Visual Query Tools
for Uncertain Data in Space and Time

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE,
UNIVERSITY OF CAPE TOWN
IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Katherine M. Malan
November 2004

Supervised by
Gary Marsden

Abstract

Uncertainty in data is a common problem. In this study we address particular challenges around the visual querying of data with vague locations in time and space. Our study is based on an archive of African art. In this archive there is a mixture of contemporary and traditional art. Many of the artworks have vague locations in time and space. By this we mean that, although there is some idea of when and where an artwork originated, the precise information is unknown.

The aim of our project was to develop query and visualisation tools for this archive of African art. Our query system had to be designed to cope with different degrees of uncertainty in data, both in time and space. One of our initial ideas was to use a geographic map to drive the interface to the system. Items were plotted as a point on the map at some random location which fell within the given location of uncertainty for the item. However, user feedback indicated that this approach was problematic, as users assumed that the point on the map was the actual location of the artwork.

A better approach was found in driving the interface around a time-line metaphor. Time is a one-dimensional attribute, making it possible to use the second dimension to display uncertainty information, while still retaining a two-dimensional interface. Objects are represented on the time-line as either a dot, if an exact time is known, or a line, if the time is uncertain (the length of the line depicts the length of the period in which the object may have been created). Results of user studies showed that the time-line based approach was clear to users.

In addition to the display of uncertain data, we also developed techniques for storing and querying uncertain data. We show how data can be stored at different levels of uncertainty, without losing any available detail, and how queries can be expressed to include both certain and uncertain data. We use the concepts of dynamic queries; add to this a 2D query tool for performing spatial queries and tools for enabling Boolean combinations of queries. Results of prototype-based user tests indicated that users could successfully perform simple queries using the visual query tools.

We have combined these techniques into a single visual query system. In this way, we show how it is possible for novice users to easily query large multimedia archives with complex uncertain attributes.

Acknowledgements

My warm thanks to the following who have helped me along the way:

Firstly to Gary Marsden, my supervisor, for your insight and patience and for telling me that you thought my work was great when I was ready to quit.

Next, to John Turest-Swartz and the rest of the CAMA team – thank you for your time and energy and for the good times at Club Africa.

I am also grateful to the NRF who supported my work through a staff development grant for a number of years.

Finally, a special thank you to my sister, Lisa, for helping me to see things in perspective and to my husband, Greyling, for his encouragement and support.

Contents

Chapter 1

Introduction

In this chapter we discuss the context of the research and the particular problem we have addressed. We also explain how the report is organised.

1.1 The context

Africa is a continent rich in arts and culture. However, this creativity is often not acknowledged, both outside and inside Africa. Part of the problem is that knowledge about African culture is not readily available. Data on traditional African art and culture, including physical artefacts, are dispersed throughout the world. Some of the data is in Africa, owned by various African cultural institutions, while other data and artefacts are owned by museums and private collections in Europe and America. Contemporary African art and culture is also not well known. Although a select few African artists and musicians have made a name for themselves outside of Africa, there are many talented artists who are not showcased outside, or even within, the continent.

Fortunately, there are organisations and individuals who are working on promoting African art and culture. One such organisation is CAMA (Contemporary African Music & Arts Archive). CAMA aims "to develop a pan-African mass-accessible arts and culture archive, and through this, the means to acknowledge creativity in the arts in Africa on a continental scale" [8]. Through the CAN (Culture Africa Network) project, they have installed computers and digital recording equipment at seven key sites throughout Africa (Mozambique, Ethiopia, Sudan, Kenya, Ghana, Mali and South Africa). In this way they are facilitating the documentation of material in existing holdings within African cultural institutions.

One of the challenges facing CAMA is the task of providing general access to the data in their archive. This involves developing tools for disseminating information to the wide range of potential users of the CAMA archive. Our research has gone some of the way to addressing this issue of the dissemination of African cultural data.

1.2 The problem

CAMA approached us with a particular need. They wanted a front-end to their archive which was a map of Africa. This map should show (in some way) all the artefacts in their archive. The user would be able to select a subset of the displayed artefacts and the results would then be displayed. On investigation of the data, we realised this would not be the simple task we had envisaged. Most of the artefacts[1] had uncertain origins. In the case of some artefacts, the origin was a particular point on the map (such as a village). In the case of most of the traditional artefacts, however, the origin was a region within a country, a whole country or even a group of countries.

A further complication was that the archive contained a mixture of traditional and contemporary African art. One of the main differences between these two kinds of art relates to the nature of the artist. In the case of contemporary art, the artist is an individual person. With traditional art, on the other hand, the artist is a culture, i.e. a group of people who exist(ed) in a particular place and time. Time therefore emerged as a second important theme to add to space (the map of Africa) as an entry point into the archive. As an attribute, the time origin of artefacts had similar degrees of uncertainty to the geographic origin. For some artefacts in the archive, we knew in which year they were made, whereas in the case of most traditional artefacts, the origin was a period. These periods could span anything from a few hundred years to a few centuries.

We therefore had to develop creative ways of visualising and querying this data. This report details what we did to solve the problem.

1.3 Outline of the report

We start, in Chapter 2, by giving more detail about CAMA and the context of the research: the different aspects of CAMA's work, the nature of the data in the archive, the envisaged users of the archive and finally, the requirements of the system.

In Chapter 3 we provide an overview of areas that relate to our research. We have drawn on related work on information visualisation and visual queries, particularly dynamic queries. We also look at the issue of uncertainty in data and how others have visualised uncertain data.

In Chapter 4 we outline proposed techniques for displaying uncertain data in space and time. We also explain our proposed visual query mechanisms, which are based on the visualisations. Finally we explain how we propose to store the data to support the application.

In Chapter 5 we explain how we evaluated some of our proposed techniques. We provide an overview of existing approaches to evaluation and then explain how we evaluated our displays of uncertain time data and uncertain space data. The questionnaire used to evaluate the time-based display is given in Appendix B. We also describe a prototype-based evaluation of our proposed visual querying mechanisms. The task sheet used in this evaluation is given in Appendix C.

[1]In this document we use either of the terms 'artefact' or 'artwork' to refer to the entities stored in the CAMA data set.

In Chapter 6 we describe how our techniques could be incorporated into a single visual query system. We describe how the system would work and then revisit the original requirements as a way of reviewing the system. Many of these features were implemented in a pilot system. Details of the implementation are given in Appendix A.

In Chapter 7 we conclude by summarizing our contributions both in general terms, as visual query tools for uncertain data, and in specific terms, as a tool for accessing African art. We also outline a number of areas for further work.

Finally, we have published a number of papers relating to this research and these are provided in Appendix D.

Chapter 2

Background: CAMA

The work that initiated this research was conducted in collaboration with CAMA, the Contemporary African Music & Arts Archive [8], based within the Department of Computer Science at the University of Cape Town. CAMA aims to establish a continental multimedia database for African arts and culture and is in the process of digitally capturing as much contemporary African culture as possible. Beyond the archive, CAMA also aims to reflect African culture back on itself. Their belief is that this will positively affect the self- and global perception of African culture and help to stimulate creativity on the continent.

The CAMA Archive is conceived as having three "faces"[45]:

- **the hunter-gatherer**, which seeks out and records African culture,
- **the keeper**, responsible for holding material safely under archival conditions,
- **the messenger**, CAMA's multimedia publishing initiative which disseminates these collected works.

2.1 The hunter-gatherer

CAMA has established the Culture Africa Network (CAN) which aims to build capacity in Africa for cultural documentation and multimedia authoring. CAN is currently operating in seven key sites (Mozambique, Ethiopia, Sudan, Kenya, Ghana, Mali and South Africa). Each of these sites has been equipped with a Macintosh G4 and a digital video camera. There are many interesting research questions associated with the capture of data in an African environment. Some of these issues are discussed by Marsden *et. al.* [33] (a copy of the paper is provided in Appendix D).

2.2 The keeper

Cultural data is ideally suited to a multimedia database, particularly in Africa where visual art, oral culture and music are richly developed and closely integrated with people's lives. The

CAMA cultural database will include multiple continual input sources (from the different CAN sites), many different types of user, a variety of output products, and sophisticated multimedia data types. A number of interesting challenges emerge as one works with the CAMA archive, due to the wide range and complex nature of the data. These challenges are discussed below.

2.2.1 The nature of the data

There is a wide range of types of data to be stored in the archive. Documentary holdings currently include: photographic material, audio and video recordings, and texts. The photographic holdings cover artworks, exhibitions, cultural events, artists and musicians, and consist of a body of original CAMA material as well as previously published material used with permission. Audio holdings consist of original DAT recordings of music and interviews with artists and musicians, as well as published work. CAMA's most extensive documentation to date has been of the Royal Academy's 1995 exhibition 'Africa – the Art of the Continent', where CAMA was able to record more than 400 of the artworks on display photographically and on digital broadcast video.

Of the seven sites in CAN, South Africa serves as the central distribution node. In other words, all the sites send new digitised material to UCT for central archiving. The archive is therefore constantly being updated so the structure of the archive and applications should be able to handle large and continually changing volumes of data.

One of the problems associated with the storage and cataloguing of African art is the lack of any agreed universal classification scheme – there is no equivalent of, say, an "impressionist" period into which art may be classified. Whilst a lack of classifications is frustrating at present, it underlines how useful and necessary this type of project is in investigating and furthering an understanding of African art.

Another notable characteristic of African cultural data is the existence of different degrees of uncertainty. Some data items have precise locations (e.g., an artefact which comes from a particular town). However, there are many items which do not have precise locations. These items originate from somewhere within an area of uncertainty, which can be anything from a province within a country to the whole of Africa (e.g., a sculpture in a gallery in Paris, which comes from somewhere in Africa). In the same way, while the year of origin of some data items is known precisely, other items have uncertain origins in time. For example, one artefact could be dated 1940, another dated "late 19th Century" and yet another "1390 - 1352 BC". This uncertainty is present on many different themes, not just time and geographic location. One of the challenges we face is how to work effectively with both certain and uncertain data simultaneously.

The influences of data items on each other is an important theme of study in African culture. These influences run in many different dimensions/themes and the existence of influence is often based on subjective opinion. Geographic and historical influences, for instance inhospitable terrain, migration, and the profusion of different African languages and cultural groupings, make for a wide diffusion of influences. These themes of study and relationships between data need to be stored in a sufficiently flexible way to incorporate new themes and divergent interpretations as they arise. In Section 4.3 we propose a data model to store the uncertainty in the data and the relationships between data, focusing particularly on the themes of space and time.

2.3 The messenger

The main focus of our research has been on the dissemination of information through providing effective interfaces to the CAMA archive. An understanding of the users and their needs was central to the success of this work.

2.3.1 Users of the archive

There are essentially two kinds of CAMA users: those who are "clued and looking for information", and those who are "clueless and looking for inspiration". Although the full potential of the CAMA archive will only become apparent in due course, these are the users we envisaged:

- **CAN network members**: These are the organisations and individuals in African countries who are members of, or are affiliated to CAN. The focus of these users will be on archiving the local art and music and making it accessible to the outside world. Data archiving and distribution tools are the products envisaged for these users.

- **Researchers of African culture**: In addition to standard querying, these users would want to explore new themes and influences and create new points of view or layers to be stored in the database. Products envisaged for these users include basic and advanced querying and visualisation tools, including audio and visual content based retrieval.

- **Educators**: These users would want access to thematic samples of data for teaching purposes. Such illustrative versions of the database could be made available in an easily accessible form (e.g. CD ROM), with customised interfaces, as well as representative samples of the data that would be accessible via the web. Tools for automatically generating these teaching products (e.g. thematic maps) are envisaged.

- **Exploratorium users**: The envisaged environment here is a museum or exploratorium where both adults and children can explore the archive through virtual environments. The product would be an interactive tool with the focus on exploration, rather than directed search.

- **Casual web users**: These are any people with a casual interest in African culture. The initial focus will be on the provision of a representative subset of the data for browsing, together with meta-data, i.e. a fuller indication of what data is available in the CAMA database. A longer-term goal is to provide further means for deeper access to the whole database

There are a number of research projects currently exploring these different views of the CAMA archive. Our study focuses on the need for query and visualisation tools. Target users would include researchers of African culture and casual web users.

2.3.2 User needs

On consultation with CAMA, we agreed on the following requirements for a query and visualisation system:

1. Typical users would include artists, historians or academics, most probably with limited expertise in using a computer. The user interface should therefore be accessible to novice users and use metaphors familiar to such users. In the case of query tools, the usual SQL-based query mechanisms would be inappropriate.

2. The users would want to explore themes and influences, so the relationships between data elements should be displayed along particular themes.

3. Some data elements which have uncertain origins, are extremely important from a cultural perspective. Therefore, it is important that uncertain items be treated with the same level of importance as items which have definite origins.

4. Since the archive is continually updated, users of the archive would typically know very little about the nature of the data stored in the archive: data volumes, areas of coverage, etc. It is therefore important that the application provide an overview of the data and visual feedback on the results of queries.

A related project which works on the same CAMA data set is the Virtual Gallery Project [24] (a copy of the paper is given in Appendix C). The generation of a successful virtual gallery is dependent on appropriate filtering – the user will need some way of deciding which subset of data to display in the gallery. The query and visualisation tools developed in our project would form the front-end to one of the virtual galleries proposed in the Virtual Gallery Project.

Chapter 3

Related work

Our aim was to develop query and visualisation tools for researchers of African culture. This chapter provides an overview of areas of work that relate to our study.

We first look at types of data: the ways in which others have classified data and how the data from our study relates to these classification systems. We then look at the broad field of visualisation: the techniques that have been used to display data in a way that promotes insight. In the next section we look at visual querying: how queries can be formulated, how querying can be combined with visualisation in the form of dynamic queries and also how we can support users in formulating combinations of queries. Finally we look at uncertainty in data: the types of uncertainty that exist and how others have visualised uncertain data.

3.1 Types of Data

In this section we look at the broad types of data that exist. The aim is to see whether or not the data from our study falls into any of the categories of data described in the literature. Once we understand more about the nature of the data, we can investigate whether there are generic visualisation approaches which could be appropriate for our use.

Data can be divided very broadly into entities and relationships [44], depending on whether the aim is to visualise the actual object, or the way that the objects relate to one another. For example, in our study, both the artefacts themselves and the way that artefacts relate in time and space are of interest to us.

Entities can have attributes, which are properties that describe the entity. Relationships can also have attributes. To classify data, we look at the properties of these attributes.

3.1.1 Attribute Types

Card *et.al.*[10] provide a classification of attributes (which they call variables). They claim that variables come in three basic types:

1. *Nominal data* (where one value is either equal or not equal to another value), for example the name of an artefact such as 'Anubis mask' or 'Statue of Queen Ankhnesmeryre II'.
2. *Ordinal data* (where the values obey an ordering), for example rating of wine (1-star to 5-star).
3. *Quantitative data* (when arithmetic can be done on values), for example the height of an artefact in cm.

Both the time attribute and the space attribute of our application are quantitative in nature. Card *et.al.* also specify two particular subtypes relevant to our data, namely:

- Quantitative Time (Q_t) for temporal variables, and
- Quantitative Geographical (Q_g) for spatial variables that are geophysical coordinates.

3.1.2 Dimension of data

A further way of describing an attribute is by specifying its dimension. The dimension of a variable is often referred to as one of the following [44]:

- *scalar*: a single quantity, such as the weight of a person;
- *vector*, such as the direction in which that person is travelling;
- *tensor*: a higher-order quantity that describes both direction and shear forces.

This distinction between scalar, vector and tensor is sometimes referred to as the *order* of the data [25]. It is important not to confuse the notion of the dimension or order of data in the scientific sense as described above with the more general notion of dimension. For example, although our space attribute is clearly two-dimensional in nature, it does not have an order (of scalar, vector or tensor) in the scientific sense.

3.2 Visualisation

Given an understanding of the nature of the data, we need to find appropriate ways of visualising the data. The main purpose of visualisation is *insight*. By gaining insight into data, one can discover new things, make better decisions and explain things to others [10]. Although visualisation is inherently application-dependent, there are visualisation techniques which are generic in nature and can be applied to different classes of data.

Much of the research in computer visualisation has originated from applications which are scientific in nature. In the case of scientific visualisation, the aim is to visualise some underlying physical phenomenon. The term 'information visualisation' refers to the broader case where there is no underlying physical phenomenon to visualise – the aim is simply to visualise the data itself. Card *et.al.*[10] make the following distinction between scientific visualisation and information visualisation:

- *Scientific visualisation* is usually based on physical data (such as the human body, the earth or molecules), where the information is inherently geometrical in nature.
- *Information visualisation* is based on non-physical information (such as financial data, or collections of items), and so does not have any obvious spatial mapping for visualisation.

The CAMA data that we are working with is non-physical in nature. (If we were, for example, working with the precise shapes of artefacts, this would be physical data.) We therefore expect information visualisation techniques to be more applicable than scientific visualisation techniques. We do, however, look (in the next section) at existing classifications of scientific visualisation techniques, as some include contexts where there is no physical entity. After that we look at a framework for describing visualisations of non-physical data (in Section 3.2.2) and then give two examples of information visualisation applications (Sections 3.2.3 and 3.2.4).

3.2.1 Scientific visualisation Techniques

The techniques for scientific visualisation can be categorized based on the characteristics of the model of the physical phenomenon being visualised. For example, Hesselink *et.al.* [25] describe a classification of techniques for flow visualisation based on: the *order of the data* (scalar, vector or tensor); the *spatial domain dimensionality* of the objects used in the visualisation (points, lines/curves, surfaces or volumes); and the *information level* of the data (elementary, local or global). This classification assumes that there is an underlying model of the physical entity that is constructed from the data. This model can be expressed as a mathematical function which can be of type scalar, vector or tensor. If (as in our case) there is no associated function, and the aim is simply to visualise non-physical data, such a classification system does not apply.

Ken Brodlie [5] also provides a classification for scientific visualisation techniques based on mathematical models (of type scalar, vector or tensor). His classification, however, includes the additional case where there is no underlying function and the requirement is simply to visualise a set of points as a scatter plot. Assuming we had no uncertainty present in our data, two of Brodlie's generic techniques for visualising point data would apply:

- *One-Dimensional Scatter Plot*. An example could be a plot of the distance of other planets from the Earth. This technique could apply to artefacts in time, where the artefacts are marked as points on a single time axis.
- *Two-Dimensional Scatter Plot*. An example could be a plot of a set of individuals' heights and weights. This technique could apply to artefacts in space, where the artefacts are marked as precise points on a latitude/longitude map.

