Data Mining in Collaborative Virtual Environments: An Integrating Framework

Simeon J. Simoff
University of Technology, Sydney, Australia
simeon@it.uts.edu.au

Robert P. Biuk-Aghai*
University of Macau, 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 communication and collaboration means, they have the potential to collect tremendous amounts of data about collaborative activities. The aim of this research is to utilise this data effectively, extract meaningful insights out of it and employ the new, semantically structured knowledge back in the environment. This paper presents a framework for integrating data mining and knowledge discovery technologies starting from the early design stage of the virtual environment. This KDD† framework includes four major groups of components, which begin with the selection of the data that should be recorded and end with a knowledge representation for incorporating dicovered information back into the system and improving the structure of the virtual space and the set of feasible actions that can be performed there. The paper contrasts the earlier ad hoc techniques to the proposed systematic approach. The applicability of the framework has been tested and demonstrated on a real environment.

---

* This research was conducted while the author was at the University of Technology, Sydney.

† "KDD" here stands for "Knowledge Discovery and Dissemination" – a term introduced by Lucio Soibelman of the Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign.

Introduction

Collaborative virtual environments have become increasingly popular in recent years. There are numerous approaches and techniques for arranging such environments for collaborative projects. The most common approach is to extend the desktop environment to include tools for meeting and sharing files. This approach takes the individual work environment and adds tools for communicating with others. Unfortunately, the toolbox approach handles only limited collaborative activity when usually people proceed effortlessly between different working styles, in terms of time, place and representations. This introduces a gap that interferes with collaborative activity. More formally, such gaps are defined as the physical or perceptual boundaries within the computer environment that either distract participants from the work they are doing, or that block them from crossing spatial, temporal or functional boundaries, inherent in collaborative work (Ishii *et al.*, 1995). Trying to bridge these gaps Gutwin and Greenberg (1998) recognised the importance of the ontology of "place" and virtual environments that follow such ontology. These environments range from simple desktop-style places to sophisticated virtual reality worlds (for an excellent taxonomy of the latter see Capin *et al.*, 1999). Despite their variety and difference in functionality these environments have several key concepts in common:

- the concept of "*inserting people*" in the networked environment, in other words, representing them as some entities. These representations span from the so-called "characters" in text based MOO/MUD and Web-based WOO environments (Maher *et al.*, 2000b) to the "avatars" in the 3D virtual worlds.
- the concept of "*structuring the space*" in the networked environment, in other words, providing some way of structuring the place, separating and handling different information within the units of this structure, and some reference system for orientation and navigation. These structures span from the "room" approach in MOO/MUDs, to the "squares of land" and "worlds" in ActiveWorlds universes;
- the concept of "*a feasible set of actions*" that can be performed in the networked environment. This set defines to what extent the environment under consideration can be used for conducting collaborative projects in a particular domain.

People

The establishment of the identity of the people in the virtual place occurs through the representation of individuals as avatars or objects that possess various properties. Object representations of a person include characteristics such as a verbal 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 representations is the variety of "rights" that can be assigned to them. Different environments use different terms for this – privileges, roles, permissions – which reflects the differences in the context in which people are represented. The context is formed by what other participants see about the room and how they see each other in that room. These representations are potential sources of preliminary information about a person's individuality.

Space structuring

The ways of structuring of the space of the environment depend on a number of factors, including the ontology (what kind of place the environment is), purpose of the environment, the embedded functionality, the preferable communication and collaboration mode (Maher *et al.*, 2000a), underlying technologies and their integration (Simoff and Maher, 2001). We illustrate the idea of different structuring of space in the example of a conference venue* (Maher *et al.*, 2000b). The section of the conference venue presented in Figure 1 is located in the MOO part of the Virtual Campus†, which is organised according to the ontology of an university campus – the space is structured in terms of "rooms" and "buildings" (groups of rooms). The reference system and the topology of the space are based on the purpose of the "buildings" and the "rooms", which constitutes a well-defined taxonomy, with the origin of the reference system at the top of the taxonomy. The section of the conference venue shown in Figure 2 is located in the Virtual Design Studio which is a part of the Virtual Campus‡, supported by ACTIVEWORLDS technology§ which is organised according to the ontology of a physical world where 3D active forms can be built for pre-fabricated building blocks. The reference system is based on a geographical coordinate system, with the origin in the "middle" of the space.

---

* Designed at the University of Sydney's Key Centre of Design Computing and Cognition specifically for the annual on-line conference on Design Computing on the Net. For more information see http://www.arch.usyd.edu.au/kcdc/conferences/DCNet00/index.html

† The Virtual Campus, Faculty of Architecture, University of Sydney, http://www.arch.usyd.edu.au:7778/

‡ In terms of underlying technologies, the Virtual Campus evolved from a MOO-based virtual environment to a loosely integrated collection of environments (Simoff and Maher, 2001).

