

Analysing Participation in Collaborative Design Environments

SIMEON J. SIMOFF AND MARY LOU MAHER

*Key Centre of Design Computing
University of Sydney, NSW-2006
{simeon, mary}@arch.usyd.edu.au*

Abstract

Computer-supported collaborative design can be realised by a broad range of collaborative environments, each facilitating a different kind of collaboration. Understanding the style of collaboration and the potential for each environment is important when choosing a particular technology. We have developed a virtual world approach to teaching design computing in which students learn through traditional lectures, online seminars, and collaborative design projects. The environment integrates both synchronous and asynchronous communication as well as shared documentation. One side effect of using this environment is the incremental development of a record of the communication and collaboration. This record can be the basis for the analysis of participation in collaboration. We show how text analysis as a part of data mining can be used to analyse different aspects of participation. Specifically, we analyse participation in synchronous communication to evaluate individual contribution. We then analyse asynchronous communication to evaluate the extent of collaboration. The methods presented can be an automated part of the collaborative environment providing information for student evaluation in an educational environment or individual contribution in a professional environment.

Keywords: collaborative design environments, data mining, text analysis, Web environments, evaluating collaboration

Computer-supported collaborative design education

The proliferation of computer media and networking has the potential to make fundamental changes in the methods, models and techniques employed to educate and train design students and professionals. The easy access to the internet requires a reconsideration of the desktop CAI/CAL¹ paradigms. Analyzing the role and impact of computers in the on-going changes in education, Lockard *et al.* (1994) specify two types of CAI/CAL applications. Applications classified as Type I employ computing resources to do things educators have previously done without computers. These applications improve such aspects of teaching like preparation of teaching materials and student management. Generally, this approach does not change the teaching strategies and schemata, thus Type I automation reduces the technical efforts but does not result in more effective teaching. The other group of applications, labeled as Type II are oriented towards bringing teaching and learning methods and experiences impossible without computers. We explore this type of environment and its use in design education and the profession.

Most of the early Web-mediated on-line environments deliver various multimedia learning materials and give students access to an enormous resource of information. However, the on-line materials alone are not sufficient for evaluation and assessment since the instructors cannot trace student participation, in the sense of interaction with course materials, other students, and tutors. The absence of an overall learning model, combined with restricted interaction and single direction of the information flow, show that these initial efforts were able to utilise only small part from the potential of internet technologies. On-line learning methods can facilitate three components of course administering - (i) *course materials* to be delivered to the students; (ii) *communication* between students and educators, and (iii) *management and*

¹ CAI/CAL - Computer Aided Instruction and Computer Aided Learning

assessment of students. In this paper we focus on automating the analysis of communication and collaboration, providing a methodology for analysing participation in any collaborative design environment.

Design, as an educational subject, is characterised by the lack of clear separation between theoretical knowledge and practical skills. In addition, design projects require intensive collaboration of numerous specialists. The instructional strategy in design studios is based on constructivist principles in which a student actively constructs an internal representation of knowledge by interacting with the material to be learned. The model implements the principles of situated cognition (Streibel, 1991) and problem-based learning (Savery and Duffy, 1995). According to these theories, both social and physical interaction enter into the definition of a problem and the construction of its solution. Neither the information to be learned, nor its symbolic description, is specified outside the process of inquiry and the conclusions that emerge from that process.

Mitchell and McCullough (1995, pp.441-462) present the general idea of the virtual design studio as an extension of conventional studios, discussing different enabling technologies and corresponding studio scenarios. Maher *et al.* (1997b) developed these ideas into a Web-based Virtual Design Studio (VDS) as a distributed environment for teaching students about collaborative design. The modification of the workspace in the VDS and the replacement of the physical space by a computer mediated educational environment enhances the studio-style learning by releasing geographical and time scheduling constraints.

Our experience in developing learning materials for virtual design studios (Simoff and Maher, 1997a and 1997b) show that the design and management of Web-mediated courses involves more than converting lectures and exercise notes to a collection of linked web pages or providing tools for communication. We identify two issues in developing effective collaborative design environments:

- the need to create a sense of place and community among the participants in the studio and
- the need for interactive and/or automatic assessment.

We address these needs in the development of our Virtual Campus [HREF1]. This is an educational environment which uses the metaphor of a physical campus as collection of rooms to organise the learning and community activities. Rooms in the virtual campus are grouped according to their type of use, such as classroom, office, or conference room. The rooms provide some of the functions of a physical room, such as privacy, security, and a place for various activities. The classrooms are equipped with a slide projector, a recorder, whiteboard, notice board and other facilities that help the student with a particular learning task. The basic communication mode is text-based, though the campus has Web-based interface for navigation through a 3D visualisation of the facilities (using VRML).

The interface, shown in Figure 1, provides facilities for viewing the part of the campus where the person is located, various tools for using the campus and a line input for communication, navigation, and creating objects. Figure 1 shows the practice studio for preparing a presentation. The room is accessible from the conference room. The top part of the window shows the room description, with a tool bar for looking at the whiteboard, the slide projector, the course materials (created using WebCT as described below), and other tools for using the campus. The bottom part of the window shows the text description of the room and responds to input from the user. The input can vary from commands to show who is in the room to communication commands such as “say”, “whisper”, and “emote”. This environment provides real time communication similar to a chat window, with the additional feature of providing a permanent place to store information and other objects. This environment effectively addresses the first need – to create a sense of place and community.

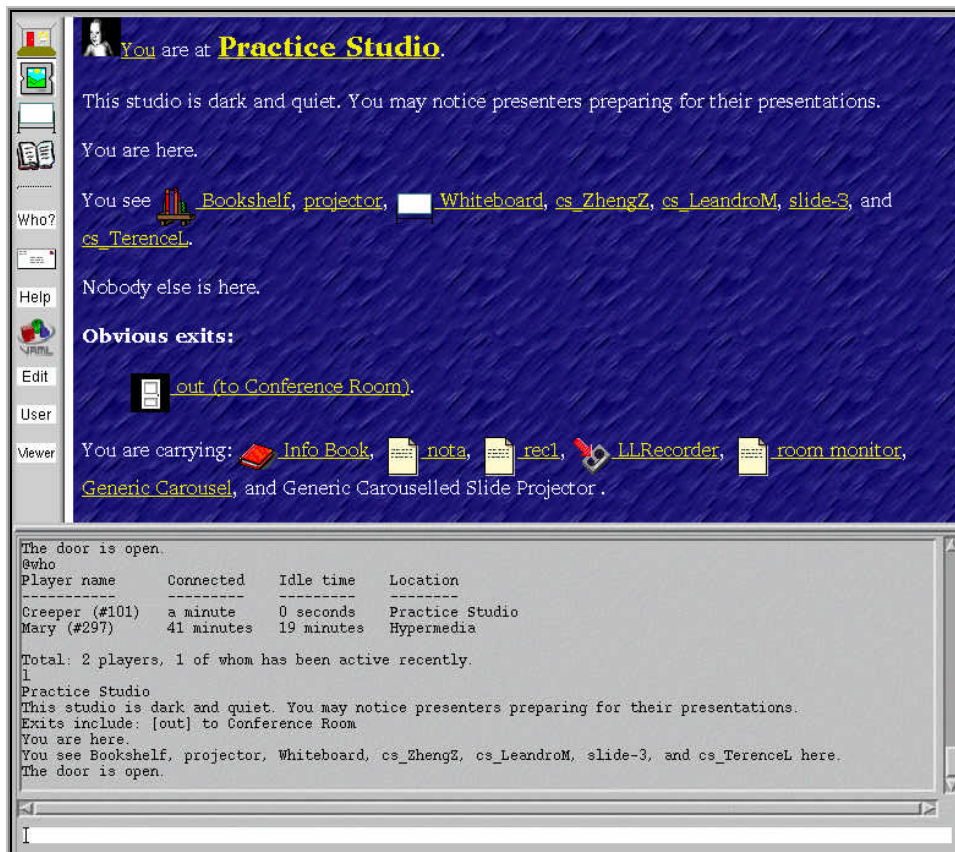


Figure 1. The Web interface to the Virtual Campus.

