Multimedia Mining of Collaborative Virtual Workspaces: An Integrative Framework for Extracting and Integrating Collaborative Process Knowledge

Simeon J. Simoff
Faculty of Information Technology
University of Technology, Sydney
P.O. Box 123
Broadway, NSW 2007
Australia
simeon@it.uts.edu.au

Robert P. Biuk-Aghai
Faculty of Science and Technology
University of Macau
P.O. Box 3001
Macau S.A.R.
China
fst.robert@umac.mo

ABSTRACT

Collaborative virtual environments are becoming an intrinsic part of professional practices. In addition to providing collaboration support, they have the potential to collect vast amounts of multimedia data about the actions and content of such collaborative activities. The aim of this research is to utilize this data effectively, extract meaningful insights out of it and feed discovered knowledge back into the environment. The paper presents a framework for integrating multimedia data mining techniques with collaborative virtual environments, starting from early conceptual development. Discovered patterns are deposited in an organisational memory, which makes these available within the virtual environment. Some of the ideas are illustrated by an example from the application to collaborative spaces developed in LiveNet, a virtual workspace design system.

Keywords

Multimedia data mining, knowledge extraction, collaborative virtual environments

1. INTRODUCTION

Collaborative virtual environments (CVE) have become increasingly popular in recent years. As a result, teams in many companies and organisations are conducting their projects via one or another form of such environments. The notion of CVEs span a broad range of distributed systems—from text-based virtual environments (text based MOO/MUD and Web-based WOO environments [15]) to desktop virtual worlds (an example of such environments with different interfaces and information organisation is presented in Figure 1 and Figure 2) and immersive virtual worlds (for an excellent taxonomy of the latter see [5]). There are numerous approaches and techniques for designing such environments to support collaborative projects [14]. These approaches and techniques use different ways of formalising the

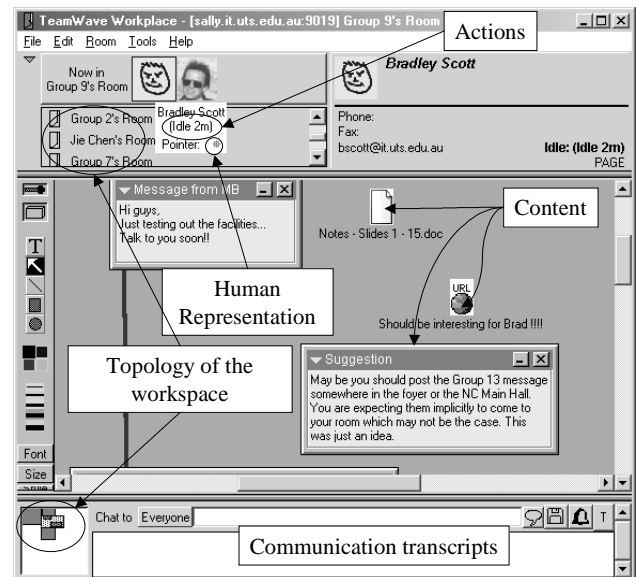requirements for the environment and the design of the project workspace.

Figure 1. Example of a typical workplace in a collaborative virtual environment with 2D interface

Common design strategies are based on the utilisation of experts' knowledge in a top-down analysis cycle. At present, requirements formalisation methods for such environments are based on two major types of methodologies:

- Domain-dependent formalisation based on an ontology of a particular metaphor, usually coming from the domain in which the environment is expected to be applied. For example, Gutwin and Greenberg [10] recognized the importance of the ontology of a "place". Simoff and Maher [18] presented the architecture of a virtual "place" that follows the ontology of a university.

- Domain-independent formalisation of project (or business process) activities (and their attributes) based on soft-systems methodology [16].
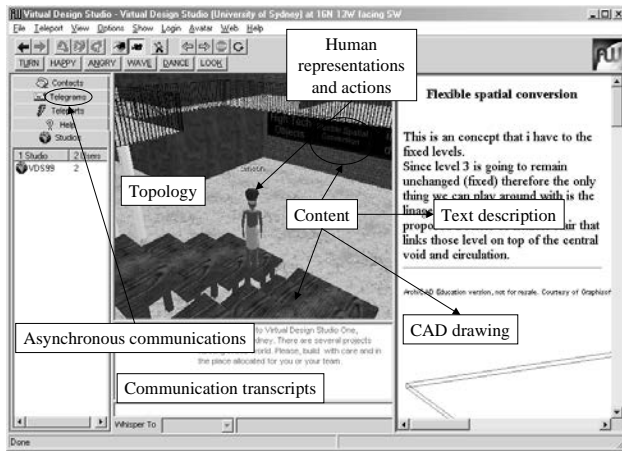

Figure 2. Example of a typical workplace in a collaborative virtual design studio (with 2D/3D interface)

Despite their variety and difference in functionality these environments have several key concepts in common:

- the concept of "*inserting people*" in the collaborative virtual environment, in other words, representing people as some entities, which range from the so-called "characters" or "avatars" in social virtual worlds [15] to sophisticated user profiles in collaborative e-market environments [17];

- the concept of "*structuring the space*" in the collaborative virtual environment, in other words, providing some way of configuring the layout of the space, separating and handling different information within the units of this structure, and some reference system for orientation and navigation;

- the concept of "*media usage*" which defines the feasible set of actions that can be performed in the collaborative virtual environment and the types of electronic media (file formats) that can be used and manipulated in the collaborative workspace [14]. The media usage defines to what extent the environment under consideration can be used for conducting collaborative projects in a particular domain.