§ http://www.activeworlds.com

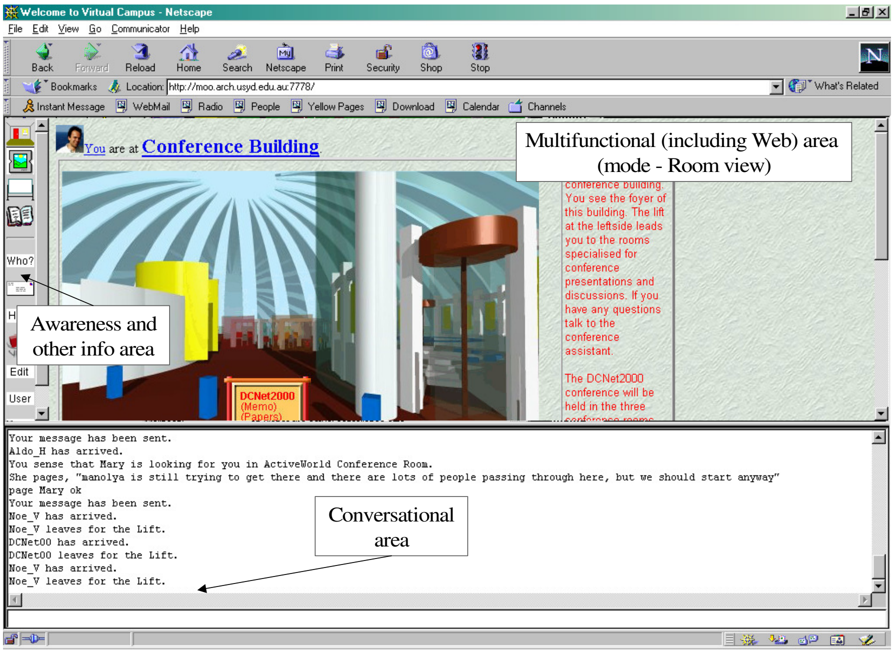


Figure 1. A "Conference building" in the Virtual Campus

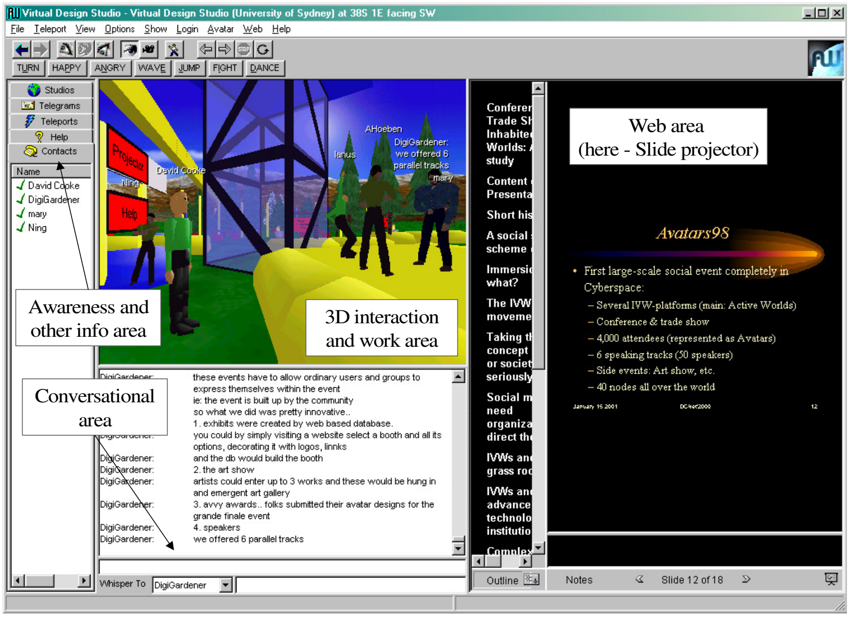


Figure 2. The Conference area in the Virtual Design Studio (ACTIVEWORLDS being the underlying technology), accessible via the Conference building in the Virtual Campus

Feasible actions

The ontology of the virtual environment can provide substantial *a priori* knowledge about the navigation and the feasible set of actions in such an environment. Furthermore, the variety of actions can provide *a priori* knowledge for the analysis and discovery of patterns of collaboration.

The design of collaborative virtual environments (CVE) remains a craftwork. In this paper we propose to combine CVE and data mining technology to develop more coherent and consistent environments.

Data mining and knowledge discovery in collaborative virtual environments

Collaborative virtual environments have the potential to provide professional working environments that can support collaborative projects in different disciplines independent of geography. Consequently, they 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 (Greenhalgh, 1999). Thus, the application of 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 data mining methods in collaborative virtual environments has been focussed mainly on the analysis of communication transcripts – whether recorded in synchronous collaborative sessions or over a bulletin board in asynchronous mode. We illustrate both cases below.

Previous ad hoc approaches

A typical scenario from a participatory design session in the Virtual Design Studio is shown in Figure 3. Such an environment can provide rich multimedia, data including data about the evolving geometry of the design and transcripts of the corresponding discussions of the ideas on each step; data about the allocation and behaviour of participants; web content of project documentation; even audio and video records. However, the transcripts of the “conversations” (the chat logs) during the collaborative sessions are the only data that can be easily recorded. A methodology for pre-processing and analysing such transcripts and for deriving measures for estimating participation in collaborative sessions was presented by (Simoff, 1999, Simoff and Maher, 2000).