The problem with this approach is that we do not know the origins of most artefacts precisely (both in time and space). The use of straight-forward scatter plot maps is therefore not sufficient for our purposes, but could provide a starting point.

3.2.2 Information visualisation Framework

The process of visualisation involves mapping data to some visual form. Depending on the type of data, different visual forms will be appropriate. In this section, we provide a framework (from Card *et al* [10]) for describing visualisations of non-physical information. The examples of visualisation systems in Sections 3.2.3 and 3.2.4 are explained in terms of this framework.

Every visualisation can be described in terms of the following components:

- spatial substrate;
- marks;
- graphical properties of the marks.

Each of these is described below.

Spatial substrate

The space in which data is plotted can be described by the axes that bound the space. There are four basic types of axes (which can be linear or radial) [10]:

- unstructured axis (when there is no axis);
- nominal axis (when a region is divided into sub regions);
- ordinal axis (when there is a meaningful ordering of sub regions);
- quantitative axis (when the region has a metric).

The most appropriate form of axis to use would depend on the nature of the data being plotted.

Marks

The elements plotted in space are called marks. The four basic types of marks are [10]: points (0D), lines (1D), areas (2D) and volumes (3D). It is important to note that although points and lines are 0D and 1D, they still take up space in a visualisation.

Graphical properties of marks

Other than the position and dimension of marks, other graphical properties include colour, texture, shape, orientation, size, crispness, resolution, transparency and arrangement. Certain properties are more appropriate than others depending on the type of data being displayed. For example, the use of shape is a more suitable representation for differentiating different nominal values than the use of greyscale [32]. On the other hand, greyscale is very suitable for representing ordinal data, but shape is unsuitable.

3.2.3 Example 1: FilmFinder System

Figure 3.1 shows the FilmFinder system by Ahlberg and Shneiderman [1]. This visualisation can be described as follows:

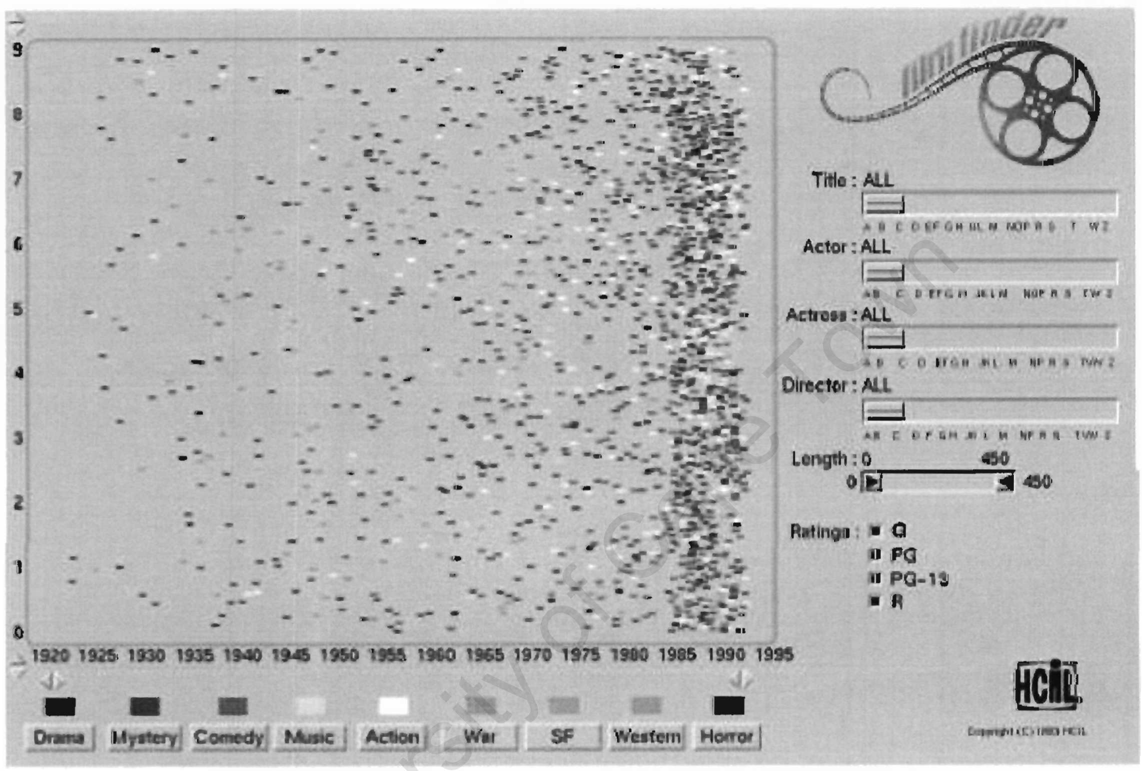


Figure 3.1: The FilmFinder system by Ahlberg and Shneiderman [1]

- **Spatial substrate:** The horizontal axis denotes time (the year in which the film was produced) and the vertical axis denotes the popularity of the film (a number between 0 and 9). Both of these axes are quantitative. If the type of film was plotted as one of the axes, this would have been an example of a nominal axis. On the other hand, if the first letter of the title of the film was used as one of the axes (sorted in alphabetical order), this would have been an example of an ordinal axis.
- **Marks:** Points are used to plot each film in the year/popularity space.
- **Graphical properties of marks:** Colour is used to indicate the type of film and a legend is shown at the bottom of the screen.

By interacting with the display (using sliders) a user can gain tremendous insight into the contents of a large film database and narrow down the dataset to the required sub-set of values.

3.2.4 Example 2: LifeLines System

LifeLines by Plaisant *et al* [38] is a general visualisation environment that can be applied to many types of biographical data. Figure 3.2 shows one example of the application of LifeLines visualising personal medical histories. The display can be described using the framework from Section 3.2.2 as follows:

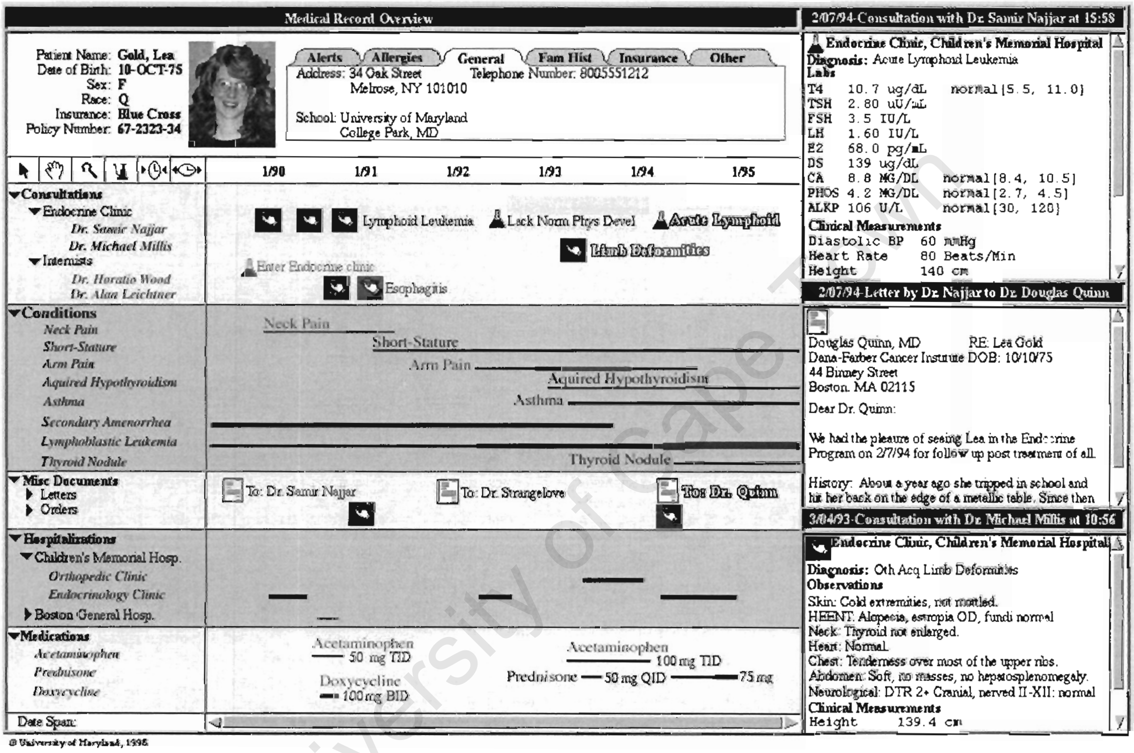


Figure 3.2: The LifeLines system by Plaisant *et al* [38]

- **Spatial substrate:** The horizontal axis shows time (over the period of the given patient's medical history) and the vertical axis shows different facets of medical history (such as consultations, conditions, medication). The time axis is quantitative and the medical history axis is nominal.

- **Marks:** The elements plotted in the time/medical history space are different, depending on the facet of medical history being displayed. Points are used to plot consultations and documents, whereas lines are used to plot conditions, periods of hospitalisation and medication.

- **Graphical properties of marks:** Different graphical techniques are also used for representing the marks in the plot area. For example, icons are used to indicate different types of consultations (e.g. a flask means that lab results are available). Colour is used to show related items in different categories (for example, a physician is allocated a colour which marks all relevant consultations, documents, etc.). Line thickness is used to indicate the severity of a condition.

In this way many different techniques can be combined into a single coherent system.

3.2.5 Example 3: Earth science data sets

Plaisant *et al* [39] developed an application to query earth science data shown in Figure 3.3 (Note: the system includes a query preview window which we will not discuss here). The underlying data in their application had a time attribute, denoting the period during which the particular earth science observation dataset was captured. We will be discussing the visualisation of these time ranges (the rectangle with lines displayed in the top left corner of the screen).

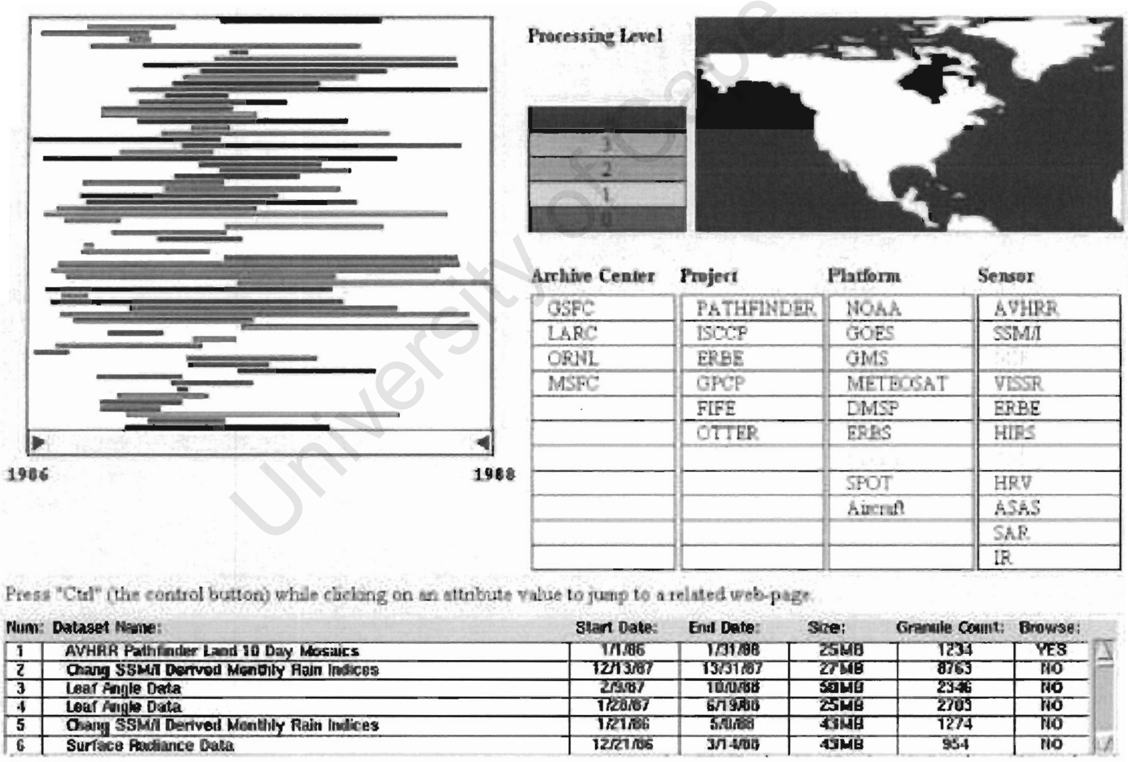


Figure 3.3: An application for querying earth science data (Plaisant *et al* [39])

- **Spatial substrate:** The horizontal axis denotes time, which is a quantitative axis. To avoid the problem of overlapping ranges, they used the vertical dimension for spacing out the time ranges. This vertical axis is therefore an example of an unstructured axis, since there is no real axis – the space is purely used to avoid overlapping.

- **Marks:** Every science dataset is represented by a line. Each line is selectable, making the display also function as a query tool. The set of time ranges displayed also changes depending on the geographic area selected on the map.

- **Graphical properties of marks:** The length of each line shows the length of time during which the data set was captured.

The link between information visualisation and visual querying is an integral part of the three example applications described above. The visualisations are not there simply to be viewed, but also to act as a basis for querying the data. In the next section we look at visual querying in general and what we can learn from what others have done in the field.

3.3 Visual querying

To query a database in a traditional way, users have to understand a number of things:

- They need to have some understanding of the structure of the database – which structures contain the relevant data?

- They need to understand how to use the database system and associated query language.

- They need some understanding of the content and size of data stored in the database. This is especially important when the database is being accessed across a network and download times are an issue, so that users can know to what level they need to narrow down their selection.

The use of visual interfaces can minimize these barriers of understanding between users and databases. In the following sections we discuss some examples of visual interfaces which can help in the formulation of queries. We also show how dynamic queries combine visualisation of the dataset with visual querying techniques. Finally we look at visual support for combining sub queries using Boolean operators.

3.3.1 Query formulation

Visual querying allows the user to express their query in a visual form which is then translated into a processable query statement. The aim is for the user to focus on the actual query rather than spending time formulating it.

Many different ways of expressing visual queries have been developed. A common way of supporting users in expressing queries is by providing a visual SQL builder, such as the one used by Stojanovic *et al* in [41]. With these kinds of tools, the user selects the tables(s) to query, then uses operator buttons (such as '<', 'AND', 'SUM', etc) to formulate the query.

Bretan *et al* [4] developed a visual query language based on the Entity Relationship model, where the user expresses a query using boxes (which contain pictorial representations of their entities) and arcs. Catarci and Santucci [13] also developed a system based on Entity Relationship diagrams, called *Query by Diagram*. Their system provides the user with a full representation of the Entity Relationship schema describing the database, which the user then interacts with to formulate a query.

With the above techniques, the user can query a database without needing to know the intricacies of a text-based query language (such as SQL) or how the database system works. A user, however, still requires some understanding of how the data is structured to know how to express the queries visually. The applications also do not provide the user with an understanding of the content and size of data stored in the database. Visual query techniques which combine information visualisation with visual querying can however hide the complexities of the database structure from the user while also providing an overview of the dataset. Dynamic queries is one such approach.

3.3.2 Dynamic queries

Ahlberg et al [2] were the first to define the concept of "dynamic queries". These allow users to formulate queries with graphical widgets, such as sliders. The goal of dynamic queries is that users see the results of query refinements as they make them. Dynamic queries depend on: presenting a visual overview, powerful filtering tools, continuous visual display of information, pointing rather than typing and rapid, incremental and reversible control of the query [40].

Two of the examples detailed in Section 3.2 use dynamic query techniques:

- **FilmFinder system** [1] (see Section 3.2.3): Figure 3.1 shows the FilmFinder system displaying all films in the dataset. The sliders on the right of the data display can be used to formulate queries and the results are shown on the data display. For example, Figure 3.4 shows how the Actor slider was used to select films featuring Sean Connery and the length slider was used to narrow down the selection to films of length 59 – 276 min. The display on the left immediately updates to show only those films that meet these requirements. To see the details of an individual film, the user selects one of the markers and a popup window gives the details (this is known as a location probe [10]).

- **Earth science datasets application** [39] (see Section 3.2.5): The system shown in Figure 3.3 uses dynamic queries in a different way from the FilmFinder system, but is based on the same principles (visual representation of the query and of the results; rapid, incremental and reversible control of the query; selection by pointing; immediate and continuous feedback). The temporal overview provides the display for the individual items in the dataset, where each item in the database is represented as a selectable line. Controls are provided so that the user can select values based on the attributes: Archive Centre, Project, Platform, Sensor and Processing level. By selecting values for these attributes, the dataset is queried and the number of lines displayed in the temporal view is updated accordingly. The results are displayed in the table on the bottom of the screen and on the

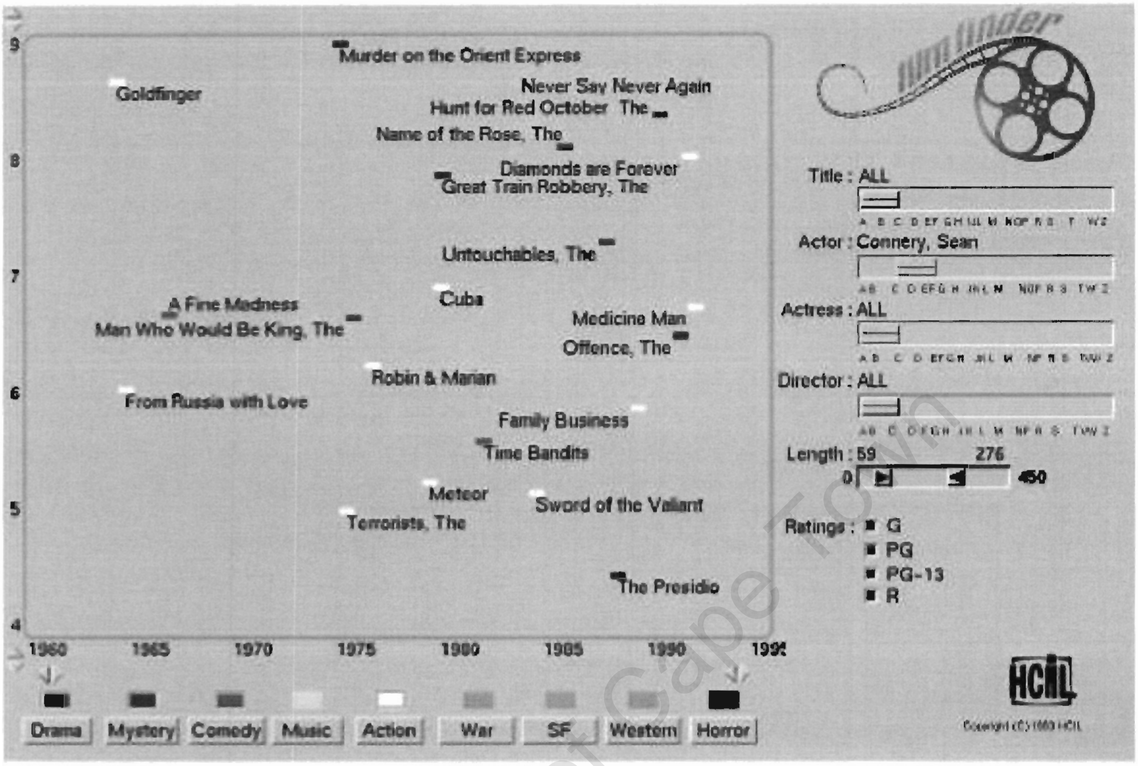


Figure 3.4: The FilmFinder system of Ahlberg *et al* [1], showing how the selection has been narrowed down using dynamic query techniques

timeline map. Figure 3.5 shows how a query was formulated by selecting two archive centres, three projects and two processing levels (selected attributes are highlighted in grey on the printout). The user probed a particular item and the details are shown in a popup window. More filtering can be done by zooming in on the timeline or on the map. In this way the timeline and map also act as query tools.

