

Using Information Visualisation to Facilitate Instructors in Web-based Distance Learning

A dissertation presented by

Riccardo Mazza

Supervised by

Prof. Marco Colombetti

Submitted to the

Faculty of Communication Sciences

University of Lugano

for the degree of

Ph.D. in Communication Sciences

June 2004

Board

The commission board comprises the following people:

Supervisor: Prof. dr. Marco Colombetti, Faculty of Communication Sciences, University of Lugano

Advisor: dr. Vania Dimitrova, School of Computing, University of Leeds (UK)

Internal examiner: Prof. dr. Lorenzo Cantoni, Faculty of Communication Sciences, University of Lugano

External examiner: Prof. dr. P. M. E. De Bra, Computer Science Department, Eindhoven University of Technology (NL)

The candidate confirms that the work submitted is his own and that the appropriate credit has been given where reference has been made to the work of others.

Abstract

The World Wide Web provides great opportunities for creating virtual classrooms of learners and instructors involved in distance education. Many software environments take advantage of the client-server communication on the Internet and support open and distance learning. Using environments called Course Management Systems (CMS), instructors can distribute information to students, produce content material, prepare assignments and tests, engage in discussions, and manage distance classes. Although CMS offer many features that benefit the learners (e.g. asynchronism, interactivity, time and space independence), some problems still remain due to the lack of effective support for instructors.

Educational research shows that monitoring the students' learning is an essential component of high quality education, which is critical in distance learning. Good on-line tutoring requires to understand the needs of individual learners and to provide adapted help. This can be achieved by monitoring regularly the students' activities and be aware of what the students are doing in the course, e.g. which concepts are known, are students participating in discussions, have they read the course material, and how well do they perform on quizzes. CMS accumulate large log data of the students' activities in a distance course and usually provide some monitoring features to enable the instructors view aspects of the data. However, this complex, multidimensional data is commonly presented in a format that is poorly structured and difficult to understand. Hence, it is rarely used by distance learning instructors.

We argue that the instructors can use the student tracking data collected by the CMS more effectively when it is represented graphically by employing appropriate visualisation techniques. Information Visualisation examines techniques for processing and pictorially representing a vast amount of abstract data, so that the data can be comprehended and interpreted by people. This thesis proposes the use of Information Visualisation to graphically represent student tracking data in a Web-based CMS. A systematic investigation is undertaken to find which visualisation techniques are appropriate to illustrate student tracking data, based on instructors' requirements and evaluation. An extension for a generic CMS was designed and a prototype, called CourseVis, was developed. It obtains tracking data from a commercial CMS (WebCT is used in the current implementation), transforms the data into a form convenient for processing, and generates graphical representations that can be explored by instructors to examine social, cognitive, and behavioural aspects of distance students.

The evaluation of CourseVis has shown that graphical representations produced with it can help instructors to identify individuals that need particular attention, discover patterns and trends in accesses and discussions, and reflect on their teaching practice.

Acknowledgments

I would like to thank my supervisor, prof. Marco Colombetti for many years of support and advice in directing my research efforts. I would also like to acknowledge the other members of the committee, Prof. Lorenzo Cantoni and Prof. Paul De Bra for their suggestions in improving this work.

I would like to thank also many other people who contributed to this thesis. First of all, thanks to Dr. Vania Dimitrova and the School of Computing of University of Leeds. She has spent much time in discussing, commenting and improving my work through all the period of work on this research. This thesis has been in part undertaken at the Informatics Research Institute of the University of Leeds thanks to a fellowship awarded by the Swiss National Science Foundation. A number of people contributed to this work: Andrea Scolari, Roberto Righetti and Giosia Poma dedicated time and efforts to implement part of the code of the prototype as a practical experience at the University of Applied Science of Southern Switzerland. The evaluation of this work was possible thanks to the kind contribute of these other people: Sharon Dowdy, Julika Matravers, Tony Lowe, and Liz Minton from Leeds; Andrea Rocci, Marco Zaffalon, Daniela Pellizzali, Luca Botturi, Jacopo Armani, and Carlo Lepori from Lugano.

I would like to thank the people who read part of this work and made interesting feedback that allowed me to improve the quality of the content and writing, among these are Giuseppe Arbia and Bhavana Rehani from NCST (Mumbai - India).

Last, but not least, thanks to my wife and sons for their support and their encouragement with love and patience.

Contents

Conventions	xviii
Abbreviations	xix
List of published papers	xx
1 Introduction	1
2 Web-Based Distance Education	6
2.1 The evolution of education	7
2.2 An introduction to Web-Based Distance Education	7
2.3 Benefits of Web-based Distance Education	9
2.4 Course Management Systems (CMS)	10
2.4.1 Educational components of a Course Management System	11
2.4.2 Available Course Management Systems	12
2.5 Problem analysis	13
2.6 The proposed solution	19
2.7 Summary	19
3 Information Visualisation and its Applications in Educational Systems	21
3.1 Introduction	21
3.1.1 Graphics for presentation	23
3.1.2 Graphics for explorative analysis	23
3.1.3 Graphics for confirmative analysis	24
3.2 Information Visualisation	24
3.2.1 Cognitive amplification	27
3.2.2 Issues to consider in Information Visualisation	31
3.2.3 Techniques for representing univariate and bivariate linear data	32
3.2.4 Techniques for representing trivariate linear data	33

3.2.5	Techniques for representing multivariate linear data	34
3.2.6	Techniques for representing spatial data	38
3.2.7	Techniques for representing hierarchical and network data	38
3.2.8	View transformations	40
3.3	Visualisations in educational software	43
3.3.1	Visualisation of user models	43
3.3.2	Visualisation of open student models	46
3.3.3	Visualisation of on-line communications	52
3.4	Summary	56
4	Gathering Instructors' Requirements	57
4.1	Introduction	57
4.2	Objectives	57
4.3	Method	58
4.3.1	Participants and setting	58
4.3.2	Procedure	59
4.3.3	Measurement	59
4.4	Results	60
4.4.1	Response rate	60
4.4.2	Survey reliability	60
4.4.3	Analysis of responses	61
4.5	Discussions with distance teachers	68
4.6	Discussion and recommendations	69
4.6.1	Social aspects of learners	69
4.6.2	Cognitive aspects of learners	70
4.6.3	Behavioural aspects of learners	70
4.7	Implication for the research and practice	71
4.7.1	About the course manipulation	71
4.7.2	About the class dimension	71
4.7.3	About the learning platform	71
4.7.4	About the information to be represented	72
4.8	Summary	72
5	The CourseVis	74
5.1	Design guidelines	75
5.2	The CourseVis system architecture	76
5.2.1	CMS	77
5.2.2	Student Data Exporter	77

5.2.3	Raw Data Repository	82
5.2.4	Domain Designer	82
5.2.5	Domain Model Repository	86
5.2.6	Data Processing and Visualisation Procedure	86
5.3	The use of OpenDX	87
5.4	An example of a graphical representation produced with CourseVis	89
5.4.1	Extraction of data from the CMS	90
5.4.2	Mapping XML student data structure into the repository	91
5.4.3	Selection of relevant data	92
5.4.4	Data transformations	93
5.4.5	Visual mappings	93
5.4.6	View Transformation	94
5.5	A model for the design space	98
5.6	Summary	100
6	Graphical Representations	101
6.1	The Java on-line course	101
6.2	Graphical representations produced in CourseVis	103
6.2.1	Social aspects	103
6.2.2	Cognitive aspects	108
6.2.3	Behavioural aspects	113
6.3	Visualisation of different stages of the course	117
6.4	Visualisation of different editions of the same course	118
6.5	Scalability of the representations	121
6.6	Case studies	121
6.7	Promoting instructor's reflection	128
6.8	Discussion	128
6.9	Summary	129
7	Evaluation	131
7.1	Introduction	131
7.1.1	Evaluation goals	132
7.1.2	Evaluation techniques	132
7.2	Criteria to assess	134
7.3	Focus group	135
7.3.1	Procedure	135
7.3.2	Results	136
7.4	Experimental study	141

7.4.1	Hypotheses	142
7.4.2	Procedure	144
7.4.3	Results	147
7.5	Semi-structured interviews	149
7.5.1	Procedure	149
7.5.2	Results	150
7.6	Results and analysis	151
7.6.1	General user feedback after evaluation	151
7.6.2	Criteria evaluated	151
7.7	Summary	154
8	Conclusions	155
8.1	Summary of this research	155
8.2	Contribution of this thesis	156
8.2.1	Contribution to the Web-based distance education	156
8.2.2	Contributions to open student models	158
8.2.3	Contribution to Human Computer Interaction and Information Visualisation	159
8.3	Future Work	160
	Bibliography	161
	Appendices	174
A	Questionnaire presented in Chapter 4	175
B	Tabulated results of the survey presented in Chapter 4	181
B.1	Countries	181
B.2	Question Q1	183
B.3	Question Q2	184
B.4	Question Q3	185
B.5	Question Q4	186
B.6	Question Q5	187
B.7	Question Q6	188
B.8	Question Q7	189
B.9	Question Q8	190
B.10	Question Q9	191
B.11	Question Q10	194
B.12	Question Q11	195
B.13	Question Q12	196

B.14 Question Q13	197
B.15 Question Q14	198
B.16 Question Q15	199
B.17 Question Q17	200
C DTD description	202
C.1 Students data DTD structure description	203
C.2 Domain data DTD structure description	207
Bibliography	207

List of Figures

2.1	Screenshots from track students window in WebCT (top) and track of content pages visited by a specific student (bottom). Students' names and user ID are blurred for privacy issues.	16
2.2	A screenshot of the interface for instructors in NetCoach. Image from Weber et al. (2001).	18
3.1	Driving directions from Lugano to Pisa provided both in graphical and textual format. Images from http://www.viamichelin.com	22
3.2	The Napoleon's army route in the Russian campaign of 1812 by Charles J. Minard, taken from Tufte (1983).	23
3.3	An 1845 map of London's Soho district plotted by Dr. John Snow showing deaths from Cholera and the location of water pumps. Dots represent cholera cases and X represent water pumps. Image from Tufte (1983).	25
3.4	A comparison chart between the Italian stock market index, MibTel (in blue) and the United States' Dow Jones market index (in red). Taken from http://www.borsanalisi.com	26
3.5	Two examples of scientific visualisations. On the left, representation of the laminar flows on the wing profile (image taken from http://www.esa.int/export/esaMI/High_School/ESAF8BG18ZC_1.html). On the right, Ozone hole over the South Pole during September 2001 (image taken from http://www.esa.int/export/esaSA/ESALRRVTYWC_earth_1.html) . .	30
3.6	Samples of 1D (left), 2D (middle) scatterplots, and histogram . Images from Spence (2001) (left and middle images) and http://www.physics.umass.edu/p151sec2s02/ (right image).	32
3.7	Some examples of 3D scatterplots. Images from Spence (2001) (left), http://www.nhm.ac.uk/entomology/bombus/introduction.html (centre), http://www.kovcomp.co.uk/xlstat/tools/t01.html (right). . .	33
3.8	Single-axis composition technique used by C. J. Minard in the Napoleon's army route in the Russian campaign diagram. Image from Mackinlay (2000).	35

3.9	A city map of the centre of Florence, streets and historical places are encoded with different shapes and colours. Image from http://www.hotelitaliani.it/Mappa/Informazioni/PiantinaFirenze.htm	36
3.10	Diagram representing sunspots from 1880 to 1980 with the sine of the latitude marking a sunspot placement. Micro-macro readings allow combining patterns, details, average and variations with the same image. The lower time-series shows the total area of the sun's surface covered by sunspots by summing over all latitudes. Diagram by David H. Hathaway, Marshall Space Flight Centre, NASA, taken from Tufte (1990).	37
3.11	Small multiples example. The coverage for mobile telephone network in Switzerland between the three operators: Swisscom (on the left), Orange (on the centre), and Sunrise (on right).	38
3.12	A data map example. Cancer mortality in the USA years 1970 - 1994 - white males by county. Image from USA National Cancer Institute http://www3.cancer.gov/atlas/2.html	39
3.13	Two examples of network drawing. On the top, a graph of a small dimension intranet (image from Graphviz drawing software Web site: http://www.research.att.com/sw/tools/graphviz/examples/), on the bottom, a graph of some Internet Web sites (image from TouchGraph drawing software Web site: http://www.touchgraph.com/bi.php?img=greenpeace_new.jpg). 41	
3.14	Zooming, scrolling, and panning operations from Descartes: a tool to support visual exploration of spatial data produced by Dialogis (http://www.dialogis.com). The middle of the right section of the window illustrates a schematic view of the entire territory. The frame shows the size and position of the territory fragment currently displayed in the map area relative to the whole territory. Using the buttons at the top right (north,south, east, west) users can scroll the territory shown in the map window.	42
3.15	The perspective wall invented by Mackinlay et al. (1991). It is a 2D layout wrapped around a 3D structure, upon which label corresponds to files on the computer. Different labels denote different type of entities (folders, text files, wav, ...), located according to precise criteria (date of creation of the file in x-axis and file type on y-axis). Image from http://www.inxight.com	42
3.16	The magic lenses techniques applied to the historical places in a city map of the centre of Florence. Image from http://www.hotelitaliani.it/Mappa/Informazioni/PiantinaFirenze.htm and modified	43
3.17	The QV tool showing a user model. Image taken from Kay (1999).	45

3.18	The visualisation in <i>VIUM</i> showing the currently selected movie. Image from Uther (2001).	46
3.19	A screenshot of ViSMod showing a fragment of a Bayesian student model in the area of biology cell. In ViSMod, both student beliefs and system beliefs have been associated to each of the concepts of the network (e.g. knowledge on <i>Chromosomes</i> is influenced by the student's personal opinion indicated in <i>My-OpinionChromosomes</i> and by the system/instructor's opinion described in <i>SystemChromosomes</i> . The control panel on bottom allows the user to manipulate the view of the model. The image is taken from Zapata-Rivera and Greer (2001a) . .	47
3.20	The main view of a student's progress in E-KERMIT. Progress bars indicate how much the student comprehends each category of the domain. The image is taken from Hartley and Mitrovic (2002).	49
3.21	A screenshot of the open student model visualised by NORMIT. In the progress bars the white part indicates how much of the domain the student has not covered yet, the light grey part indicates how much of the covered part of the domain the student knows correctly, and the dark grey part indicates how much of the covered part of the domain the student knows incorrectly. The image is taken from Mitrovic (2003).	49
3.22	The visualisation of KnowledgeSea (on the left) and WebVIBE (on the right). Image from Willms (2003).	50
3.23	The Bull's and Nghiem open learner model presented in two forms: tabular (top-left window) and graphical (bottom-left window). Image from Bull and Nghiem (2002).	51
3.24	The comparison of two individual student models in the Bull's and Nghiem open learner model. Image from Bull and Nghiem (2002).	52
3.25	The graphical display of student understanding displayed by Chaatra. Image from Rehani and Sasikumar (2002).	53
3.26	The communication graph of the e-mail exchanged within groups in Simuligne. Image from Reffay and Chanier (2002).	54
3.27	The PeopleGarden visual representations of participation on a message board. Image from (Xiong and Donath, 1999).	55
4.1	Type of involvement in courses for participants in the survey	61
4.2	Number of courses managed by participants of the survey	62
4.3	Number of students in the largest class	62
4.4	Educational platforms and their facilities used by instructors	63
4.5	Assessment techniques and tool used in courses	63

4.6	Information about students instructors would be interested to have	64
4.7	Interest in time-dependent information	65
4.8	Interest in information related to concepts of the domain of the course	66
4.9	Type of interesting students to monitor	66
4.10	How instructors want to use students' information	67
5.1	The CourseVis system architecture	76
5.2	The primary elements of the student data structure generated by the CourseVis Student Data Exporter.	78
5.3	Course DTD elements in the student data structure generated by the CourseVis Student Data Exporter.	79
5.4	Student DTD elements in the student data structure generated by the CourseVis Student Data Exporter.	80
5.5	Discussion DTD elements in the student data structure generated by the CourseVis Student Data Exporter.	81
5.6	Domain designer - user interface for the definition of the concepts of the course .	83
5.7	Domain designer in CourseVis - user interface for the definition of pages and quiz questions related to a concept. The instructor selects the corresponding pages for a concept from the window in the right.	84
5.8	The domain data structure DTD generated by the CourseVis Domain Designer . .	85
5.9	Visualisation pipeline followed in CourseVis, derived from the <i>reference model for visualisation</i> proposed by Card et al. (1999).	86
5.10	Example of the OpenDX visual program editor and the image produced.	88
5.11	Some examples of interactors used in OpenDX to change the visual image produced by a program.	89
5.12	The OpenDX visual program for the visualisation example discussed in this section.	95
5.13	OpenDX visualisation produced for the example discussed in this section.	95
5.14	Visualisation produced for a course with 231 students enrolled.	96
5.15	Zooming and panning applied to the representation produced	97
6.1	Home Page of the Java on-line course	102
6.2	The discussions plot, represents discussion threads made using the discussion board tool in CMS. The diagram shows a ball for each thread, positioned on the student who started the thread, the date when the thread started, and the topic of the discussion. The size and the colour of the ball are proportional to the number of follow-ups.	105
6.3	Results of two rotations of a discussions plot, top image focuses on students, bottom image focuses on the topics.	106

6.4	The discussion graph in CourseVis.	108
6.5	The cognitive matrix, the colour in each cell represents performance level of a student on a specific concept. A blank (uncoloured) cell means that student has not done the relative quizzes yet.	110
6.6	The cognitive graph, for each concept of the course is measured the mean of performance values achieved by students on evaluation profs.	112
6.7	The students accesses plot, each circle means that a student logged into the course on the corresponding date. The histogram reports the number of global accesses made by students on each date.	114
6.8	The student behaviour graph. This representation include dense information about a single student using the composition technique: access to the content pages, global accesses to the course, progress with the schedule of the course, messages read, posted and replied to discussions, and submission of quizzes and assignments.	115
6.9	Three cognitive matrices representing students' level of knowledge on concepts in three different periods of the course. On left is reported a matrix corresponding to the student performance at about 3 weeks after the course started, matrix on the centre at about 7 weeks after the start and matrix on the right at the end of the course, at about 10 weeks after the beginning.	118
6.10	Cognitive matrices for the 2001 edition (on the left) and the 2002 edition (on the right) of the Java on-line course.	119
6.11	Discussion plot for the 2001 edition (top) and 2002 edition (bottom) of the Java on-line course. The 3D scatterplots are rotated to focus on students.	120
6.12	The students' accesses plot for the course Verbal Semiotics, having 231 students enrolled	122
6.13	The discussion plot for the course Verbal Semiotics.	123
6.14	The student behaviour for the student <i>michele</i> . The representation include a grid for a better location of objects in diagram.	124
6.15	The student behaviour for the student <i>sebastiano</i>	126
6.16	The student behaviour for the student <i>salvo</i>	127
7.1	Participants at the experimental study were split in two groups. Group A was assigned to a course provided by a standard WebCT installation, group B was assigned to a course provided by a WebCT installation integrated with the CourseVis graphical representations.	145
8.1	The process of generation of useful insights on instructor's mind with CourseVis	159
B.1	Submits by countries	182

B.2	Number of courses managed by instructors	183
B.3	Type of involvement in courses	184
B.4	Level of instruction of courses	185
B.5	Number of students in the largest class	186
B.6	length of involvement in on line courses (years)	187
B.7	Educational platforms used by instructors	188
B.8	Facilities used in courses	189
B.9	Assessment techniques and tool used in courses	190
B.10	Interesting information about students	192
B.11	Interesting information about students (summarised version)	193
B.12	Interest in time-dependent information	195
B.13	Interest in information related with concepts of the domain of the course	196
B.14	How instructors want to use students' information	197
B.15	Type of interesting students to monitor	198
B.16	Interest in different typology of students' interaction	199
C.1	Student data structure	203
C.2	Domain data structure	207

List of Tables

3.1	Specific techniques for Information Visualisation identified by Card et al. (1999)	31
5.1	Mapping of the XML into relational tablewhichs as part of the architecture of CourseVis. Primary keys and foreign keys are omitted for simplicity.	92
5.2	Data table produced by the Data transformation process for the Java programming course described in Section 6.1.	93
5.3	The “visual vocabulary” defined by Card and Mackinlay (1997) used to formally describe the design space of representations in Information Visualisation and followed in CourseVis.	100

Conventions

In this thesis, we will make use of the following notational conventions:

Example	Explanation of the conventions
NLG	Acronyms are indicated in small capitals.
typewrite	Typewriter font is used for fragments of program code and for the content of files.
<i>italic</i>	Italics are used for emphasis in the body of the text, or to denote a specific part of a articulate object, or to denote a direct quote.

We also assume the following:

- The meaning of *learner* and *student* are assumed equal. The meaning of *teacher*, *instructor*, *facilitator* are assumed equal.
- Throughout this thesis we will use *male* gender only for pure convenience, the use of *his* should be taken to mean *his* or *her*, and the use of *he* should be taken to mean *he* or *she*.
- In the exposition, *we* refers to the *author* and *our* refers to *author's*.
- All URL addresses cited in this thesis were visited between August 1st 2003 and December 20th 2003.

Abbreviations

Following abbreviations were used throughout the thesis:

Term	Description
AI	Artificial Intelligence
AIES	Adaptive and Intelligent Educational Systems
CAL	Computer Assisted Learning
CMC	Computer Mediated Communication
CMS	Course Management System
DBMS	Database Management Systems
DTD	Document Type Definition
HCI	Human Computer Interaction
ITS	Intelligent Tutoring System
IV	Information Visualisation
WBDE	Web-Based Distance Education
XML	eXtensible Markup Language

In this thesis, abbreviations are used both for the singular and the plural. E.g. ITS is used as abbreviation for “intelligent tutoring system” and “intelligent tutoring systems”.

List of published papers

Some parts of the work presented in this thesis have been published in the following articles:

Some ideas of **chapters 2** and **3** are presented in:

- Federico Flueckiger, Riccardo Mazza, Carlo Lepori. Virtual Campus in Southern Switzerland, multilingual open distance learning with strong local roots and closed relationships to other universities. *NETTIES2001 - 7th International Netties Conference/ 3rd International Conference on New Learning Technologies*. Fribourg, 2001 September 13-15. pp. 2.7.1 - 2.7.8.
- Riccardo Mazza. Using Open Student Models in Distance Learning Environments to help teachers provide adaptive tutoring. *Studies in Communication Sciences*. Special Issue New Media in Education. 2003. pp. 245 - 251

The results of the survey described in **chapter 4** have been included in:

- Riccardo Mazza, Vania Dimitrova. Informing the Design of a Course Data Visualisator: an Empirical Study. *Proceedings of the 5th International Conference on New Educational Environments*. Lucerne. 2003 May 26-28. pp. 215 - 220

Early ideas of **chapter 5** have been summarised in the following papers:

- Riccardo Mazza. Diagnosing the state of the student's knowledge in a web-based learning environment. *Proceedings of the 4th International Conference on New Educational Environments*. Lugano, 2002 May 8-11, pp. 3.1/15 - 3.1/18.
- Riccardo Mazza. External Representation of the Student's Knowledge in a web-based Learning Environment. *International Conference on Intelligent Tutoring Systems, Young Researchers Track*. Biarritz - France, 2002, June 5-7, pp. 29 - 31

Some of the graphical representations discussed in **chapter 6** have been presented in:

- Riccardo Mazza, Vania Dimitrova. CourseVis: Externalising Student Information to Facilitate Instructors in Distance Learning. In J. Kay, H. U. Hoppe and M. F. Verdejo (editors) *Proceedings of Artificial Intelligence in Education: shaping the future of learning through intelligent technologies - 11th International Conference on Artificial Intelligence in Education (AIED 2003)*, Sydney, Australia, 2003 July 20-24. IOS Press. pp. 279 - 286.
- Riccardo Mazza, Vania Dimitrova. Visualising Student Tracking Data to Support Instructors in Web-Based Distance Education. *13th International World Wide Web Conference (WWW 2004) - Educational Track*. 2004 May 17-22, New York. **Best alternate track paper award.**

The CourseVis system and its graphical representations will be presented in a poster-demonstration session at the ED-MEDIA 2004: World Conference on Educational Multimedia, Hypermedia & Telecommunications to be held in Lugano, Switzerland, 2004 June 21-26.

The results of this work will be implemented also in the research project “*EDUKALIBRE, Libre software methods for E-Education*” founded by the European Union SOCRATES programme in the years 2003 - 2005 (Project N. 110330-CP-1-2003-1-ES-MINERVA-M). Participating institutions are: Universidad Rey Juan Carlos, Madrid (ES) - coordinator, School of Engineering of the University of Porto, Porto (PT), University of Karlsruhe (TH), Institute for Industrial Production / French-German Institute for Environmental Research, Karlsruhe (DE), Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Prague (CZ), University of Leeds, Leeds (UK), and the University of Lugano.

To Beatrice, Vincenzo and Giulia

Chapter 1

Introduction

Tutoring a distance on-line course is often a difficult task. Recently instructors are acknowledging that the competencies necessary to manage a course asynchronously can be different from those required in a regular classroom (Bates, 1999). Distance education led to changing the role of the teacher from “lecturer” into “facilitator” of learning. This new role requires new competencies and abilities in order to support distance students effectively. Activities like answering questions, monitoring and promotion of discussions, monitoring the learners’ progress and testing the acquired knowledge and skills on a regular basis are essential for a good on-line tutoring practice (Helic et al., 2000; Cotton, 1988; Ragan, 1998). Computing technologies, such as Course Management Systems (CMS), have been developed to support the design and delivery of distance learning courses over the Internet. Although generic CMS are very effective in facilitating the delivery of distance courses, they provide very little help to instructors to gain understanding of cognitive and social processes in distance classes. For example, questions like: “*what level of mastery on concept C has the student S achieved*”, or “*did student S access the whole course materials, and when*” or “*has student S participated to discussions regularly*” are difficult to answer appropriately using conventional CMS.

A vast amount of data is collected by the CMS to keep a thorough record of the students’ activities in a distance course, but this is usually presented in a complex tabular format, often incomprehensible and difficult to use by instructors. Even worse, commonly there is no relationship between concepts taught in a course and pages or quizzes created. As a result, instructors know very little (sometimes even nothing) about what is happening in their distance classes. Instructors “*may become confused or incapable of monitoring the individual learning needs or the misconceptions of the distance students*” (Tsinakos and Margaritis, 2000).

A field of Artificial Intelligence (AI) applies AI techniques to educational applications to enable adaptivity. Knowledge-based educational systems, like Intelligent Tutoring Systems (ITS), have been studied and proposed since the eighties, but the real explosion of the use of computing

technologies in education appeared only during the late nineties with the coming of the World Wide Web. CMS populated the universities and, until now, they are the main on-line course delivery systems through the Internet.

Some researchers tried to apply AI techniques to educational applications to tackle the problem of monitoring learners and supporting instructors during the learning process. Well established techniques, like Student Models (Self, 1994; Greer, 1996) try to acquire information about the learner and create a “*computer system’s belief about the learner’s knowledge*” (Stauffer, 1996). Student models use probabilistic theories like Bayesian networks (Villano, 1992; Murray, 1998, 1999; Mayo and Mitrovic, 2000) or Fuzzy Logic (Gisolfi et al., 1992; Hawkes and Derry, 1996) to remedy misconceptions, generate feedback, make systems adaptable to individual learners and represent the learners’ level of mastery of a specific concept or topic. In these systems, learners facts, like the level of knowledge on concepts, the learning style, or the attendance to the course are computed by AI algorithms, and need a thorough expert model to embed the necessary knowledge about the domain. They could be used to generate appropriate advices to instructors about students who need particular attention, as made in Kosba et al. (2003). Student models can also be graphically rendered and made available to instructors for inspection and tuning. This case is referred to as *open (or scrutable) student models* and can be used by instructors to adapt their teaching to the individual (Kay, 1995; Zapata-Rivera and Greer, 2001a; Bull, 1997) or to better understand their students and help individuals with particular problems (Bull and Nghiem, 2002). However, information generated by student models is inferred by AI algorithms, and instructors have to rely on their accuracy when using it.

An alternative approach is proposed in this thesis where we explore the use of Information Visualisation (IV) (Card et al., 1999; Spence, 2001; Tufte, 1990) to render the data collected from a CMS into an appropriate graphical manner. By applying IV techniques, data from CMS can be presented in a format suitable for the instructors to gain understanding of their learners and become aware of what is happening in their classes. Unlike student models, students’ data gathered from the CMS is not processed by AI algorithms, but is processed and used to generate graphical representations which may reveal facts, trends or situations with students. With the representations produced by IV techniques, information such as “*student S needs help because he is performing very badly at evaluation proofs*” are not suggested by a computer program, but can be inferred in the mind by the instructors looking at graphical representations of the students’ performance on quizzes. Instructors may take advantage of their knowledge of the domain of the course to infer reliable information about the progress of students or to improve specific parts of the course that may be problematic for students.

This thesis proposes a systematic research on using visualisation techniques to externalise data gathered by a generic CMS, and intends to propose appropriate graphical representations taking into consideration instructors’ needs. This is accomplished by performing the following tasks:

- An empirical study that investigates the instructors' needs when teaching at distance using CMS. This study is accomplished with a survey submitted to a sample of instructors. The results of the study elicit recommendations for the implementation of graphical representations aimed at facilitating distance teachers.
- Design and construct a software system that works in conjunction with a conventional CMS. This software produces a series of pictorial representations to help instructors to gain understanding of social, cognitive and behavioural aspects of distance students. The graphical representations are produced taking into consideration the requirements gathered with the survey.
- Validate the effectiveness of the proposed visualisations with a sample of instructors using group interviews (focus group), experimental study and semi-structured interviews.

This work produced a course visualisation tool (*CourseVis*) which extracts tracking data from an on-line course maintained with a CMS, transforms the data into a form convenient for processing, and generates graphical representations that can be explored by course instructors. A set of pictorial representations to depict information required by distance teachers have been produced. These can be explored by course instructors to examine social, cognitive and behavioural aspects of distance students.

The advantages of *CourseVis* are:

- Instructors and students continue to use the CMS in the usual manner. *CourseVis* has been built as a tool that runs in conjunction with a commercial CMS. This enables the institution running a distance course (university, school, ...) to avoid new investments, training in learning a new learning environment, and to avoid transferring courses to a new platform.
- The representations produced comply with the instructor's requirements, following a user study with instructors of on-line courses.
- The tool can be easily adopted to support other CMS, as a software layer has been implemented to give a universal interface with a generic CMS.

The results of the work give a number of original contributions to:

1. Web-based distance education, in particular the issue of student monitoring and tracking, and facilitating instructors.
2. Information visualisation, for the application of visualisation techniques into an educational domain.

3. Artificial intelligence in education, in particular the use of visualisations techniques in open student modelling.

This thesis consists of eight chapters. In Chapter 2 we will give an introduction to Web-based distance education and the Web-based course management systems. The aim of this chapter is to place the research reported here within the relevant context. Some problems concerning CMS that we can find in literature (isolation, frustration, lack of motivation, etc.) are presented. We argue that these problems can be in part avoided if instructors have a better understanding of what is happening to their students and intervene accordingly. In particular, we illustrate why CMS are particularly problematic in monitoring the students' activities and their level of knowledge acquired during the running of the course. Difficulties in using the data provided by CMS for the tutoring activities will be pointed out. We will then propose the use of graphical representations of data collected by CMS to help instructors identify the use of information needed for they tutoring activities.

In Chapter 3 we will describe some techniques of Information Visualisation that can be helpful for representing student tracking data from CMS. Educational systems that exploit some graphical representations of students' data will be analysed as well.

The first step in developing usable systems is gathering design requirements. Chapter 4 will analyse the instructors' requirements based on a survey we have conducted. The results are illustrated starting from an analysis of responses grouped into 3 themes: social, cognitive and behavioural aspects. Based on the results of the study, we will draw recommendations in the development of a tool to help teachers gain a better understanding of cognitive, social and behavioural aspects about individual students in on-line distance teaching. The recommendations are used for the design and implementation of CourseVis.

Chapter 5 will describe the architecture and will provide details of the implementation of CourseVis, a Course graphical Visualisation tool which acquires tracking data from a CMS and generates graphical representations that can be explored and manipulated by instructors to examine specific aspects of distance students.

Chapter 6 will illustrate a set of graphical representations created with CourseVis to support instructors in their teaching activities. Data from a course on the programming in Java given in 2002 will be used to illustrate some examples of the proposed representations. We will demonstrate how CourseVis can be employed in different domains; to compare different editions of the same course; and to promote instructors' reflection.

To examine whether the proposed representations comply with the instructors' requirements, in Chapter 7 we will present an empirical evaluation with a sample of potential users of CourseVis. Several instructors participated in a focus group, and some other instructors participated in an experimental study with sessions of CourseVis. The evaluation process will demonstrate that instructors can take advantage of graphical representations in CourseVis in their tutoring activities.

The evaluation will reveal also some possible improvements of CourseVis and how they can be addressed in future work.

Chapter 8 will summarise the work done in this research and will draw our conclusions. The main contributions of the work will be described, and possible extensions for future work will be outlined.

Chapter 2

Web-Based Distance Education

Distance learning is the most significant phenomenon occurring in higher education today. Everywhere one looks, whether in community colleges, 4-year institutions, Ivy League colleges, research institutions, or technical colleges, distance education is on the rise, and the rise is occurring at a rapid pace. Distance education and technology are major factors in the contribution to current and expected changes in the postsecondary education enterprise. (Judith, 2003).

This chapter presents an introduction to Web-based distance education and a class of tools that enable the Web-based distance education: the Course Management Systems (CMS). Here we present some problems concerning CMS that we can find in literature (e.g. isolation, frustration, and lack of motivation), possibly caused by the physical distance between the teacher and the learners. We argue that these problems could be, to some extent, avoided if the instructors have a better understanding of what is happening with their students (are they reading materials; are they participating in discussions; what is their performance in exercises; etc.) and intervene accordingly. Instructors should be provided with appropriate means to diagnose when a problem might arise, or has arisen, so that they can take immediate actions to prevent or overcome that problem.

2.1 The evolution of education

Education is known as “*the process, either formal or informal, of acquiring knowledge and skills, leading to the development of understanding, attitudes and values.*”¹. From ancient Egypt to the modern schools, education played a key role in the progress of the society. The globalisation of economy, the evolution of communications, as well as the rapidly changing technologies with increasing mobility on workplace demand constantly new qualifications from any member of society. Education is now not only an activity made in schools or universities, but also in companies, organisations, prisons, etc. Companies want employees to learn a new machinery, the free-lance consultant wants to learn the latest innovations so that he can do the job better and avoid becoming obsolete in few years. Education is available not only at school age, but it appears in the whole cycle of the life of humans. Education will be the only possibility for citizens to adapt their abilities and their knowledge to the current developments in society. In this way, the term “*lifelong learning*” becomes a keyword of postmodernism - a success factor in the labour market of the starting millennium. (Flueckiger and Mazza, 2001).

Today we are confronted with a new situation. The vast expansion of knowledge and the use of the Internet is changing the way we are learning. For example, if one would like to find the newest developments on a topic, one starts surfing the Internet, sends e-mails to make links, participates in Internet discussions, and so forth. People use learning technologies and learning strategies to reach a vast and more appropriate amount of knowledge. The ability to handle the new means and methods of electronic communication becomes the way pointing factor on the path to the tree of finding. The role of the lecturer is changing too: he no longer pursues a pure teacher-directed education, but he more and more undertakes the role of a tutor that accompanies the student within his finding process - he becomes the “*Primus inter Pares*” in a virtual class (Flueckiger and Mazza, 2001).

Even though technology enables new potential to the delivery of the education, as Berge and Collins (1995) point out, old technologies are augmented, but not totally replaced. Therefore, it is crucial to provide a similarly fertile atmosphere in the newly developed virtual environments. In this way the learning with new educational technologies will not become a troublesome duty but a joy and will stimulate individual successes (Flueckiger and Mazza, 2001).

2.2 An introduction to Web-Based Distance Education

Today new needs are becoming apparent in the process of learning. While classroom lessons still play a critical role in our society, new forms of means and media by which the instruction is

¹World Encyclopedia. George Philip Ltd, 2003. Oxford Reference On-line. Oxford University Press. 9 October 2003.

conveyed are appearing. Education must be “*available anytime and anywhere*”: it could seem a TV’s commercial, but this is what the new era education is moving to, especially in companies. Lessons available 24 hours 7 days per week, with delivery to home, office, hotel, available in remote and inaccessible locations are demanded more and more (Rosenberg, 2001).

For years the new media in general, and the computer with its features in particular, are suitable for using in teaching. Computer-assisted learning (CAL) is a discipline appeared in seventies that uses computers to aid the education and training of people. It is also referred to with other terms such as computer-aided instruction (CAI), computer-based learning (CBL), computer-based training (CBT) and computer-managed instruction (CMI). CAL can test attainment at any point, provide faster or slower routes through the material for people of different aptitudes, and can maintain a progress record for the instructor. Computers were used to produce lessons based on student history and test performance. Components comprising each lesson unit are selected in accordance with learners’ needs, as indicated by test performance on previous units, academic history, etc.

Significant effort was put on CAL in seventies and eighties, but technology changed as well during the years. CAL programs revealed their limitations in growing instability of content, development cost and time, and lack of awareness of modern instructional design approaches (Rosenberg, 2001).

In nineties, the large diffusion of computers and telecommunication networks opened the Computer-mediated communication (CMC) technologies as an alternative modality of education delivery, particularly with the advent of the Web (Cannone-Syrcos and Syrcos, 2000). Those technologies use computer networks for exchanging of information between people, even in remote locations. CMC tools include e-mail, WWW, newsgroups, groupware, chat rooms, and Internet-based audio and video applications. There were other technologies that allow the delivery of education at distance, e.g. videotaped lectures have been used in universities and professional courses for the last two decades (Valentine, 2002), while television and radio have extensively been used to distribute education to the learners’ house as well². But the real novelty brought by CMC was that this information in the format of text, images, audio, and video can be exchanged in both directions in real time (or *synchronous*) communication, which means that people are communicating with each other at the same time, or can be exchanged at different times (or *asynchronous*) and enabled sharing of information. The availability of Internet to everyone (not only at universities and large companies) and powerful computers at low cost “*have opened the way to formats other than pen-and-paper correspondence courses and allow for a more interactive, integrated learning environments ... CMC and networking in general can promote long-distance collaboration among students and content specialists in many different areas. The integrated use of technology*

²We can cite some examples of higher education delivered by Italian television in Garito (1999), and the English language teaching programmes on the radio delivered by BBC (<http://www.bbc.co.uk/worldservice/learningenglish/index.shtm>).

offers many educational opportunities and possibilities when driven by sound visions of learning.” (Berge and Collins, 1995).

Educators around the world soon found new opportunities for learners through CMC. Cannone-Syrcos and Syrcos (2000, p. 174) indicate the “*ability to liberate educational instruction from the constraints of distance and time*” as the greatest benefit of CMC in education. Cannone-Syrcos and Syrcos (2000) indicate also further advantages that might derive from the use of CMC for the education, such as interactivity and collaborative learning (Dillenbourg, 1999). CMC is not a technology that suits all types of learning, courses or learners, but it is useful only in a certain learning context. For instance, Bates (1995) illustrates some situations where CMC is suitable, particularly when there is a need for information transmission and discussions. Educators, psychologists and technologists started to study and apply a new multidisciplinary field that can be referred to as “*Web-based distance education*” (WBDE). This term can be applied to the modern concept of distance learning, defined by Greenberg (1998) as “*a planned teaching/learning experience that uses a wide spectrum of technologies to reach learners at a distance*”, where the technologies used are those provided by CMC tools over the Internet.

In the literature, a plethora of terms related to the delivery of the education through the Internet can be found: *computer-based distance education*, *Internet-based training*, *On-line learning*, *Web-based learning*, *Web-based distance education*, *Web-based instruction*. For the sake of simplicity and consistency we will use the term *Web-based distance education* through the thesis.

2.3 Benefits of Web-based Distance Education

The literature shows that Web-based distance education may bring several advantages in the process of education (Vidyakash, 2000; Barron, 1999; Cantoni, 2001; Wegner et al., 1999; Arsham, 1995; Galusha, 1997):

- **Flexibility of time and place:** learners are not forced to attend a classroom in a geographical location, or to follow time constraints. Learners may attend on-line courses whenever they wish: in the night, during the weekend, at work.
- **Interactivity:** the educational activity is not only listening to a lecture, but can offer an increased interaction with (or between) the students. Some type of learners may benefit from this form of communication. For instance, introverted students who usually don't ask questions in classrooms because they are too shy may be more encouraged to exchange e-mails or discussions.
- **Accessibility:** the Internet medium gives a great potential to the delivery of education: rural

zones may benefit in having qualified instruction³, disabled person can follow courses without travel problems, students on business may follow courses given by the most important overseas teacher on economics.

- **Self paced study:** the process of learning can be paced and controlled by the learner. If a student grasps a part of the content, he can go on with the course without waiting for other peers to understand too (as it may happen in a regular classroom), while if a student finds a part hard to understand, he may take the course at a slower pace, perhaps asking support from peers or from the instructor.
- **Multimedia content:** course materials can be enriched with audio, video, images, interactive animations ... If the content of a course is given in different formats at the same time (for instance, a concept may be explained with a plain text paragraph, or it can be illustrated with an animation, or it can even be illustrated with an audio/video picture with the teacher who explains the concept), the learner can select from which media to study, depending on his personal learning style.
- **Convenience:** under some conditions, an on-line course may be cheaper than a face-to-face course. In the short period costs are higher because the preparation and the implementation of a course are heavy, but in a long period (usually several years) this cost is paid off by the low costs of maintaining the course. Also for the learners, not having to travel is quite convenient.

2.4 Course Management Systems (CMS)

The use of CMC for education on the Internet started at universities in the late eighties. At that time the Internet was used mainly to exchange documents and e-mails between students and between students and teachers, and it was limited to a small part of technologically advanced departments. The invention of the World Wide Web and the development of Web clients (the browsers) (Berners-Lee et al., 1992) revolutionised education in schools, organisations and universities. Teachers produced home pages with information and materials of their courses. There were also teachers who started to create first newsgroups, interactive exercises, simulations, on-line notes, student evaluations, glossary and bibliography with Internet (Goldberg, 1996). The World Wide Web provided new opportunities for distance education to create a virtual classroom to bring together a community of learners for interactive education.

³With regard of this we can mention the “Progetto Poschiavo” of the “Istituto Svizzero di Pedagogia per la Formazione Professionale” (<http://www.ispfp.ch>) that brought high level education courses in Val Poschiavo e Val Bregaglia (two remote regions on southern Switzerland distinguished for their difficulties in transports with the rest of the Switzerland due to the morphology of Alps) from 1996 to 2001, with a good success and important benefits for the population of these regions.

Some researchers in middle nineties perceived the benefits of using CMC for educational purposes, and started to develop specific tools dedicated to the delivery of courses through the Internet that make extensive use of CMC tools. First experiences on integrated environments were made at universities in the USA and Canada, which lead to the development of *Course Management Systems* (CMS). These are software environments that take advantage of the client-server communication on the Web to support open and distance learning (Avgeriou et al., 2003). Using these environments instructors can distribute information to students, produce content material, engage in discussions, prepare assignments and tests, and manage distance classes (McCormack and Jones, 1997). From a technical point of view, the core of the application is a Web server (the same used on common Web sites) upon which several tools are built to provide services to learners and instructors.

2.4.1 Educational components of a Course Management System

A CMS is a tool that provides a comprehensive set of functionalities that can be used by instructors and learners to perform education related tasks. The use of these functionalities depends on the instructor's teaching strategy and didactic objectives. A CMS can be used to supplement a course taught in ex-cathedra modality (for instance a teacher may use it to engage discussions on issues of the course) or to create a course completely on-line. In both situations, CMS provide basic functionalities that, according to Brusilovsky and Miller (2001) and Vidyakash (2000), can be structured around four main components:

- **Presentation:** it comprises all functions related to the delivery of the course materials. Content may be in the form of text, images, audio, video, etc. They are usually stored in open formats suitable to be read with a Web browser (e.g. HTML pages, GIF, JPEG and MPEG files).
- **Activities:** it comprises activities that involve students in learning. Typical activities are assessment of student progress through quizzes, tests and assignments. Assessments serve both the instructor to assess the students' understanding of the materials or compare performance of students in the class, and the students to evaluate their understanding of the subject. Some tools for assessment, such as the multiple choice questions, can provide automated evaluation and an automated feedback to students, while others, such as essay type questions, require manual grading. Students may work in groups on joint assignments.
- **Communication:** it comprises all forms of interactions between instructor and students, or between students. Interactions can be synchronous (all participants are communicating at the same time, such as in a telephone conversation) or asynchronous (participants are contributing to the communication at different times). Communication is essential, especially

in courses given mainly on-line, because “*it is an important way that teachers and students diagnose and remedy problems*” (Brusilovsky and Miller, 2001, p. 168). The most common interaction tools used in CMS are:

- *chat*: allows participants to synchronously communicate with each other in real time. A text box visualises the conversations and every user may write a sentence on the text box that will be seen by the others. A display of participants currently participating on chat is available.
- *whiteboard*: it is a drawing board shared between members during an on-line discussion. It allows everyone to synchronously draw geometric figures such as polygons, circles, but also text and pictures. It can be useful for drawing diagrams and allows anyone to contribute in the drawing.
- *e-mail*: it is the common e-mail tool that allows participants in an on-line course to asynchronously exchange private messages, and to include files in attachments when needed.
- *discussion*: it is an asynchronous type of communication media that enables course members to post and read messages. Discussions are usually divided into different topics that identify the subjects upon which messages are grouped. Topics can be public or private. Private topics were accessible only to a restricted set of students decided by the instructor, and can be useful to set up a discussion environment for a group assignment, while public topics are accessible to everyone in the course. Discussion is an important communication tool that allows students who have problems to ask the instructor or other students and they may receive feedback, for instance with additional explanation or suggestions by others.
- **Administration**: it comprises students’ administration activities such as restricting the access to part of the course, payment and course cancellation. This is performed by administrative personnel. The instructor could also have some class management functionalities, such as adding new students to his course, modifying their details, assigning grades, monitoring performance and tracking student activities.

2.4.2 Available Course Management Systems

The interest of researchers in education technologies is recently focusing on the Web as the medium of interaction and learning (Vidyakash, 2000). For this reason, in the last five years a plethora of CMS tools have appeared. CMS can be divided in two categories according to their origin (Brusilovsky and Miller, 2001):

- **University systems:** are products developed by universities following some research projects. Some of these systems are advanced prototypes implemented to demonstrate innovative ideas and distribute (or giving license of use) for free. An example of these is ARIADNE (Forte et al., 1996). Others are university-supported products implemented from their home universities having less technological features, but in return being well-developed and tested. Examples are Mallard from University of Illinois at Urbana-Champaign (Graham et al., 1997) and Virtual-U from Simon Fraser University (Fisher et al., 1997).
- **Commercial products:** are products commercially available from private companies. Some of them were originally developed at universities and then exploited as spin-off commercial companies. This is the case with WebCT, one of the most popular CMS in universities all around the world, originally developed by the University of British Columbia in Canada in 1995 (Goldberg et al., 1996) and then recently shipped by a company (WebCT, 2003), while the University is continuing the development of this tool on an industrial basis. Others CMS tools are produced exclusively by commercial companies, among these are the BlackBoard Learning System (BlackBoard, 2003), Lotus Learning Space (LearningSpace, 2003) and TopClass (TopClass, 2003).

We have to acknowledge that a number of CMS that follow the open source and free software initiative (Stallman et al., 2002) are being developed and maintained by a group of voluntary during last years. Some of them have reached a good level of maturity and several institutions are adopting them as alternative platforms to the commercial ones, which end-user license are usually very expensive. Among these the Moodle software (Moodle, 2003) is having a discrete success in the academic community.

2.5 Problem analysis

CMS have the great merit of having introduced interactive Web-based distance learning in places where teaching activities are carried out for centuries in forms of lectures from the Aristotle's age, and bring new advantages to learners and teachers. However, current research reveals that education delivered through CMS suffers from some problems. Some researchers (Smith-Gratto, 1999; Galusha, 1997; Rivera and Rice, 2002; Valentine, 2002; Hara and Kling, 2000) list a number of problems encountered by students studying on-line courses, including the students' feeling of isolation due to lack of contact with the instructor, disorientation in the course hyperspace, loss of student motivation, and lack of institutional support (including help with technical problems). This is in part due to the nature of computer mediated communication which lacks some specific modalities of interaction like gesture, face expression, direct dialogue, etc. which are considered

significant reasons for the higher drop-out rates of Web-based courses than regular classes (Smith-Gratto, 1999).

To help instructors to overcome these problems some researchers (Smith-Gratto, 1999; Galusha, 1997) indicated as a possible solution the monitoring of student learning and providing learners with appropriate and prompt feedback.

Smith-Gratto (1999) recommend that to have quality Web-based instruction “*instructors should evaluate student products and provide corrective information that allows the students to learn where their understanding of the material is faulty*”. Galusha (1997) indicates a critical success factor for the distance student as “*it is important that the student receive prompt feedback in any institutional setting, particularly in distance learning where the learner is impaired by the lack of casual contact with the teacher and other students*”.

Educational research literature shows how monitoring student learning is crucial in lessons given in classrooms. Cotton (1988) reports on this issue: “*the practice of monitoring student learning is an essential component of high-quality education. The careful monitoring of student progress is shown in the literature to be one of the major factors differentiating effective schools and teachers from ineffective ones.*” The same criteria are applicable to on-line courses as well. In fact, Helic et al. (2000) suggest: “*a good tutoring practice requires monitoring of a learner’s progress with material and testing of acquired knowledge and skills on regular basis*”. Also Ragan (1998) states that: “*In a distance education model, assessment and measurement become even more critical in the absence of the face-to-face interactions which enable teachers to gauge student response, feedback, and progress toward goals*”.

This rises from the previous considerations that the effective use of CMS in distance education requires that instructors should be provided with appropriate means to diagnose that a problem might arise or has arisen, so that they can take immediate actions to prevent or overcome that problem.

Vidyakash (2000) reports on this issue: “*since the instructor and students are not in the traditional face-to-face environment there must be some extra provision through which the students taking the course and their performance can be tracked*”. This has been done in eighties in courses that utilise print-based material packages and quizzes delivered via mail and students are supported by telephone. For example, TRIX (Holt et al., 1987) is a system that keeps track of students and calculates potential problem indicators upon several criteria, such as if the student is failing behind schedule more than X days, or the student has been in the course for X days and completed no assignments. TRIX produces timely a variety of problem reports for tutors, coordinators and administration for a better course and student management. More recently, Rehani and Sasikumar (2002) implemented Chaatra: a student monitoring and learner modelling system integrated into a commercial CMS to give instructors qualitative information about the students’ overall understanding of the subject, if a student lags in some particular topic, what kind of modification to the

course materials should be carried out, etc. Kosba et al. (2003) apply framework based on fuzzy logic called TADV on a CMS to provide teachers with appropriate advice to help them manage their distance courses. Monitoring student performance in on-line courses has been reported also in Pappas et al. (2001) and Thomas and Paine (2002).

While CMS are supposed to help tutors to monitor students' activities and performance, they often provide complex, confusing, and difficult to use information. In fact, the interaction mediated by the digital media makes it difficult for the instructor to verify elements essential in didactics, e.g. to understand what part of the course an individual student or a group of students are working on, or the level of mastery achieved by a student for specific concepts of the course.

CMS's administration components accumulate large log data of the students' activities in a distance course and usually have built-in student monitoring features that enable the instructor to view some statistical data, such as a student's first and last login, the history of pages visited, the number of messages the student has read and posted in discussions, marks achieved in quizzes and assignments, etc.

Instructors may use this information to monitor students' progress and to identify potential problems. However, tracking data is usually provided in a tabular format, is often incomprehensible, with a poor logical organisation, and is difficult to follow. As a result, tracking data is very rarely used by distance learning instructors. See Figure 2.1, top for an example of students track page provided by WebCT and Figure 2.1, bottom for the tracking of content pages visited by a specific student. Moreover, CMS do not provide information about the actual learning that is taking place (e.g. the level of understanding achieved by a student on a particular concept, some indication of which concepts from the course material the students face difficulties with), albeit such information is very important for instructors in distance learning.

Advanced methods for monitoring student activities have been employed in Intelligent Tutoring Systems (ITS) (Holt et al., 1994; Barr and Feigenbaum, 1982; Brusilovsky, 1994; Ragnemalm, 1995; Tsinakos and Margaritis, 2000; Stauffer, 1996; Brown and King, 1998; King, 1998). An ITS is a computer-based instructional system that provides individualised tutoring thanks to a detailed model of the knowledge domain, a student model, and a pedagogical model. The purpose of the *student model* is to describe the student in some manner and represent the student's understanding of material to be taught. It is used to guide decisions concerning how to adapt the tutoring, how to tailor explanations or hints or to provide the right level of coaching. A number of student modelling techniques have been developed (Dillenbourg and Self, 1992; Holt et al., 1994). They mainly describe the student's knowledge in terms of beliefs and are based on Artificial Intelligence inference. For instance, Logic-ITA (Merceron and Yacef, 2003) is a teaching tool for formal proofs in propositional logic that makes use of a Web-based ITS (Logic Tutor) to keep the instructor informed about the progress and the problems encountered by the class. Logic-ITA collects data from students' exercises in its internal database, and provides tools to mine this database



Figure 2.1: Screenshots from track students window in WebCT (top) and track of content pages visited by a specific student (bottom). Students' names and user ID are blurred for privacy issues.

and analyse pedagogical relevant information, such as identifying the most common mistakes and the logic rule causing the most problems.

The student model can be an internal component (black-box) of a complex system, or can be available to students and tutors for inspection and tuning. This case is referred to as *open student models*, or *inspectable student models*, or also *scrutable student models* (Holden and Kay, 1999) and is mainly aimed at improving the quality of the learner model and promoting students' meta-cognition. Others have gone beyond the simple exploration, and have combined the externalisation process with the possibility for the students to negotiate and tune the student model (Dimitrova et al., 2000; Kay, 1995). Recent studies have demonstrated that opening the student model to individuals can encourage the learner reflection, as the awareness about what they know and what they don't know enhances the learning process (Bull, 1997; Brna et al., 2003). However, student modelling approaches require a narrow fine-grained model of the knowledge domain and are appropriate when the learning is made largely through interaction with the system, such as in ITS systems.

During the latest years a great number of Intelligent Web-based learning platforms have been developed. One of the most challenging research areas is represented by the advanced Web-based Adaptive and Intelligent Educational Systems (AIES) (Brusilovsky, 1999). They are new Web-based educational applications that can offer some adaptivity and intelligence, with the purpose to adapt the content and/or the presentation of the course to individual students and to provide personalised tutoring. Although the AIES produced a plethora of educational platforms (Brusilovsky et al., 1996, 1998; De Bra and Ruiter, 2001; Specht et al., 1997), these systems have been used only in very specific and limited cases. For example, the course "Hypermedia Structures and Systems" at The Technische Universiteit Eindhoven (TU/e) taught by prof. Paul De Bra use the AHA! platform (*An open Adaptive Hypermedia Architecture*). Other examples of AIES systems are NetCoach (Weber et al., 2001), a tool for the creation of adaptive Web-based courses where instructors can observe users in the course. They get a list of all currently active users and have access to all users in the course. Instructors can inspect the current learning state of a user. They can see how long users have been working on the course, which concepts they have worked at, how many errors they have made, and other interesting information that may help instructors to understand the difficulties learners have with the course in order to help these learners. Figure 2.2 represents a screenshot of the interface for instructors in NetCoach.