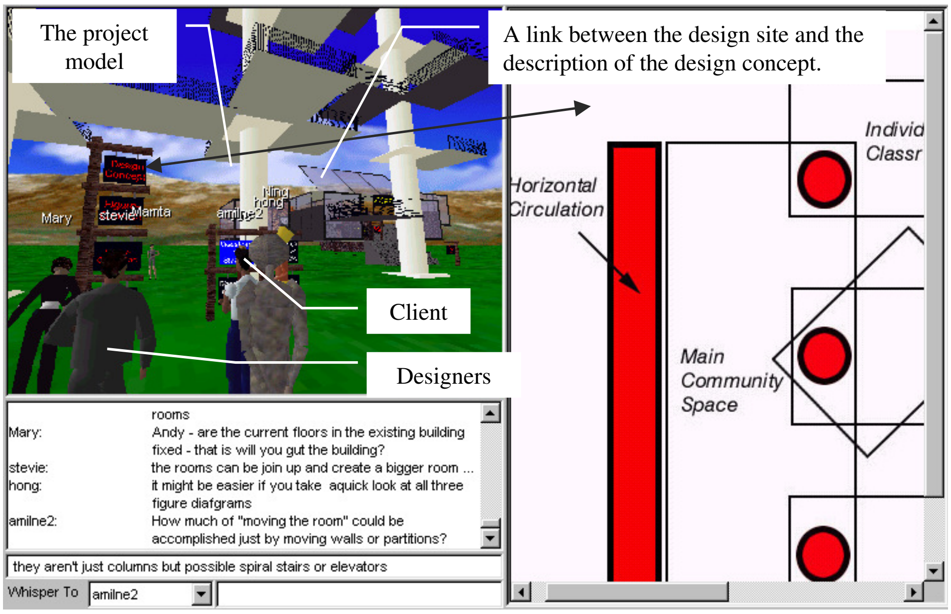


Figure 3. An example of a collaborative project in the virtual design studio

This approach has been extended to incorporate the on-line analysis of bulletin board transcripts of asynchronous collaboration. Personal contributions to a collaborative session can be evaluated using text analysis of seminar transcripts (Simoff and Maher, 2000) and multimedia analysis of related web pages.

Figure 4 presents a fragment from a team bulletin board. The messages on the board are grouped in threads. A threefold split of the thread structure of e-mail messages in discussion archives in order to explore the interactive threads was proposed in (Berthold *et al.*, 1997, Berthold *et al.*, 1998). It included (i) reference-depth: how many references were found in a sequence before this message; (ii) reference-width: how many references were found, which referred to this message; and (iii) reference-height: how many references were found in a sequence after this message. The threefold split was extended in (Sudweeks and Simoff, 2000) to include the time variable explicitly. This model, expressed graphically as a tree, allows the comparison of the structure of discussion threads both in a static mode (for example, their length and width at corresponding levels) and in a dynamic mode (for example, detecting moments of time when one thread dominates another in multi-threaded discussions).

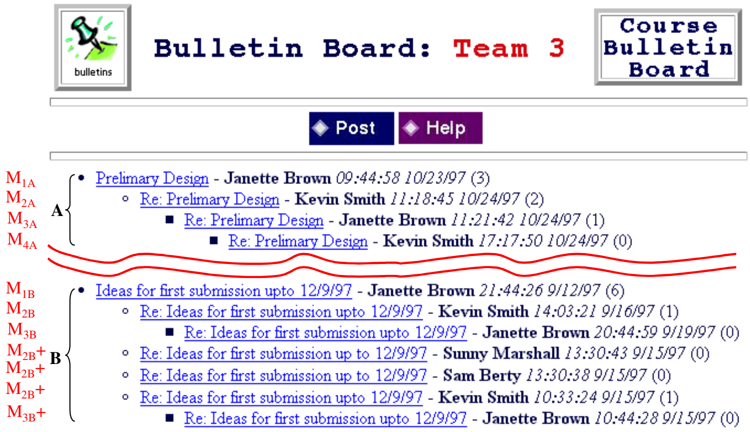
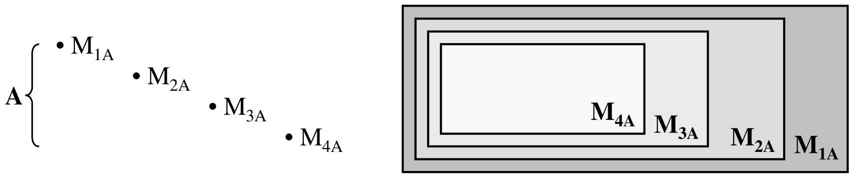


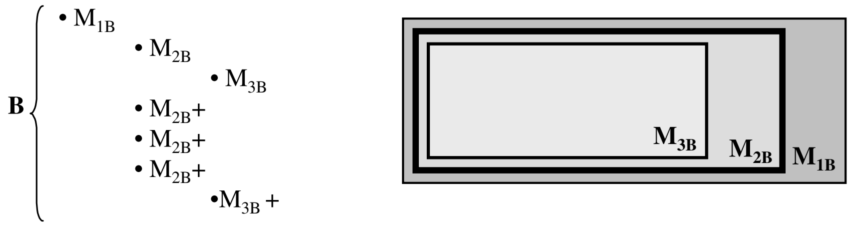
Figure 4. Fragments from an asynchronous communication on the bulletin board in a virtual world

Based on this model, on-line visualisation techniques have been developed, which are modified versions of the nested set visualisation of tree structures (Knuth, 1973, pp. 311-312). Figure 5 shows an example of such visualisation applied to threads “A” and “B” from Figure 4.

Each first message in a level is represented by a corresponding rectangle, labeled in this example to illustrate the message correspondence. Thus, there are four nested rectangles in Figure 5a. When messages are at the same level, the thickness of the line is estimated based on the content-analysis of the message, including the text, graphics and images. As a reasonable approximation, each of the relevant messages on the same level can be represented as additional 0.5 pt to the base line thickness. In Figure 5b the base line thickness is 1 pt, thus rectangle “M2B” has thickness 2.5 pt.