1.1 People

Object representations of a person include characteristics such as a text description, messages about their movements in the place, and links to web pages to help establish their identity and personality. An important aspect of people's representation is the variety of "rights" that can be assigned to them. Different environments use different terms for this—privileges, roles, permissions.

Thus, the representations are potential sources of preliminary information about a person's individuality. However, in collaborative projects it is important to be able to make judgments about the individual's working preferences and the collaboration styles of the people in a team or in different teams and to reuse such knowledge when forming teams in other collaborative projects. The preliminary information is not always sufficient for establishing successful work. Data mining techniques can be applied for extracting information about the individuals, the functioning of groups of individuals and to discover patterns of collaboration based on project communication between them. This knowledge can be reused for configuring groups in new projects.
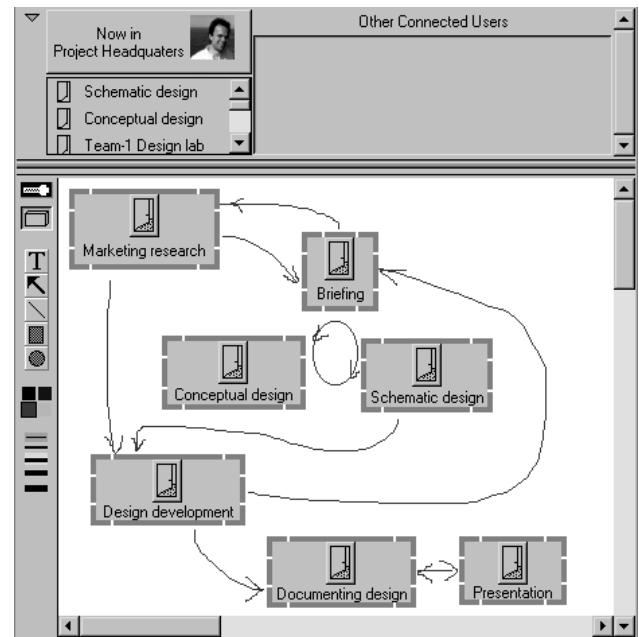

Figure 3. Pre-defined topology of a collaborative virtual environment for design projects

1.2 Space Structuring

The ways of structuring of the environment's space depend on a number of factors, including the ontology (what kind of place the environment is), purpose of the environment, the embedded functionality, the preferred communication and collaboration mode [14], underlying technologies and their integration [18]. Figure 3 illustrates the topology of a design environment, predefined by the model (ontology) of the design process. Each "doorway" leads to a room (workspace), which contains the relevant information for each design stage. The model used is derived from the existing models of the design process in the research literature (e.g. [13]). Further, such schema can be used as a prototypical workspace. However, such top-down approach does not capture the knowledge from the actual use of the virtual environment—which parts of it were used more intensively, what are the

"neighbouring" relations (e.g., co-visited rooms) and other relations, document contents during different phases, etc. Data mining techniques can be applied for discovering such relations. Discovered knowledge can be reflected in variations of the space structuring of the "design prototypes", resulting in building a library of such prototypes and reusing them according to the requirements of the new project space.

1.3 Media Usage

The ontology of the virtual environment can also provide substantial *a priori* knowledge not only about the possible navigation, but also about the set of feasible actions in such an environment. Usually the initial set of actions is derived from the design requirements. The real set of actions used in different projects, however, may vary substantially. The overlapping set of actions forms the common kernel set, and the remainder is the individual component. In the long term, this provides a potential for designing pro-active prototypes supporting different types of projects. Digital media is the basis on which project information can be shared across a network. In collaborative virtual environments the focus is on the following types of digital media: communication and action data, images, CAD models, text, links, audio and video components, and virtual reality representations. Data mining techniques can be applied for composing the action sets. Discovered action sets and space structures will form pro-active design prototypes, resulting in a library of such prototypes and their reuse according to the requirements of a new project. The content mining techniques delivers content features that then are related to the actions.

In this paper we present a framework for integrating multimedia data mining in the design framework of CVEs, in a way that facilitates not only the data collection and analysis, but also the application and integration of discovered knowledge. We illustrate some aspects of the application of the framework for monitoring and extracting knowledge from collaborative activities and incorporating that knowledge back into the collaborative environment.