However, these are some of the few cases where an AIES is being used in a real life course. Most Web-based courses make use of commercial CMS that provide mainly a simple network of static hypertext pages. They adopt a limited presentation schema approach, where the content pages are just HTML or PDF files. In this way, it is very easy for an instructor who wants to create an on-line course to convert his/her material produced, for instance, with a word processor in HTML format, and put this material on-line very quickly. This is probably the reason why

example

show user statistics -->

load old user -->

remove user -->

Total of 4 users
3 active users in the last
60 minutes

show

Show guests

Users and User Groups

Discussion List

User Infos

edit

edit

edit

Active users in example (4)

Last Change	Load Time	User Name	Note	Act Item
14:01:22 June 20, 2001	13:54:50 June 20, 2001	john	0	Chapter-2-1-2
13:54:25 June 20, 2001	13:54:14 June 20, 2001	joe	0	Chapter-3-1
13:51:59 June 20, 2001	13:51:42 June 20, 2001	student	0	Chapter-2-1

Name: john

Visited pages: 8
Learning time: 4 min, 36 sec
Current concept: Chapter-2-1-2

Tasks worked on	Current Page: Chapter 2-1-2	Chapter: Chapter 2-1	Total Course
Exercises	1.0 of 1.0 100.0% correct	2.0 of 2.0 100.0% correct	4.0 of 4.0 100.0% correct
Tests	-	-	-
Total	1.0 of 1.0 100.0% correct	2.0 of 2.0 100.0% correct	4.0 of 4.0 100.0% correct
Learning State (Points)	1.0 of 1.0 100.0% learned	2.0 of 2.0 100.0% learned	4.0 of 5.0 80.0% learned

User: john

Course	Concept	CritV	Value	Solv	Infer	Hits	Errs	NCorr	NFalse	U-No	ConceptName
2	2	1.0	1.0	1	0	1	0	1	0	0	Chapter-1-1
2	3	1.0	1.0	1	0	1	0	1	0	0	Chapter-1-2
2	6	1.0	1.0	1	0	1	0	1	0	0	Chapter-2-1-1
2	7	1.0	1.0	1	0	1	0	1	0	0	Chapter-2-1-2

Figure 2.2: A screenshot of the interface for instructors in NetCoach. Image from Weber et al. (2001).

commercial tools dominate the market of CMS, used at universities all over the world. On the other hand, AIES very often require the designer of the course to cope with some technical issues, such as the conversion of the course materials into an XML format.

2.6 The proposed solution

This work proposes to build a flexible student monitoring and tracking system for a generic CMS where the information needed to build a student model is provided by the log data of the students' activities collected by the CMS, rather than being inferred by the system by the means of some Artificial Intelligence inference. This can be done by representing CMS's activities and communications students have been engaged into in some *graphical representations*. A novel approach is proposed here whereby Information Visualisation (IV) techniques (Card et al., 1999; Spence, 2001; Tufte, 1990) are adopted to graphically render the vast amount of student tracking data collected by CMS. Using sophisticated IV tools, instructors can manipulate the graphical representations generated, which will help instructors gain an understanding of their learners and become aware of what is happening in distance classes. Inferring student and group models using Artificial Intelligence techniques can be fairly challenging, especially when large numbers of students are dealt with. In contrast, Information Visualisation techniques merely represent data collected by CMS in a visual format with minimum data processing. By managing student tracking data with appropriate visualisation techniques instructors form mental models (Spence, 2001) of individual students as well as mental models of groups of students. In this case models are inferred in the instructor's mind, instead of being inferred by algorithms.

Next chapter will explore some forms of graphical representations of students' data in educational systems and we will describe techniques for Information Visualisation that could be applied in students' data from a CMS. These techniques will be used to derive *CourseVis*, a prototype of graphical visualisation tool that acquires tracking data from a CMS and generates graphical representations that can be explored by instructors.

2.7 Summary

In this chapter we have discussed how Web-based distance education is evolving, and have described the Course Management Systems as the most popular tools that enable the delivery of education with Internet. We have argued that some of the problems occurring when using these tools, as revealed in literature, can be minimised if the instructor of the course is aware of what is happening to the students and intervene promptly with appropriate actions. To be aware of what is happening in the course the instructors may be facilitated by some graphical representation of students' data collected by an appropriate tool. The next chapter will review graphical representa-

tions of students' data performed by educational tools and will explore some techniques that may help in design effective graphical representations.

Chapter 3

Information Visualisation and its Applications in Educational Systems

Last chapter illustrated some problems that might arise to students when studying at distance using CMS. We have pointed out some works that suggest that these problems can be minimised if the instructor of the course is active in monitoring student learning and provides the learners with appropriate and prompt feedback when a problem occurs. We have argued that there are no systematic investigations on providing instructors with appropriate student monitoring and tracking system, and have proposed to explore some pictorial representations to the students' tracking data collected by CMS. This chapter will explore the usage of graphical representations to externalise students' data in educational environments. We will also present some techniques of Information Visualisation (IV) that can be useful to present complex multi-dimensional data to users with simplicity, clarity and efficiency.

3.1 Introduction

There are several situations in the real world where we try to understand some phenomena, data, and events by graphics. Some aspects, such as when people need to find a route on a city map, the stock market trends during some period, the unemployment diffusion in Europe, etc. may be understood better using graphics rather than text (Ruddle et al., 2002). Let us consider for example a situation when a person needs to find a route to go from Lugano to Pisa. This information can be represented both in a pictorial way, such as a map, and in a textual way, such as the textual description of a route. See this example represented in Figure 3.1.

Some specific aspects that may be interesting when one is finding a route, such as the possibility for finding an alternative route, or the existence of historical places in the vicinity of the route, may be understood better from the graph rather than the textual format. Graphics, if well



Figure 3.1: Driving directions from Lugano to Pisa provided both in graphical and textual format. Images from <http://www.viamichelin.com>

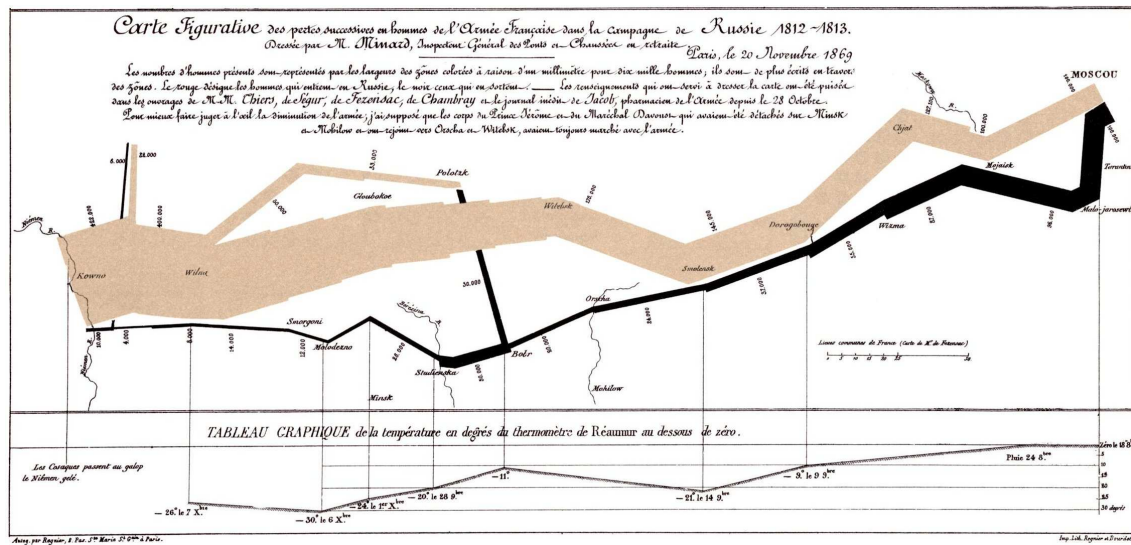


Figure 3.2: The Napoleon's army route in the Russian campaign of 1812 by Charles J. Minard, taken from Tufte (1983).

constructed, may make the human cognitive process of constructing a mental image of the route easier (Ruddle et al., 2002).

3.1.1 Graphics for presentation

Graphics are a mean to display facts about the data in a way that others can see and understand the underlying structure and the hypothesis about the data (Rober, 2000). Tufte (1983, p. 13) writes, “*graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency*”. See for instance Figure 3.2. This map by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the Polish-Russian border, the thick band shows the size of the army at each position. The path of the Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. Edward Tufte, a highly reputed visual designer, statistician and academic, comments on this image “*It may well be the best statistical graphic ever drawn*” (Tufte, 1983, p. 40).

3.1.2 Graphics for explorative analysis

Graphics also are a mean for finding and identifying structures and properties in a given data set (Card et al., 1999; Tufte, 1983; Rober, 2000). The special properties of visual perception of data may facilitate the finding of relationships, trends, revealing hidden patterns, or as Bertin

(1981, p. 16) says, “*it is the visual means of resolving logical problems*”. To illustrate, Figure 3.3 represents a map of London’s Soho district where an outbreak of cholera appeared in 1845. Black dots represent individual deaths from cholera, and **x** marks the position of the water pumps. This map allowed Dr. John Snow to observe that most of the deaths of the area were concentrated around the Broad Street water pumps, which was discovered to be a cause of the diffusion of cholera in the area. (Tufte, 1983; Spence, 2001).

3.1.3 Graphics for confirmative analysis

Graphics are also the visual mean to confirm or reject some hypothesis about the data. For example, operators in the stock market exchange know that stock market indexes between different countries influence each other. This can be illustrated by the picture in Figure 3.4 representing the values of MibTel, the Italian stock market index and the US Down Jones market index within one year. It is easy to recognise that the increasing and decreasing values of both indexes come together. This relation, explicitly presented in the graphics, would be represented in the textual representation of the values by symbolic formulas which are less expressive and intuitive than a picture (Larkin and Simon, 1987).

3.2 Information Visualisation

Graphical representations are often referred by eminent authors with the term “*visualisation*” (or *visualization* in the more diffused American version of the term). For instance, Card et al. (1999, p. 6) define the term visualisation as “*the use of computer-supported, interactive, visual representations of data to amplify cognition*”. It has been noted by Spence (2001) that there is a diversity of uses of the term “*visualisation*”. For instance, in a dictionary the following definitions can be found:

*Visualize: form a mental image of..*¹.

*Visualization: The display of data with the aim of maximizing comprehension rather than photographic realism*².

*Visualization: the act or process of interpreting in visual terms or of putting into visible form*³

These definitions reveal that visualisation is an activity in which humans are engaged, as an internal construct of the mind (Spence, 2001; Ware, 2000). It is something that cannot be printed

¹The Concise Oxford Dictionary. Ed. Judy Pearsall. Oxford University Press, 2001. Oxford Reference Online. Oxford University Press.

²A Dictionary of Computing. Oxford University Press, 1996. Oxford Reference Online. Oxford University Press.

³Merriam-Webster Online dictionary <http://www.webster.com>

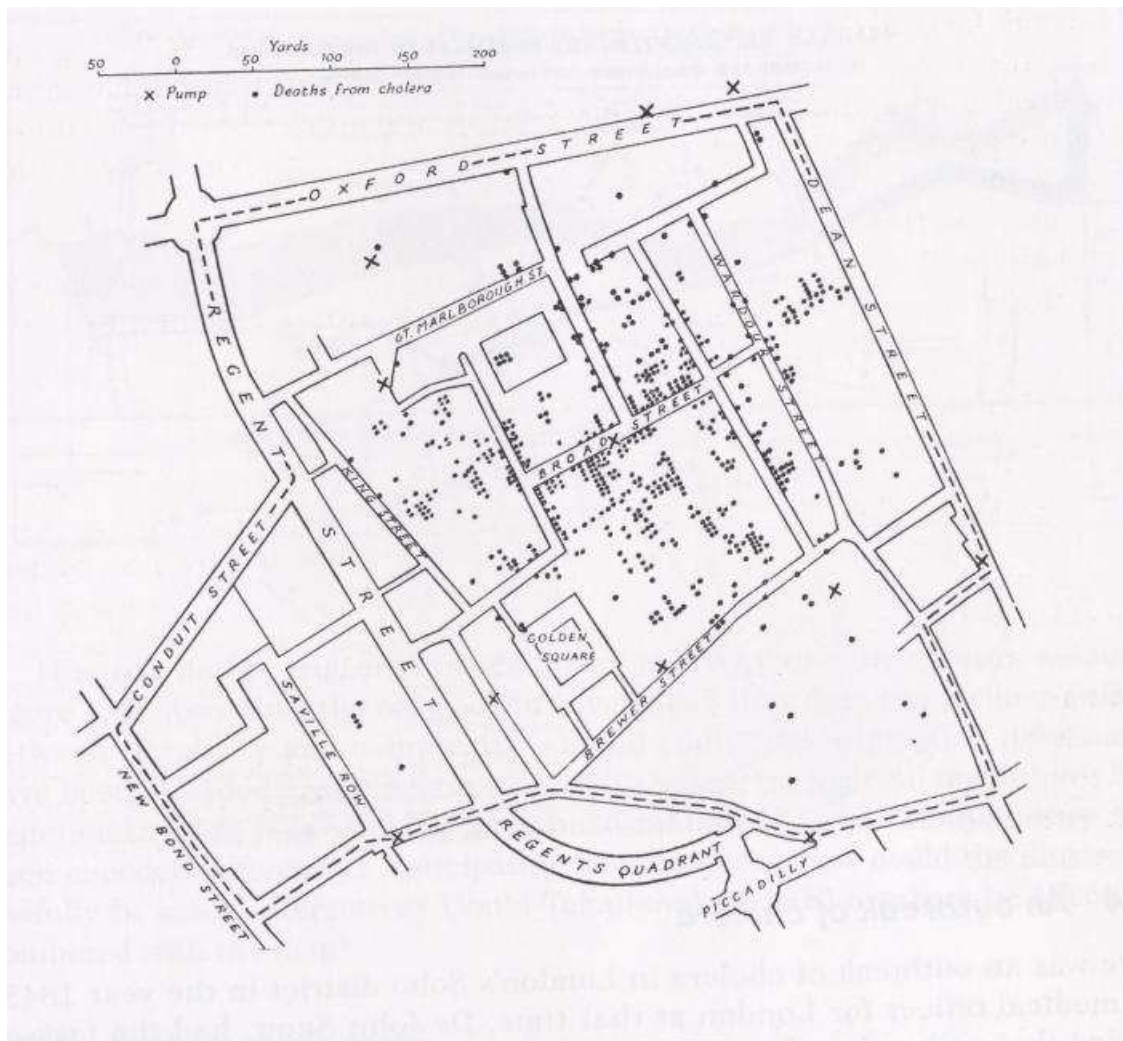


Figure 3.3: An 1845 map of London's Soho district plotted by Dr. John Snow showing deaths from Cholera and the location of water pumps. Dots represent cholera cases and X represent water pumps. Image from Tufte (1983).

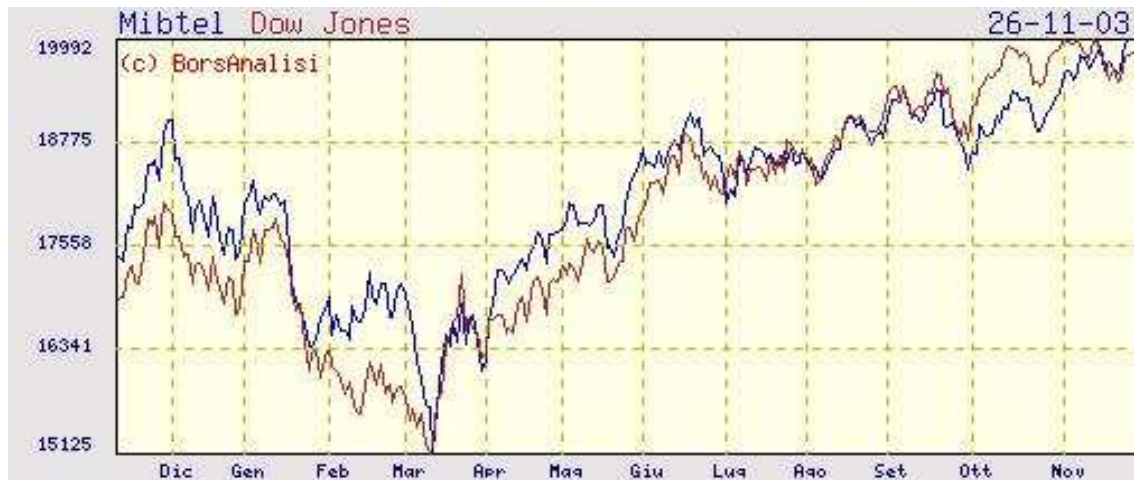


Figure 3.4: A comparison chart between the Italian stock market index, MibTel (in blue) and the United States' Dow Jones market index (in red). Taken from <http://www.borsanalisi.com>

on a paper or displayed on a computer screen. With these considerations we can summarise that visualisation is a cognitive activity, facilitated by graphical external representations from which people construct internal mental representation of the world (Ruddle et al., 2002; Spence, 2001; Ware, 2000). Computers may facilitate the visualisation process with some visualisation tools. This is especially true in the latest years with the use of more and more powerful computers at low cost. However, the above definition is independent from computers: although computers can facilitate visualisation, it still remains an activity that happens in the mind. Some authors use the term “visualisation” to refer to both the printed graphical representation and the cognitive process of understanding an image. In this thesis, we maintain the distinction between the creation of a pictorial representation of some data and the formation of an internal mental model of the data when interpreting the pictorial representation.

Information Visualisation is a relatively new discipline concerned with the creation of visual artefacts aimed at amplifying cognition. A number of new definitions have been produced to define the scope of this discipline. The Information Visualisation Research Group at the Institute for Software Research at University of California, Irvine cite on its Web pages:

Information visualization focuses on the development and empirical analysis of methods for presenting abstract information in visual form. The visual display of information allows people to become more easily aware of essential facts, to quickly see regularities and outliers in data, and therefore to develop a deeper understanding of data. Interactive visualization additionally takes advantage of people's ability to also identify interesting facts when the visual display changes, and allows them to

*manipulate the visualization or the underlying data to explore such changes*⁴.

Similarly, the User Interface Research Group Web-site of the Palo Alto Research Centre (PARC-XEROX) defines that:

*Information Visualization is the use of computer-supported interactive visual representations of abstract data to amplify cognition. Whereas scientific visualization usually starts with a natural physical representation, Information Visualization applies visual processing to abstract information. This area arises because of trends in technology and information scale. Technically, there has been great progress in high-performance, affordable computer graphics. At the same time, there has been a rapid expansion in on-line information, creating a need for computer-aid in finding and understanding them. Information Visualization is a form of external cognition, using resources in the world outside the mind to amplify what the mind can do*⁵.

The above definition is appropriate nowadays with the computers being a constant part of our life, but as Spence (2001) and Hearst (2003) pointed out the role of computers is merely of a mean that facilitates visualisations. Hearst (2003) summarises that IV is:

The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system.

IV has a long history having its origin from the historical works of J. H. Lambert [1728-1777] and William Playfair [1759-1823] who were the first to introduce graphics, in contrast with the tabular presentation of data and are considered the inventors of the modern graphics design (Tufte, 1983). Starting with Playfair, data were represented with methods of plotting. Other important contributions were made more recently by Jacques Bertin and Edward Tufte. Bertin, a French cartographer, was the first who tried to define, in 1967, a theory of IV by the identification of the basic elements of diagrams and described a framework for graphics design (Bertin, 1983). In 1983 Tufte published his theory of data graphics focused on the maximisation of the density of useful information in graphics (Tufte, 1983). Both Bertin and Tufte's theories have influenced the current development of IV.

3.2.1 Cognitive amplification

Graphics aid thinking and reasoning in several ways. For example, let us take a multiplication (a typical mental activity) e.g. 27×42 in our head, without having a pencil and paper. This will

⁴<http://www.isr.uci.edu/research-visualization.html>

⁵<http://www2.parc.com/istl/projects/uir/projects/ii.html>

take usually at least five times longer than when using a pencil and paper (Card et al., 1999). The difficulty in doing this operation in the mind is holding the partial results of the multiplication in the memory until they can be used:

$$\begin{array}{r}
 27 \times 42 \\
 \hline
 54 \\
 1080 \\
 \hline
 1134
 \end{array}$$

This is an example which shows how visual and manipulative use of the external representations and processing amplifies cognitive performance. Graphics use the visual representations that help to amplify cognition. They convey information to our minds that allows us to search for patterns, recognise relationship between data and perform some inferences more easily. Card et al. (1999) propose six major ways in which visualisations can amplify cognition:

1. by increasing the memory and processing resources available to users;
2. by reducing the search for information;
3. by using visual representations to enhance the detection of patterns;
4. by enabling perceptual inference operations;
5. by using perceptual perception mechanisms for monitoring;
6. by encoding information in a manipulable medium.

IV is becoming an increasingly important discipline because the availability of powerful representations may facilitate the way we present and understand large complex datasets. Larkin and Simon (1987) argued in their seminal paper “*Why a diagram is (sometimes) worth ten thousand words*” that the effectiveness of graphical representations is due to their spatial clarity. Larkin and Simon compared the computational efficiency of diagrams and sentences in solving Physics problems, and concluded that diagrams helped in three basic ways:

1. **Locality** - is enabled by grouping together information that is used together. This avoids large amounts of search and allows different information closely located to be processed simultaneously. For example, Figure 3.4 puts together information about the history of two different stock market indexes and allows to process their evolution immediately.
2. **Minimising labelling** - is enabled by using location to group information about a single element, avoiding the need to match symbolic labels and leading to reducing the working

memory load. For example, driving directions from Lugano to Pisa provided in graphical format in Figure 3.1 use visual entities such as lines depicted in red with a yellow stripe in the middle to denote a highway. Turning points (such as in Parma in the example) are clearly indicated by a crossing of the roads. Symbolic textual representations used in the textual format of the map are unnecessary because the connections are explicitly represented in the graphics.

3. **Perceptual enhancement** - is enabled by supporting a large number of perceptual inferences which are easy for humans to perform. For example, in Figure 3.3 the link between deaths from cholera and the location of a water pump responsible for the spread of cholera could be recognised immediately.

IV definitions introduce the term “abstract data”, for which some clarification is needed. The data itself can have a wide variety of forms, but one can distinguish between data that have a physical correspondence and is closely related to mathematical structures and models (e.g. the airflow around the wing of an aeroplane, or the density of the Ozone layer surrounding earth) and data that is more abstract in nature (e.g. the stock market fluctuations, or the effects of temperature on the Napoleon’s army movements in the Russian campaign). The former is known as *Scientific Visualisation*, and the latter as *Information Visualisation*. (Spence, 2001; Uther, 2001; Hermann et al., 2000). Scientific Visualisation was developed in response to the needs of scientists and engineers to view experimental or phenomenal data in graphical formats (examples are given in Figure 3.5), while Information Visualisation is dealing with unstructured data sets as a distinct flavour (Hermann et al., 2000). This thesis deals with Information Visualisation, and hence Scientific Visualisation is beyond the scope of this review.

A key question in IV is how we convert abstract data into a graphical representation, preserving the underlying meaning and, at the same time, providing new insight (Hearst, 2003). There is no “magic formula” that helps the researchers to build systematically a graphical representation starting from a raw set of data. It depends on the nature of the data, the type of information to be represented and its use, but more consistently, it depends on the creativity of the designer of the graphical representation. Some interesting ideas, even if innovative, have often failed in practice. Tufte (1983) and Bertin (1981) list a number of examples of graphics that distort the underlying data or communicate incorrect ideas. Tufte indicates some principles that should be followed to build effective well designed graphics. In particular, a graphic should:

- show the data;
- avoid distorting what the data have to say;
- present many data in a small space;

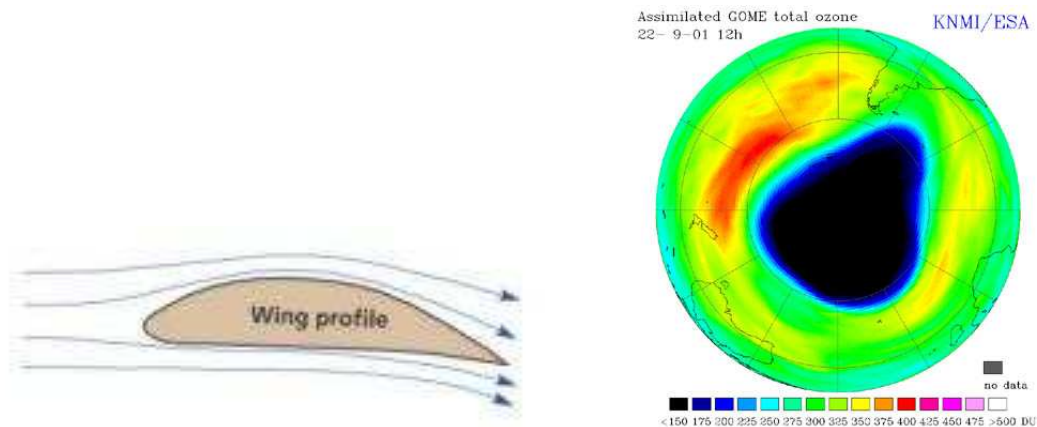


Figure 3.5: Two examples of scientific visualisations. On the left, representation of the laminar flows on the wing profile (image taken from http://www.esa.int/export/esaMI/High_School/ESAF8BG18ZC_1.html). On the right, Ozone hole over the South Pole during September 2001 (image taken from http://www.esa.int/export/esaSA/ESALRRVTYWC_earth_1.html)

- make large data sets coherently;
- encourage inferential processes, such as comparing different pieces of data;
- give different perspectives on the data - from broad overview to the fine structure.

Graphics facilitate IV, but a number of issues must be considered (Shneiderman, 2002; Tufte, 1983; Spence, 2001):

1. Data is nearly always multidimensional, while graphics represented on a computer screen or on a paper are presented in a 2D dimensional surface;
2. Sometimes we need to represent a huge dataset, while the number of data representable on a computer screen or on a paper is limited;
3. Data may vary during the time, while graphics are static;
4. Humans have remarkable abilities to select, manipulate and rearrange data, so the graphical representations should provide users with these features.

The above issues are considered in the IV field, and a number of methods and techniques have been proposed to meet these requirements. Card et al. (1999, p.33) give a comprehensive list of eight types of data, eleven visual structures, four views, three types of human interaction, eleven tasks and eleven levels that a user might want to accomplish with a visualisation tool (see Table

Data Tables	Visual Structures	Views	Human interaction	Tasks	Level
Spatial (scientific) Geographic Documents Time Hierarchies Networks World Wide Web	Position Marks Proprieties: -Connection -Enclosure -Retinal -Time Axes: -Composition -Alighment -Folding -Recursion -Overloading	Brushing Zooming Overview+ detail Focus+context	Dynamic queries Direct Manipulation Magic lenses	Overview Zoom Filter Details-on-demand Browse Search Read fact Read comparison Read pattern Manipulate Create	Delete Reorder Cluster Class Promote Average Abstract Instantiate Extract Compose Organize

Table 3.1: Specific techniques for Information Visualisation identified by Card et al. (1999)

3.1). This is a multitude of possibilities. Next sections will describe some techniques that can be used to display pictorial representations and that could be (or have already been) considered to display students' data in educational environments.

3.2.2 Issues to consider in Information Visualisation

Before using one or more IV techniques we have to consider several issues (Spence, 2001; Card et al., 1999; Hearst, 2003; Reed and Heller, 1997):

1. **The problem.** This relates to what has to be presented, found, or demonstrated.
2. **The nature of the data.** Data types could be *numerical* (e.g. list of integers or reals), *ordinal* (non-numerical data having a conventional ordering, such as days of the week), and *categorical* (data with no order, such as names of persons or cities).
3. **Number of data dimensions.** Depending on the number of dimensions (also called *attributes* or *variables* (Spence, 2001)), representations are said to be handling *univariate* (one dimension), *bivariate* (two dimensions), *trivariate* (three dimensions), and *multivariate* (four or more dimensions) data. We perceive our world in three spatial dimensions, so it is easy to map and interpret up to three dimensions. However, handling more than three dimensions is very frequent in real world situations and represents one of the most challenging tasks in IV.
4. **Structure of the data.** This could be *linear* (data coded in plain data structures such as arrays, tables, alphabetical lists, sets, etc.), *temporal* (data which changes during the time),

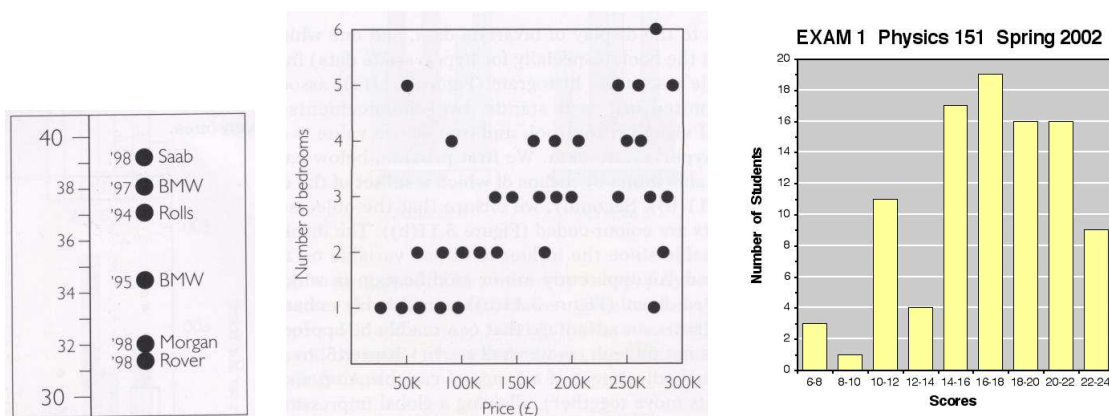


Figure 3.6: Samples of 1D (left), 2D (middle) scatterplots, and histogram . Images from Spence (2001) (left and middle images) and <http://www.physics.umass.edu/p151sec2s02/> (right image).

spatial or *geographic* (data which has a correspondence with something physical, e.g. maps, floor plans, 3D CAD; usually this is a subject of scientific visualisation and is not considered to be IV in the strict sense), *hierarchical* (data that naturally arise in taxonomies, the structures of organisations, disk space management, genealogies, etc.), *network* (data describing graph structures, i.e. nodes and links, nodes representing a data point, and a link representing a relationship between two nodes).

5. **Type of interaction.** Whether the resulting graphical representation is *static* (e.g. a print or a static image on a display screen), *transformable* (users can manipulate how the representation is rendered, such as zooming or filtering), or *manipulable* (users may control parameters during the process of image generation, i.e. restricting the view to certain data ranges)

Each one of the previous can suggest the use of one or more techniques, the most important (and used) ones are outlined next.

3.2.3 Techniques for representing univariate and bivariate linear data

Univariate and bivariate linear data are easily representable as a single image which relates data values with some scale in a plain. Conventional approaches are *scatterplots*, *charts*, and *histograms*. When used with bivariate data, scatterplots are helpful for encouraging the viewer to assess the possible causal relationship between the two plotted variables (Tufté, 1983). Figure 3.6 on the left represents a one-dimensional scatterplot illustrating the prices of some cars. The only attribute here is represented by the numeric value of a car, mapped onto the scale of the

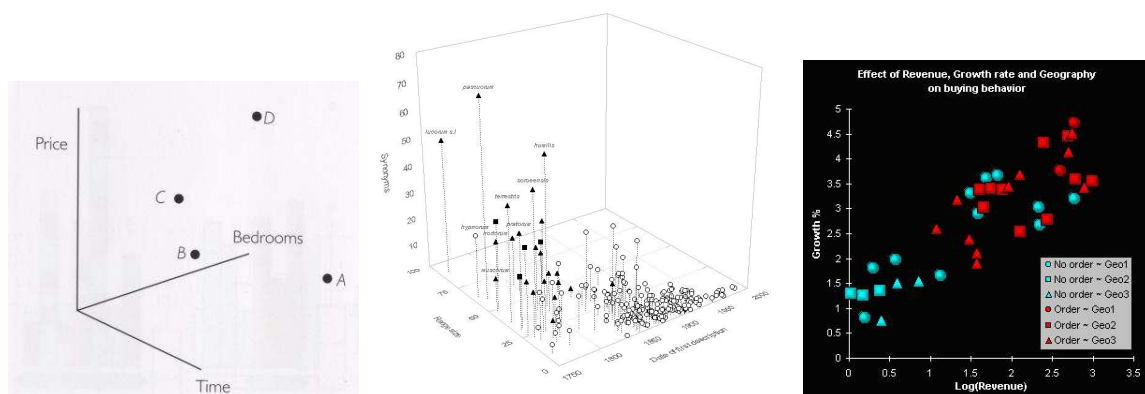


Figure 3.7: Some examples of 3D scatterplots. Images from Spence (2001) (left), <http://www.nhm.ac.uk/entomology/bombus/introduction.html>(centre), <http://www.kovcomp.co.uk/xlstat/tools/t01.html>(right).

axes. This image could be helpful to easily locate the lowest, the highest values, and the general distribution of values and branching on data (Spence, 2001). Figure 3.6 in the middle represents a two-dimensional scatterplot which relates prices of houses with the number of bedrooms from a house-hunting archive. Users can clearly identify global trends (price of houses increases with the number of bedrooms), as well as items far from other having similar characteristics (e.g. a one-bedroom house selling for more than 250.000 GBP) (Spence, 2001). Another form of representation is the histogram (Figure 3.6, right). The example reports distribution of scores on a Physics exam. Histograms are suitable to facilitate comparisons between the values of a dimension, while scatterplots are helpful to convey overall impression of relationships between two variables (Hearst, 2003).

3.2.4 Techniques for representing trivariate linear data

Three-dimensional linear data is easily representable in a three dimensional space, as we are living in a three-dimensional world. However, the problem still arises because we basically can represent in a printed image or on a computer screen a two dimensional representation of a three-dimensional space, hence some approximations are necessary (Spence, 2001). Figure 3.7 left illustrates an example of a 3 dimensional scatterplot where each dimension is linearly mapped on an axis. The impassable barrier of the 2D image makes impossible to distinguish between the number of bedrooms of C and B. To overcome this problem some solutions have been proposed, such as to project items on the plans to identify the exact position (see Figure 3.7, centre) or to replace one dimension with a different shape of the items according to this value (see Figure 3.7, right).

3.2.5 Techniques for representing multivariate linear data

Very often real world cases cope with situations where relationships between more than three variables must be analysed. Let's think about an analysis of factors such as age, living place, job, and sex on the appearance of cancer on a sample of patients. This is a challenging aspect in IV, because some proprieties of images have to be explored to distinguish between several variables in a 2D drawing plane. For this purpose several methods have been proposed. Sachinopoulou (2001) suggested a classification into six groups, summarised in the following table:

Methods	Description	Some known techniques
<i>Geometric</i>	Transforming and projecting data in a geometric space.	Scatterplot matrix, Hyperslice, Prosection views, Surface and volume plots, Parallel coordinates, Textures and rasters.
<i>Icon</i>	Relies on a geometric figure (the icon) where the values of an attribute is associated with one features of this, such as the colour, a shape, the orientation.	Chernoff faces, Stick figure, Colour icon, Glyphs and Autoglyph.
<i>Pixel</i>	Use pixel as basic representation unit, and manipulate pixels to represent data.	Space fillings and Mosaic plots.
<i>Hierarchical</i>	Include trees and hierarchies and are useful when the data has some hierarchical or network structure.	Hierarchical axes, Dimension stacking, Threes, Worlds within worlds, Infocube.
<i>Distorsion</i>	Propose to distort the tree-dimensional space to allow more information to be visualised.	Perspective Wall, Pivot table and table lens, Fish eye view, Hyperbolic trees, Hyperbox.
<i>Graph based</i>	Represent data using nodes and edges and is adopted when the large graphs should be represented.	Basic graph, Hyperbolic graph.

Techniques cited in the rightmost column of the table above are described in details in Sachinopoulou (2001); Card et al. (1999); Spence (2001); Chen (1999). These techniques are able to represent data in multidimensional space. A comprehensive description of multidimensional techniques is beyond the scope of this work, rather we focus on those techniques that have been used in CourseVis.

A number of additional techniques for multidimensional data representations exist. Differently from the techniques cited in the table above, they don't offer the possibility to establish relationship between each variable, but may represent several attributes allowing relationships on a subset of variables. Some of them are *composition*, *layering and separation*, *micro-macro readings*, and *small multiplies*.

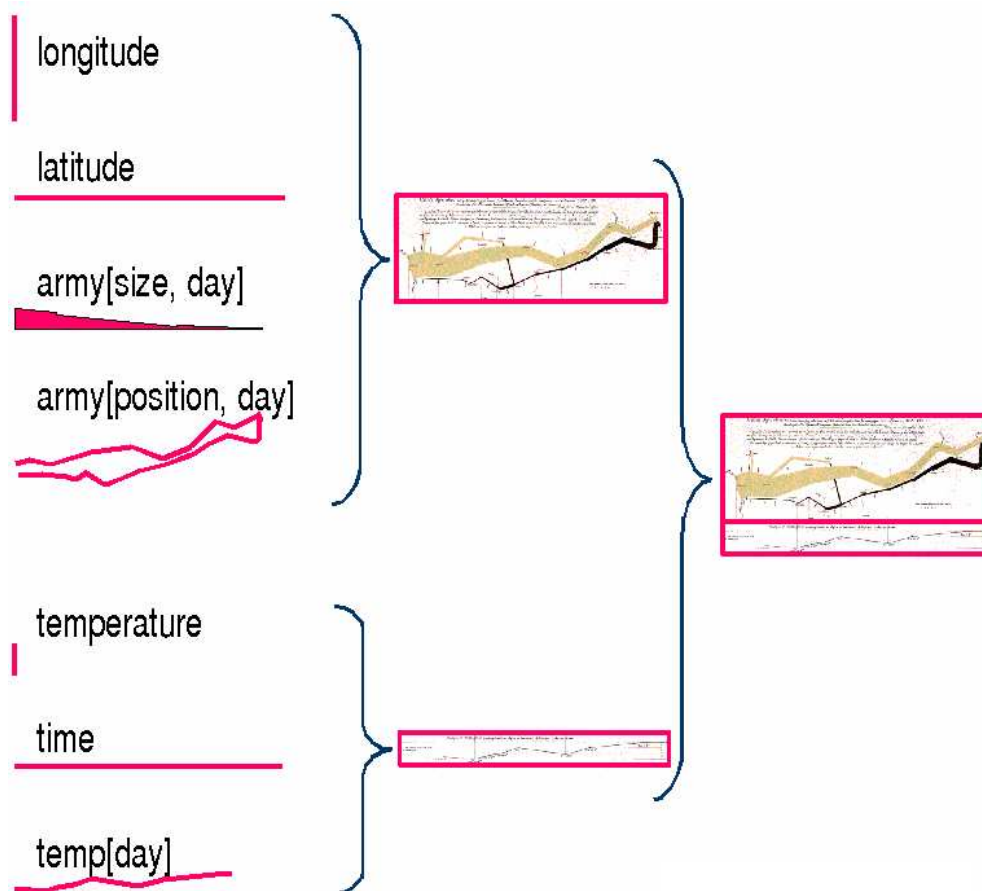


Figure 3.8: Single-axis composition technique used by C. J. Minard in the Napoleon's army route in the Russian campaign diagram. Image from Mackinlay (2000).

Composition

The basic idea of the composition technique (Mackinlay, 1986; Card et al., 1999) is the orthogonal placement of axes that encode the same information, creating a 2D metric space of multidimensional data. An example of this technique is the diagram of the Napoleon's army route in Figure 3.2. This is an example of *single-axis composition* technique (Mackinlay, 1986) where attributes *army size*, *army longitude*, *army latitude*, and *temperature* have identical horizontal axis which is the *time*. Figure 3.8 illustrates how C. J. Minard composed these variables in the famous diagram.

Layering and separation

Layering and separation is a technique illustrated, among others, by Tufte (1990) and concerns the visual differentiation of various aspects of the data. Tufte argues that “*confusion and clutter are*



Figure 3.9: A city map of the centre of Florence, streets and historical places are encoded with different shapes and colours. Image from <http://www.hotelitaliani.it/Mappa/Informazioni/PiantinaFirenze.htm>.

failure of the design, not attributes of information ... the point is to find design strategies that reveal detail and complexity - rather than to fault the data for an excess of complication" (Tufte, 1990, p. 53). He proposes layering and separation as one of the most powerful devices for reducing noise and enriching the content in graphics, and is achieved by distinction of colour, shape, size, addition of elements that direct the attention via visual signals, or ordering data to emphasise layer differences (Ruddle et al., 2002). An example of this technique is visible in Figure 3.9. It shows a city map of the centre of Florence. This picture illustrates at least 2 layers of reading: the location of historical places and the map of the streets to find a path. Colours have been used to distinguish between parks (in green), interesting places for tourists (brown), river (blue), and major streets (yellow). Shapes have been used to direct attention of tourists in historical places, and give an idea about what sort of building it is. This map can be used also by someone who is not interested in historical monuments in Florence but has to find the route to move from the *train station* to *Piazza Signoria*.

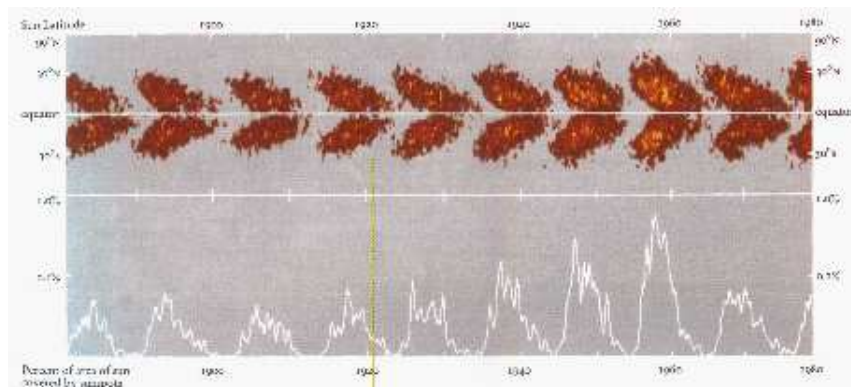


Figure 3.10: Diagram representing sunspots from 1880 to 1980 with the sine of the latitude marking a sunspot placement. Micro-macro readings allow combining patterns, details, average and variations with the same image. The lower time-series shows the total area of the sun's surface covered by sunspots by summing over all latitudes. Diagram by David H. Hathaway, Marshall Space Flight Centre, NASA, taken from Tufte (1990).

Micro-macro readings

Micro-macro reading “is a method for presenting large quantities of data at high densities in a way that a broad overview of the data is given and yet immense amount of detail is provided” (Ruddle et al., 2002). It encodes information at different levels of detail, such as the same image can be used to detect fine-grained level on information encoded (micro processing) as well as large-grained level of information (macro processing). An example of micro-macro reading is depicted in Figure 3.10. Here the picture allow the comparison of between-cycle variation (macro) and within-cycle variation (micro). Also shows an apparent growth trend in the wingspan of recent cycles (Tufte, 1990).

Small multiples

The small multiples technique consists of the same graphical design structure repeated several times (Tufte, 1983). It is used to compare at a glance a series of graphics showing the same combination of variables while another variable changes. This method considers positioning similar graphical elements together in order to emphasise changes of the data (Ruddle et al., 2002). Figure 3.11 illustrates a small multiples design for representing the coverage of mobile phones in Switzerland for the tree main operators. Comparisons are possible to decide which operator utilises on a specific region for best signal reception.



Figure 3.11: Small multiples example. The coverage for mobile telephone network in Switzerland between the tree operators: Swisscom (on the left), Orange (on the centre), and Sunrise (on right).

3.2.6 Techniques for representing spatial data

With the term spatial (or geographic) data we refer to representation techniques which have a correspondence with something physical. For example, data maps (Tufte, 1983) represent one or more attributes associated with areas of a map. See for instance the map in Figure 3.12 representing the cancer mortality in the USA at Lung, Trachea, Bronchus and Pleura for white males in years 1970 - 1994 in counties. Data maps can carry a huge volume of data in small space, but carry an intrinsic problem: they may give a wrong illusion that the importance of the information is geographic, rather than factual (Ruddle et al., 2002; Tufte, 1983). For instance, the figure doesn't allow ordering the counties according to the number of cancer deaths, or the number of people living in the county.

Space-times narrative design is another technique for representing spatial data. Its basic idea is to add spatial dimensions to the design of the graphics, so that the data are moving over the space as well as over the time (Tufte, 1983, 1990). The Napoleon's army route in Figure 3.2 is an excellent example. The graphics shows, among other things, how the crossing of the Berenzina river caused lots of deaths on Napoleon's troops, the latitude and longitude of movements, and the direction of the movements. It depicts the dynamics of the whole march with a static image.

3.2.7 Techniques for representing hierarchical and network data

There are many situations where data to be represented is linked in a graph structure, for example the organisation chart in a company, pages in a Web site, or DNA sequences. This is often referred as *network information visualisation*, and involves gaining insight into a structure that may consist of many data items (Reed and Heller, 1997). These are usually graphically represented with a drawing graph. Graphs are sometimes problematic, because a graph with few nodes is easy to draw and to comprehend visually, but real world situations often need to handle large data sets. See for example Figure 3.13, where two examples of network drawing are depicted. The image on the top is a representation of an intranet: the nodes are pages and links are URL connections

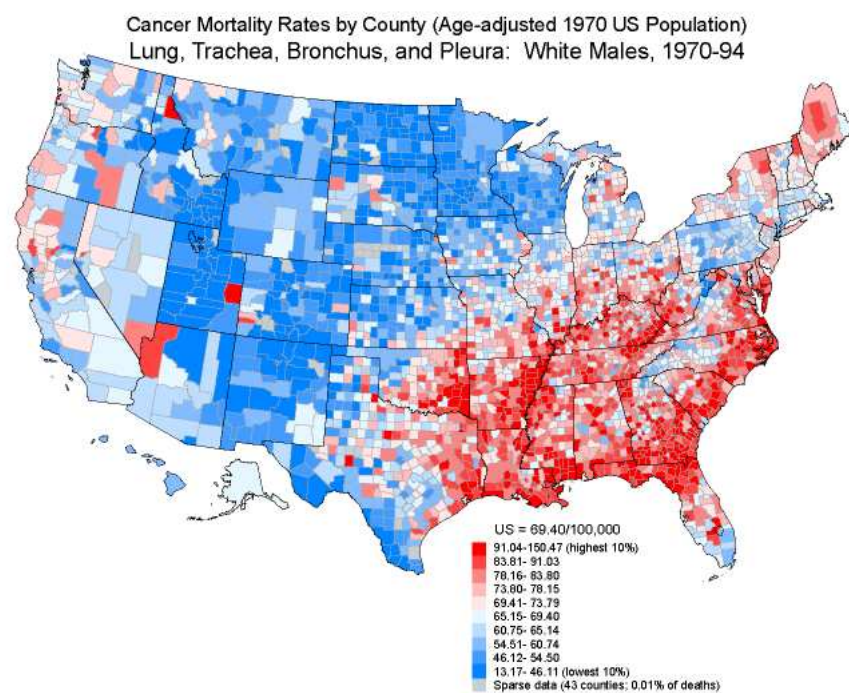


Figure 3.12: A data map example. Cancer mortality in the USA years 1970 - 1994 - white males by county. Image from USA National Cancer Institute <http://www3.cancer.gov/atlas/2.html>.

between pages. This network consists of 24 nodes easily represented in a 2D-picture. However, some situations have to deal with a large number of nodes. Figure 3.13 on the bottom represents an instance of a network with many nodes and links represented in a 3D picture. In this case, the limitation of the design pane requires rendering the network as interactive graphs. By engaging their visual image, a user is able to navigate through large networks, and to explore different ways of arranging the network components on the screen.

3.2.8 View transformations

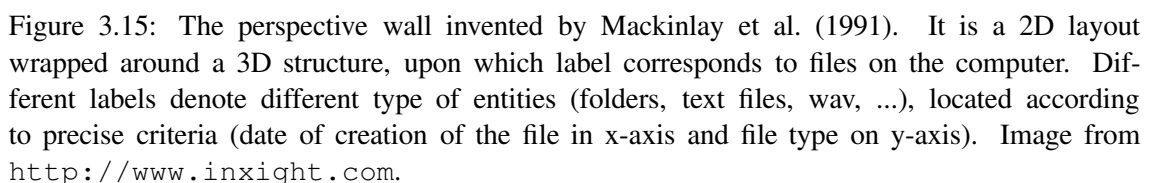
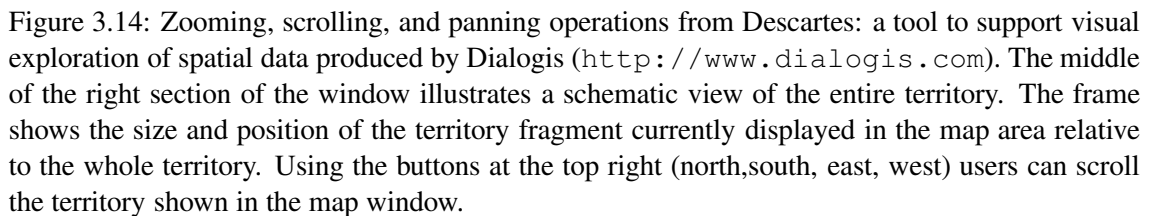
A problem humans are experiencing in their everyday life is to have too many things placed in a limited space: books on shelves, addresses in agenda, windows on a computer screen, data to display in a Personal Digital Assistant (PDA). The information explosion phenomena of last years leads to the existence of more data than what can easily be displayed at once. "*Too much data, too little display area*" is a common problem in Information Visualisation (Spence, 2001). There are several techniques proposed to solve this problem, some of which are zooming, panning, scrolling, focus+context and magic lenses (Spence, 2001).

- **Zooming** is the increasing magnification of a decreasing (or increasing) fraction of a two-dimensional image.
- **Panning** is the smooth movement of a viewing frame over a two-dimensional image of greater size.
- **Scrolling** is the movement of data past a window able to contain only a part of it (such as we are doing with the scrolling of a long document in a word processing program).
- **Focus+context**'s basic idea is to illustrate at the same time the overall picture (the context) and to see details of immediate interests (the focus). This technique allows users to expand and contract selected sections of a large image, thereby displaying simultaneously the contents of individual sections of a document as well as its overall structure.
- **Magic lenses** follow the metaphor of reading a text by the means of a lens that enlarges the size of the text. In IV it can be used to place a lens upon the area of interest and receive more detailed information on the data amplified with the lens. For instance, magic lenses could be applied to Figure 3.9: an application could show this map to tourists and a lens placed over parts of the map could show details about the historical place selected.

An example of zooming, scrolling, and panning is illustrated in Figure 3.14. Focus+context technique is illustrated in an example in Figure 3.15. An example of magic lenses is illustrated in Figure 3.16



Figure 3.13: Two examples of network drawing. On the top, a graph of a small dimension intranet (image from Graphviz drawing software Web site: <http://www.research.att.com/sw/tools/graphviz/examples/>), on the bottom, a graph of some Internet Web sites (image from TouchGraph drawing software Web site: http://www.touchgraph.com/bi.php?img=greenpeace_new.jpg).



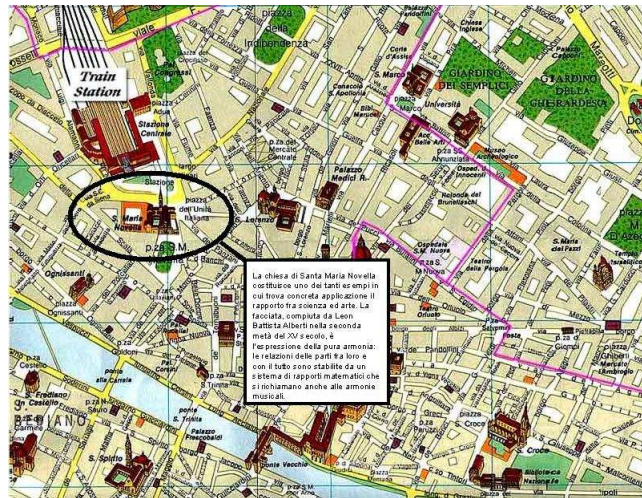


Figure 3.16: The magic lenses techniques applied to the historical places in a city map of the centre of Florence. Image from <http://www.hotelitaliani.it/Mappa/Informazioni/PiantinaFirenze.htm> and modified

3.3 Visualisations in educational software

Graphical representations of students' data have been explored by some educational systems. Here we will analyse some of them in order to have an overview of the nature of data represented and the sort of information they want to convey. The analysis here is not restricted to Web-based CMS, discussed in the previous chapter, but we consider any software application which aims to provide learning to students and gives the instructors some feedback on actions and improvements undertaken by students with the subject.

3.3.1 Visualisation of user models

Recently, tools for visualising user models are being developed. A user model is a representation of a set of beliefs about the user, particularly their knowledge in various areas, and their goals and preferences. These models are enabling the increasing personalisation of software, particularly on the Internet, where the user model is the set of information and beliefs that is used to personalise the Web site (Uther, 2001). Methods for user modelling are often exploited in educational systems, hence we consider also this sort of applications in our analysis of the state of the art.

UM/QV

QV (Kay, 1999) is an overview interface for UM (Kay, 1995), a toolkit for cooperative user modelling. A model is structured as a hierarchy of elements of the domain. QV uses a hierarchical

representation of concepts to present the user model. For instance, Figure 3.17 gives a graphical representation of a model showing concepts of the SAM text editor⁶. It gives a quick overview whether the user appears to know each element of the domain. QV exploits different types of geometric forms and colour to represent known/unknown concepts. A square indicates a knowledge component, diamond a belief, a circle indicates a non-leaf node, and crosses indicates other component types. The filling of the shape is used to indicate the component value. For instance, in the example the white squares show that the user knows that element, while the dark squares indicate lack of knowledge. Nested shapes, such as *default_size_k* or *undo_k* indicate that the system has not been able to determine whether the user knows it or not (e.g. if there is inconsistency in the information about the user). The view of the graph is manipulable, in particular, clicking on a non-leaf node it causes the sub-tree to be displayed, useful in case of models having a large number of components to be displayed.

VIUM

Several projects explore visualisation techniques to show specific perspectives of student data. One of these is *VIUM* (Uther and Kay, 2003), a tool for visualising large user models. Originally developed to help medical students find topic areas they need to study, it has also been used to show movie recommendations. *VIUM* takes a user model described in a simple RDF (enabling the display of any user model that can be expressed in this format), and allows the user to explore the model in certain ways. The model is structured as a graph of related concepts. Each concept may contain a *title*, a *score* value and a *certainty* value that indicates how strongly the evidence for the model supports the score value (Uther and Kay, 2003).