The second issue is the need to provide for interactive and automated assessment. Since the students are not in a physical classroom, in-class tests do not make sense. There are several difficulties with implementing interactive assessment within Web-mediated course delivery mechanism in comparison to stand-alone CAI/CAL programs. The main disadvantage comes from the batch mode of information exchange. This makes the implementation of genuine, immediate and meaningful feedback problematic. Typically, a form is filled out then sent to the server which processes it and sends back another form or web page. Sutherland (1997) presents new architectures of enhanced Java/CGI implementations and Applet/Servlet remote method invocations, which have the potential to overcome this problem of the client/server model. These innovations allow greater interactivity and permit more complex client operations. However, this is at the technological front end and, consequently, there is a lack of tools to simplify the implementation.

The response to this need is a new generation of Web-authoring environments, which provide a set of tools for student communication and tools for course designers to create and structure courses. There are numerous "working" pioneers, among which WebCT [HREF2], developed at the University of British Columbia, has been successfully employed for design course management and delivery at the University of Sydney. Figure 2 illustrates WebCT in course design mode. These new environments include on-line quizzes, student participation and progress tracking facilities and enhanced course management facilities. These features are mainly based on the analysis and bar-graph visualisation of server statistics. They do not analyse and assess student collaboration - an essential part of project-based design courses.



Figure 2. WebCT: designing a Web-based course on Design Computing

The combination of the WebCT course design environment with the Virtual Campus provides a comprehensive environment for collaborative learning. The combined environment is illustrated in Figure 3. The WebCT course server includes a bulletin board system with archiving facility and the campus classrooms provide a place for meetings, presentations, and seminars. Project-based courses, conducted in this environment, use a variety of synchronous and asynchronous computer-mediated teaching techniques. Transcripts of on-line subject discussions, conducted in the Virtual Campus and messages on bulletin boards a rich source of data. We show how text analysis and data mining methods can be used for the analysis and evaluation of student participation and collaboration.

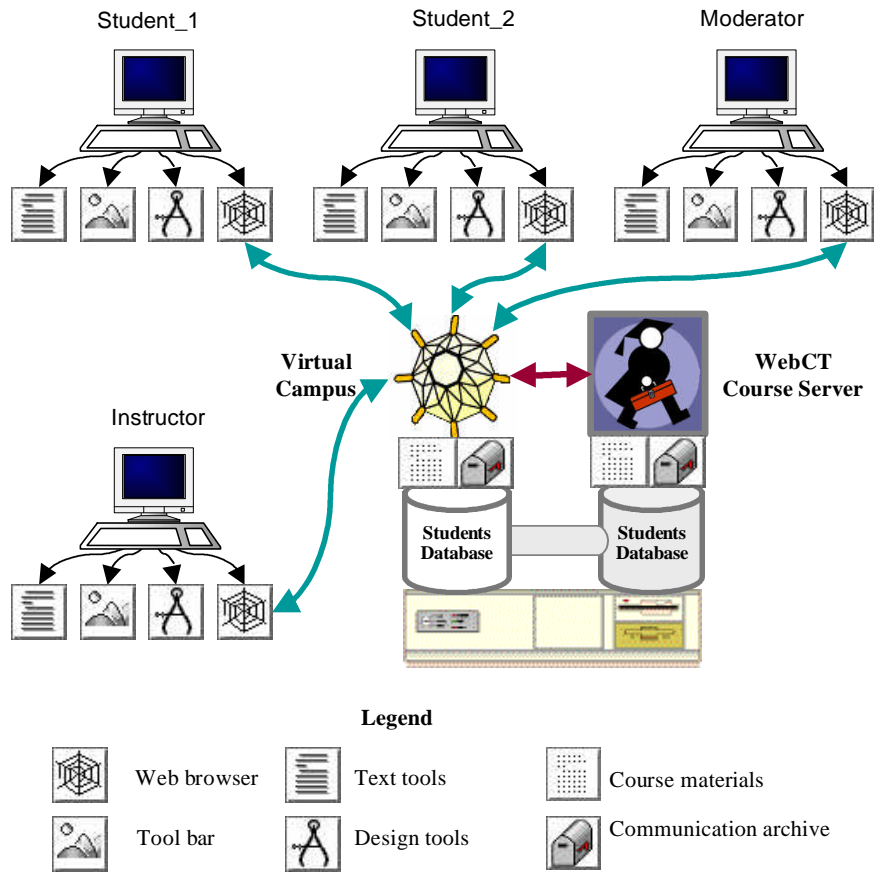


Figure 3. Integrated Web-based design educational environment.

Evaluating individual contribution using text analysis and data mining

Data mining (DM), known also as knowledge discovery (KD), is the overall process of examining a data source for implicit information and recording this information in explicit form, in other words, the extraction of a high-level knowledge from a low-level data. The overall process of data mining is shown in Figure 4. DM involves the identification of potentially useful and understandable patterns in this data (Fayyad *et al.*, 1996, Chapter 1), spanning the entire spectrum from discovering information of which one has no knowledge to where one merely confirms a well known fact. Data mining methods have been developed in machine learning, statistics, data visualization and deductive databases to examine the content of large databases (Chen *et al.*, 1996).

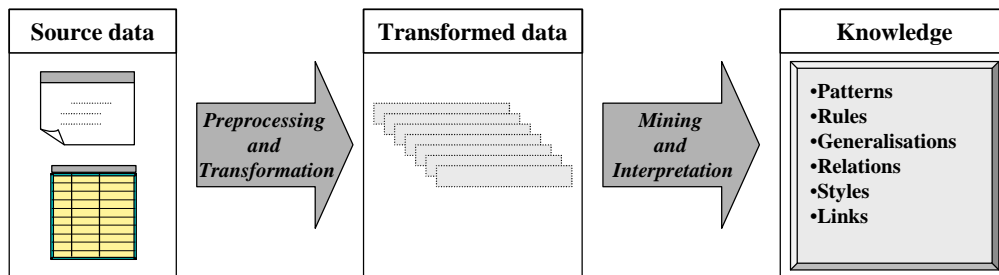


Figure 4. The data mining lifecycle.

Historically, data mining was initiated in large databases. The number of records in such databases can be estimated in millions, and the information is well-structured in table form according to a particular data model. However, a vast amount of design knowledge is coded in machine-readable text form or as electronic dictionaries, manuals and references, as CAD drawings or digital images. Most of the design data is unstructured, in comparison with the large databases used in conventional data mining. A characteristic of data mining and knowledge discovery is that the source data was created for purposes other than knowledge discovery. In our case, the source data is the text transcripts from the seminar discussions in the virtual campus.

The on-line discussions in the virtual campus are based on the idea of a computer-mediated moderated discussion. A computer-mediated moderated discussion can be held both in synchronous and asynchronous mode. Asynchronous discussions can be realised through a computer-mediated bulletin board, a discussion list-servers, or even by e-mail. Here we focus on synchronous seminar discussions taking place in the virtual campus classroom. Each seminar is devoted to particular theme as part of the course, complementary to lectures or course material, which students have previously attended or read. Figure 5 illustrates virtual seminar scenario.