2. THE FRAMEWORK

The presented framework embeds knowledge discovery in CVEs. Its two primary goals are:

1. To influence the design of CVEs so as to provide the data necessary for mining and analysis of collaboration.

2. To feed extracted knowledge back into the use of CVEs.

As a result, data design and design of the collaborative virtual environment are seen as complementary and parallel activities, offering the opportunity to control data collection to a greater extent. Knowledge obtained from collaboration data is a further contributor to CVE design. A number of related research efforts are underway in the direction of controlled data collection, carried out mainly in the field of e-commerce and Web data mining [19].

The framework for multimedia data mining in CVEs is shown in Figure 4. It includes four major groups of inter-woven components:

1. *Collaborative virtual environments.*

2. *Collaboration data.*

3. *Knowledge discovery.*

4. *Organisational memory.*

Moreover, the three components appearing in the upper part of the figure consist of three parts, at different levels of abstraction:

1. *Conceptual level.*

2. *Structural level.*

3. *Collaboration level.*

Below, we discuss the components of the framework in more detail.

2.1 Collaborative Virtual Environments

CVEs are the support systems within which collaboration is carried out. CVEs are becoming increasingly part of professional practice. Such environments aim to support certain work practices, hence are domain-specific. For each domain, an understanding of the domain-dependent requirements for the CVE has to be obtained.

On the conceptual level this activity identifies the concepts to be supported by the environment: the structuring metaphor employed, navigation facilities, representation of people and their abilities, artefacts and tools provided in the environment, etc. On the structural level, this initial step is followed by the actual design of the CVE when the relationship between the identified concepts is established and their detail is elaborated. Once designed (and implemented), the CVE is utilized by its users on the collaboration level.

2.2 Collaboration Data

The activities related to CVEs are paralleled by those related to collaboration data. Within the presented
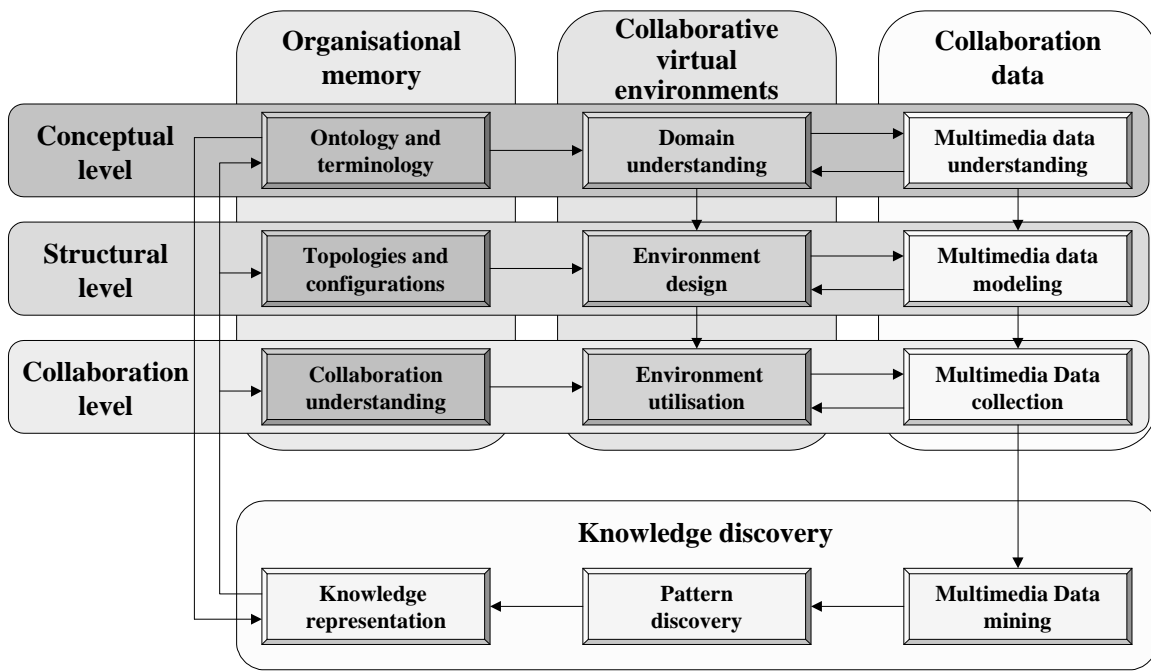
Figure 4. Framework for integrating data mining in the design and application of collaborative virtual environments and knowledge extraction from them

framework, collaboration data is that portion of data which facilitates knowledge discovery within the domain of collaboration, regardless of whether it is of direct use within the CVE. Traditionally, virtual collaboration systems did not provide any particular support for data collection aimed at knowledge discovery. Data was seen as an internal aspect of the system and only internally required data was maintained. The presented framework emphasizes the need for additional data that can enable knowledge discovery in the collaboration domain, and therefore within this framework collaboration data is treated separately from the CVE.

On a conceptual level, domain understanding within the CVE sphere and data understanding within the sphere of collaboration data are mutually complementary: once domain understanding identifies a concept to be supported, data understanding identifies the necessary data elements. Such data elements include fields of log files, the different multimedia document and file formats involved in collaboration, the transcripts from synchronous chat communications, the thread structure and content of asynchronous bulletin board discussions, the audio/video communications, and the integrating data descriptions, which define the relations between the different types of media data.

