ArchMatrix: a Visual Interactive System for Graph-Based Knowledge Exploration in Archaeology

Barbara Rita Barricelli
Università degli Studi di Milano
Via Comelico, 39/41
20139 Milano, Italy
+390250314007
barricelli@dico.unimi.it

Stefano Valtolina
Università degli Studi di Milano
Via Comelico, 39/41
20139 Milano, Italy
+390250314009
valtolin@dico.unimi.it

Matilde Marzullo
Università degli Studi di Milano
Via Festa del Perdono, 7
20122 Milano, Italy
+390250312323
matilde@infinito.it

ABSTRACT
This paper presents the design and development of ArchMatrix, a visual interactive system that supports archaeologists in archiving, managing and studying the findings collected during archaeological excavations. This work allowed to deepen the study of participatory design in interdisciplinary communities and was carried out by computer scientists in close collaboration with archaeologists. To overcome the several issues that characterize the communication among different domain experts, participatory design’s principles have been applied. The ArchMatrix system implements the Harris Matrix method, used both in archaeology and in architecture to describe the position (absolute and relative) of stratigraphic units. ArchMatrix relies on a graph database that enables to apply advanced information retrieval strategies. In this way, archaeologists are supported in developing new opportunities for investigation to increase their knowledge, improve their traditional working practices and to develop new ones.

Categories and Subject Descriptors

General Terms
Design, Documentation, Human Factors.

Keywords
Visual Interactive Systems, Participatory Design, Knowledge Management, Interdisciplinary Projects, Harris Matrix, Archaeology.

1. INTRODUCTION
Information visualization strategies are applied in archaeology for assisting domain experts in the examination and interpretation of the stratigraphy of excavation sites, and identifying both natural and cultural strata. The British archaeologist Edward Cecil Harris in 1973 invented the Harris Matrix that is used to provide a means to view stratigraphic sequences in diagram form [1]. To support knowledge management and decision making by means of a Harris Matrix, we designed and implemented a visualization system able to meet archaeologists’ real needs in handling contents and structures. ArchMatrix is the result of a work that started with an analysis of limits and weaknesses of existing software solutions and that aimed at defining an innovative visualization tool for efficiently storing and managing archaeological excavation site knowledge. We enable the archaeologists to visualize the knowledge and to search for it using a graph-based visual environment, by offering a visual representation of archaeological assets and their relationships in order to support intuitive and useful explorations. This approach to the knowledge visualization responds to the need of facilitating the exploration and analysis of very vast amount of data: moreover, very often the research on excavation sites is performed not only by archaeologists but also in collaboration with other communities of experts such as architects, geologists, chemists, and biologists. The design and development of ArchMatrix have been performed by applying participatory design’s principles and involving a team of archaeologists.

2. HARRIS MATRIX
According to the experience we developed in years of collaboration with Etruscologists [2], the archaeological practice is mainly exploited in two phases. The first is carried out on the field and consists of excavation and collection of archaeological findings and related information. The second regards the analysis of the collected material and information and takes place only when the excavation campaign is over. During the excavation phase, each time that a stratigraphic unit is detected, the archaeologists fill in a proper form to keep track of its characteristics. First of all each unit is identified by a type: layer stratigraphic unit (US) or structural stratigraphic unit (USS). Then, the number of the box in which the findings retrieved in the unit are stored, the maximum and minimum altitude of the unit, and a description of the unit itself are given. In the rest of the form, the relationships among the unit and other units are made explicit. Three types of relationship could exist: active, passive, and neutral. Active relationships are a) covers, b) fills, c) leans, and d) cuts; passive relationships are a) is covered, b) is filled, c) relies on, d) is cut; neutral relationships are a) is equal to, and b) binds. Once the form is filled in, the archaeologists represent the stratigraphic unit in a drawing on paper that will be used in successive data analysis. At the end of the excavation campaign, all this material is collected and studied for creating the correspondent Harris Matrix. The resultant Harris Matrix enables the archaeologists to determine the chronology of the various units (an example is shown in Figure 1). The rectangles represent...
3. RELATED WORKS