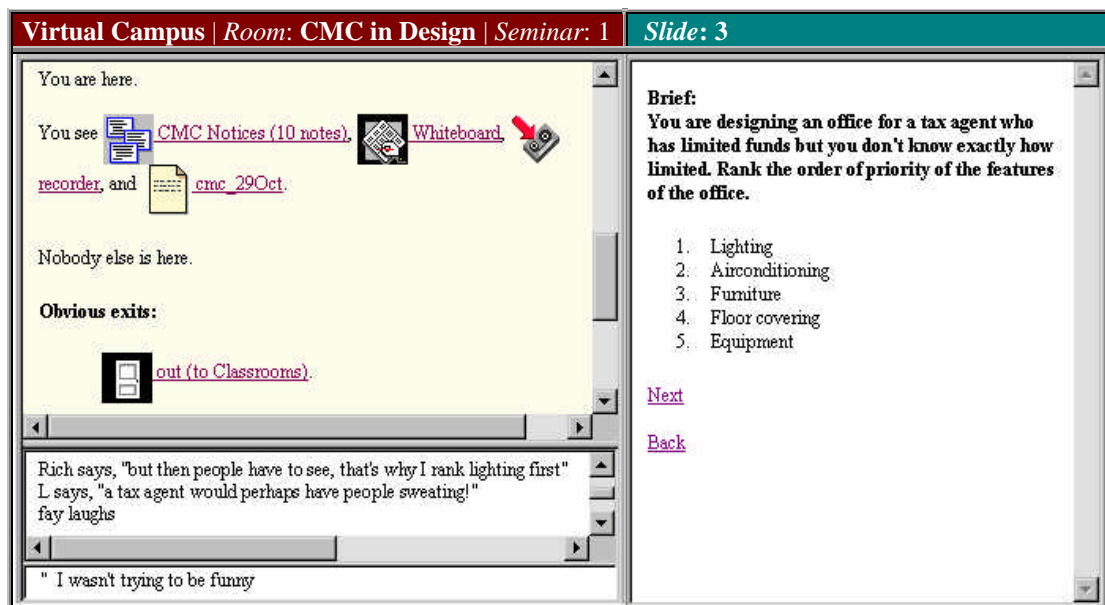


Figure 5. Virtual seminar in virtual campus environment.

Discussions in the virtual campus have an initial start-up period, when participants are getting comfortable. Then the discussion continues as in a physical seminar, except the participation rate tends to be higher and it can be more difficult to control the topic. The role of the instructor(s) is to prevent distractions and maintain the focus of the discussion. One of instructors (usually the specialist in the topic and the potential moderator in the discussion) prepares a preliminary agenda, corresponding to the course objectives. Mentors also watch that all students should have the opportunity to participate in the discussion.

Online seminars can form part of the student's assessment by including the amount and content of the student's participation. The virtual campus provides a means for recording in explicit and descriptive form all activities during the discussion in a format suitable for further quantitative and qualitative analysis. Thus we have the ability to analyse students' and instructors' contribution, providing feedback in both directions.

A virtual seminar can be represented as a sequence of activities, as shown in Figure 6. Each activity is described by an expression. An expression consists of a subject, who performs

the action or the utterance in the activity, a verb, which describes the action or utterance, an object towards which the action is directed and the content of the action. Using this formalism we can represent, analyse and compare virtual seminars implemented through different underlying environments. In this paper we consider only the virtual seminars implemented via text-based environments. A raw excerpt from the transcript of a seminar discussion on cyberspace is shown in Figure 7.

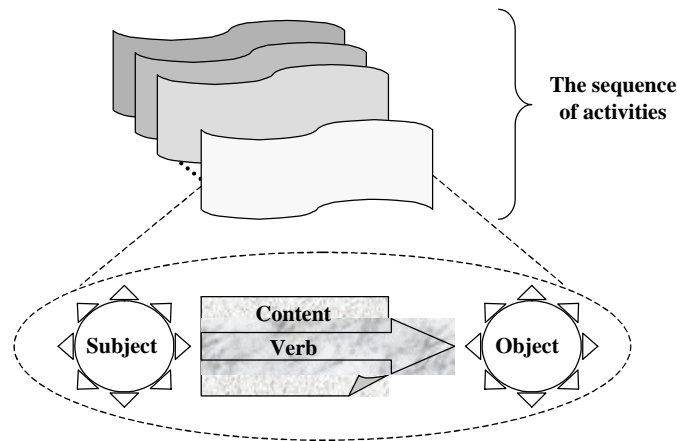


Figure 6. Formal representation of virtual seminar.

Course: Design - Theory and Practice Seminar: 2 Venue: TandP Room in Virtual Campus
Joan teleports in.
Jane says, "How eill we be doing this "
Martha says, "You will be creating an office"
Robert says, "More like a car crash tonight."
Bart says, " how crowded is this place"
Martha says, "meeting people here to dicuss projects"
Oliver says, "cyberspace is the new kind of architecture"
Bob says, "like a traffic jam in N.Y.C"
Herbert says, "somebody please tell me which slide we are in"
Rusty says, "I feel like talk to machine without personal contact""
Robert says, "Yeah!!!"
Martha [to herbert]: slide 5
Joan says, "slide 5"
Joan says, "but it does help those who were normally shy"
Oliver says, "that's right"
Martha says, "it takes time to get used to this kind of interaction"
Robert says, "There are no boundaries in cyberspace."
Martha says, "its not better or worse, it is just different"
Jane says, "Will these offices be communal?"
Sharon says, "perhaps you could"perhaps bearing in mind what works for conversation?"
Martha says, "you will have a personal office, so no crowds"

Figure 7. An unprocessed excerpt of a seminar discussion on cyberspace.

As a description of the sequence of activities during the seminar, the transcript is a mixture of quantitative and qualitative data. Initially, we format the transcript in a way that each line corresponds to a single activity. Then such characteristics as the total number of lines, the number of lines, which start with the same subject, are examples of quantitative data. The concepts that constitute the content of each activity are example of qualitative data. These data, provided in the records of the virtual seminar, are useful source for investigating the way the seminar has been run, students' and moderator's participation, the appropriate discussion styles and other interesting patterns that will help to understand and better apply virtual seminars in design education.

The raw seminar transcript may make sense to the participants in the discussion and not to someone analysing the content. An observer would not understand the flow of the conversation in Figure 7, or who Charles is in Figure 8. Is Charles the moderator of the seminar, or an external expert invited for a discussion on particular topic? The names of the participants do not indicate their roles in the discussion.

Course: AI in Design Seminar: 3 Venue: AI Room in Virtual Campus	
9	Mary has arrived.
10	Mary drops Charles.
11	Mary says, "Charles is here"
12	Anne says, "hi charles"
13	Charles [to Anne]: Hello there! How's it going?

Figure 8. Who is Charles?

Each transcript undergoes preliminary processing to create the corresponding data sets for statistical and content analyses, and a more meaningful copy of the transcript archived and used for further research and analysis. The preprocessing includes several stages. It starts with initial cleaning of some data headers and service information, and automatic numbering of all lines in the remaining body of the transcript. On the one hand, the numbering provides reference label for each line, i.e. seminar activity, assuming that we have already done the formatting step discussed before. These numbers then allow us to refer to a particular part of the discussion. For instance, the line numbers in Figure 8 show that this is part of the beginning of the seminar. On the other hand, the total number of lines gives us an idea about the length of the seminar discussion as a sequence of activities, rather than as a time interval. Considering that the length of the seminar is restricted in time by the course schedule, the number of lines also provides some idea about how active the participants were in the discussion.

The next step is connected with the line structure and content of the transcript record. In this paper we consider only transcripts which consist of natural language expressions. Some expressions can include one or more completed sentences, some only part of a sentence. Following the model in Figure 6, we split each expression into two parts. The subject, the verb and the object are labeled as a *left-hand side* and the content remains in the *right-hand side*. This is a conventional separation, merely for distinguishing between descriptive and actual content data.

From the initial transcript we generate two additional data sets: one in which we modify only the left-hand side, that gives us clear idea who is who through the whole seminar. We keep the right-hand side unchanged for quantitative analysis. The second data set is derived from the first one by normalising and coding the left-hand side in a way that both sides become consistent. The data transformation process includes two basic techniques: reference normalisation and role coding techniques.