On the structural level, during environment design, data modeling identifies details of and relationships among the collaboration concepts and data.

Finally, on the collaboration level, the CVE is utilized in an actual project, which generates the collaboration data. These data are collected for subsequent data mining.

2.3 Knowledge Discovery

The knowledge discovery in this framework differs slightly from the classical schema [7]—the selection and data pre-processing stages are implicitly embedded in the data design. Therefore, collected data is expected to be ready for the application of multimedia (or rich media) data mining methods. As a further difference to the classical knowledge discovery schema, a step of knowledge representation is explicitly included at the end. Its purpose is to map discovered knowledge back into the CVE's representation.

Knowledge discovery in this framework aims to produce a better understanding of computer-mediated collaboration, and to enable the usage of discovered knowledge to improve structural features of the workspace configuration and media content when new projects are conducted in the environment. For example, through the analysis of the structuring of virtual environments, templates of structures of these environments can be collected, implying certain navigation behaviour. Collecting data about actual navigation within the environment can provide a source for discovering traversal patterns, which can provide indicators for improving the topology (structuring) of the environment.

Other possibilities for improvement of the environment exist according to particular collaboration and business

process needs. This is something difficult to know ahead of time. The development of such environments follows emergent and adaptive strategies, rather then predefined topologies. In both cases, some necessary indicators for improvement of the structure are required.

2.4 Organisational Memory

Over the past decade, the CSCW community and related areas have taken a keen interest in organisational memory (OM) [1,3,6]. This suggests that there is value in retaining and later drawing on historical records of virtual collaboration. Such records may be referenced when setting out on new virtual collaboration, to "see how others have done it", and perhaps to reuse and re-enact parts of those collaboration instances. Unlike conventional work settings where details of collaboration have to be collected manually through effort-intensive and sometimes intrusive methods, CVEs are an ideal source of data on collaboration, particularly when work is predominantly or entirely carried out virtually, as such environments can automatically record a great amount of detail on the collaboration.

While much work in organisational memory concerns itself with the content of collaboration, or the *declarative memory*, little work has been done on harnessing the *procedural memory*, or knowledge about how work has been carried out. The importance of utilising this aspect of organisational memory in groupware systems has been pointed out relatively early on [6], and again more recently within the context of virtual team effectiveness [8]. The presented framework makes the procedural portion of organisational memory an integral part of the collaboration support environment by maintaining knowledge extracted from collaboration environments and making it available within the environment.

On the collaboration level, this knowledge relates to an understanding of the collaboration. For example, it can identify what main types of activities were conducted within a virtual environment, how the activities were carried out over time, what differences exist in the activity of different people within the environment, etc. This knowledge can be utilized within the environment itself, leading for instance to an adaptation of the environment itself and/or its interface in order to facilitate the execution of predominant activities. It can also serve as a management and control instrument, which is of particular value when collaboration is completely virtual and traditional management methods are severely limited.

On the structural level, representations of the environment's topology are maintained. Where structural patterns are discovered in a set of environments, this too is deposited in the organisational memory in the form of different topologies and configurations available for reuse. Such information may feed back into environments in use,

for instance to rearrange the environment's topology if its current arrangement is discovered to encumber work.

Use of CVEs may, over time, also lead to the emergence of new concepts, or an application of existing concepts in ways that were not previously anticipated. These are deposited on the conceptual level as modifications to the underlying ontology, and feed into the ongoing development of a CVE. An example of this is where an environment lacks a certain feature, but where users discover workarounds that, though cumbersome, allow the feature to be supported. Discovery of such cases can be of use in the development of the next version of the CVE to explicitly support the feature.

In the next section we present an example of the application of the framework for extracting knowledge about different aspects of computer-mediated collaboration.

3. EXTRACTING KNOWLEDGE FROM COLLABORATIVE ACTIVITIES

The example of the application of the framework is related to a particular system for developing collaborative virtual workspaces—LiveNet. The framework was particularly applied in the areas of CVEs, collaboration data, and knowledge discovery. We start by introducing the LiveNet system, then show how the framework was applied.

3.1 LiveNet as a Collaborative Virtual Environment

LiveNet is a virtual workspace design system, developed at the University of Technology, Sydney [12]. It supports mainly asynchronous collaboration of distributed groups of people, i.e. different-time, different-place interactions, although its design does not limit it from other modes of collaboration. A central server is accessed across the network through one of several client interfaces, most commonly through a Web interface (as shown in Figure 5). LiveNet provides virtual workspaces which bring together people, artefacts (e.g. multimedia documents), communication channels, awareness facilities, and a collection of tools, all tied together through a configurable governance structure. A simplified ontology of the basic CVE design unit in LiveNet is shown in Figure 6. In terms of the ontology, workspaces contain roles, occupied by participants (i.e. actual people), who perform actions. Some actions may operate on artefacts, others may be interactions with other workspace participants through discussions. However, most workspace elements such as documents, discussions and participants, may be shared between workspaces. Thus workspaces are not just stand-alone entities (units) but nodes in a network of inter-connected collaboration spaces. Neither are structures of workspaces in LiveNet collaboration space static—once