Figure 3.18 shows an application of *VIUM* to database of movies from the Internet Movie Database (IMDB) (<http://www.imdb.com>) and model a hypothetical user viewing preferences for those movies. In this example we have a contrived model of movie recommendations from the IMDB. In the display the biggest node is the currently selected movie. If the user selects “action -> show” from the menu bar on the top-left the IMDB page for this movie will be shown. Other nodes are sized according to how far they are from the current selection in the graph. Nodes are coloured according to how much the (hypothetical) recommendation service thinks you will like the movie - greener is better, redder is worse. The title is indented further from the left side of the display as the evidence for that decision decreases. The slider at the top of the left pane enables the user to determine the boundary between the classification of a component as true (displayed as green) or false (displayed as red). If the user is only looking for very strongly recommended movies, he can move the slider further to the right, and then only very strongly recommended

⁶SAM is an editor under the Plan 9 operating system, originally designed in the early 1980s for the DMD 5620, Bell Labs’ windowing terminal. It aims to emulate the *ed* editor, but with extended capabilities, such as editing multiple files at once. (Definition from [http://en.wikipedia.org/wiki/Sam_\(program\)](http://en.wikipedia.org/wiki/Sam_(program))).

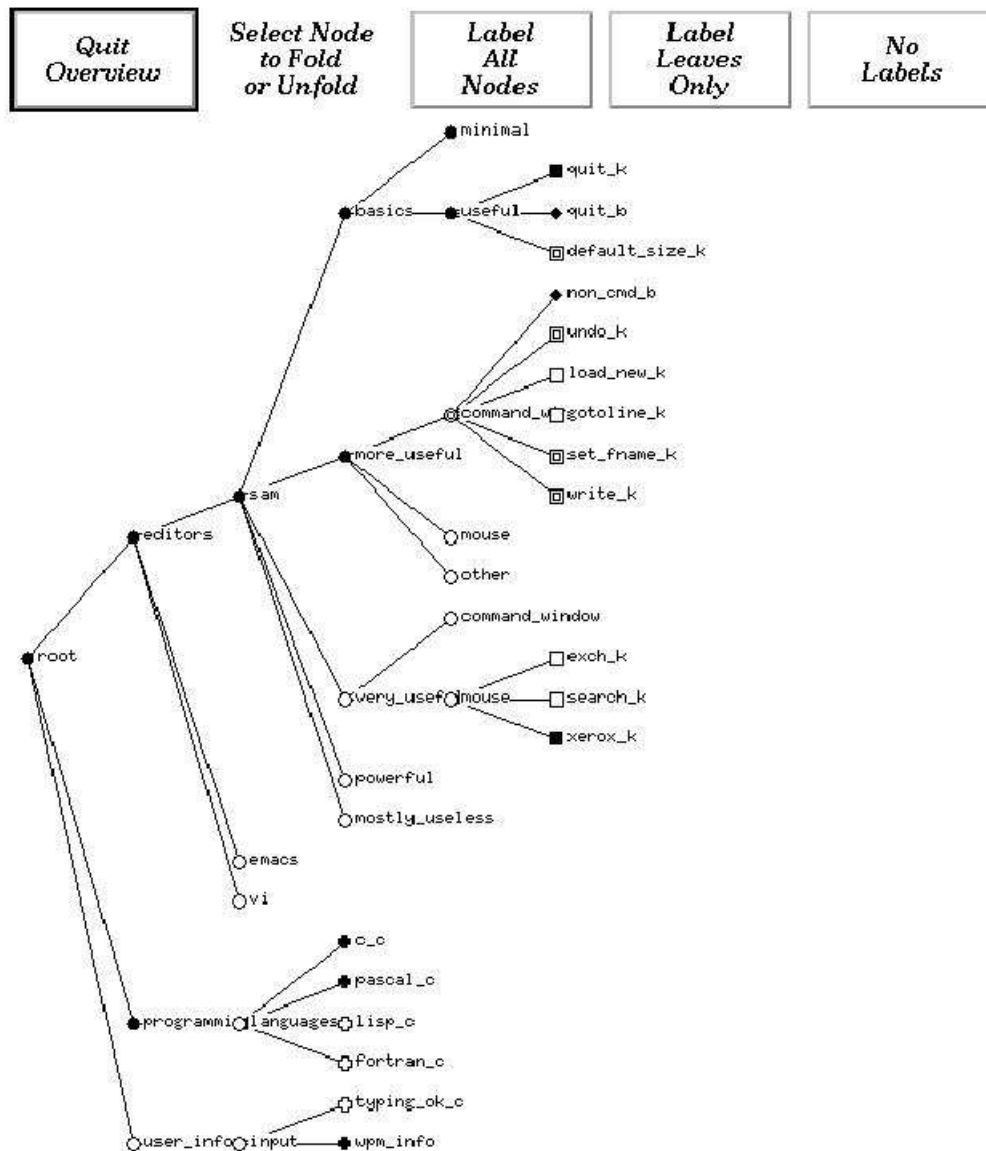


Figure 3.17: The QV tool showing a user model. Image taken from Kay (1999).

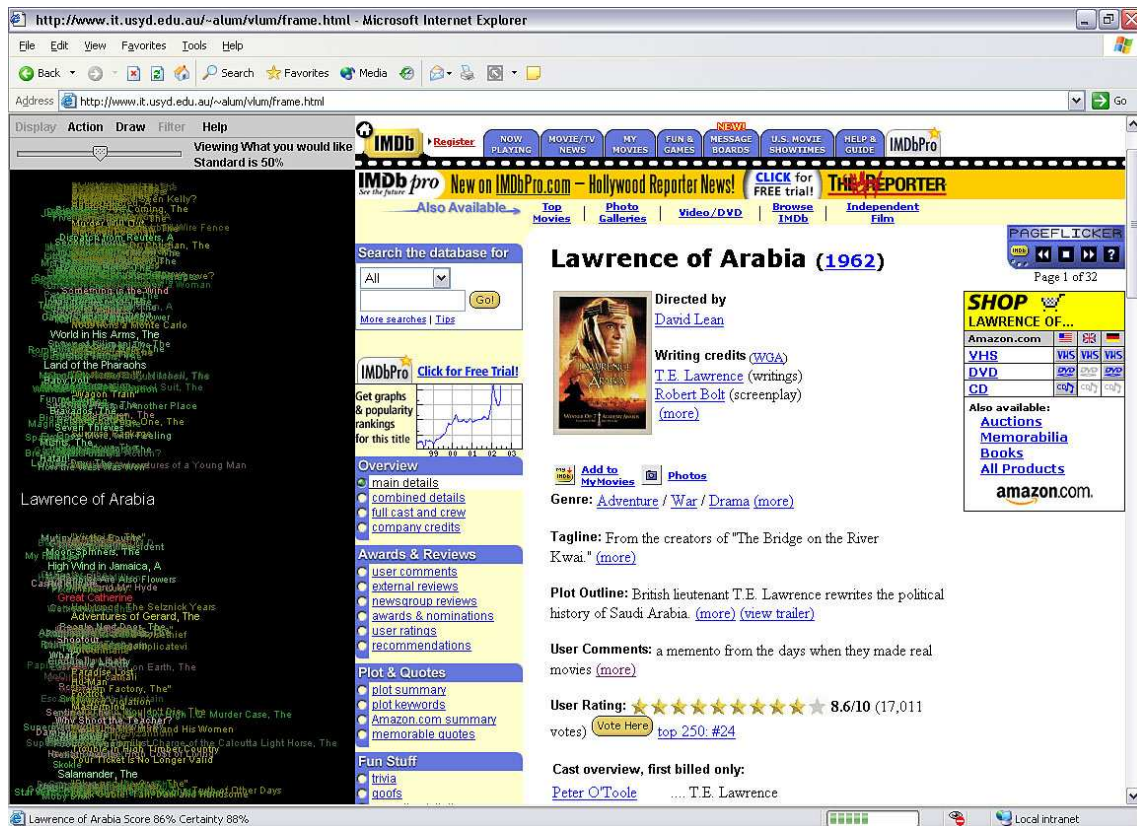


Figure 3.18: The visualisation in VIUM showing the currently selected movie. Image from Uther (2001).

movies will be green.

The VIUM visualisation uses the *focus+context* technique, as it shows the whole graph at once, with focus on the selected component and related components.

3.3.2 Visualisation of open student models

Some forms of representing cognitive aspects of students have been explored in open student models. As already mentioned in Section 2.5, student models are a key component of intelligent educational systems used to represent the student's understanding of material taught. Recently student models are made available to students and tutors for inspection, tuning, and to promote reflection in the learning process (for the learners) and in teaching (for the instructors). Especially for students, in recent years research has begun to address the impact of allowing learners to inspect the view that the system maintains of them. In this case there is a challenging representation problem, because open student models have to represent complex information with clarity and

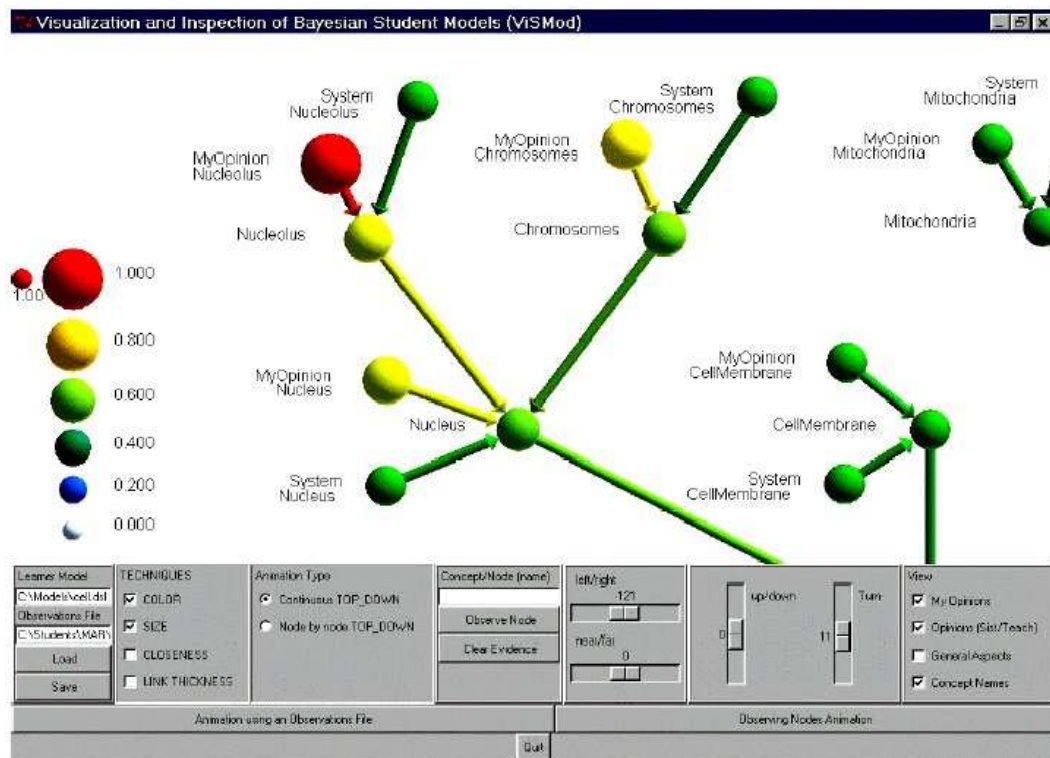


Figure 3.19: A screenshot of ViSMod showing a fragment of a Bayesian student model in the area of biology cell. In ViSMod, both student beliefs and system beliefs have been associated to each of the concepts of the network (e.g. knowledge on *Chromosomes* is influenced by the student's personal opinion indicated in *MyOpinionChromosomes* and by the system/instructor's opinion described in *System Chromosomes*). The control panel on bottom allows the user to manipulate the view of the model. The image is taken from Zapata-Rivera and Greer (2001a)

simplicity. We report here examples of open student models that have exported graphical representations to externalise a student model built by the system (based on some Artificial Intelligence inference).

ViSMod

ViSMod (Zapata-Rivera and Greer, 2001a,b) is an interactive visualisation tool for the representation of Bayesian learner models ⁷. In ViSMod learners and instructors can inspect the learner model using a graphical representation of the Bayesian network. ViSMod uses concept maps to render a Bayesian student model and various visualisation techniques such as colour, size prox-

⁷Bayesian learner models are one approach to model student's knowledge through Bayesian networks. These networks probabilistically (based on learner interactions with the system) assess the learner's knowledge state. Each Bayesian network node has a probability determining learner knowledge concerning this piece of subject (Chec, 2003).

imity link thickness, and animation to represent concepts such as marginal probability, changes in probability, probability propagation and cause-effect relationships. One interesting aspect of this model is that the overall belief of a student knowing a particular concept is captured taking into account the students' opinion, the instructors' opinion, and the influence of social aspects of learning on each concept.

By using VisNet it is possible to inspect complex networks by focusing on a particular segment (e.g. zooming or scrolling) and using animations to represent how probability propagation occurs in a simple network in which several causes affect a single node.

KERMIT/E-KERMIT/NORMIT

KERMIT (Knowledge-based Entity Relationship Modelling Intelligent Tutor)(Suraweera and Mitrovic, 2002) is a Knowledge-based ITS aimed at teaching conceptual database design at university level students. KERMIT teaches the basic Entity-Relationship (ER) database modelling by presenting to the student the requirements for a database, and the student has to design an ER diagram for it. E-KERMIT is an extension of KERMIT developed by Hartley and Mitrovic (2002) with an open student model. In E-KERMIT the student may examine with a dedicated interface the global view of the student model (see Figure 3.20). The course domain is divided in categories, representing the processes and concepts in ER modelling. In the representation of the open student model concepts of the domain are mapped with histograms. The histogram shows how much of the concrete part of the domain the student knows correctly (in blue) or incorrectly (in grey) and the percentage of covered on the concepts of the category. For instance, the example shows that the student covered 32% of the concepts of the category *Type*, and has scored 23% out of a possible 32% on this category. This means that the student's performance on category *Type* so far is 77% ($23/320 * 100$).

NORMIT is another ITS implemented by Mitrovic (2003) that teaches data normalisation in database design⁸. It is a problem-solving environment that provides help about the basic domain concepts when the system deduces that the student does not understand them or the student has difficulties in applying the knowledge acquired. NORMIT was integrated with an interface that opens the student model to the learner (see Figure 3.21). Progress bars indicate how much of the domain the student has not covered yet (shown in white), and how much of the covered part of the domain the student knows correctly (shown in light grey), or incorrectly (shown in dark grey).

⁸Data normalisation is the process of refining a relational database schema in order to ensure that all relations are of high quality (Elmsari and Navathe, 2001)

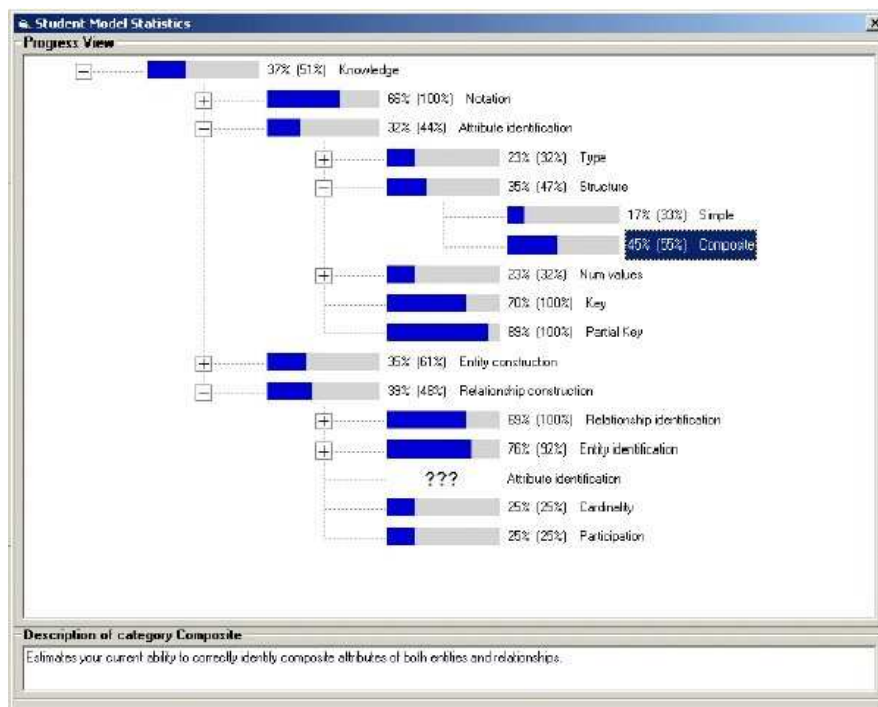


Figure 3.20: The main view of a student's progress in E-KERMIT. Progress bars indicate how much the student comprehends each category of the domain. The image is taken from Hartley and Mitrovic (2002).

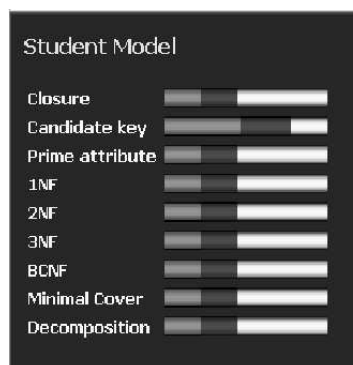


Figure 3.21: A screenshot of the open student model visualised by NORMIT. In the progress bars the white part indicates how much of the domain the student has not covered yet, the light grey part indicates how much of the covered part of the domain the student knows correctly, and the dark grey part indicates how much of the covered part of the domain the student knows incorrectly. The image is taken from Mitrovic (2003).

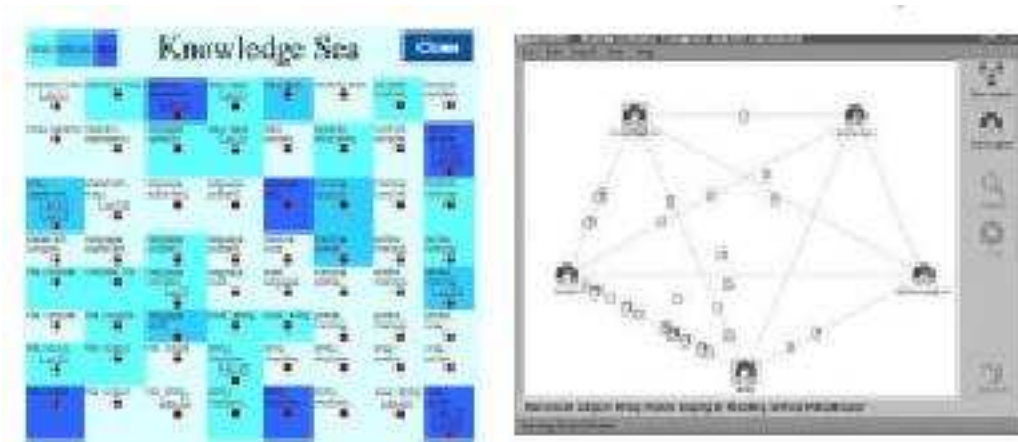


Figure 3.22: The visualisation of KnowledgeSea (on the left) and WebVIBE (on the right). Image from Willms (2003).

KnowledgeSea and WebVIBE

Another example of a graphical representation of a student model is proposed by Willms (2003). It uses a spatial visualisation system, called KnowledgeSea and WebVIBE (Figure 3.22) to help students to more effectively locate relevant documents to study based on their interests and knowledge acquired. Its basic idea is that *“the student is able to find relevant documents faster when they have a visualization of their interest and knowledge available ... “* and *“... students are more likely to make use of external material when they can judge its relevance with the help of visualization”*. Students’ interest is computed by asking them to highlighting part of a document that is relevant to them, and the knowledge is computed on test scores. The user model consists of components which include the student’s interest and his knowledge called Point of Interest. Figure 3.22 gives an example of the representations used. The WebVIBE graphical representation (Figure 3.22, right) represents Points of Interest by magnets as a metaphor of attraction on a virtual document space. The user can interactively select and place the Point of Interest on the document space. The knowledgeSea graphical representation (Figure 3.22, left) displays a map of the educational documents, with a set of keywords corresponding to each cell that describe the documents in this cell. The users use the WebVIBE display to see how each document is related to the others with respect to the Point of Interest.

Bull’s and Nghiem’s General Learning Environment

Bull and Nghiem (2002) reported a work in progress on a General Learning Environment designed to help learners to better understand their learning and instructors to support students in their learning with an open student model. In this work: (a) the model was intentionally very simple to



Figure 3.23: The Bull's and Nghiem open learner model presented in two forms: tabular (top-left window) and graphical (bottom-left window). Image from Bull and Nghiem (2002).

be easily and quickly deployed in a variety of course type; (b) it is one of the few works where the student model has been designed to be open to students, peers, and instructors. The learner model is open to students to encourage reflection, to peers to help them compare their performance to that of peers in their group, and finally it is opened to instructors to better understand their students, help individuals with particular problems, or to target their teaching to the general difficulties of specific groups. Figure 3.23 illustrates an example for the domain of Japanese particles. The right window contains a test where the students have to select the grammatically correct particle for each sentence. The left windows contain the estimated level of knowledge of different concepts of the domain represented in tabular and graphical format. The graphical version gives at a glance a comparison of correct versus incorrect attempts at questions, and indicates the student's overall performance.

Bull and Nghiem explored also the possibility to enable learners to compare their own student model with that of other learners. Some researches suggest this extension in order to give learners the possibility to retrieve good peers, appreciate the level of knowledge gained and realise whether they are going well (or badly), and promote peer tutoring (Bull and Nghiem, 2002; Kay, 1997; Bull and Broady, 1997). Figure 3.24 from Bull and Nghiem (2002) gives an example of this.

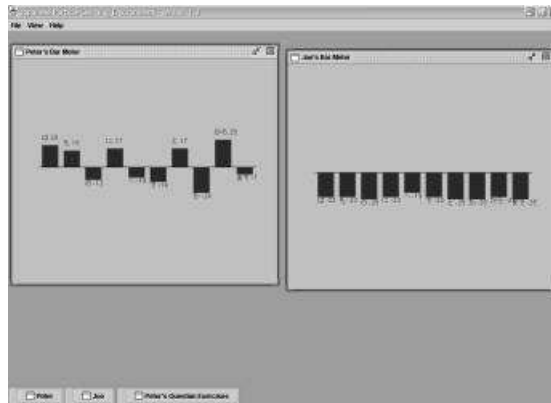


Figure 3.24: The comparison of two individual student models in the Bull's and Nghiem open learner model. Image from Bull and Nghiem (2002).

Chaatra

Chaatra is a student monitoring and learner modelling system implemented by Rehani and Sasikumar (2002). It uses students' responses to questions testing multiple concepts to evaluate the level of understanding of students in each concept of the domain. Each quiz question is mapped to a subset of concepts through a "question-concept map", and the level of knowledge acquired by a student on a concept is computed with a weighted sum of marks received in quizzes questions. The output is shown in Figure 3.25. Based on predefined thresholds understanding values on concepts are classified "good", "medium (need improvement)", "weak", or "not conclusive" (the latter in case of the number of answered questions is below a threshold and thus assigned an unreliable measure). This display can be used by learners to view the estimate of their understanding or to compare their understanding with the average class understanding for each concept.

3.3.3 Visualisation of on-line communications

One important aspect that should be considered in distance education is the social one. There are various types of interactions that may occur in distance learning, such as interactions between individual students and interactions between students and the instructors. In Internet, the main tools that engage students and instructors in communicative activities are discussion forums, e-mail and chat. Some past works attempts to visualise the communications exchanged in a educational settings to discover patterns, roles, and engagement of students in social activities.

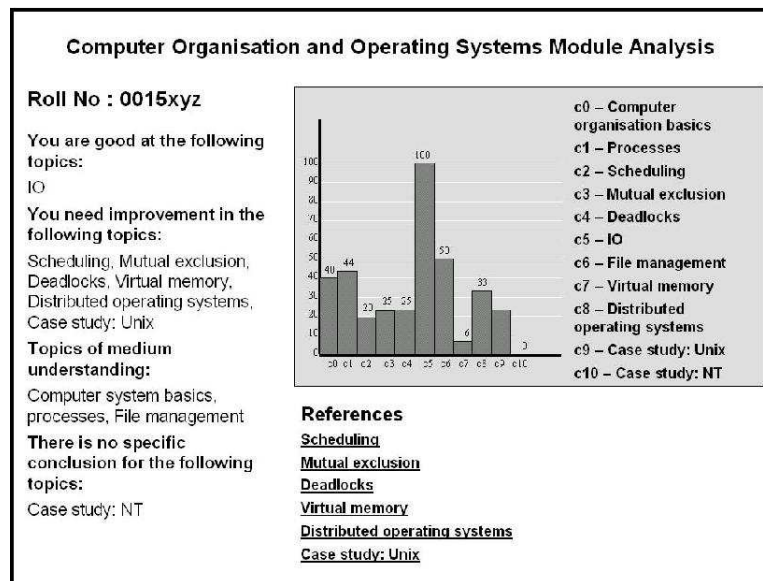


Figure 3.25: The graphical display of student understanding displayed by Chaatra. Image from Rehani and Sasikumar (2002).

Simuligne

Simuligne (Reffay and Chanier, 2002) is a research project that uses Social Network Analysis (Scott, 1991) to monitor group communications in distance learning in order to help instructors detect collaboration problems or slowdown of group interactions. Social Network Analysis is a research field which “characterise the group’s structure and, in particular, the influence of each of the members on that group, reasoning on the relationship that can be observed in that group” (Reffay and Chanier, 2002, p. 34). It provides both a graphical and a mathematical analysis of interactions between individuals. The graphical version can be represented with a network, where the nodes in the network are the individuals and groups while the links show relationships or flows between the nodes. The Social Network Analysis can help to determine the prominence of a student respect to others, and other social network researchers measures, such as the *cohesion* factor between students. The cohesion is a statistical measure that represents how much the individuals socialise in a group which shares goals and values. Reffay and Chanier applied this theory to a list of e-mails exchanged in a class of 40 distance learners, split into four learning groups (Aquitania, Lugdunensis, Narbonensis, and Gallia). Figure 3.26 illustrates the graphical representation of the e-mail graph for each group⁹. We can see for instance that there is no communication with An2 and An3 in the group Aquitania, or the central role of L15 in the Lugdunensis group. Other numerical measures, namely the cohesion factor, is computed starting from these graphs.

⁹The instructor, and all the messages he is concerned with, are not considered in these graphs.

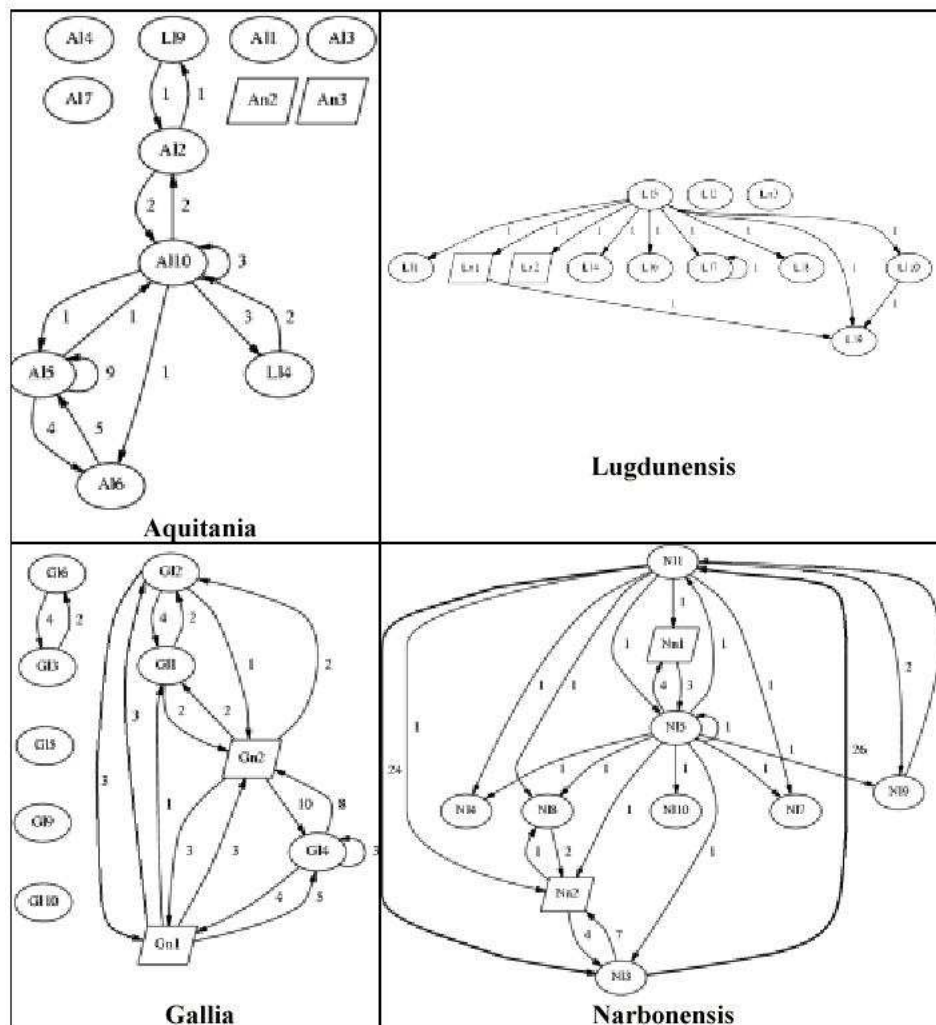


Figure 3.26: The communication graph of the e-mail exchanged within groups in Simuligne. Image from Reffay and Chanier (2002).

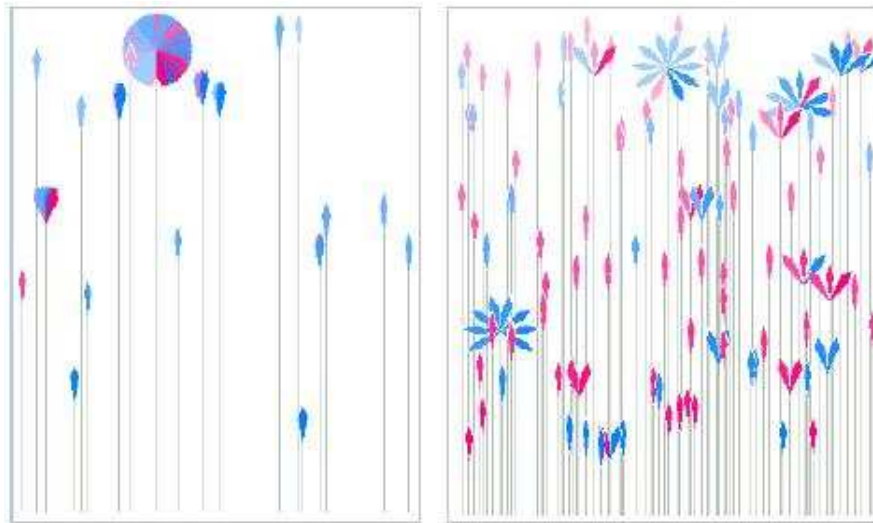


Figure 3.27: The PeopleGarden visual representations of participation on a message board. Image from (Xiong and Donath, 1999).

PeopleGarden

PeopleGarden (Xiong and Donath, 1999) uses a flower and garden metaphor to visualise participations on a message board. The message board is visualised as a garden full of flowers. Each participant is represented by a flower. The height of flower denotes amount of time a user has been at the board, and its petals his postings. Initial postings are shown in red, replies in blue. An example is represented in Figure 3.27. The figure can help the instructor of a course to quickly grasp the underlying situation, such as a single dominant member in discussion on the left or a group with many members at different level of participation on the right.

To summarise, a number of educational systems apply some Information Visualisation techniques to graphically render data about students. The survey shows that most of the projects consider one graphical representation to render data collected by a specific application. The graphics are selected somehow ad-hoc, there is a lack of systematic studies of what information is needed, as well as what graphics are appropriate and when. In contrast, such a systematic approach is proposed in Information Visualisation research but has not been directly followed in educational systems. Due to their complexity, the visualisation tools in educational software are mostly at a prototype stage, and there is a lack of empirical studies of the appropriateness and usefulness of the graphical representations. There is no use made of existing visualisation software in educational systems, which will speed up the development process, facilitate the deployment in real life situations, and enable evaluative studies with users. This thesis embarks upon the challenge of ap-

plying Information Visualisation techniques to render graphically student tracking data collected from CMS used in Web-based distance courses. A systematic approach will be followed based on IV research, several graphical representations will be produced and their appropriateness will be examined with real users.

3.4 Summary

This chapter illustrated some IV techniques that can be used to build graphics that may help to understand complex multidimensional data and to amplify cognition. Some of these techniques can be used to create a graphical representation of students' data collected by CMS that might help instructors in monitoring student learning in order to provide learners with appropriate and prompt feedback when a problem occurs. Graphical representations of students' data have already been explored in several educational environments, particularly in ITS, but these are external representations of student models based on some Artificial Intelligence inference. However, there are no systematic investigations on what sort of information instructors are interested in and how this information could be presented. IV instead is powerful in providing insights from raw data with minimum data processing. By managing students' tracking data with appropriate visualisation techniques instructors could interpret data and infer useful information that may help instructors in managing the class. The innovative idea of this proposal is that student models are inferred in the instructor's mind, instead of being inferred by algorithms. Next chapter will investigate by the means of a survey the instructors' needs when teaching at distance, and the results will be used to inform the design of a course data visualisator that will be built taking into consideration the instructors' requirements.

Chapter 4

Gathering Instructors' Requirements

4.1 Introduction

As stated at the beginning, the main goal of this work is to produce a system that generates a set of graphical representations of students' information from tracking data collected by a CMS. The first step in developing usable systems is gathering design requirements. This process is usually based on identifying the needs for the system and includes studying characteristics of the work environment and the performed tasks of the target user group (Dix et al., 1998). This chapter presents an empirical study that investigates the instructors' needs when teaching at distance with the help of a CMS. We submitted a survey, which is a common technique for understanding the needs of a broad base of potential users, to distance learning instructors. The objective of the survey is to identify what kind of student information instructors require in their activities and to inform the design of a system to help instructors to gain understanding of their students (including both individual and group of students). To do so, the system will fetch information from the CMS and will presents it in graphical format.

This chapter presents a summary of the major findings from a questionnaire. Based on the results of the study, we will draw recommendations how to help teachers gain a better understanding of cognitive, social and behavioural aspects of individual students in on-line distance teaching. These requirements are used to inform the design of CourseVis, which is discussed in Chapter 5.

4.2 Objectives

The study aims at exploring what kind of information is required by instructors during the running of course delivered at distance on the Web. Our objective is to find out which sort of information is required by them, and then to implement suitable visualisation techniques. We submitted a questionnaire to people involved in distance learning to address the following themes:

- *Social aspects*, which concern communicative activities that students are engaged in. There are various kinds of interactions that occur in a Web-based course run with a CMS, such as interactions between individual students, interactions between students and the instructor, interactions between the students and the system, and vicarious interactions. Vicarious interactions take place when a student actively observes and processes the interaction between other students or between other students and the instructor, without taking an active part in the interaction (Sutton, 2001).
- *Cognitive aspects*, which concern the students' understanding of the course material. Cognitive aspects are usually monitored in distance learning by considering the students' overall course performance, their results on a selected evaluation proof (e.g. quiz or assignment), the students' familiarisation with a specific topic of the course, etc.
- *Behavioural aspects*, which are behavioural indicators to judge factors such as active learning, motivation, engagement, and, in general, to assess the success or failure of a distance course. These concern the students' course attendance, students performing well or badly, students who are progressing too fast or too slowly with the course schedule, etc.

4.3 Method

4.3.1 Participants and setting

The participants in this survey were instructors involved in distance learning courses, all together 98 responses were processed. We contacted people involved in the Swiss national research program "Swiss Virtual Campus¹", the ICLT group² in Leeds, the IFETS mailing list³, some people that participated in past seminars organised by the Global Educators' Network of the Canadian TeleLearning Network of Centres of Excellence (TL*NCE)⁴, and, finally, some people involved

¹The Swiss Virtual Campus is a programme of the Swiss Confederation entitled under the 1999 Swiss Federal Law on University Development to support new information and communication technology in higher education. At the program participates almost each Swiss University with 50 projects and an overall grant of over 30 million Swiss francs. In general, each project should develop a course that can be followed via the Internet that includes teaching material, exercises, seminars or practical work as well as on-line or direct aids and assessment. Additional informations in <http://www.virtualcampus.ch>

²ICLT stands for "*Interaction, Communication and Learning Technologies*", this group brings together people with a common research interest in the interactions that occur between people and between people/computers when technology is used for learning and teaching. More information in <http://cbl.leeds.ac.uk/~aisha/iclt/home.html>

³International Forum of Educational Technology & Society (IFETS) is a subgroup of the IEEE Computer Society Learning Technology Task Force (LTF). The purpose of this task force is to "...contribute to the field of Learning Technology and to serve the needs of professionals working in this field". Details about LTF and IFETS can be found in <http://lttf.ieee.org> and <http://ifets.ieee.org>

⁴The Global Educators' Network (GEN) is "... an international community of on-line educators aimed to encourage educator exchange and leverage experience and expertise in e-learning models and methods". The main activity of

as instructors in on-line courses in the UK and in Switzerland and whom we know personally.

4.3.2 Procedure

The questionnaire was prepared in a Web form and located at a Web server in Lugano. An e-mail message was sent out to 188 individuals and to the IFETS mailing list. The e-mail explained the purpose of the study, encouraged recipients to respond to the survey on a voluntary basis, and provided the address of the Web page with the questionnaire. The access to the questionnaire was very simple and required only an Internet browser.

The questionnaire was comprised of 17 questions. Almost all questions were in a multiple choice format and respondents answered by clicking on the choice that corresponded to their response. Some of the questions also had a small open-text box where respondents were enabled to specify in details other items not covered by the set of choices provided in the question. Participants were allowed to write general comments about the study. They could also request a copy of the results, and in this case they had to provide an e-mail address.

The responses to the survey were sent anonymously⁵ to the Web server, where a small program in charge of receiving the responses and putting the answers in a database was implemented. The responses were collected for two weeks from December 2nd 2002 until December 15th 2002, and then the Web page containing the questionnaire was replaced with a page indicating the end of the survey and providing the e-mail of the author.

In addition, we conducted some discussions with distance teachers in Leeds to uncover aspects that had not been captured by the questionnaire.

4.3.3 Measurement

The questionnaire was structured across four major categories:

- *User related questions*, asking the type of involvement in on-line courses, the length of this involvement, and the size and type of the classes.
- *Platform related questions*, asking some information about the technological platforms used and the facilities provided by these platforms.
- *Students and assessment related questions*, that aimed to identify what assessment techniques were utilised by the instructors and how the assessment information was used in

GEN is to promote monthly on-line seminars on topics related to learning and teaching on-line. The Global Educators' Network can be found at <http://vu.cs.sfu.ca/GEN/welcome/welcome.html>

⁵It should be mentioned that we collected the IP addresses of the submissions. This information was used to have an overview of the countries from where the questionnaires were submitted (see B.1 in the appendix), and never used to identify the submitter.

their teaching activity. In addition, some questions addressed the information about the students the instructors demanded.

- *Feedback related questions*, a couple of open text boxes where participants could provide comments and feedback and another one to type their e-mail address in case interested in receiving a summary of the results.

A copy of the questionnaire submitted is given in Appendix A.

4.4 Results

The analysis that follows presents the key issues and results extracted from the collected responses. Appendix B shows more details about the responses to each question.

4.4.1 Response rate

At the cut-off date (December 15th 2002) we received 98 answers, among which two had posted some comments but did not answer specific questions. If we consider that we sent to the IFETS mailing list that had 4210 members in December 2000, the response rate was:

$$\frac{98 \text{ respondents } (-2 \text{ not considered})}{188 + 4210 \text{ surveyed}} = 2.18\%$$

Considering this low response rate, we must take into consideration that it is not often the case that people are prepared to put effort in surveys and dedicate 10-20 minutes of their time to answer questions when they would not have any immediate gain from this. Moreover, it must be stressed that in a wide mailing list with thousands of members, only a small fraction of them may be interested in the topic of the survey. Besides, very often e-mail addresses provided are old and no longer valid.

4.4.2 Survey reliability

The question that arises here is: are the respondents to the survey a representative set of all the users of the system we are designing? There was a relatively large number of respondents from at least 12 different countries (see Appendix B.1). The survey revealed that the majority of the respondents were instructors (Appendix B.3) involved in undergraduate and postgraduate university courses, and continuous education or adult courses as well (Appendix B.4). Also the majority of respondents had experience of more than 1 year in on-line courses (Appendix B.6).

However, it must be pointed out that the majority of respondents were from Switzerland and the United States (e.g. none of the respondent was from an Asian country, results are therefore biased toward the experience of European (Swiss) and American (USA) tutors). We ought to

acknowledge that instructors in the sample were not randomly selected, therefore results of this survey will not be extendable to the whole population of instructors of distance courses. The sample selected is a so-called “reasoned sample”, by which this survey will draw some tendencies of participants, but the data from the survey may not be fully representative of the views of all instructors involved in on-line teaching, which is out of the scope of this work. The findings discussed below provide an interesting range of responses that can be used to extract some trends, rather than being a definitive picture of all distance learning instructors.

4.4.3 Analysis of responses

In this section we will give a summary of the figures extracted from the survey, extended tabular data is given in Appendix B. The topics are grouped by the four major categories that closely follow the objectives of the survey. In the next section, we will discuss the findings and will give some recommendations for the design of CourseVis. Percentages given in this paragraph are rounded to the nearest integer, exact values are reported in Appendix B.

Participants' profile

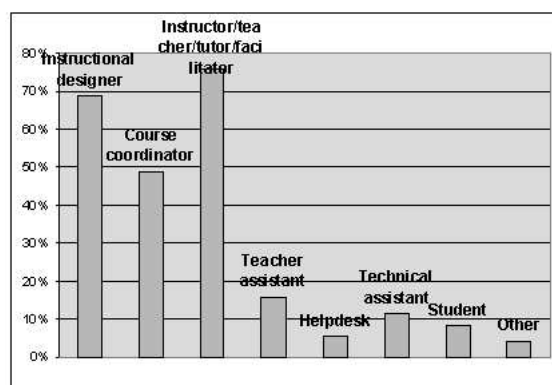


Figure 4.1: Type of involvement in courses for participants in the survey

The majority of respondents are classified as instructors or instructor related (76%), course coordinators (48%) or instructional designers (69%) (see Figure 4.1). Less than 16% were classified in other categories like teacher assistants or helpdesk. These results give us some confirmation that the population addressed in the survey could be identified as the desired class of users for CourseVis.

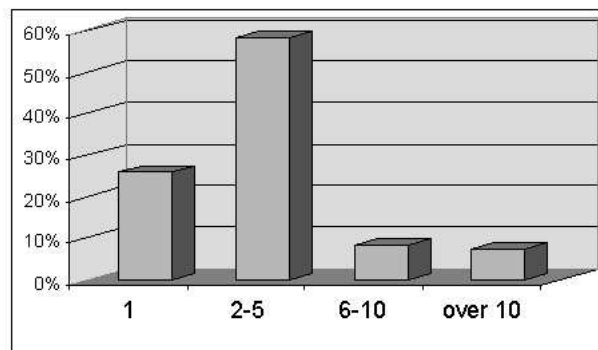


Figure 4.2: Number of courses managed by participants of the survey

Most of the respondents are involved in two or more courses (74% - the sum of the last 3 columns in the graph in Figure 4.2), but only the 15% of them are involved in 6 or more courses. The 26% are involved in just one course.

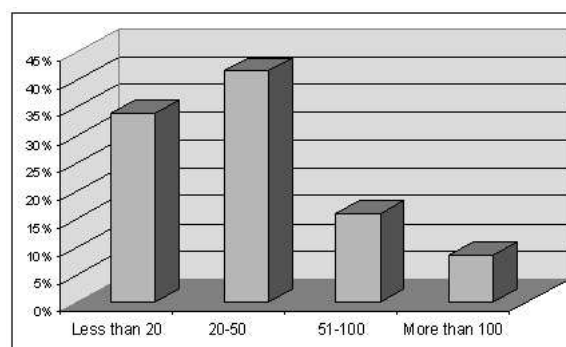


Figure 4.3: Number of students in the largest class

Classes are composed in 75% of the cases by less than 50 students (considering the first 2 columns in the Figure 4.3), 16% of the classes involve between 50 and 100 students. Only 8% of the classes are composed by more than 100 students.

Platform used

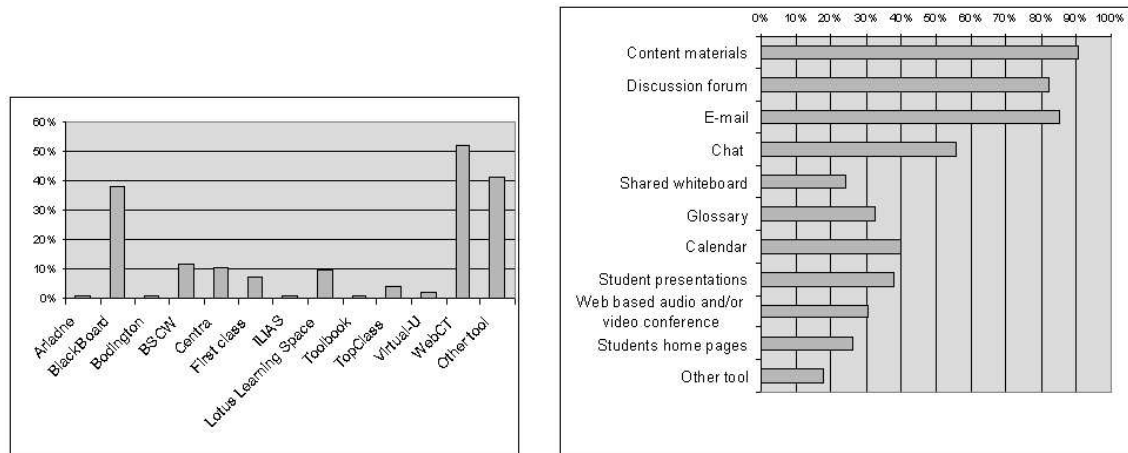


Figure 4.4: Educational platforms and their facilities used by instructors

The vast majority of platform used were WebCT and BlackBoard (see Figure 4.4 left). This two combined together give about the 90% of the overall platform usage. We also asked about the utilisation of various facilities provided by the Course Management Systems used. The results where that predominately used are the content materials, discussion forums, e-mail, all of them have been used by about 82% of the respondents. Chat is popular with a half of them (56%), while other tools are occasionally used with a ratio of less than 40%.

Assessment data used

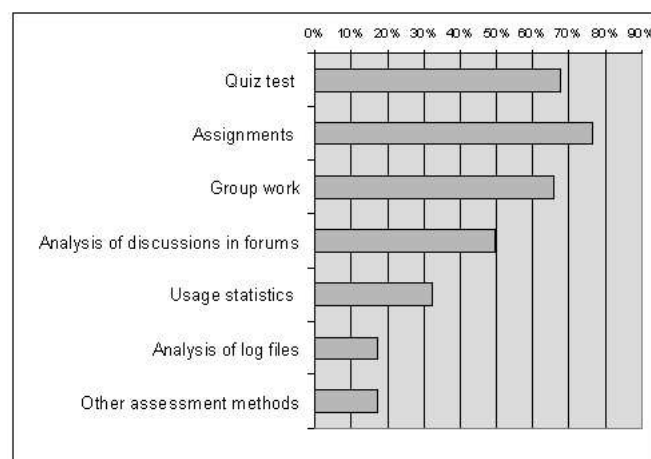


Figure 4.5: Assessment techniques and tool used in courses

The questionnaire revealed that the most used assessment tools were quiz, assignments and group works, between 65% and 76% (Figure 4.5). The analysis of discussions in forums is less used, however, it has a relevant percentage of usage (49%).

We asked instructors which information regarding the students they would like to have, and how this information could be used in their teaching activities. They had to specify the level of interest in some evident issues, such as the access to the course, the participation in discussion, etc. In order to specify the level of interest, we provided a categorical scale to this question, having the following values: “*Extremely interesting*”, “*Very interesting*”, “*Interesting*”, “*Somewhat interesting*”, “*Not at all interesting*” and “*I don’t know*”. Moreover, we left to the respondent the possibility to specify further issues not covered by the questionnaire, using a free text box. Detailed results can be found in Appendix B.10. Here we restrict our analysis splitting the categories of issue in four different classes: issues rated by the respondents as “*Extremely interesting*” or “*Very interesting*”, issues rated by the respondents as “*Interesting*” or “*Somewhat interesting*”, and finally other classified using remaining categories, i.e. “*Not at all interesting*” and “*I Don’t know*”.

	Access to the course	Quiz and assignment grade	Participation in discussions	Posting e-mail to colleagues	Frequency of re-visiting of the same page	Participation in chat	Participation in group exercises
Extremely interesting + Very interesting	54,74%	65,26%	65,26%	33,68%	26,32%	27,37%	67,37%
Interesting + Somewhat interesting	33,68%	25,26%	23,16%	44,21%	37,89%	42,11%	21,05%
Not at all interesting	4,21%	5,26%	5,26%	13,68%	22,11%	16,84%	4,21%
I don't know	3,16%	3,16%	2,11%	4,21%	3,16%	10,53%	7,37%

Figure 4.6: Information about students instructors would be interested to have

As shown in Figure 4.6, the following issues:

- Access to the course
- Quiz and assignment grade
- Participation in discussions
- Participation in group exercises

have been rated by the respondents as highly desirable (*Extremely interesting* or *Very interesting*), while the following issues:

- Posting e-mail to colleagues
- Frequency of revisiting the same page

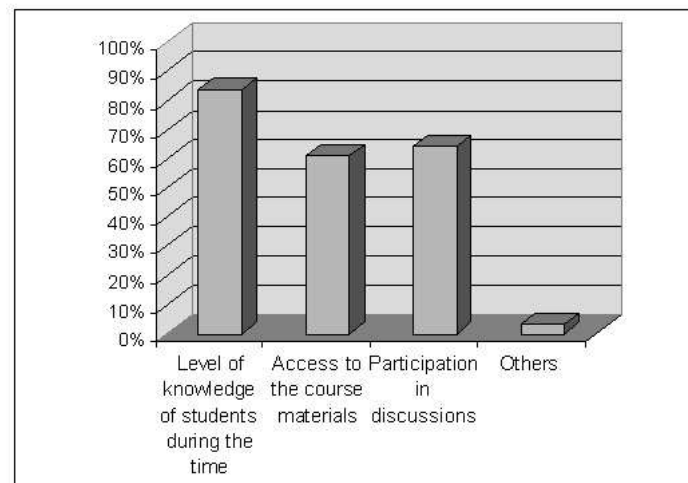


Figure 4.7: Interest in time-dependent information

- Participation in chat

have been rated in the second level of interest (*Interesting* or *Somewhat interesting*).

The additional information provided by the respondents given in the free text box (see question B.11 in the Appendix B) reveals that some participants underlined the importance of the qualitative analysis of postings.

Regarding what kind of time-dependent information about students instructors were interested in, responses (see Figure 4.7) revealed that most respondents were interested in the students' level of knowledge during the time (84%), but also a good percentage were interested in the access to the course materials (61%), as well the participation in discussions during the time (64%).

In the informal discussions some instructors pointed out that it could be interesting for them to have the possibility to see the concepts taught related to information about the students, e.g. the level of understanding of the students for a specific concept or if students read the course materials concerning that concept.

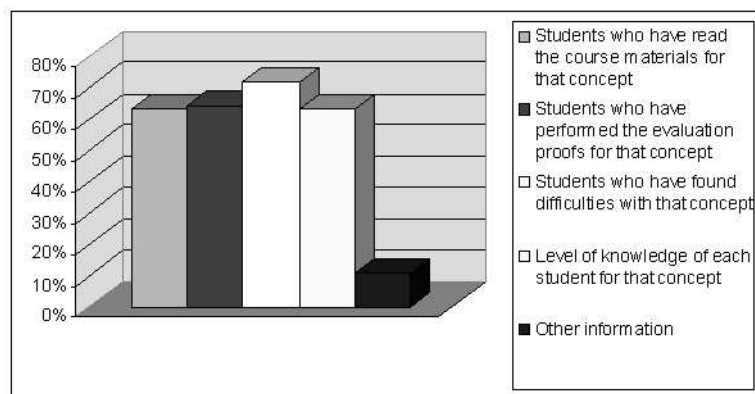


Figure 4.8: Interest in information related to concepts of the domain of the course

With regard to information related to concepts, the responses were equally distributed in the list of provided answers (see Figure 4.8), showing that instructors are interested in students who have read the course materials for a concept, students who have performed the evaluation proofs for a concept and the level of knowledge of each student for that concept. The reading of course materials and the level of knowledge of students for a concept received were regarded as less interesting compared to the knowledge of students having difficulties on a concept, which was identified as the most interesting.

Good tutoring practice usually requires monitoring learners on a regular basis, and this may be valid for distance learning as well. We therefore asked which type of students should be monitored.

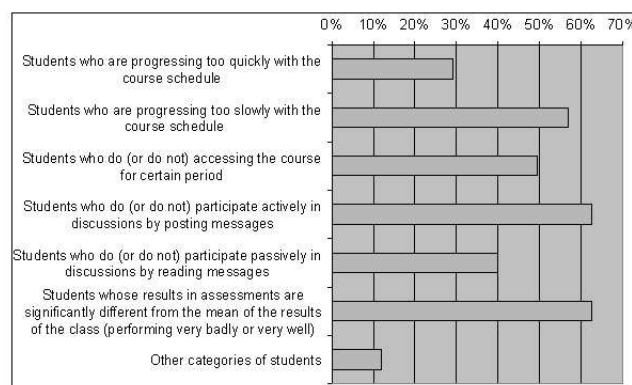


Figure 4.9: Type of interesting students to monitor

Answers (see Figure 4.9) showed an emerging interest in the participation in discussions (62%), confirmed also by comments placed by some respondents in the free text option, and in students whose results in assessments are considerably diverting from the mean (62%). Students

who are progressing too slowly in the course have been of interest to the instructors, as well (57%).

It could be worth understanding why the participants to the survey decided to give a particular answer to a question. This is not easy to understand simply with a questionnaire. We need to find a way to infer the reason for an answer from the respondent. Therefore, participants were asked, assuming that they had all desired information about their students, how would they use that information for their teaching activities.

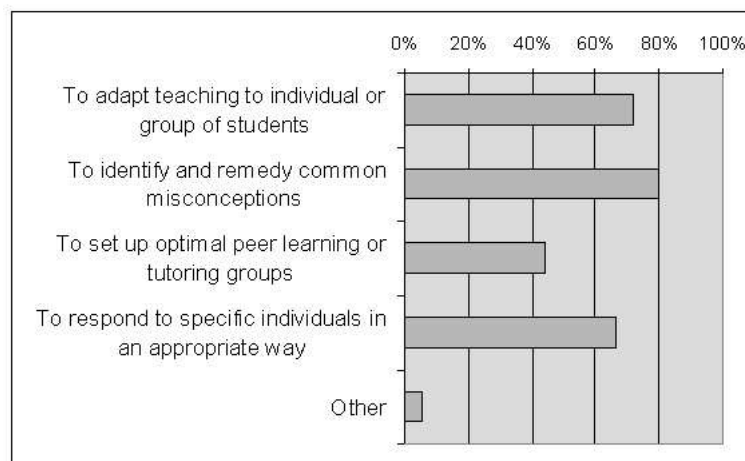


Figure 4.10: How instructors want to use students' information

80% of the respondents answered that they wanted to use this information to identify and remedy some common misconceptions that could arise in the class (Figure 4.10); more than 71% of the participants might use this information to adapt their teaching and to provide individualised feedback; 66% of the respondents thought that this information could be useful to reply to specific individuals in an appropriate way.