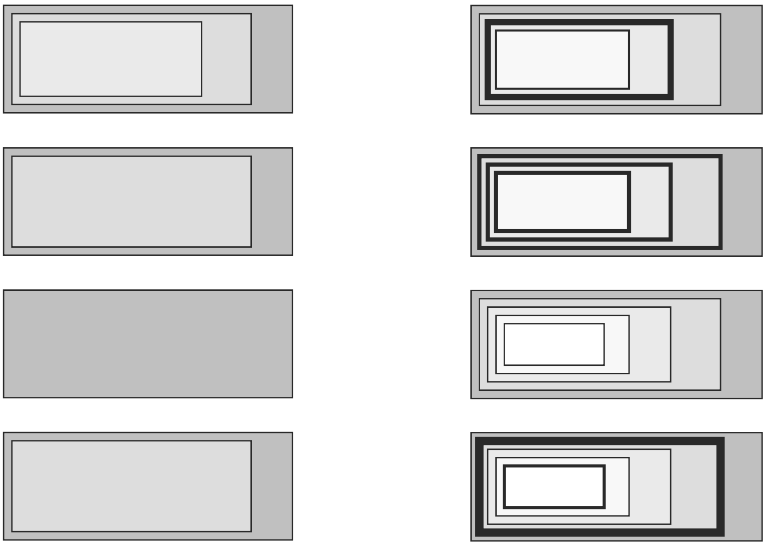
a. Nested rectangles for single message per level.



b. Nested rectangles when there are multiple messages on some levels.

Figure 5. Visualisation of discussion threads

Figure 6 illustrates the application of the technique for monitoring collaborative teams. Collaboration can be considered at different levels of task sharing. Two extreme approaches to sharing tasks during collaboration are identified in (Maher *et al.*, 1997): single task collaboration and multiple task collaboration. During single task collaboration, the product is a result of a continued attempt to construct and maintain a shared conception of the task. In other words, each of the participants has his/her own view over the whole problem and the shared conception is developed during intensive discussions. The basic assumption is that collaboration style influences the communication pattern. An example, of the visual pattern of such type of collaboration is presented in Figure 6b. It is characterised with relatively large amounts of nested rectangles, usually indicating also several messages in response to a particular message. During multiple task collaboration, the problem is divided among the participants so that each person is responsible for a particular portion of the product. Thus, multiple task collaboration does not necessarily require the creation of a single shared conception, thus messages are usually related to the project management. Isolated messages and short threads dominate this collaboration style, as illustrated in Figure 6a.



a. Collaboration without creating a shared understanding of the problem.

b. Intensive collaboration for creating a shared understanding of the problem

Figure 6. Patterns of collaboration

Apart from identifying participation and collaboration patterns, it has been difficult (if not impossible) to extract and analyse data that can provide insights about structuring the environment and the feasible set of actions. Therefore, we further propose a framework for the design of collaborative virtual environments that support mining of data about collaboration recorded by the environments. This framework differs from the approach of Chen (1999), who uses graphical capabilities of the virtual environment to support the visual exploration of external data within the environment itself.

Framework for embedding knowledge discovery in collaborative virtual environments

The framework is oriented towards designing environments that provide the option to collect data for the purpose of its mining and analysis. As a result we have the opportunity to control data collection to a larger extent. There are a number of research efforts in the direction of controlled data collection, carried out mainly in the field of e-commerce and Web data mining (see Spiliopoulou and Pohle, 2001).

The framework is shown in Figure 7. It includes four major groups of components:

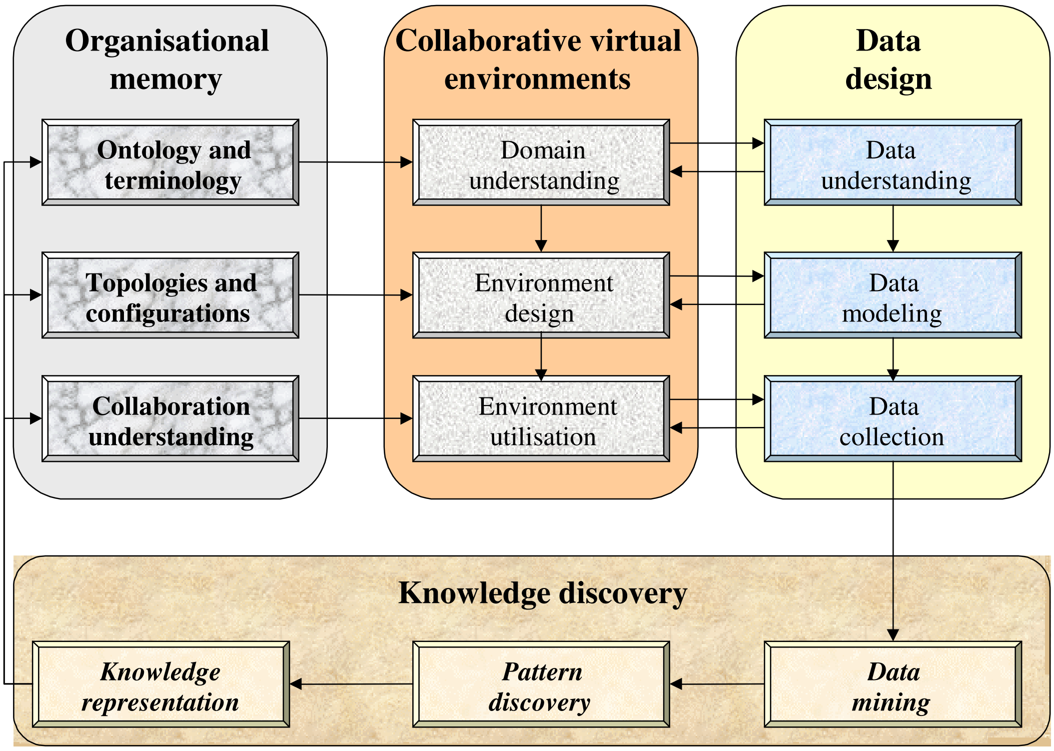


Figure 7. Framework for integrating data mining in the design, application and research in collaborative virtual environments.