There have been many other applications that employ dynamic queries and the benefits of using this approach have been well established [2, 15, 37, 40, 46]. Ahlberg et al [2] experimented with different types of interfaces, one of them being a dynamic query interface. Although qualitative results did not show that the use of dynamic queries resulted in superior user performance, subjective evaluation clearly indicated that users preferred using dynamic queries. Shneiderman [40] shows that the use of dynamic queries results in significant improvements in user performance and user enthusiasm, in three separate applications. Several experiments have demonstrated that dynamic queries also help users find trends and exception conditions [40, 37, 46].

In summary, some of the benefits of dynamic queries over traditional query systems are [46]: it is quicker to express queries; novices learn to use the system quickly; the content of the

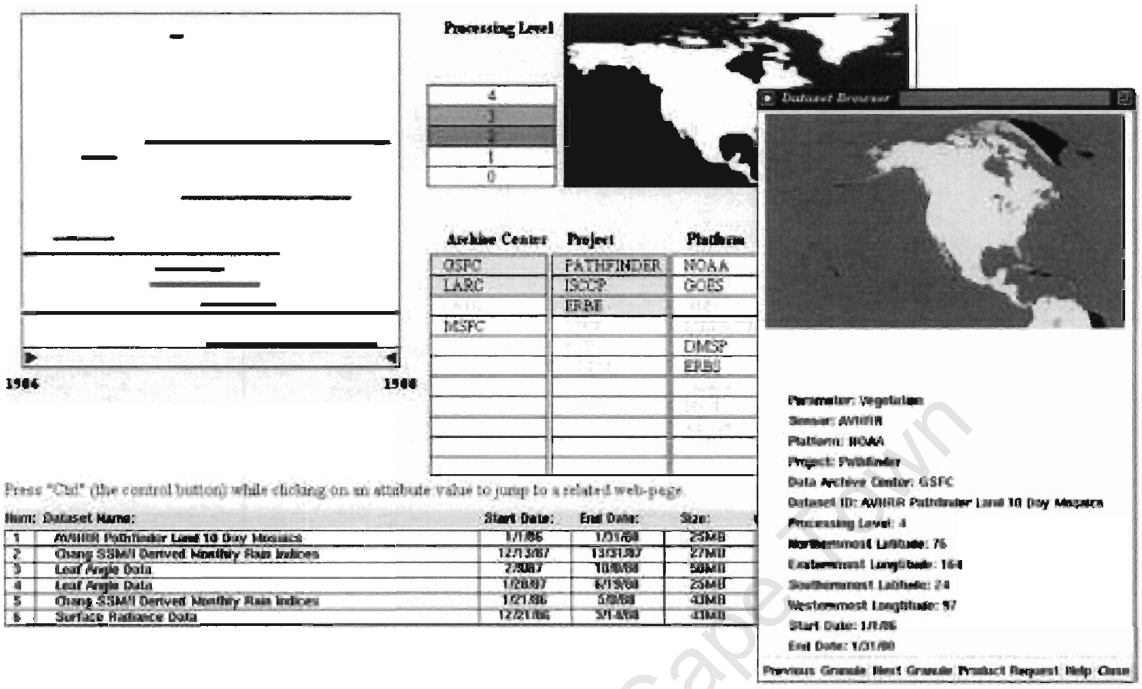


Figure 3.5: An application for querying earth science data (Plaisant *et al* [39]). This screen shows how a query is formulated using the attribute controls.

database and trends are easily inferred. Drawbacks of using dynamic queries include [46]: the data needs to have some form of ordering and it is difficult to express more complex Boolean queries.

In Section 3.5 we discuss how dynamic querying techniques could be applied to the CAMA dataset.

3.3.3 Combination queries

One of the challenges of using dynamic queries is how to express combinations of sub queries in different ways so that the user is not limited by the interface. Expressing Boolean combinations of visual queries is recognized as an area needing further work, e.g. Ahlberg et al [2] highlight as an area for further research: "the ability to select Boolean combinations of sliders". Applications have been developed which use dynamic queries combined with Boolean operators, but these have had limited Boolean functionality.

In the FilmFinder system (Figure 3.4) all the query sliders work together using intersection ('AND'). It is not possible, for example, to display films which have either the actor Sean Connery or are directed by Steven Spielberg (actor and director are two different sliders). It is also not possible to express 'OR' relationships between values on the same slider (for example, films starring either Sean Connery or Clint Eastwood). The earth science data application

(Figure 3.5) is slightly more expressive in that it uses the OR operation within attribute value groups (e.g. the user expresses archive centre "GSFC" or "LARC" by selecting them both) and the AND operation between attribute value groups (e.g. archive centre "GSFC" and processing level 4).

Shneiderman [40] suggests a way of using Boolean operators with dynamic queries, by placing filter widgets on a flow diagram. This filter/flow technique was evaluated, to show that subjects were able to solve a Boolean query more accurately using a filter/flow interface, than when using SQL [47].

Others have suggested ways of supporting the user in expressing combinations of queries, which have not been combined with dynamic queries. Jones [27, 28] implemented a visual representation for query specification based on Venn diagrams, as an alternative to a text-based query interface. This provided an effective way for users to specify Boolean combinations of queries, but was limited in that it did not support selection of the universal set (all items in the selection). Mirel [34] investigated the kinds of problems that users encounter in expressing complex combination queries for business problems. She proposed a toolbar of options for selecting data, which included replace (=), add to (+), subtract from (-) and intersect (&). She claims that with this toolbar, users' performance improved dramatically.

Combining dynamic queries with such a toolbar for performing combination queries, could present a powerful solution to combining both expressiveness and ease of use.

3.4 Uncertainty in data

In this section we look in more detail at the issue of uncertainty in data: why it is relevant, the types of uncertainty that exist and how it can be visualised.

3.4.1 The relevance of uncertainty

Uncertainty is evident in many forms in data and it presents challenges to those working with the data, as expressed by Gershon [19]:

> *Life is not perfect. We confront imperfection (and uncertainty) every day. Similarly, data and information as well as their representations will never be perfect. We thus need to get accustomed to and make peace with this fact, not expect decision making based on perfect data, information and presentation. In that regard, we need to develop principles and methods of imperfection ("uncertainty") management - how to get and understand imperfect information using imperfect representations and reach sound decisions in real-world conditions.*

In the CAMA archive, there are different degrees of uncertainty in the data. The uncertainty originates from incomplete information. For example, an artwork could originate in "late 19th Century" from "Ivory Coast/Ghana". If we knew exactly when it was made and where it came from, there would be no uncertainty. We need to develop ways of effectively working with this 'incompleteness'.

3.4.2 Types of uncertainty

To effectively handle the uncertain data in the CAMA archive, we need to understand more about the nature of the uncertainty. There are different sources of uncertainty in data. Three common types of uncertainty include [36]:

- *statistical uncertainty*: either given in the form of a calculated confidence interval or an actual distribution of the data;
- *error*: a difference among estimates or between a known value and an estimate;
- *range of uncertainty*: an interval in which the data value must exist.

Our form of uncertainty is clearly a *range of uncertainty* for both time and space.

Uncertainty is often quantified as a value from 0..1 and is either probabilistic or fuzzy in nature. In literature, there is some confusion around the distinction between fuzziness and probability. Burrough [6] claims that fuzziness is not a probabilistic attribute, but rather 'an admission of the possibility that an individual is part of a set'. Fisher [17] also claims that the two terms are conceptually different, and that the confusion in the use of the terms can probably be attributed to the fact that both fuzzy sets and boolean sets (probability) work in the domains of 0..1. He claims that probability is the likelihood of an object belonging to a set or not.

To explain the difference between probability and fuzziness, Fisher uses an example of the viewshed function, used for analysing surface elevation in Geographic Information Systems (GIS). For any given location on a geographic surface, other points on the surface are either visible from that location or not (i.e. is there a direct line of sight to the location or not). This viewshed function can be represented as a binary map, where 1 indicates that a particular point is visible and 0 that it is not visible. By definition, the viewshed is Boolean, since only two alternatives exist: any particular location on the map is either visible or not visible. There may be some doubt as to whether a point is in reality visible or not and this doubt is then expressed as a probability in the range 0..1.

In contrast, consider the scenario where what was being modelled was not whether a location is visible or not, but the degree to which something can actually be seen. The degree to which an object is visible is dependent on many factors such as atmospheric factors, eyesight, object-background contrast, etc. This alternative would be a form of fuzzy uncertainty as opposed to a probabilistic uncertainty, since in reality it is not certain whether it is visible or not.

In our case, if complete data were available on the origins of cultural objects in Africa, there would be no uncertainty in the data. Using Fisher's distinction above, the uncertainty is therefore most suitably modelled using *probability* rather than fuzzy sets. By nature, each object has a definite origin in time and space, but there is doubt as to what this exact origin is. This doubt can be expressed as a probability of an artefact originating in any given country or in any given year.

For example, consider a date range of 1650 – 1700. If the nature of the uncertainty was fuzzy, this could mean that there is some form of normal distribution of certainty, with an artefact more likely dating from 1675 than other dates within the range. If the nature of the uncertainty was probabilistic, however, any date within this range could have an equal probability. In the case of African art, an artefact with the origin "19th Century" (artefacts with this sort of range are very common) is equally likely to have originated in 1800 as 1899.

In the area of GIS, studies of uncertainty in data have mainly focussed on areas which have fuzzy boundaries [6, 14, 17, 43]. Our study is however focussed on uncertain point data, rather than areas which have vague boundaries.

3.4.3 Visualising uncertain data

The aim in visualising uncertainty in data is to ensure that users are made aware of the presence and degree of uncertainty so that they are able to make more informed decisions. In our study, we could have chosen to ignore the uncertainty. To produce a picture of artefacts in time, we would have had to allocate every artefact to a precise time. An artefact originating from, say, 'late 19th century' could be given a precise value of 1875. The resulting picture would give the user the impression that these are precise dates. Users would also lose the information of how the origins of artefacts could have overlapped.

Techniques for representing uncertainty

Gershon [19] suggests that the degree of imperfection of the information on an object is merely associated information, which could be represented intrinsically or extrinsically. Intrinsic representations of imperfection could be achieved through the use of traditional visual variables (position, size, brightness, texture, colour, orientation and shape), as well as boundary (thickness, texture and colour), blur, transparency, animation and extra dimensionality. Extrinsic representation of imperfection could be achieved through objects close to the real objects, and he suggests the use of objects such as question marks, dials, thermometers, arrows, bars, objects of different shapes and complex objects (such as pie charts). Some of these ideas of intrinsic and extrinsic representations could be relevant to our study, but the effectiveness would have to be confirmed.

A number of studies have been done in visualising uncertainty in scientific applications. For example, the use of uncertainty glyphs, animation and envelopes for visualising uncertainty in fluid flow [30]; the use of animation for visualising uncertainty in sea surface temperature data [20]; and the use of sonification for visualising geometric and fluid flow uncertainty [31]. A whole range of uncertainty visualisation approaches are discussed by Pang *et.al.* in [36]. The application areas covered include radiosity (comparing results from algorithms), animation (comparing human vs. modelled motion data), interpolation (the impact of parameters on algorithms) and flow visualisation (the effect of different integration methods and parameters).

The type of uncertainty in the scientific applications described above is more fuzzy in nature than probabilistic. In all the cases above, specific values are available and these values are compared with other values to show the variation (i.e. the level of uncertainty or error). These values are based on some underlying model that describe the set of points. In our case, we frequently do not have actual values to plot, but rather a range within which the values could occur, so the uncertainty is probabilistic in nature, rather than fuzzy.

3.5 Summary

In this chapter we looked at categories of data and techniques for visualising data. We then presented a framework for describing any information visualisation and we described three example applications in terms of this framework. We looked at visual techniques for supporting users in querying databases and how visual querying can be combined with information visualisation in the form of dynamic queries. Finally, we looked at the issue of data uncertainty and how others have managed uncertainty in data. To end this chapter we summarize our thoughts on how this related work could apply to our scenario.

3.5.1 Relevance to CAMA

One of the requirements of our application (see Section 2.3.2) was that the user gain insight into the volume and area of coverage of the data in the archive. Two important themes along which the data would be explored are time and space. In designing visualisations, we would need to decide on appropriate spatial substrates, marks and graphical properties. Both time and space are quantitative in nature so would be suitable for mappings to quantitative axes. Each artefact could be mapped to a point in an appropriate time or space map. An alternative could be to use a line for each artefact to display a range of uncertainty.

To support querying, a dynamic query approach could be particularly suitable for novice users. However, on their own, dynamic query tools are not expressive enough to formulate combinations of queries such as:

> Find items that occur in the period 1000BC to 1000AD in West Africa, excluding Ivory Coast.

Some visual support for Boolean combinations of queries will be needed (such as the toolbar suggested by Mirel [34]).

To manage the uncertainty in our data some of the broader ideas of using intrinsic and extrinsic representations of uncertainty suggested by Gershon [19] could be applied to our data. Intrinsic representation will probably be more appropriate, because extrinsic representations are dependent on the real object being displayed. Displaying the actual object without a built-in indicator of uncertainty could be misleading since we have such a wide range or uncertainty and these ranges are often very large (e.g. in one case the geographic origin is known down to village precision, whereas in another case, the origin could be anywhere within a set of a countries).

In the following chapter we investigate these ideas further.

Chapter 4

Techniques

Our aim was to develop effective techniques for visualising and querying African cultural data. Dynamic queries (discussed in Section 3.3.2) seem to present an attractive solution by combining the display and query mechanisms in a way that is intuitive for novice users. In this chapter we present proposed techniques for displaying and querying uncertain time and space data. We also propose ways of storing the data to support the visualisation and query mechanisms.

4.1 Display of uncertain data

In this section we look at techniques for visualising both one-dimensional and two-dimensional data. As discussed in Section 2.3.2, there are a number requirements for such a visualisation:

- The user interface should be accessible to novice users and use metaphors familiar to such users.
- Users would want to explore themes and influences, so the relationships between data elements should be displayed along particular themes.
- Uncertain items should be treated with the same level of importance as items which have definite origins.
- The application should provide an overview of the data (data volumes, areas of coverage, etc.).

Two important themes along which data is explored is time and space (geographic location). In the following two sections, we look at techniques for displaying the data on both time (one-dimensional) and space (two-dimensional) maps.

4.1.1 Display of 1-D time data

In African culture, the *origin* of cultural artefacts is an important theme of study. Researchers want to see *when* artefacts were made and to see this in relation to other artefacts. A time-line is a familiar metaphor used frequently in historical studies and so would be suitable as the basis of a visualisation for novice users. Displaying the artefacts on a time-line would provide a picture of the relative positions of artefacts in time and would assist researchers in exploring the influence of time on culture. Such a 'time map' could also be valuable in providing an overview of the data in the archive, both in terms of seeing how much data there is and in seeing which periods have greater coverage of data.

As explained before, the archive contains a mixture of certain and uncertain data. Even though some artefacts may have uncertain origins (such as the 'Anubis mask' with the origin '6th - 4th century BC or later'), these items should be treated with the same level of importance as artefacts with definite origins (such as the artwork 'Illusions of Permanence', created by Bonita Alice in July, 1998). We therefore need some novel way of displaying both certain and uncertain items simultaneously on the same time-line.

The approach we propose for displaying the time data is illustrated in Figure 4.1. This visualisation can be described using the framework detailed in Section 3.2.2 as follows:

- **Spatial substrate:** The horizontal axis denotes time which is a quantitative axis (a time-line is shown at the top of the display). The vertical axis represents uncertainty with the axis going from least uncertain (top) to most uncertain (bottom). We choose to use an ordinal axis, where there is a meaningful ordering, but there is no metric.

- **Marks:** Each item in the archive is represented as a bar on the display. The certain items are plotted close to the time-line, whereas the more uncertain items are plotted further away from the time-line.

- **Graphical properties of marks:** The length of each bar is used to indicate the level of uncertainty. Items which are certain are depicted as points, whereas items which are less certain are depicted as longer bars, depending on the range of uncertainty. This is an example of an intrinsic representation of uncertainty (see Section 3.4.3), since the property of the mark itself is used to show the level of uncertainty.

At a glance the user can gain insight into the data stored in the archive. It is easy to spot prolific periods (or periods for which much data was collected) as these will exist as vertical clusters. It is also relatively easy to compare relative amounts of uncertainty by comparing the length of the bar for each individual artefact.

This representation of uncertain time ranges is superficially similar to a representation used by Plaisant *et. al.* [39], where the authors employed dynamic queries on earth science data (see Section 3.2.5). In their application, data items in the form of time periods were plotted below a time-line. Although the two displays look similar, they are fundamentally different in what they encode. The underlying data visualised in Figure 3.3 has a time attribute, which is a real time period (denoting the period during which the particular earth science observation dataset was captured). Our time ranges on the other hand, denote a period of uncertainty, within which the item *could* occur. Our dates are single events, whose occurrences fall within a range of uncertainty, as opposed to dates that are fundamentally extensive ranges.