Questionnaire feedback

We received feedback about the questionnaire from 25 participants, see Appendix B.17. Eight of them were disregarded as interesting, e.g. *"I can't answer to your questionnaire"* or *"Thanks and good luck"*. However, some participants provided very valuable feedback, such as:

"The personal and social side of student interaction between tutor and peers is crucial to forming good relationships - often ignored in research into cognitive development - also how do they feel about the learning environment and relationships"

The personal and social aspects of student interaction is a key aspect that must be considered in every educational environment, and this is especially valid in distance education. Probably the par-

ticipant noted above was referring to some aspects of learning like “emotions” or “motivations”, explored in some recent works (Conati and Zhou, 2002; De Vicente and Pain, 2002), and represents the extreme frontier in educational research. However, the analysis of this aspects would require qualitative analysis of dialogue, which is out of the scope of this work. Qualitative analysis of discussions provides deep insight into social aspects in distance classes but is usually laborious and time consuming. Instead, by using suitable visualisation techniques, quantitative analysis can be performed to discover general tendencies and phenomena about social aspects of students as well as to highlight parts of the interaction for further qualitative analysis.

Another participant questioned the usefulness of vicarious interactions:

“I am a foreign language teacher, and interaction is fundamental for language acquisition, that is why vicarious interaction is not the kind of interaction I would promote in my courses.”

This comment points out that vicarious interaction may not be useful in some domains, even though they it represent one of the possible interactions in educational systems. However, this could not be generalised for all domains. Indeed, vicarious interactions were considered important by many respondents (see summary of the question in B.16). The question above also stresses the importance for instructors to monitor who is actively participating in discussions which in some domains, such as language learning, is crucial.

4.5 Discussions with distance teachers

A questionnaire alone is not able to collect all relevant aspects and requirements that can be needed by distance tutors who may use CourseVis. To uncover aspects that had not been captured by the questionnaire, we conducted several discussions with distance teachers in Lugano, and presented our ideas with an early prototype of CourseVis to the ICLT group at Leeds. Discussions where conducted informally, here we outline some aspects discussed in these conversations which have not been covered by the questionnaire.

The most argued topic was the role of discussions in distance learning. Teachers argued that discussions have a fundamental role for a successful on-line course. This is also confirmed by the recent research interest in analysing social aspects of discussions in distance learning and, more generally, in every kind of computer mediated communication modality (Pilkington, 2001). There was a consensus that a textual analysis of the communication exchange between students had to be performed by the instructors. Participants stressed the difficulties in implementing automatic systems dedicated to support instructors in dialogue analysis. One of them even suggested to count the number of words in postings as a simple indicator of the degree of involvement in discussions, although the benefits of such an approach are debatable.

An interesting comment was made about the opportunity to compare data from different runs of the same course. CMS give the opportunity to reuse the content when appropriate. This enables teachers to reuse content material and to quickly develop new versions of a course by maintaining the same global structure. It was argued that comparing students' performance on specific topics of the course in different classes (if available by the system) may hint the teacher to improve specific parts of the course where students face certain difficulties.

4.6 Discussion and recommendations

In line with the objectives of the study, we will discuss three themes - social, cognitive and behavioural aspects which tutors are interested in. We will identify requirements of distance teachers related to each of these themes and will give some recommendations for the implementation of a tool for externalising appropriate information to instructors.

4.6.1 Social aspects of learners

The main tools in an on-line course to engage students and instructors in communicative activities are discussion forums, e-mail and chat. The former two are used by more than 80% of the respondents in the study, while chat is popular only with a half of them (Figure 4.4 right). Group work is also an interesting opportunity for students to cooperate together and socialise, and is used by 66% of respondents (Figure 4.5). This facility is fundamental for the creation of a successful on-line course.

Recommendations

A tool that represents the students' social activities in the course should consider the data provided by the communication facilities included in that course. The participation in discussions is considered as the most interesting and valuable in the evaluation of both the course and the students' activities. Several respondents made comments about the usefulness of the qualitative analysis of discussions, but dialogue analysis is laborious and, at the current state of research, is done mainly manually by humans. Our recommendation is that appropriate quantitative visualisation of discussions data should be provided. The instructor can use this information as an indicative element to decide which parts of the interaction have to be analysed qualitatively and to understand further certain social aspects in distance classes. Data from other communication tools (e-mail and chat), even if very common in on-line courses, are rated with a low interest for the teaching purposes. We, therefore, will not recommend including these data in CourseVis. The participation in group work is also crucial for the instructors and appropriate visualisations of data about this kind of activity should be considered.

4.6.2 Cognitive aspects of learners

The most interesting aspect for instructors appears to be the cognitive one. This is probably valid both in distance and in face-to-face education. Students enroll in a course because they want to learn the concepts of the course. A good teacher has to check that this has been accomplished in their course and, if not, to remedy by adjusting the teaching or adapting the material to the students.

Recommendations

It is highly recommended to give a suitable external representation of the students' performance in the course. The survey indicated that the assessment techniques typical of on-line courses (quiz, assignments and group work) were used by the majority of the respondents (Figure 4.5). The participants in the questionnaire expressed a high interest in having information about the overall performance in the course as well as the level of knowledge achieved by each student for each domain concept of the course (Figures 4.7 and 4.8). The importance of this information was also stressed by participants in the discussions we had. Almost all instructors wanted to use this information in order to identify and remedy learners' common misconceptions in their courses (Figure 4.10). A visualisation system should provide instructors with a clear and immediate external representation of students having difficulties with a concept. Moreover, this representation should carry out a comparison of a student with the whole class, as desired by the majority of respondents (Figure 4.9).

4.6.3 Behavioural aspects of learners

Teachers tend to look at behavioural indicators and use them to judge the factors like active learning, motivation, engagement, and in general to assess the success or the failure of a course. With computer mediated communication some of these indicators are difficult to analyse, others are not relevant. We asked the participants in the survey about the various behavioural indicators distance teachers require.

Recommendations

One of the most needed information that should be represented externally in a course is the students' access to the course (Figure 4.6). Akin face-to-face teaching, poor course attendance is considered as a symptom of a possible problem with the learners. Other information that should be represented is the reading of materials (Figure 4.8), e.g. access and time spent, the performing of evaluation proofs (Figure 4.8) and the participation in discussions (Figure 4.6), as already mentioned. Some instructors use these activities as basis for judging the mandatory participation in the

course⁶. Moreover, some indication of the students' progressing with the schedule of the course should be provided (Figure 4.9).

4.7 Implication for the research and practice

This survey gave some useful indications about what kind of information is required by distance teachers. These indications have arranged a set of recommendations that will be considered in the implementation of CourseVis. Our next step is to exploit visualisation techniques to present this information to instructors in an appropriate manner that facilitates the required understanding.

In particular, CourseVis will be implemented taking into consideration the following issues derived by the recommendations. We give a list of them and, for each one, the list of questions from the questionnaire which have suggested it.

4.7.1 About the course manipulation

Distance teachers usually have been involved in more than one on-line course (see Figure 4.2). For this reason CourseVis has been designed in order to provide the facility to manage a list of courses, instead of a single one. Each course is usually identified in a CMS by a unique string. This string is the ID of the course, which corresponds usually to the identification code assigned by the school or university to the course. We use this ID to distinguish between different courses in CourseVis.

4.7.2 About the class dimension

Classes in distance learning are composed for the big majority of less than 50 students (see Figure 4.3). CourseVis could be restricted to support classes of 50 students at maximum. However, the support of larger sizes of classes should be considered as well, because classes with high number of students, even if they are infrequent, are practised in some cases.

4.7.3 About the learning platform

Section 4.4.3 revealed that the most popular course management systems are WebCT and Blackboard. This is not an unexpected results, as BlackBoard affirms to have more than 2.400 clients and WebCT more than 2.200 clients between universities, colleges, schools and other education providers worldwide⁷. CourseVis should be implemented in order to support the most popular

⁶One of the respondent to the survey included the following comment: "20-30% of a student's grade is based on active, meaningful participation in asynchronous group discussion databases. If students do not earn 60% on this component, they fail the course."

⁷Information given by their corporate websites (<http://www.blackboard.com>, <http://www.webct.com>) in January 2003.

CMS. A possible solution to this problem could be the definition of a common interface between CMS which enable CourseVis to be independent by from the specific CMS. We will see in Chapter 5 how this independence has been implemented in CourseVis.

4.7.4 About the information to be represented

The information that must be represented is described in the form of recommendations in Section 4.6. These can be summarised in the following schema, in parenthesis are reported the list of questions which underline each recommendation.

Social aspects

- quantitative visualisation about discussions (Q7, Q9, Q11, Q14)
- visualisation of participation in group work (Q8, Q9)

Cognitive aspects

- student overall performance in the course (Q9)
- level of knowledge achieved by each student for each domain concept of the course (Q11, Q12)
- representation of students having difficulties with a concept (Q12)
- comparison of a student's progress with that of the whole class (Q14)

Behavioural aspects

- student's access to the course (Q9, Q11, Q15)
- student's reading of course materials (Q7, Q12, Q15)
- performing of evaluation proofs (Q12)
- progress with the schedule (Q14)
- participation in discussions (Q7, Q9, Q11, Q14)

4.8 Summary

In this chapter, we presented a survey conducted with instructors involved in distance learning. This study drew indications of the type of information required by instructors teaching at distance by focusing on three themes: social, cognitive and behavioural aspects of the students. The instructors' requirements elicited recommendations to be taken into account in the design of CourseVis. CourseVis has been designed as a system that illustrates the use of Information Visualisation to

depict the data in an on-line course, including how to use data from CMS and how to identify suitable visualisation techniques. The findings of this survey can be extended to the design of other software tools aimed at facilitating distance teachers.

In the next chapter, we will describe how the recommendations outlined here have been addressed in the design and implementation of CourseVis.

Chapter 5

The CourseVis

In Chapter 2 of this thesis we described the problem for distance teachers in having clear, reliable and usable information about their students, and we anticipated that some forms of graphical representations could help the instructors in gaining insight and understanding of what is happening to the students participating in the course. A survey conducted with instructors involved in distance education identified requirements for the graphical representations, described in the form of recommendations in Chapter 4. The next step in this research is, by following the recommendations, to develop a visualisation tool to help instructors involved in distance learning courses delivered by CMS tools to better understand what is happening in their classes.

In this chapter, we will describe the design and implementation details of CourseVis, a *Course* graphical *Visualisation* tool which acquires tracking data from a CMS, transforms the data into a form convenient for processing, and generates graphical representations that can be explored and manipulated by instructors to examine specific aspects of distance students. CourseVis is built on top of OpenDX, a general purpose open source tool for the generation of graphical representations for scientific and analytical data¹. OpenDX handles the processes of image abstraction and rendering which allows a rapid implementation of the CourseVis prototype.

This chapter will give a technical description of CourseVis. It begins with a description of the main design guidelines which guided the system design, then a thorough description of system architecture is given. A detailed example which explains the complete visualisation process will be given. Some examples of usage of CourseVis on a real on-line course and the representations produced are discussed in the next chapter.

¹OpenDX was initially developed by IBM, and has now become an open source software (<http://www.opendx.org>).

5.1 Design guidelines

Before designing the general software architecture, we will consider some general rules that will guide the overall process of system design and development. These rules will be drawn from the author's experience as instructor of courses delivered via Internet, our involvement in some projects of the Swiss Virtual Campus, and informal discussions with distance teachers.

Ease to use. During some tutorials on the use of new technologies in education at the University of Applied Sciences of Southern Switzerland, it became clear that the majority of teachers had serious difficulties with using CMS (even the basic functionalities of WebCT) particularly with uploading files to the course and creating self-evaluating quizzes. This appears to be more accentuated with professors who work in non technical domains, such as Literature or Social Studies. Problems in using CMS in other universities have also been documented in Sharp (2000). The usability of a learning system is of primary importance in a user-oriented application (Dix et al., 1998) and should be considered as a high priority. CourseVis should be able to support instructors with limited technical skills.

Integration with a CMS. CMS are widely used to support distance learning courses. In fact, commercial CMS dominate the market nowadays (see Chapter 2). These systems adopt a very simple publishing approach and maintain a network of static hypertext links. Such systems are easy to learn and use, which is probably why they are preferred by distance learning instructors. For this reasons, we propose the implementation of CourseVis as an additional tool which runs in conjunction with a commercial Web-based CMS. Hence, the instructor does not need to learn another system, additional institutional investments are avoided, and thousands of Web-based courses already available on the Web and created using commercial solutions can be used.

Domain independence. Some of the traditional advanced Intelligent Educational Systems are tailored to a specific domain to be taught, hence, their control mechanisms are often domain dependent. Only recently they are trying to separate architectural, methodological, and control issues from the domain of knowledge. Domain dependence is quite necessary to create very dynamic and adaptive courses, but it is not necessary if we want to build a tool like CourseVis.

CMS independence. The implementation of CourseVis should not be restricted to a specific course management system. In fact, despite differences in the format of the tracking information provided by various CMS, there are commonalties in the content and the structure, e.g. history of the pages visited, marks the students receive for each quiz, messages posted to discussion forums, etc. This justifies the possibility for designing a general course visualisation tool that can be applied to a

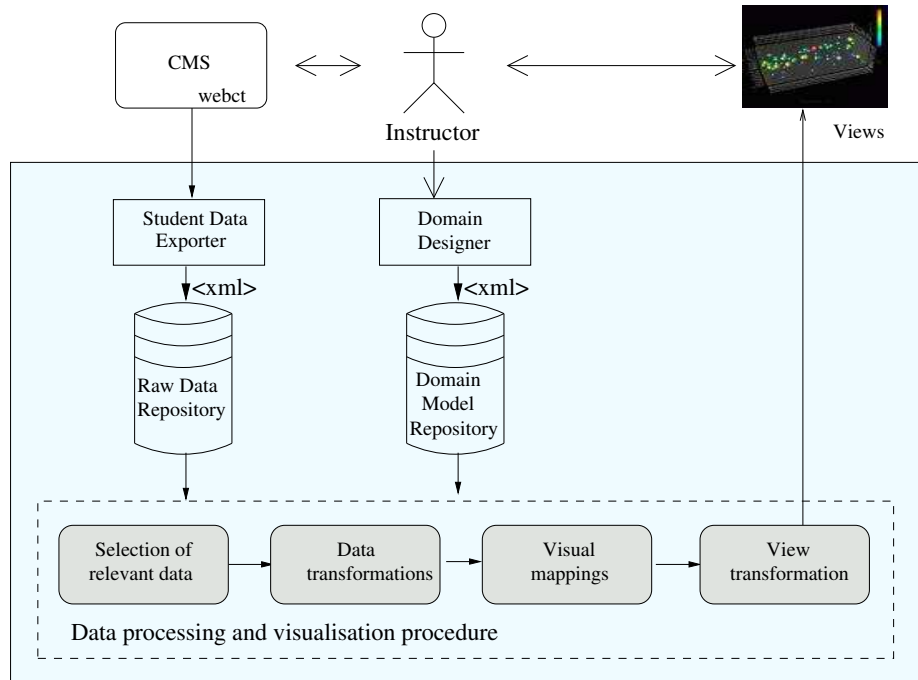


Figure 5.1: The CourseVis system architecture

wide range of distance learning environments.

Free open source development. The free open source initiative is becoming widely supported (<http://www.gnu.org>; <http://www.opensource.org>). The basic idea of the open source is that the software must be free to be read, modified, improved, reused and redistributed. The CourseVis prototype follows the open source idea and is implemented using only free, open source software. This criterion was decisive for the choice of OpenDX as visualisation software package, among the others available (see Section 5.3).

5.2 The CourseVis system architecture

The overall architecture of the CourseVis system is represented in Figure 5.1. It shows that CourseVis has been designed as a tool that extends an existing CMS adding new functionalities in the form of graphical representations.

Figure 5.1 depicts each functional block of CourseVis. In the following paragraphs we will describe each block in details.

5.2.1 CMS

We used WebCT (<http://www.webct.com>) as the reference learning platform for building our prototype. The University of Applied Sciences of Southern Switzerland (SUPSI) has licensed this learning platform since 1999, and it is now the main Web-based distance learning platform in the universities of southern Switzerland. Moreover, WebCT was recommended by the Swiss University Council as the best platform to be used in projects of the Swiss Virtual Campus Programme (<http://www.virtualcampus.ch>). The expertise we had with this product, and the availability of a large set of courses given during last years led us to choose this learning environment as the base for the implementation of the CourseVis prototype. In the current implementation we are using WebCT version 3.6.3.8 - Standard Edition installed on a Linux RedHat 7.2 platform.

5.2.2 Student Data Exporter

The only module in CourseVis that is dependent on the CMS is the *Student Data Exporter* which aims to collect and transform the student tracking data provided by the CMS in a well-defined format. A specific implementation of the Student Data Exporter for the WebCT course management system was built.

Although CourseVis is currently based on data imported from WebCT, it can be adapted to support other CMS. In fact, the Student Data Exporter module is committed to fetch some data that is usually present in a wide range of CMS, such as discussions, access to the course, marks students receive in quizzes, and so on. Other Student Data Exporter for other CMS (e.g. BlackBoard) are considered as a possible extension for a future work.

The Student Data Exporter collects any possible data regarding the activities done by students (tracking data) in a specific course given with the CMS and converts these data in a standard, well-defined representation in XML. The Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML created by the W3C consortium² as recommendation for exchange format of a wide variety of data (The W3C Consortium et al., 2000; Goldfarb and Prescod, 2002). This XML representation of student tracking data is used as an input to other modules in CourseVis. This enables the other modules to be based on a unified data syntax and semantics, and to be independent from the particular CMS used.

The task that this module has to carry out is not trivial. WebCT is a copyrighted commercial software, therefore users know almost nothing about its internal data structure and, in particular, the program code cannot be modified or used in whole or in part outside of WebCT, otherwise we would break the Software License rules. However, WebCT's data files are mostly text-based,

²The W3C (<http://w3c.org>) is a consortium of academic institutions and industrial companies founded in 1994 by Tim Berners-Lee, the person who created the HTTP protocol for the Web. Among his activities is the definition of standards (called recommendations) to promote the interoperability and the evolution of the Web. The XML standard is a recommendation released by the W3C for the first version in 1998.

which makes it easy to build an external procedure that collects the students' data file generated by WebCT and recreate the original data structures using a reverse engineering approach. In this way we did not break the WebCT license regulations, because we just used the data produced by the tool, and did not work with the source program code.

A specific XML syntax has been implemented for the data being produced by this module. In the next section we will describe this syntax.

Conceptual description of the student data structure

Here we will give a brief conceptual description of the XML data format generated by the Student Data Exporter module. This data format is defined as a syntax, whose grammar is known as a Document Type Definition, or DTD (Goldfarb and Prescod, 2002). The DTD describes the tag names, content model, and attributes of elements in use in an XML document. We refer to the data structure generated by the Student Data exporter module with the term *student data structure*. DTD for this data structure was created as a reference to help developers in creating further Student Data Exporter modules specific for other CMS. Other developers can make use of this DTD and validate their XML documents to ensure that the documents generated are consistent with the syntax specified in it.

Note that this chapter contains only a partial description of DTD created for the CourseVis, an exhaustive explanation is given in Appendix C.

The primary elements of the student data structure are shown in Figure 5.2.

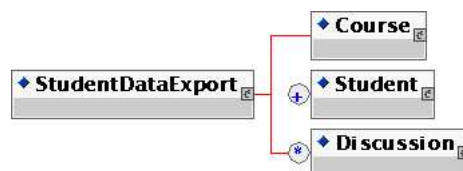


Figure 5.2: The primary elements of the student data structure generated by the CourseVis Student Data Exporter.

`StudentDataExport` is the first, top-level object (in the DTD's terminology these objects are called *elements*) of the DTD tree. This encloses all the other elements of the structure. The other elements of the student data are grouped in three main groups :

Course - contains generic information about the course;

Student - stores all the data about each student. This is a multiple element³ replicated for each student of the course;

³In DTD a plus (+) indicates that an element may be repeated more than once but must occur at least once. The question mark (?) indicates that it is optional (it may be absent, or it may occur exactly once). The asterisk (*) indicates

Discussion - stores information about the discussions made on the discussion board tool of the course⁴.

Elements contained in `Course` are expanded in Figure 5.3.

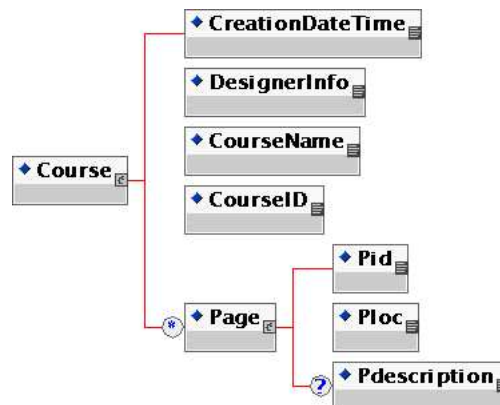


Figure 5.3: `Course` DTD elements in the student data structure generated by the CourseVis Student Data Exporter.

`Course` contains the date and time of the course creation (`CreationDateTime`), the name of the instructor (`DesignerInfo`), the name of the course (`CourseName`) and the unique identification string of the course (`CourseID`) which is the same that is used within the CMS. `Page` is the list of the pages which compose the content materials of the course, they are ordered according to the table of contents of the course (`Pid` reflect the order which appears in the table of contents), have a unique identifier (`Ploc`) and a description (`Pdescription`).

The `Student` element contains specific data about each student of the course. They are depicted in Figure 5.4. Here we can identify further information about the personal data of the students (`PersonalInfo`), data about the student's access to the course (`AccessInfo`), data about the e-mail this student has exchanged (`Mail`), data about the quiz performed by the student (`Quiz`), and finally data about the assignments given to students (`Assignment`). The `AccessInfo` element contains further information about the first access, last access, and hits number to the course (respectively `FirstAccess`, `LastAccess` and `Hits`), the history of the course's pages visited (`History`). The `Mail` element contains all the e-mail (included the full text of the message in the `MMessage` element⁵) sent and received by the student using the e-mail

that the content is optional and may be repeated (may occur zero or more times). An element name with no punctuation, such as `Course`, must occur exactly once.

⁴Note that this element is marked with an asterisk which means that the content of this element is optional. This is because the instructor may decide not to include a discussion board in his course.

⁵Full text of e-mail is not considered in the current version of CourseVis. It has been considered by the student data structure generated by the CourseVis Student Data Exporter for completeness and possible future work.

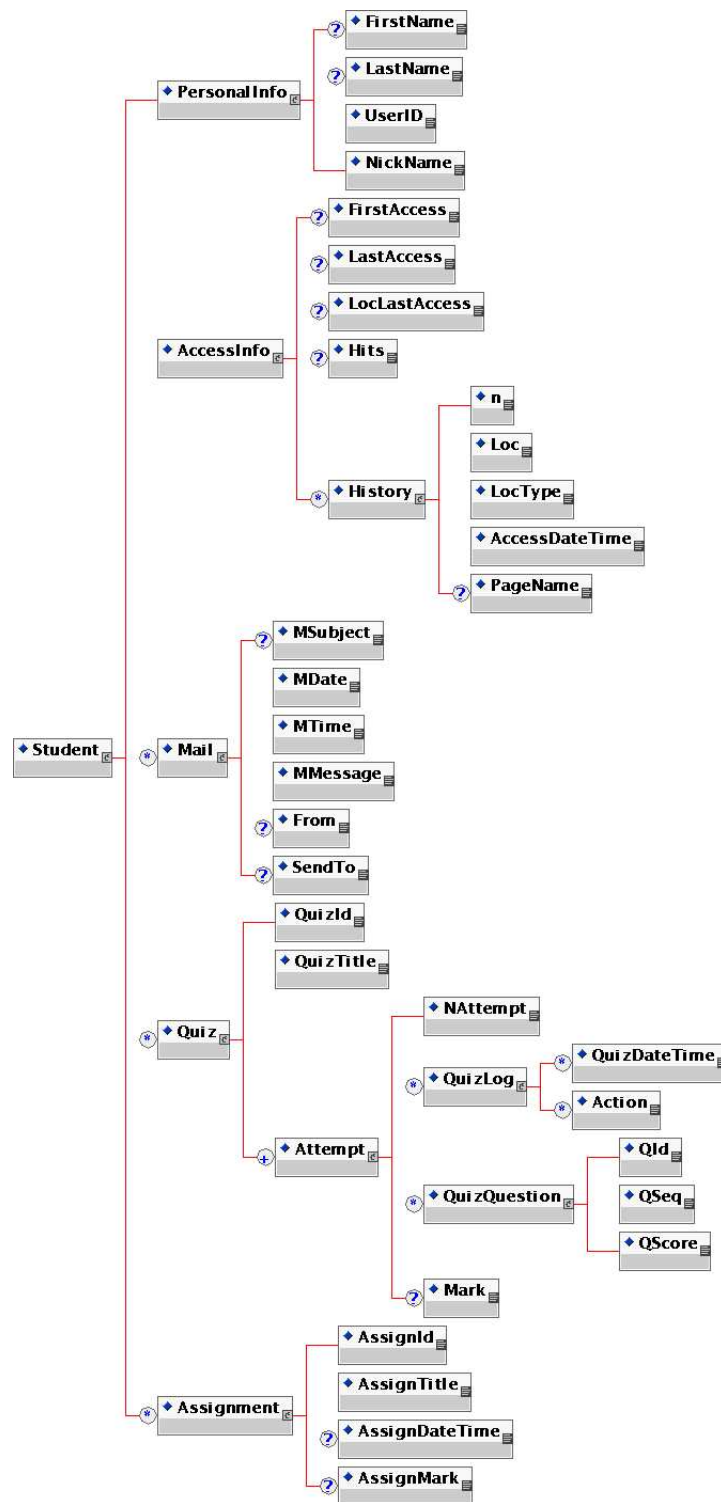


Figure 5.4: Student DTD elements in the student data structure generated by the CourseVis Student Data Exporter.

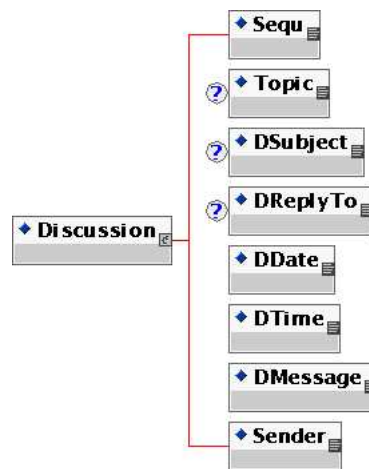


Figure 5.5: *Discussion* DTD elements in the student data structure generated by the CourseVis Student Data Exporter.

tool provided by the CMS. The *Quiz* element collects details about the attempts to the quizzes done by the student and stores information about the global marks achieved for a quiz set (*Mark*) as well as information about the mark received for each single question contained in a quiz set (*Qscore*). The *Assignment* element collects information about the assignments submitted by the student.

The last of the primary elements of the students data structure is *Discussion* (see details in Figure 5.5). This is a multiple element (there is one *Discussion* element for each message which appears in the discussion tool), and it contains data about each message posted in discussions including the text of the message (*DMessage*), the sender (*Sender*), the topic (*Topic*) and, eventually, the identification number (*Sequ*) of the message that this is replying to (*DReplyTo*).

The above explanation of the DTD structure is aimed at helping the implementer of a Student Data Exporter to produce the correct XML data output for this module. Like any other specification of a conceptual data schema, each element must be precisely defined in each technical aspect (data type, value space and range, explanation, etc.). A detailed description of each element involved in the students data structure DTD hierarchy is given in Appendix C. The technical details described in the appendix are essential to other developers for the implementation of a Student Data Exporter.

It must be stressed that at the current stage not all of the elements existing in the student data structure are used by CourseVis to produce the representations. For instance, the element *Mail* included in the *Student* element was never used by the subsequent modules of CourseVis. Future extension of CourseVis may produce additional representations that may require the data

provided by elements that are not actively used now.

The Student Data Exporter has been implemented in the Perl programming language (Wall et al., 2000) (WebCT is mostly implemented in Perl, and the same language was chosen for the implementation of the Student Data Exporter). The implementation of this module was particularly challenging because it required a big effort to reverse engineering the information data produced by the CMS. The implementation of the Student Data Exporter for WebCT was possible due to its particular data file format, which is text-based. It must be noticed that implementation of Student Data Exporter for other CMS could be technically challenging without the contribution of the producer of the software.

5.2.3 Raw Data Repository

The XML data produced by the Student Data Exporter is stored in the *Raw Data Repository* (see Figure 5.1). This module is essentially a relational database management system (DBMS) implemented using MySQL (<http://www.mysql.com>), a popular open source database. The mapping between the XML documents and the database's relational structure was implemented using the XML-DBMS Java Packages library, a middleware for transferring data between XML documents and relational databases developed by Ronald Bourret (<http://rpbouret.com/xmldbms>). We have chosen this middleware due to its simplicity and because it provides a complete set of Java classes ready to be used. Documents produced by the Student Data Exporter are passed to the Raw Data Repository by the means of regular data files stored into the file system of the computer where CourseVis is running. In a further release of the program the Student Data Exporter could send the generated XML data through a network connection. In this manner the CourseVis could be split in 2 parts: a server part (the CMS and the Student Data Exporter) and a client part, allowing the instructors to install the client part of CourseVis on their own computers, hence, avoiding the installation of the complete system on the machine where the CMS is running.

5.2.4 Domain Designer

The *Domain Designer* (see Figure 5.1) is a module dedicated to instructors to define the *domain model* of the course. The domain model describes the entities of the course domain (that we called *concepts*) and their relationships with the pages of the course and questions in quizzes. It is necessary to define a model of the domain in order to relate information about concepts of our course with students. For instance, the numeric results on quizzes could be used to judge the level of comprehension of the students on the concepts of the domain. If we want to represent the level of knowledge achieved by each student on a specific concept of the course we need to define a model of the domain. Without this model, it would be impossible to relate evaluation proofs (i.e. quizzes) with concepts of the course.

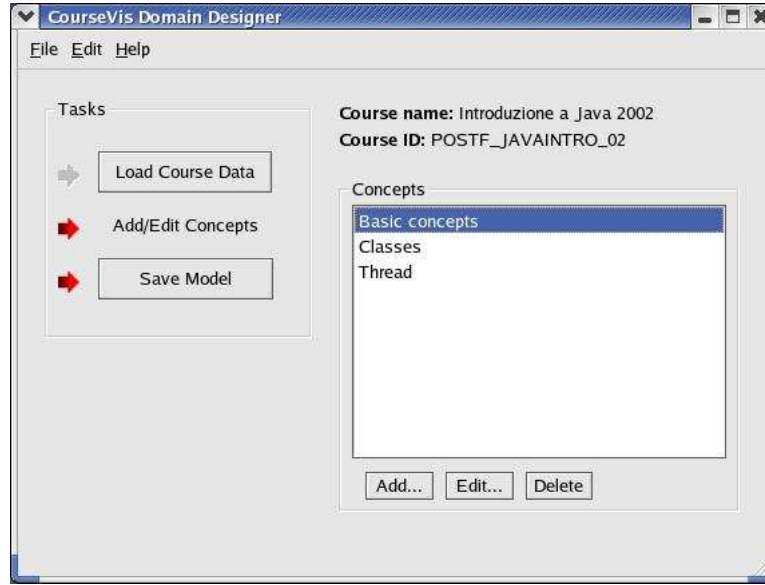


Figure 5.6: Domain designer - user interface for the definition of the concepts of the course

The model of the domain can be implemented in different ways, ranging from in-depth representations of the domain of the course using an *ontology* of the domain (Guarino, 1997) to a simple definition of a plain list of concepts. The ontology, as “*an intensional semantic structure which encodes the implicit rules constraining the structure of a piece of reality*”(Guarino, 1997), is the most comprehensive. However, it requires significant knowledge engineering effort. One of the general principles that directed the design of CourseVis is *ease to use* (see Section 5.1). Thus, we want to provide the users (i.e. the instructor) with a simple approach to construct a model of the domain sufficient for the successful run of CourseVis. For this reason, the model of the domain in CourseVis is intended as a plain list of concepts, where each concept is associated with a set of *pages* and *evaluation proofs* (i.e. quizzes, group exercises). Note that we have not included any dependencies among concepts. Possible inclusion of hierarchical concept relations is being considered as a future development of CourseVis (see Chapter 8).

The model of the domain can be seen as a collection of concepts $\{C_1, \dots, C_n\}$ where each concept is composed by a list of pages (P) belonging to the content materials of the course and a set of questions (Q) selected among the quizzes created for the course, in this format:

$$C_i = (\{P_1, \dots, P_h\}, \{Q_1, \dots, Q_k\})$$

Most of the pages of the content materials and questions in quizzes would deal with multiple concepts, hence a single page or a single question could belong to different concepts.

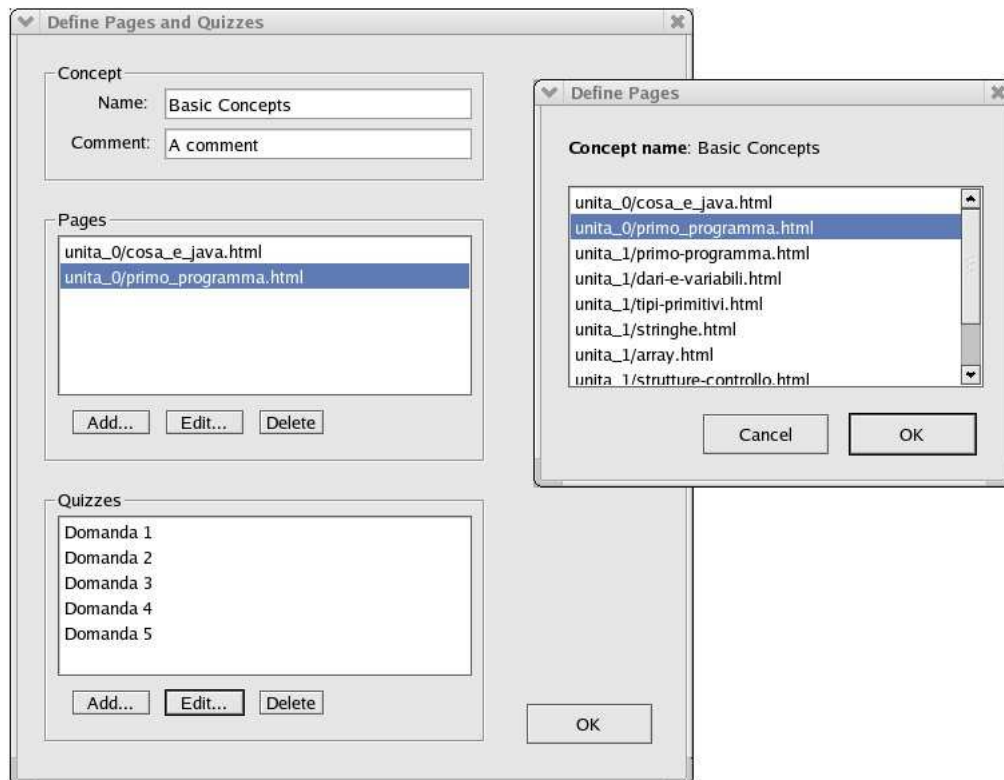


Figure 5.7: Domain designer in CourseVis - user interface for the definition of pages and quiz questions related to a concept. The instructor selects the corresponding pages for a concept from the window in the right.

The Domain Designer is implemented using the C++ programming language⁶. In the current version it displays a simple interface to the instructor to define the list of the concepts of the course (see Figure 5.6). For each concept the instructor has to select among the content pages existing into the course, which are related to the concept. The instructor also has to select which questions in quizzes are related to the concept. The interface used by instructors to relate concepts with pages and questions is illustrated in Figure 5.7.

To help instructors in authoring the model of the domain, the list of content pages and the list of question of quizzes existing in a course is automatically gathered by the Student Data Exporter module of CourseVis. In this manner the instructor can match concepts against pages and quiz questions simply by selecting and clicking on the corresponding page or question, as illustrated in Figure 5.7 for the selection of content pages.

⁶The choice of the programming language for the implementation of the Domain Designer was due to convenience matter.

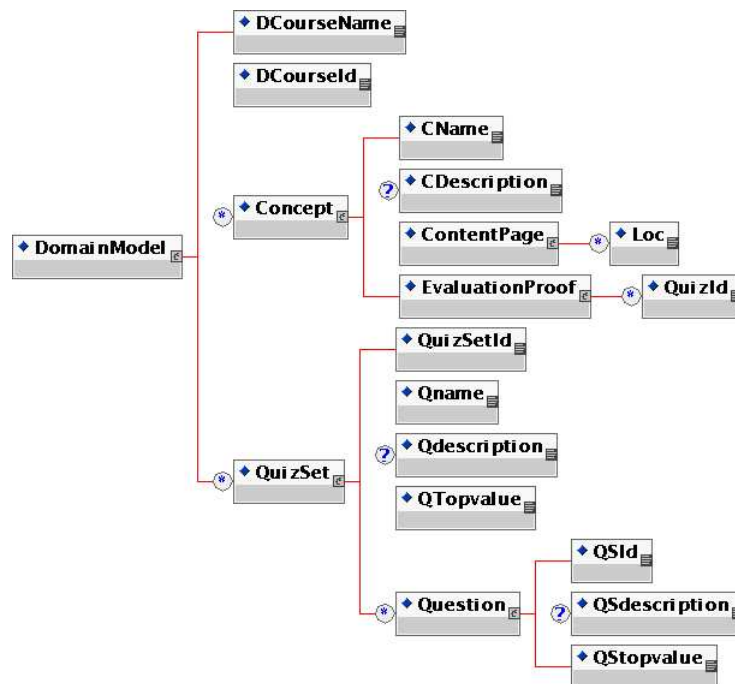


Figure 5.8: The domain data structure DTD generated by the CourseVis Domain Designer

Conceptual description of the domain data structure

The Domain Designer produces an output that is an XML data structure which describes the domain of the course upon the information provided by the instructor. We identify this XML data with the term *domain data structure*. The DTD of the domain data structure is presented in Figure 5.8.

In this schema we can identify the following elements:

- Elements containing generic data about the course, i.e. the name of the course (DCourseName) and a unique identification string defined by the CMS as the course identifier (DCourseId)
- Elements containing the list of the concepts (Concept), replicated for each concept of the domain. This element has a name (CName) and a description (CDescription). The Concept element contains further elements having the list of content pages related to it (ContentPage) identified by a location (Loc) and the list of evaluation proofs related to it, currently represented by a list of quiz questions (QuizId).
- A list of the quiz set inserted into the course (QuizSet), which is a collection of questions (Question) proposed to students in a format that can be graded by the system. Typical question formats are multiple choice, matching, fill-in the blank, etc. For both a quiz set and

a single question the instructor has to define the top grade value (respectively, $Q_{Topvalue}$ and $Q_{Stopvalue}$). These values will be used by other modules in CourseVis to perform some computations.

See Appendix C for a detailed explanation of each element involved in the domain data structure DTD.

5.2.5 Domain Model Repository

The *Domain Model Repository* (see Figure 5.1) is dedicated to store the XML data produced by the Domain Designer. Its implementation is quite identical to that of the Raw Data Repository. In fact, they share the same DBMS and share the same primitives that map XML documents onto a relational structure. See Section 5.2.3 for details on this module.

5.2.6 Data Processing and Visualisation Procedure

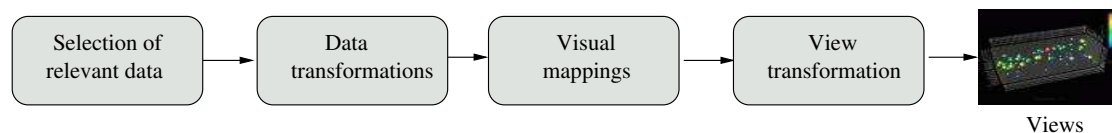


Figure 5.9: Visualisation pipeline followed in CourseVis, derived from the *reference model for visualisation* proposed by Card et al. (1999).

The CourseVis modules described so far contain the necessary data about students and domain that can be used to analyse the activities performed by students into the course. Data collected and stored in the repositories must now be processed and presented to the instructor in a format that is easy to understand and communicated with clarity, precision and efficiency (Tuft, 1990). Applying Information Visualisation to data stored in repositories is a way to gain insights and understanding of the data (Spence, 2001), and in this case the activities of the course from the data collected by CourseVis.

To represent data in visual format we need to do some computations and transform the data into visual structures and representations. This step is performed in CourseVis by the “*Data Processing and Visualisation Procedure*” (see Figure 5.1). This “*visualisation pipeline*”, which has been replicated in Figure 5.9 for the reader’s convenience, follows the well known “*reference model for visualisation*” proposed by Card et al. (1999). This model describes the creation of a graphical representation as mapping of data to visual form through a pipeline of four stages:

1. *Selection of relevant input data*: repositories contain a huge quantity of data in relational format (data is stored in a DBMS). Depending on the representation we want to produce,

only a subset of this data is relevant. This stage extracts from the repositories only data relevant to the particular representation. For each representation we have defined a specific set of data that is being extracted from this module.

2. *Data transformation*: this step manipulates the selected data extracted from the repository and generates an intermediate format which may add further metadata of the abstract representation. Data provided by the CMS and stored into the repositories often need to be transformed to add additional information (for instance the data contain errors or missing values that must be addressed before the data can be visualised), or to perform some calculations (such as statistical calculations to add additional information).
3. *Visualisation mapping*: transforms the pre-processed filtered data into geometrical primitives with appropriate attributes, such as colour or opacity. Visualisation mapping is the core of the whole visualisation process. This process transforms the data from a format of relational tables into graphical structures to be processed by the human vision.
4. *View transformation*: generates the image by using the geometric primitives from the mapping process by specifying graphical parameters such as position and scale. These graphical parameters are controlled by the user interaction, to perform some graphical manipulation such as zooming, scaling, panning, etc.

First two modules are implemented in Perl, while the last two stages of the pipeline (*Visualisation mapping* and *View transformation*) are implemented by using the OpenDX visualisation tool described in Section 5.3. In Section 5.4 we will illustrate the complete image generation process in CourseVis with an example.

5.3 The use of OpenDX

OpenDX (<http://www.opendx.org>) is a powerful open source package for the visualisation of analytical (as well as scientific and engineering) data. Originally developed at IBM in 1991 (the original name was Visualization Data Explorer), it has been continuously enhanced and expanded and has now become an Open Source product with its full source code freely available to anyone. This software package has been chosen both for the functionality it provides and for its Open Source nature, which is in line with the design principles of CourseVis (see Section 5.1).

OpenDX allows the construction of complex graphical representations both by a text-based scripting language and a visual programming language. The visual programming environment uses a graphically oriented editor which allows the generation of graphical images as output simply by combining functional modules together into an extended data flow architecture. Specifically, to build a visual program one has to select some modules, place them in an editor panel using a

drag-and-drop modality, and connect the modules together into a network. Each module can be considered as a function which takes a set of input data and produces some output data. When a module receives its input, it performs its specified operations to the input data, and distributes its results to the modules connected to the output downstream. Figure 5.10 shows an example of the visual program editor (on the left) and the image produced by this data-flow (on the right).

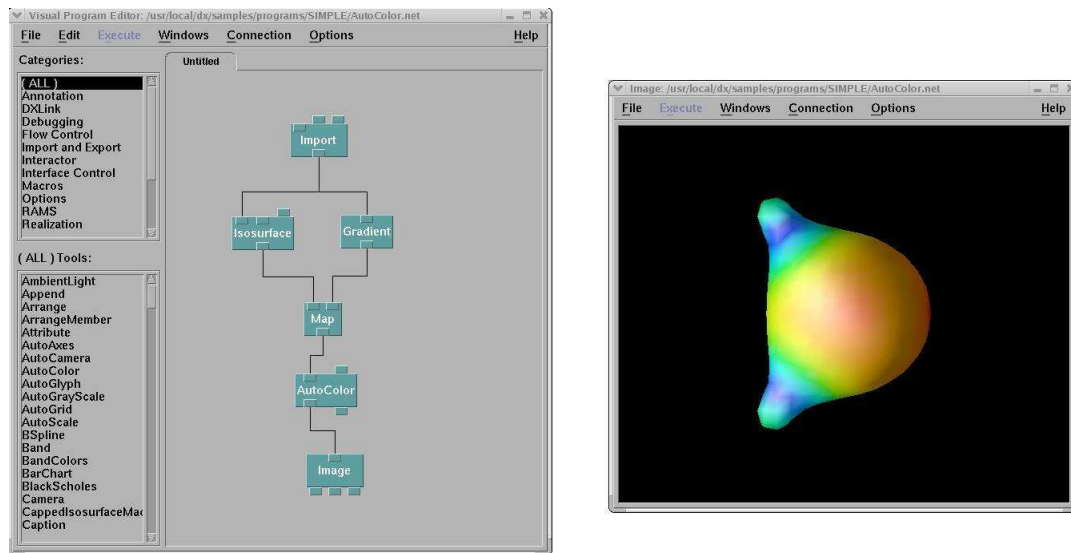


Figure 5.10: Example of the OpenDX visual program editor and the image produced.

In the visual editor modules are represented graphically as rectangular boxes with tabs on the top (the inputs) and on the bottom (the outputs). A connection between two modules means that the data in output from a module is passed as input to other modules. The data flow in Figure 5.10 starts with the *Import* module (in charge of reading data stored in an external data file), propagates the data to other modules, just like a flow-chart diagram, and produces the resulting image on the right of the figure.

OpenDX provides a simple GUI builder that enables an user to directly manipulate images (e.g. rotate or zoom) and to control the appearance of a graphical representation using dials, switches, buttons, and sliders. These latter are called *interactors*, some examples of interactors are shown in Figure 5.11.

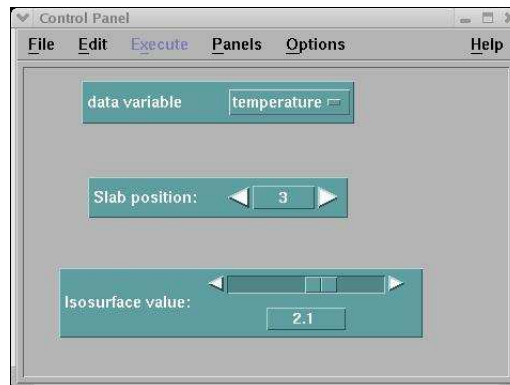


Figure 5.11: Some examples of interactors used in OpenDX to change the visual image produced by a program.

OpenDX can run as a stand-alone program or can be integrated in other applications. It includes interface library that allow OpenDX functionality to be incorporated into third-part applications. Moreover, OpenDX was designed to work in a client/server environment and to run in parallel on a shared-memory multiprocessor.

CourseVis uses OpenDX as the visualisation tool, implementing the two latter stages of the data processing and visualisation procedure depicted in Figure 5.9, namely the visual mapping and the view transformations, implemented by specific OpenDX modules, as described in the next section.

5.4 An example of a graphical representation produced with CourseVis

The process of images creation in CourseVis can be described by a detailed example which explains the complete image creation process starting from the data extracted from the CMS and ending with the images produced. To describe the genesis of a representation we will use the data extracted from a Java on-line course (see Section 6.1). The representation here is intended to explain the details of the process of image's generation in CourseVis. In the next chapter we will show some representations which meet the instructors requirements identified in the survey presented in Chapter 4.

The example below addresses the following *instructor question*:

What is the global number of messages read and posted by students throughout the course?

It is often an instructors' responsibility to “*monitor the quality and regularity of learner postings*” (Rossman, 1999) and, according to Statz (2001), “*Students need to be encouraged to provide*

thoughtful and relevant ideas that enhance the discussions". This activity requires at the very basic level that the instructor keep the tracking of the number of messages posted and read by students in the course. This information must be easily visible to each instructor when running a course at distance (see Chapter 4). However, commercial CMS provide this information in a format that is often impractical and difficult to understand. There are some commercial CMS that don't even provide this information (this is the case of BlackBoard, for example). For illustration, the page of students tracking provided by WebCT is shown in Figure 2.1 on top. The number of articles read and posted by each student are reported in numerical tabular format in last 2 columns of the table.

This table corresponds to the question defined earlier. In fact, the information requested, namely the number of messages read and posted by each student, is reported in the table. However, the tabular modality may be difficult to follow. It provides a lot of information which is not requested by the question (e.g. first and last access date), which may distort the user's attention. According to Tufte (1983, p.91): "*Data graphics should draw the viewer's attention to the sense and substance of the data, not to something else*". This also sacrifices a lot of precious space on the display area, as indicated by Spence (2001) for this kind of graphical representations. Moreover, the instructor may be interested in performing some comparisons between students, like the identification of the most active students in posting and reading discussions, which is difficult to perform using the representations in Figure 2.1. We will illustrate the use of Information Visualisations in CourseVis to envisioning this information.

5.4.1 Extraction of data from the CMS

We propose the use of a simple two dimensional scatterplot (Spence, 2001) to graphically represent the numbers of messages read and posted in discussions. We start with the first stage of CourseVis: the extraction of the data from the CMS. This is done by the Student Data Exporter module which is in charge of extracting all the course data from the CMS and converts this data in an XML format compliant with the DTD structure defined in Appendix C. Full details concerning the process of data extraction from the CMS will be omitted here due to complexity. The next step is the output generated by the Student Data Exporter module. The fragment below extracted from an XML document generated by the Student Data Exporter:

```

.....
<AccessInfo>
  <FirstAccess>2002-01-16 10:13:59</FirstAccess>
  <LastAccess>2002-04-08 13:59:00</LastAccess>
  <LocLastAccess>unita_7/comunicazione.html</LocLastAccess>
  <Hits>177</Hits>
  <History>
    <n>1</n>
    <Loc>WEBCT__homepage</Loc>
    <LocType>Home_page</LocType>
    <AccessDateTime>2002-01-16 10:13:59</AccessDateTime>
  </History>
  <History>
    <n>2</n>
    <Loc>unita_1/tipi-primitivi.html</Loc>
    <LocType>Notes</LocType>
    <AccessDateTime>2002-01-16 10:20:39</AccessDateTime>
  </History>
  <History>
    <n>3</n>
    <Loc></Loc>
    <LocType>Articles_Read</LocType>
    <AccessDateTime>2002-01-16 10:25:12</AccessDateTime>
  </History>
.....

```

The fragment illustrates some information concerning the first and last accesses made by a student in the course (`FirstAccess` and `LastAccess` elements) as well as few elements concerning the history of course pages visited by the student (`History` elements). This is only a very small fragment of a document generated by the Student Data Exporter. Depending on the activities done in the course, and the number of students, the complete XML document can reach several Mega Bytes of dimension.

The generation of this document could be scheduled by the operating system which hosts the CMS to create an updated document on a regular basis. For instance, the operating system can be configured to launch the Student Data Exported once a day in order to update daily the exported data.

5.4.2 Mapping XML student data structure into the repository

The next step in CourseVis is the mapping of the XML document into a relational database: the Raw Data Repository. The table-based mapping methodology proposed by Ronald Bourret (<http://www.rpbouret.com/xmldbms>) has been used. Following this methodology, an XML document, structured as in the example showed above, could be mapped in the following tables:

Table AccessInfo

FirstAccess	LastAccess	LocLastAccess	Hits
2002-01-16 10:13:59	2002-04-08 13:59:00	unita_7/comunicazione.html	177

Table History

n	Loc	locType	AccessDateTime
1	WEBCT__homepage	Home_page	2002-01-16 10:13:59
2	unita_1/tipi-primitivi.html	Notes	2002-01-16 10:20:39
3		Articles_Read	2002-01-16 10:25:12
..

Table 5.1: Mapping of the XML into relational table whichs as part of the architecture of Course-Vis. Primary keys and foreign keys are omitted for simplicity.

The Processes described so far are independent from the representation to be produced. Now we have to select the type of representation to produce, because the operations performed by the following modules are dependant on the specific representation to create.

As can be seen in Table 5.1, the information about the number of messages read and posted by each user is contained in the `History` element of the XML data structure (see Appendix C for a complete description of the DTD). In particular, this element contains a list of pages visited⁷ by each student, with the type (`locType`) of the visited page. The type of page visited is “Articles_Read” if the student read an article, “Original_Posts” if the student posted a new message (not replying to another one), “Follow_ups” if the student answered to a message originally posted by another student.

5.4.3 Selection of relevant data

The first stage of the visualisation pipeline (see Figure 5.9) is the ‘selection of relevant data’. This chooses tables and records in the Raw Data Repository and, eventually, the Domain Model Repository, that are eligible for the representation to be produced. At an implementation level, this corresponds to several SQL statements run onto the repositories to extract the data necessary to produce the desired representation. In this example, statements which select the history data belonging to each user and extract only the records having in the `locType` element the values “Articles_Read”, “Original_Posts” or “Follow_ups” are invoked.

⁷The term “pages visited” refers not only to the content pages with the materials of the course, but also to every sort of page visualised by the student in his browser, such as a message in discussion, an exercise, the main page of the course, etc.

5.4.4 Data transformations

The second stage of the visualisation pipeline (Data transformations) transforms the data to add additional information with some calculations. In this example, the complete list of messages read or written by the user (this information is already provided by the CMS's discussion tool) has not been considered, but rather a global number of messages throughout the course is being used. For this reason we calculate the sum of messages read and posted by each user and report this values in the following data table:

student	read	posted
luisa	263	21
michele	14	0
giorgio	134	20
enzo	90	10
luigi	89	3
orazio	117	13
ada	117	7
...

Table 5.2: Data table produced by the Data transformation process for the Java programming course described in Section 6.1.

5.4.5 Visual mappings

The third and the fourth stage of the visualisation pipeline (Visual mapping and View transformation) are implemented using the OpenDX tool. Card et al. describe this step of the process: *“In Visualization, Data Tables are mapped to Visual Structures, which augment a spatial substrate with marks and graphical properties to encode information [...] Visual Structures are based on graphical properties effectively processed by human vision”* (Card et al., 1999, p. 23). This sentence explains this process which is the core of the entire visualisation. Although the data represented in Table 5.2 give the values requested by the instructor's question (for example, see the illustration in Figure 2.1 on top), by adding some spatial components this information can be represented in a pictorial way, such as a diagram or a plot. There is an argument that graphics, if well constructed, may help the human cognitive process and can convey information in human enabling the recognition at essential features quicker than the textual or tabular format (Ruddle et al., 2002). Hence, graphical representation may help the instructor to make important inferential processes.