Reference normalisation

We identify all login names used by each participant and all references within the discussion for each participant. After the name reference normalisation, each participant is represented by only one name spelling in the seminar. The case of reference normalisation within the content of the discussion (the left-hand side) is illustrated in Figure 9, where the person participating under the name "davidS" (line 129) is referred in the content of the discussion by another participant as "david" (line 132).

Course: CMC in Design Seminar: 2 Venue: CMC Room in Virtual Campus	
128	carol says, "the hearer only hears what he/she wants to hear"
129 *	davidS says, "say if thats the case then its not communication"
130	Sim [to Rob]: what about what?
131	Rob says, "that's very typical of a social situation"
132 *	carol says, "it's communication, david, just not very effective communication"
133	anna says, "if the hearer only hears what they want to hear, that's related to time"

Figure 9. The name reference within the discussion can differ from the login name.

Another part of the normalisation process is converting where necessary the right-hand side of each expression to the structure "subject-verb-object". For example, lines 4 and 7 in Figure 10 show that Tim talks to the people in the room and then to a particular person, respectively. Both lines are examples of the same activity:

Line 4: Subject (tim) - Verb (says) - Object () - Content (Yes, I guess so, but I'm not sure)

Line 7: Subject (tim) - Verb () - Object (Prof) - Content (can we bring Charles in here?)

Course: AI in Design Seminar: 3 Venue: AI Room in Virtual Campus	
4	Tim says, "Yes, I guess so, but I'm not sure "
5	Mary stands up from the big table.
6	Mary goes out.
7	Tim [to Prof]: can we bring Charles in here?
8	anmore says, "yes i've encountered him but have spent more time with other conversing bots"
9	Mary has arrived.
10	Mary drops Charles.
11	Mary says, "Charles is here"
12	Anne says, "hi charles"
13	Charles [to Anne]: Hello there! How's it going?
14	Blue_Guest says, "pick that robot up"
15	cookie sits down at the big table.
16	Orange_Guest says, "Put that robot down"
17	Tim [to Charles]: slow so far
18	Blue_Guest says, "maybe you should say something intellectual tim"
19	Tim says, "Someone else try to enhgage Charles in conversation"
20	Anne scrutinizes Charles's key word list...

25	Anne smiles at charles
26	Charles [to Anne]: It's always nice when somebody smiles.
27	Anne hugs charles

436	Anne . o O (and fast before he gets gagged)

Figure 10. Various constructs which require normalisation and role coding.

This activity is explicitly specified in line 4 by the verb "says", when in line 7 the verb is omitted. The object in line 4 is the group of people in the seminar, thus it is also omitted. Formally, normalisation requires substitution of the missing components so we have a uniform data structure. Another type of normalisation is the substitution of a word for a set of separated symbols. Line 436 in Figure 10 is example of such case, where we need to replace ". o O" with "thinks". Normalisation includes also correction of typing errors.

Normalisation considerably improves the structure of the transcript and partially its readability. However, it does not analyse the content of the seminar. The reader probably has noticed that the lines in Figure 8 are part of the excerpt presented in Figure 10. Yet, who is Charles?

Role coding

Once the normalization is done, we code the reference to each participant with their formal role. The aim of this operation is to obtain qualitative data sets, which, on the one hand, are suitable for quantitative analysis, and, on the other hand, are readable and meaningful to researchers, and do not violate the anonymity of the data source. We define two sets of role categories: basic and special.

Course: AI in Design Seminar: 3 Venue: AI Room in Virtual Campus	
4	Moderator says [to audience], "Yes, I guess so, but I'm not sure "
5	Instructor_1 stands up from the big table.
6	Instructor_1 goes out.
7	Moderator says [to Instructor_1]: can we bring Robot_1 in here?
8	Student_1 says [to audience], "yes i've encountered him but have spent more time with other conversing bots"
9	Instructor_1 has arrived.
10	Instructor_1 drops Robot_1.
11	Instructor_1 says [to audience], "Robot_1 is here"
12	Student_1 says [to audience], "hi Robot_1"
13	Robot_1 says [to Student_1]: Hello there! How's it going?
14	Student_4 says [to audience], "pick that robot up"
15	Student_6 sits down at the big table.
16	Student_3 says [to audience], "Put that robot down"
17	Moderator says [to Robot_1]: slow so far
18	Student_4 says [to audience], "maybe you should say something intellectual Moderator"
19	Moderator says [to audience], "Someone else try to engage Robot_1 in conversation"
20	Student_1 scrutinizes Robot_1's key word list...
25	Student_1 smiles at Robot_1
26	Robot_1 says [to Student_1]: It's always nice when somebody smiles.
27	Student_1 hugs Robot_1
436	Student_1 thinks [to audience] (and fast before he gets gagged)

Figure 11. Normalised and role-coded version of the transcript in Figure 10.

The basic set encompasses the role categories, which are common for the course discussions. In the case of virtual seminars, which are part of a regular course program, the basic set of role categories includes *student*, *moderator*, *instructor* and *expert*. The last category accommodates external professionals and world-known experts, both from industry and academia that could participate in the virtual seminar. Each participant is assigned a name according to the following format: "Category_Number", for example, "Instructor_2", "Student_1", "Student_2". In practice, only one person is assigned the role of a moderator, thus in our coding we use "Moderator" instead of "Moderator_1". Such naming convention explicitly shows the role of each participant and allows easily the automation of analysis procedures whether we make investigations at individual or category level. The basic set of roles provides the foundation analysis both within a single course and across courses. The special set includes categories, which reflect the roles required by the specific features of each course program.

Now a simple glance at Figure 11 is enough to unravel the mystery of Charles. The discussion, devoted to the principles of knowledge-based design, included an on-line design of a touring robot for a text-based virtual environment. The data in Figure 11 also gives far better vision of what is happening at each particular stage of the discussion.

Data mining through text analysis

The data from the transcripts can be analysed at least in two ways: confirmatory and exploratory. When the initial hypothesis shape the schema of the confirmatory analysis, the

exploratory data analysis (EDA) and data mining (DM) are performed to discover interesting patterns or structures that require explanation. In this way, EDA/DM² provides us with information for generating hypothesis for later confirmatory study. Exploratory analysis is suitable for new areas of research. This section illustrates how we can analyse the data sets that we have constructed from the transcript of a virtual seminar. Our aim is to present the methodology rather than the results of complete research.

Earlier we have emphasized that the transcripts can be viewed both as quantitative and qualitative data. Sudweeks and Simoff (1998) introduced CEDA³ methodology, which provides a framework for combining quantitative and qualitative methods for doing research in internet-mediated communications. CEDA employs quantitative methods to extract reliable patterns, whereas qualitative methods are incorporated for content analysis of the essence of phenomena. In virtual seminars we can combine statistics derived from the quantitative data with results obtained by text data mining and content analysis of the qualitative data from expressions.

An essential part of the quantitative analysis of the transcript data is the development of the coding scheme. Current coding schemes used in the research in application of virtual environments in education are stand-alone single level coding schemes. They are designed to extract information about considered features at particular level of details. Such schemes do not support drill-down analysis - a popular data mining techniques, used in new research areas when there is a lack of knowledge about the object of investigation.

To be able to implement drill-down data mining techniques consistently we propose to use open hierarchical coding schema, designed to conduct investigations on increasing level of detail and utilise the results obtained on previous levels. The initial step in this approach is similar to the initial data-driven coding of the two-step coding schema used for the analysis of the documented designs in computer-mediated collaborative design experiments (Maher *et al.*, 1997a, 1997c). However, its further steps are in some sense the opposite to the two-step schema, in which the coded output of the first step is regrouped according to the coding schema using in the next step.