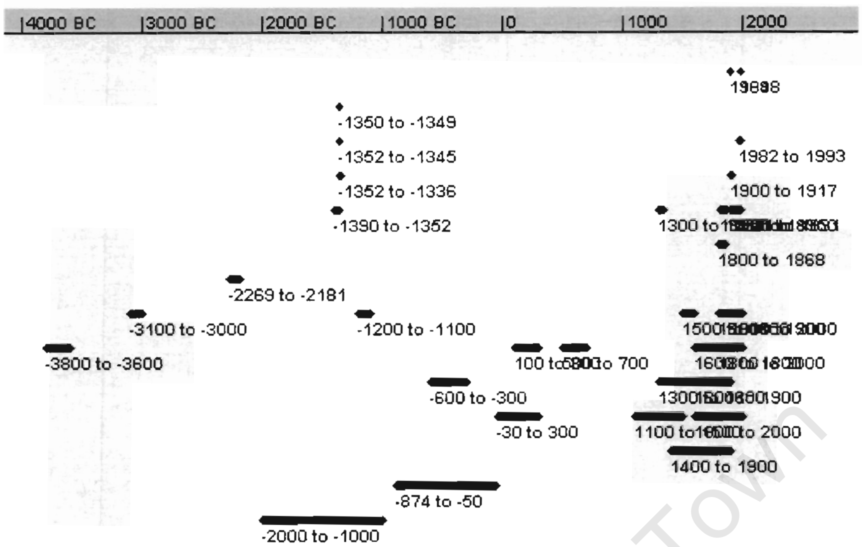


Figure 4.1: Display of uncertainty ranges on a time map. Each item in the archive is represented as a bar. More uncertain items are plotted further away from the time-line.

4.1.2 Display of 2-D uncertain data

In the same way that time is an important attribute of any cultural artefact, the geographic origin of an artefact is also an important theme of study. Researchers want to see *where* artefacts come from and to study the influence of space on culture. The obvious metaphor for a visualisation of the geographic origins of African cultural artefacts is a map of Africa. Displaying the artefacts on a map would show where individual artefacts originated and could support researchers in investigating the influence of space on culture. A map of Africa showing all the artefacts currently in the archive would also provide a useful geographic overview of the data – indicating the countries which are well represented in the archive and highlighting the countries for which the archive has little data.

Once again, the problem we face is that the archive contains a mixture of certain and uncertain data. This is also evident in the geographic origins of artefacts. For example, one sculpture of a figure of a man (currently held in Lagos Museum) is known to originate from Tada village, Kwara State in Nigeria, whereas another sculpture of a male figure (currently held in a private collection in Paris) is known to originate from somewhere in the area of Burkina Faso/North Eastern Ivory Coast. These two examples show that the origin of some artefacts is known down to village precision, whereas others are not even known to definitely originate from a single country. This makes it difficult to place all the items on the same map.

We propose an approach that uses a dot density map for displaying an overview of the data on a map of Africa. Individual artefacts can then be viewed on a zoomed in view of each country. This approach is explained in the following sections.

4.1.3 Dot density overview map of Africa

If the map is viewed at a low resolution (i.e. zoomed out to the whole of Africa), the distribution of the cultural artefacts is shown as a dot density map. An example of a dot density map is shown in Figure 4.2 (note: this is a hypothetical example and is not based on data from the CAMA archive). The map shows the relative number of items in each country, where one dot is used to represent two items in the archive. If, for example, 20 cultural artefacts originate from country A, then 10 dots will be displayed in random positions on the map within the boundaries of country A. The ratio of dots to artefacts can change depending on the number of artefacts in the archive. The map in Figure 4.2 shows, for example, that the (hypothetical) archive contains more artefacts originating from Egypt than Madagascar and that the archive seems to contain no data on Angola.

The spatial substrate of this visualisation differs from the other visualisations described so far in that the space does not have a vertical or horizontal axis. The space is not used in the traditional geographic way (with latitude and longitude axes, which are both quantitative), but is rather defined in terms of the boundaries of countries. There is no meaningful ordering, but as one moves around the space, one is either inside a particular country or not. The 'axes' are therefore the boundaries of countries.

Figure 4.2: Dot density map (1 dot = 2 items)

To display the data as a dot density map, we need a way of calculating the number of cultural artefacts originating from each country. This is straightforward when artefacts have definite origins, but not when artefacts have uncertain origins. To handle this uncertainty, we propose an approach that assigns probabilities to artefacts falling into countries. This process of calculating probabilities is illustrated in Figure 4.3. The following examples illustrate how the process works:

- Say the origin of an artefact is "Bulawayo" (a city in Zimbabwe). In this case, the artefact has only one place of origin, which is a point. A query will then be issued to find the enclosing country (Zimbabwe). Since there is only one country of origin, the probability of the artefact falling into Zimbabwe is 1. The process will be similar for provinces or kingdoms.
- Say the origin of an artefact is "Angola/Zaire". In this case, the artefact has two possible places of origin, which are both countries. The probability of being in either country will shared, giving 0.5 in each case.
- Say the origin of an artefact is "West Africa". In this case, the artefact has a region as the origin. A query will find all countries defined as being in the sub-region "West Africa". The probability of the item being in each of the countries of the sub-region will be shared equally between the countries.

To determine the number of artefacts in a country, we add the probabilities of all items which may originate from that country.

We believe that a dot density overview map, such as the one we are proposing, will provide valuable insight into the geographic spread of data in the archive. However, on its own such a map does not provide any detail on individual artefacts in the archive. We therefore propose an additional country-level view of the map which will plot individual artefacts in a geographic context.

4.1.4 Individual detail country-level view

We propose that the user be able to click on any country on the overview map and a zoomed-in view of the country will be displayed. Within the borders of the country, icons will be displayed representing each artefact in the archive which originates from that country. If there is sufficient screen space, thumbnail images of the particular artefact will be displayed.

Items which have definite point origins (such as a village) can be placed at the point of origin. Where there are multiple items originating from the same point, a suitable representation will be used to indicate this.

In the case of items which have imprecise origins, but which are known to originate from one country only, we propose placing representations of the items within the boundaries of the country using a suitable spread. We plan to experiment with different placement routines to find which technique results in a representation which is attractive, while at the same time not misleading to the user (it is important that users do not assume that the origin of the artefact is known precisely when it is not the case).

In the case of an item which has a shared probability of originating from different countries (e.g. an item with origin "Zimbabwe/Zambia" or "West Africa"), we propose placing a representation of the item in all possible countries of origin. Since the map will contain a mixture

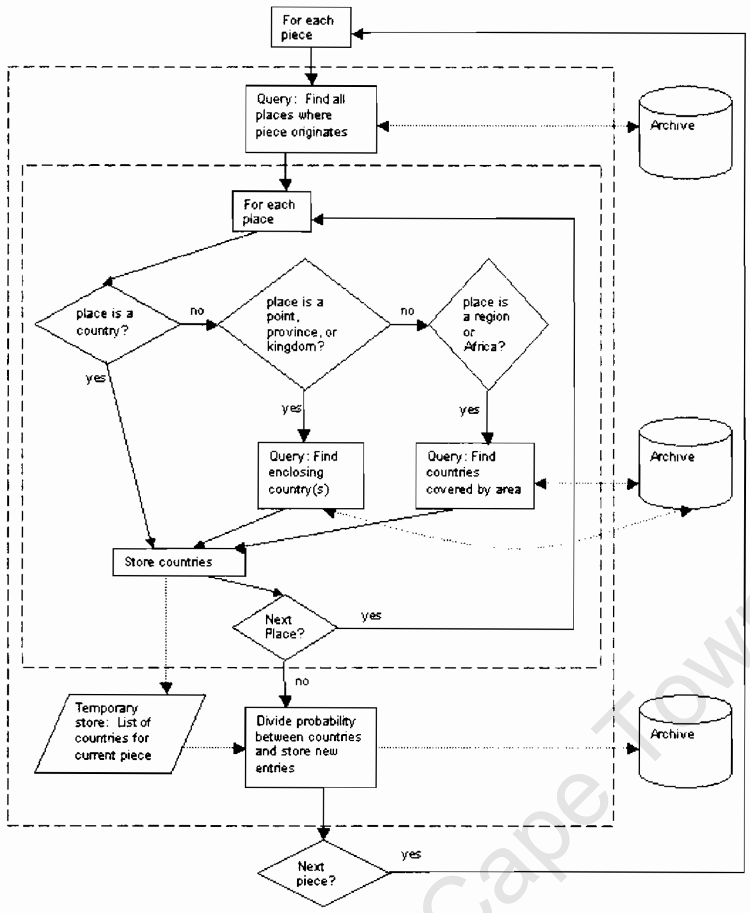


Figure 4.3: Process of calculating probabilities of artefacts (called pieces on the diagram) falling into countries

of representations, i.e. items with certain and uncertain origins, we will need some way of indicating the level of uncertainty.

4.2 Querying uncertain data

Given the proposed techniques for displaying data, we need mechanisms for querying the data, so that the user is able to select a subset of the data. These query mechanisms would need to be visual in nature, rather than the usual SQL-based mechanisms.

In this section we propose techniques for querying the data based on the 1D time attribute as well as the 2D space attribute. We also propose a way of performing combinations of queries based on multiple attributes.

4.2.1 Querying 1D data

When querying the archive, users would typically be interested in selecting items within a particular time range, such as "the 18th century" or "1650-1700". To support this form of querying in a visual way, we propose the use of double sliders on the time-line of the time map (Figure 4.1). These double sliders will allow the user to select items within a particular time range. Feedback will be given in the form of vertical dashed lines (see Figure 4.4).

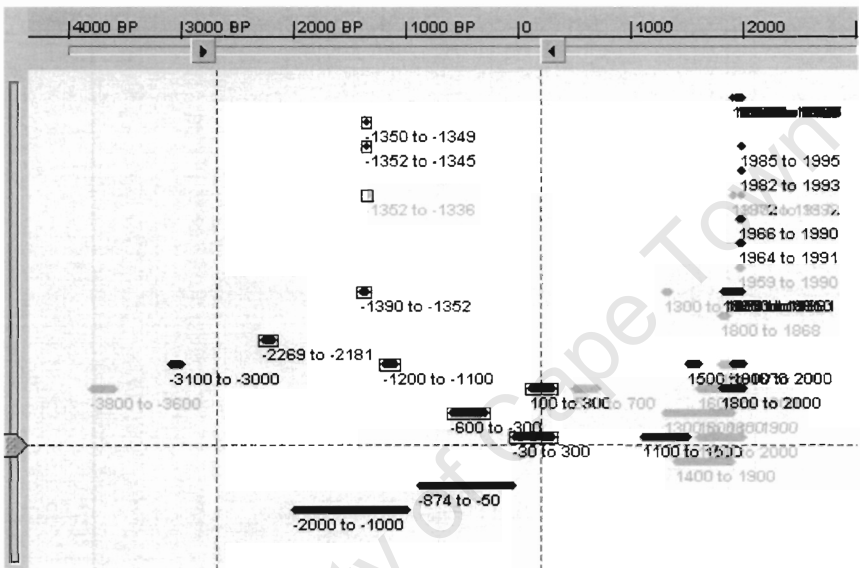


Figure 4.4: Double sliders on the time-line allowing the user to select items within a particular time range.

Since many of the items have uncertain origins in time, an interesting question arises as to whether an uncertain item should be selected or not when the range does not fall completely within the boundary of the query. Figure 4.5 shows the different possibilities:

- an item may fall completely inside the range (e.g. item 1), indicating that it definitely originated from within the range of the query;
- an item may fall completely outside the range (e.g. item 2), indicating that it definitely originated from outside the range of the query;
- an item may fall partly inside the range (e.g. items 3 and 4), indicating that the item *could possibly* have originated from within the range of the query.

Since we would like our query tool to return both certain and uncertain elements, we propose that items such as items 3 and 4 in Figure 4.5 be included, by default, in the result of the query. Therefore, items with uncertainty ranges which 'touch' either of the vertical dashed lines of the

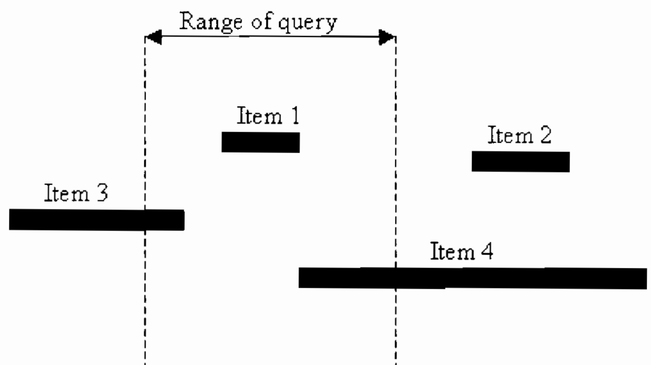


Figure 4.5: Given the range of a query, an item may fall completely inside the range, completely outside the range or may fall partly inside the range.

time slider will be selected. We also propose the use of a horizontal slider to allow the user to exclude the more uncertain elements (see Figure 4.4) from the query.

A similar technique of using double sliders for querying based on time is presented by Plaisant et. al. [39]. However, since the nature of the data is different (a real time period in [39], as opposed to an uncertainty range in our case) the ways they can be queried differs. Although it is appropriate in our case for the user to filter out the more uncertain items (those with the "longer" range), this would not be appropriate in the case of real time periods.

4.2.2 Querying 2D data

Given the visualisation strategy described in Section 4.1.2, the user would need a way of selecting groups of items based on geographic location. We propose two approaches, one based on the dot density overview map and one based on the individual detail country-level view.

In the case of the dot density overview map, we propose that the user be able to select a subset of items based on a selected country or sub region (group of countries). This will be done by choosing a mode and clicking on the map. For example, if the user chooses a mode of selecting items based on country and clicks on the map inside Mali, then all items originating from Mali will be selected. The user will also be able to indicate whether the system should include uncertain items in the selection or not (i.e. items which could have originated from Mali).

In the case of the zoomed-in country level view map, the user will see representations of individual elements. We propose that in this view the user be able to select individual elements or groups of elements using a rubber-band rectangle or circle tool.

4.2.3 Combinations of queries

Given the different query tools proposed above, we will need to provide mechanisms for the user to express combinations of these queries. For example, say the user wanted to formulate the following query:

> Find items that occur in the period 1000BC to 1000AD in West Africa, excluding Ivory Coast.

This query requires the use of the intersection and difference Boolean operators.

To support the user in expressing the different ways of combining sub-queries into a single query, we propose using Mirel's toolbar [34] (as discussed in Section 3.3.3). In addition, we introduce the notion of 'tagging' elements before they are selected. Through manipulation of the visual query tools (such as the double time sliders), items on the data display are 'tagged'. The Boolean operators are then used for combining the current selection with the tagged items in different ways. This is illustrated in Figure 4.6. Figure 4.6(a) shows a display of 11 elements where four of the elements have been tagged (surrounded by rectangles). If the '=' button is pressed, the selection is replaced by the tagged elements as shown in the updated display in Figure 4.6(b). Four elements are now selected (coloured grey). Through a new query process, five elements are tagged (see Figure 4.6(c)). Pressing the intersect ('&') button will then update the selection to include those elements which were both tagged and selected, resulting in a selection of two elements (see Figure 4.6(d)).

The idea is therefore that the user tags items using the query tools and then adds these to (or subtracts from, replaces, intersects these with) the current selection using the Boolean operator buttons. Feedback on which items are selected and which are tagged is provided through cues. In addition, the number of items tagged and the number of items selected will be fed back to the user.

4.3 Data model to store uncertainty

Finally, we required a model of the data to support the visualisation and query needs of the user. A simplified version of the entity-relationship diagram for the data model is given in Figure 4.7. The overall guiding principle when modelling the data was to store as much detail as was available. Later, this could be aggregated to less detail, as required. Each artefact has a time associated with it. The date qualifier is a textual description of the time, whereas the from and to fields denote time as integers. In this way, we store the time as a range, where the range indicates the minimal time period in which the origin is known to fall. We call this the range of uncertainty. For example, the date of an artefact could be: "18th Dynasty, Amarna period". This means that we do not know the precise year of origin, but we know that it originated at some time in that period.

In order to do comparative queries on the time attribute, it has to be expressed numerically, so the from and to fields associated with this period would be -1352 and -1345, respectively, while the date qualifier would contain the original string. When the date is precise (i.e., known

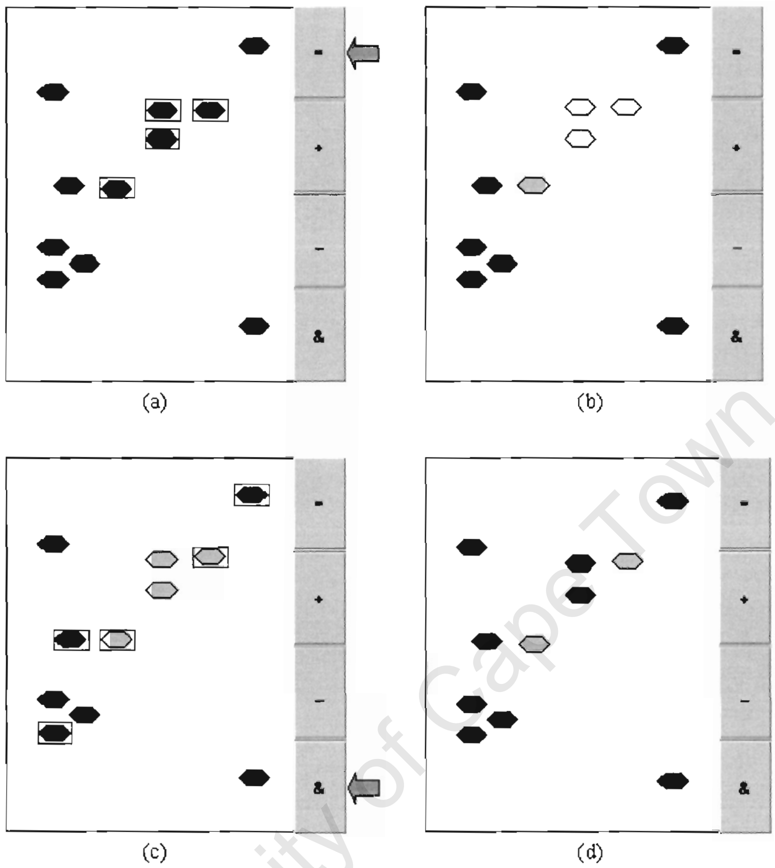


Figure 4.6: Diagram (a) shows four tagged elements. Diagram (b) shows the four elements selected as a result of pressing the replace button. Diagram (d) shows two elements selected after intersecting the selection in (b) with the tagged elements of (c).

to within a year), the from and to fields store the same date. For example: an artwork which was made in 1985 would have 1985 stored in both fields.