created, a workspace can be dynamically adapted to evolve—incorporate work subspaces, or attach to other workspaces, together with the project collaboration carried out in it. As a result entire "ecologies" of inter-connected workspaces can co-evolve. Mining the multimedia data in such evolving workspaces can provide invaluable insights about the development of such "ecologies".
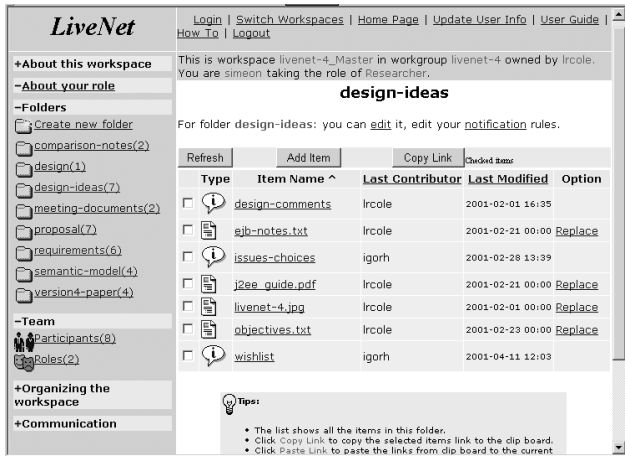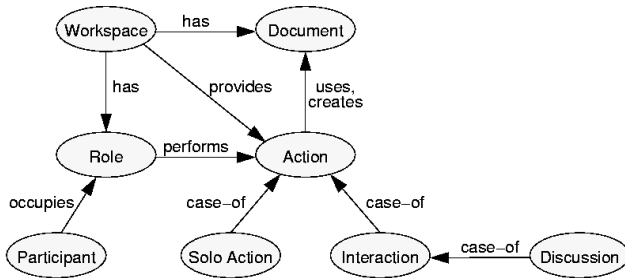

Figure 5. Typical LiveNet screen (web interface)

Figure 6. Simplified ontology of LiveNet

3.2 Integrating Collaboration Data

Early versions of LiveNet had not been developed with the support of data mining and knowledge extraction activities in mind. Consequently, only a limited amount of collaboration data was available, namely only data which was necessary for the internal operation of the system. While this allowed some structural-level features to be extracted, there was no data to support the extraction of collaboration-level or conceptual-level features.

Subsequently, the provision of suitable collaboration data was "retrofitted" onto an earlier version of LiveNet. Building on the existing domain understanding, conceptual data requirements were developed, followed by data modeling. These were integrated into LiveNet by appropriately adapting its design (and implementation)—corresponding to the flow from data modeling to environ-ment design in our framework. Finally, collection of the

new collaboration data and knowledge discovery followed.

The first iteration of this cycle lead to some knowledge extraction, both on the collaboration and the structural levels. However, analysis of the collaboration also revealed that certain data elements, which were not captured at the time would be needed to provide a more complete picture of the collaboration. This had not been accounted for in the first cycle of integration of collaboration data. Consequently, a second cycle was initiated in which data understanding and data modeling were refined, and environment design was brought up-to-date with the new data model. The subsequent data collection and data mining lead to a more comprehensive analysis of collaboration and a richer knowledge discovery. Following this second cycle, new data require-ments are already emerging which, once implemented, will lead to a yet richer body of collaboration data. This confirms to us the validity of our framework in feeding discovered knowledge back into the ongoing development of the collaborative environment. It also highlights the fact that this is likely not achieved in a single effort, but is an iterative process, with insight from each iteration triggering a new iteration.

3.3 Knowledge Discovery in LiveNet

Collaboration data in LiveNet consists of two parts: a database contains the internal data of the CVE, maintaining the current state of all workspace elements (documents, roles, participants, etc.). The second part is a set of log files that are external to the system itself and which record all user actions carried out in the system over time. Although the vast majority of users interact with LiveNet through a web interface, the log records captured by the LiveNet server are on a semantically much higher level than those in the corresponding web access log. While a web log includes IP addresses, document names, timestamps and http request types, the LiveNet log records information in terms of the LiveNet CVE's conceptual model. Thus every record includes the name of the workspace and its owner, the name of the participant carrying out the action, his/her role name, the LiveNet server command requested, etc. This allows analysis to exploit metadata available in the application and to capture higher-level actions than a mere web log does (this corresponds to the approach of [2]).

The analysis we carried out focused primarily on the log of collaboration actions, and to a lesser extent on the workspace database. It involved pre-processing of the log, visualization of workspace data, and actual data mining. The pre-processing step normalizes session numbers, aggregates lower-level events into higher-level actions, and calculates session summaries. In this context, a session is the sequence of actions carried out by a user from login to logout time. Data pre-processing is

considered part of collaboration data collection and is usually automatically performed.