Let us assume that the instructor is interested in the global number of messages read and posted in order to split a class in four types of students on the basis of the activities in discussions: *proactive* (students who are very active in proposing new discussions), *active* (students who especially contribute in discussions), *vicarious* (students who passively observes discussions without

taking an active part in interaction) and *inactive* (students who don't use the discussions tool). A simple classification could be done using the data provided by the tool in Figure 2.1. However, the instructor has to analyse the values and perform some classification on their own, which will require additional time. We propose a graphical representation which renders these values on a 2D diagram (number of messages read on the X axis and posted on the Y axis) and positioning the student's names on each correspondence. As shown later, this representation could help the instructor to classify the students straightforwardly.

Let us go back to the visualisation pipeline shown in Figure 5.9. The Data transformation process produced the data table represented in table 5.2. This data table can be given as an input to a special program implemented with OpenDX to produce the desired representation. The OpenDX can be programmed to produce the visual abstraction of this data and generates the graphical structures that are necessary to produce the representation. This is the third step of the visualisation pipeline: the Visual mapping process.

For the example, the OpenDX program illustrated in Figure 5.12 was implemented. This program, implemented using the visual program editor available with OpenDX, loads the data table produced by the Data transformation process with the ImportSheet module⁸. The visual program propagates the computation throughout other modules connected to the output downstream (for the sake of simplicity, the details of each single module of the visual program are omitted). The general computations involved in the program are the conversion of the values in the columns *post* and *read* of table 5.2 into coordinates of a plan. In particular, the module *Compute* computes a pointwise expression over a field, and generates a 2D vector in the format:

$$\{[263, 21] [14, 0] [134, 20] \dots\}$$

Values in this vector cannot be rendered directly. These values are converted into geometry with the OpenDX's *AutoGlyph* and *Glyph* modules. These modules create a couple of scatterplot representations that can be rendered and displayed by the *Image* module.

5.4.6 View Transformation

OpenDX's *Image* module implements the last step of the visualisation pipeline, i.e. the View transformation. It generates the image using the geometric primitives defined by the previous modules, and enables the user to manipulate the image to perform some graphical manipulation such as zooming, scaling, panning, rotating, etc. (see Section 3.2.8). Figure 5.13 illustrates the final result of this visual production. This diagram has some advantages compared to the interface provided by the CMS (illustrated in Figure 2.1 on top):

⁸In OpenDX, a data table in a columnar format as the one in Figure 5.2 is called a "*spreadsheet*".

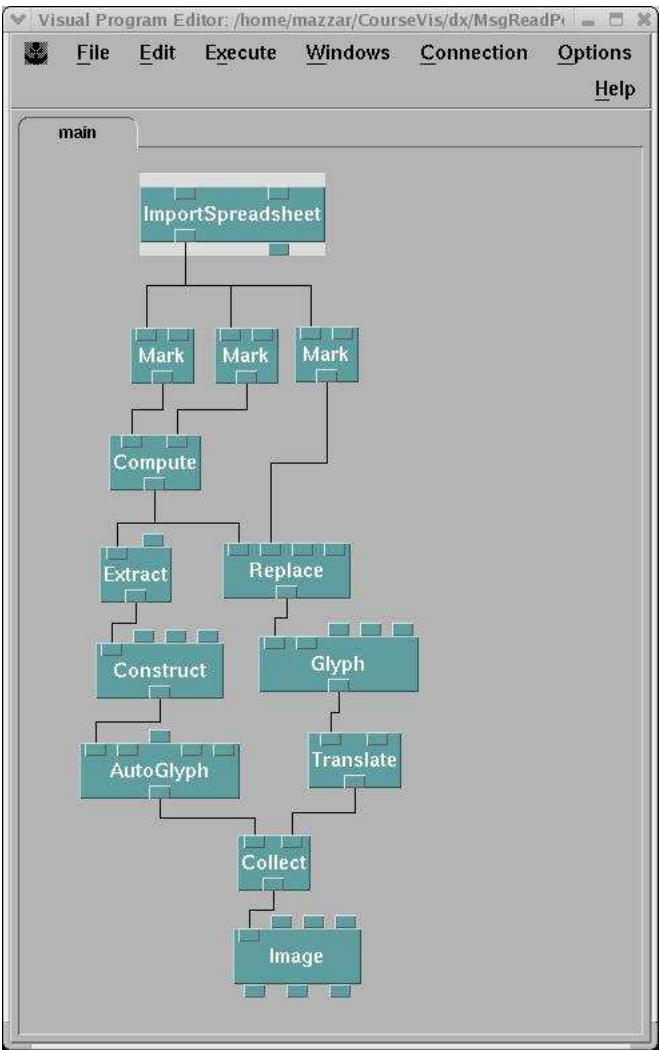


Figure 5.12: The OpenDX visual program for the visualisation example discussed in this section.

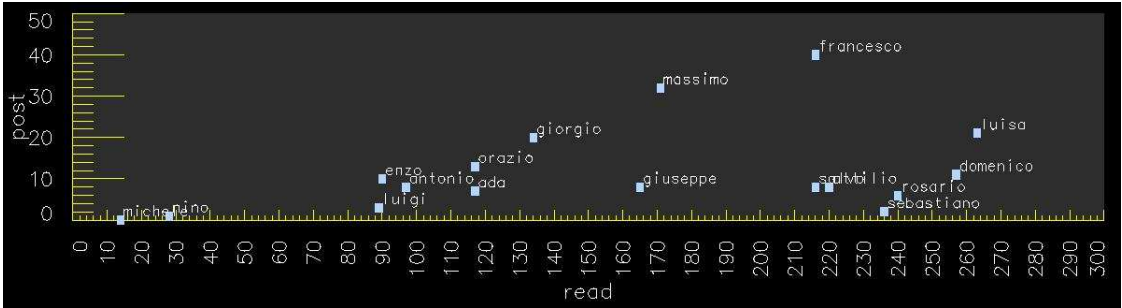


Figure 5.13: OpenDX visualisation produced for the example discussed in this section.



Figure 5.14: Visualisation produced for a course with 231 students enrolled.

- The diagram contains only essential information.
- The instructor can recognise essential features quicker than when using the textual or tabular format. For instance, it can be seen at a glance that the student *francesco* was the most active student in posting messages, and the student *sebastiano*, even though he read a lot of messages, had posted almost no messages.
- The instructor can infer some interesting information which would have been very difficult to do with the interface provided by the CMS. For instance, the students *michele* and *nino* can be classified as *inactive* (very few messages read and almost none sent), the students *francesco*, *massimo*, *giorgio* and *luisa* can be classified as *proactive* (high number of messages posted), the student *sebastiano* can be classified as *vicarious* (high number of messages read and almost none posted) and the remaining students can be classified as *active*. This information can be very valuable in a process of decision making for the instructor to decide further actions in discussions.

The OpenDX's *Image* module enables the user to directly manipulate images and to control the appearance of a representation by the means of some graphical manipulations such as zooming, scaling, rotating, panning etc. Image manipulations could be useful if, for some reasons, one has to rearrange the image's presentation in order to focus on a specific part of the image or because there is too much data to display in a single picture.

For instance, the previous diagram involves only 17 students. In the case of a course having many more students, this diagram could be confusing or unreadable due to the limited space in the diagram. See for instance in Figure 5.14 the same graphical representation for a course having 231 students. There is a clear presentation problem in this case, because we have more data that can easily be displayed at once. This problem has been explained in Section 3.2.8. Among the IV techniques dealing with this problem, zooming and panning are probably the most appropriate in the case of the picture represented in Figure 5.14. Zooming and panning combined together

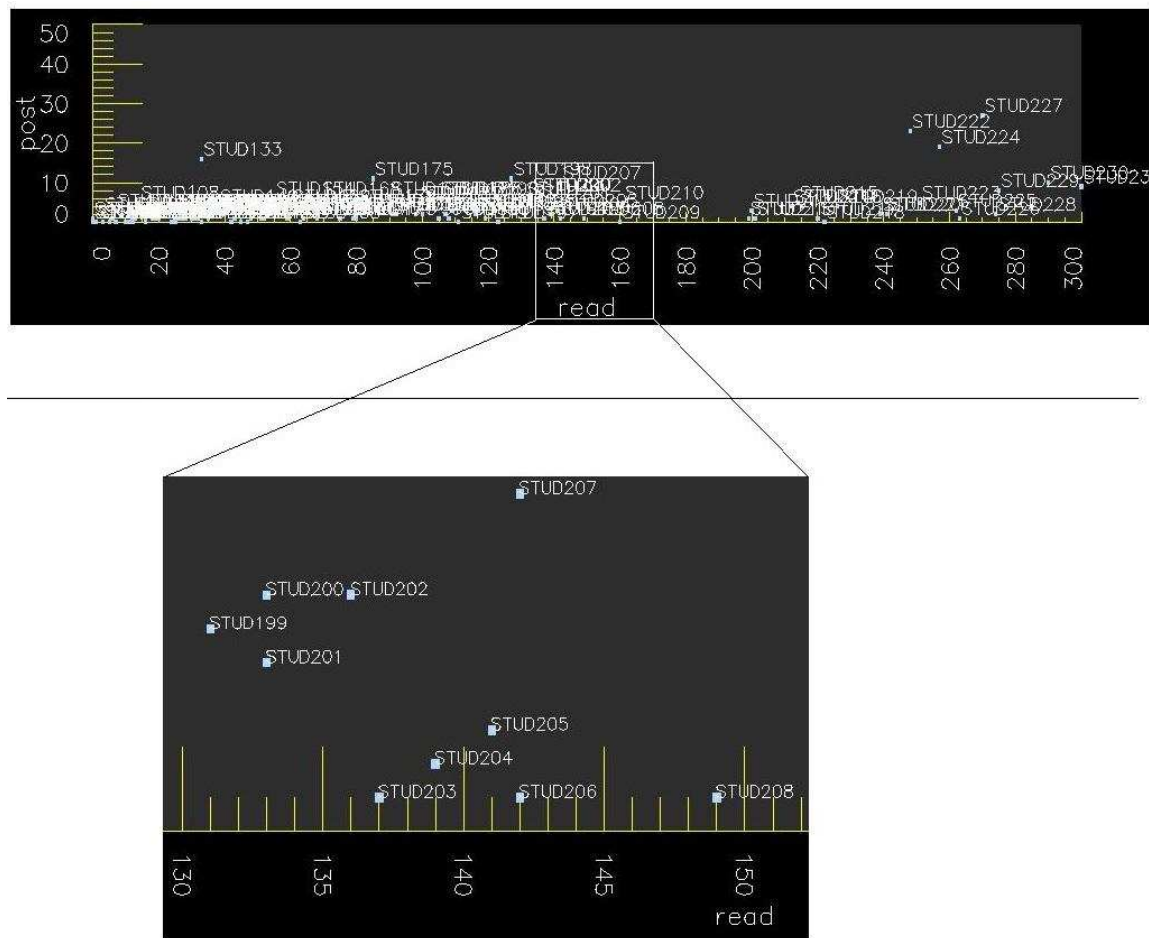


Figure 5.15: Zooming and panning applied to the representation produced

could, in part, help the user clarify parts of the representation in Figure 5.14. The OpenDX's Image module provides the user with these two transformation functionalities. This is part of the last stage of the visualisation pipeline - "*View transformation*". Figure 5.15 illustrates a zooming applied to the image to magnificate a frame. This frame can be moved over the upper image, while the zoomed image below reports the magnification of the corresponding frame.

It must be noted that the representation example described here does not use the Domain model repository, as the type of information to depict in the image does not have relation with the domain of the course. In the next chapter, we will illustrate other representations which use information stored in the Domain model repository to depict students' cognitive aspects related to the concepts taught in the course.

As said before, the objective of this paragraph is to describe the process of creation of a representation in CourseVis. The image in Figure 5.13 is just an exemplification of what can be produced using CourseVis and is not a complete example for the analysis of the messages generated by students. In fact, the usefulness of the picture in Figure 5.13 is questionable, and the analysis of the appropriateness of this image for the instructor's teaching activities is out of the scope of this chapter.

In the next chapter, we will discuss some representations produced in CourseVis taking into account the needs of instructors discussed in Chapter 4, and will illustrate how the requirements expressed by instructors led to the creation of a set of pictorial representations. The validation of these representations will be discussed in Chapter 7.

5.5 A model for the design space

When planning to implement graphical representations, it is important to consider the different data types and characteristics, because depending on them we can apply different transformations and visualisation mappings. We describe the structure of the design space using the design space model proposed by Card and Mackinlay (1997) and expanded in Card et al. (1999). This analysis, inspired by studies of Bertin (1983) on the semiotics of graphics, distinguish data in three basic types:

N = Nominal (an unordered set, like a list of names or film titles)

O = Ordinal (an ordered set. obey a $<$ relation, like months names)

Q = Quantitative (numeric values that can be manipulate arithmetically)

According to Bertin (1983) and Card and Mackinlay (1997), a visual presentation is originated by data and consists of **Marks**, **Positions** in space and time, and a set of **Retinal** proprieties. Marks could be one of the following:

Point (**P**)

Line (**L**)

Area (**A**)

Volume (**V**)

Position could be encoded on the graphical space coordinates: **X**, **Y**, **Z** in space and **T** for the time.

Retinal proprieties are proprieties processed by retina of the eyes which is sensitive to them independent of position. Bertin (1983) identify 6 retinal proprieties:

Colour (**C**)
 Size (**S**)
 Orientation (**O**)
 Gray Scale (**GS**)
 Texture (**TE**)
 Shape (**SH**)

Card and Mackinlay (1997) also add two proprieties which are considered in Information Visualisation:

Connection (-)
 Enclosure ([])

Connections are typical in topological structures like Graph and Trees and allow to show relations among objects (such as the connections between nodes in a graph). Enclosure is a particular way to visualise trees with a technique called *treemaps*, invented by Johnson and Shneiderman (1991) and described also in Card et al. (1999).

Card and Mackinlay consider other proprieties which can be present in Information Visualisations which require human **Controlled Processing (CP)**. They need a human internal process to be decoded. For instance, a text in a graphic must be read and processed by the mind before being comprehended. They are opposite to **Automatic Processing** properties, which work on visual proprieties such as position and colour, and are “*superficial, parallel, ... has high capacity, is fast, cannot be inhibited, is independent on load, unconscious, and characterised by targets 'popping out' during the search*” (Card et al., 1999, p. 25).

Thus, visualisations are composed from the following **visual vocabulary**, reported from Card and Mackinlay (1997):

Let us consider previous example reported in Figure 5.13. *Read* and *post* are variables mapped onto X and Y axis:

$$\begin{aligned}
 read &: Q \rightarrow X : P \\
 post &: Q \rightarrow Y : P
 \end{aligned}$$

This notation says that the *read* data variable, which is a quantitative data type, is mapped onto the X coordinate value of a point P, and the *post* data variable, which is also a quantitative data type, is mapped onto the Y coordinate of the point P.

$$student : N \rightarrow CP : tx$$

This latter notation says that the *student* variable, which is a nominal data type (i.e. not intrinsically spatial), is visualised directly with a text (*tx*) - a controlled processing propriety.

- Marks: (Point, Line, Area, Volume)
- Automatically processed Graphical Properties:
 - Position:(X, Y, Z, T)
 - Retinal encodings: (Colour, Size, Orientation, Gray Scale, Texture, Shape)
 - Connections
 - Enclosure
- Controlled Processed Graphical Properties

Table 5.3: The “visual vocabulary” defined by Card and Mackinlay (1997) used to formally describe the design space of representations in Information Visualisation and followed in CourseVis.

Using these distinctions, we can use a common table format for these proprieties. The following table reports description of representation in Figure 5.13:

Variable	D	X	Y	Z	T	R	-	[]	CP
read	Q	P							
post	Q		P						
student	N								tx

This notation will be used for the remaining part of this document to describe and analyse the structure of the design space, used in graphical representations.

5.6 Summary

This chapter presented the general structure of the CourseVis tool and illustrated how it worked with a sample representation generated using the tracking data of a course given with WebCT. The CourseVis architecture was inspired taking into consideration the principles of simplicity, integration and independence with an existing learning environment, domain independence, free and open source development. The architecture described is a generic software that can be programmed to generate specific representation of the data collected by the distance learning environment. In the next chapter we will explore some representations that have been generated using CourseVis to meet the instructors’ requirements we identified in the survey described in Chapter 4. In Chapter 7 the representations produced by CourseVis will be evaluated with distance learning instructors.

Chapter 6

Graphical Representations

In the previous chapter, we described the architecture and implementation details of CourseVis, together with a sample graphical representation which illustrated one of the possible images that can be created. CourseVis is a generic tool which can be programmed on a systematic approach for selecting appropriate pictorial representation which may be used to facilitate instructors involved in distance learning.

This chapter will illustrate how the requirements expressed by instructors discussed in Chapter 4 led to the creation of a set of pictorial representations expressly depicted to support instructors in their teaching activities. We will discuss several graphical examples based on data extracted from an existing distance learning course. Each representation is commented in details, and some particular information on students and facts derived from representations are analysed. We will also show how CourseVis is able to analyse data of a course in different periods, how the data of a course repeated in different editions could be compared, and how a possibility to manage courses in different domains can be provided.

6.1 The Java on-line course

To start, we need a set of example data to be applied with the illustrations we want to create. The University of Applied Sciences of Southern Switzerland (SUPSI) started the course *Introduzione a Java* (“*Java on-line*”) in March 2001 (Mazza, 2001). This course, dedicated to students participating in the postgraduate course on “Advanced Computer Science”, is taught every year, and backup copies of each edition have been collected for possible studies on students’ data and interactions. The course has been implemented using the WebCT Course Management System, and in each edition between 16 and 34 students participated. The author of this thesis has been the instructor of the course, and has full access to the data collected in the different editions of the course.

The second edition of the course, given between January and March 2002, lasted for 10 weeks



Figure 6.1: Home Page of the Java on-line course

and included 16 students. Students were professionals with computing experience who enrolled the course to complete or bring up to date their skills on the Java language. The course was structured in a mixed modality: interactions were completely on-line except for some face-to-face sessions with the instructor and other students on the first day of the course (dedicated to meet each other and illustrate the structure of the course and the requirements of the study), at the middle of the course (to check the students' progress), and at the last day of the course (to participate at the final test in classroom). The course was organised in 8 units, each of them comprising:

- Content materials in HTML or PDF format,
- A set of quizzes marked automatically by the system, with questions of types multiple choice, matching, fill-in the blank,
- A topic on discussion forums dedicated to the issues discussed in the unit,
- Some exercises.

The course included also some individual assignments and a session of group exercises (each group involved 3 students, who had to collaborate together to solve the exercise). These activities were compulsory. The main communication tool used was the discussion board, where students and the instructor engaged in discussions on topics of the course and, in some cases, exchanged messages to collaborate on group exercises. An e-mail tool was available as well, and some chat sessions on pre-defined topics were organised (but they had only few participants).

This edition of the Java on-line course will be used to illustrate the graphical representations of student data in CourseVis. This course was backed-up from its original WebCT environment and restored in another installation of WebCT integrated with CourseVis. The backup included not only the course materials, but also all the student data, discussions, quizzes, submissions, marks, etc. Students' names are replaced with alias for privacy issue. Graphical representations will show, if not indicated otherwise, data gathered at the end of the course.

6.2 Graphical representations produced in CourseVis

This section illustrates pictorial representations generated in CourseVis. We start from the requirements expressed by instructors in the survey discussed in Chapter 4 and will show some representations which meet the instructors requirements identified in the survey. In line with the themes of the survey, the three aspects - social, cognitive and behavioural, are considered in the representations. We will recall the requirements of distance teachers related to each of these themes (in particular, the implication for the research and practice discussed in Section 4.7) and will give some pictorial representations which externalise appropriate information to instructors.

6.2.1 Social aspects

Survey analysis (see Section 4.7.4) indicated that representations of students' social activities in the course should consider the data provided by the communication facilities included in that course, i.e. the discussions board and data of participation in group work. Participation in group work can only be measured based on the messages exchanged between members of a group, as students usually work on their own at distance on a specific part of the work, and then put each part together to compose a final solution¹. Specific private areas of groups in discussions must be created if we want to distinguish discussions about the group exercise from regular discussions on other course's topics.

For the social aspects of learners, we produced two graphical representations: the '*Discussion plot*' and the '*Discussion graph*', described in the following section.

Discussions plot

Discussion plot represents data collected from the discussion board. The discussion board is a tool which allows students to exchange (read and post) messages. Each message has a sender, a

¹In the Java on-line course the group work concerned the implementation of program. Each student in a group worked on some Java classes of the program, which were previously defined during a project phase by the group. Then one of the students, designed as coordinator, put all the classes together to compile the expected program and submitted the solution to the instructor. All discussions between members of the group were made on a private area of the discussion board, using e-mails, or at presence. Discussions through the discussion board have been encouraged by the instructor in order to monitor group activities.

date, and a topic. Messages are usually displayed by threads - a list of messages arranged by a subject and in chronological order is a thread. The length of a thread (i.e. the number of messages written on a specific subject) is called **follow-up** and can be derived. We want to create a graphical representation that illustrates for each thread the student who started the thread, we call him the issue “**originator**”, along with the **date** it was created, the **topic** where the thread is located, and the follow-ups received. Hence, the variables to consider in the representation are originator, date, topic and follow-up. A 3D scatterplot is used to represent data about student discussion threads, see Figure 6.2. The first three variables (originator, date, topic) are mapped in the three dimensions of the scatterplot:

$$\text{originator} : N \rightarrow X : P$$

$$\text{date} : O \rightarrow Y : P$$

$$\text{topic} : N \rightarrow Z : P$$

The final variable, follow-up, is mapped onto retinal proprieties, Colour and Size, which is represented with a sphere:

$$\text{follow-up} : Q \rightarrow C, S$$

A sphere denote a specific thread of discussions, and the size and the colour of the spheres represent the number of follow-ups in a discussion thread. Sphere colour scale, from blue (min) to red (max), is reported on the top right of image.

The following table summarises these proprieties:

Variable	D	X	Y	Z	T	R	-	□	CP
originator	N	P							
date	O		P						
topic	N			P					
follow-up	Q					C,S			

The instructor may use operations, such as rotating and zooming, to manipulate the image. Figure 6.3 shows two rotations that enable the analysis of different relations by examining the same data set. The top scatterplot presents the discussion threads open by each student during the course. Certain social characteristics of students can be discovered. For instance, it can be seen that the discussions were predominantly initiated by students, and two individuals (*Francesco* and *Massimo*) dominated in opening threads followed by others. The instructor might delegate communication tasks to such students, e.g. to moderate discussions. *Domenico* has also been active in opening discussion threads but has not been followed by others. The instructor may decide to look at some of the threads initiated by this student and encourage longer discussions, if

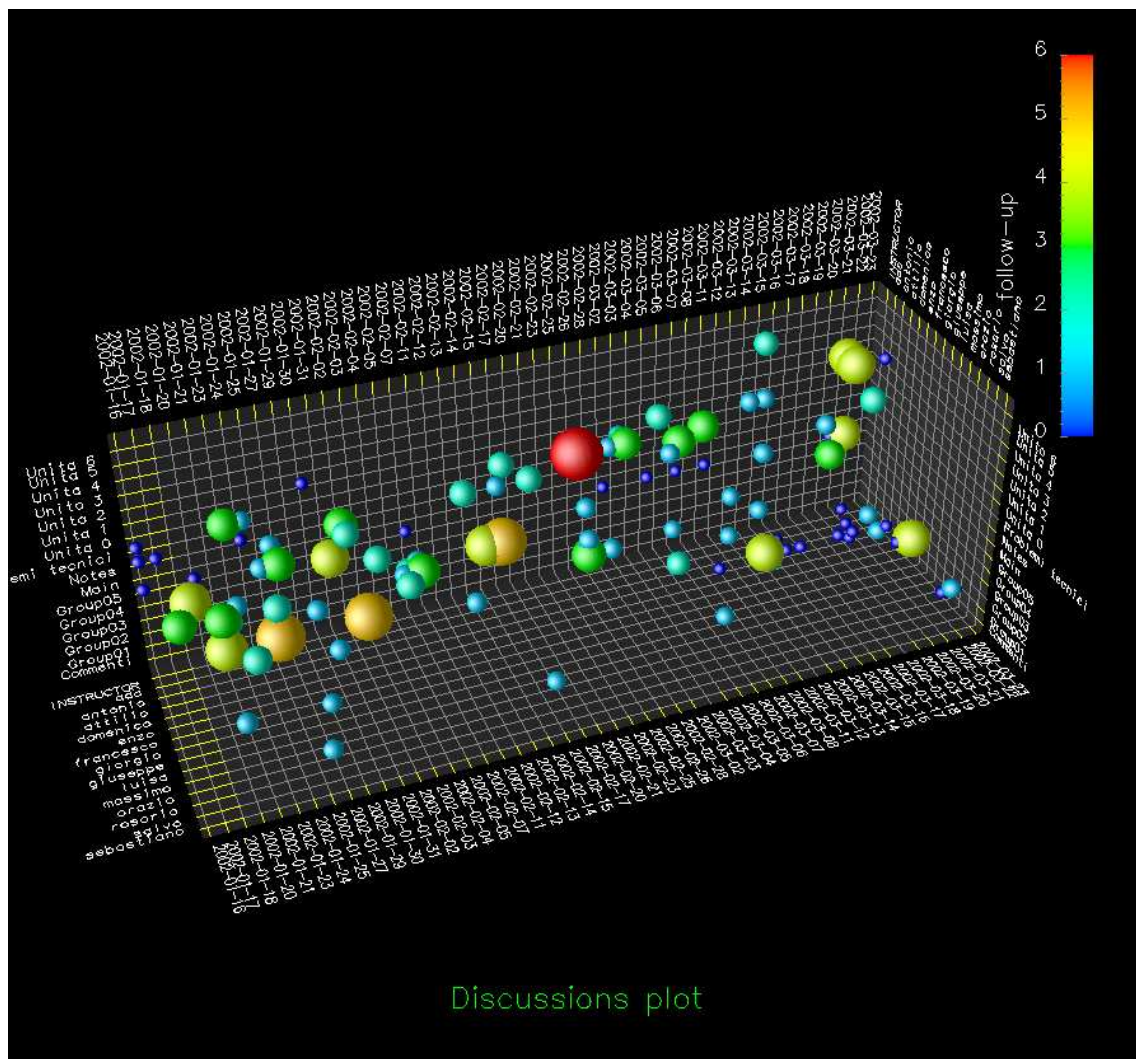


Figure 6.2: The discussions plot, represents discussion threads made using the discussion board tool in CMS. The diagram shows a ball for each thread, positioned on the student who started the thread, the date when the thread started, and the topic of the discussion. The size and the colour of the ball are proportional to the number of follow-ups.

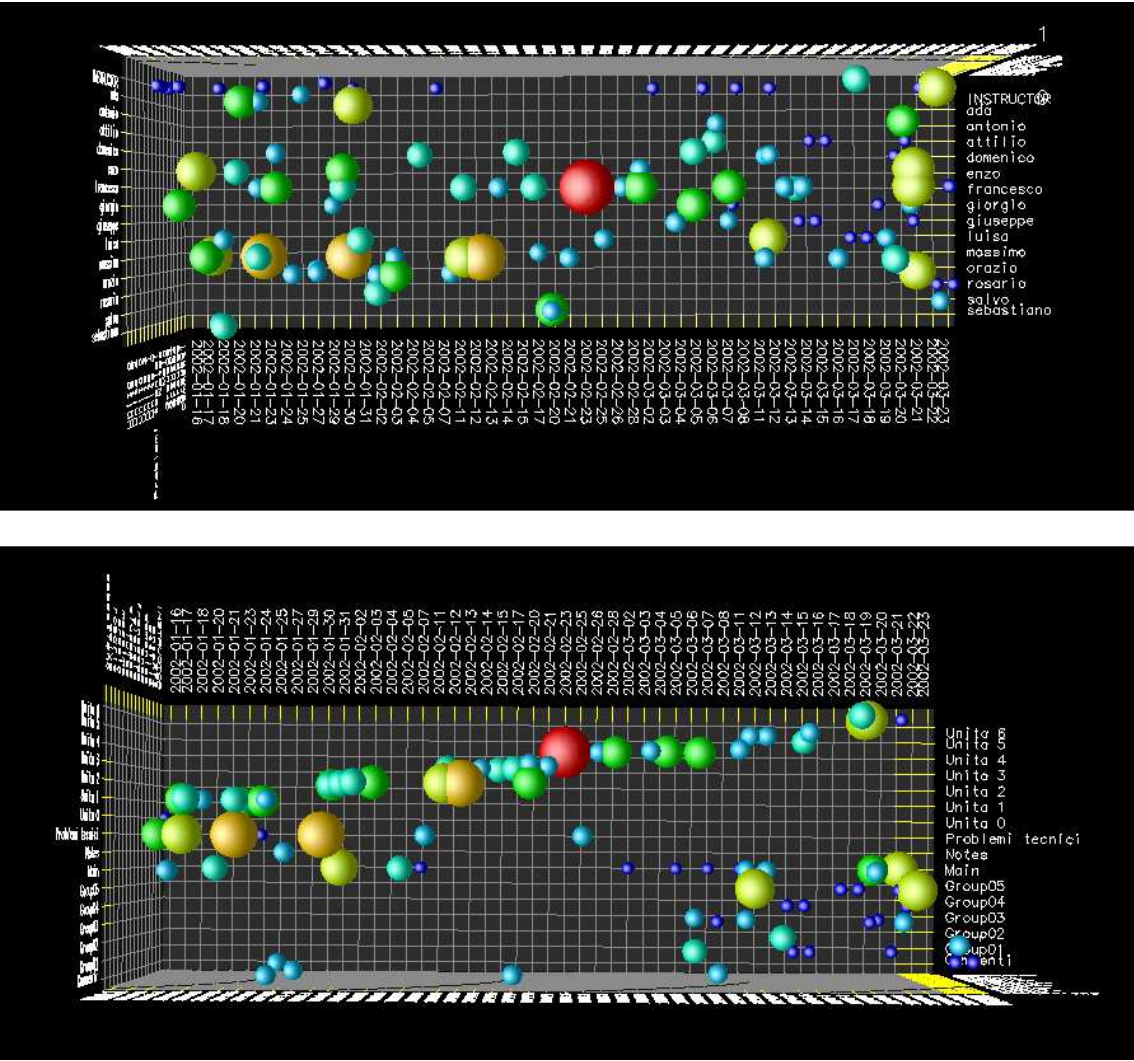


Figure 6.3: Results of two rotations of a discussions plot, top image focuses on students, bottom image focuses on the topics.

appropriate. The figure identifies students who have not been active in opening discussions. The instructor may need to pay more attention at these students.

Using the bottom representation in Figure 6.3, the instructor can gain insights into the discussed topics. Some threads can be identified, e.g. there are lengthy discussions about technical problems (*problemi tecnici*) at the beginning of the course, most units have had fairly active discussions. There has been a group work assigned, the top lines of the representation show the communication in each group. While group 5 has communicated intensively, groups 3 and 4 have not communicated at all (at least not using the tools in the course). The instructor may intervene in the group work by encouraging the students to communicate and collaborate more effectively.

Discussion graph

For each student, the instructor may be interested to know the **number of threads started** and **number of follow-ups received for those threads**. In this case, information reported in the previous diagram can be condensed in a trivariate representation describing these variables:

Variable	D	X	Y	Z	T	R	-	□	CP
student	N	P							
threads started	O		P						
follow-ups	N		P						

These variables can be represented in a 2D-diagram using the *x-axis composition* technique (see Section 3.2.5). See Figure 6.4 for an example². Stars denote the number of threads started by a student, while the triangle denote the number of follow-ups received in threads originated by the student.

Using the representation in Figure 6.4, the instructor can gain insights on the feedback received by colleagues on threads generated by students. For instance, *massimo* created about 18 new threads, which received more than 40 messages in follow-up. This might mean that discussions proposed by *massimo* stimulated the interest of others who got engaged in discussions. There are other students, such as *attilio* and *rosario*, who promoted few new threads and received even less follow-ups. The instructor may be interested to check messages posted by those students to verify the reason of such low feedback and decide some actions to take.

The two graphical representations that illustrate the discussions engaged in the course, the Discussion plot and the Discussion graph, represent in a graphical format the same source of data collected from the discussion tool of the CMS. The Discussion plot contains information about dates of the course and topic of discussion, which is not considered in the Discussion graph.

²This representation was created in a first version using a continuous line plot. It was then substituted with a scatterplot after the comment reported by the focus group (described in chapter 7).

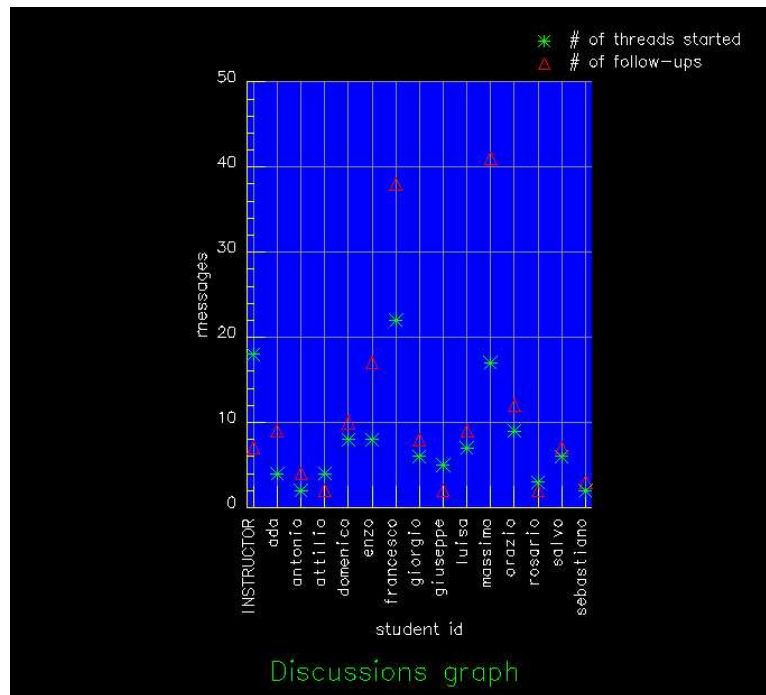


Figure 6.4: The discussion graph in CourseVis.

Hence, the Discussion graph is a summary of data already represented in the Discussion plot, but gives a quick overview on the number of threads started and follow-up received by each student.

Some other information available by the discussion tool of the CMS, such as the list of students who participated in a thread and the date which a thread finished, is not taken into account in these two representations because considered not of primary importance by instructors during some informal discussions.

6.2.2 Cognitive aspects

Survey analysis (see Section 4.7.4) indicated that representations of students' cognitive level should consider for each student the overall course performance and on specific concepts of the domain. Representations should highlight students having difficulties with a concept and facilitate comparison of a student's progress with that of the whole class.

For the social aspects of learners, we produced two graphical representations: the '*Cognitive matrix*', and the '*Cognitive plot*', described below.

Cognitive matrix

The source of information we considered to assess the students' understanding of the subject are the results from quizzes. For each student the CMS registers the answers to each question, the time taken to answer the question, how many times a question was answered and the marks he received for each answer to a quiz question. Marks are expressed in numerical format with a natural number. Each question is related to one or more domain concepts by the means of the domain model of the course. As discussed in Section 5.2.4, each concept (C) is composed by a list of pages (P) belonging to the content materials of the course and a set of questions (Q) selected among the quizzes created for the course that test for concept C , i.e. we have:

$$C_i = (\{P_{i_1}, \dots, P_{i_h}\}, \{Q_{i_1}, \dots, Q_{i_k}\})$$

Pages of the content materials and questions in quizzes would deal with multiple concepts, hence a single page or a single question could belong to different concepts.

This simple approach allows CourseVis to link marks on questions with concepts of the domain. However, at a particular time of the course, a student may have not answered all questions $\{Q_{i_1}, \dots, Q_{i_k}\}$ related to a concept C_i . For this reasons, if we consider a **student** S , a **concept** C_i , at the time t of calculation, we consider the following:

$$A_i = \{Q'_{i_1}, \dots, Q'_{i_r}\} \subseteq \{Q_{i_1}, \dots, Q_{i_k}\}$$

A_i is the set of questions that test a concept C_i answered by the student at the time of calculation t .

Let $M_{S,Q_{i_j}}$ the mean of the marks³ received by S on question Q_{i_j} which tests the concept C , and $T_{Q_{i_j}}$ the maximum value for marks on question Q_{i_j} , we can define a **performance** value P_{S,C_i} of the student S con the concept C_i as the following result:

$$P_{S,C_i} = \frac{\sum_{j=1}^r \frac{M_{S,Q'_{i_j}}}{T_{Q'_{i_j}}}}{r}$$

In other words, the performance value achieved by student S on concept C_i is the mean of normalised values of marks received by S for questions which test the concept C_i . If there are no question testing for a concept C_i then the student's understanding of C_i cannot be determined using this model.

The variables to consider in the graphical representation and their mapping are the following:

³We recall that a student can answer each question several times. The Java on-line course put a limit to three times a question can be answered. The mean of the marks received for each attempt is considered in this model.

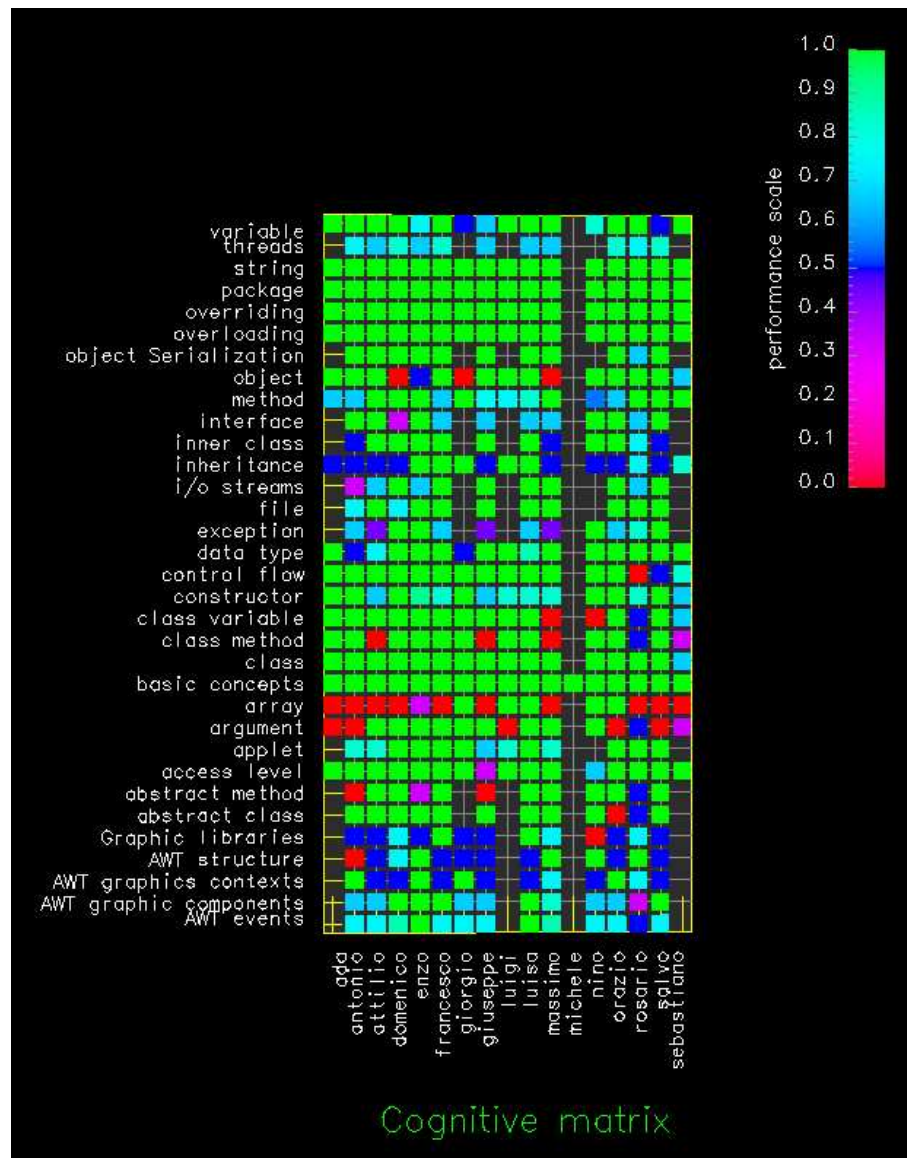


Figure 6.5: The cognitive matrix, the colour in each cell represents performance level of a student on a specific concept. A blank (uncoloured) cell means that student has not done the relative quizzes yet.

Variable	D	X	Y	Z	T	R	-	[]	CP
student	N	P							
concept	N		P						
performance (P_{S,C_i})	Q					C			

A coloured matrix is used to represent data about students' performance on the course quizzes, see Figure 6.5. The students are mapped on the x-axis and the concepts are mapped on the y-axis of the matrix. The performance values are mapped on the colour of the mark corresponding to a student and a concept, which is represented with a square. If a student didn't try to answer any questions related to a concept, the corresponding cell is left blank. The colour scale is, from red (min value=0) to green (max value=1)⁴.

The matrix enables global analysis of the overall performance of students on the course topics and comparison between topics. This can promote the instructors' reflection on their practice (see also Section 6.7). For instance, surprisingly, *array* was found a difficult concept by most of the students, and looking at the course material, the instructor found vague presentations and quizzes that had to be improved. The matrix also allows local analysis of the performance of a particular student on a specific topic and comparison between students. For example, it can be identified that even though *massimo* was successful in opening long discussion threads (see Figure 6.3), his performance at quizzes wasn't particularly brilliant. This may point out that this student could be encouraged to pay more attention at reading course material.

The representation in Figure 6.5 uses the *micro-macro* design approach (Tuft, 1990). The example also shows the use of colour for *layering and separation*. Both micro-macro and layering and separation methods are described in Section 3.2.5.

Cognitive graph

Performance values registered by students on a specific topic of the course can be summarised in a **mean value**. We define *performance mean* (PM) of students on concept C_i as the following value:

$$PM_{C_i} = \frac{\sum_{j=1}^n P_{S_j, C_i}}{n}$$

where $S_1 \dots S_n$ are the students enrolled in the course who have done at least one quiz question on concept C_i . With this calculation the information on student names is lost, hence the list of variable and their mapping is the following:

⁴In a first version of the program the range of the scale colours for the performance rate was from blue (min) to red (max). In the actual version of the program we changed the colours accordingly to the indications received by instructors in the evaluation of the program.

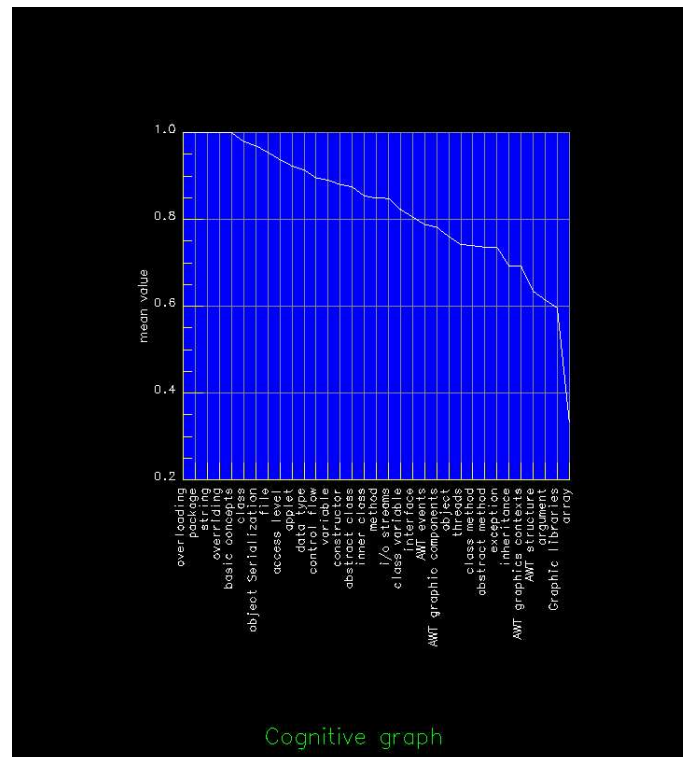


Figure 6.6: The cognitive graph, for each concept of the course is measured the mean of performance values achieved by students on evaluation profs.

Variable	D	X	Y	Z	T	R	-	[]	CP
concept	N	P							
performance mean (PM_{C_i})	Q		P				L		

This bivariate data can be represented with a simple plot relating the two variables, as shown in Figure 6.6. Course concepts are mapped in the x-axis and the performance mean is mapped in y-axis. Concepts are ordered by the resulting mean and linked with lines (L). This simple plot contains less information than the cognitive matrix, in particular the local analysis of the performance of a particular student on a specific topic has been lost, but it allows a quicker and more precise evaluation of problematic concepts. For instance, at a glance the instructor may notice that the concept *array* received a very bad performance (this was already noticed in the cognitive matrix). Moreover, it can be noticed that concepts *graphic libraries*, *argument*, and *AWT structure* received a mean performance near the 60%, which is relatively poor. The instructor may decide at this point to look at the quizzes and course materials and eventually improve parts which are unclear.

The two representations shown, namely the Cognitive matrix and the Cognitive graph, are

just two of the numerous representations that can be produced with this data. The choice of the representation depends on several issues, illustrated in Section 3.2.2, but also on some subjective issues, such as aesthetics or understandability. Among all the possible representations that can be drawn with data on quizzes, the representations in Figure 6.5 and Figure 6.6 have been chosen taking into consideration some subjective issues given by the experience of the author in distance education, and considering some informal discussions with distance instructors.

6.2.3 Behavioural aspects

Survey analysis (see Section 4.7.4) indicated that representations of students' behaviour should consider the students' attendance to the course, the reading of materials, the performing of evaluation proofs, the participation in discussions, and the students' progress with the schedule of the course.

For the behavioural aspects of learners, we produced two graphical representations: the '*Students accesses plot*', and the '*Student behaviour graph*'.

Students accesses plot

At each access of the **student** in the course, the CMS registers in its internal data structures the **date** and the **time**. We can take advantage of this information to create visual representations that summarise the overall students' attendance to the course. The instructors may be interested also in knowing how many times the course has been accessed by all students on a particular day D of the course. Let $NA_{S_i,D}$ is the number of accesses made by a student S_i into the course on date D . We compute the **global accesses** (GA) of all students on the day D in the following way:

$$GA_D = \sum_{i=1}^n NA_{S_i,D}$$

where $S_1 \dots S_n$ is the list of students enrolled into the course. Currently the time of access is not considered and such data may be used in future versions of CourseVis.

The variables to consider in the representation and their mapping are:

Variable	D	X	Y	Z	T	R	-	[]	CP
student	N		P						
date	O	P							
global accesses	Q		P						

We choose to represent students' access to the course with a image derived from a composition of a scatterplot and an histogram, see Figure 6.7. The scatterplot represents the bivariate data of

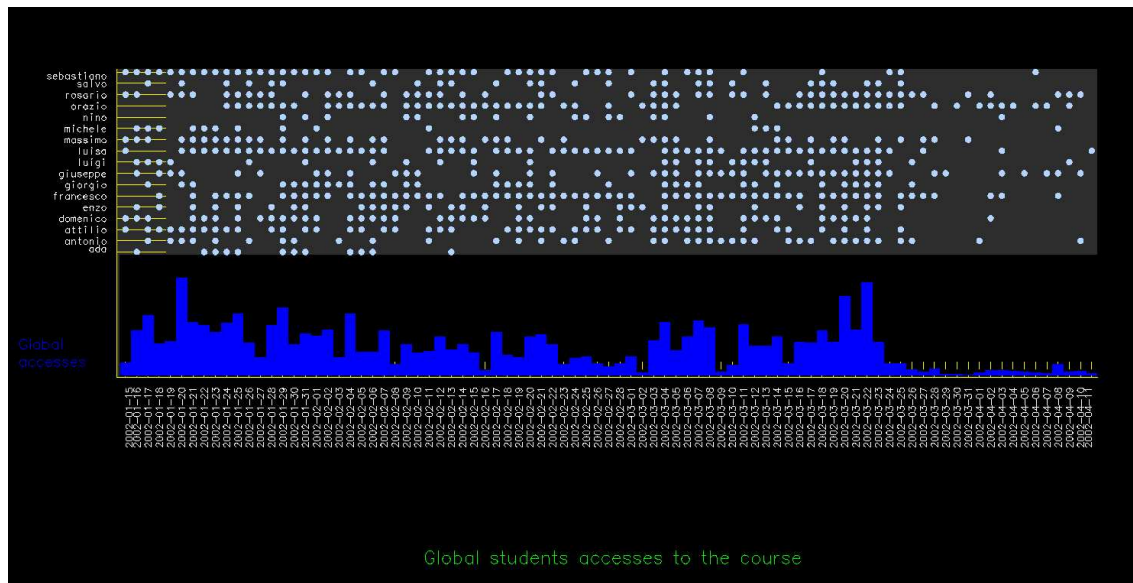


Figure 6.7: The students accesses plot, each circle means that a student logged into the course on the corresponding date. The histogram reports the number of global accesses made by students on each date.

students and dates of the course, a corresponding bullet represents at least one access to the course made by the student on the selected date; the histogram represents the number of pages of the course accessed by all students on a date.

The instructor has an overview at a glance of the global students' access to the course with a clear identification of patterns and trends. For instance, there is a high concentration of access between the 4th and 23rd of March, corresponding to the group activities, while little access is made after March 23rd, when the course is over. The same image gives attendance information about a specific student, e.g. *michele* made few visits to the course, while *francesco* was quite consistent.

This representation doesn't give the information about how many accesses have been made by a student on a day of the course, which will be included in the next representation.

Student behaviour graph

Figure 6.7 illustrates the students' attendance to the course. Other behavioural aspects are covered in the representation we called '*Student behaviour graph*', illustrated in Figure 6.8. It represents information regarding a particular student, and takes advantage of a *single-axis composition* method (see Section 3.2.5) for presenting large number of variables in a 2D metric space. With a common x-axis mapping the dates of the course, a number of variables are represented in Figure

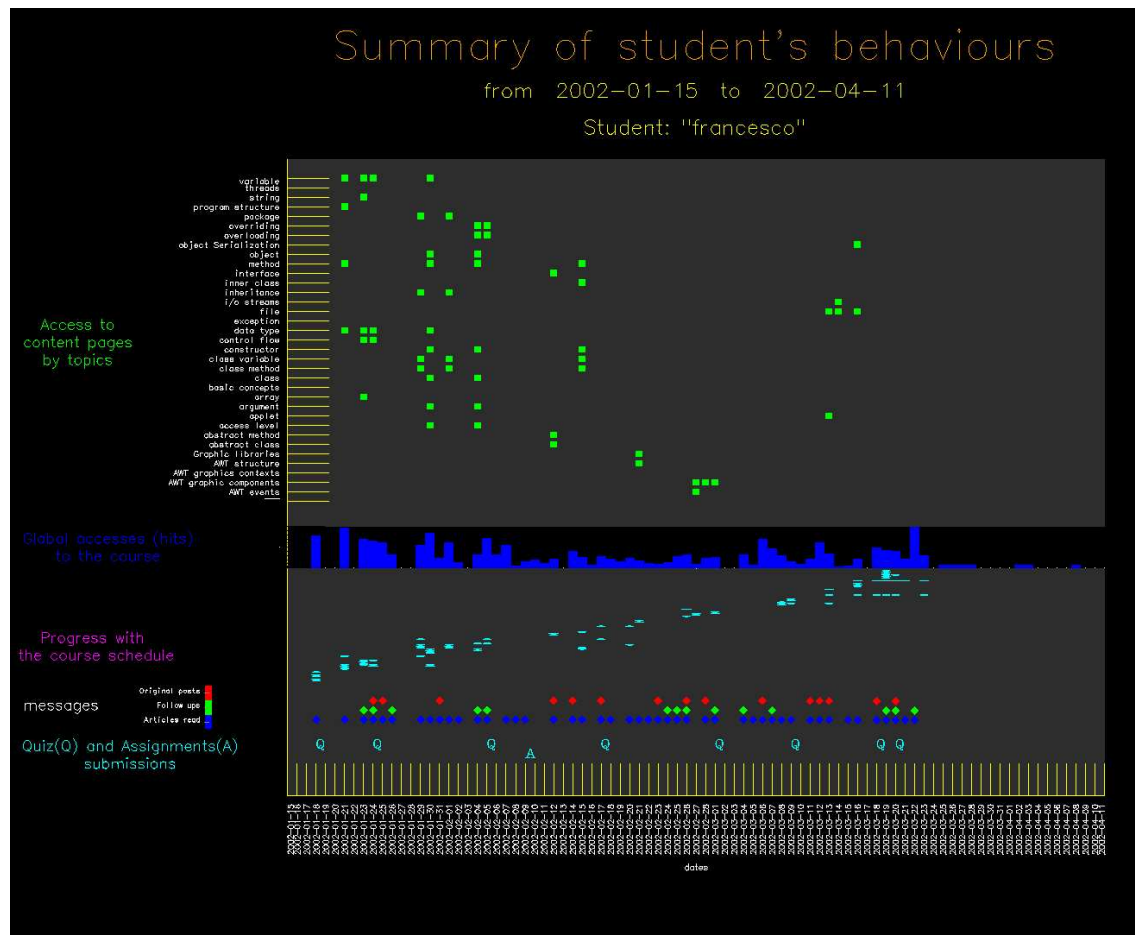


Figure 6.8: The student behaviour graph. This representation include dense information about a single student using the composition technique: access to the content pages, global accesses to the course, progress with the schedule of the course, messages read, posted and replied to discussions, and submission of quizzes and assignments.

6.8:

- *Accesses to content pages by topics* - if a student accessed a content page of materials dealing with a particular concept, it is marked with a green square. The instructor can be informed whether the concepts are studied by the student, as well as whether a student repeatedly accesses particular concepts in different dates.
- *Global accesses to the course* - the number of accesses to the course (not only content material, but to any part of the course) on a single day are computed and this value is reported on the blue histogram. The length of the bar of the histogram on each day is proportional to this number. The instructor can detect days or periods with high activities or with low accesses.
- *Progress with the course schedule* - we assume that the course has a linear structure, i.e. it has a syllabus with a list of content pages having a schedule organised in a linear sequence. We put in the y-axis the page ordering according to this sequence, the cyan mark then represents a page accessed by the student on a day, where the y-location of the mark is proportional to the sequence of the pages in the schedule (first page on bottom, last page on top). The instructor may have an overview of the progress made by the student with the schedule during the time, which is useful to determine the learning style of the student (e.g. a student who accessed many pages at once and then doesn't read anything for a long period, versus a student who follows a more constant and progressive reading of materials). The representation in Figure 6.8 shows that the student *francesco* was particularly persistent and methodical in reading the course materials following the schedule of the course. We found that other students of the java on-line course weren't very constant or even didn't read some parts of the content materials (see for instance student *salvo* in Section 6.6).
- *Messages* - some information about the participation in discussion has been already considered in the discussion plot (see Figures 6.2 and 6.3). However, some further information on discussion attendance could be given. In particular information about the posting of a new message, the reply to a message, or simply the reading of a message is denoted in the representation with a diamond coloured respectively in red, green and blue. The instructor may use this information to determine whether a student is particularly proactive, and if he is reading messages regularly (like the student *francesco* in the example showed), or if he didn't read messages at all.
- *Quiz and assignment submission* - the event of submission of a quiz or an assignment are reported in the graphic, as on the date it was submitted a *Q* or *A* is included. The instructor may use this information to assess the submissions of a quiz and assignment.