Each artefact could come from a number of places, where place is either a point, a province, a country, a region or Africa as a whole. Place, in this context, refers to the origin of the artefact rather than where it is currently kept. For example, an artefact originally from Ethiopia could currently be held in the British Museum, London. Where it comes from is a more important attribute for relative querying than where it is currently being kept.

The relationships between these sub-places are also stored. For instance, a point, Tada Village, falls into a province, Kwara State, which falls into the country Nigeria, which falls in the region, West Africa, which falls into Africa. In this way, if the place of an artefact is Tada Village, a query on West Africa would also return this artefact. Some artefacts originate from a number of places (usually because the original location is uncertain). For example, an artefact could come from Ghana/Ivory Coast. For these cases, we calculate and store a probability for

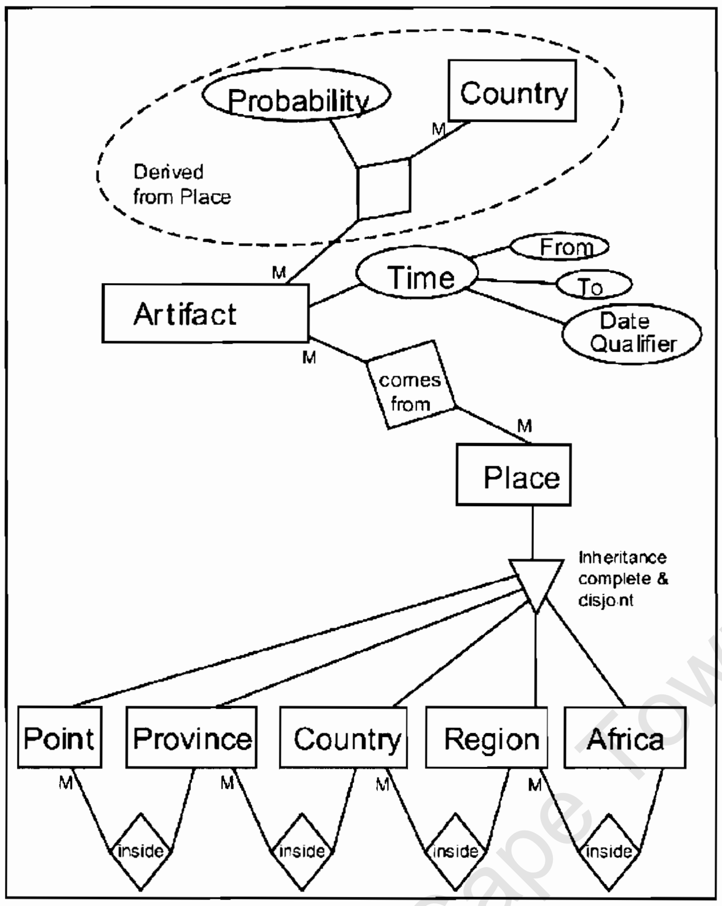


Figure 4.7: A simplified version of the entity-relationship diagram for the CAMA archive.

the artefact and each country concerned. The probability of an item falling into a country is a function of the number of countries within which that item could possibly fall.

4.4 Summary

In this chapter we proposed the following:

- A visualisation for displaying items based on their time attribute (see Figure 4.1).
- Two visualisations for displaying items based on their geographic origin: a dot density map for low resolution (see Figure 4.2) and a more detailed representation of individual artefacts at country-level resolution.
- A query mechanism based on the time map which uses double sliders to select based on time and a single slider for excluding uncertain elements (see Figure 4.4) as well as a

mechanism for querying based on geographic origin.

- A mechanism for tagging elements using query tools and then selecting using Boolean operator buttons.
- A data model which accommodates uncertain data by storing as much detail as possible.

In the next chapter we evaluate our proposed visualisations and some of the proposed visual querying techniques.

Chapter 5

Evaluation of techniques

Our aim was to develop effective techniques for visualising and querying African cultural data. In this chapter we look at the effectiveness of some of the techniques proposed in the previous chapter. We start by providing an overview of existing approaches to evaluation of software. We then describe the evaluation of two of our proposed visualisations: the display based on the time origin attribute and the display based on the geographic origin attribute. Based on these results we developed a prototype for visual querying of the data. In Section 5.4 we describe how we tested this prototype.

5.1 Approaches to evaluation

In this section we look at some of the theory and related work on evaluation of software. We start by looking at usability evaluation, an area which has been well researched by the Human Computer Interaction community. In Section 5.1.2 we look at some of the issues specific to the evaluation of visualisation software.

5.1.1 Usability evaluation

An important part of evaluating software involves evaluating the usability: the extent to which the software system enables users to achieve specific goals effectively and efficiently, while promoting feelings of satisfaction [26]. There are a wide range of usability evaluation techniques which are applicable at different stages and in different contexts. Several taxonomies of usability evaluation methods have been proposed.

Evaluation method classes

Ivory & Hearst [26] classified usability evaluation methods using five broad categories (which they called *method classes*):

- *Testing*: This form of evaluation occurs when an evaluator observes users interacting with a user interface. Evaluation methods which would fall into this class include thinking-aloud protocol, shadowing method, log file analysis.
- *Inspection*: Inspection evaluation covers methods where an evaluator uses a set of criteria (or heuristics) to identify potential problems. Cognitive walkthrough and heuristic evaluation are examples which fall into this category.
- *Inquiry*: This class covers evaluation methods where users provide feedback via interviews, surveys, questionnaires, etc.
- *Analytical Modelling*: With analytical modelling, an evaluator employs models to generate usability predictions. Example methods include GOMS analysis and UIDE analysis.
- *Simulation*: This class covers evaluation methods where an evaluator employs models to mimic user interactions. Petri Net modelling and genetic algorithm modelling would fit into this category.

To evaluate the proposed displays of ID time data and 2D space data, we performed an initial inquiry-based evaluation. We used questionnaires and interviews to obtain feedback from users (described in Sections 5.2 and 5.3). We followed these initial evaluations with testing using a prototype. A form of pluralistic walkthrough was used where an evaluator observed users working in pairs while they completed predefined tasks (described in Section 5.4).

Evaluation strategies

Fitzpatick [18] proposes a different taxonomy of usability evaluation methods based on the appropriateness of methods. He defines the following four strategies to usability evaluation:

- *Real World*: This strategy is used when real users and real computers are available.
- *Virtual Engineering*: This applies to the case when both the user and computer are representational.
- *Soft Modelling*: involves real users and representational computers.
- *Hard Review*: involves representational users with real computers and software product.

The same evaluation method can be employed using different strategies. For example, observation or questionnaire methods could be used with either a Real World or Soft Modelling strategy; heuristic methods could be used with either a Hard Review or Virtual Engineering strategy.

Fitzpatrick argues that the appropriate use of a strategy will depend on the phase of the software development lifecycle. On the two extremes of the software lifecycle, a Virtual Engineering strategy is appropriate earlier on at the design phase and a Real World strategy at the final installation phase. Soft Modelling and Hard Review would be appropriate somewhere in between.

Our initial evaluations were in the form of Soft Modelling where we involved real users with representational computers (questionnaires with pictures of visualisations and interviews with computer-generated visualisations). Later, we performed a hard review evaluation with representational users interacting with a prototype on computer.

5.1.2 Evaluation of visualisation software

One of the main aims of information visualisation is to support the user in making sense of the data [11]. To measure whether a visualisation achieves this aim involves two checks:

- Is the visualisation valid?
- Does it support the user in making sense of the data? In other words, does it result in more understanding.

These are discussed below.

Validity

Validation involves determining whether the visualisation accurately represents the data. Experiments with human subjects are essential to validate visualisation techniques [21]. Validity is an extremely important issue in our study. We are proposing visualisations of uncertain data which could potentially be misleading. We need to ensure that the presence of uncertainty is understood by users and not interpreted as something else.

Sense making

How can we evaluate whether a visualisation results in more user understanding? Globus and Uselton [22] have the following to say about this:

> *The grand hypothesis of the visualisation community is that scientific visualisation improves human insight. Several methods have been traditionally used to prove this hypothesis:*
>
> - *proof by repeated assertion*
> - *proof by vigorous gesticulation*
> - *proof by pretty picture*

Clearly there are better approaches to proving that a visualisation improves insight. Although insight itself is difficult to measure, task performance can more easily be measured [22].

To determine whether our proposed visualisations were valid and improved insight, we asked users questions around the content of the archive. If they deduced the correct answers, then we could assume that the visualisations were valid representations of the data and led to insight into the data.

Evaluation by comparison

Many information visualisation systems are evaluated by comparing them to some other (usually non-visual) system [2, 3]. One of the ways of comparing two systems is to set user tasks which are performed in both systems. The evaluation is then based on the comparative performance of users on the tasks, focussing on aspects such as the time to complete a task, number of errors, learning time, etc.

An alternative way of evaluating two comparative systems is to measure the ease (or difficulty) of accessing information. Card *et.al* [12] propose a metric for measuring the cost of accessing information in the form of a "Cost-of-Knowledge- Characteristic Function". This technique can be used to conduct empirical comparisons between different systems with the same function.

Experiments comparing visualisation systems are easier to perform than experiments to evaluate a particular system [22]. Such comparative experiments, however, are dependent on the existence of an alternative system. This is not always feasible.

5.2 Evaluation of display of 1D time data

We wanted to see if the visualisation of uncertainty ranges as proposed in Section 4.1.1 was effective. In particular, we wanted to test if the representation conveyed information on the contents of the archive in an easy-to-understand way. As explained in the previous section, we proposed using an inquiry based evaluation with users using an incomplete system. The purpose of inquiry methods is to gather subjective impressions about various aspects of a user interface, rather than studying specific tasks or measuring performance [26]. To achieve this we used screen dumps of the visualisation and asked users questions around their interpretation of the visualisation.

5.2.1 Hypotheses

We wanted to measure if the representation shown in Figure 4.1 was valid. In particular we wanted to see whether the following would be clear to the user:

1. Each artwork is represented by a bar on the display.
2. There is a timeline above the display and the bars are positioned below this timeline depending on their time attribute.
3. A bar represents a time-frame within which that artwork could have been made (i.e. a range of uncertainty).

5.2.2 Subjects

There were two factors we considered when choosing the subjects to take part in our study. Firstly, we wanted reasonably realistic users. In Section 2.3.2 we identified our primary users as artists, historians or academics, including casual web users. Globus [22] issues the following warning regarding the choice of subjects for experiments:

> *A final word of warning for those considering controlled experiments of visualisation systems, be careful to choose experimental subjects from the target user audience. For example, if insight into moderately experienced users is desired, don't use programmers or novice users (or freshman psychology students) as experimental subjects.*

The second factor we considered was to include some expertise in human processing of visual information. In [23], Korfhage stresses the importance of using visual experts during the design stages of any visual interface:

> *Given the centrality of the user, it is imperative that professionals who are familiar with human processing of visual information be consulted in the design and implementation of any visual interface. These include professionals in cognitive science and graphic design, among others.*

Twelve people from different disciplines were approached and took part in the experiment. Four academics were approached for their interest and involvement in African art: an African art lecturer, an archaeologist, an art history lecturer and a historian. Four people were approached for their interest in visual information and representations: a graphic designer, a communications lecturer, a computer science lecturer and a web designer. The four remaining subjects were selected as 'general public' users: a bookkeeper, a businessman, a marketer and a secretary.

None of the subjects had any prior knowledge of the project.

5.2.3 Questionnaire

We designed a simple questionnaire to test our hypotheses. Given that evaluation by comparison is easier to perform, we based our evaluation on a comparison of the contents of two archives. The questionnaire we used is shown in Appendix B. It starts with a basic description of the project and a brief explanation of the representation used. The purpose of the experiment is explained and the representations of two different archives are displayed as two figures.

In the explanatory text of the questionnaire (see Appendix B), the only information given on the visualisation is the following:

- The information stored in the archive is about African art.
- The two figures show the contents of two different archives of art.
- The main display shows the contents of the archive

- A single artwork is highlighted on the display of each archive and the corresponding artwork is displayed on the right, below the map.

To test our hypotheses, the data from the two archives displayed on the questionnaire differed: Archive A contained more artworks, whereas Archive B contained fewer and older artworks. The following comparative questions were then asked:

1. **Which archive has more artworks:**
   (a) **Archive A**
   (b) **Archive B**
   (c) **It is not clear which archive has more artefacts**

   With this question we were testing whether it was clear that each artwork is represented by a bar on the display. If this was clear, then it should have followed that (a) was the correct answer, since Archive A was represented by more bars than Archive B.

2. **Which archive contains the oldest artworks:**
   (a) **Archive A**
   (b) **Archive B**
   (c) **It is not clear which archive contains the oldest artefacts**

   With this question we were testing whether the subjects noticed and understood the timeline and that it was clear that the position of bars below the timeline indicated each artwork's associated time attribute. The correct answer is (b), since there are bars in the representation of Archive B that appear earlier than 2000BC, whereas all the bars in Archive A occur after 1000BC.

3. **What do you think it means if a bar in either one of the figures is longer than another bar?** This question tests whether it is clear that a bar represents a range of uncertainty. The question is open-ended in order to elicit as many different interpretations as possible.

5.2.4 Method

All the subjects involved in the experiment had no knowledge of the project. The process we followed was as follows:

1. An appointment was made with the subject.
2. The subject was visited in their workplace.
3. The questionnaire was handed to the subject and they were required to read it and answer the questions. No further explanations were given.
4. If the subject asked questions relating the answers, it was explained that questions cannot be answered until after the questionnaire has been filled in and handed back.

5.2.5 Results

The results are shown in Table 5.1.

Specialisation of person	*Question 1*	*Question 2*	*Question 3*
African art lecturer	a	b	Longer time span (the length of cultures)
Archaeologist	c	b	Continuation of a 'tradition' a trend in the chronological sense
Art history lecturer	c	b	That the culture lasted longer
Bookkeeper	a	b	Longer time period
Businessman	a	b	There is no specific determined date on the artwork. It was made between date X & date Y
Communications lecturer	a	b	It implies that 'the bar' (section) covers a longer time period
Computer Science lecturer	a	b	Express uncertainty of when an artefact was produced
Graphic designer	a	b	The artworks have been done between or over that period of time, it's unsure
Historian	a	a	It gives an indication of the chronological period and possible contents
Marketer	a	b	The period is longer / more artworks
Secretary	a	b	It was over a longer period
Web designer	a	b	A longer time period

Table 5.1: Results from experiment: description of person with the answers for each question

Ten out of the 12 people (83%) answered question 1 correctly (Archive A has more artefacts). Two people said that it was not clear which archive has more artefacts. Eleven out of the 12 (92%) answered question 2 correctly (Archive B has older artefacts). One person said that Archive A has older artefacts.

There were a variety of answers for question 3. These are analysed in the next section.

5.2.6 Discussion

From the responses to both questions 1 and 2 we can assume that:

- To ten out of the 12 people it was clear that each artwork is represented by a bar on the display.

- Eleven out of the 12 people noticed the timeline and deduced that the positioning of bars was dependant on their time attribute.

Question 3 is slightly more complicated. We have grouped the responses into four categories shown in Table 5.2.

Category	*Answers to Question 3*
Uncertain (3 answers)	– There is no specific determined date on the artwork. It was made between date X & date Y – Express uncertainty of when an artefact was produced – The artworks have been done between or over that period of time, it's unsure
Cultural era (4 answers)	– Longer time span (the length of cultures) – Continuation of a 'tradition' a trend in the chronological sense – That the culture lasted longer – It gives an indication of the chronological period and possible contents
Longer period (4 answers)	– Longer time period – It implies that 'the bar' (section) covers a longer time period – It was over a longer period – A longer time period
Incorrect (1 answer)	– The period is longer / more artworks

Table 5.2: Answers to question 3 of the questionnaire grouped into categories

The first group of responses (which we have called 'Uncertain') are what we originally thought of as a correct answer. The second group we have called 'Cultural era'. These responses were all given by people who work with history in some form (African art, archaeology, art history, history). These people have a deeper understanding of where the estimate of time comes from. When presented with a time period such as '6th to 4th century BC', they had the additional knowledge that this is an estimate based on a cultural era (usually represented by a combination of time, place and ethnic group). For example, some of the other time estimates in the CAMA database are:

- 22nd Dynasty, reign of Osorkon II, c. 874-50 BC
- predynastic, Amratian (Naqada Ic-IIa) c. 3800-3600 BC
- Meriotic period, 2nd-3rd Century AD

An answer to question 3 of a longer bar representing "that the culture lasted longer" (third answer under 'Cultural era') is therefore completely correct. The answer: "It gives an indication of the chronological period and possible contents" shows that the person (in this case the historian) is seeing a bar as representing a period and the corresponding artwork is merely an example of the kind of artwork that could have originated in that period. We therefore also view the second category of answer as correct.

The third category of answer to question 3 we have called 'longer time period'. These responses are true, but do not give us much of a clue as to whether they understood what this longer time period actually meant.

The only answer which we can confidently say is incorrect is the one which states: "The period is longer / more artworks". This person thought that a longer bar represents more artworks than a shorter bar.

In summary we can say the following about question 3: 7/12 answers were correct, 4/12 answers were unclear and 1/12 answers was incorrect. We can therefore say that most of the people understood that a bar represents a time-frame within which that artwork could have been made.

5.3 Evaluation of display of 2D space data

We wanted to see if it was feasible to place items on a map when the origins of the items are uncertain (as proposed in Section 4.1.4). To test this, we implemented a system and asked for feedback from users. This work formed part of a previous study [35] and is described briefly in the following sections.

5.3.1 Scope of study

To narrow down the problem we worked with a subset of CAMA's data. All items had the same level of uncertainty: they were known to originate from a single country, but the precise origin within the country was unknown.

We implemented three different techniques for placing the uncertain items within the borders of each country:

- random placement;
- square grid placement (with and without random offsets);
- circular grid placement (with and without random offsets);

The aim was to find a way of placing the icons (representing artefacts) on the map so that the position of the icon is not assumed to be the exact origin of the artefact.