The data used originated from students and instructors of a number of courses at the University of Technology, Sydney, who used the LiveNet system both to coordinate their work, and to set up workspaces as part of the students' assignments. The data covers a three month period, with a total of 571,319 log records, They were aggregated into 178,488 higher-level actions in a total of 24,628 sessions involving 721 workspaces and 513 users.

3.4 Space Structuring

During knowledge discovery, using visualization certain of the relationships existing within and between workspaces can be discovered. This particularly aids exploratory analysis, when the purpose is to get an understanding of the structure of, and patterns in, the data. We selected data originating from students of one course who used LiveNet during the mentioned period. There were a total of 187 student users, organised into 50 mostly 3-5 person groups, whose use accounted for about 20% of the above-mentioned log data.

Initial visualization focused on networks of workspaces, to discover how individual student groups partitioned their work in terms of distinct workspaces, and to what extent these workspaces were linked to one another. This exploratory analysis revealed two distinct patterns: the majority of users preferred to use just one workspace to organise all their course work (such as posting drafts of assignment documents, discussing work distribution and problems, etc.). This workspace tended to contain many objects—or have a high *absolute workspace density* [4]. We term such groups *centralizers*. On the other hand, a few groups tended to partition their work across a collection of connected workspaces, usually with a separate workspace for each major course assignment. These workspaces tended to contain fewer objects (having a lower absolute workspace density) than the ones of the centralizers. We term these groups *partitioners*.

Figure 7 shows a map of LiveNet workspaces with colours highlighting absolute workspace density—lighter colour indicating lower density, darker colour indicating higher density. Branching out from the central node at the top are networks of workspaces for three groups. Nodes represent workspaces, edges represent hierarchical relationships between workspaces. What the map reveals is that the group on the right, Team40, has a very high density in the workspace used for facilitating its work (the workspace Team40_Master). Moreover, it uses only one workspace for this purpose. Thus the right group is a typical example of a centralizer. On the other hand, workspaces in the group at the centre have a much lower density. Out of the eight workspaces in this group, six are used for facilitating

aspects of the group's work. This is indicative of a partitioner group.
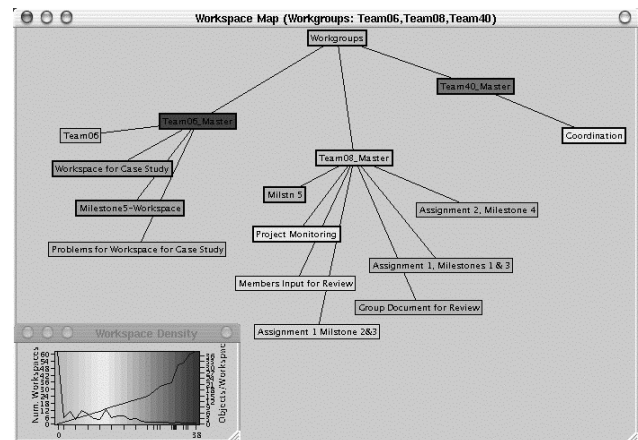

Figure 7. Workspace densities of three different groups

There are plausible explanations for both the centralizer and partitioner cases. Both approaches have their own advantages: in the centralizer case, it is convenience in not having to create multiple workspaces, to switch between them, and in addition to have everything available to all participants in a single location. In the partitioner case, the advantage is increased clarity, structuring according to task, and consequently reduced cognitive load. Furthermore, some groups may bring certain preferences as to the way to organise their work into workspaces and enact these preferences in the way they structure their virtual working environment. When such preferences are recognized during knowledge discovery, and deposited in the organisational memory, they can feed back into the design of new virtual collaboration environments, thus helping to offer more adequate support to cooperative groups with diverse working styles.

3.5 Media Usage

A further area we investigated was focused on identifying which actions different groups mainly carried out within LiveNet. All in all, 80 different actions are available in LiveNet. The majority of student groups used only about half of these. The major actions carried out are related to the main LiveNet conceptual elements: workspaces, roles, participants, documents, and discussions. A taxonomy of these actions is shown in Figure 8.

While all groups had been given the same task—to prepare a number of assignments and to set up a collection of workspaces to support a given process—the way they implemented this task varied markedly. This was evident in a number of aspects of their use of the LiveNet system, such as intensity of use, number of workspaces created, number and length of sessions, number of actions per session, etc. One area of our analysis focused on the proportional distribution of main actions. This revealed

that strong differences existed among different groups. To illustrate two examples, Figure 9 shows action distributions among the major high-level actions of the taxonomy of Figure 8 for one group whose distribution of actions was fairly even across categories (with the exception of the participant category): the five major action categories did not vary greatly, none of them exceeding 0.29 of the total (circle size signifies proportion out of the total).
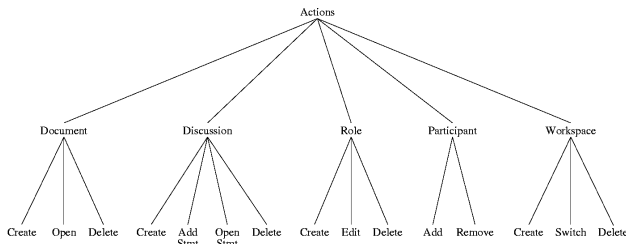