The *Student behaviour graph* representation makes use of the *layering and separation method* (see Section 3.2.5) to visually differentiate various aspects of the data by distinctions of colour, shape and size. The mapping of variables involved in this representation is shown in the following table:

Variable	D	X	Y	Z	T	R	-	[]	CP
date	O	P							
global accesses	Q		P						
concepts	N		P						
hits	Q		P						
content pages	O		P						
messages			P			C			
quiz submissions			P						
assignment submissions			P						

6.3 Visualisation of different stages of the course

So far, all graphical representations illustrated courses already completed. However, often the students' data is analysed from instructors while a course is running, rather than when it is already completed. CourseVis allows instructors to represent student data during a particular stage of the course. Moreover, graphical representations could be saved as image files, and compared to analyse the course progression during the time.

An example is given in Figure 6.9. It reports the cognitive matrix of the Java on-line course at three different periods of the course. From left to right, the image reports cognitive matrices for students data on days, respectively, February 10th, March 10th, and March 23th⁵. Concepts not yet covered in quizzes are not reported on y-axis.

The images in Figure 6.9 show the use of '*small multiples*' (see Section 3.2.5) - a method used in Information Visualisation to compare at a glance a series of graphics showing the same combination of variables while another variable (the time, in our case) changes. These images allow to compare students' progress with performance on quizzes while the course is ongoing. For instance, student *sebastiano* performed very badly on several concepts at the beginning of the course (concept *object*, *class variable*, *class method*, *class*, and *array* got the lowest mark). However, quizzes in the Java on-line course can be repeated up to three times, and marks are combined in a mean in the case of repetition. The image in the centre reports that *sebastiano* improved his performance value on some concepts as he repeated some quizzes, achieving better results. However, the concept *array* remained in red colour up to the end of the course. This means that *sebastiano* didn't try to repeat quizzes on *arrays* or still performed badly on further attempts.

⁵The Java on-line course started on January 16th and finished on March 23th.

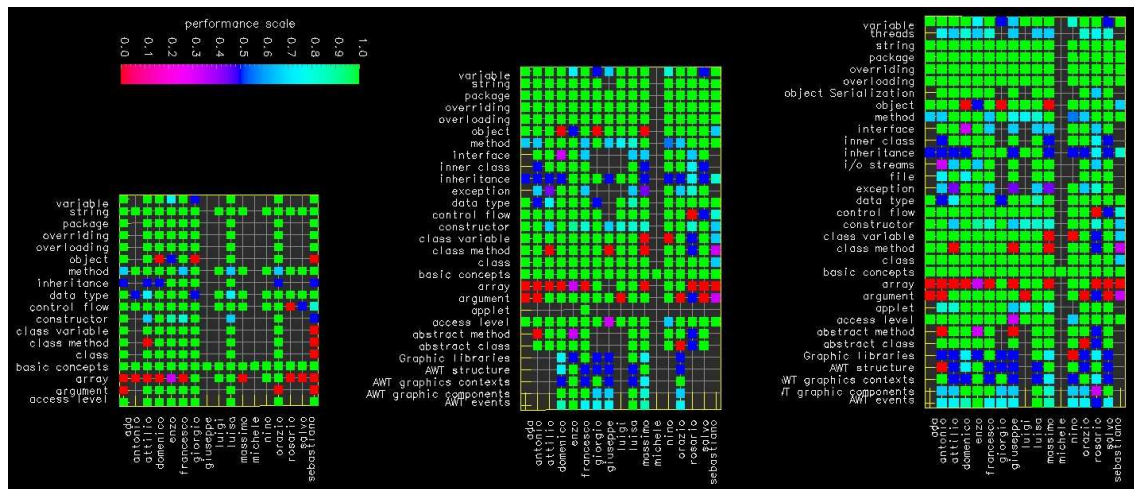


Figure 6.9: Three cognitive matrices representing students' level of knowledge on concepts in three different periods of the course. On left is reported a matrix corresponding to the student performance at about 3 weeks after the course started, matrix on the centre at about 7 weeks after the start and matrix on the right at the end of the course, at about 10 weeks after the beginning.

Other information about students' participation in quizzes is identifiable in Figure 6.9. For instance, at the beginning of the course, the students *giuseppe* and *michele* answered only quizzes related to *basic concepts*. However, it can be seen that after 7 weeks (the matrix on the centre) the student *giuseppe* answered quizzes on concepts not yet done at the beginning, while the student *michele* still remained on the same situation revealed 4 weeks before. The graphics enables that the instructor becomes aware of the situation with *michele* and he may encourage the student to work on the quizzes.

6.4 Visualisation of different editions of the same course

On-line courses have the characteristic of being repeated several times, with little (or even nothing) changes between different editions. The first edition of the Java on-line course was proposed between March and May 2001, then a second edition of the same course was proposed between January and March 2002. The content and the structure of the course remained the same, with few changes on the course materials, leaving the syllabus and the structure of units unchanged. This enables a comparison of students' results on different editions of the same course, and is useful for the “*monitoring and evaluation of the impact of changing content and assessment methods within courses*” (QAA, 2001, p. 3).

Figure 6.10 shows the cognitive matrices at the end of the course for the two editions of the Java on-line course. The 2001 edition (image on the left) enrolled 34 students, the twice of the

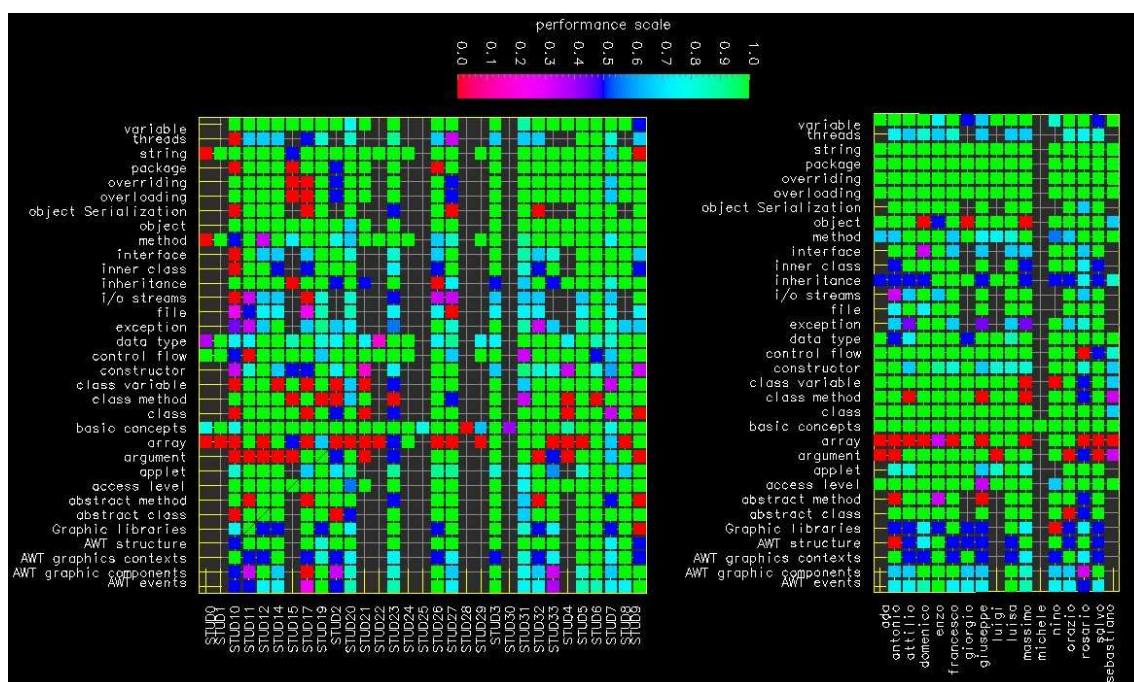


Figure 6.10: Cognitive matrices for the 2001 edition (on the left) and the 2002 edition (on the right) of the Java on-line course.

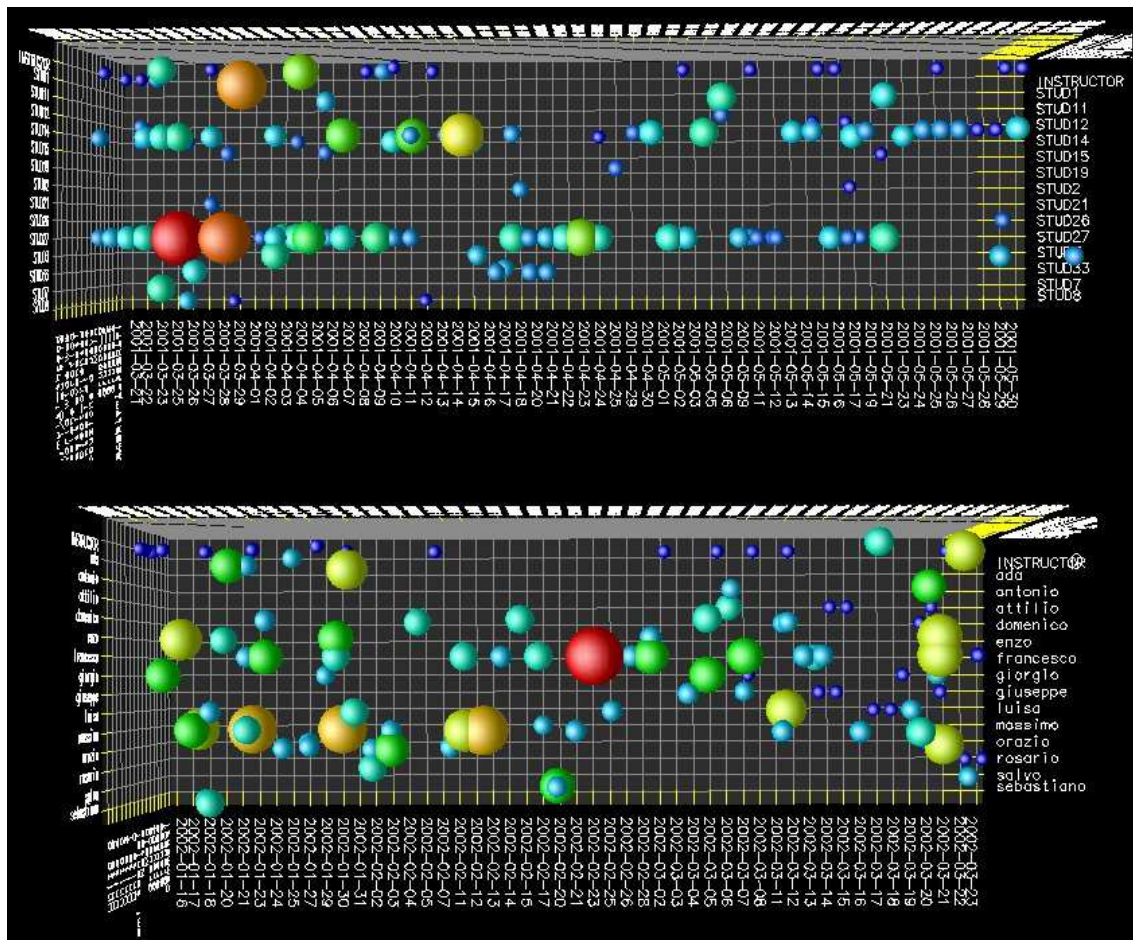


Figure 6.11: Discussion plot for the 2001 edition (top) and 2002 edition (bottom) of the Java on-line course. The 3D scatterplots are rotated to focus on students.

2002. At a glance we can see that one of the concepts with which students had problems, *array*, was problematic on both editions. This may prompt that there may be something wrong with the quizzes related to this concept. The matrix of the 2001 edition reveal a high number of students doing very few quizzes, while in the 2002 edition all students, except one, where more active in doing quizzes.

Another useful comparison between editions of the course can be made on discussions. In Figure 6.11 the 3D scatterplot of discussions for the two editions of the course, focusing on students, are illustrated. It can be seen immediately that in the 2001 edition of the course (image on top) there was a big predominance of two students (*STUD14* and *STUD27*) in starting new discussion threads, while other students where just reading or replying to discussions started by those two students. Instead, in the 2002 edition of the course discussion threads where initiated by different

students, and threads were discussed by more messages.

6.5 Scalability of the representations

All graphical representations illustrated so far reported students' data from the Java on-line course. However, CourseVis is a generic tool that can be applied to different courses and domains. In principle, every course that can be easily broken up into constituent concepts can be handled by CourseVis. Nevertheless, we need a model of the domain which creates a mapping between questions of quizzes and content pages with concepts (see Section 5.2.4).

In Section 5.4.6 it was illustrated the production of a diagram for a course with 231 students (see Figure 5.14). That diagram illustrates data from the course "*Fondamenti di Semiotica verbale*" (Verbal Semiotics) taught to the undergraduate students at the Faculty of Communication Sciences of University of Lugano. That course involved a large number of students, hence some graphical representations may reveal some presentation problems. Information Visualisation techniques like *zooming*, *panning*, *scrolling*, and *filtering* may help in this case.

See for instance in Figure 6.12 the accesses plot for the course Verbal Semiotics. The high number of students may cause problems with the readability of the data, because the diagram is too big to fit in the display of a computer. The zooming and panning techniques already used for the image in Figure 5.14 may help in this case.

As a further example Figure 6.13 reports the discussion plot for the course Verbal Semiotics, with focuses on students.

6.6 Case studies

In this section we will discuss potential and benefits of visualisations produced by CourseVis based on some case studies of students of the Java on-line course. To start our analysis, we consider the cognitive matrix illustrated in Figure 6.5.

About student Michele: demotivated, passive, drop-out

In the cognitive matrix we can see that *michele* has done only the quizzes regarding basic concepts (that was the first concept explained in the course materials). At a first glance, it seems that he dropped out the course very soon. But, if we look at the students' accesses plot (Figure 6.7) we can see that he sporadically accessed the course in the first half and at the end of the course. It is interesting to have a look at the behaviour representation, illustrated in Figure 6.14. It can be seen that he has done some activities (particularly quizzes and messages read) at the beginning of the course, then nothing up until April 28th, when the course was already over, he accessed some pages (he probably printed some content pages he was interested for). Surprisingly, during the accesses



Figure 6.12: The students' accesses plot for the course Verbal Semiotics, having 231 students enrolled

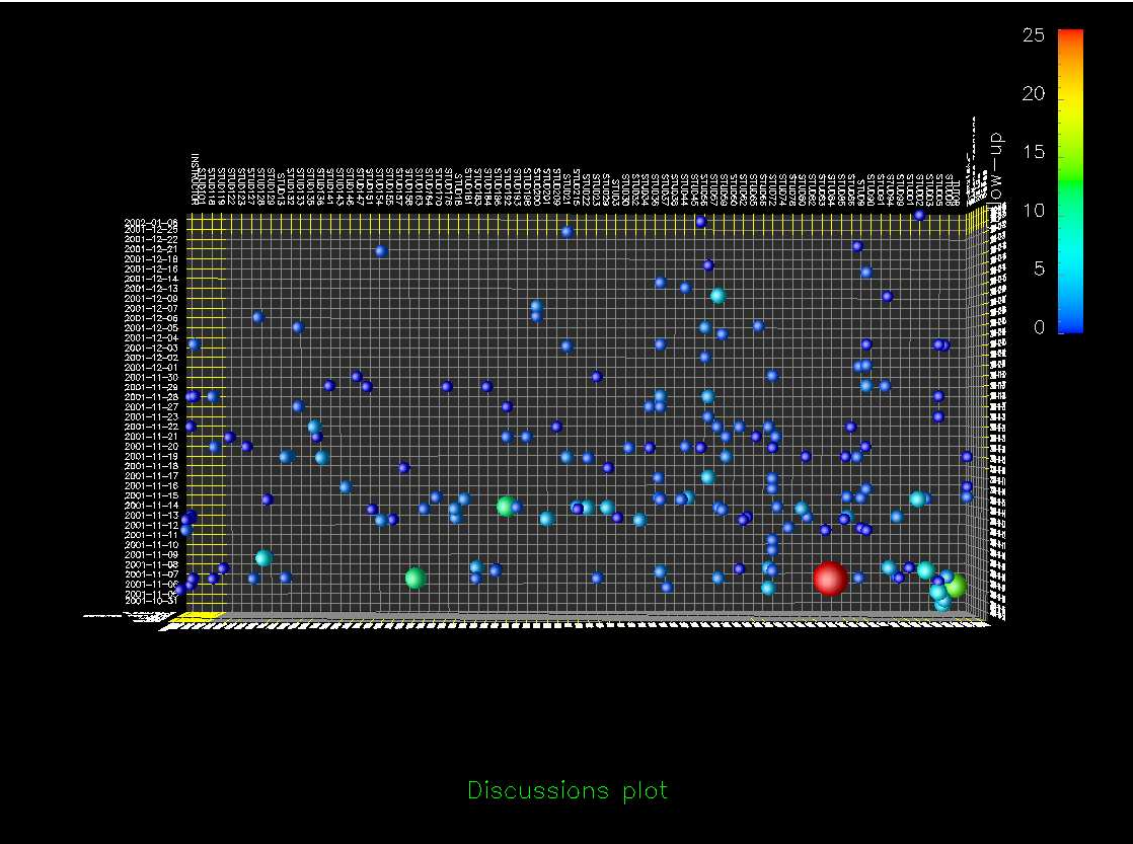


Figure 6.13: The discussion plot for the course Verbal Semiotics.

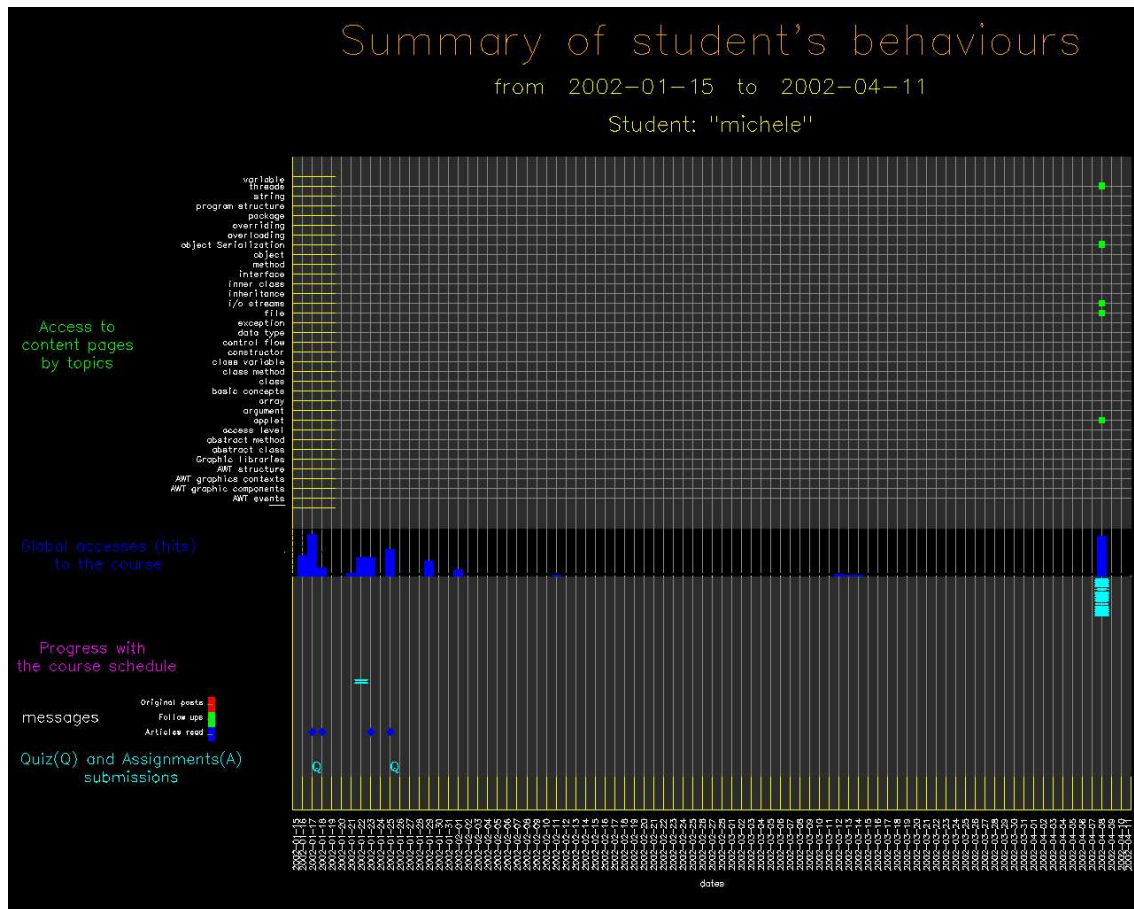


Figure 6.14: The student behaviour for the student *michele*. The representation include a grid for a better location of objects in diagram.

at the beginning he did some quizzes and read a few messages without reading the content pages for any concept. The student didn't participate in the final test in classroom.

This suggests that the student is unmotivated in participating to the course. In the cognitive matrix registered 3 weeks after the course started (see Figure 6.9) the instructor had the opportunity to check that *michele* was doing very little in the course, and the behaviour graph could have pointed that *michele* probably tackled the course in a wrong way, or he had different expectations, or he hadn't had time to study because of some unexpected commitments at work. In any case, the instructor could have had the opportunity to detect this situation in time to do appropriate actions, for instance personal e-mail communication.

About student Sebastiano: passive reader in discussions, sudden drop of motivation

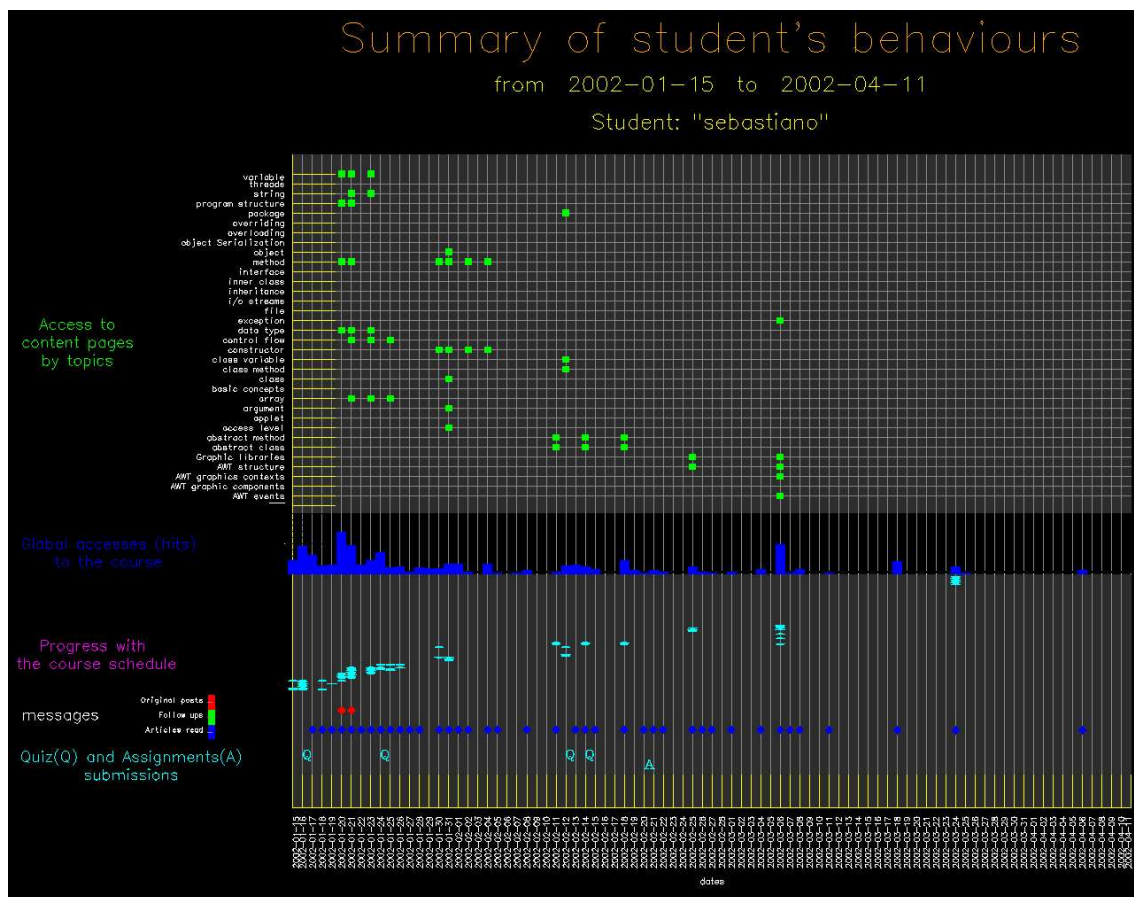
With respect to other students, *sebastiano* didn't do quizzes for many concepts, and received limited performance values on concepts started. The progress made on concepts illustrated by the three cognitive matrices in different periods of the course shows however that he improved his performance on quizzes, but during the second half of the course his attendance (see the students accesses plot in Figure 6.7) was progressively decreasing. His behaviour (see Figure 6.15) illustrate that he was very regular on accessing, reading materials, and reading messages; but after March 8th he didn't attend the course that regularly. Except for some days at the beginning of the course, *sebastiano* didn't post messages in discussions. He didn't participate in the final test.

We don't know the reason why *sebastiano* suddenly lost interest in the course, but visualisations show that he actively observed the interactions between others, without taking an active part in discussions. The instructor could have done something to stimulate this student in being more active in interactions with other students. If *sebastiano* was more active in contributing to discussions some problems could have probably been addressed, and he might have taken advantage of the help of peers or the instructor to overcome these problems.

About student Salvo: bright student, little interest in the course, perhaps bored

The cognitive graph shows that *salvo* is doing quite well on concepts performance. In discussion plot reveals that he opened a couple of discussion threads, with low feedback, and the accesses plot shows that the attendance was very scattered (but constant), and it was more consistent during the period corresponding to the group activities. His behaviour illustrated in Figure 6.16 is very different from that seen before. While other students, such as *francesco* and *sebastiano* accessed the course almost every day and read the content pages regularly, *salvo* accessed few days, leaving sometimes a gap of several days between accesses. Is it surprising that he was able to answer quizzes with good performance having read very few course materials. At the final exam in classroom he received one of the highest marks among the students.

Salvo was one of the special cases (almost no accesses to materials vs. one of the best students in tests and final exam). This situation was unknown to the instructor before seeing graphics produced by CourseVis. This can be explained with the fact (verified by the instructor in a following interview with the student) that *salvo* enrolled the course having already a good knowledge of the subject. This situation, discovered by the instructor only when the course was already over, could have been revealed by using CourseVis in the first stage of the course. The instructor could have provided more advanced, stimulating material to good students, such as *salvo*.

Figure 6.15: The student behaviour for the student *sebastiano*.

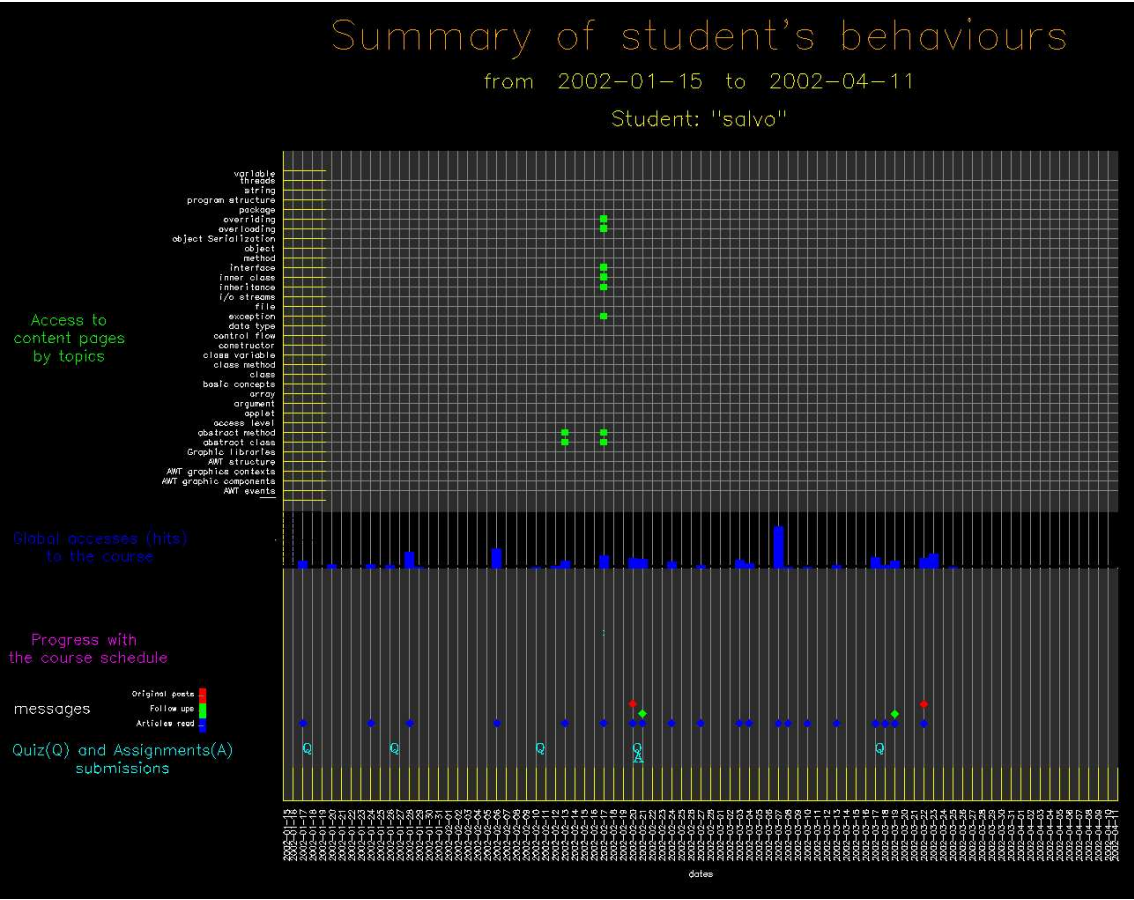


Figure 6.16: The student behaviour for the student *salvo*.

6.7 Promoting instructor's reflection

Examples illustrated a situation in the Java on-line course where information provided by the cognitive matrix can be a useful source of information on students progress, but also may provide the instructor with important information on the course quality and organisation. This is the case of the concept “*array*”, which had a very bad performance on the majority of students in both editions of the Java on-line course. The instructor found an ambiguous question in the quiz testing this concept, and fixed the problem for the future editions of the course. The instructor made use of the graphical representations to reflect on the course, and improving the quality of course. Recent research revealed how opening the learner model to students and instructor may support reflection and improve the knowledge of the content and the structure of the domain (Morales et al., 1999; Bull and Nghiem, 2002; Brna et al., 2003; Zapata-Rivera and Greer, 2001a). The instructor may also find interesting information in the discussion plot. For example, it is shown that threads started by the instructor received little follow-ups (see Figure 6.3). This indicates that probably the instructor didn't promote interesting discussions, and may encourage him to be more proactive in future editions of the course.

6.8 Discussion

Graphical representations illustrated so far rely on the CourseVis model of mapping between questions in quizzes and concepts, and a mapping between content pages and concepts. While the main advantage of this approach is that it is very simple and practical, it requires the course materials and lessons to be created in a format suitable to be processed by CourseVis procedures. In particular, the instructor has to identify a set of concepts which constitute the subject of the course (some subjects cannot be easily broken into constituent concepts), and the knowledge of these concepts has to be tested by means of quizzes. The instructors may face the problem of deciding at what level of granularity should they define concepts. The structure of the course material and quizzes then depends on the decisions taken about course concepts. Moreover, CourseVis is appropriate for courses having materials taught in linear order.

If there are few questions testing a concept C then the estimation of performance of students on concept C may not be reliable. This was the case with the concept *array* in the Java course discussed in previous chapters. Array was tested with only one question that was formulated ambiguously. How many questions are necessary to have an accurate measure of the student's understanding in that concept is an open issue, and the decision must be taken by the instructor upon his knowledge of the domain and the experience gained in teaching. Moreover, at a particular time of the course, a student may have answered only a subset of the questions that test for a concept. For this reason, in the future work of CourseVis is advisable to provide a threshold

mechanism to avoid partial results in quizzes that might produce misleading performance values.

Inevitably, some concepts and questions are more difficult than others. This level of difficulty is not considered by CourseVis, but the concepts could be evaluated by a set of questions having different levels of complexity. A more advanced model could associate for each question the concepts it tests together with the level of complexity and consider it on the computation of the performance. However, this introduces a further complexity element on the definition of the domain model (see Section 5.2.4).

Courses having too many students or concepts may present some scalability problems. However, IV techniques such as focusing, zooming brushing (see Section 3.2.8 for an extended discussion) can help examine large amounts of data produced.

As discussed earlier, data from CMS can be represented using several techniques and representations. There is no standard procedure to determine the most appropriate graphical representations for a particular set of data. For a data set to be represented, two approaches for finding suitable graphical representations can be followed: one is trying to represent data with as many different representations as possible, and then to explore with the end users which one is the most appropriate for their needs; the other is trying to analyse the problem and the user requirements with appropriate techniques, and then designing a representation that is expected to meet the needs of the users. The first approach requires the generation of a number of different representations and may be time consuming. However, it will be effective, as the users will be able to choose representations that best match their needs, which can be unclear to the designers who produce graphical representations. The second approach instead is more efficient, but it relies heavily on the designers' interpretation of the users' needs, and may not always produce the most appropriate representation for every individual user. The author of this work has a considerable experience in teaching Web-based distance courses, as well as contacts with teachers using CMS. For this reason it was feasible to follow the second approach: an empirical investigation was undertaken to gather user requirements, and an evaluative study involving potential users of CourseVis was conducted to examine the appropriateness of the proposed representations. The evaluative study will be described in the next chapter.

6.9 Summary

This chapter described the graphical representations produced by CourseVis with the data of an existing Java on-line course. We have described the graphical representations for social, cognitive and behavioural aspects created in CourseVis following the requirements expressed by the instructors in the survey. These representations have been commented in detail and some facts about students and case studies have been demonstrated that could provide useful information for instructors. We showed also how representations can be used to compare a course in different

stages and editions. The next chapter will present the evaluation of CourseVis with potential users to assess the effectiveness, efficiency, and usefulness of the representations produced.

Chapter 7

Evaluation

In previous chapters, we presented the general architecture of CourseVis and a set of graphical representations to help instructors to gain insight of what is happening to the students participating in an on-line course. To evaluate whether those representations comply with the instructors' requirements, and whether instructors may take advantage of the proposed representations, we conducted an empirical evaluation with a sample of potential users of CourseVis.

The CourseVis evaluation took place in three stages: a focus group, an experimental study, and a semi-structured interview. The focus group and the semi-structured interview aimed to assess the usefulness of the graphical representations generated in CourseVis, while the experimental study aimed at verifying the effectiveness and the efficiency of the representations. The process of evaluation revealed also some problems and possible improvements of graphical representations that may be taken into consideration in future developments of CourseVis.

7.1 Introduction

CourseVis is an instructor-centred tool, and it is important to evaluate how this system impacts its intended users. The process of evaluation assesses if the graphical representations produced with CourseVis meet the requirements gathered from the survey with instructors (see Chapter 4) and to see if the users have some benefits in using the tool in their teaching activities.

Evaluation of user-oriented systems has a long history in the field of the Human Computer Interaction (HCI) (Dix et al., 1998; Johnson, 1992) and is quite different from evaluation in other areas of Computer Science, mainly because it is sometimes hard to construct experiments or observations that give definitive quantitative answers regarding the merit of one system over another. For this reason, evaluation of user-centred tools often makes use of analytic and informal techniques with the objective to investigate the usability and functionality of an interactive system (Dix et al., 1998).

7.1.1 Evaluation goals

According to Dix et al. (1998), a typical evaluation task in HCI is intended to assess the following:

- assess the extent of the system's functionality;
- assess the effect of the system on the user;
- identify any specific problems with the system.

The first one is intended to explore if the system provides the required functionality, and if it accords with the user's requirements. The second one is typical of the implementation of user interfaces, where evaluations are performed to detect aspects of the system related to human factors, such as usability, learnability, acceptability, etc. The last one is related to the identification of eventual problems with the design. For instance, the identification of an unexpected result, or a command that generates a wrong interpretation among users which can prompt changes to the system's design.

One of the objectives of the CourseVis evaluation presented here is to examine if CourseVis succeeded in supporting instructors with the required functionality (i.e. the first point of the evaluation task in HCI). Moreover, the evaluation will examine the impact of the graphical representations on the users and discover eventual problems with the system (i.e. the last two points of the evaluation in HCI systems).

7.1.2 Evaluation techniques

There are a number of established evaluation techniques that correspond to the two types of evaluation: *analytic* evaluations and *empirical* evaluations.

Analytic evaluation methods come from psychological models of human information processing and are based on studies of human cognition and behaviour, e.g. one of the most known models is GOMS (*Goals, Operations, Methods and Selection of rules*) proposed by Card et al. (1983). This class of evaluation techniques are more appropriate to the evaluation of systems at the design-specification stages. Analytic evaluations occur during the system's design and are oriented to identify problems and guide modifications during the development of a system.

For the evaluation of a system already implemented as a prototype empirical evaluations are more appropriate, and are suitable to make formal claims about this system. Empirical evaluation is also called *summative* evaluation, in contrast to *formative* evaluations that examines a system under development. Summative evaluation is undertaken in a variety of systems. For instance, Intelligent Tutoring Systems are evaluated to assess the educational impact on students (Scriven, 1967; Mark and Greer, 1993), while in Adaptive Hypermedia the effectiveness of user models (Chin, 2001), student models (Weibelzahl and Weber, 2003) and adaptive features (Eklund and

Sinclair, 2000) is examined. Empirical (or summative) methods of evaluation are quantitative studies that consist of an analysis of determinate hypotheses tested through direct measurements (Dix et al., 1998). Examples of hypotheses can be the user's performance in relation to a specific task, or the number of trials required in order to accomplish a specific task. This requires the definition of one or more variables related to the hypotheses examined and a metric associate to each of them. The evaluation is carried out usually by the means of *experimental studies* (Johnson, 1992). It consists of asking the user testers to run a task and performing some measurements using observation, questionnaires or interviews.

In our case, we ran an empirical study with a sample of potential users of CourseVis and measured the effectiveness and the efficiency of the graphical representation using data from a past course. Experimental studies have been performed in intelligent educational systems to determine the effects of particular factors or aspects of the systems to the learners (Mark and Greer, 1993; Dimitrova, 2002; Shute and Regian, 1993; Vasandani and Govindaraj, 1991). Though our investigation addresses the instructors instead of the learners, we rely on the validity of the experimental approach.

Qualitative methods can also be quite useful for empirical evaluation of the graphical representations. Qualitative research involves the use of qualitative data, such as interviews, documents, and participant observation data, to understand and explain social phenomena¹. They are opposite to quantitative methods used in experimental studies for their ability to analyse phenomenon from the point of view of the participants that it is largely lost when textual or analytical data are quantified (Kaplan and Maxwell, 1994). Focus group (Krueger, 2000; Greenbaum, 1998; Morgan, 1988) is a form of qualitative research which involve group interviewing of individuals selected and assembled by researchers to discuss and comment on, from personal experience, the topic that is the subject of the research (Gibbs, 1997). This is less formal than controlled experimental study, but has the advantage of getting the user's viewpoint directly and "*may reveal issues which have not been considered by the designer*" (Dix et al., 1998, p. 431). A focus group with some instructors has been used for the evaluation of CourseVis. Moreover, additional semi-structured interviews were conducted with the people involved with the experimental study. This combination of both qualitative (focus group and individual interviews) and quantitative methodologies (experimental study) is known as *triangulation* (Krueger, 2000) and it is relevant to our research because it allows to examine data gathered in a variety of ways, establish commonalities or differences and to provide rigour to the study (Denzin and Lincoln, 1998).

The remaining of this chapter has been organised in the following way: Next section describes a list of criteria being assessed in the CourseVis evaluation. Section 7.3 describes the focus group conducted with a sample of instructors. Section 7.4 discusses the experimental study carried out.

¹See *The Qualitative Research in Information System website and journal* <http://www.auckland.ac.nz/msis/isworld/> for an interesting source of information on the conduct, evaluation and publication of qualitative research.

Section 7.5 describe the semi-structured interviews with participants at the experimental study, along with a summary of the most interesting comments. In Section 7.6 the results from each evaluation stage are combined and some conclusions are drawn.

7.2 Criteria to assess

CourseVis aims at improving the instructors' understanding of what is happening in their courses by providing appropriate graphical representations. As already mentioned in Chapter 3, visualisation is a cognitive activity that forms an internal representation of a certain world in the mind of the user (we say that the user forms a mental, or internal, model based on some graphical representations). The world we want to represent is, in our case, information about the students attending an on-line course. The main objective of this evaluation is therefore to understand if the instructors' mental models about their classes are better when using CourseVis rather than when CMS standard functionalities are used.

The following questions were addressed:

- *Suitability of CourseVis functionality* - is the required functionality available within the system? A survey in Chapter 4 identified some specific requirements that instructors needed for their on-line teaching activities. A prototype has been implemented that represents in a pictorial format information identified as requirements in the survey. The first step is to assess if these requirements are fulfilled by the prototype.
- *Effectiveness* - do the graphical representation produced by CourseVis provide instructors with a better understanding of what is happening to the students of their courses? More specifically, can the instructor understand some facts about the students with more *accuracy* and *precision* with the graphical representations provided, rather than without having them? Do the graphical representations provide *extra information* not already provided by the CMS?
- *Efficiency* - can representations help the instructors infer some information about the students more *quickly* than other tools already provided by the learning platform?
- *Usefulness* - To what extend the information provided by CourseVis is useful to the instructors? Can this information help instructors identify specific aspects in the course?

The evaluation was carried out in three stages: a focus group and an experimental study in conjunction with a semi-structured interview. The focus group is intended to explore the usefulness of the proposed representations. The experimental study aimed to explore the system's functionality, effectiveness and efficiency, while the semi-structured interviews investigated the usefulness and

gave us the opportunity to find out eventual problems with the proposed representations. All these investigations combined together will give a unified analysis.

7.3 Focus group

Focus group is a technique that can help to investigate user attitudes, feelings and beliefs on proposed representations with a group interview (Nielsen, 1997; Gibbs, 1997; Greenbaum, 1998). It is conducted by bringing together some representative users to discuss issues and concerns about the features of the system being evaluated. The discussion is lead by a moderator whose role is to facilitate the interaction between group participants and to keep them focused on the topic of investigation. This investigation technique is useful because *“attitudes, feelings and beliefs ... are more likely to be revealed via the social gathering and the interaction which being in a focus group entails”* (Gibbs, 1997) rather than other investigation methods, such as observation, interviews or questionnaire surveys. It has been employed by some researchers to evaluate educational issues related to the use of a CMS, e.g. Greenwood (2001).

7.3.1 Procedure

The focus group took place at the University of Leeds on February 28th 2003. It involved five people who had experience with on-line learning or computer-based educational systems. The selected participants had a complementary background, to avoid that diverse opinions and experiences may not be revealed with a group too homogeneous with regard to specific characteristics (Gibbs, 1997). The experimenter acted as a moderator for the discussion.

The participants were given an introductory explanation about the aim of CourseVis, about the purpose of the study and their role. Then, each graphical representation was displayed to the group on a wide plasma screen and the participants were asked to discuss that representation. The moderator made the interactions with the CourseVis system, i.e. the participants did not use the interface, they just looked at the graphical representations showed on the wide screen and requested at times appropriate manipulations (zooming, rotating).

To promote a debate, the moderator asked some open questions, e.g.

- What kind of information can you gather looking at these images, what can you understand about the students and the class as a whole?
- Can you identify general patterns or tendencies?
- Is the representation useful, and why? Is it interesting to you? Would you be interested to have this information displayed?
- Do you think it must be presented differently and how?

- Is there anything else which can be useful that is not captured in this representation?
- In which tutoring activities do you think that information represented here may be useful?
- What kind of activities may you use this representation for?

The session lasted about one hour and was recorded on a videotape. The videotape helped the data analysis, as it provided a complete record of the session, capturing all verbal and nonverbal behaviour. The recording was analysed by two people (the experimenter and the advisor of this research), and the results will be explained in the next paragraph. Participants were informed that the discussion was recorded and were assured that it would be used confidentially for the sole purpose of evaluating and improving CourseVis.

The focus group revealed some problems concerning details of the representations that were judged as misleading from all members of the group. These problems were fixed in the subsequent version of the representations, which is outlined in Chapter 6.

We have to stress some limitations of the focus group. In literature there is no universal indication about the recommended number of people per group. Even though some authors give precise suggestions about the group size (Krueger (2000) recommended 7 to 10, Morgan (1988) suggested as a range of acceptable size from 4 to 12) we have to acknowledge that the few number of participants to the group led to a partial analysis. The number of groups is questionable as well. We used only one group, while some experts suggest that the number of groups should be at least two (Morgan, 1988) or three (Krueger, 2000). The difficulties in finding instructors available for a group study led us restrict our analysis to only one group. Moreover, involved individuals belong to specific context, within a specific culture (although there were representatives from two countries, UK and USA). In particular, all of them teach courses at a university level. Their point of view may differ from others teaching in different level, such as in secondary school.

7.3.2 Results

Each graphical representation illustrated in Section 6.2 was proposed to the group and was subject to discussion. A summary of predominant statements and relevant participant thoughts will be provided. Results are divided for representations to help the reader to contextualise the discussion on a particular image.

Discussion plot

This representation was discussed in Chapter 6 and was illustrated in figures 6.2 and 6.3. A participant's first impression when looked at this image was:

“is like having an awful lot of information, initially overwhelming what you really need to”,

due to the many dimensions represented. This proved that, at a first stage, the 3D scatterplot created some confusion. One participant commented:

“It looks to me like a swimming pool”.

Predominant criticisms was the difficulty to locate the exact position of the balls with respect to the axis, and the rotation of the scatterplot was considered too complicated to decide to what different angle to position it. In fact, the positioning of the scatterplot which hides one of the dimensions (and which transforms the 3D image in a 2D scatterplot) was considered the only usable. But, at the same time, the same participant who criticised the rotation of the scatterplot appreciated the flexibility given by the graphical representation to exploring the dots *“to find something interesting [...] because you never know what you need to see”*.

This graphical representation was appreciated because it may help instructors identify discussions that need attention, to see further what the discussion was:

“If I was the instructor and I had the discussion ball in front of me, I would want to look at some of these and I would focus on the red ball for instance. The image may help me focusing on which piece of discussion I want to look at and analyse”.

but also:

“I’d look at the little one, the once having no follow-ups, to see what’s wrong with those time periods what’s going on in that discussion, why it is not exciting, so I can change that, or take part as appropriate”.

The connection with the topics of discussion could be useful to infer particular information:

“I noticed that the technical problems discussion topic are at the beginning, they would be very helpful to me. When you are looking at the students to see if they find problems I can see that they know what to do now, as the technical problems seem to have been solved”.

Discussion graph

We showed also the 2 dimension plot related with this image, as illustrated in Figure 6.4. Immediately participants appreciated this representation for its clarity and simplicity. A participant recognised almost immediately who the main contributors in discussions were and who weren’t doing very much: *“Plot is very useful because it immediately gives this information”*. With regard to this graphical representation, participants revealed that several 2D images seem more useful than the 3D scatterplot, e.g.:

"I think most instructors would prefer a number of 2D representations rather than having to look at a complex representation like this. It takes you a lot of time to find out what it is of about. I think I would want to use it (the 3D image) as the last resource".

A suggestion to improve the graphical representation was to change the 2D plot using histogram, and give a user the possibility to go straight to the discussion thread by clicking on a particular ball of the 3D scatterplot:

"It's quite an interesting navigation tool for the discussion room, if I want to go to a particular discussion it would be potentially easy for me to click on and open the discussion thread, rather than going to the discussions with all the hierarchy and trying to search the relevant thread".

Participants noticed that it was not possible to see the names of the students who participated in a thread. This information, not being captured by the graphical representation, was considered desirable. As future work the addition about who participated in a thread (not just who started a thread) in this graphical representation is worth investigating.

Cognitive matrix

This representation was illustrated in Figure 6.5. A participant reacted immediately:

"I certainly would be intrigued by michele, he is only doing quizzes on basic concepts, but he is doing quite well, he is probably unsatisfied about the basic concepts? or he probably wanted to know only what the basic concepts were, and then he was satisfied about what he knew and then he left ?".

This shows that instructors could be interested in monitoring anomalous situations, like the student *michele*. The cognitive matrix might promote instructors' reflection not only on specific students, but also on parts of the course. A participant expressed doubts about the quality of quizzes and teaching when looked at this graphical representation:

"It may be possible to tell something about the quality of your quizzes or of you teaching. Simply because the "AWT Events" for example are not done extremely well by anybody, the same was for "threads", so I have some doubt about the quality of your teaching material. This doesn't tell you way they do it badly, this tells you about the quizzes, but it may be the teaching of the subject. You identified the "array" for instance with few reds, but this could be because your quizzes are bad, or your teaching materials are bad".

As already explained in Section 6.7, students answered wrongly to quizzes related to *array* because the questions were vague and misleading. This representation revealed a problem on quizzes formulation discovered by inspecting the cognitive matrix.

Another important application of the cognitive matrix was captured in a comment about students whose results were good:

“If people are doing extremely well and your quizzes picture that they are high level, I’d probably look at the application form of those students and see what kind of background they have. It will certainly help you to adapt your teaching in general.”.

This comment revealed how information reported in this graphical representation could be used to help teachers provide adaptive teaching, according to individual and class progress, as suggested in Mazza (2003).

Some possible improvements of this graphical representation suggested were: (a) to put the list of concepts in a chronological order to show how the level of knowledge of each student is evolving during the time of the course; (b) to include (or link) in the representation other sorts of proofs (e.g. assignments), so the instructor can better examine the student’s performance with other data; (c) and finally to change the scale colours².

Cognitive graph

We showed to participants the 2 dimension graph illustrated in Figure 6.6. It was stressed that this representation may help instructors to more easily quantify the difference between concepts. However, the cognitive matrix was preferred over the 2D graph, because it allowed “*a quicker glance*” to compare concepts and individual students’ performance at the same time. Participants suggested that a histogram is more appropriate than a continuous line plot for this case.

Students accesses plot

This representation was illustrated in Figure 6.7. All participants expressed interest in having this graphical representation in their courses. They would use this graphical representation to identify students’ accesses and find early problems, as one of the participant commented:

“If I was running this course I would want to use it in order to see if a particular student hadn’t accessed in to very much, to find this early on”.

²In the version of the program used in the focus group the range of the scale colours for the performance rate was from blue (lowest) to red (highest). People in western countries usually consider red as colour of alert, and green as a colour of relaxation. In a following version of the program we changed the colours accordingly with a range from red (lowest) to green (highest).

It was their opinion that this representation could give instructors the opportunity to have a clear picture of students who are actively participating in the course and those who are not. Instructors could encourage students not very persistent to be more active, with the intent to do appropriate activities before it is too late:

“you want to provoke them by e-mail at some point and say: - Look, you’ve got a week free and there is no access -“,

and

“I think that the image helps me to realise that something be wrong. I should be doing something before it is too late”.

One participant argued that some instructors may be more interested in having not only the information about the access on a date, but also how long students accessed the course, how often, and even the period of the day (e.g. morning, afternoon, night). It was pointed out that one dot may represent a number of accesses and this may be misleading. However, other participants thought that, even without this information, this graphical representation is useful to assess some aspects of students’ participation.

Participants identified some patterns looking at the dots, which referred to the group exercises given at $\frac{3}{4}$ of the course and the termination of the course, which occurred 15 days before last date reported in the graph.

Participants made also some suggestions to improve the graphical representation: (a) to put a scale on the global access histogram to better clarify what it represents; (b) to indicate week ends and holidays with different colours on dates and mark topics of study on dates to see if some topics required long time to work on them. Future development of CourseVis could investigate how the representations could incorporate these recommendations.

Student behaviour graph

This representation was illustrated in Figure 6.8. Participants found it very useful to have the possibility to monitor individual students, and most of them appreciated it:

“I think that the green stuff [the access to content pages by topics] and the blue stuff [the global access to the course] are very useful.”

but also, so much information may not be needed:

“It is interesting to have an overview with this kind of picture on the student, but it is also has so much information that I would not need”.

and:

“I think it is best to just provide as little information as clearly as possible. If you provide 2 or 3 variables at time, you are probably more successful than trying to compete and provide as much information as possible”.

Some participants found very difficult to look at the access to content pages to discover if a student accessed the material on every concept:

“If there is one thing I will be looking for here is how the student access material on every concepts, as it is very hard for me to tell looking at this. I see the access to a lot of things, I’ve got to study the lines carefully to working on it.”

A possible solution to this problem is to put a summary after the pink area and tell what pages are accessed, which is recommended in future development of CourseVis.

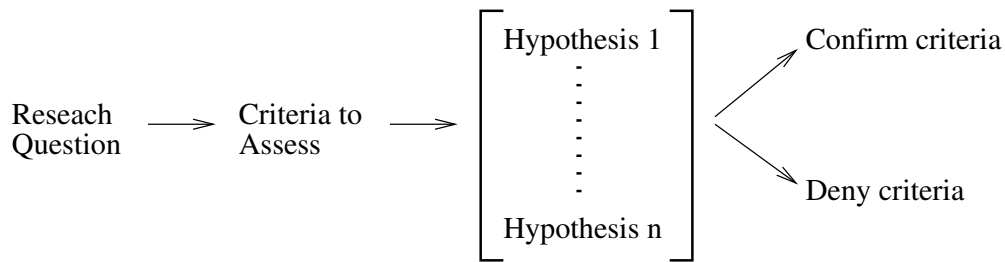
Another suggestion made by participants was to modify the quizzes representation to show also the score achieved by each student, so that a better picture of the students’ progress is extracted. As a future implementation the participants suggested to substitute “Q” with the score, or to put the score next to “Q”.

7.4 Experimental study

The focus group revealed some opinions of instructors about graphical representations, but a further study is required to analyse whether the representations meet the instructors’ requirements expressed in the survey and to see if the representations may provide new information or could give instructors a better understanding of what is happening in their classes. To investigate those issues we ran an experimental study.

Experimental research is quite common in psychology and education, and is suited to an instructor support system like CourseVis because it enables researchers to examine relationship between the proposed representations and the instructors’ understanding of what is happening in their classes, and to obtain quantitative measures of the significance of this relationship (Mark and Greer, 1993). We follow the experimental methodology suggested by Johnson (1992), which is quite similar to other experimental approaches suggested by other experts in the evaluation of HCI systems (Dix et al., 1998; Faulkner, 1998) and educational systems (Mark and Greer, 1993).

According to Johnson (1992), the first step in an evaluation study is the identification of the **research questions** to be examined. The research questions can be formulated as a series of **criteria to assess**, that can be validated or negated by a series of **hypotheses**, defined to be empirically or numerically testable; they concern with specific criteria we want to assess and their application on an experiment can confirm or deny the tested criteria.