The coding schema shown in Figure 12 illustrates this idea. The seminar activities, described in a transcript can be classified initially into two general categories: ones that are related to the theme of the seminar and ones that are unrelated (like 'passing a mug of coffee'). This classification is sufficient if we are interested only in a measure of whether the participant was able to stay on the topic of the seminar. The automation of this procedure uses sorting and text analysis.

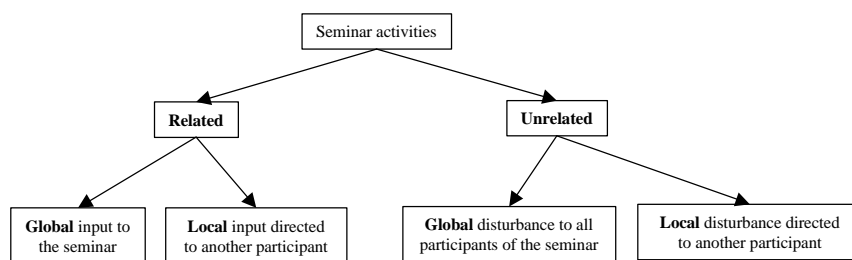


Figure 12. An example of open hierarchical coding schema

We investigate the transcripts of three seminars within the elective course "Computer-Mediated Communications for Designers". In this example, the course included four students and three instructors, labeled "Student_1", ..., "Student_4" and "Instructor_1", ..., "Instructor_3", respectively. For each seminar students were given preliminary readings – a major paper on the topic and some additional selected readings. In each seminar one of the instructors act as

² Exploratory data analysis and data mining share the same functional objective - the discovery of unknown patterns and relationships in collected data, thus we use joint EDA/DM abbreviation.

³ CEDA - Complementary Explorative Data analysis.

a moderator, having the freedom to conduct the discussion in accordance with the topic. The other instructors participated with their expertise.

Table 1 represents the results of the coding to three seminars according to the first level of the coding scheme in Figure 12. We consider only the lines between two margin activities performed by the moderator: the opening and the closing of the seminar. The shaded cells indicate the instructor who was the moderator of the corresponding seminar. Null values in column "Related" indicate that the person has not participated in the seminar. A simple bar-graph visualisation of the data in Table 1 is presented in Figure 13. It provides a rough pattern of the concentration of each participant during the seminar. All three seminars are intensive, with the leadership of the third seminar and the smallest concentration during the second.

Table 1. Seminar activities categorised into "Related" and "Unrelated" to the seminar theme

Participants	Seminar 1		Seminar 2		Seminar 3	
	Related	Unrelated	Related	Unrelated	Related	Unrelated
Instructor_1	3	0	79	0	86	0
Instructor_2	109	1	36	3	28	0
Instructor_3	20	7	41	4	0	0
Student_1	102	6	40	12	47	0
Student_2	144	4	67	2	35	0
Student_3	38	10	60	10	0	0
Student_4	0	0	15	7	0	0
Total	416	28	338	38	196	0

We consider only the left branch in the coding scheme further, considering the activities related to the topic of the seminar. The "discussion pies" in Figure 14 and Figure 15 illustrate the individual (a) and categorised (b) participation. The visual balance of Figure 14a indicates that the participation in the discussion by Student_1, Student_2 and the Moderator (Instructor_2) are nearly the same. As shown in Figure 15a the participation of the other instructors is relatively small compared to the Moderator's contribution. Patterns in Figure 14b indicate the different character of the discussion during Seminar 2. Student_2 continues to be the most active among the students. Figure 15a shows that more than two thirds of the utterances came from the students; Figure 15b indicates that student participation was only a bit more than 50%; and Figure 15c indicates that student participation was less than 50%. This could be a warning sign that student participation is dropping.

Considering the role and participation of the instructors and moderators is also of interest. Instructor_3 has considerably less participation than the other instructors. A possible hypothesis is that the moderator didn't have enough experience, thus leaving the initiative to the other instructors. Considering the effectiveness of the moderator by comparing the pie graphs, we see that Figure 14a and Figure 14b show more uniform participation from the students. Consequently, we can state that the amount of control imposed from the moderator shapes the visual pattern of the discussion, therefore, the visual pattern of the discussion allows us to infer the experience and quality of moderation.

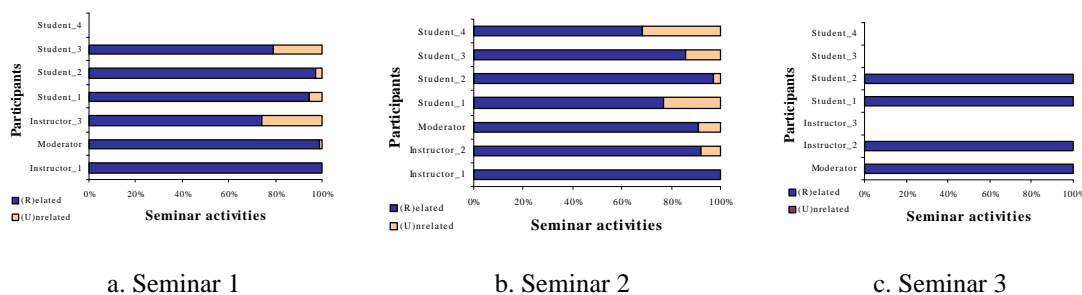


Figure 13. Participants profile: related versus unrelated seminar activities.

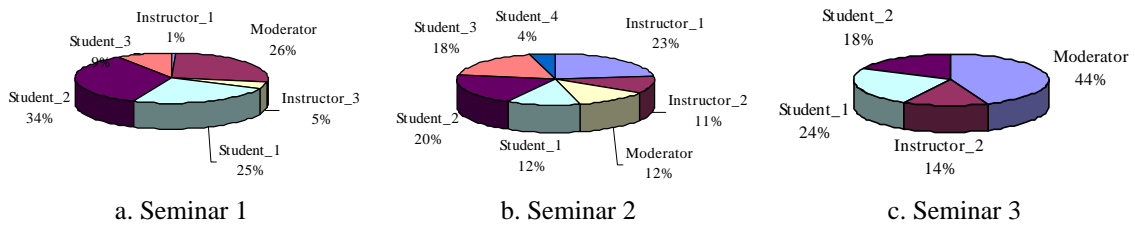


Figure 14. Line-based estimate of individual activity.

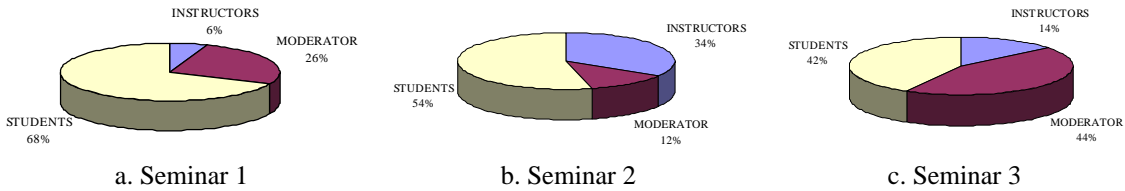


Figure 15. Line-based estimate of group activity.

Relying on line-based estimators of the participants' activity could be misleading in cases when participants used either very short expressions - one or two words, as in lines 411-413 in Figure 10, or very long (approximately more than 10-15 words), for example, when defending particular design solutions. The estimators can be corrected by introducing weights, based on the average length of expressions and length variance. However, word-based estimators, derived from the total amount of words, alphanumeric and other characters provide idea about individual and group activities. We illustrate this approach using the number of words in the seminars, presented in Table 2.

Figures Figure 16 and Figure 17 provide visualization of the participants' activity based on the data in Table 2. Note that segmentations of the pie graphs in Figures Figure 14 and Figure 16, and Figure 15 and Figure 17, respectively, are very similar, indicating that there average number of words per expression across the seminars is neither in the "too short" area, nor on the "too long" side. The estimated average values are presented in Table 3. Thus, visual patterns of discussion are applicable for assisting on-line student assessment and instructor/moderator adjustment. The information about the amount of other characters used in expressions gives an idea about the use of punctuation and other symbols in expressing additional cues in the text. The relative usage based on the data in Table 2 is shown in Figure 18.