Figure 8. Taxonomy of major high-level LiveNet actions

Figure 10, on the other hand, shows a highly uneven distribution of actions in another group, where one action category (role) strongly dominates with 0.56 of the total, and two other action categories (document and discussion) barely register.
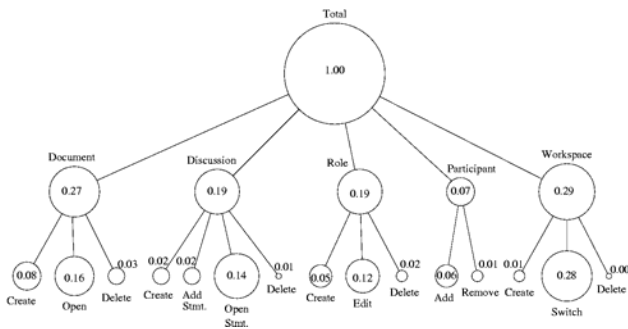

Figure 9. Relatively even distribution of actions in group 1
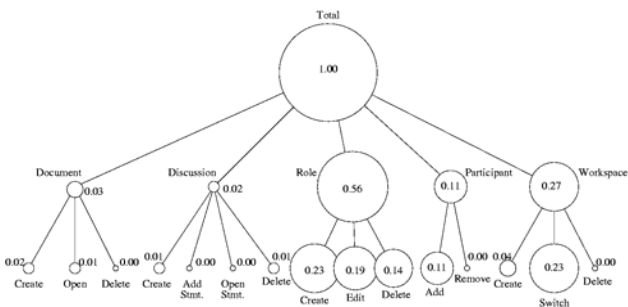

Figure 10. Highly uneven distribution of actions in group 50

This difference may be explained when considering that group 1 (Figure 9) had a total of 627 sessions consisting of a total of 7446 actions, while group 50 (Figure 10) had only 36 sessions and 633 actions. Not only did group 1 use LiveNet much more intensively, but they also made much greater use of the system to facilitate their own work (as manifested in the solid proportion of actions in the document and discussion categories). Thus the skew in action distribution towards role-related actions on the part of group 50 is caused by the under-utilization of other LiveNet features, not by an absolute high number of actions related to roles (in absolute terms, group 1 carried out 431 role-related actions, while group 50 carried out only 142 such actions). It should be noted that the choice of these two groups for illustration was not coincidental: group 1 was the best-performing group in the course, while group 50 was the worst-performing group, as measured in the marks obtained for their assignments in the course, one of which involved heavy use of LiveNet. The situation was comparable in other similarly scoring groups.

When such cases are identified and included in the organisational memory as part of a record of collaboration, they can be of use in evaluating virtual work. This can be particularly useful with fully virtual teams that never meet face-to-face, where conventional management methods for project monitoring and control are severely limited or absent. The organisational memory thus takes on the additional role of a management instrument.

The presented knowledge extraction related to media usage has already yielded interesting results through the application of relatively simple data analysis methods. As this research continues, we plan to more fully exercise a range of data mining methods on the available body of data to obtain further insights. For instance, clustering may assist us in identifying patterns of behaviour in different user groups, which can help categorize the usage of different collaboration spaces used for different purposes. It may also aid in the construction of "group profiles" which allows more personalized support to be provided.

4. INTEGRATING EXTRACTED KNOWLEDGE IN THE ORGANISATIONAL MEMORY

An important part of the framework is the way knowledge is returned back to the environment. An example of such feedback is the collection and generalisation over the workspace graph structure. The procedure to some extent is similar to building a case base of workspace configurations. Case indexing and retrieval is based on matching graph structures. The new collaborative process is formalised into a graph structure using concepts from a modified form of the soft systems methodology [11], with activities, roles and artefacts as node types and particular rules for the connections between the nodes (for example, a participant and an artefact cannot have a direct

connection). The formal representation is usually a result of a high-level (i.e. not detailed) description of the process. This representation is matched against the graph representation of the workspace configurations. Retrieved cases provide the initial configuration for further adaptation.

The framework also allows a feedback from the organisational memory towards modification of the knowledge representation schema, used for representation and incorporation of discovered knowledge. The detailed discussion of the issues related to the modification of the knowledge representation schema, however, are beyond the scope of this paper.

5. CONCLUSIONS

CVEs can provide researchers with enormous amounts of data about various aspects of computer-mediated collaboration. Unfortunately, the design of earlier environments did not pay much attention to the issues of data collection [9]. Thus, the application of multimedia data mining methods had to struggle with translating data collected for other purposes, for example, a server log used usually for correct recovery after a failure, into data useful for the goals of data mining. Consequently, the earlier application of multimedia data mining methods in CVEs has been focused mainly on the analysis of communication transcripts—whether recorded in synchronous collaborative sessions or over a bulletin board in asynchronous mode, and over project document content.

