, actions, patterns of virtual collaboration), in the form of an ontology of the given collaboration system. This is followed by the specification of the transformations that produce an information item from one or more information items on the level below it. All these specifications become part of an ontology that covers all levels of the Information Pyramid from the base level up, and the transformations between them.

These transformations link consistently the patterns of virtual collaboration between levels, as patterns of virtual collaboration on a given level (with the exception of the lowest level) are aggregations of patterns of virtual collaboration on the level below, as illustrated in the example in Figure 5. Thus an instance of a higher-level pattern of virtual collaboration corresponds to multiple instances of lower-level patterns of virtual collaboration. In this way there is a chain of correspondences of patterns of virtual collaboration from the lowest level to the highest level of the Information Pyramid.

At the end of the reverse engineering cycle, a set of models is available which reflect certain essential process features, expressed in terms of the Information Pyramid of the virtual workspace system from which they were obtained. These can be deposited in a knowledge repository, such as an organizational memory, as expressions of how collaboration has occurred, i.e. as procedural memory, complementing other information on the outcomes of the collaboration. Such process models thus become available for future retrieval and reuse, supporting the design of new virtual workspaces.
Information Visualization in Process Derivation

The derivation of collaboration processes from the data collected by a virtual workspace system is a process that can be partly automated. The above section has introduced one method we have developed to accomplish this. However, process derivation is not focused on identifying the role of individual components in the collaboration processes, for example when a number of overlapping processes occur within the same set of virtual workspaces. Visual exploration of the workspace data can be used to complement the reverse engineering method, revealing relationships among the components of patterns of virtual collaboration.

<table>
<thead>
<tr>
<th>Level</th>
<th>Action Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>Product-Concept-Development</td>
</tr>
<tr>
<td>Task</td>
<td>Product-Brainstorming, Concept-Development, Final-Report-Preparation, Market-Study, Financial-Analysis</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Document-Preparation, Group-Discussion, Document-Sharing</td>
</tr>
<tr>
<td>User</td>
<td>Post-Discussion-Statement, Open-Discussion-Statement</td>
</tr>
<tr>
<td>System</td>
<td>add_statement, get_back_tree</td>
</tr>
</tbody>
</table>

Figure 5. Example of chain of correspondences of patterns of virtual collaboration provided by the Information Pyramid to the reverse engineering method

A specialized tool, the Workspace Visualizer, was developed by us for the visualization of instances of virtual workspaces. The tool utilizes the “heterogeneous network” metaphor, where nodes with different colours represent different types of entities, or different properties of entities. Such a metaphor makes it possible to accommodate three main types of information in the same representational form: (1) details of the relationships among a network of virtual workspaces (shown in inter-workspace maps); (2) the internal composition of these workspaces (shown in intra-workspace maps); and (3) details of the communication patterns within a discussion forum of a workspace (shown in communication maps). An example of an inter-workspace map, displaying relationships between workspaces, is shown in Figure 6.
This map reveals a number of clusters of workspaces that appear to be closely related and could be part of the same work process. Later process analysis can show whether this assumption can be supported. Firstly, task analysis is performed for all workspaces. To illustrate this, Figure 7 shows an intra-workspace map (also produced by our Workspace Visualizer), displaying the relationships among the elements internal to a workspace, such as roles, documents, and discussion forums. Figure 8 shows the EMOO diagram that has been derived from this intra-workspace map.

![Intra-workspace map of the “Propose Change” workspace](image)

Both figures show that almost all assignments of documents and discussion forums to roles in the workspace are identical. The only differences exist in the creation/modification of the Problem and Proposed Change documents (arrow pointing from the role to the document), which may only be performed by the Client and Coordinator roles, respectively. Coupled with the presence of the discussion forums for commenting on the design and discussing changes, this indicates a participatory work process: all roles may read all documents and join in the discussions, while changes to documents are coordinated by having only one role in charge of making such changes.