Table 2. Number of words used during the seminars.

Participants	Seminar 1			Seminar 2			Seminar 3		
	Words	Alphanumeric characters	Other characters	Words	Alphanumeric characters	Other characters	Words	Alphanumeric characters	Other characters
Instructor_1	80	345	3	462	1877	28	728	2783	39
Instructor_2	715	3179	135	145	647	13	215	886	32
Instructor_3	181	723	32	275	1205	24	0	0	0
Student 1	754	3169	155	304	1386	54	228	919	69
Student 2	1157	4946	203	503	2213	47	271	1046	31
Student 3	203	836	12	252	1113	10	0	0	0

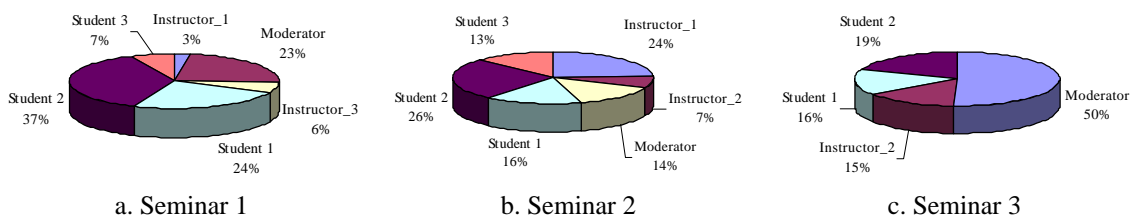


Figure 16. Word-based estimate of individual activity.

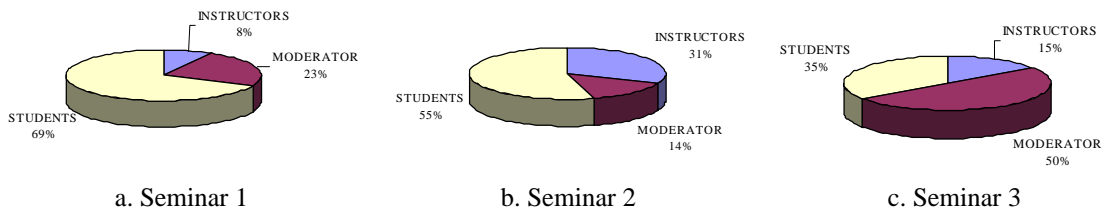


Figure 17. Word-based estimate of group activity.

Table 3. Average number of words per expression.

Participant	Seminar1	Seminar2	Seminar3
Instructor_1	8.0	8.4	8.5
Instructor_2	6.6	4.1	7.7
Instructor_3	7.9	6.7	0.0
Student 1	5.5	7.4	4.9
Student 2	7.0	7.6	7.5
Student 3	5.1	4.5	0.0
Seminar Average	6.7	6.5	4.8

The exploration of the discussion structure and collaboration during the seminar requires the use of the next level in the coding scheme in Figure 12. In each seminar, from the activities related to the discussion, the lines that respond to a specific person are classified as local threads, the rest are marked as global. Shaded lines 248, 250 to 254 in Figure 18 are examples of local threads. The other threads we label as global.

Course: CMC in Design Seminar: 2 Venue: CMC Room in Virtual Campus	
248	Student_3 says [to Moderator]: thats just what were doing
249	Student_4 says, "Student_1 more a level of transparency in the technology..speed helps.."
250 *	Student_3 says [to Student_2]: yup
251 *	Student_2 says [to Student_3]: do you think that using a text based chat solves the problem for today?
252 *	Student_3 says [to Student_2]: no way but it's a start
253 *	Student_2 says [to Student_3]: or is the text based chat still a hindrance?
254 *	Student_3 says [to Student_2]: a start in understanding
255	Student_2 says, "I realise that synchronous comms for designers are very good"
256	Instructor_1 says, "I'm a little lost"
257	Instructor_1 says, "It seems that Student_2 has hijacked the discussion"
258	Student_4 says, "text based doesnt allow us to draw, like whiteboards on vc.."
259	Student_1 says, "particularly over long distances."
260	Instructor_1 says, "Let's pick a slide and start on those topics"
261	Instructor_1 says, "how about slide #2"
262	Student_2 says, "I was just getting lonely - no one was talking to me...."

Figure 18. An example of a long local thread which can dominate the seminar

The result of the coding is presented in Table 4. The shaded cells indicate the instructor who was the moderator of the corresponding seminar.

Figure 19 illustrates the results. Figure 19a shows very few local threads, which we consider to be an indicator of tight collaboration during the discussion. An increase in the amount of local threads can be caused by problems in moderation (Figure 19b) or if the seminar theme includes a number of issues and the discussion of these goes in parallel (Figure 19c). Sometimes it could be a result of preferred style of behaviour during the virtual seminar.

The analysis above does not provide much information about whether the discussion was related to the topic, or how much particular participant contributed to the topic of the

discussion or not. Text-based virtual seminars provide the necessary data for extracting such information.

Table 4. Related seminar activities classified into “Local” and “Global” threads.

Participants	Seminar_1		Seminar_2		Seminar_3	
	Global	Local	Global	Local	Global	Local
Instructor_1	3	0	24	17	56	30
Instructor_2	98	11	73	6	15	13
Instructor_3	17	3	27	9	0	0
Student_1	102	0	38	2	39	8
Student_2	139	5	55	12	11	24
Student_3	38	0	47	13	0	0
Student_4	0	0	15	0	0	0
Total	397	19	279	59	121	75

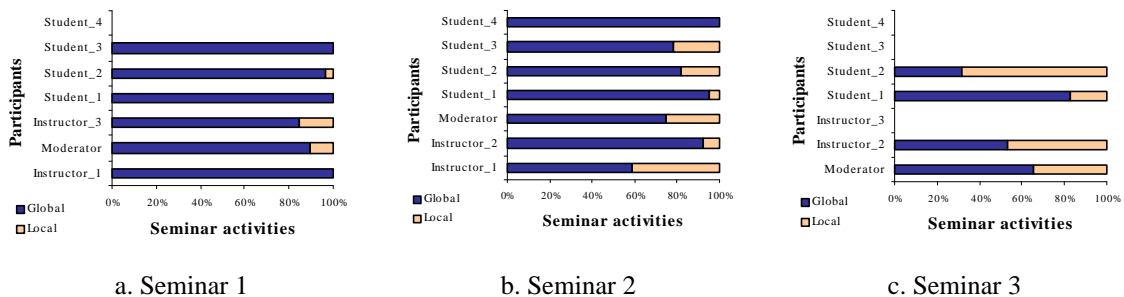


Figure 19. Participants profile: Related activities: Local versus global threads in seminar discussions.

Computer-based content analysis is a state-of-the-art activity, considerably improved with the overall progress in computing. In virtual seminars it provides a quantitative analysis of the qualitative seminar data, offering objective results to supplementing instructor's subjective evaluation of the discussion. The content analysis of seminar transcripts starts with the specification of a list of words that distort the content of the transcripts. The list of words, excluded from the analysis, consists of prepositions, conjunctions and disjunctions, articles, pronouns, and other particular words. The last group is used for analysing the content. Content analysis software counts the occurrences of the remaining words and excludes the infrequently occurring words. The investigator defines the threshold and also a list of thematic key words, whose occurrence has to be considered even if it is below the threshold. In addition to term frequencies, automated content analysis investigates the co-occurrence of the frequently occurring words and their clustering based on these co-occurrences.