- "Collaborative virtual environments" is the label for the support related to understanding the domain for which we design the virtual environment, the actual environment design, and its deployment.
- "Data design" groups the issues related to the data manipulation during each of the stages described above. It includes understanding of the nature of the domain data, building a model of the domain that can be expressed in terms of collected data, and the actual data collection.
- "Knowledge discovery" in our case differs slightly from the classical schema (see Fayyad *et al.*, 1996) – the selection and data pre-processing stages are implicitly embedded in the data design, in other words, collected data is expected to be ready for data mining. Another difference is the inclusion of the knowledge representation – among the goals of the data mining is the better understanding of computer-mediated collaboration and the usage of discovered knowledge to improve the structure of the environment. For example, we can provide a template structure of a virtual place, which implies some navigation behaviour. Collecting data about the navigation within the place can provide a source for discovering traversal patterns, which can provide indicators for improving the topology (structuring) of the virtual place. Another possibility for improvement is according to particular collaboration needs. This is something difficult to

know ahead of time. In both cases we need some measure to obtain the necessary indicators for improvement of the structure.

- "Organisational memory": Over the past decade, the CSCW community and related areas have taken a keen interest in organizational memory (OM for short) (Ackerman, 1994; Ackerman and Halverson, 2000; Bannon and Kuutti, 1996; Conklin, 1993). This suggests that there is value in retaining and later drawing on historical records of virtual collaboration. Such records could be referenced when setting out on new virtual collaboration, to "see how others have done it", and perhaps to reuse and re-enact those collaboration instances. Unlike conventional work settings where details of collaboration have to be collected manually through effort-intensive and sometimes intrusive methods, a collaborative virtual environment is an ideal source of data on collaboration, particularly when work is predominantly or entirely carried out virtually, as such an environment can automatically record a great amount of detail on the collaboration.

Application of the framework

To illustrate the ideas presented above, we show how they were applied to a certain groupware system, LIVENET, whose design is oriented towards data collection. We first introduce the system, then show how knowledge about collaboration was extracted from it.

LIVENET

The LIVENET system is a virtual collaboration system prototype developed at the University of Technology, Sydney (Hawryszkiewycz, 1999). 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 8). LIVENET provides virtual *workspaces* which bring together people, artefacts (e.g. documents), communication channels, awareness facilities, and a collection of tools, all tied together through a configurable governance structure. The relationship of LIVENET's main conceptual elements is represented in the meta-model in Figure 9 (Biuk-Aghai, 2000). In terms of the meta-model, workspaces contain roles, occupied by participants (i.e. actual people), who carry out actions. Some actions may operate on document artefacts, others may be interactions with other workspace participants through discussions. Most workspace elements such as documents, discussions and participants, may be shared between workspaces. Thus workspaces are not just stand-alone entities but nodes in a network of inter-connected collaboration spaces. Neither are structures of workspaces in LIVENET static—once created, a workspace can be dynamically adapted to evolve together with the collaboration carried out in it, while likewise entire "ecologies" of inter-connected workspaces can co-evolve.

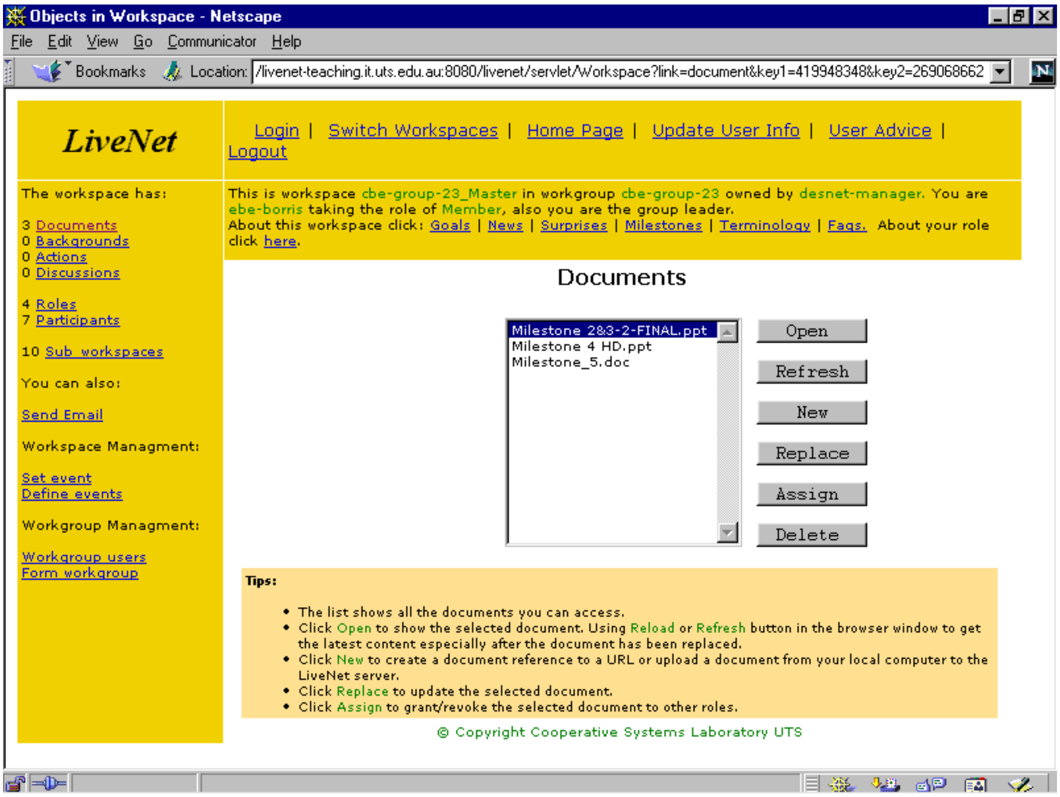


Figure 8. Typical LIVENET screen (web interface)