5.3.2 User Feedback

We presented the techniques to users from a range of academic disciplines. These included geographers, computer scientists (with expertise in visualisation), psychologists and artists. The system was demonstrated to the users after which they were interviewed and asked to provide feedback on specific issues, including:

- the placement of items (which technique they felt was more effective);
- the assumed actual origin of artefacts.

Feedback indicated that users preferred the layout generated by a circular grid with a random offset, than the other layouts. Although users' said that they understood that each of the icons was associated with the whole country and not a specific point, this conflicted with comments made later on in the interviews. Some of the users said that they expected the icons to be concentrated around significant places, such as particular towns and museum sites, since more artefacts originate from these places. This view that more icons should be concentrated in particular places indicated that some of the users had the perception that the location of the icon represented a precise origin.

5.3.3 Discussion

Preliminary feedback from users therefore indicates that it could be misleading to place icons on a map if the origin is unknown. It seemed that the mental models users constructed of how accurate maps were in the real world caused a cognitive dissonance when looking at the maps generated by our system. Further studies are needed to confirm this.

Based on our results, we developed a prototype using the time-based map, rather than a geographic map as the central display. A 2D map of Africa was retained as a means of querying, but not visualising individual elements. In the next section we describe this prototype and the evaluation of the prototype.

5.4 Evaluation of mechanisms for querying uncertain data

We have shown through inquiry based evaluation that our proposed time-based visualisation is reasonably clear to users. Using this time-based visualisation as a basis for displaying artefacts, we developed a prototype for testing our proposed visual querying techniques. In this section we describe the prototype and the evaluation of the techniques.

5.4.1 The prototype

The user interface of the prototype is shown in Figure 5.1. The prototype is based on a subset of 28 artefacts. To simplify the problem, each artefact originates from only one country (i.e. we did not include any artefacts which could originate from more than one country). The user interacts with the prototype as follows:

- Clicking on a single bar on the display results in a pop-up window depicting an image of that particular artefact. Textual information associated with the artefact (including the

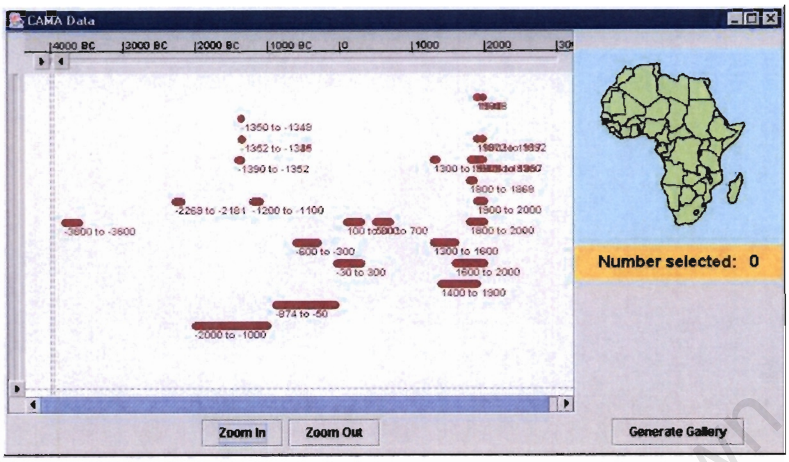


Figure 5.1: User interface of the prototype based on a subset of 28 artefacts

time and geographic origin) is also displayed in the window (Figure 5.2). While the popup window is open, the associated artefact is tagged with a red rectangle and the country is outlined in red on the map.

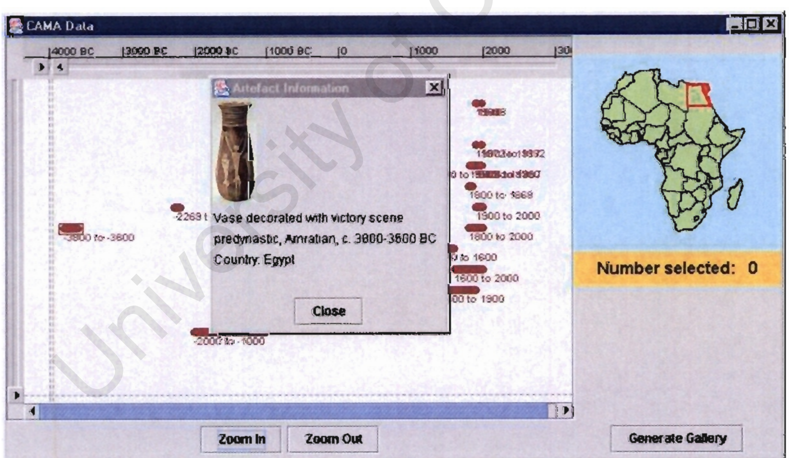


Figure 5.2: Popup window showing information on a particular artefact. The associated artefact is tagged with a red rectangle and the country of origin is outlined in red on the map.

- To select all artefacts originating from a particular country, the user can click on the relevant country on the map. In response, all the artefacts originating from that country are highlighted in yellow on the main display. The country is also highlighted in the same yellow colour and a label below the map (with the same yellow background) provides feedback on the number of artefacts selected. In Figure 5.3 South Africa is selected, resulting in four selected artefacts in the main display. To unselect these artefacts, the user simply clicks on the country again. Multiple countries can be selected, which results in artefacts from any selected country being selected.

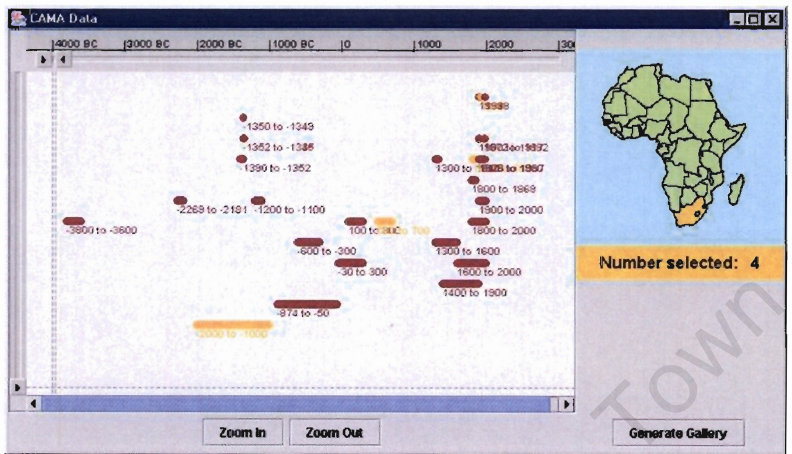


Figure 5.3: On the map, South Africa is highlighted in yellow. Artefacts originating from South Africa are highlighted in yellow on the display and the number selected is displayed below the map (on a yellow background)

- To select artefacts based on a time range, the user can position the time range selector buttons by moving them along the time line. In Figure 5.4 the selected time range is specified as 2000BC to 1000AD, resulting in 11 selected artefacts.
- The time range selector buttons can be used in conjunction with the map query tool to

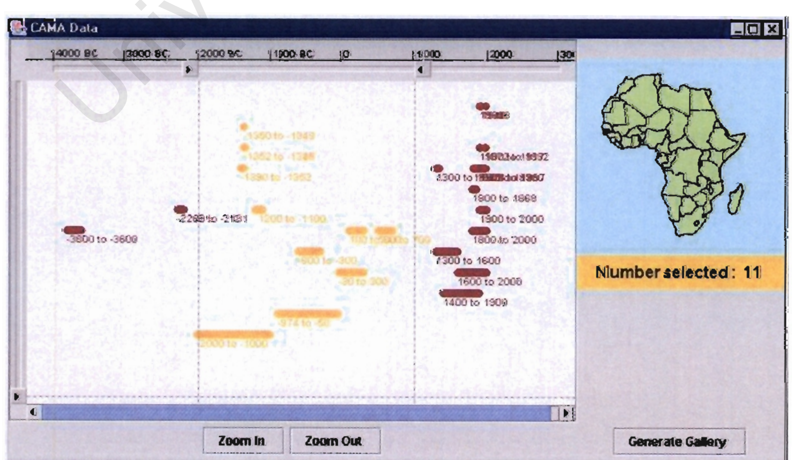


Figure 5.4: The time range selectors have been used to select all artefacts originating in the period 2000BC to 1000AD.

perform a union query.

- The uncertainty slider can be used to exclude the more uncertain items. In Figure 5.5 the user has selected items originating from 2000BC to 1000AD as well as items originating from Egypt. The uncertainty slider has excluded the lower two artefacts.

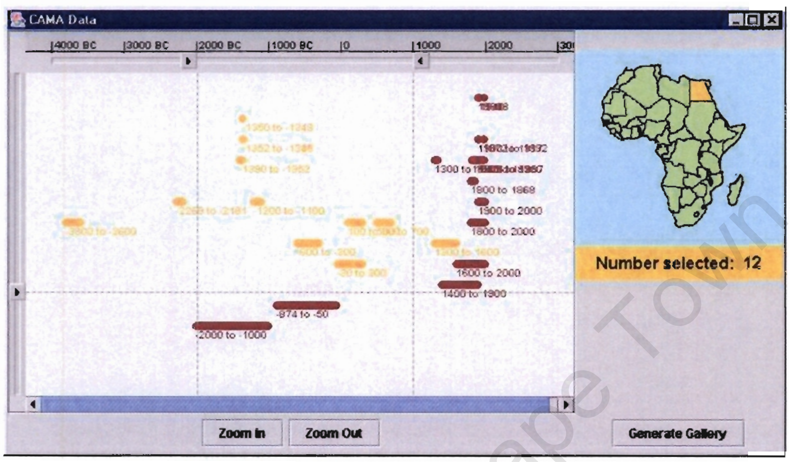


Figure 5.5: Selected artefacts include those originating in the period 2000BC to 1000AD as well as those originating from Egypt (highlighted in yellow on the map). The lower two artefacts have been excluded using the uncertainty slider.

- The user can also zoom in and zoom out on the display.

5.4.2 Aim of the experiment

The aim of the experiment was to determine whether a person with no knowledge of the project and without help could correctly perform basic querying tasks using the prototype. Unlike in the case of the inquiry-based evaluation of the display of 1D time data where we attempted to find realistic users, for this testing we used representational users.

For the experiment, users were required to perform specified tasks while interacting with the prototype. The questions covered the following basic tasks:

- identifying a particular artefact on the display based on its time attribute;
- selecting and unselecting all artefacts which originate in a given country using the map query tool;
- selecting all artefacts which originate in a given time period using the timeline query tools;
- selecting a set of artefacts originating from a given country in a given time period using a combination of the timeline query tools and the map query tool.

We also wanted to determine whether a person, after interacting with the prototype for a short period of time, could answer some questions about the visualisation. The questions we posed were the same three questions used in the initial inquiry-based evaluation of the 1D time data display and one further question. The additional question related to the purpose of the uncertainty slider. The task sheet is given in Appendix C. We discuss each of the questions in Section 5.4.4 below.

5.4.3 Method

We designed a set of eight short tasks for testing the effectiveness of our proposed querying techniques. Sixteen people took part in the experiment. None of these people had any prior knowledge of the project. There was no overlap between the subjects used for this experiment and any prior experiments.

The experiment took the form of a pluralistic walkthrough, where subjects worked in pairs. Each pair completed the task sheet, while interacting with the prototype on computer. While working through the tasks, they were observed by a single observer and their discussion was recorded in written form. No help was provided by the observer.

The actual process followed for each pair of subjects was as follows:

1. An appointment was made with the pair of subjects.
2. On arrival at the appointed time, the subjects were told the following:
   - that the project related to an archive of African artworks;
   - that the purpose of the experiment was for us to determine whether our ideas were effective or not;
   - that we wanted them to attempt to complete the tasks on the sheet of paper;
   - that they should discuss their thoughts with each other and that this would be recorded;
   - that they should not ask for help.

5.4.4 Results

The full task sheet is given in Appendix C. In this section we discuss each question and how it was answered by all eight subject pairs.

Question 1

> *There is an artefact that originates in 3800 – 3600BC. Click on this artefact to obtain more information on it. Where does this artefact come from?*

With this question we were testing whether the subjects were able to find a particular artefact based on its time attribute. All eight pairs of subjects performed this task without hesitation and all wrote down the correct answer (Egypt). In the case of two of the subject pairs, there was some discussion around what a negative value means (as it appears on the label printed next to the bar representing the artefact (-3800 to -3600)). Both pairs came to the conclusion that the minus referred to Before Christ (BC). One subject event commented that "*0 is when Christ was born*".

From these results we can deduce that the users had no problems finding particular artefacts based on their time attribute.

Question 2

> *How would you select all artefacts that originate from South Africa? How many artefacts from South Africa are currently stored in the archive?*

With these questions we were testing whether using the map as a query tool was an intuitive task or not. We were also checking that subjects would deduce that artefacts highlighted in yellow were "selected".

All eight pairs of subjects clicked on South Africa to select the artefacts and correctly deduced that four artefacts originate from South Africa. In all cases, clicking on South Africa was the first option tried. Typical comments included "*See if you can click on the map*", "*Probably hit the map*", "*Let's click on the map - Oh, that's clever* [on observing the result]".

To determine the number of artefacts selected, some of the subjects noticed that the number of selected artefacts was printed below the map, while others counted the number of yellow items on the display. One of the pairs did not notice the label below the map and were at first not sure whether each marker represented a single artefact or many artefacts. The question posed was "*Is there only one artefact in each little dot?*". To answer their question they explored by clicking on different markers. They quickly deduced that each marker represented a single artefact and so wrote down the answer 4.

Some of the subjects tested whether their answer was correct by clicking on the selected markers to check if the information on the artefact stated that it was indeed from South Africa. One subject commented that "*the yellow ones are highlighted – see the colours match* [on the map and on the display]". To check that their deduction was correct, they clicked on other countries.

From these results we can see that users had no problems selecting a subset of artefacts from a particular country using the map query tool.

Question 3

> *How would you unselect these artefacts?*

This question was simply to test if they could use the map query tool to unselect items already selected. All eight pairs of subjects managed this successfully by clicking on South Africa again. Most achieved this on their first attempt. One pair had already tried it while answering the previous question. Two subjects (from different pairs) first suggested that they try clicking on the sea. When that did not work, they tried clicking on South Africa again.

In summary, it was immediately clear to most users how to unselect items (currently selected based on the country of origin) using the map query tool. To a minority of users, this became clear on their second attempt.

Question 4

> *How would you select all artefacts that originate in the period 2000 BC to 1000 AD? How many artefacts are there that originate in this period?*

With this question we were testing whether users would know to use the time query tool to select items in a given time period. For this to be possible, they would need to notice the time line above the display as well as the selector buttons and would need to figure out how to use the two buttons together to select a range.

All of the eight pairs of subjects figured out that they had to use the buttons on the timeline to select the artefacts. Five of the pairs correctly indicated that 11 artefacts originated in the given period. Two of the pairs correctly positioned the sliders, but counted the yellow markers as 10 instead of 11 (there was some overlap between the selected markers). The final pair correctly used the sliders, but incorrectly positioned the right slider on 1000BC instead of 1000AD, producing an incorrect answer of 6 selected artefacts.

Although three of the pairs of subjects immediately used the buttons on the timeline to select a range, the other pairs tried other options before noticing the timeline slider buttons. Three pairs of subjects first tried to click and drag directly on the timeline (i.e. without using the buttons) from the start date (2000BC) to the end date (1000AD), before trying the buttons. One other pair first tried zooming in to the given range. When they realised that was not working they tried to click directly on the timeline and after that noticed the buttons and used them correctly. The final pair went through a longer process. They first tried dragging directly on the timeline, then tried right-clicking on the timeline, then tried zooming in to the given range. At this stage they made the observation that "*we are not marking them*". Their next attempt was to try holding down the Ctrl key while clicking and finally they noticed the buttons on the timeline and used them successfully.

In summary, all users managed to use the time query tool to select a subset of artefacts based on a time range, although some took longer than others to find the query tool. 3/8 of the pairs immediately found the query tool. 3/8 of the pairs found the tool on their second attempt. 2/8 of the pairs took longer to find the tool. It is worth noting, however, that once users found the buttons, they had no problems in using them to formulate the query.

Question 5

> *How would you add to this selection all artefacts that originate in Egypt? How many are selected now?*

With this question we were testing whether users would be able to use the timeline query tools with the map query tool to formulate a simple compound query. This was limited to the union operator (i.e. adding items to the selected set).

All pairs of subjects correctly combined the positioning of the timeline with clicking on Egypt to add to the selection. Six pairs wrote down the correct answer of 14 artefacts (the number selected). One pair deduced that there were 13, since their previous answer was incorrectly specified as 10 and they counted 3 more outside the selected time frame. Another pair had the incorrect answer of 13 because the one timeline slider was positioned on 1000BC instead of 1000AD (as explained in the previous question).

In summary, all users were able to correctly use the time query tool with the map query tool to formulate a simple union query.

Question 6

What do you think it means if a bar on the display is longer than another bar?

This question tests whether it is clear that a bar represents a range of uncertainty. The question is identical to the third question posed in our initial inquiry-based evaluation (described in Section 5.2). In our initial evaluation, subjects were given an introductory explanation (see Appendix B) which included the following information:

- The main display shows the contents of the archive.
- A single artwork is highlighted on the display of each archive and the corresponding artwork is displayed on the right, below the map.

By asking this question again (to different subjects), we wanted to see whether users would be able to answer this question without having been told the above facts. We wanted to see whether they could deduce the correct answer from their short experience of interacting with the system. As in the initial evaluation, we have grouped the responses into categories as shown in Table 5.3. Note the actual answers are shown in italics and author comments are given in square brackets.

In summary, we can say the following about Question 6: 4/8 of the answers were correct, 3/8 were not clear and 1/8 answers were incorrect (although through the discussion we could see that this pair later changed their understanding to a correct answer).

In the initial questionnaire-based evaluation, there were 12 subjects. Four of these subjects worked with historical data in their occupations (African art, archeology, art history, history). These four subjects all gave correct answers in the 'cultural era' category (see Table 5.2). In the current prototype-based evaluation, none of the users worked with historical data and it is significant that none of the answers could be categorised as 'cultural era' responses. If we ignore the 'cultural era' responses in the questionnaire-based evaluation, we are left with 8 users, which is comparable with the 8 pairs in the current evaluation. The number of answers in each of the remaining categories for both evaluations is then very similar (see Table 5.4).

Question 7

What do you think the purpose is of the vertical slider on the left of the display (the slider which moves up and down with associated horizontal dotted line)?