One of the problems in applying content analysis methods to virtual seminar data is the relatively low word frequencies. Qualitative techniques are usually applied for the analysis of focused interviews. The word frequencies in such interviews are about 100-300. In a seminar transcript these frequencies are usually no greater than 10.

Evaluating collaboration through data analysis and visualisation

Collaborative design denotes design activity when more than one person works on a single design problem, having a common goal or intent. Collaboration is possible when the collaborators share activities and information to achieve the common goal. Effective collaboration occurs when the collaborators share design tasks, communication, representation and documentation (Maher *et al.*, 1997b).

To facilitate collaboration between team members, the virtual campus provides each design team with a separate bulletin board and design logs viewer. These bulletin boards are

independent of the major studio bulletin board. To protect the privacy of teamwork and information exchange, each bulletin board and design log viewer were password protected. The management of these components is implemented through the VDS Team manager, whose front panel is shown in Figure 20.

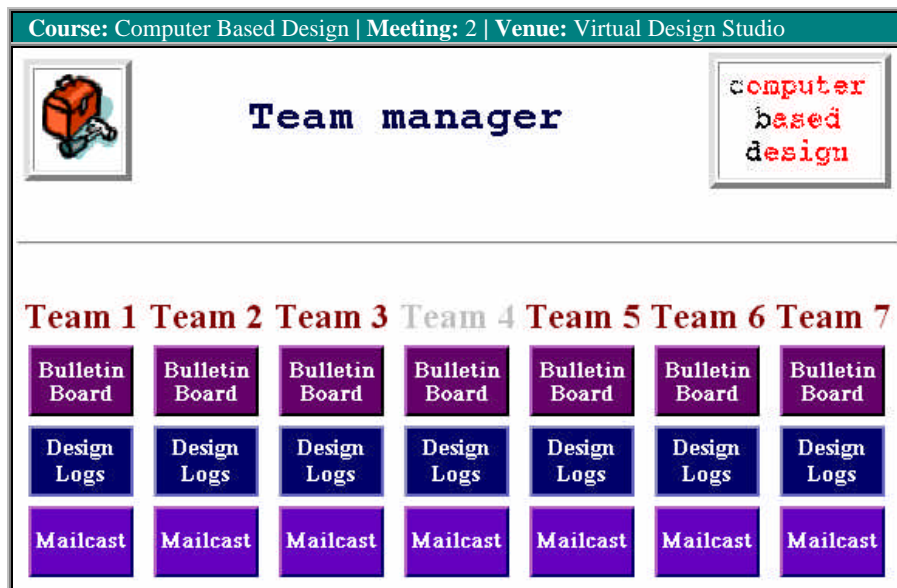


Figure 20. The "front panel" functionality of Team manager.

We assume that the major exchange of information between designers is realised through the bulletin board. Then we consider that both the content and the pattern of the messages on the bulletin board reflect team collaboration. In this paper we consider message patterns assuming that the content analysis has established a correspondence between the subject and the content of the message.

The messages on the board are grouped in threads. Berthold *et al.* (1997, 1998) proposes a threefold split of the thread structure of e-mail messages in discussion archives in order to explore the interactive threads. 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. These parameters are important for a detailed analysis of the messages and the structure of the discussion. For qualitative judgements we introduce *thread-length* of a thread as the maximum of the sum of reference-height and reference-width over the messages in the thread, and *thread-width* as the maximum among the sums of reference-widths at each level.

In addition to the threefold split, Sudweeks and Simoff (1998) included the time variable explicitly. Each message is completely identified by two indices - one for its level and one for its position in time in the sequence of messages at this level. The *time envelope*, defined as the time interval between the Figure 21 shows the visual representation of the thread reference tree. "M" and "t" denote message and time points, respectively. This model 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-thread discussions).

We categorise bulletin board messages in project-based design courses into two major groups: related to design tasks and management. Task-related messages consider particular topic of the project, when management messages are connected with the arrangement of meetings, changes in schedules and places.

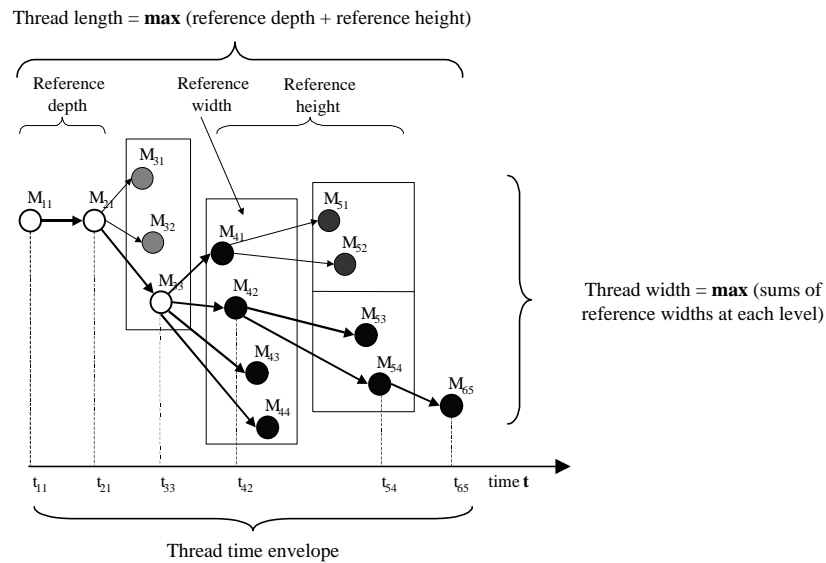


Figure 21. Visualisation of a thread reference tree (adapted from Sudweks and Simoff (1998)).

The graph structure reference in Figure 21 shows the type of design collaboration that occurs in a team environment. Once the representation is structured hierarchically there are several techniques to display its structure. Donald Knuth, in the first volume of his "bible for programmers" described few notations for visualising tree structures - graph tree, identity graph, and a nested set (Knuth, 1973, pp. 305-313). The graph visualisation is more helpful for smaller threads and when we are interested in quantitative information. Graph tree is also useful when we consider the dynamic properties of the discussion. A nested set visualisation provides better illustration when we are interested more in the type of collaboration and can omit the time.

Collaboration on a shared design task can be considered at different levels of abstraction and "degrees" of task sharing. Maher *et al.* (1997b) identify two extreme approaches to sharing design tasks during collaboration: *single task collaboration* and *multiple task collaboration*.

During single task collaboration the resultant design is a product of a continued attempt to construct and maintain a shared conception of the design task. In other words each of the participants has his/her own view over the whole design problem and the shared conception is developed, roughly speaking, by the "superposition" of the views of all participants. This type of collaboration is reflected in the bulletin board message structure by the existence of task-related threads running in parallel. The collaboration is closer when there are large thread widths and lengths. The time envelopes also can be larger, compared to the average time envelope - participants usually revise some of their earlier views. The bulletin board also includes several long threads usually connected with crucial issues in the project.

During multiple task collaboration the design problem is divided among the participants so that each person is responsible for a particular portion of the design. Thus, multiple task collaborative design does not necessarily require the creation of a single shared design conception, though designers work cooperatively in a common electronic workspace. This collaboration style is reflected by isolated task-related messages or short threads. Most of the messages or the threads are related to the management of the project.

Student collaboration in course design projects is usually a combination of these approaches. Some of the teams usually followed the single task collaboration until they created a joint understanding and view on the problem and then they had some separation of the implementation tasks. Other teams started with distributing the initial research tasks with some

degree of single task collaboration during the composition of the final design. Figure 22 illustrates a bulletin board of a single task collaborating team. Graph-tree and nested set visualisation of this bulletin board fragment is shown in Figure 23 a and b, respectively. The graph-tree in Figure 23a illustrates the idea of parallel threads corresponding to different project tasks, although the threads are relatively short (the greatest thread-length is 3, but for most of them the thread-length is 2) and "narrow" (thread-width for each one is 1). The nested set in Figure 23b provides better qualitative picture of collaboration. That is, more nested rectangles indicates more collaboration.