When a set of hypotheses has been defined, the experimental test can be carried out with a representative sample of users of the system. The test is ran in *controlled conditions*, which means that the experimenter systematically manipulates one or more conditions of the experiment (called *independent variables*) in the domain under investigation, and all the other conditions that could affect the outcome of the experiment (called *dependent variables*) are monitored. For each hypotheses one or more independent variables (the factors that we vary in the experiment) and one or more dependent variables (the facts that we want to measure when varying the conditions of the experiment) are defined. The validation procedure is to vary of one or more independent variables and measure the effects of those variables on relevant dependent variables which will either confirm or reject the experimental hypotheses. Mark and Greer (1993) define this type of studies in ITS as a *control group studies*.

The analysis of the data and observations resulting from the controlled experiment can then be performed, and we can check if the results confirm or reject the research hypothesis. If the hypothesis is not confirmed, possible explanations might be drawn to identify specific problems or limitations and, eventually, to suggest improvements.

This sort of experimental studies are very convenient for the empirical evaluation of a specific criteria to assess in our system, but they are inappropriate to reveal problems which may occurs during the observation, or to elicit information about user preferences, impressions and attitudes (Dix et al., 1998). For this reasons, at the end of the experiment we interviewed users about their experience on the basis of a set of questions prepared in advance. The details about these interviews are described in a following section.

7.4.1 Hypotheses

First of all we have to define some hypotheses we want to validate, and then derive a set of dependent and independent variables. For example, we want to assess the hypothesis that the instructor gains more insights in how many students access the course using CourseVis rather than with the CMS standard interfaces. The independent variable in this example will be the interface used (CourseVis' representations or CMS' interfaces) and the dependent variable could be instructors' knowledge about how many students accessed the course during a specific period of time. This hypothesis can be confirmed (i.e. the prediction is true) if the experiment reveals that

when using CourseVis instructors can define in details how many students accessed the course during that period, or define exactly the number of times the user accessed the course, while using CMS standard interface the instructor has only a vague or none knowledge about this value.

Hypotheses are constructed taking into consideration the evaluation criteria specified in Section 7.2, the instructors' requirements specified in Section 4.6, and the representations described in Section 6.2. Dependent variables chosen are directly dependent to the hypotheses we want to validate, and correspond to functionalities, effectiveness and efficiency of representations (see Section 7.2). Usefulness, comments and eventual problems are being investigated by semi-structured interviews and will be described later in Section 7.5. In this study we consider the same independent variable, the use of a CMS standard interfaces versus adding CourseVis representations to the CMS. Hypotheses are grouped in social, cognitive and behavioural as described below.

Social aspects

We start with the social aspects of the learners, and the graphical representation illustrated in Section 6.2.1. The survey indicated that quantitative representations about discussions and participation in group work should be provided (see Section 4.7.4). CourseVis produces a graphical representation which intends to give a quantitative representation about the students' discussions throughout the course and discussion activities in group works (the discussion plot on 6.2). We argue the following:

hypothesis: Instructors have a better understanding about who the most predominant students in opening new discussion threads are, and the number of follow-up which these threads generated, when they are supported by CourseVis than when just a CMS is used.

dependent variables: Knowledge about the names of three students who initiated most new threads in the course, and knowledge of which students received more feedback by others. The length of time to complete these tasks is also considered.

Cognitive aspects

Instructors' requirements about the cognitive aspects of learners suggested the creation of the graphical representation illustrated in Section 6.2.2. This illustration provides a 2 dimensional graphical representation of the students' performance on the course quizzes. Instructors requested to have information about the single concept performance and overall performance of students in the course, and the possibility to compare a student with the rest of the class (see Section 4.7.4). Graphical representation in 6.5 intend to address these requests. We argue the following:

hypothesis: Using CourseVis instructors can estimate the performance of a particular student on a particular topic, can compare students with respect to single topic performance or overall

performance, and finally can identify students (or topics) that needs further consideration and investigation.

dependent variables: Evaluation about the level of performance achieved by a specific student on a particular topic, comparison on the overall performance on the course between two students. List of the 3 topics where students performed very well and list of the 3 topics where students performed very badly in quizzes. Length of time to complete these tasks.

Behavioural aspects

Instructors' requirements about the behavioural aspects of learners suggested the creation of a couple of graphical representations presented in sections 6.2.3 and 6.2.3. Instructors requested to have information about the student's access, the reading of materials and the participation in discussions. Moreover instructors are interested whether students perform evaluations and whether they are progressing on the course schedule. Graphical representations discussed in 6.7 and 6.8 intend to address these requests. We argue the following:

hypothesis: Using CourseVis instructors have an almost immediate understanding about which days a student accessed the course. Instructors have a better vision about the pages read by a student and the participation in discussions. Instructor can identify patterns and trends of student's access and participation in discussions.

dependent variables: The number of days a student accessed a course. Identify the most (and least) accessed week during the course. Identify patterns and trends in the course's access, content pages and participation in discussions. Length of time to complete these tasks.

7.4.2 Procedure

An experimental study requires the existence of an actual implementation of the system we want to test. The ideal situation in this sort of study is to carry out some experimental studies involving real courses with real instructors. We described in Chapter 5 the implementation of CourseVis as a prototype running upon WebCT. The data gathered from the Java on-line course has been used in Chapter 6 to illustrate the graphical representations produced. We can use the same prototype and the same course to run our experiment.

The java on-line course was backed-up from its original WebCT environment and restored in another installation of WebCT integrated with CourseVis. In this way we had two platforms to be used for the investigations: a WebCT standard platform and a WebCT platform integrated with CourseVis as an additional tool. The two platforms provide the same identical course.

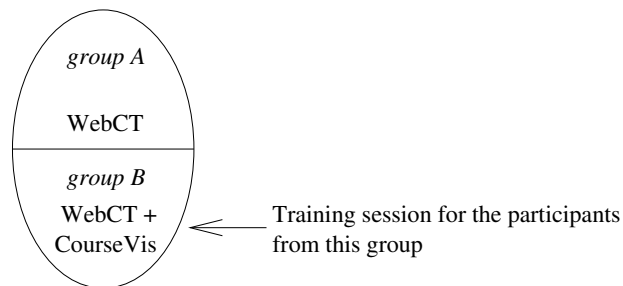


Figure 7.1: Participants at the experimental study were split in two groups. Group A was assigned to a course provided by a standard WebCT installation, group B was assigned to a course provided by a WebCT installation integrated with the CourseVis graphical representations.

Participants

There is no universal indication about the selection of people to participate in an experiment. However, there are basic principles in experimental designs that give some indications about the participants' selection.

First of all, the number of participants. Clearly, the sample size is often determined by pragmatic considerations. Especially in this case, where the availability of subjects is very limited. General indications suggest having not less than 6 participants in a controlled experiment³ (Johnson, 1992). The sample should be selected as representative of the intended target population (i.e. distance learning instructors, in our case).

Another big issue is the experimental method to use. Roughly, experimental designs can be carried out in *between-groups* or in *within-groups* (Johnson, 1992). In between-groups participants are split in different groups, and each group is assigned to different condition of the experiment. In within-groups the same group of the experiment is used in all conditions. The latter one is known to produce some transfer of experience effect, so we decided to run a between-group experiment.

For the experimental test we asked six instructors with past experience in teaching Web-based courses to participate in the evaluation of CourseVis. Participants selected are distance teachers from the University of Lugano and the University of Applied Sciences of Southern Switzerland having experience as instructors of distance courses given with the WebCT CMS and having some knowledge of the domain of the course to be taught in our experiment: the Java programming language.

Running of the experiment

To conduct the between-groups experiment, we divided participants into two groups (see Figure 7.1). The first group (group A) was assigned to a standard WebCT installation and the other (group B) assigned to the WebCT integrated with the graphical representations provided by CourseVis for that course.

The participants were asked to pretend that they were the instructors of the java on-line course and to try to get as much information as possible about the students of that course. In order to measure what understanding about the course the participants gained, we asked them a set of questions, related to the dependent variables described above. Questions were identical for both groups. Therefore, the only difference between the two groups (i.e. the independent variable) was that the second group used CourseVis integrated with the CMS.

Each participant attended an individual session. Prior to the session, the participants were given a description of the purposes and objectives of the study (about 5 minutes), and the course “Introduzione a Java” was presented and explained in sufficient details. Participants of group B were trained also on the graphical representations produced by CourseVis before running the experimental session. To avoid a bias on the data of the representations, the initial training was given with the data of the course given in year 2001. For the experimental session we used instead the data of the same course given in year 2002. Each year has distinct students enrolled, hence the information on students discovered in training session didn’t influenced the experimental session. Training session for participants of group B lasted about 15 minutes.

During July 2003 we ran some experimental sessions where participants were asked to answer a set of questions prepared to validate the hypotheses previously described, accordingly to their dependent variables. Sessions consisted in giving participants the access to a computer having a CMS platform installed and asking questions. Participants of group A used CMS functionalities to answer the questions, while participants of group B explored the CourseVis representations to answer the questions.

For each question it was measured:

1. the time necessary to answer,
2. the tools used,
3. the accuracy of the answer.

Each session was followed by a semi-structured interview that will be discussed in a following section. Generally, each full session (interview included) lasted about one hour and a half.

³However, Johnson (1992) suggest a number of 10 or more to increase the ability to detect differences (if there is availability of subjects).

7.4.3 Results

Participants were asked a set of questions on the basis of hypotheses to be validated. Questions are classified on the particular aspect (social, cognitive, behavioural) to be investigated.

Social aspects

We asked the participants who were the three most active students in opening new threads in discussions in the course. Participants of group A used the discussion board to answer to the question, and took about 2 minutes to answer to it. Participants of group B used the discussion plot and the discussion graph and answered in less than 1 minute (one of them answered in 15 seconds). All participants of group B answered correctly, as the numeric values reported on the 2D discussion graph are straightforward. Participants of group A instead gave only approximate values, as they had to browse all discussion messages and gave an estimation on the students' names which appear in the new threads messages. One of participants from group A was unable to give more than 2 students' names, as he was "*confused by so many messages and names*".

Then we asked who, among the three students cited by participants in previous question, received more feedback by others. Participants of group B answered in about 3-5 seconds, while participants of group A had to browse among threads and didn't answer before 30 seconds. All participants answered to the question correctly.

Validation of the hypothesis on social aspects: With previous results we can affirm that, using representations provided by CourseVis, instructors have a better understanding about who were the predominant students in opening new discussion threads, and who received more feedback. The most important benefit provided by the representations is the rapidity of having this information, provided very quickly by the discussion graph and the discussion plot.

Cognitive aspects

We asked the participants what is the level of knowledge achieved by *massimo* on the concept *applet*. Participants of group B answered correctly (with a specific number in the scale 1 to 10) in about 8 seconds using the cognitive matrix. Participants of group A couldn't answer the question, as the CMS did not provide instructor with this sort of information. One of the participants from group A tried to browse quizzes questions in order to identify answers provided by the student on the concept regarding applet, but after about 10 minutes he gave up. We have to acknowledge that the real instructor of the course who knows in detail each single question in quizzes may be able to find this information, but nevertheless, this will require considerable effort.

We asked instructors to identify who had better results in quizzes between *domenico* and *antonio*. Two participants of group B answered correctly in 10 seconds, the other in 1 minute. Among

participants of group A one answered correctly in 30 seconds, the others took above 1 minute and 30 seconds to read the numeric values, do the sum of marks for each quizzes, and provide the answer.

Next question was the indication of the three topics where students performed very well and respectively very badly in quizzes. The cognitive graph helped the participants from group B who answered correctly in few seconds. The participants from group A didn't even try to answer, because they needed to map (manually) quiz questions with concepts and to do some statistics. In any case, this operation would require a very long time.

Validation of the hypothesis on cognitive aspects: The results of the study demonstrated that when using CourseVis the instructors could estimate with a good precision the performance of a particular student on a particular topic almost immediately. Students' comparison on global results on quizzes can be made both with CMS native interfaces and with CourseVis in about the same time and same precision, but the comparison of students on the basis of quizzes' results on a specific concept can't be easily done with the CMS standard interfaces, while it can be done quite easily with pictorial representations provided by CourseVis. Using representations provided by CourseVis instructors may estimate which topics of the course are problematic on the basis of results on quizzes and need further investigations, while this information is not given by the CMS.

Behavioural aspects

We asked the participants to give an approximate value about how many days the student *antonio* accessed the course. All participants from group A commented that this was hard to do with the tracking data provided by the CMS, because it was presented as a log file in a textual format (see Figure 2.1 at bottom which shows an example of this log report). In fact, one participant from group A affirmed that it "*needs to be processed by a computer algorithm to be usable*". Participants of group B gave a fairly precise answer (with an error below of the 15%) using the students accesses plot in about 15 seconds.

We asked which period of the course was the most accessed and least accessed. Again, participants from group A didn't answer the question, due to the impossibility to elaborate in mind all the tracking data of all students; the participants from group B identified correctly most accessed period corresponding to the beginning and the last accessed period corresponding around the Easter vacations.

We asked then if the student *antonio* accessed the content pages of the course regularly and if he participated to discussions regularly. All participants of group A answered that they did not have any functional tool provided by the CMS that may give information about the accesses to content pages (apart the tracking log which was already considered difficult to understand by instructors). The participation in discussions can be rated with the number of messages posted, and

the constancy and regularity of posting can be estimated only by a person who actively participated at discussions and read the messages. Participants of group B, in about 15 -30 seconds, recognised that *antonio* read materials of the course irregularly. They also realised that he was regular in reading messages, and this took them between 2 seconds and 30 seconds.

Validation of the hypothesis on behavioural aspects: The experiment demonstrated that using CourseVis instructors could detect which days a student accessed the course. This information is provided by the CMS in an inconvenient and complex manner, while the representation provided by CourseVis give a quick and easy access to this information. Instructors can also identify patterns and trends in the content pages read by a student and in the participation in discussions, which aren't provided by the CMS.

7.5 Semi-structured interviews

The experimental study described in the previous section demonstrated that some desired information is provided by the representations, but it didn't address if the instructors consider this information interesting or not. A way to find out whether provided representations are useful and whether the users are satisfied is to ask them. For this reason, along with the experimental study, we conducted interviews with the instructors involved in the experiment. Interviews are a well-established technique often used in social science research, market research and human-computer interaction (Preece et al., 2002).

The type of interview method used was the semi-structured interview: each instructor was asked a predetermined set of questions, leaving the interviewee free to discuss and comment on the representations. The interviews addressed specific questions to elicit information about users' impression and general comments, as well as their opinion about the *usefulness* of the proposed representations.

7.5.1 Procedure

During the experimental study, each instructor was asked for each question to judge the usefulness of the information gathered by the question. For instance, in the experimental study the first question was to indicate who were the three most active students in opening new threads in discussions. Once he had answered the question, instructors were asked whether this information could be useful in their teaching activities, and in which situations. Participants from group A who did not use the representations produced by CourseVis were given graphical representations at the end of the experimental session, in order to be able to discuss about their usefulness. User comments were analysed and most interesting issues are presented below.

7.5.2 Results

In line with previous sections, the results are presented in three aspects: social, cognitive, behavioural.

Social aspects

All participants, except one, found interesting to have the information on the number of threads opened and the follow-up received. Both representations regarding discussions, namely the discussion plot (Figures 6.2 and 6.3) and the discussion graph (Figure 6.4), were found useful. However, the discussion graph was preferred for its clarity and simplicity, while the discussion plot was judged helpful for the relationship of discussions with the dates. Three participants commented that the number of threads opened in discussion could be an indicator of the initiative of the student, and it can be useful to judge the participation of the students in the course. They also said that the feedback received could be an indicator of the significance of the question but, for a more accurate estimation, it must be necessary to analyse the content of discussions.

Cognitive aspects

All participants considered information provided by the graphical representations on cognitive aspects useful. The cognitive matrix (Figure 6.5) was judged interesting by two participants to have an indication about the state of the progress of the class. Four participants commented that information provided by the cognitive matrix and cognitive graph (Figure 6.6) could be useful to improve the next edition of the course, because representations give a “*clear vision about which concepts was problematic, or which quizzes was too easy as their colour is green to each student, these could be useful to adapt the content or quizzes in future editions of the course*”. Cognitive matrix was also found interesting by a participant because it gives the information about who have done the quizzes while the course is running, and suggested to put concepts ordered on time of appearance on quizzes, instead of alphabetical order.

Behavioural aspects

We have to distinguish between the “student accesses plot” (Figure 6.7) and “student behaviour graph” (Figure 6.8). The student accesses plot was considered useful by most participants because it may indicate a drop-out student, or the students’ commitment to the course. However, it was stressed that this representation did not express the real engagement of a student to the course, for example a student may print all the course materials and study on his own without accessing the course for several days. One instructor suggested to put visual indicator on dates to distinguish “interesting dates” such as week-ends, days which schedule test, face-to-face lessons, group

exercises (if any of them are planned in the course).

The representation student behaviour graph was appreciated especially for giving indication about the accesses to content pages by topics and about the messages in discussions. Some instructors pointed out the usefulness to discover the behaviour of a specific student, especially *“if combined with results in quizzes it could indicate the learning style of the student”*. An instructor found interesting to locate *“students who needs particular attention”*, such as those who participate with regularity in discussion but didn’t read course materials, it could be an indicator to intervene directly asking why he didn’t read those course materials.

7.6 Results and analysis

7.6.1 General user feedback after evaluation

Generally, all participants to the three phases of the evaluation process (focus group, controlled experimental study, and semi-structured interview) appreciated the proposed graphical representations and judged “interesting” the information they may provide. The most appreciated feature was the “immediacy” of having some information that, otherwise, it may require too much time and effort to be derived with CMS standard tools.

A follow-up of the use of the graphical representations was some reflections of instructors on aspects students that haven’t been considered before. For instance, in the semi-structured interview an instructor, when asked if a particular information was considered useful for their courses, answered: *“I never considered this information in my course before but, if it is easily provided by a tool like CourseVis, I will certainly use it because it could be very helpful ”*. Many other instructors expressed similar affirmation. An instructor concluded the experimental session with this sentence: *“this information help you to think”*, which suggest that graphical representations may provide reflection.

7.6.2 Criteria evaluated

In Section 7.2 we described the criteria that are being considered in the evaluation process. Evaluation techniques so far described (focus group, experimental study and semi-structured interview) are indented to investigate on the criteria proposed. We are ready now to comment on each criterion on the basis of the results achieved on each evaluation technique adopted.

About the suitability of CourseVis functionality

We want to assess whether instructor requirements gathered with the survey, in term of functionalities available with the system, are met by the proposed graphical representations. The experimen-

tal study intended to investigate on this issue. We recall from Section 4.7.4 that the information to be represented is the following:

- Social aspects: quantitative visualisation about discussions, visualisation of participation in group work
- Cognitive aspects: student overall performance in the course, level of knowledge achieved by each student for each domain concept of the course, representation of students having difficulties with a concept, comparison of a student's progress with that of the whole class
- Behavioural aspects: student's access to the course, student's reading of course materials, performing of evaluation proofs, progress with the schedule, participation in discussions

We saw in Section 7.4.3 (social aspects) that instructors who used CourseVis had quantitative information about the number of threads opened by students and the follow-up received. Moreover, if an instructor dedicates a topic to the discussions within a group work, there is an indication about the quantity of threads and the number of messages exchanged within the group (see discussion plot in Section 6.2.1 for topics “group01, group02, ...”). However, the evaluation uncovered some shortcomings, such as missing of the name of the students who participated in a thread (see Section 7.3.2). We, therefore, consider that this criterion was partially fulfilled in the version of CourseVis evaluated.

Section 7.4.3 (cognitive aspects) demonstrated that instructors who used CourseVis had information about the performance of the students on the whole course and in each specific concept, may compare students on their global performance and have indications about which concepts are problematic. Sufficient information required for the cognitive aspects is provided by the proposed representations.

Section 7.4.3 (behavioural aspects) demonstrated that instructors who used CourseVis had an indication about which days a student accessed the course, which content pages had been read (and when), and some information about the participation in discussions. Moreover, the representation student behaviour graph illustrated in Figure 6.8 integrates the performing of evaluation proofs and the progress with the schedule. Users in the focus group pointed out some desired information about the students access which were not illustrated in the students accesses plot (see Section 7.3.2 - students accesses plot), such as the number of accesses in a single day per student, or the period of the day in which a course is accessed. This shows that this criteria, for the behavioural aspects, was partially fulfilled.

About the effectiveness

The second criterion analysed in the survey is the “effectiveness” of the proposed representations. We defined in Section 7.2 that the representations would be considered *effective* if they gave to

instructors a better understanding of what is happening in their courses or, in other words, if representations provided the instructors with extra information not supplied by the CMS or are provided with more accuracy and precision than the CMS. The experimental study has been used to investigate these issues.

The results of the experimental study revealed some information provided by representations that is not available with the CMS standard interface. These are: the performance of a student on a topic of the course, the indication of concepts that need further investigation because students performed very bad on that, the content pages read by a student per topics, and the progress with the schedule of the course.

The experimental study revealed also that some information already provided by the CMS could be determined with more accuracy when provided by CourseVis. For example, instructors who used the discussion graph had a precise indication on the number of thread opened and the follow-up received by each student.

About the efficiency

We defined in Section 7.2 that the CourseVis representations are *efficient* if the information is provided by the representation quicker than the CMS standard interfaces. It is known that some information are provided by the CMS, but obtaining this information would require very long time and may be inconvenient for the instructor's activities.

The experimental study demonstrated that using representations produced with CourseVis instructors may take advantage of the rapidity of having information provided by the discussion graph and the discussion plot, may compare students' quizzes performance almost immediately, and have indication on days the students accessed the course with a quick and easy representation.

About the usefulness

Even if particular information is provided in an effective and efficient way, instructors will have a real advantage only if the information provided is useful to them. The focus group and the semi-structured interview were intended to investigate this issue. These revealed some situations where the information provided by the graphical representation was regarded as useful by participants, but also a number of problems and suggestions were identified.

Discussion plot and discussion graph were appreciated because they may help instructors to identify threads that need particular attention, and the number of threads opened could be used to judge about the participation of student in the course.

The cognitive matrix and the cognitive graph were appreciated because they may help instructors in: monitoring anomalous students, promote reflection on parts of the course, adapt the teaching to the individual, compare students' performance and individuate which concepts was

problematic.

The students accesses plot and the student behaviour graph were considered interesting to instructors as they could indicate the drop-out students, the students' commitment to the course, the content pages accessed and the participation in discussions. The combination of several information on a single representation for the student behaviour graph was regarded useful because this could help instructors to identify students who need particular attention, or infer their learning styles.

7.7 Summary

In this chapter a thorough study on the graphical representations proposed with CourseVis has been conducted. The combination of different quantitative and qualitative methodologies allowed examining data gathered in different perspectives. The combination of these studies led to the investigation on the following criteria: extend of required functionalities, effectiveness, efficiency, and usefulness of the graphical representations.

Generally, participants in the studies appreciated the proposed representations. Results revealed that most of the information needed by instructors for their teaching activities, identified with the survey, was provided. Representations were found effective in proposing new information uncovered by the functionalities provided by commercial CMS, efficient in giving needed information quickly, and useful in helping instructors to gain an understanding of what is happening on their courses.

Chapter 8

Conclusions

This final chapter summarises the work described in this thesis and the results achieved. We will first describe a summary of the research, the main achievements of this work, and will outline its contribution to relevant research fields. We will then describe limitations of our approach and will sketch out possible areas for future investigation.

8.1 Summary of this research

This work is an investigation into student monitoring and knowledge modelling using visualisations to help instructors become aware of what is happening in distance learning classes. Specifically, techniques from Information Visualisation have been used to graphically render complex, multidimensional student tracking data collected by CMS. A system, called CourseVis, illustrates the proposed approach. CourseVis has been developed as a generic tool for graphically representing student tracking data - it is required that the data from CMS are converted into an XML format, which is then imported into the CourseVis graphical render to produce various representations of the data. An instantiation of CourseVis to extend the functionalities of WebCT, a widely used commercial CMS, has been developed. Some representations of student tracking data from this CourseVis instantiation have been presented in Chapter 6.

A crucial part of this work is the exploration of the instructors' needs when they are managing a class at distance. A survey which involved instructors with experience in distance learning was carried out (see Chapter 4). The results have been taken into consideration to inform the design of the proposed graphical representations.

CourseVis follows the Card's reference model for visualisation to generate graphical representations from student tracking data collected by CMS (see Chapter 5). The attributes of each data set were identified and a systematic approach from IV was applied to select appropriate visual stimuli in the graphical representations.

CourseVis has been applied to different stages and different editions of a Java on-line course and to a course on verbal semiotics (see Chapter 6). Some IV techniques, such as zooming and panning, have illustrated also the possibility to apply these representations in courses having a high number of students enrolled.

To evaluate whether the representations generated in CourseVis comply with the instructors' requirements, and whether instructors may take advantage of the proposed representations, we conducted an empirical evaluation with a sample of potential users of CourseVis (see Chapter 7). The results of the evaluation (carried out in three stages: a focus group, an experimental study, and a semi-structured interview) showed that the work achieved its original aim. In particular, the evaluation has shown that the representations help instructors to quickly and more accurately grasp information about social, cognitive, and behavioural aspects of students. The provided information was regarded by the instructors as very useful for managing successful distance courses. It was noted that the graphical representations provided in CourseVis may help instructors identify early, and even prevent, some of the problems with distance learning, e.g. students who do not communicate and may feel isolated, students not visiting the course material who might be confused or be potential drop-outs, and long discussion treads on a topic that may highlight problems experienced by learners.

8.2 Contribution of this thesis

This work gives a number of original contributions that will be illustrated below. The validity of the approach and the significance of the results have been demonstrated with publications at highly ranked international conferences (see the list of published papers on page xx).

8.2.1 Contribution to the Web-based distance education

The work presented in this thesis is a progression in the issue of students' monitoring and learner modelling in Web-based distance education systems. Chapter 2 showed how monitoring student learning is crucial in Web-based distance education, and we described how CMS provide a primitive students tracking system that usually make available to instructors students' data in a tabular format, often incomprehensible, with a poor logical organisation, and difficult to follow. As a result, students tracking data is used only by skilled and technically powerful distance learning instructors. Moreover, CMS do not provide information about the actual learning that is taking place, such as the level of understanding achieved by a student on a particular concept. A novel approach has been described here whereby Information Visualisation techniques are adopted to graphically render a vast amount of students tracking data collected by CMS. Using sophisticated IV tools, and a model of the domain taught in the course, instructors can manipulate the graphical representations generated, which can help instructors gain an understanding of their learners and

become aware of what is happening in distance classes. Graphical representations developed in CourseVis have allowed some useful student information to be found, and the relevant issues on building a visualisation environment for students' data in a CMS were identified. An extensive evaluation study with instructors of courses taught at distance compared their views of students in a course given with a standard CMS with the same learning environment integrated with the graphical representations provided by CourseVis for that course (see Chapter 7). Results of this evaluation revealed that representations were found effective in proposing new information uncovered by the functionalities provided by commercial CMS (e.g. the performance of students on a topic of the course, course attendance, patterns in reading the course materials, and participation in discussions). Moreover, the evaluation demonstrated that using representations produced with CourseVis the instructors may take advantage of the rapidity of having information provided by graphical representation that otherwise may require too much time and effort to be derived using CMS standard tools.

Advanced educational tools such as the Web-based Adaptive and Intelligent Educational Systems have an internal Student Modelling feature that take into account students' progress and level of knowledge attained, but this type of applications often require the instructors to convert course materials in some formats (such as XML) and require the users to install (and learn) another learning platform. CourseVis, instead, is coupled with a commercial Web-based CMS as an additional tool. Hence, the instructors use the learning platform which they already know in the usual manner, use already available Web-based courses created with commercial CMS, and preserve investments that have already been made.

Other works propose some forms of graphical representations of students' data. For instance, the Bull's and Nghiem's General Learning Environment and Chaatra (described in Section 3.3.2) propose a graphical representation of students' understanding in a course. They use a simple model based on the accumulation of correct and wrong answers to derive students' understanding of course concepts. However, these works do not take into consideration other factors considered as important from instructors in distance courses, such as attendance, reading of materials, and participation in discussions. Moreover, the graphical representations proposed have not been evaluated with the users, while in CourseVis a comprehensive approach that takes into consideration multiple factors in a course (not only the results on quizzes) has been followed. When using CourseVis, teachers can identify anomalous situations in a course, such as the low interest of student *salvo* and the drop out of student *sebastiano* in the Java on-line course described earlier in the thesis. The graphical representations provided in the General Learning Environment and Chaatra would not enable such analysis.

8.2.2 Contributions to open student models

Most of the approaches to monitoring students' understanding make use of AI techniques (e.g. *UM/VIUM*, and *ViSMod* described in Section 3.3.2). Open student models is an emerging field originated from the advances of ITS, becoming increasingly important because it offers a tractable way to deal with the dynamics of the student's knowledge and provides the means for encouraging student reflection and enhancing learning (Bull, 1997). Open student models may also help instructors to reflect on their teaching processes and materials (Bull and Nghiem, 2002). This thesis demonstrated that graphical representations in *CourseVis* allowed the instructor to reflect on the quality of the teaching materials prepared. The main contribution to open student models provided by this thesis is a systematic approach to produce appropriate visualisations of student data. The pictorial representations used in open student models have been chosen rather ad-hoc and no systematic analysis has been done to find out which graphical form is suitable in what case. In contrast, *CourseVis* is based on a systematic approach for selecting appropriate graphical representations and may be extended to suggest appropriate environments for externalising student models. Most open learner modelling systems deal with a representation of an individual student that hardly supports comparison with other students. Representations used in *CourseVis* could be employed in multi-user AIED systems to open models of groups to teachers and learners. Moreover, *CourseVis* provides multiple representations of the same set of data to show different perspectives. Distinctively from the approaches that use AI techniques to infer the knowledge of the student, *CourseVis* allows the instructors to gain insight of their students from exploring graphical representations of student tracking data, combined with their knowledge of the domain of the course, materials, and quizzes and the direct interaction with students in discussions or conversations that may allow to infer student models in the instructors' mind (see Figure 8.1), instead of being calculated with some statistical or AI methods (such as *E-KERMIT*, *knowledgeSea* and *WebVIBE*, Bull's and Nghiem's General Learning Environment illustrated in Section 3.3.2).

However, we have to acknowledge that while other works on open student models, e.g. *VisMod*, enable changes of the graphical representations to be propagated to change the student model, *CourseVis* is just a viewing tool and does not allow changes to the student data to be made from the graphics.

Furthermore, whilst user model visualisers built their own renders which are suited for the specific user model format considered, we explore a powerful, generic IV tool (*OpenDX*). This has enabled us to quickly adapt *CourseVis* to distance learning courses and to deploy it in realistic settings. The generality of the visualisation tool used enables the deployment of *CourseVis* in other courses and CMS platforms.

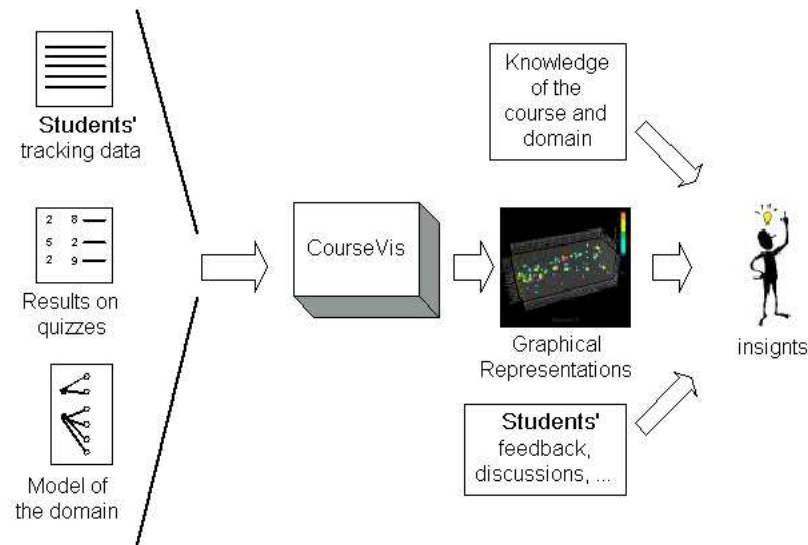


Figure 8.1: The process of generation of useful insights on instructor's mind with CourseVis

8.2.3 Contribution to Human Computer Interaction and Information Visualisation

Human-Computer Interaction is “*a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them*” (Hewett et al., 2002, p. 5). It studies the design of user-centred technologies and the evaluation of specific Input/Output devices such as visual user interfaces (Borner and Chen, 2003). The work presented in this thesis contributes to the design of user-centred educational technologies and to the use of Information Visualisation techniques to the deployment of visual interfaces for instructors of CMS. HCI methods were applied to investigate the instructors' needs when teaching on-line courses, and a number of graphical representations were designed to meet those needs. The evaluation of CourseVis addressed the usefulness of the graphical representations and was conducted in a realistic context involving Web instructors who are potential users of the tool.

We have applied Information Visualisation technologies and procedures to address real life problems in an educational domain. The thesis has demonstrated that visualisation tools can be used to improve the effectiveness of Web-based distance learning. The work, therefore, has contributed to the application of IV by showing the potential of this field into a new domain. Although recent educational systems employ some graphical representations, there is no systematic approach of how to apply IV to educational systems, and the techniques used are rather primitive with most educational systems having developed their versions of visualisation tools that lack generality and are hard to reuse (see Chapter 3). This work took a different approach by following IV procedures for visualising data and employing IV tools. This enabled generality - the method-

ology for designing and implementing CourseVis can be replicated easily into other educational systems. In this line of argument, the work represents not only a novelty of the use of IV to facilitate instructors in Web-based distance learning, but also shows a methodology of how IV can be applied into a wider range of user-centred systems (which may address educational problems as well as problems from other domains).

8.3 Future Work

The work carried out in this research produced a runnable prototype of CourseVis that has been used to illustrate and validate the proposed graphical representations. CourseVis has been regarded as a very useful tool by the participants in the evaluative studies we have conducted, and further improvement has been suggested.

From the technical point of view, we have to acknowledge that a significant amount of work has to be done in order to promote CourseVis from a raw prototype to a usable application. CourseVis' representations have been implemented with the OpenDX visualisation tool. This tool can be used both as a standalone application and as library functions that can be programmed to be integrated with a third-part application. The current version of CourseVis use OpenDX as standalone application which loads input data from an external data file. This is sufficient for proof of concept, but for a wider application of CourseVis it should be a whole application integrated with the CMS, and the users shouldn't deal with data file or launching programs. The full integration with a CMS is one of the highly recommended improvements of this work. An interesting solution could be splitting the CourseVis in a client-server architecture, and visualising the views in a Web page. In this manner all the necessary software would remain on the server platform, and the user will interact with CourseVis through a Web browser.

CourseVis has been designed to support a generic CMS, but we implemented a "student data exporter" module only for the WebCT CMS. Modules to support other CMS should be deployed; this would strength also the generality of the CourseVis for other CMS.

Scalability of CourseVis is another issue worth mentioning here. We showed in Section 6.5 a couple of representations where techniques like zooming and panning helped in courses having an high number of students enrolled. However, the empirical evaluation did not include very large data sets and further investigation may be required of how instructors get insights into large data sets using CourseVis. It must be stressed though that IV techniques adopted in CourseVis facilitate mainly the work with large data sets.

CourseVis allows exporting representations in image files and comparisons of different stages of the course (see Section 6.3). In a fully fledged version of the system it may be sensible to do this process automatically, instead of manually by saving pictures, as illustrated in Section 6.3.

In order to keep the preparation for the deployment of CourseVis simple and to require as

little instructors' effort as possible to use the tool, the model of the domain used in the current implementation of CourseVis includes only a list of concepts and each concept is mapped to a subset of quiz questions and content pages. In contrast to models of the domains in most Intelligent Tutoring Systems, CourseVis does not explore relationships between concepts. Possible inclusion of hierarchical concept relations may have to be considered for more sophisticated domain models.

Questions in quizzes are not all the same: some concepts and questions are more difficult than others. The level of difficulty of concepts and questions is not considered by CourseVis. A more advanced model could associate for each question the concepts it tests together with their level of complexity, and take this into account in the computation of the performance.

Last but not least, the privacy issue ought to be mentioned. CMS collect and store in their repositories a large quantity of users' personal data which is rendered graphically in CourseVis. Before using CourseVis, and CMS in general, appropriate security mechanisms should be confirmed. Furthermore, there has to be a clear policy on how and for what purposes the students' personal data is used.

The above aspects will be addressed in the next version of CourseVis being implemented within the framework of the recently started EU funded Edukalibre project. The project examines the use of free, open source software to develop a truly free and open learning environment that will not only mimic commercial CMS but will also further extend the CMS functionality. The visualisation tools offered in CourseVis is one such extension. The generality of the tool enables its use with a variety of CMS and makes it possible to incorporate CourseVis into the open source CMS being developed as part of the Edukalibre project. This will enable the distribution of CourseVis to the wide community of educators involved in Web-based distance learning.

Bibliography

- Arsham, H. (1995). Interactive education: Impact of the internet on learning & teaching. <http://ubmail.ubalt.edu/~harsham/interactive.htm>.
- Avgeriou, P., Papasalouros, A., Retalis, S., and Skordalakis, M. (2003). Towards a pattern language for learning management systems. *Educational Technology & Society*, 6(2):11–24. <http://ifets.ieee.org/periodical/6-2/2.html>.
- Barr, A. and Feigenbaum, E. A. (1982). *The HandBook of Artificial Intelligence*. William Kaufmann.
- Barron, A. (1999). A teacher's guide to distance learning. Technical report, Florida Center for Instructional Technology College of Education, University of South Florida, Tampa, Florida.
- Bates, A. W. (1995). *Technology, open learning and distance education*. Routledge, London.
- Bates, A. W. (1999). *Managing Technological Change: Strategies for Academic Leaders*. Jossey Bass Publishers, San Francisco.
- Berge, Z. and Collins, M. (1995). Computer-mediated communication and the online classroom in distance learning. *Computer-Mediated Communication Magazine*, 2(4):6. <http://www.ibiblio.org/cmc/mag/1995/apr/berge.html>.
- Berners-Lee, T., Cailliau, R., Groff, J.-F., and Pollermann, B. (1992). World-wide web: The information universe. *Electronic Networking: Research, Applications and Policy*, 1(2):74–82. <http://citeseer.nj.nec.com/berners-lee92worldwide.html>.
- Bertin, J. (1981). *Graphics and Graphic Information Processing*. Walter de Gruyter, Berlin.
- Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. The University of Wisconsin Press. English Edition.
- BlackBoard (2003). Learning system, blackboard inc. <http://www.blackboard.com>.

- Borner, K. and Chen, C. (2003). Visual user interfaces: Challenges and opportunities. *IEEE Technical Committee on Digital Libraries Bulletin (TCDL)*, 1(1). <http://www.ieee-tcdl.org/Bulletin/current/borner/borner.html>.
- Brna, P., Bull, S., and Dimitrova, V., editors (2003). *AIED03, Proceedings of the workshop on Learner Modelling for Reflection*.
- Brown, W. C. and King, R. L. (1998). Artificial intelligence in education: Current progress and future prospects. Technical report, Engineering Research Center for Computational Field Simulation, Mississippi state University.
- Brusilovsky, P. (1994). Student model centered architecture for intelligent learning environments. In *Proc. of Fourth international conference on User Modeling, 15-19 August, Hyannis, MA, USA*.
- Brusilovsky, P. (1999). Adaptive and intelligent technologies for web-based education. *Kunstliche Intelligenz, Special issue on Intelligent Systems and Teleteaching*, 4:19–25.
- Brusilovsky, P., Eklund, J., and Schwarz, E. (1998). Web-based education for all: A tool for developing adaptive courseware. In *Computer Networks and ISDN Systems. Proceedings of the Seventh International WWW conference 14-18 April*, pages 291–300.
- Brusilovsky, P. and Miller, P. (2001). *Access to Knowledge: New Information Technologies and the Emergence of the Virtual University*, chapter Course Delivery Systems for the Virtual University, pages 167–206. Elsevier Science and International Association of Universities, Amsterdam. <http://www2.sis.pitt.edu/~peterb/papers/UNU.html>.
- Brusilovsky, P., Schwarz, E., and Weber, G. (1996). Elm-art: An intelligent tutoring system on world wide web. In Frasson, C., Gauthier, G., and Lesgold, A., editors, *Intelligent Tutoring Systems*, volume 1086 of *Lecture Notes in Computer Science*. Springer Verlag, Berlin.
- Bull, S. (1997). See yourself write: a simple student model to make students think. In Jameson, A., Paris, C., and Tasso, C., editors, *User Modeling: Proceedings of the Sixth International Conference*, pages 315–326. Springer.
- Bull, S. and Broady, E. (1997). Spontaneous peer tutoring from sharing student models. In Boulay, B. D. and Mizoguchi, R., editors, *Artificial Intelligence in Education - AIED97*, pages 143 – 150.
- Bull, S. and Nghiem, T. (2002). Helping learners to understand themselves with a learner model open to students, peers and instructors. In Brna, P. and Dimitrova, V., editors, *ITS 2002 Workshop on Individual and Group Modelling Methods that help Learners Understand Themselves*, pages 5–13.

- Cannone-Syrcos, B. and Syrcos, G. (2000). *International Perspectives on Tele-Education and Virtual Learning Environments*, chapter 10. Ashgate.
- Cantoni, L. (2001). Nuove tecnologie nei processi di formazione, notes of the course. Univesity of Lugano, Faculty of Communication Sciences.
- Card, K. S., Mackinlay, J. D., and Shneiderman, B. (1999). *Readings in Information Visualization, using vision to think*. Morgan Kaufmann, Cal. USA.
- Card, S. K. and Mackinlay, J. (1997). The structure of the information visualisation design space. In *InfoVis97, IEEE symposium on Information Visualisation, Phoenix, AZ*, pages 92–99.
- Card, S. K., Moran, T. P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates.
- Chec, J. (2003). Use of artificial intelligence in new learning environments. In *T.E.L. Technology Enhanced Learning Conference, Milan, Nov.20 - 21 2003*.
- Chen, C. (1999). *Information Visualisation and Virtual Environments*. Springer.
- Chin, D. N. (2001). Empirical evaluation of user models and user-adapted systems. *User Modeling and User-Adapted Interaction*, 11:181–194.
- Conati, C. and Zhou, X. (2002). Modeling students' emotions from cognitive appraisal in educational games. In *ITS02: Proceedings of 6th International Conference on Intelligent Tutoring Systems*, number 2363 in LNCS, pages 944–954. Springer Verlag.
- Cotton, K. (1988). Monitoring student learning in the classroom. School Improvement Research Series (SIRS). Northwest Regional Educational Laboratory, U.S. Department of Education. <http://www.nwrel.org/scpd/sirs/2/cu4.html>.
- De Bra, P. and Ruiter, J. P. (2001). Aha! adaptive hypermedia for all. In *Proceedings of the WebNet Conference*, pages 262–268.
- De Vicente, A. and Pain, H. (2002). Informing the detection of the students' motivational state: An empirical study. In Verlag, S., editor, *ITS02: Proceedings of 6th International Conference on Intelligent Tutoring Systems*, number 2363 in LNCS, pages 933–943.
- Denzin, N. K. and Lincoln, Y. S. (1998). *Collecting and interpreting qualitative materials*. Sage, Thousand Oaks, CA.
- Dillenbourg, P. (1999). *Collaborative-learning: Cognitive and Computational Approaches*. Elsevier. Oxford.

- Dillenbourg, P. and Self, J. (1992). A framework for learner modelling. *Interactive Learning Environment*, 2(2):111–137. <http://tecfa.unige.ch/tecfa/publicat/dil-papers-2/Dil.7.2.4.pdf>.
- Dimitrova, V. (2002). Style-olm: Interactive open learner modelling. *International Journal of Artificial Intelligence in Education*, 13. To appear.
- Dimitrova, V., Brna, P., and Self, J. A. (2000). Involving the learner in diagnosis potentials and problems. Tutorial presented at Web Information Technologies : Research, Education and Commerce, Montpellier, France.
- Dix, A., Finlay, J., Abowd, G., and Beale, R. (1998). *HUMAN-COMPUTER INTERACTION*. Prentice Hall, second edition.
- Eklund, J. and Sinclair, K. (2000). An empirical appraisal of the effectiveness of adaptive interfaces for instructional systems. *Educational Technology & Society Journal*, 3(4). http://ifets.ieee.org/periodical/vol_4_2000/eklund.html.
- Elmsari, R. and Navathe, S. B. (2001). *Fundamentals of database systems*. Benjamin/Cummings, Redwood.
- Faulkner, C. (1998). *The essence of human-computer interaction*. Prentice Hall, London; New York.
- Fisher, B., Conway, K., and Groeneboer, C. (1997). Virtual-u development plan: Issues and process. In *Ed-Media/ED-Telecom'97 World Conference on Education Multimedia/Hypermedia and World Conference on Educational Telecommunications*. Calgary, Canada, pages 352–357. <http://citeseer.nj.nec.com/466114.html>.
- Flueckiger, F. and Mazza, R. (2001). Open distance learning at southern switzerland universities. a comprehensive and multilingual approach involving universities, teachers, students, and administration. In *EUNIS 2001 - 7th International Conference of European University Information Systems, Berlin, Humboldt-University*.
- Forte, E., Forte, M. W., and Duval, E. (1996). Ariadne: A framework for technology-based open and distance lifelong education. In *7th EAAEIE International Conference "Telematics for Future Education Training"*. Oulu, Finland, June 12-14, pages 69–72.
- Galusha, J. M. (1997). Barriers to learning in distance education. <http://www.infrastructure.com/barriers.htm>.
- Garito, M. A. (1999). The television university new models of distance teaching and learning. In *Online Educa Berlin - 5th International Conference on technology Supported Learning*.

- Gibbs, A. (1997). Focus groups. *Social Research Update, University of Surrey, UK*. Num. 19. <http://www.soc.surrey.ac.uk/sru/SRU19.html>.
- Gisolfi, A., Dattolo, A., and Balzano, W. (1992). A fuzzy approach to student modeling. In *Computers & Education*, number 4, pages 329–334.
- Goldberg, M. W. (1996). Calos: An experiment with computer-aided learning for operating systems. In *Proceedings of the ACM's 27th SIGCSE Technical Symposium on Computer Science Education*.
- Goldberg, M. W., Salari, S., and Swoboda, P. (1996). World wide web - course tool: An environment for building www-based courses. *Computer Networks and ISDN Systems*, 28:1219–1231.
- Goldfarb, C. F. and Prescod, P. (2002). *The XML Handbook*. Prentice Hall PTR, fourth edition.
- Graham, C. R., Swafford, M. L., and Brown, D. J. (1997). Mallard: A java enhanced learning environment. In *WebNet'98, World Conference on the WWW, Internet and Intranet. Toronto November 1-5 1997*, pages 634–636.
- Greenbaum, T. L. (1998). *The Handbook for Focus Group Research*. SAGE publications, CA, USA.
- Greenberg, G. (1998). Distance education technologies: Best practices for k-12 settings. *IEEE Technology and Society Magazine*, 17(4):36–40.
- Greenwood, S. (2001). University of bristol focus group on teaching with virtual learning environments. 2001, October 24. http://www.ltss.bris.ac.uk/vle_focus_oct01.htm.
- Greer, J. (1996). Student modelling and intelligent tutoring systems. Tutorial notes on The Fifth International Conference on User Modeling (UM96) Kailua-Kona, Hawai'i, January 2-5, 1996.
- Guarino, N. (1997). Understanding, building, and using ontologies: A commentary to - using explicit ontologies in kbs development-, by van heijst, schreiber, and wielinga. *International Journal of Human and Computer Studies*, 46:293–310.
- Hara, N. and Kling, R. (2000). Student distress in web-based distance education course. *Information, Communication & Society*, 4(3):557–579.
- Hartley, D. and Mitrovic, A. (2002). Supporting learning by opening the student model. In *Intelligent Tutoring Systems, 6th International Conference, ITS 2002, Biarritz, France and San Sebastian, Spain, June 2-7, 2002, Proceedings*, volume 2363 of *Lecture Notes in Computer Science*, pages 453–462. Springer.

- Hawkes, L. W. and Derry, S. J. (1996). Advances in local student modeling using informal fuzzy reasoning. *International Journal on Human-Computer Studies*, (45):697–722.
- Hearst, M. (2003). Information visualization: Principles, promise, and pragmatics. Handouts of the tutorial at CHI 2003 Conference on Human Factors in Computing Systems.
- Helic, D., Maurer, H., and Scherbakov, N. (2000). Web based training: What do we expect from the system. In *Proceedings of ICCE 2000, Taiwan*, pages 1689–1694. <http://coronet.iicm.edu/denis/pubs/icce2000.pdf>.
- Hermann, I., Melancon, G., and Marshall, M. S. (2000). Graph visualisation and navigation in information visualisation. *IEEE Transaction on Visualization and Computer Graphics*, 1(6):24–43. <http://www.cwi.nl/InfoVisu/Survey/StarGraphVisuInInfoVis.html>.
- Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., and Verplank, W. (2002). Acm sigchi curricula for human-computer interaction. <http://sigchi.org/cdg/index.html>.
- Holden, S. and Kay, J. (1999). The scrutable user model and beyond. In *Proceedings of 9th International Conference on Artificial Intelligence in Education, Le Mans, France, July 19–23*.
- Holt, P., Dubs, S., Jones, M., and Greer, J. (1994). The state of student modelling. In Greer, J. E. and McCalla, G. I., editors, *Student Modelling: The Key to Individualized Knowledge-Based Instruction*, volume 125 of *NATO ASI Series F, Computer and Systems Sciences*, pages 3–35. Springer-Verlag.
- Holt, P., Rubin, E., and Schmittroth, L. (1987). Student monitoring in distance education. *ICDE Bulletin*, 14:54–58.
- Johnson, B. and Shneiderman, B. (1991). Tree-maps: A space filling approach to the visualization of hierarchical information structures. In *Proceedings of IEEE Visualization 1991 Conference, San Diego*, pages 284–291.
- Johnson, P. (1992). *Human-computer interaction: psychology, task analysis and software engineering*. McGraw-Hill, London.
- Judith, L. (2003). *Distance Education: The Complete Guide to Design, Delivery and Improvement*. Teachers College Press, Columbia University, New York.
- Kaplan, B. and Maxwell, J. A. (1994). Qualitative research methods for evaluating computer information systems. In Anderson, J. G., Aydin, C., and Jay, S. J., editors, *Valuating Health*

- Care Information Systems: Methods and Applications*, pages 45–68. Sage, Thousand Oaks, CA,.
- Kay, J. (1995). The um toolkit for cooperative user modelling. *User modeling and user adapted interaction*, 4:149–196.
- Kay, J. (1997). Learner know thyself: Student models to give learner control and responsibility. In Halim, Z., Ottomann, T., and Razak, Z., editors, *Proceedings of International Conference on Computers in Education. Association for the Advancement of Computing in Education (AACE)*, pages 17–24.
- Kay, J. (1999). *A scrutable user modelling shell for user-adapted interaction*. PhD thesis, Basser Department of Computer Science, University of Sydney, Australia.
- King, R. L. (1998). Classification of student modeling approaches for intelligent tutoring. Technical report, Engineering Research Center for Computational Field Simulation, Mississippi state University.
- Kosba, E., Dimitrova, V., and Boyle, R. (2003). Fuzzy student modelling to advise teachers in web-based distance courses. In Hoppe, U., Verdejo, F., and Kay, J., editors, *Proceedings of the 11th International Conference on Artificial Intelligence in Education (AIED2003)*, volume 97 of *Frontiers in Artificial Intelligence and Applications*, pages 458–460. IOS Press.
- Krueger, R. A. (2000). *Focus Groups: A Practical Guide for Applied Research. Third Edition*. Sage Publishing, Newbury Park, CA,.
- Larkin, J. H. and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. In Glasgow, J., Narayanan, H., and Chandrasekaram, B., editors, *Diagrammatic reasoning - cognitive and computational perspectives*, pages 69–109. California: AAAI PressThe MIT Press. 1995. Reprinted from *Cognitive Science*, 11, 65-100, 1987.
- LearningSpace (2003). Ibm lotus corp. <http://www.lotus.com>.
- Mackinlay, J. D. (1986). Automating the design of graphical presentations of relational information. *ACM Transaction on graphics*, 5(2):110–141.
- Mackinlay, J. D. (2000). Information visualisation. Course notes. University of Aarhus, Denmark. <http://www.daimi.au.dk/~mackinla/iv-course/index.html>.
- Mackinlay, J. D., Robertson, G. G., and Card, S. K. (1991). The perspective wall: Detail and context smoothly integrated. In *ACM Conference on Human Factors in Computing Systems (CHI '91)*, pages 173–179.

- Mark, M. A. and Greer, J. E. (1993). Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence and Education*, pages 129–153. N. 4.
- Mayo, M. and Mitrovic, A. (2000). Using a probabilistic student model to control problem difficulty. In *Intellignet Tutoring Systems. 5th International Conference, Montreal, Canada, June 2000*, Lecture Notes in Computer Science, pages 524–533. Springer.
- Mazza, R. (2001). Formazione continua e classe virtuale: l'apprendimento del linguaggio java tramite tecnologie di formazione a distanza. In *XXXIX Congresso annuale AICA, Tecnologie, Innovazione e Società, Como*, pages 19–27.
- Mazza, R. (2003). Using open student models in distance learning environments to help teachers provide adaptive tutoring. *Studies in Communication Sciences. Special Issue New Media in Education*, pages 245–251.
- McCormack, C. and Jones, D. (1997). *Building a Web-Based Education System*. Wiley, New York.
- Merceron, A. and Yacef, K. (2003). A web-based tutoring tool with mining facilities to improve learning and teaching. In Hoppe, U., Verdejo, F., and Kay, J., editors, *Proceedings of the 11th International Conference on Artificial Intelligence in Education (AIED2003)*, volume 97 of *Frontiers in Artificial Intelligence and Applications*, pages 201–208. IOS Press.
- Mitrovic, A. (2003). Supporting self-explanation in a data normalization tutor. In Aleven, V., Hoppe, U., Kay, J., Mizoguchi, R., Pain, H., Verdejo, F., and Yacef, K., editors, *Supplementary proceedings of AIED 2003*, pages 565–577.
- Moodle (2003). Moodle course management system. <http://www.moodle.org>.
- Morales, R., Pain, H., Bull, S., and Kay, J., editors (1999). *Open, interactive, and other overt approaches to learner modelling (Proceedings of the workshop at AIED 99)*, volume 10.
- Morgan, D. L. (1988). *Focus group as qualitative research*. Sage, London.
- Murray, W. R. (1998). A practical approach to bayesian student modeling. In *Intellignet Tutoring Systems. 4th International Conference, San Antonio, Texas, USA, August 1998*, Lecture Notes in Computer Science, pages 424–433. Springer.
- Murray, W. R. (1999). An easily implemented, linear-time algorithm for bayesian student modeling in multi-level trees. In *9th International Conference on Artificial Intelligence in Education, Le Mans, France July 19–23, 1999*.
- Nielsen, J. (1997). The use and misuse of focus groups. *IEEE Software*, 14(1):94–95.