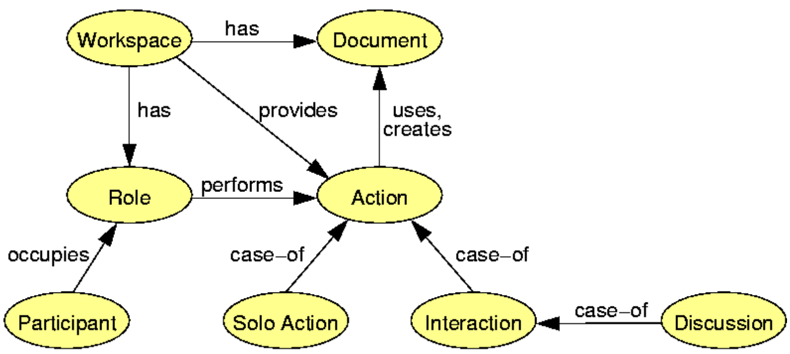


Figure 9. Meta-model of main conceptual elements in LIVENET

Knowledge discovery in LIVENET

Data about workspaces in LIVENET captures two aspects: a database maintains the current state of all workspace elements (documents, roles, participants, etc.), while log files 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 LIVENET'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 (Ansari *et al.*, 2000).

The analysis we carried out focused primarily on the usage log, 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.

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 their 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.

Space structuring

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, organized 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

* We used the Weka data mining workbench available at: http://www.cs.waikato.ac.nz/~ml/weka/

analysis revealed two distinct patterns: the majority of users preferred to use just one workspace to organize 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* (Biuk-Aghai and Hawryszkiewycz, 1999). We term such groups *centralizers*. To a certain extent, this mode corresponds to the single-task collaboration mentioned earlier. 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*. Their collaboration style corresponds to the multi-task collaboration.

Figure 10 shows a map of LIVENET workspaces with colours highlighting absolute workspace density—green meaning low density, red meaning high 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.

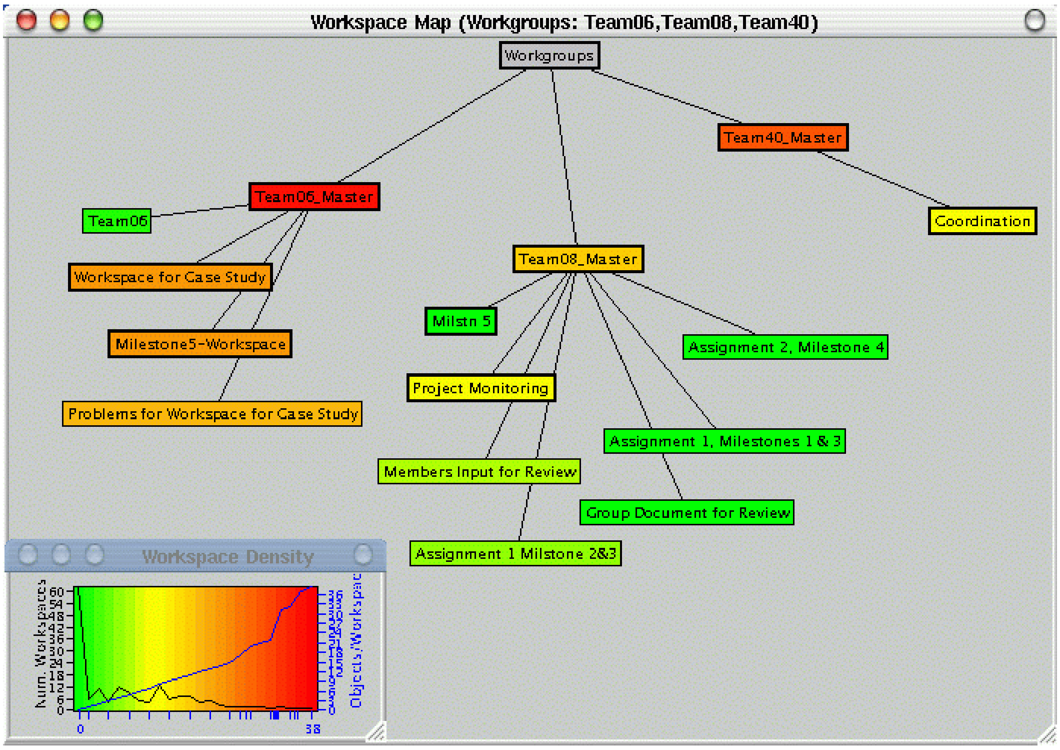


Figure 10. Workspace densities of three different groups

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 in the case of multi-task collaboration. Furthermore, some groups may bring certain preferences as to the way to organize their work into workspaces and enact these preferences in the way they structure their virtual working environment. To recognize such preferences, using KDD methods, and to feed them back into the setup of virtual collaboration environments could thus help offer more adequate support to cooperative groups with diverse working styles.

Feasible actions

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 modeling elements: workspaces, roles, participants, documents, and discussions. A taxonomy of these actions is presented in Figure 11.