The framework presented in the paper looks at the integration of data mining technologies in CVEs at the early design stages of the virtual environment. A key issue at the design stage is the selection of the data that should be recorded. These records are complementary to the standard logs of the web server. They include activity log data, and the dynamics of media usage and changes in workspace content. Careful design and analysis of the activity log data have the potential to lead to improvements of the structure of the space and tuning the set of feasible actions with respect to the purpose of the environment. The applicability of the framework has been tested and demonstrated on a real environment. The new generation of environments has the potential to produce vast amounts of data about collaboration. The proposed combination of CVE and multimedia data mining technology will allow more coherent and consistent CVEs to be developed.

ACKNOWLEDGMENTS

REFERENCES

1. Ackerman, M.S. Augmenting the organizational memory: A field study of Answer Garden. In *Proceedings of the Conference on Computer Supported Cooperative Work*. (Chapel Hill, NC, USA, 22-26 Oct. 1994), 243-252.

2. Ansari, S., Kohavi, R., Mason, L., and Zheng, Z. Integrating e-commerce and data mining: Architecture and challenges. In *WEBKDD 2000 Workshop: Web Mining for E-Commerce – Challenges and Opportunities*. (Boston, MA, USA, 20 Aug. 2000).

3. Bannon, L.J., and Kuutti, K. Shifting perspectives on organizational memory: From storage to active remembering. In *Proceedings of the Twenty-Ninth Hawaii International Conference on System Sciences*. (Hawaii, USA, 3-6 Jan. 1996), vol. 3, 156-167.

4. Biuk-Aghai, R.P., and Hawryszkiewycz, I.T. Analysis of virtual workspaces. In Kambayashi, Y., and Takakura, H., Eds., *Proceedings of the 1999 International Symposium on Database Applications in Non-Traditional Environments (DANTE '99)*. (Kyoto, Japan, 28-30 Nov. 1999), 325-332.

5. Capin, T.K., Pandzic, I.S., Magnenat-Thalman, N., and Thalman, D. *Avatars in Networked Virtual Environments*. John Wiley and Sons, Chichester, 1999.

6. Conklin, E.J. Capturing organizational memory. In Baecker, R.M., Ed., *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*. Morgan Kaufmann Publishers, 1993, 561-565.

7. Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery: An overview. In Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., Eds., *Advances in Knowledge Discovery and Data Mining*. AAAI Press/MIT Press, Menlo Park, California, USA, 1996.

8. Furst, S., Blackburn, R., and Rosen, B. Virtual team effectiveness: A proposed research agenda. *Information Systems Journal 9*, 4 (1999), 249–269.

9. Greenhalgh, C. *Large Scale Collaborative Virtual Environments*. Springer-Verlag, London, UK, 1999.

10. Gutwin, C., and Greenberg, S. Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. (Seattle, WA, USA, 14-18 Nov. 1998), 207-216.

11. Hawryszkiewycz, I.T. Analysis for Cooperative Business Processes. In Zowghi, D., Ed., *Proceedings of the Fifth Australian Workshop on Requirements Engineering*. (Brisbane, Australia, 8-9 Dec. 2000), 3-11.

12. Hawryszkiewycz, I.T. Workspace networks for knowledge sharing. In Debrency, R., and Ellis, A., Eds., *Proceedings of AusWeb99, The Fifth Australian World Wide Web Conference*. (Ballina, Australia, 18-20 Apr. 1999), 219-227. (Available at: http://ausweb.scu.edu.au/aw99/papers/hawryszkiewycz /paper.html).

13. Lesley, H.G. and McKay, D.G. Towards an information and decision support system for the building industry. In Mathur, K.S., Betts, M.P. and Tham, K.W. (eds), *Management of Information Technology for Construction*, World Scientific, Singapore, 1993, pp. 101-111.

14. Maher, M.L., Simoff, S.J., and Cicognani, A. *Understanding Virtual Design Studios*, Springer-Verlag, London, UK, 2000.

15. Maher, M.L., Simoff, S.J., Gu, N., and Lau, K.H. Designing virtual architecture. In *Proceedings of CAADRIA2000*. (2000), 481-490. (Available at: http://www.arch.usyd.edu.au/~chris_a/MaherPubs/200 0pdf/caadria2000.pdf)

16. Patching, D. *Practical Soft Systems Analysis*, Pitman, London, 1990

17. Schafer, J.B., Konstan, J., and Riedl, J., Electronic Commerce Recommender Applications. *Journal of Data Mining and Knowledge Discovery*, 5 (1/2), 115-152.

18. Simoff, S.J., and Maher, M.L. Loosely integrated open virtual environments as places. *Learning Technology 3*, 1 (Jan. 2001). (Available at: http://lttf.ieee.org/ learn_tech/issues/january2001/ index.html#3).

19. Spiliopoulou, M., and Pohle, C. Data mining for measuring and improving the success of web sites. *Data Mining and Knowledge Discovery 5*, 1-2 (Jan. 2001), 85-114.