We analyzed several existing software tools for Harris Matrixes creation. In what follows we briefly discuss pros and cons of two of the most used ones. Harris Matrix Composer is a Java application that offers a simple graphical editor for stratigraphic units and related relations drawing. US and USS can be made distinguishable by the use of different shapes to represent them. Also colors could be used but they are not assigned to a specific semantic. The application allows to define only one relationship, called “above”, that means that one unit lies over another one. The position of the units depends on the direction of the arrow drawn to depict the “above” relationship. Once all the units are connected, the software allows to group them into boxes that represent different overlays. Even if Harris Matrix Composer is one of the most used tools in this field, it is affected by several limits. Among the others, the system does not rely on databases to retrieve information about stratigraphic units and there is no possibility to associate extra information to the units. Just one type of relationship between units could be expressed and it is not allowed to manage overlays (nor manually or automatically). Proleg MatrixBuilder implements the “Chrono-stratigraphic” algorithm, aimed at automatically generate a Harris Matrix from a set of stratigraphic units connected by some relationships. The user can insert the data related to the stratigraphic units and their relationships by filling in some tables. Once the input phase is over, the software shows the correspondent Matrix. The user can choose shapes and colors to be used to describe the units (e.g. circle, rectangle, and hexagon). Proleg MatrixBuilder stores all the inserted information about the stratigraphic units and the relationships among them. However, the queries that are allowed are very simple and limited: just one unit at time could be retrieved and the only attribute that could be used as key is its identification number. Several limitations affect Proleg MatrixBuilder. Among the others, once the Harris Matrix is generated it is impossible to change the stratigraphic units’ position already inserted. The different types of relationships cannot be associated to different shapes or colors. The types of information that could be assigned to the units are limited. It does not offer tools for units’ retrieval on the basis of relationships or with complex queries and overlays are not manageable. Even if these two software applications provide useful features to draw and/or generate Harris Matrixes, they do not represent complete solutions to support archaeologists’ practice. The most important limitation is the lack of tools for analysis of the data – in particular the absence of complex query implementation and links to data regarding the findings that belong to the stratigraphic units.

4. PARTICIPATORY DESIGN

By bringing together different cultures, experiences, skills, working practices, and working habits, like happens in the case of an archaeological excavation site, the way in which collaborative work is carried out is deeply influenced. The design of an efficient tool for storing, handling, and visualizing archaeological knowledge by means of a Harris Matrix needs to mainly take into account how archaeologists in their activities use a specialized notation, which reflect both their ways of reasoning and the criteria they adopt to document their work. To design a system that meets the archaeologists’ requirements, we applied participatory design (PD) techniques. PD [3] is a design approach that considers the participation of several stakeholders having expertise, backgrounds and knowledge grounded in different domains. PD techniques are used to integrate the skills of various stakeholders to obtain information from experts about their practices, their methods, their notations and everything about their work. Among the existing PD’s techniques, for ArchMatrix project we applied cooperative prototyping, PICTIVE method, and ethnographic method. Cooperative prototyping was introduced in [4]. It differs from traditional prototyping in which users’ only role is to provide feedback to the work done by designers. Indeed, the cooperative prototyping approach establishes a process where both the users – in our case archaeologists – and the designers are participating actively and creatively exploiting their different roles and expertise. Moreover, the result of cooperative prototyping is not just a software prototype but represents a learning process.
In fact, the archaeologists learned to use the system even before it was completed. Similarly, the participants in this process learned how to teach the use of ArchMatrix to future users. PICTIVE method (Plastic Interface for Collaborative Technology Initiatives through Video Exploration), was developed by Bell Communications Research. It is based on collaborative prototyping and visual communication paradigms. PICTIVE employs a combination of deliberately low-tech design tools (e.g., paper and pencil, colored labels, sticky notes) with high-tech video recording facilities (e.g., webcam). Mock-up prototypes of ArchMatrix were created using low-tech design tools. Recording facilities have been used to capture the process of prototyping, in order to use the recordings as a guide for the implementation of the system or to explain the logic that exists behind the design decisions. Finally, ethnographic method tries to connect ethnography with design of computer systems. It was proposed in [5] and is based on the fact that most human practices, and among these also work practices, cannot be valued outside of the social and cultural context in which they take place. The ethnographic method embeds and adopts the principles that guide traditional ethnographic approaches. Following this approach, designers of ArchMatrix observed archaeologists in their activities and they worked with them side by side for some weeks in order to understand their terminology, practices and the tacit knowledge applied in their context of use.