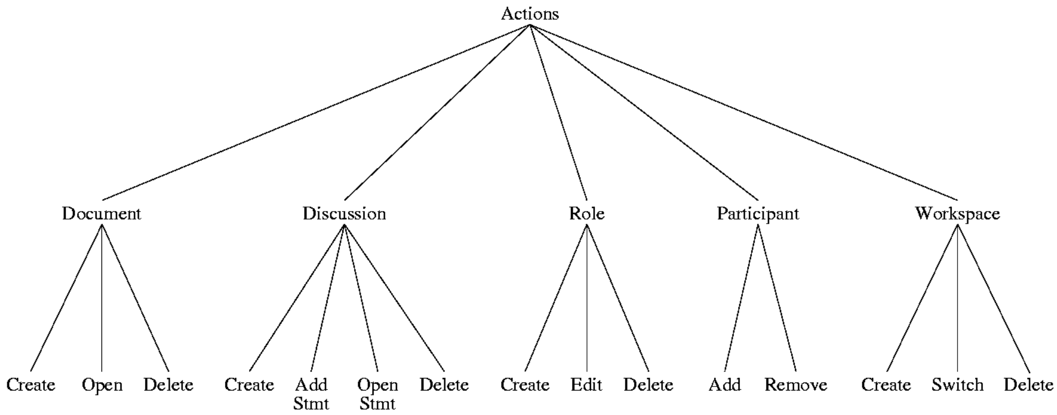


Figure 11. Taxonomy of major high-level LIVENET actions

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 12 shows action distributions among the major high-level actions of the taxonomy of Figure 11 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).

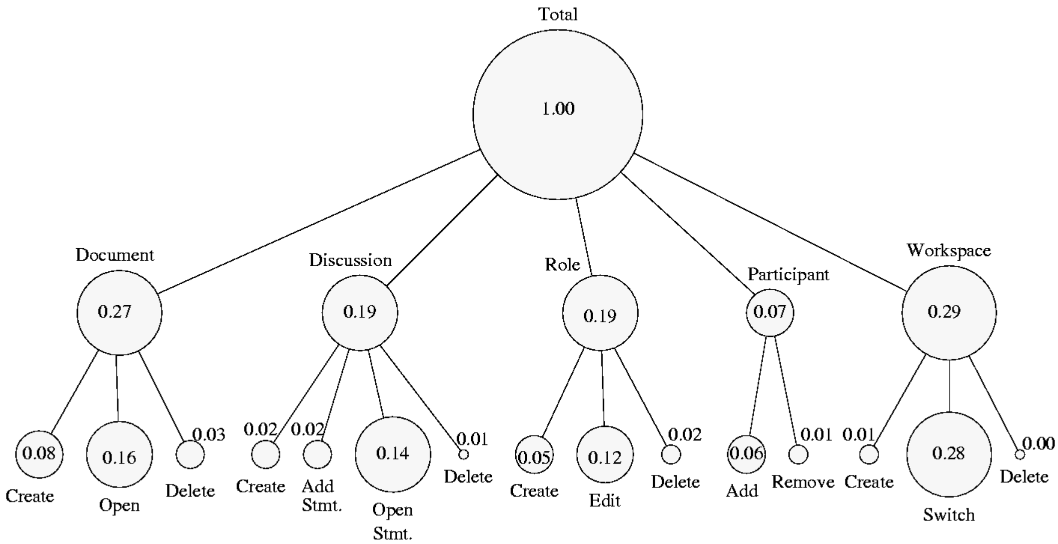


Figure 12. Relatively even distribution of actions in group 1

Figure 13, 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.

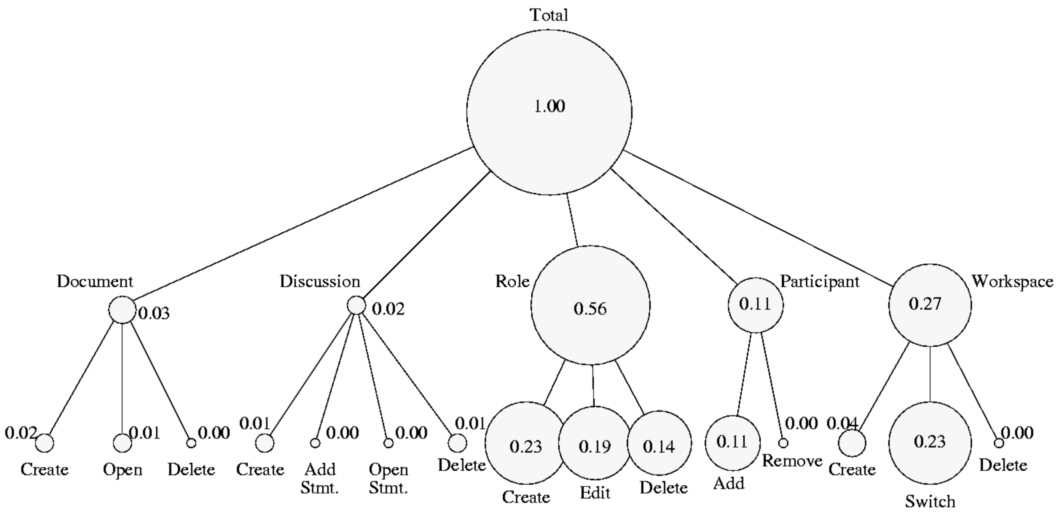


Figure 13. Highly uneven distribution of actions in group 50

This difference may be explained when considering that group 1 (Figure 12) had a total of 627 sessions consisting of a total of 7446 actions, while group 50 (Figure 13) 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 is caused by the under-utilization of other features of LiveNet, 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.

Identification of such cases can be of use in evaluating virtual work. This can be of particular use with fully virtual teams that never meet face-to-face, where conventional management methods for project monitoring and control are severely limited or absent.

Conclusions

Collaborative virtual environments have the potential to change the way we work. Unfortunately, earlier observations of human activities uncovered very few aspects of computer-mediated collaboration. The new generation of environments has the potential to produce tremendous amount of data about collaboration. The development of data mining technologies offers a complementary instrumentation capable of extracting semantic information and turning the collected data into invaluable asset. The integration of such environments with data mining technologies provides unique opportunities to unveil some secrets in the art of human collaboration.