Category	*Answers to Question 6*
Uncertain (4 answers)	– *Uncertainty of time* – *They don't know exactly – accuracy problem* – *Not sure of time precision of time* – *Longer period (unknown time for specific artefact)* [although their answer was 'longer period', they qualified that they meant that it was unknown]
Longer period (3 answers)	– *Longer period* – *The time period is more than the other* – *Longer time period*
Incorrect (1 answer)	– *more artefacts* [This pair initially gave the answer 'more artefacts'. However, later during the experiment (after having interacted for longer with the system) one of them stated: "*So, it's not more artefacts*" and later said: "*Ah, for each of these artefacts, they're estimating – they don't know precisely*". The original answer was, however, not changed.]

Table 5.3: Answers to question 6, grouped into categories

Category	*Number of responses: questionnaire-based evaluation*	*Number of responses: prototype-based evaluation*
Uncertain	3/8	4/8
Longer period	4/8	3/8
Incorrect	1/8	1/8

Table 5.4: Comparison of results for questionnaire-based and prototype-based evaluations (ignoring the 'cultural era' category

The purpose of asking this question was to see whether it was clear to users that the vertical axis represented uncertainty (more uncertain items are plotted lower down than less uncertain items). In addition, we wanted to see whether users understood that they could use the uncertainty slider to unselect the more uncertain items from the existing selection.

Although most of the users understood that the uncertainty slider was used to unselect items, only 1 of the user pairs understood that it was the more uncertain items that were being unselected. Table 5.5 lists the answers in three categories: 'fully understood', 'understood that slider used to unselect' and 'confused'. For each user pair we have given the answer to the question they wrote down on the task sheet as well as verbal comments made (which were recorded by the observer), shown in brackets after the written response.

These results seem to indicate that although most of the users understood that the uncertainty slider is used to unselect artefacts, they did not understand on what basis this was being done. Only 1 pair clearly understood that it was the more uncertain items that were being unselected. It seems therefore that it was not clear to most users that the vertical axis represents uncertainty.

Category	*Written answer (verbal comments in brackets)*
Fully understood (1 pair)	– *Distinguish least accurate and more accurate dating* (“*Eliminate objects where the dates are not so certain*”)
Understood that slider used to unselect (5 pairs)	– *Clears selections* – *Indicates how many artefacts there are* (“*Takes a subset of them*”) – *It represents the number of artefacts* (“*Used to eliminate*”) – *To select to smaller period - do not know* (“*It works with these*” [time sliders],“*It highlights all of them in the rectangular thing*” – *Used to select a specific range* (“*It's used to unselect*”)
Confused (2 pairs)	– *Zoom marker* (“*Doesn't select anything*” [none were selected before], “*What can this scale be?*” [pointing to uncertainty axis], “*Maybe they're from the same country*”, “*Makes no sense to me*”) – [No answer written on task sheet], (“*Does it have something to do with North and South?*”,“*It works if something is selected*”,“*There must be some or other reason why it is up or down*”, “*There must be a reason why they are on different levels*”)

Table 5.5: Summary of results for Question 7

Question 8

The two screen shots on the following page show the contents of two different archives. Tick the relevant answers:

1. *Which archive contains more artefacts?*
   (a) *The top archive*
   (b) *The bottom archive*
   (c) *It is not clear which archive has more artefacts*
2. *Which archive contains the oldest artefact?*
   (a) *The top archive*
   (b) *The bottom archive*
   (c) *It is not clear which archive contains the oldest artefacts*

The purpose of asking these questions was to see whether it was clear that each artwork is represented by a bar on the display and whether users understood the position of artefacts in relation to the timeline (i.e. each artefact's time attribute). These questions were identical to questions 1 and 2 in our initial questionnaire-based evaluation. By repeating these questions (with different users) we wanted to see whether users could answer these questions without the background information given in our initial evaluation (as described in Question 6 above).

All user pairs ticked the correct answers for both questions. It seems therefore that to all 8 pairs of users (through interaction with the prototype) it was clear that:

- each artefact is represented by a bar on the display, and
- the positioning of bars on the display is dependent on their time attribute.

5.4.5 Discussion

From the results of the prototype-based pluralistic walthrough evaluation, we can deduce the following:

- Users correctly performed the following tasks with ease:
  - querying a particular artefact based on its time attribute,
  - selecting and unselecting artefacts based on the country of origin using the map query tool,
  - performing a simple union query by combining the time and map query tools.
- Some users experienced difficulty in locating the time query tool sliders, but once located, they used them correctly with ease.
- Users experienced difficulty in understanding the basis of and purpose for the uncertainty slider for excluding more uncertain artefacts from the current selection.
- A repetition of the three questions posed in the initial questionnaire-based evaluation of the 1D time-based visualisation yielded similar results in this prototype-based evaluation.

5.5 Conclusion

In this chapter we described how we evaluated some of our proposed techniques for visualising and querying African cultural data.

To evaluate the visualisation based on 1D time data, we performed an inquiry-based evaluation using realistic users. Results were encouraging and seemed to indicate that the visualisation of uncertainty ranges on a time map was sufficiently effective. A second inquiry-based evaluation on the display of uncertain items on a geographic map did not yield such positive results. Feedback from users indicated that it could be misleading to place icons on a map if the origin is unknown.

To test our proposed querying techniques, we implemented a prototype using the time-based visualisation. We used a form of pluralistic walkthrough where pairs of users completed a sheet of tasks while interacting with the prototype. Results indicated that users correctly used the time and map query tools with reasonable ease. Users, however, experienced problems in understanding the uncertainty slider (for excluding more uncertain items from the selection). If deemed important enough to include in the final system, this tool would require user training or help in the form of user hints.

In the following chapter we discuss how the time-based visualisation and query mechanisms (as implemented in our prototype) could be combined with other proposed techniques into a single visual query system. These additional techniques include:

- the notion of tagging and selection to support multiple boolean queries (not just the union query as in our prototype);
- support for different levels of uncertainty in space.

Chapter 6

Visual query system

Using the techniques proposed in Chapter 4 and based on the evaluation of these techniques (Chapter 5), we show in this chapter how our techniques can be combined into a single visual query system. We describe how the system would work and then revisit the original user requirements as a way of reviewing the proposed system. Some of these techniques were implemented in a pilot system. The details of this implementation are given in Appendix A.

6.1 User interface

The overall user interface of our visual query system is shown in Figure 6.1. All items in the database are shown in the central data display area. The contents of the archive should also be shown on the map as a dot density map. That way the user can see how the items are distributed geographically. By manipulating query tools, items on the data display are selected. The final aim is to automatically generate a virtual gallery of the artworks, which will be populated by the selected items by the user clicking on the 'Generate Gallery' button.

6.1.1 Probing

Any item can be queried (by right clicking on the mark on the main display), and a thumbnail image of the relevant artefact is displayed on the bottom right corner. Associated attributes should also be displayed next to the image as shown in the figures of Appendix B. More detailed probing could be done on a given item by displaying a window with all the details from the database. A further possibility is that the geographic origin of the 'probed' element is shown on the map as either a highlighted dot (for items with a certain origin) or a highlighted country/countries.

6.1.2 Selecting a subset of items

To select a subset of items, the user first "tags" items in the data display. Elements are tagged by clicking on individual items, by using the time query tool (the slider above the data display) or the space query tool (the map on the right – described in Section 6.1.3). Once some of the elements have been tagged, Boolean operators (see the vertical button bar to the right of the data display) are used for combining the current selection with the tagged items in different ways. Initially, no items will be selected, so the user will select the set of tagged items using the replace button ('='). This will result in all tagged items becoming selected. Other items can then be tagged and these can be added to (or subtracted from, or intersected with) the current selection.

Each tagged item is surrounded by a red rectangle for user feedback, while selected items are shown in yellow (note that an item can be both tagged and selected). The number of items tagged and the number of items selected are fed back to the user below the data display area.

Figure 6.1 shows these two tools in operation. Some of the elements have been selected (those shaded yellow, or grey in black and white print). The timeline controls are used to tag elements in the given range (tagged items are surrounded by rectangles), while the uncertainty control excludes the two very uncertain elements below the horizontal dashed line.

In this way, using the query tools, the user is able to continually refine their query until they have the desired selection and number of artefacts. The visual representations of tagged items and selected items ensures that the user knows which items will be involved in any operation before the operation is done.

6.1.3 The space query tool

Besides being a display mechanism, the map can also be used to specify queries. (This idea is known as "equal opportunity" and was first proposed by Thimbleby [42]. It allows users to refine queries, just as with Query-by-Example techniques described by Zloof [48].) The map can also be used as a query tool.

There are three buttons above the map display, which act as tools. The first button (i) is used for querying the map. If this tool is selected and the user clicks on the map, information about that country will be displayed: the name of the country and the number of items which occur in that country. The other two buttons are used for selection.

The country selector button (C) is used for selecting countries, whereas the sub-region selector button (S) is used for selecting sub-regions. The idea is that the user first chooses a tool for selecting areas. Depending on the tool selected, they will be in a particular selection mode. If they are in a country selection mode, and they click on the map, the relevant country will be highlighted. If they are in a sub-region selection mode, and they click on the map, the relevant sub-region will be highlighted. In this way, the user can select multiple areas on the map. As countries are selected, the items which fall into that region, are tagged on the data display area.

The user is also given the option of either including items with uncertain geographic attributes in the query, or not. The user does this by using the check box below the map tool. If this option is selected, additional items, which could possibly occur in the selected area, are also tagged. It is important to remember that a query on a region will include all items, at different levels of precision, which fall into that region. For example, if the user selects Zimbabwe and Mozambique, then the items which will be tagged will include: items which originate from

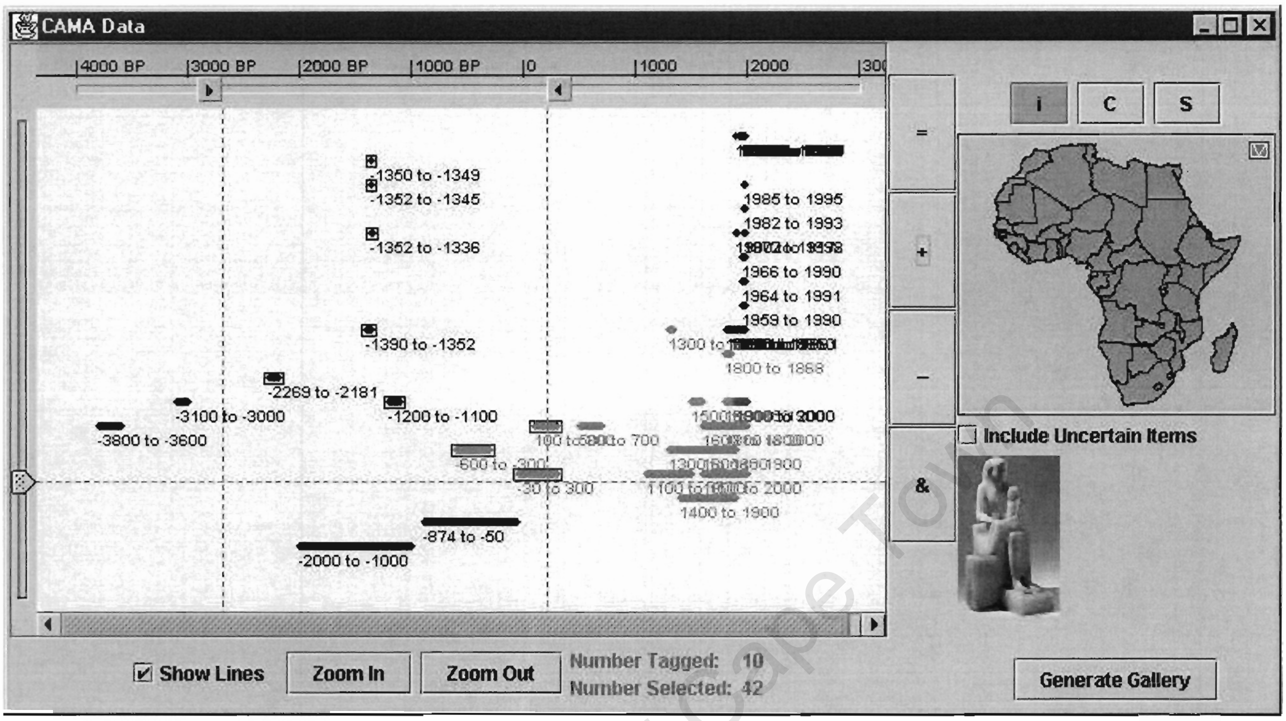


Figure 6.1: *Pilot Visual Query System.* Tagged elements are enclosed in rectangles (coloured red in the system), while selected items are highlighted in yellow (light grey on black and white printout). The time query tools (arrowed buttons above the data display), map query tool (on the right) and Boolean operator buttons, are used to tag elements.

either country, items which originate from a point in either country and items which originate in a province in either country. In addition, if the user has chosen to include uncertain elements, then items which have an area which is a superset of either country as their origin (Southern Africa, for example), will also be included.

6.2 Sample query

In this section we provide an example query and step through the process involved in solving the query using our system.

> **Query:** Find all artefacts that occur in the period 1000BP to 1000 in Southern Africa, including uncertain items (i.e. items that could occur in the given time and place).

1. Tag items on data display using time query tool sliders (tagged items are surrounded by red rectangles).

2. Drag uncertainty slider to the bottom to include all uncertain items.
3. Select the tagged elements by pressing the replace (=) tool.
4. Use map query tool to tag items that occur in Southern Africa (first select the sub region tool, then click on Southern Africa on the map).
5. Check the box below the map to include uncertain items.
6. Intersect the selected items with tagged items using the intersect button.

This example illustrates how a fairly complex query can be achieved relatively easily simply by clicking on a few buttons.

6.3 Review of the system

In deciding whether our overall approach is adequate, we revisit the requirements of the application (see Chapter 2):

1. *Users would typically be artists or academics, most probably with limited expertise in using a computer. The user interface should therefore be accessible to novice users and use metaphors familiar to such users. In the case of query tools, the usual SQL-based query mechanisms would be inappropriate.*

   The techniques we propose use metaphors familiar to most users (a time line and a geographic map). This, in conjunction with visual query mechanisms, makes the approach accessible to novice users.

   As an example, consider the sample query expressed in Section 6.2. We explained how this query could be expressed in six simple steps using the visual tools. The equivalent query in SQL would look something like this:

   ```
   Select Catalogue
   From Artworks, Artwork_Country_Link
   Where (Artworks.CatalogueID = Artwork_Country_Link.ArtworkID)
      And (year2 >= -1000 and year1 <= 1000)
      And placeID in
      (Select Place
       from Place
       where type = country
       and inside = "Southern Africa")
   ```

   Novice users cannot be expected to express queries using such complex syntax. The use of dynamic queries, extended to handle uncertain data, with the possibility of fully combining queries, means that users can do all their querying visually.

2. *The users would want to explore themes and influences, so the relationships between data elements should be displayed along particular themes.*

   By displaying the items on a time map, users can explore the theme of time as an influence on art. The problems with displaying uncertain geographic data on a map, made it infeasible to do the same with space as a theme.

3. *Some data elements which have uncertain origins, are extremely important from a cultural perspective. Therefore, it is important that uncertain items be treated with the same level of importance as items which have definite origins.*

   We have shown how both uncertain and certain items can be displayed on the same time-based map. We have devised query mechanisms which catch both certain and uncertain data in the same query. In these ways, the uncertain items are given the same level of importance as certain data elements.

4. *Since the archive is continually updated, users of the archive would typically know very little about the nature of the data stored in the archive: data volumes, areas of coverage, etc. It is therefore important that the application provide an overview of the data and visual feedback on the results of queries.*

   Through the time-based visualisation, an overview of the data is provided. This display not only indicates how much data is in the archive, but also where the data is situated in time. Through user tests (Section 5.2) it was clear that users were able to tell when one archive had more artefacts than another archive. The dot-density overview map also provides details of the spread of data over the geographic area.

6.4 Conclusion

In this chapter we illustrated how our proposed visualisation and query techniques could be combined into a single system. The main visual display is based on a time-based visualisation of all the artefacts in the archive. The user is able to select a subset of items by firstly tagging items using the time and space query tools and then selecting them using one of the boolean operator buttons. We have shown how a sample query could be expressed using our proposed visual tool. We also revisited the original requirements of the application to satisfy ourselves that these requirements would be met by such a system.

Although many evaluative studies have shown the benefits of using a dynamic query approach [2, 15, 37, 40, 46], further work is needed to properly evaluate the combination of the boolean operator buttons with dynamic queries.

Chapter 7

Conclusion

The contributions of this work can be seen from two angles. Firstly, in specific ways we have contributed to making African art accessible through the development of visualisation and query tools. Secondly, in general terms, we have contributed to the development of visual query tools for probabilistic uncertain data in space and time. We look at both of these aspects in the following sections.

7.1 Tools for accessing African art

Storing and retrieving digitized African art introduces complications not found in systems designed for Western art. Time and location uncertainty, coupled with a lack of standard classification categories, mean that new interfaces need to be developed if we are to make African artefacts accessible to an on-line audience. We have built a system which stores uncertain data and have built a novel interface which allows this data to be queried effectively, despite the uncertainty. Requirements of the system were that:

- the user interface should be accessible to novice users;
- relationships between data elements should be displayed along particular themes (such as space and time);
- uncertain items should be treated with the same level of importance as items with definite origins;
- the application should provide an overview of the data.

Our proposed techniques meet all of these requirements. Through the interface, users can perform all their querying in a visual way, making the process of querying accessible to novice users. Relationships between data elements based on time are depicted by our time-line visualisation. User studies on the time-based visualisation indicated that the overall metaphor was understandable to users. Further tests showed that users were able to effectively perform simply queries using both the time and space visual query tools. With our approach, both certain and uncertain items are displayed on the same map. The storage and query mechanisms also ensure that uncertain items are given the same level of importance as certain elements. Finally, an overview of the data is provided in time through the main visualisation and space through the dot-density map.

7.2 Visual query tools for uncertain data in space and time

In general terms, we have presented approaches to storing, visualising and querying uncertain data in archives, which ensure that uncertain data are given the same level of importance as certain data. We have shown how data with uncertain 1-D attributes (such as time) can be visualised by using a scatter-plot using uncertainty as the second dimension. The level of uncertainty can be depicted using explicit representations, such as the lengths of lines. Using a dynamic query approach, we have shown how data based on this scatter-plot can be queried using double sliders.