5. ARCHMATRIX
ArchMatrix allows to store Harris Matrixes and to display them by means of a Web-based application. For storing, displaying and handling Harris Matrixes, we studied a solution based on a graph database due the useful structure it offers for modeling US and USS, their relationships to each other, and the conceptual structures wherein they lie. Graph databases have been designed to meet a lack of the relational databases in storing relationship data. This weakness, described in [6], appears evident when it is needed to deal with many-to-many relationships or to store data about the relationships themselves. In RDBMSs (relational database management systems) the solution to this issue is the use of join tables, but this method has some evident weaknesses: it can greatly increase the number of tables in the database and to query relationships’ data is extremely unwieldy because may leads to recursive joins, based on very complicated SQL statements and unpredictable performances. Moreover, compared with relational databases, graph databases are often faster for handling relationship datasets. For ArchMatrix, we adopted a Neo4J database. Neo4J is a full ACID-transaction compliant graph database written in Java. The data structure is optimized for graph networks. We implemented the database using the REST server of Neo4J and a PHP-based Web application integrated with a Neo4J REST PHP library through which querying the REST services. The Harris Matrix is stored as a graph $G = (V, E)$ and modeled as a set of vertices (V) associated with a set of edges (E). The nodes represent US or USS whereas each edge is expressed as a pair of two units connected according to a specific type of relationship (covers, fills, leans, cuts, ties, or is equal). The graph is directed: $E_{ij}$ indicates an edge that is going from the node $i$ to $j$, and $E_{ji}$ indicates an edge going from the node $j$ to $i$ and it represents the inverse relationship of $E_{ij}$. Let us give an example to clarify the point: if the relationship of the edge $E_{ij}$ is “covers”, then it describes that unit $i$ covers unit $j$. In our Neo4J implementation both nodes and relationships can have properties. A property is a key-value pair where the key is a string and the value can be either a primitive type or an array of one primitive type. These properties are used to describe the characteristics of each node (US or USS and the phase to which it belongs) or the type of the relationship. To visualize a graph in ArchMatrix we searched for a visualization tool able to work with a Neo4J graph through a Web application and to enable useful exploration of data and analysis of graphical relationships. We also focused our research on tools that enable dragging, dropping and re-organizing operations with automatic updates of the database. We adopted Medusa a Java Applet for visualizing and handling graphs through a very intuitive user interface. Medusa displays a very high number of edges between nodes using Bezier curves and allows users to add and delete nodes and edges by simply clicking the mouse. The graphs are visualized according to a set of parameters retrieved in the database and each node is defined as

\[ \text{node\_name} : \text{R} : \text{G} : \text{B} : \text{X} , \text{Y} : \text{a} : \text{c} : \text{“url”} : \text{“a”} : \text{“c”} : \text{“I”}, \]

where: node\_name: is the name of the node; R, G, B: is the RGB combination of the color used to represent the node; s: is an integer between 0 and 3 that indicates the geometric form used to display the node; url: indicates the URL associated to the click on the node; a: is the label that appears over the node; c: is an integer greater than 0 that indicates a cluster which the node belongs to; I: is a string between quotes that indicates the layer linked to the node. Each edge is passed in the form

\[ \text{node\_name} : \text{R} : \text{G} : \text{B} : \text{X} , \text{Y} : \text{a} : \text{c} : \text{“url”} : \text{“a”} : \text{“c”} : \text{“I”}, \]