![Corresponding EMOO diagram capturing essential aspects of the “Propose Change” workspace](image)
To show the internal structure of a discussion forum, and thereby reveal patterns of communication within a work group, our Workspace Visualizer provides communication maps. An example of three communication maps is shown in Figure 9. Communication patterns revealed include: intensity of participation in communication by different group members, lurkers (users who observe but do not post messages), pairs of communicating users (showing “who talks to whom”), directionality of communication and imbalance of communication (e.g. one user sending more messages to another user than receiving messages from that user), etc. Such information can be of value in giving additional information on the nature of the work performed within virtual workspaces by a given group of users.

![Figure 9. Communication maps revealing different communication patterns](image)

Given an ontology of the collaboration system and a set of rules, the task analysis and derivation of corresponding patterns of virtual collaboration can to a great extent be performed automatically. Information visualization of various aspects of the collaboration can hereby play a valuable role in “filling the gaps” where derivation needs to be performed manually.

**CONCLUSIONS**

The ability to extract information about virtual collaboration from virtual workspaces, in the form of patterns of virtual collaboration, offers the opportunity for utilization of the data collected in the virtual workspace system, and the future reuse of extracted information. This paper has presented a novel methodology for reverse engineering of virtual collaboration, expressed as collaboration processes performed through virtual workspaces. It produces design models at micro (task) and macro (process) levels of these collaboration processes using notations from a design methodology intended for virtual collaboration. Through the presented methodology, it not only becomes possible to trace the evolution of processes from an initial design, in the case where such a design has been performed, but also allows the discovery of ad-hoc and emergent processes for which no such initial designs were created. In both cases, processes obtained through reverse engineering can be retained in a library, or case-base, of reusable process templates.

The methodology is based on the combination of a new model of vertical information integration related to virtual collaboration, the Information Pyramid of Virtual Collaboration, which encompasses information about objects and actions that constitute patterns of virtual collaboration at different levels of granularity. Information on different levels ranges from fine-grained system events to entire collaboration processes, integrated through defined transformations that specify how an information item on one level can be derived from information items on a lower level. Definitions of information at different levels, and of transformations, are specified in an ontology of a given virtual workspace system. The complete information model makes it possible to abstract from fine-grained events to large-scale collaboration activities and to drill down from larger-scale activities to their constituent smaller-scale activities.

The integration of the Information Pyramid and the ‘Space-Data-Memory’ framework provides the basis of a process reverse-engineering methodology for discovering knowledge about collaboration processes in virtual workspaces. The presented methodology is independent of the underlying virtual workspace system employed, and only requires knowledge of its ontology. Only the concrete implementation of the data mining methods used needs to be adapted to the given virtual workspace system so as to capture different elements needed in the calculation of process predictors. Likewise, the interpretation of discovered patterns will need to be framed in the context of the virtual workspace system utilized.

The proposed approach has the potential to influence:

1. the way virtual workspace systems are designed, or redesigned. Insights obtained through the analysis of collaboration processes can, for example, reveal deficiencies in the levels of support provided by a particular collaboration system implementation, leading to a redesign of a future version of the system;

Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA August 11th-14th 2005
2. the way multi-agent systems are designed in order to extract and utilize information about the behaviour of individual agents and their societal structures. The principle of the Information Pyramid can be applied to the design of multi-agent systems, where the data mining technology is embedded in the systems and discovered interaction patterns point to appropriate predefined agent configurations.

In this way, the approach can become the backbone of a new design methodology—design of collaboration and multi-agent systems by adaptation.

Finally, the illustrated combination of data mining techniques and reverse engineering, and the availability of a rich source of data on actual collaboration practices in industry and academia can lead to a better understanding of the influence of computer mediation on collaboration processes. Our future work will focus on the analysis of the artefacts and their content involved in the collaboration process, and on the interaction between the entities engaged in collaboration processes.

ACKNOWLEDGMENTS

This research has been supported by the Australian Research Council, the University of Technology, Sydney and the University of Macau.

REFERENCES