Course: Computer Based Design Bulletin Board: Team 2 Venue: Virtual Design Studio		
4	Lighting etc - Derek 08:46:10 10/16/97 (1)	(M ₁₄)
	• Re: Lighting etc - Sophie 10:43:18 10/17/97 (0)	(M ₂₄)
3	Seating - Derek 15:18:57 10/14/97 (2)	(M ₁₃)
	• Re: Seating - marky 17:22:56 10/14/97 (1)	(M ₂₃)
	• Re: Seating - Sophie 09:03:27 10/15/97 (0)	(M ₃₃)
2	Product Research. - Derek 14:37:43 10/14/97 (1)	(M ₁₂)
	• Re: Product Research. - mark 17:20:16 10/14/97 (0)	(M ₂₂)
1	Another idea - Sophie 14:24:18 10/14/97 (1)	(M ₁₁)
	• Re: another idea - Derek 14:40:00 10/14/97 (0)	(M ₂₁)

Figure 22. Bulletin board fragment with task-related messages, visualised as indentation graph.

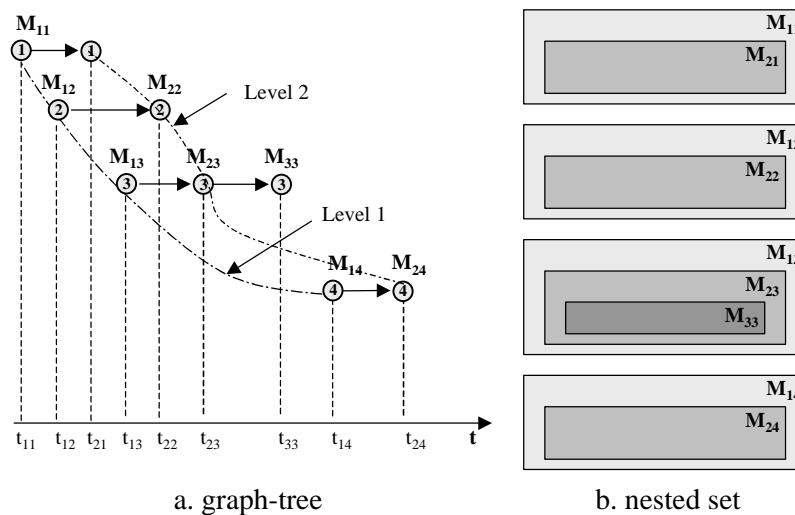


Figure 23. Formal visualisation of the bulletin board fragment in Figure 22

Summary

Creating collaborative design environments usually involves the consideration of the technology that can enable communication and information sharing. These environments can provide for both synchronous and asynchronous communication. We demonstrate how the record of the communication can be used to analyse participation and contribution to the collaborative project. We present our Virtual Campus which provides an educational environment for on-line seminars and threaded discussions on electronic bulletin boards.

The analysis of the record of on-line seminars can show who has participated and the extent of their participation. Evaluating individual participation can identify not only the amount of contribution, but the content of the contribution. The analysis of the bulletin board discussion can provide indications of the type of collaboration and the extent of the interaction on ideas and management. Although we present the benefits of this kind of analysis for

educational environments, the analysis of professional collaboration can provide feedback on project participation and the effectiveness of the collaborative team.

The method for analysing participation is based on text analysis and data mining. The analysis framework provides extensive support for research both in educational and collaborative design environments. Presented techniques can be employed in the protocol analysis of collaborative design in Web-based design environments. Current environments do not include on-line data analysis and analytical processing features. Data analysis, mining and visualisation could become integral part of design educational and collaborative environments. Providing means for on-line analysis of communication and collaboration is invaluable for the active learning and research in computer-mediated design.

References

- Berthold, M. R., Sudweeks, F., Newton, S. & Coyne, R. (1997). Clustering on the Net: Applying an autoassociative neural network to computer-mediated discussions. *Journal of Computer Mediated Communication*, 2(4), (<http://www.ascusc.org/jcmc/vol2/issue4/berthold.html>).
- Berthold, M. R., Sudweeks, F., Newton, S. & Coyne, R. (1998). It makes sense: Using an autoassociative neural network to explore typicality in computer mediated discussions. In F. Sudweeks, M. McLaughlin & S. Rafaeli (Eds), *Network and Netplay: Virtual Groups on the Internet*, Menlo Park, CA: AAAI/MIT Press, pp.191-220.
- Chen, M-S., Han, J. and Yu, P.S. (1996). Data mining: an overview from a database perspective. *Knowledge and Data Engineering*, 8 (6), 866-883.
- Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery: An overview. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (eds), *Advances in Knowledge Discovery and Data Mining*, AAAI Press, Boston, MA, pp. 1-34.
- Knuth, D. (1973). *The Art of Programming*.
- Lockard, Abrams, and Many (1994). *Microcomputers for Twenty-First Century Educators*
- Maher, M. L., Cicogniani A. and Simoff, S. J.(1997a), An Experimental Study of Computer Mediated Collaborative Design, *International Journal of Design Computing*, 1, <http://www.arch.usyd.edu.au/kcdc/journal/index.html>
- Maher, M. L., Simoff, S. J. and Cicogniani A. (1997b). Potentials and limitations of Virtual Design Studio. *Interactive Construction On-line*, January, a1.
(also available at <http://www.arch.usyd.edu.au/~mary/VDSjournal/index.html>)
- Maher, M.L., Simoff, S.J. and Cicognani, A. (1997c). Observations from an experimental study of computer-mediated collaborative design. In Maher, M.L., Gero, J.S. and Sudweeks, F.(eds) *Preprints Formal Aspects of Collaborative CAD*, pp.165-185.
- Mitchell, W. J. and McCullough, M. (1995). *Digital Design Media*. Van Nostrand Reinhold, New York.
- Savery, J. R. and Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 35(5), 31-38.
- Simoff, S. J. and Maher, M. L. (1997a). Design education via web-based virtual environments. In T. Adams (ed.), *Computing in Civil Engineering*, Proceedings of the Fourth Congress of Computing in Civil Engineering, ASCE, N.Y., pp. 418-425.
- Simoff, S.J. and Maher, M. L. (1997b) Web-mediated courses: the revolution in on-line design education. In H. Ashman, P. Thistlewaite, R. Debreceeny, A. Ellis (eds), *Into the Mainstream - the Web in Australia*, Southern Cross University Press, Lismore, pp. 143-154. (available also on a CD AusWeb97, ISBN (Mac) 1 875 855 23 8, ISBN (Win.) 1 875 855 24 6, By A. Ellis and R. Debreceeny, Southern Cross University Press. On-line: <http://ausweb.scu.edu.au/proceedings/simoff/index.html>
- Sudweeks, F. and Simoff, S. J. (1998). Complementary explorative data analysis: the reconciliation of quantitative and qualitative principles. In Jones, S. (ed.), *Doing Internet Research*, Sage Publications, Thousand Oaks, CA (in print).

Streibel, M.J. (1991). Instructional plans and situated learning. In G. J. Anglin, (ed.), *Instructional Technology, Past, Present, and Future*. Libraries Unlimited, Englewood, CO, pp. 117-132.

Sutherland, J. (1997). The Java Revolution. *Sun Expert*, **8** (1), pp. 43-54.

Taylor, D. and O'Connor, K. (1997). Experience with remote collaboration for concurrent engineering. In Maher, M.L., Gero, J.S. and Sudweeks, F.(eds) *Preprints Formal Aspects of Collaborative CAD*, pp.29-47.

WWW References

[HREF1] The Virtual Campus at the University of Sydney
<http://www.arch.usyd.edu.au:7778/>

[HREF2] World Wide Web Course Tools
<http://homebrew1.cs.ubc.ca/webct/webct.html>