- Pappas, G., Ledermann, E., and Broadbent, B. (2001). Monitoring student performance in online courses: new game-new rules. *Journal of Distance Education*, 16(1). <http://cade.athabasca.ca/vol162/pappasetal.html>.
- Pilkington, R. (2001). Analysing educational dialogue interaction: Towards models that support learning. *International Journal of Artificial Intelligence in Education, Special Issue on Analysing Educational Dialogue Interaction*, 12.
- Preece, J., Rogers, Y., and Sharp, H. (2002). *Interaction Design*. John Wiley & Sons.
- QAA (2001). Academic review: subject review, earth, environmental sciences and environmental studies. Technical report, The Quality Assurance Agency for Higher Education. University of Aberdeen. http://www.qaa.ac.uk/revreps/acrev/sr017_01.pdf.
- Ragan, L. C. (1998). Good teaching is good teaching: an emerging set of guiding principles and practices for the design and development of distance education. *DEOSNEWS, The American Center for the Study of Distance Education, Pennsylvania State University*, 8(12). <http://www.educause.edu/ir/library/html/cem9915.html>.
- Ragnemalm, E. L. (1995). Student diagnosis in practice; bridging a gap. Ida technical report, Department of computer and information science, Linköping, Sweden. <http://www.ida.liu.se/publications/cgi-bin/tr-fetch.pl?r-95-36+ps>.
- Reed, M. and Heller, D. (1997). Olive, on-line library of information visualization environments. <http://otal.umd.edu/Olive>.
- Reffay, C. and Chanier, T. (2002). Social network analysis used for modelling collaboration in distance learning groups. In Verlag, S., editor, *ITS02: Proceedings of 6th International Conference on Intelligent Tutoring Systems*, number 2363 in LNCS, pages 31–40.
- Rehani, B. and Sasikumar, M. (2002). Chaatra: A student monitoring and learner modelling system. In Mudur, S. P., Sasikumar, M., and Dhamapurkar, D. S., editors, *Online Learning: Selected Readings, Proceedings of the International Conference, Vidyakash-2002*. Vikas Publishing House PVT LTD.
- Rivera, J. C. and Rice, M. L. (2002). A comparison of student outcomes & satisfaction between traditional & web based course offerings. *Online Journal of Distance Learning Administration. The State University of West Georgia.*, 5(3).
- Rober, N. (2000). Multidimensional analysis and visualization software for dynamic spect. Master's thesis, Otto-von-Guericke-Universitat, Magdeburg.

- Rosenberg, M. J. (2001). *e-Learning - Strategies for developing knowledge in the digital age*. McGraw-Hill.
- Rossmann, M. H. (1999). Successful online teaching using an asynchronous learner discussion forum. *Journal of Asynchronous Learning Networks*, 3(2):91–97.
- Ruddle, R., Brodrie, K., and Dimitrova, V. (2002). Communication, visualisation and interaction. University of Leeds, School of Computing. Handbook of the course.
- Sachinopoulou, A. (2001). Multidimensional visualization. Technical report, Espoo 2001. Technical Research Centre of Finland, VTT Tiedotteita, Meddelanden, Research Notes 2114. <http://www.inf.vtt.fi/pdf/tiedotteet/2001/T2114.pdf>.
- Scott, J. (1991). *Social Network Analysis. A Handbook*. SAGE Publication, London.
- Scriven, M. (1967). The methodology of evaluation. In Tyler, R., Gagne, R., and Scriven, M., editors, *Perspectives of curriculum evaluation*, pages 39–83. Rand McNally, Chicago.
- Self, J. (1994). The role of student models in learning environments. *Transaction of the Institute of Electronics, Information and Communication Engineers*, pages 3–8.
- Sharp, J. E. (2000). E-teaching simply with e-mail. In *FIE 2000 - The 30th IEEE Frontiers in Education Conference, October 18-21, 2000 Kansas City, MO*. IEEE Education Society.
- Shneiderman, B. (2002). Tutorial on information visualization. INFORMS Roundtable. November 2002, San Jose, CA.
- Shute, V. J. and Regian, W. J. (1993). Principles for evaluating intelligent tutoring systems. *Journal of Artificial Intelligence and Education*, pages 245–271. N. 4.
- Smith-Gratto, K. (1999). Best practices and problems north carolina a & t state university. Report to the Distance Education Evaluation Task Force Distance Education. <http://qed.ncat.edu/ir&p/report.htm>.
- Specht, M., Weber, G., Heitmeyer, S., and Schoch, V. (1997). Ast: Adaptive www-courseware for statistics. In Brusilovsky, P., Fink, J., and Kay, J., editors, *Proc. of Workshop Adaptive Systems and User Modeling on the World Wide Web at 6th International Conference on User Modeling, UM97, Chia Laguna, Sardinia, Italy*, pages 91–95.
- Spence, R. (2001). *Information Visualisation*. Addison-Wesley.
- Stallman, R. M., Lessig, L., and Gay, J. (2002). *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Free Software Foundation.

- Statz, B. L. (2001). Managing on-line classrooms: suggestions for assign student participation and providing timely and meaningful feedback. *Academy of Educational Leadership Journal*, 5(2).
- Stauffer, K. (1996). Student modeling and web-based learning systems. Technical report, Athabasca University.
- Suraweera, P. and Mitrovic, A. (2002). Kermit: A constraint-based tutor for database modeling. In *ITS 2002, Proceedings of 6th International Conference*, LNCS 2363, pages 377–387.
- Sutton, L. A. (2001). The principle of vicarious interaction in computer-mediated communications. *International Journal of Educational Telecommunications*, 7(3):223–243.
- The W3C Consortium, Bray, T., Paoli, J., Sperberg-McQueen, C. M., and Maler, E. (2000). Extensible markup language (xml) 1.0 (second edition). W3C Technical Recommendation.
- Thomas, P. and Paine, C. (2002). Monitoring distance education students' practical programming activities. *Education Technology & Society*, 3(5). http://ifets.ieee.org/periodical/vol_3_2002/thomas.html.
- TopClass (2003). Wbt systems. Dublin, Ireland. <http://www.wbt systems.com/>.
- Tsinakos, A. A. and Margaritis, K. G. (2000). Student models: The transit to distance education. *European Journal of Open and Distance Learning*.
- Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut, USA.
- Tufte, E. R. (1990). *Envisioning Information*. Graphics Press, Cheshire, Connecticut, USA.
- Uther, J. (2001). *On the visualisation of large user models in web based systems*. PhD thesis, School of Information Technologies, University of Sydney.
- Uther, J. and Kay, J. (2003). Vlum, a web-based visualisation of large user models. In Brusilovsky, P., Corbett, A., and De Rosis, F., editors, *Proceedings of the 9th International Conference on User Modeling, 22 to 26 of June 2003, Johnstown, USA*, LNAI, 2702, pages 198–202. Springer.
- Valentine, D. (2002). Distance learning: Promises, problems, and possibilities. *Online Journal of Distance Learning Administration*, 5(3). State University of West Georgia, Distance Education Center.
- Vasandani, V. and Govindaraj, T. (1991). An experimental evaluation of an intelligent tutor for diagnostic problem solving. In Birnbaum, L., editor, *The Internal Conference on the Learning Sciences*. Association for the Advancement of Computing in Education.

- Vidyakash (2000). Web-based environments for online learning. Technical report, Educational Technology Group National Centre for Software Technology Mumbai. http://www.ncst.ernet.in/vidyakash/reports/lms_comparison_report.html.
- Villano, M. (1992). Probabilistic student models: Bayesian belief networks and knowledge space theory. In *Intelligent Tutoring Systems, Second International Conference, ITS'92, Montreal, Canada*.
- Wall, L., Christiansen, T., and Orwant, J. (2000). *Programming Perl, 3rd ed.* O'Reilly & Associates.
- Ware, C. (2000). *Information Visualization, perception for design*. Morgan Kaufmann Publishers, San Francisco, USA.
- WebCT (2003). World wide web course tool, webct educational technologies. Vancouver, Canada. <http://www.webct.com>.
- Weber, G., Kuhl, H. C., and Weibelzahl, S. (2001). Developing adaptive internet based courses with the authoring system netcoach. In *Proceedings of Eight International Conference on User Modeling, Southofen, Germany, July 13-17*.
- Wegner, S. B., Holloway, K. C., and Garton, E. M. (1999). The effects of internet-based instruction on student learning. *Journal of Asynchronous Learning Network*, 3(2):98–106.
- Weibelzahl, S. and Weber, G. (2003). Evaluating the inference mechanism of adaptive learning systems. In Brusilovsky, P., Corbett, A., and De Rosis, F., editors, *Proceedings of the 9th International Conference on User Modeling, 22 to 26 of June 2003, Johnstown, USA*, LNAI, 2702. Springer.
- Willms, S. (2003). Visualising a user model for educational adaptive information retrieval. In Brusilovsky, P., Corbett, A., and De Rosis, F., editors, *Proceedings of the 9th International Conference on User Modeling, 22 to 26 of June 2003, Johnstown, USA*, LNAI, 2702, pages 432–434. Springer.
- Xiong, R. and Donath, J. (1999). Peoplegarden: Creating data portraits for users. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*. http://smg.media.mit.edu/papers/Xiong/pgarden_uist99.pdf.
- Zapata-Rivera, J. D. and Greer, J. E. (2001a). Externalising learner modelling representations. In *Workshop on External representations in AIED: Multiple forms and multiple roles*, pages 71–76. Held with the AI-ED 2001 - 10th International Conference on Artificial Intelligence in Education, San Antonio, Texas.

- Zapata-Rivera, J. D. and Greer, J. E. (2001b). Visualizing and inspecting bayesian belief models.
In *International Joint Conference on Artificial Intelligence IJCAI*, pages 47–49.

Appendix A

Questionnaire presented in Chapter 4

This appendix reports a copy of the questionnaire we submitted to some distance teachers in December 2002. Aim of this survey is to understand which peculiar information are needed by instructors when they are teaching a course at distance. Detailed results of this survey are illustrated in appendix B, while the most interesting results have been discussed in chapter 4 and taken in consideration for the implementation of CourseVis.

On-line Learning Environment Survey

Thank you for taking part in this study. Object of this research is to provide assistance to tutors in distance learning environments. The aim of this survey is to learn more about which information distance tutors need in their teaching activity. Your answers will help us to determine how use the distance learning environment as a teaching resource more effectively.

There are 17 questions which may take approximately 20 minutes to answer. If you do not know an answer, give the best answer you can, or leave it blank.

The study respects the confidentiality and anonymity of participants. No personal or confidential data will be collected.

Click on Submit when ready to send

Part A: User related

1. Please indicate how many on-line courses you are involved in.

- 1
- 2-5
- 6-10
- over 10

2. Please indicate your involvement in on-line courses (you can select more than one option)

- Instructional designer
- Course coordinator
- Instructor/teacher/tutor/facilitator
- Teacher assistant
- Help desk
- Technical assistant
- Student
- Other

3. What level of students have you worked with in on-line courses?

- Secondary school
- University/college undergraduate
- University/college postgraduate
- Continuous education/adult courses

4. How many students were in your largest class taught at a distance?

- Less than 20
- 20-50
- 51-100
- More than 100

5. How long have you been involved in distance learning courses?

- 1 year or less
- 2 years
- 3 years
- 4 years
- 5 years or more

Part B: Platform related

6. Which of the following web-based course management tools have you used in distance learning courses? (you can select more than one option)

- Ariadne
- BlackBoard
- Bodington
- BSCW
- Centra
- First class
- ILIAS

- Lotus Learning Space
- Toolbook
- TopClass
- Virtual-U
- WebCT
- Other tool (Please specify)

7. Which of the following facilities have you used in on-line courses? (you can select more than one option)

- Content materials
- Discussion forum
- E-mail
- Chat (synchronous communication)
- Shared whiteboard
- Glossary
- Calendar
- Student presentations
- Web based audio and/or video conference
- Students home pages
- Other tool (Please specify)

part C: Assessment related

8. Which of the following assessment techniques and tools have you uses in on-line courses? (you can select more than one option)

- Quiz test (e.g. multiple-choice, fill-in the blank, match, etc.)
- Assignments (evaluated by the teacher)
- Group work
- Analysis of discussions in forums
- Usage statistics (e.g. pages visited, track of student activities, etc.)
- Analysis of log files
- Other assessment methods (please specify)

9. Which of the following information regarding a student are you interested in for your teaching activity?

	Extremely inter- esting	Very inter- esting	Interesting	Somewhat interesting	Not at all interesting	I don't know
Access to the course (e.g. access frequency, number of pages read)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quiz and assignment grade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participation in dis- cussions (number of articles posted/read)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Posting e-mail to col- leagues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Frequency of re- visiting of the same page	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participation in chat	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participation in group exercises	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Is there any other information that you are interested in the previous question? Please specify the information and its importance for managing on-line courses.

11. Some information changes during the course. Which of the following time-dependent information are you interested in? (you can select more than one option)

- ☐ Level of knowledge of students during the time
- ☐ Access to the course materials
- ☐ Participation in discussions
- ☐ Others (please specify)

12. For each concept of the domain of your course, what information about your students with respect to the concept are you interested in? (you can select more than one option)

- ☐ Students who have read the course materials for that concept
- ☐ Students who have performed the evaluation proofs (e.g. assignments, quiz, etc.) for that concept
- ☐ Students who have found difficulties with that concept
- ☐ Level of knowledge of each student for that concept
- ☐ Other information (please specify)

- To adapt teaching to individual or group of students
- To identify and remedy common misconceptions
- To set up optimal peer learning or tutoring groups
- To respond to specific individuals in an appropriate way
- Other (please specify)

--

- o Students who are progressing too quickly with the course schedule
- o Students who are progressing too slowly with the course schedule
- o Students who do (or do not) accessing the course for certain period
- o Students who do (or do not) participate actively in discussions by posting messages
- o Students who do (or do not) participate passively in discussions by reading messages
- o Students whose results in assessments are significantly different from the mean of the results the class (performing very badly or very well)
- o Other categories of students (please specify)

[illegible]

Vicarious interaction takes place when a student actively observes and processes the interaction between other students or between other students and the instructor, without taking an active part in interaction.

Part D: feedback related

16. If you are interested in receiving a summary of this survey, please put in your e-mail address. You will be notified as soon as this summary is ready. You can leave it blank if you do not want information about the results of this study.

17. Any additional comments

Thank you very much for your participation, please press Submit Survey button to send your answers.

Appendix B

Tabulated results of the survey presented in Chapter 4

In this appendix we report the tabular results of the survey we submitted to instructors. The figures reported in this appendix have determined the recommendation we elicited in chapter 4 which have been used to lead the implementation of CourseVis. Each question existing in the questionnaire is described both in numerical tabular format and in graphical format. For each question we recall the text of the question, the type of the question (single answer, multiple choice, free text), the number of people who answered to the question and finally the number of people who leave the question blank. A copy of the questionnaire is available in the appendix A.

B.1 Countries

It could be interesting to know which and how many countries we addressed with this survey. For this reason a small program was developed to detect the host name from where each response was submitted. Using the host name we can derive the country of the submitter.

The following table represents a list of countries reached by our survey, and the number of submissions received from each country. The same information is also represented in a graph.

Note that that this is only a partial list. Some questionnaires were submitted from a host whose name hasn't a resolution in the DNS. In other words, some host has only the numerical IP address (es. 195.176.176.173) and does not have an alphanumerical host name (es. www.unisi.ch). Because it is not possible to derive the country of a host which hasn't an alphanumerical host name, this sort of addresses was not considered.

Host name extension	Percentage of answers from this country
au	2
ca	4
ch	23
com	7
de	1
fr	1
edu	18
es	2
hk	1
ie	1
it	1
mil	1
net	11
tr	1
uk	2
us	1

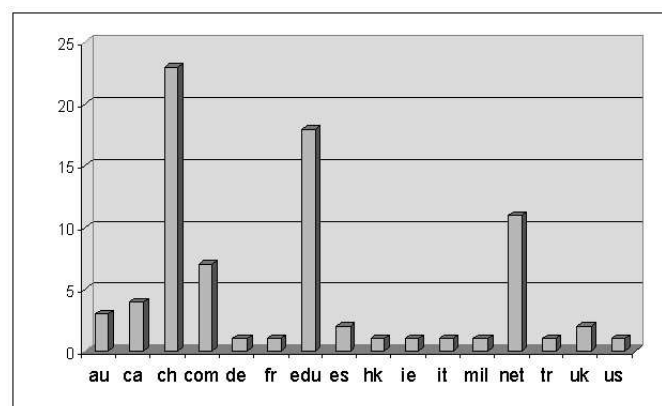


Figure B.1: Submits by countries

B.2 Question Q1

Question: Please indicate how many on-line courses you are involved in.

Type of question: Single

Number of respondents to this question: 96

Number of blank: 2

Response	How many	%
1	25	26.04
2-5	56	58.33
6-10	8	8.33
over 10	7	7.29

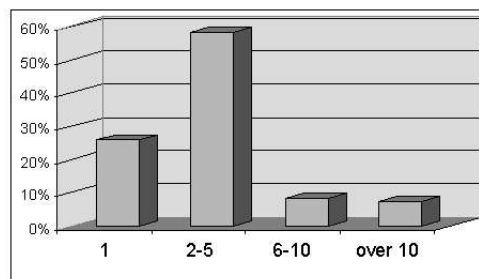


Figure B.2: Number of courses managed by instructors

B.3 Question Q2

Question: Please indicate your involvement in on-line courses.

Type of question: Multiple choice

Number of respondents to this question: 93

Number of blank: 2

Response	How many	%
Instructional designer	65	68.75
Course coordinator	45	48.96
Instructor/teacher/tutor/facilitator	72	76.04
Teacher assistant	14	15.63
Helpdesk	5	5.21
Technical assistant	10	11.46
Student	8	8.33
Other	4	4.17

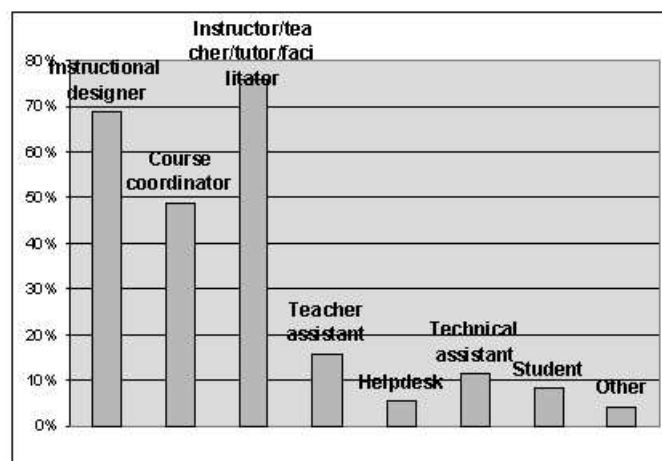


Figure B.3: Type of involvement in courses

B.4 Question Q3

Question: What level of students have you worked with in on-line courses?

Type of question: Multiple choice

Number of respondents to this question: 95

Number of blank: 3

Response	How many	%
Secondary school	7	7.37
University/college undergraduate	54	56.84
University/college postgraduate	51	53.68
Continuous education/adult courses	29	30.53

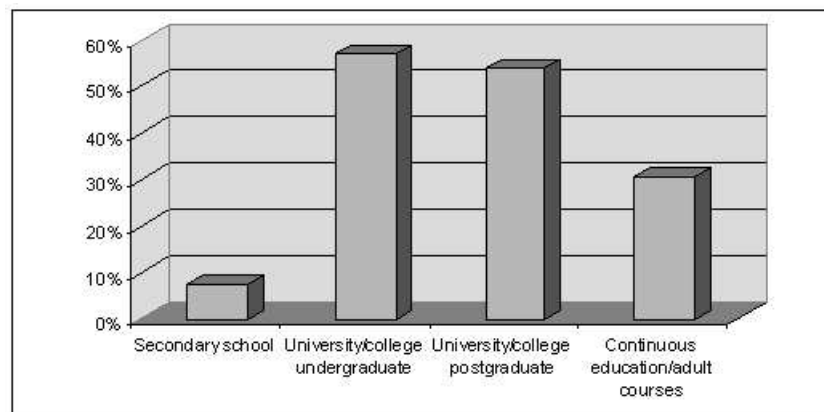


Figure B.4: Level of instruction of courses

B.5 Question Q4

Question: How many students were in your largest class taught at a distance ?

Type of question: Select one

Number of respondents to this question: 94

Number of blank: 4

Response	How many	%
Less than 20	32	34.04
20-50	39	41.49
51-100	15	15.96
More than 100	8	8.51

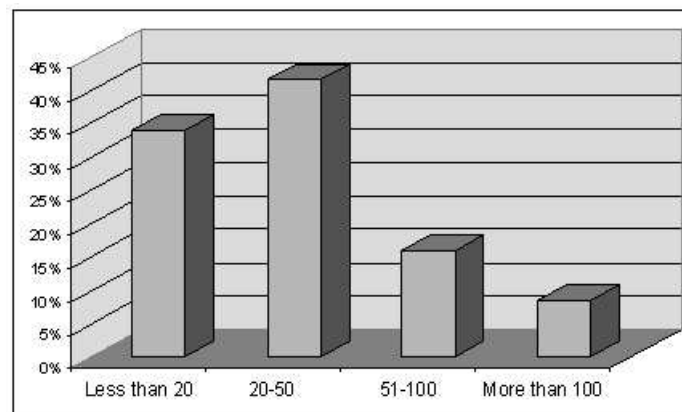


Figure B.5: Number of students in the largest class

B.6 Question Q5

Question: How long have you been involved in distance learning courses?

Type of question: Select one

Number of respondents to this question: 96

Number of blank: 3

Response	How many	%
1 year or less	16	16.67
2 years	21	21.88
3 years	23	23.96
4 years	8	8.33
5 years or more	28	29.17

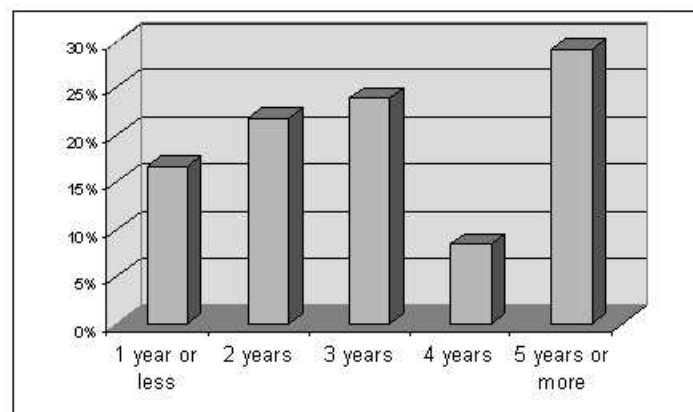


Figure B.6: length of involvement in on line courses (years)

B.7 Question Q6

Question: Which of the following web-based course management tools have you used in distance learning courses?

Type of question: Multiple choice

Number of respondents to this question: 94

Number of blank: 4

Response	How many	%
Ariadne	1	1.06
BlackBoard	36	38.30
Bodington	1	1.06
BSCW	11	11.70
Centra	10	10.64
Fist Class	7	7.45
ILIAS	1	1.06
Lotus Learning Space	9	9.57
Toolbook	1	1.06
TopClass	4	4.26
Virtual-U	2	2.13
WebCT	49	52.13
Other tool	39	41.49

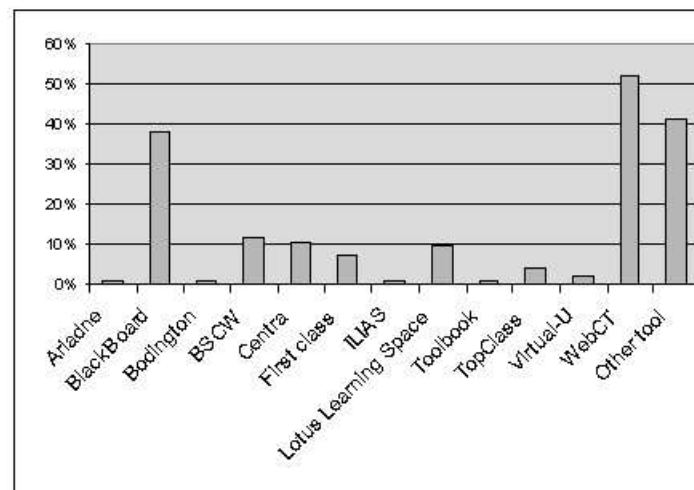


Figure B.7: Educational platforms used by instructors

B.8 Question Q7

Question: Which of the following facilities have you used in on-line courses?

Type of question: Multiple choice

Number of respondents to this question: 95

Number of blank: 3

Response	How many	%
Content materials	86	90.53
Discussion forum	78	82.11
E-Mail	81	85.26
Chat	53	55.79
Shared whiteboard	23	24.21
Glossary	31	32.63
Calendar	38	40.00
Student presentations	36	37.89
Web-based audio and/or video conference	29	30.53
Students home pages	25	26.32
Other tool	17	17.89

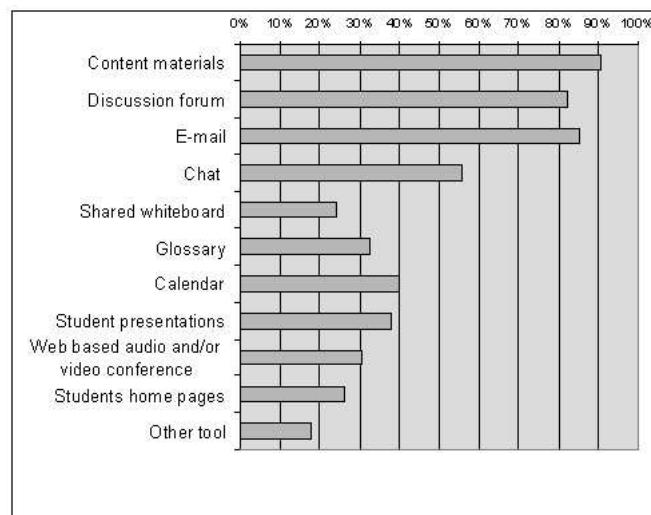


Figure B.8: Facilities used in courses

B.9 Question Q8

Question: Which of the following assessment techniques and tools have you uses in on-line courses?

Type of question: Multiple choice

Number of respondents to this question: 93

Number of blank: 5

Response	How many	%
Quiz test	63	67.74
Assignments	71	76.34
Group work	61	65.59
Analysis of discussions in forums	46	49.46
Usage statistics	30	32.26
Analysis of log files	16	17.20
Other assessment methods	16	17.20

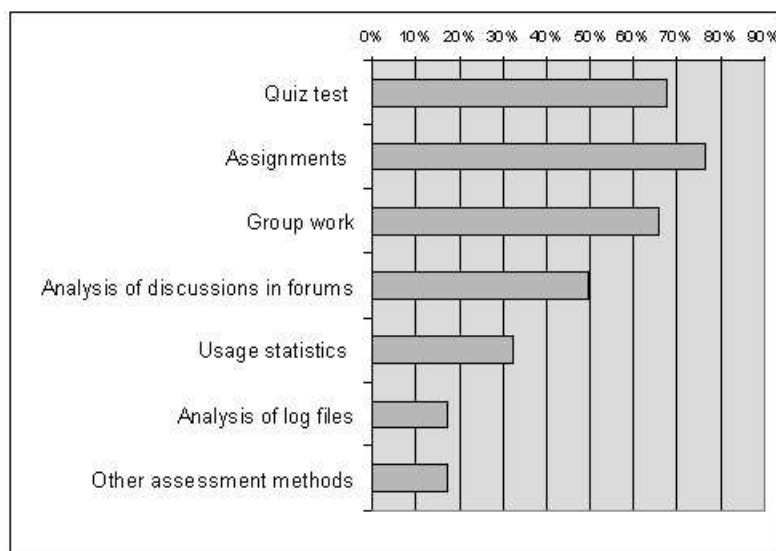


Figure B.9: Assessment techniques and tool used in courses

B.10 Question Q9

Question: Which of the following information regarding a student are you interested in for your teaching activity?

Type of question: Multiple choice

Number of respondents to this question: 95

Number of blank: 3

Response	Extremely interesting		Very interesting		Interesting		Somewhat interesting		Not at all interesting		I don't know	
	#	%	#	%	#	%	#	%	#	%	#	%
Access to the course	18	18.95	34	35.79	22	23.16	10	10.53	4	4.21	3	3.16
Quiz and assignment grade	34	35.79	28	29.47	14	14.74	10	10.53	5	5.26	3	3.16
Participation in discussions	32	33.68	30	31.58	17	17.89	5	5.26	5	5.26	2	2.11
Posting e-mail to colleagues	15	15.79	17	17.89	29	30.53	13	13.68	13	13.68	4	4.21
Frequency of revisiting of the same page	8	8.42	17	17.89	18	18.95	18	18.95	21	21.11	3	3.16
Participation in chat	11	11.58	15	15.79	21	21.11	19	20	16	16.84	10	10.53
Participation in group exercises	33	34.74	31	32.63	15	15.79	5	5.26	4	4.21	7	7.37

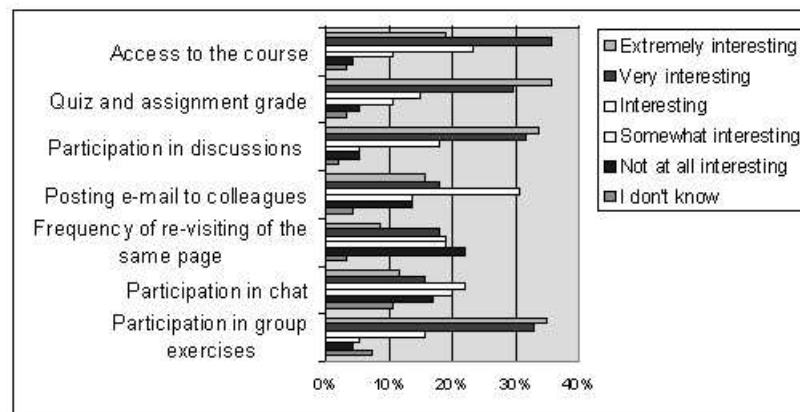


Figure B.10: Interesting information about students

This table reports a summary of the previous values.

Response	Extremely in- teresting + very interest- ing	Interesting + Somewhat interesting	Not at all in- teresting + I don't know
	%	%	%
Access to the course	54.74	33.68	7.37
Quiz and assignment grade	65.26	25.26	8.42
Participation in discussions	65.26	23.16	7.37
Posting e-mail to colleagues	33.68	44.21	17.89
Frequency of revisiting of the same page	26.32	37.89	25.26
Participation in chat	27.37	42.11	27.37
Participation in group exercises	67.37	21.05	11.58

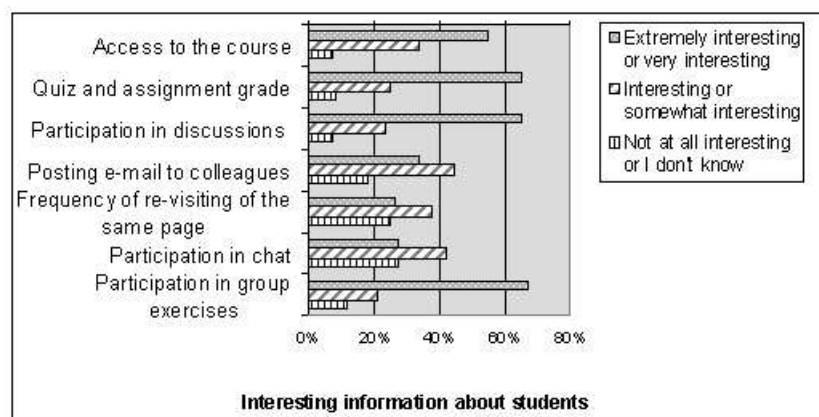


Figure B.11: Interesting information about students (summarised version)

B.11 Question Q10

Question: Is there any other information that you are interested in the previous question? Please specify the information and its importance for managing on-line courses.

Type of question: Free text

Number of respondents to this question: 26

Number of blank: 72

List of responses:

- Peer evaluations
- The previous question was confusing.
- Ability to control the access to the exam questions after the exam.
- Quality of contribution - much more important than frequency
- Quality of discussion/contributions. Ability to facilitate discussion with others.
- Asking questions, answering questions, discussion continuation
- We pay most attention to the quality of the postings made by students, and consider the number of postings as secondary to that element. That being said, we do expect students to have at least 6-10 meaningful postings per week in an 8 week course. Students are assigned to group sizes of 8-10 for the duration of the course.
- Student's feedback about lessons is very important in order to improve the on-line course.

B.12 Question Q11

Question: Some information changes during the course. Which of the following time-dependent information are you interested in?

Type of question: Multiple choice

Number of respondents to this question: 93

Number of blank: 5

Response	How many	%
Level of knowledge of students during the time	78	83.87
Access to the course materials	57	61.29
Participation in discussions	60	64.52
Others	3	3.23

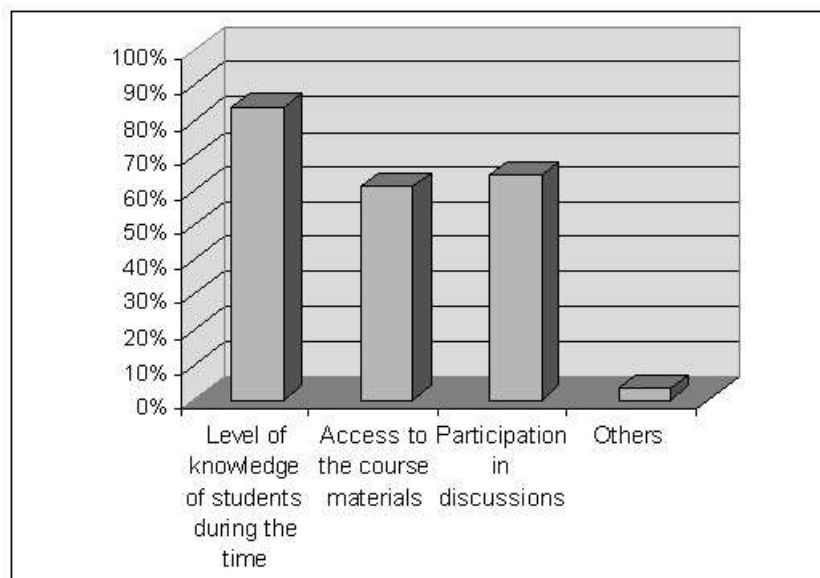


Figure B.12: Interest in time-dependent information

B.13 Question Q12

Question: For each concept of the domain of your course, what information about your students with respect to the concept are you interested in?

Type of question: Multiple choice

Number of respondents to this question: 95

Number of blank: 3

Response	How many	%
Students who have read the course materials for that concept	60	63.16
Students who have performed the evaluation proofs for that concept	61	64.21
Students who have found difficulties with that concept	68	71.58
Level of knowledge of each student for that concept	60	63.16
Other information	10	10.53

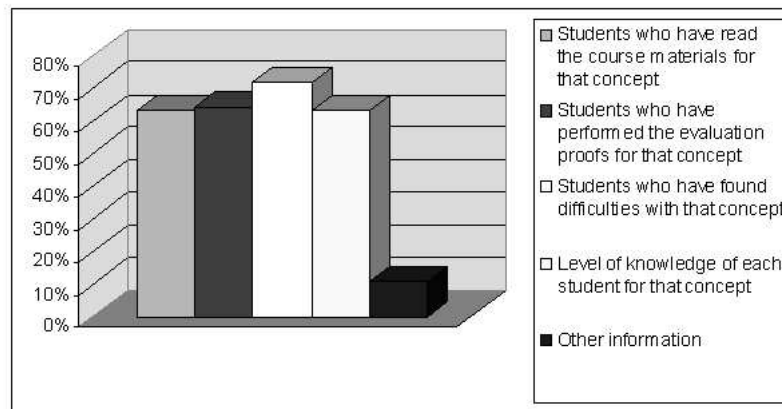


Figure B.13: Interest in information related with concepts of the domain of the course

B.14 Question Q13

Question: Assuming that your learning environment can give you information about your students, how would you use this information in your courses

Type of question: Multiple choice

Number of respondents to this question: 95

Number of blank: 3

Response	How many	%
To adapt teaching to individual or group of students	68	71.58
To identify and remedy common misconceptions	76	80.00
To set up optimal peer learning or tutoring groups	42	44.21
To respond to specific individuals in an appropriate way	63	66.32
Other	5	5.26

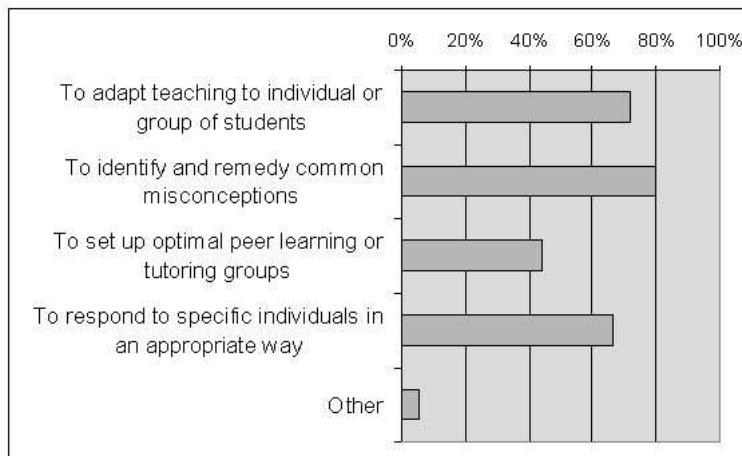


Figure B.14: How instructors want to use students' information

B.15 Question Q14

Question: In your on-line teaching activities, which of the following types of students are you particularly interested in: In your on-line teaching activities, which of the following types of students are you particularly interested in

Type of question: Multiple choice

Number of respondents to this question: 93

Number of blank: 5

Response	How many	%
Students who are progressing too quickly with the course schedule	27	29.03
Students who are progressing too slowly with the course schedule	53	56.99
Students who do (or do not) accessing the course for certain period	46	49.46
Students who do (or do not) participate actively in discussions by posting messages	58	62.37
Students who do (or do not) participate passively in discussions by reading messages	37	39.78
Students whose results in assessments are significantly different from the mean of the results of the class (performing very badly or very well)	58	62.37
Other categories of students	11	11.83

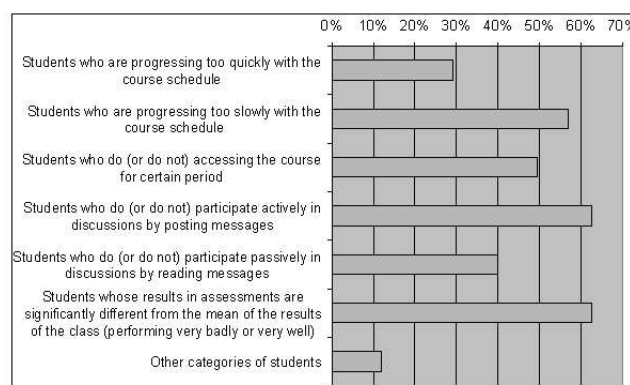


Figure B.15: Type of interesting students to monitor

B.16 Question Q15

Question: Previous research has identified five kinds of interaction that affect the learning process in distance education (L.A. Sutton, 2001). For each one of those, could you please tell how important they are with respect to monitoring your students' activities and progress?

Type of question: Multiple choice

Number of respondents to this question: 94

Number of blank: 4

Response	Extremely interesting		Very interesting		Interesting		Somewhat interesting		Not at all interesting		I don't know	
	#	%	#	%	#	%	#	%	#	%	#	%
Learner-content interaction	40	42.55	33	35.11	17	18.09	1	1.06	1	1.06	2	2.13
Learner-instructor interaction	29	30.85	35	37.23	19	20.21	7	7.45	2	2.13	2	2.13
Learner-learner interaction	29	30.85	33	35.11	16	17.02	10	10.64	2	2.13	4	4.26
Learner-interface interaction	13	13.83	24	25.53	31	32.98	17	18.09	1	1.06	6	6.38
Vicarious interaction	8	8.51	29	30.85	26	27.66	15	15.96	6	6.38	5	5.32

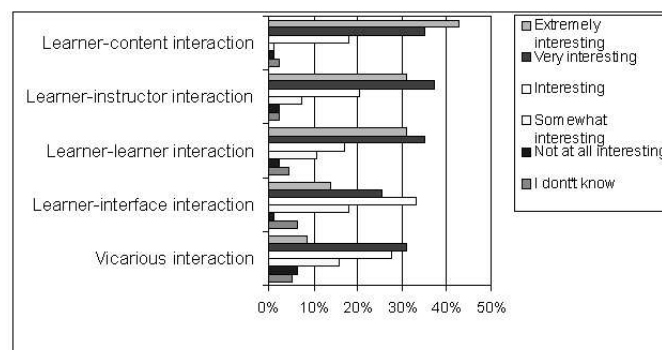


Figure B.16: Interest in different typology of students' interaction

B.17 Question Q17

Question: Any additional comments

Type of question: Free text

Number of respondents to this question: 26

Number of blank: 72

List of responses:

- A lot of preparations are required for instructor. Several people (grading assistant technical assistant) should be in classroom every time the instructor delivers his/her lecture. What if some video/audio technical difficulty problems arise during the lecture?
- Many instructors do both Web-only and Web-supported distance learning courses. It may be important to distinguish between these or have instructors tell you which they use.
- I am very interested in working further with on-line learning. However, I feel that the training and guidelines for teachers and instructors on how to set up and run such service are still lacking.
- I got confused with your terminology - I am teaching "distance" - it is synchronous - my class is broadcast via satellite to students at distance locations who watch me on a TV. A lot of your questions are about "on-line" - my class is not on-line, but is distance so a lot of your questions don't apply to me.
- Sorry I can't answer these. I've had training for teaching on-line, but haven't done so.
- I am a foreign language teacher, and interaction is fundamental for language acquisition, that is why vicarious interaction is not the kind of interaction I would promote in my courses.
- In fact the experiment is in development and implies 4 partner universities
- I do not understand the point of Q15. These activities are all related to the process, not the outcomes.
- The personal and social side of student interaction between tutor and peers is crucial to forming good relationships - often ignored in research into cognitive development - also how do they feel about the learning environment and relationships
- I would like to know what kind of a research are you doing and do you do a qualitative or quantitative research.
- I found the drop out rates varied dramatically depending on whether the course is structured or unstructured with structured having less of a drop-out rate.
- The answers above concern current development in progress of an on-line course rather than hands-on experience with existing course.

- I'm involved in blended systems... but the questionnaire seems to consider only on-line learning ... so I decide to take in account only one of the systems I'm participating in
- How do we show enthusiasm over the web?

Appendix C

DTD description

This appendix provides a conceptual, informative description of the elements in the DTD used by CourseVis.

Elements fields are reported in tabular format for an easy identification. The columns in these tables refer to:

Name: it is the name by which the data element is referenced.

Explanation: a brief description of the element.

Req: indicates if the element is required. Possible values are:

M = mandatory element that must be included,

O = optional element

Multi: the multiplicity of the element. If an element is repeated then all sub-elements of the tree are repeated with the same element. Possible values are:

1 a single instance

n the element is repeatable

Datatype: indicates, when applicable, whether the values are string (**S**), integer (**I**), real (**R**), date(**D**), time(**T**) or DateTime(**DT**) .

Value space: is the set of allowed values for the data element (when applicable). This depends by the datatype used. In case of strings, this defines the maximum number of char of the string; in case of integer or real this specifies the range of allowed values. All DateTime must be specified in the format YYYY-MM-DD hh:mm:ss, the date in the format YYYY-MM-DD, the time in the format hh:mm:ss <<<<<specify the standard>>>>.

C.1 Students data DTD structure description

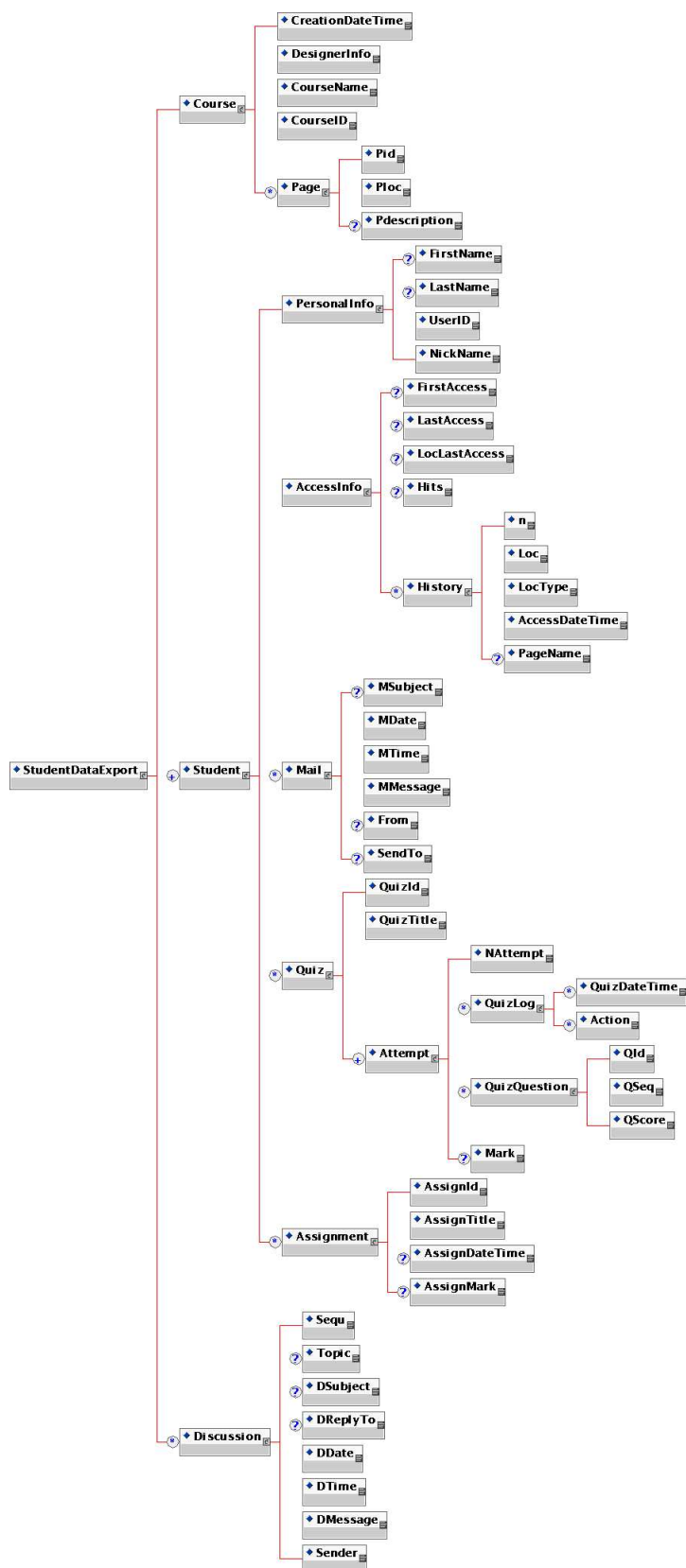


Figure C.1: Student data structure

Name	Explanation	Req	Multi	Datatype	Valuespace
StudentDataExport	Is the top level element. It's used as root element to enclose all the other data.	M	1		
Course	Collects a list of descriptive elements about the course data exported.	M	1		
CreationDateTime	The date and the time when the course has been created in the CMS.	M	1	DT	
DesignerInfo	Personal information about the designer of the course (name, address, contact info, ...).	M	1	S	1..1024
CourseName	The name of the course, as used in the CMS.	M	1	S	1..256
CourseID	The unique string identifying the course, as defined in the CMS.	M	1	S	1..256
Page	Collects a list of pages exactly as defined in the table of the contents of the course. The pages must be listed in the same order which appears in the table of the contents of the course. If the course is composed by more table of contents, these must be grouped in a single table, and reported in this collection.	O	n		
Pid	Sequence number of the page. Starts by 1. Pages are ordered using this sequence number, according to the table of contents of the course	M	1	I	1..1000
Ploc	The location of the page within the course. This must be a unique identificative value. The path of the page in the filesystem could be used for that.	M	1	S	1..256
Pdescription	A textual description of the page. The title of the page, if existing, can be used.	O	1	S	1..1024
Student	Collects a list of elements related to a specific student. This element is repeated for each student of the course.	O	n		
PersonalInfo	Collects a list of elements regarding the students personal information.	M	1		
FirstName	The first name of the student.	O	1	S	1..256
LastName	The last name of the student.	O	1	S	1..256
UserID	A single string which identify uniquely a student in the CMS. Usually this correspond to the username for the login into the course.	M	1	S	1..256
NickName	This is merely an alias to the UserID that is used within CourseVis to guarantee the anonymity of the students when presenting the data or the visualisations outside the scope of the course.	M	1	S	1..256
AccessInfo	Collects a list of elements regarding statistical details about student tracking	M	1		
FirstAccess	The date and the time that the student first accessed the course	O	1	DT	
LastAccess	The date and time that the student last accessed the course	O	1	DT	
LocLastAccess	The location (path) of the last page accessed by the student	O	1	S	1..256
Hits	The total number of times that the student accessed the pages of the course or tools (E-mail, discussion, quiz, etc.)	O	1	I	1..10000
History	Collects a list of elements related to the history of the pages of the course visited by the student. This collect not only the content pages, but also the other pages of the course, like the homepage, the tools, etc. This element is repeated for each page visited by the student.	O	n		
n	Sequence number. Starts by 1	M	1	I	1..10000

Loc	The location of the page within the course. This must be a unique identificative value. The path of the page in the filesystem could be used for that. This field is optional if the type of the page visited is not a Notes.	O	1	S	1..256
LocType	This element describe the type of the page visited. It assumes one of the following values: Home_page Notes Tool_pages Articles_Read Original_Posts Glossary Annotations Follow_ups	M	1	S	1..256
AccessDateTime	The date and the time that the student made this access	M	1	DT	
PageName	A textual description of the page visited. The title of the page, if existing, can be used.	O	1	S	1..1024
Mail	Collects a list of elements which describe the mails sent or received by this student. This element is repeated for each mail sent or received by the student.	O	n		
MSubject	The subject of the mail	M	1	S	1..1024
MDate	The date the mail was sent	M	1	D	
MTime	The time the mail was sent	M	1	T	
MMessage	The body of the message	M	1	A	undef
From	If present, this element contains the name of the sender of this e-mail. If absent, means that this mail was sent by the current student	O	1	S	1..256
SendTo	If present, this contains the name of the user who receive the message; if blank, means that the current student is the addressee of this e-mail. If a single e-mail is sent to more receivers, the whole e-mail is repeated for each receiver. The SentTo field always contains only one name.	O	1	S	1..256
Quiz	Collects a list of elements regarding the quiz performed by the student. This element is repeated for each quiz performed by the student.	O	n		
QuizId	An identificative number assigned by the CMS to the quiz.	M	1	I	1..1000000000
QuizTitle	The title of the quiz	M	1	S	1..256
Attempt	Collects a list of elements regarding each attempt made by the student to perform this quiz.	M	n		
NAttempt	Sequence number of the attempt. Starts by 1.	M	1	I	1..1000
QuizLog	Log information regarding the current attempt of the quiz	O	n		
QuizDateTime	The date and the time of the current action on the quiz	O	1	DT	
Action	A textual description of the action made by the student on the current attempt to perform a quiz. It assumes one of the following values: quiz started quiz submitted save question: X	O	1	T	1..1024
QuizQuestion	A quiz is composed by a list of questions. This element collects information about each single question	O	n		
QId	An identificative number assigned by the CMS to the quiz question.	M	1	I	1..1000000000
QSeq	Sequence number of the question in the current quiz. Starts by 1.	M	1	I	1..1000
QScore	The score achieved by the student on the current question	M	1	R	0..1000

Mark	The mark achieved by the student of the quiz. If not present, means that the student hasn't completed the quiz.	O	1	R	0..1000
Assignment	Collects a list of elements regarding the assignments done by the student. This element is repeated for each assignment performed by the student	O	n		
AssignId	An identificative number assigned by the CMS to the assignment	M	1	I	1..1000000000
AssignTitle	The title of this assignment	M	1	S	1..256
AssignDateTime	The date and the time that the student submitted this assignment to the system. If not present means that the student hasn't yet submitted the assignment	O	1	DT	
AssignMark	The grade received for this assignment. If not present means that the instructor hasn't yet graded the assignment	O	1	R	0..1000
Discussion	Collects a list of elements regarding the discussion in the course made with the discussion tool. This element is repeated for each message in the discussions	O	n		
Sequ	A sequence number assigned by the tool	M	1	I	1..10000
Topic	A textual description of the topic where this message was addressed	O	1	S	1..256
DSubject	The subject of the message	O	1	S	1..1024
DReplyTo	If this message was sent as reply to another message, this element contains the Sequ number of the message this message is replying	O	1	I	1..10000
DDate	Date message was sent	M	1	D	
DTime	Time the message was sent	M	1	T	
DMessage	The body of the message	M	1	S	undef
Sender	The UserID of the sender of this message	M	1	S	1..256

C.2 Domain data DTD structure description

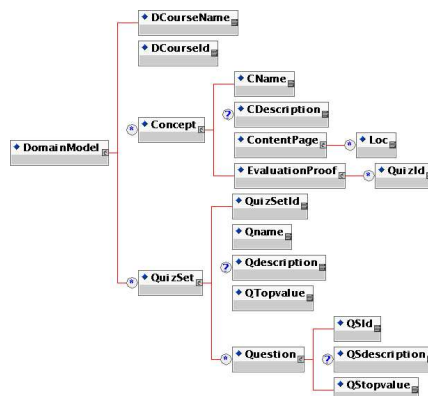


Figure C.2: Domain data structure

Name	Explanation	Req	Multi	Datatype	Valuespace
DomainModel	Is the top level element. It's used as root element to enclose all the other data.	M	1		
DCourseName	The textual name of the course	M	1	S	1..256
DCourseId	The unique string identifying the course, as defined in the CMS.	M	1	S	1..256
Concept	Collects a list of elements regarding a concept in the course. This element is repeated for each concept of the domain of the course	O	n		
CName	Name of the concept	M	1	S	1..256
CDescription	Textual description of the concept	O	1	S	1..1024
ContentPage	Collects a list of content page of the course related with the current concept.	M	1		
Loc	The location (path) of a page. This element is repeated for each page.	O	n	S	1..256
EvaluationProof	Collects a list of evaluationProofs (quiz) related with the current concept.	M	1		
QuizId	The CMS identification number of the question in a quiz related to this concept	O	n	I	1..1000000000
QuizSet	A quizset correspond to a quiz created on the CMS for the course. It is composed by a set of questions. This element is repeated for each quiz existing in the course	O	n		
QuizSetId	The identification number given by the CMS to this quiz.	M	1	I	1..1000000000
Qname	The title of the quiz	M	1	S	1..256
Qdescription	A description of the quiz	O	1	S	1..1024
QTopvalue	The maximum grade that can be assigned to this quiz	M	1	R	1..1000
Question	Collects the list of questions that compose a quizset. This element is repeated for each question existing in the quiz	O	n		
QSid	The identification number given by the CMS to this quiz.	M	1	I	1..1000000000
Qdescription	A description of the question	O	1	S	1..1024
QStopvalue	The maximum grade that can be assigned to this question	M	1	R	1..1000