Figure 2. The units are positioned in the respective overlays.
related to specific units to increase the explicit knowledge different layers) and defined in a specific node property. This "chronological phase" which the units belong to (placed in graph according to their needs. For example in Figure 2 the user Matrix. Moreover, ArchMatrix allows the users to re-organize the sub-structures of the Harris Matrix. This allows to define queries locality for some specific graph operations by means of designing related ones. This subset visualization is aimed at improving the is visualized showing the relationships between this unit and the of the Harris Matrix indicating a specific unit, and then the graph annotations, positions and shapes. The user starts the exploration of the Harris Matrix indicating a specific unit, and then the graph is visualized showing the relationships between this unit and the related ones. This subset visualization is aimed at improving the locality for some specific graph operations by means of designing sub-structures of the Harris Matrix. This allows to define queries related to specific units to increase the explicit knowledge displayed in the graph but avoiding to visualize the whole Harris Matrix. Moreover, ArchMatrix allows the users to re-organize the graph according to their needs. For example in Figure 2 the user has rearranged the units in a Harris Matrix according to the "chronological phase" which the units belong to (placed in different layers) and defined in a specific node property. This property describes the chronological phase of an US or USS according to the chronologies of the findings discovered in the units and retrieved from the excavation database connected to the Neo4J database. Interacting with nodes and edges, users can modify some properties of the US/USS. For example, in case the findings discovered in a USS are dated to a later chronology in respect to the chronology of the USS itself (because it is a monument like a well) the archaeologist needs to modify its chronology dragging and dropping the USS among different layers (representing different chronological phases). These interactions apply direct modifications into the Neo4J database through the REST services.

6. EVALUATION

To evaluate ArchMatrix, we performed a user test that involved 5 archaeologists all experts in Etruscoology. Each user was required to perform a set of tasks. The tasks were the same for all the groups listed in the same order. We used the think-aloud approach and for each task performed by the user, one observer took note of starting time, duration of the execution (if the user does not complete a task in the maximum time allowed or by giving up spontaneously, the observer has to ask the user to pass to the next task), ist of errors committed by the user (if an error is committed more than once, the number of repetitions has to be noted), questions, comments and suggestions, satisfaction or frustration expressions, and usability problems. After a brief presentation of ArchMatrix and the goal of the test, the users were asked to fill in an initial questionnaire used to determine their profile. All the users were of Italian mother tongue, they use the computer both for work and for personal use and apply the Harris Matrix method often in their work activity. The tasks assigned to the users are all activities that the archaeologists need to perform in analyzing the US/USS and the relationships among them. Given the novelty in the management of the overlays, particular attention has been given to operations made on them, to investigate about the satisfaction of the users in having this kind of functionality available. The main problems in performing the tasks were related to the position of the main menu on the bottom of the interface. All the users seemed to search for it on the top of the screen and this caused a loss in terms of time of execution. Other issues were related to the lack of a right-click contextual menu. For example, almost all the users search for it to rename or delete overlays. Apart from these two main problems, ArchMatrix results a usable system. A final questionnaire has been submitted to the users at the end of the test. It was composed by three sets of questions: A first set was a SUS (system usability scale) questionnaire, a ten-item attitude Likert scale. The other set was a CUSQ (computer usability satisfaction questionnaires) questionnaire, a nineteen-item attitude Likert scale. A last part was dedicated to a comparison between ArchMatrix and other analogous systems used by the archaeologists and to a request of proposals for ArchMatrix improvement. What emerges from the user test performed and from the final questionnaire that the users were asked to fill in, is that ArchMatrix is more powerful than the other systems used by the archaeologists. They all recognize that ArchMatrix offers more possibilities in terms of modification of a Harris Matrix and that is more helpful in their daily work practice.

7. CONCLUSION

ArchMatrix is the result of a participatory process of design that involved both computer scientists and archaeologists. The application of Participatory Design techniques led to the design and development of a visual interactive system that deeply improved the practice of the archaeologists thanks to the direct manipulation features that ArchMatrix offers.

8. ACKNOWLEDGMENTS

This work is supported by “DESIRE: Creative Design for Innovation in Science and Technology” (PITN-GA-2008-215446-DESIRE).

9. REFERENCES