The framework presented in the paper looks at the integration of data mining technologies in collaborative virtual environments 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 complimentary to the standard logs of the web server. Careful design and anlysis of this logs 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 paper contrasts the earlier ad hoc approaches to collaboration analysis to a systematically designed approach, demonstrating the potential of proposed integration.

An important part of the framework is the way knowledge is returned back to the environment. The framework allows also 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.

Acknowledgments

This research has been supported by the Australian Research Council and University of Technology, Sydney. The support of the University of Macau to Mr. Biuk-Aghai during his stay at the University of Technology, Sydney is gratefully acknowledged.

References

Ackerman, M.S. (1994): "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, pp. 243-252, ACM Press.

Ackerman, M.S. and Halverson, C.A. (2000): "Reexamining Organizational Memory", *Communications of the ACM*, 43(1), Jan. 2000, pp. 58-64.

Ansari, S., Kohavi, R., Mason, L. and Zheng, Z. (2000): "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. (Available at: http://robotics.stanford.edu/users/ronnyk/WEBKDD2000/papers/ronny.pdf).

Bannon, L.J. and Kuutti, K. (1996): "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, pp. 156-167, IEEE Computer Society Press.

Berthold, M.R., Sudweeks, F., Newton, S. and Coyne, R. (1997): "Clustering on the Net: Applying an Autoassociative Neural Network to Computer-Mediated Discussions", *Journal of Computer Mediated Communication*, 2(4). (Available at: http://www.ascusc.org/jcmc/vol2/issue4/bert-hold.html).

Berthold, M. R., Sudweeks, F., Newton, S. and Coyne, R. (1998): "It makes sense: Using an autoassociative neural network to explore typicality in computer mediated discussions", in Sudweeks, F., McLaughlin, M. and Rafaeli, S. (eds.) *Network and Netplay: Virtual Groups on the Internet*, AAAI/MIT Press, Menlo Park, CA, pp. 191-220. (Available at: http://www.ascusc.org/jcmc/vol2/issue4/bert-hold.html).

Biuk-Aghai, R.P. (2000): "Virtual Workspaces for Web-Based Emergent Processes", in: *Fourth Pacific Asia Conference on Information Systems: Electronic Commerce and Web-Based Information Systems*, Hong Kong, China, 1-3 Jun. 2000, pp. 864-880.

Biuk-Aghai, R.P. and Hawryszkiewycz, I.T. (1999): "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, pp. 325-332, IEEE Computer Society.

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

Chen, C. (1999): *Information Visualisation and Virtual Environments*, Springer-Verlag, London, UK.

Conklin, E.J. (1993): “Capturing Organizational Memory”, in: Baecker, R.M. (ed.) *Readings in Groupware and Computer-Supported Cooperative Work: Assisting Human-Human Collaboration*, pp. 561-565, Morgan Kaufmann Publishers.

Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth, P. (1996): “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.

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

Gutwin, C. and Greenberg, S. (1998): “Design for Individuals, Design for Groups: Tradeoffs Between Power and Workspace Awareness”, in: *Proceedings of the ACM Conference on Computer Supported Cooperative Work,* 14-18 Nov. 1998, Seattle, WA, USA, pp. 207-216, ACM Press.

Hawryszkiewycz, I.T. (1999): “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, pp. 219-227. (Available at: http://ausweb.scu.edu.au/aw99/papers/hawryszkiewycz/paper.html).

Ishii, H., Kobayashi, M. and Grudin, J. (1995): “Integration of Interpersonal Space and Shared Workspace: Clearboard Design and Experiments”, in: Greenberg, S., Hayne, S. and Rada, R. (eds.), *Groupware for Real-Time Drawing: A Designer's Guide*, McGraw-Hill Book Company, Berkshire, England, pp. 96-125.

Knuth, D. E. (1973): *The Art of Computer Programming, Vol 1: Fundamental Algorithms*, Addison-Wesley, Reading, MA, USA.

Maher, M.L., Simoff, S.J. and Cicognani, A. (1997). Potentials and Limitations of Virtual Design Studio, Interactive Construction On-line, Jan. 1997, a1.

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

Maher, M.L., Simoff, S.J., Gu, N. and Lau, K.H. (2000b): “Designing Virtual Architecture”, in: *Proceedings of CAADRIA2000*, pp. 481-490. (Available at: http://www.arch.usyd.edu.au/~chris_a/MaherPubs/2000pdf/caadria2000.pdf)

Simoff, S.J. (1999): “Monitoring and Evaluation in Collaborative Learning Environments”, in: *Computer Supported Collaborative Learning (CSCL99)*, Stanford, CA. (Available at: http://kn.cilt.org/cscl99/A83/A83.html).

Simoff, S.J. and Maher, M.L. (2000): “Analysing Participation in Collaborative Design Environments”, *Design studies*, 21, 119-144.

Simoff, S.J. and Maher, M.L. (2001): “Loosely Integrated Open Virtual Environments as Places”, *Learning Technology*, 3(1), (Available at: http://lttf.ieee.org/learn_tech/issues/january2001/index.html#3).

Spiliopoulou, M. and Pohle, C. (2001): “Data Mining to Measure and Improve the Success of Web Sites”, *Data Mining and Knowledge Discovery*, (to appear).

Sudweeks, F. and Simoff, S.J. (2000): "Complementary Explorative Data Analysis: The Reconciliation of Quantitative and Qualitative Principles", in: Jones, S. (ed.) *Doing Internet Research*, pp. 29-55, Sage Publications.