In the case of uncertain 2-D attributes (such as geographic location), we have found that it is potentially misleading to place individual items on a 2-D map based on this attribute. We have shown, however, how such a map can be used to provide an overview of the data (in the form of a dot-density map) and can be used as a 2-D query tool on a visualisation based on an alternative attribute.

Finally, we have shown how visual query tools can be combined with Boolean operator buttons so that users can express complex queries in a simple way.

7.3 Contributions of this work

This work contains several novel contributions:

1. We have developed a technique for visualising probabilistic uncertain one-dimensional data (such as time). This technique uses a scatterplot where the uncertainty is plotted as the vertical axis. The degree of uncertainty is depicted using an intrinsic representation of the length of the marker. User tests show that this technique is understandable.

2. We have introduced a technique for using a map as a 2D query tool in a dynamic query interface. This notion allows the map to be used as both a display mechanism and a query mechanism.

3. We have developed a technique for combining Boolean operators with dynamic queries by distinguishing between tagged and selected items.

7.4 Further work

Further work is required to evaluate our proposed techniques for expressing complex combinations of queries. Although users were able to correctly use both the time and space query tools, further evaluation is needed to confirm the effectiveness of the Boolean operators used with these query tools.

Another area for further work lies in finding effective ways of visualising uncertain 2-D data to support visual queries. In our application, we have visualised the 1-D uncertain data on a 2-D time map, rather than displaying the items on a 2-D space map. Initial user feedback indicated that visualising uncertain 2-D data on a 2-D map was a problem. Further studies are needed to confirm this. Ideas which could be considered include:

- Using a 3-D representation, where the third dimension represents uncertainty. The uncertain items would then be plotted "above" the map – the further the item is from the map, the more uncertain its location. A problem with this approach is that the objects will still have to float above some exact location. It is also questionable whether this can be done without misleading or confusing novice users.
- Plotting items on a 2-D map, but with some novel way of indicating which items are certain and which are uncertain.

One of the requirements of our system was that the certain and uncertain data be managed together. An interesting thought is whether it would be effective to manage the contemporary data (which is normally the certain data) and the traditional data (which is normally uncertain) as separate query tools. Further work could include investigating this possibility and seeing if anything is lost by approaching the problem this way.

In a wider perspective, we believe that there is a lot more work to be done within the field of creating interfaces to uncertain data. Our application has introduced us to time and location uncertainty, but there exist many other types of uncertainty which could be addressed – for example uncertainty about the creator of the artefact. We believe this is a fertile area for study.

Appendix A

Implementation

A prototype was developed including some of the features described in Chapter 6. In this appendix we give a brief overview of the technical details of the implementation of the prototype.

A.1 Overview

At the time of the study, CAMA did not have a database. Most of their data was stored on tapes or in static HTML pages. For this study, a database was developed in Microsoft Access (based on the model shown in Figure 4.7) and populated with sample data from CAMA's web site. The program itself was implemented in Java. The map tool was based on a GIS coverage in shapefile format [16] and was incorporated as a `JShape` component [29].

A.2 Java implementation

Figure A.1 shows a UML class diagram containing the most important classes of the system. The purpose of each of these classes is briefly explained below.

A.2.1 class `Visualize`

This class extends Java's `JFrame` class and is used to create the main application window. It contains the timeline (which is a `DataRuler` object), the map tool (which is a `JShape` component), the Boolean operator buttons and the main visualisation panel (which is a `PlotArea` object). It also reports on the number of objects tagged and selected.

A.2.2 interface `VisDataElement`

This interface specifies the functionality required of a data element to be visualised by the system. In this way different data sets can be visualised using the `Visualize` class, as long as the elements conform to this interface.

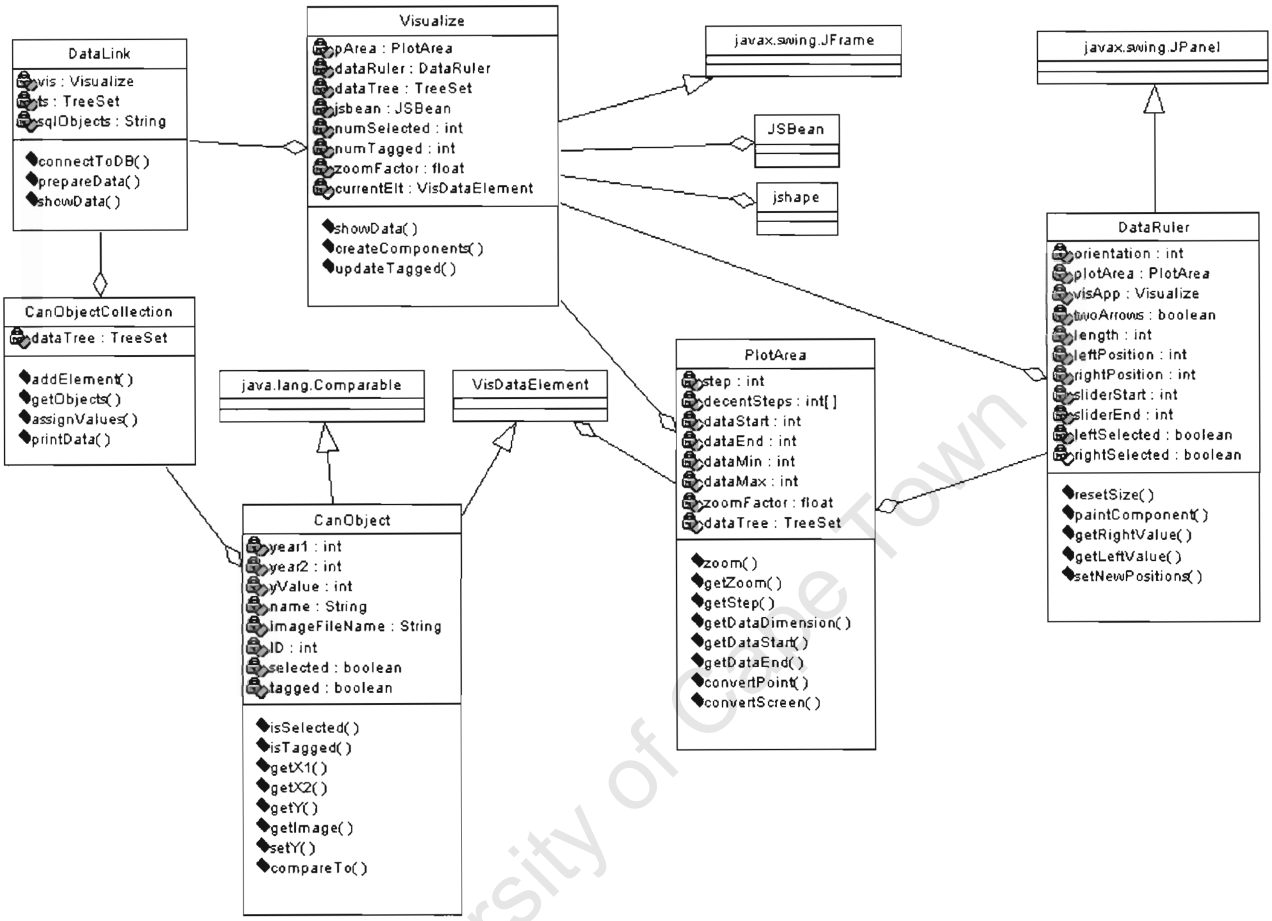


Figure A.1: UML class diagram of the classes used to implement the prototype

A.2.3 class `CanObject`

This class is an implementation of the `VisDataElement` interface and describes an element in the CAMA archive. By implementing `VisDataElement`, it becomes an eligible type to be visualised by the system. `CanObject` also implements Java's comparable interface so that `CanObject` objects can be compared with each other. The comparison is based on the level of uncertainty with respect to time.

A.2.4 class `CanObjectCollection`

This class contains a collection of `CanObject` objects in a tree structure.

A.2.5 class `PlotArea`

This class describes the main visualisation panel of the application window. It includes functionality for displaying any set of `VisDataElement` objects. Based on the data it is given, it's methods calculate a sensible range for the timeline, calculate screen positions for each item in the database and handle the zooming in and out of the display.

A.2.6 class `DataRuler`

This class extends Java's `JPanel` class and defines the range selector for performing queries on the display. It contains an inner class, `ArrowListener`, which defines the event handlers to the left and right arrow buttons.

A.2.7 class `DataLink`

This class is used to access the database of artefacts. The method `connectToDB` establishes a connection, executes a query and initialises a `CanObjectCollection` object with values from the database.

Appendix B

Questionnaire used for inquiry-based evaluation

We used the questionnaire shown on the following three pages to test the effectiveness of our display of time-based data. The test is described in Section 5.2.

Experiment

As part of a research project we are trying to create pictures (or representations) which convey information on archives of African art. We would like to see whether the representations we have used are intuitive — in other words, do the representations we use convey information on the contents of the archive in an easy-to-understand way?

The two screen shots shown in Figures B.1 and B.2 on the next page illustrate a tool used for querying archives of African artwork. The main display (the white rectangle taking up most of the screen) shows the contents of the archive. Figure B.1 shows the contents of archive A and Figure B.2 shows the contents of a different archive B. In archive A, a single artwork is highlighted in a lighter colour (towards the bottom of the display) and the corresponding artwork made in 12th-15th Century of a head is displayed on the right (below the map). In the same way, in Figure 2, the artwork called *Anubis mask* originating from 6th - 4th century BC or later is highlighted on the display.

The purpose of the questions we will ask you (on the last page) is not to test you, but rather for us to test whether our representation is clear. It is not necessary to understand the detail shown on the pictures, but rather that you gain an overall impression of the contents of the archive.

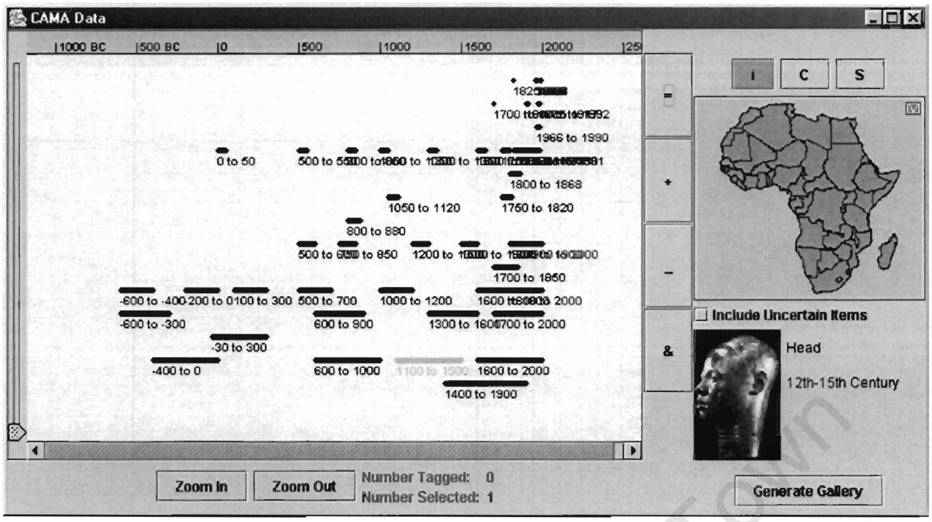


Figure B.1: Representation of Archive A

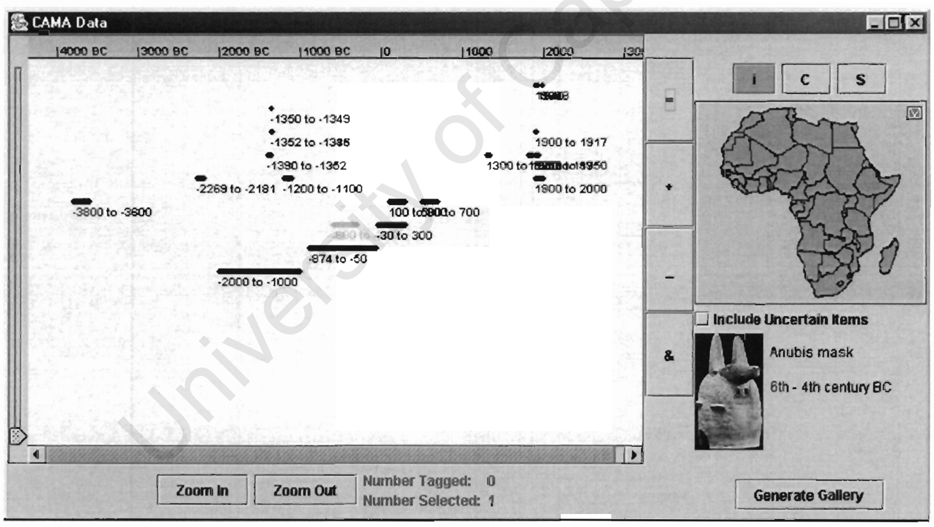


Figure B.2: Representation of Archive B

Please circle the letter corresponding to your answer.

1. Which archive has more artworks:
   (a) Archive A
   (b) Archive B
   (c) It is not clear which archive has more artworks
2. Which archive contains the oldest artworks:
   (a) Archive A
   (b) Archive B
   (c) It is not clear which archive contains the oldest artworks
3. What do you think it means if a bar in either one of the figures is longer than another bar?

   ____

   ____

   ____

   ____

   ____

Occupation: ______________________________

Thank you for your time!

Appendix C

Task sheet used for prototype-based evaluation

To evaluate the visual querying mechanisms we developed a prototype and used this prototype to perform a task-based pluralistic walkthough experiment. The experiment is described in Section 5.4. The following two pages show the tasks as they were given to the experiment subjects.

Experiment

Please try to answer the following questions by interacting with the system:

1. There is an artefact that originates in 3800 – 3600BC. Click on this artefact to obtain more information on it. Where does this artefact come from?______________________________
2. How would you select all artefacts that originate from South Africa? How many artefacts from South Africa are currently stored in the archive? ______________________________
3. How would you unselect these artefacts?
4. How would you select all artefacts that originate in the period 2000 BC to 1000 AD? How many artefacts are there that originate in this period?______________________________
5. How would you add to this selection all artefacts that originate in Egypt? How many are selected now?____
6. What do you think it means if a bar on the display is longer than another bar?

   ____

   ____

   ____

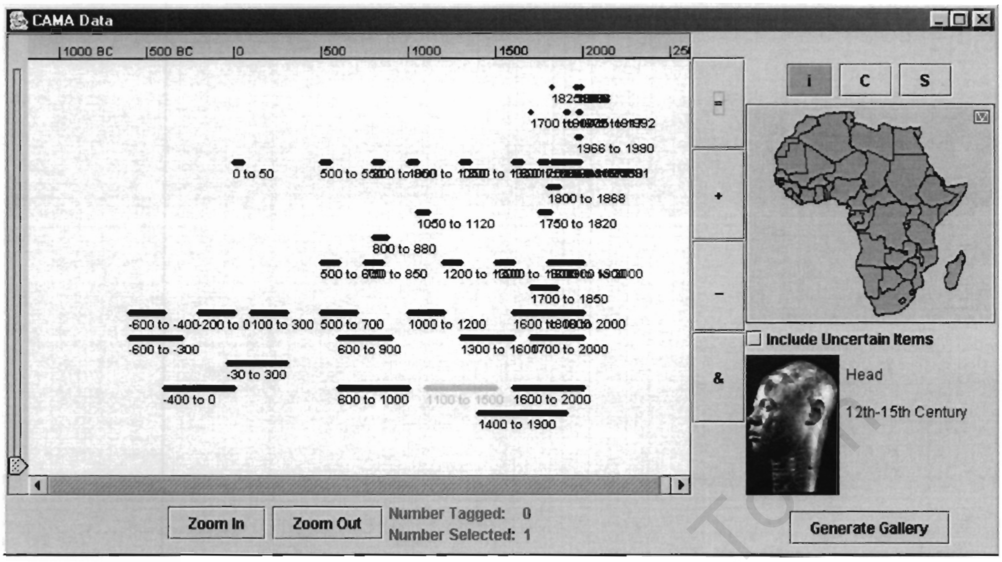
7. What do you think the purpose is of the vertical slider on the left of the display (the slider which moves up and down with associated horizontal dotted line?)

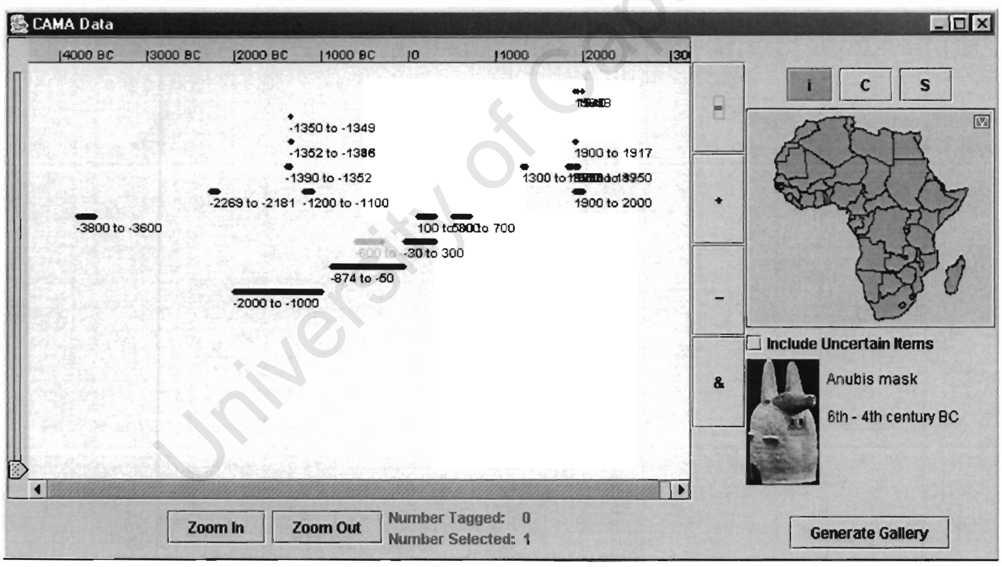
   ____

   ____

   ____

8. The two screen shots on the following page show the contents of two different archives. Tick the relevant answers:
   (a) Which archive contains more artefacts?
      i. The top archive
      ii. The bottom archive
      iii. It is not clear which archive has more artefacts
   (b) Which archive contains the oldest artefact?
      i. The top archive
      ii. The bottom archive
      iii. It is not clear which archive contains the oldest artefacts


CAMA Data
|1000 BC |500 BC |0 |500 |1000 |1500 |2000
i
C
S
Include Uncertain Items
Head
12th-15th Century
Zoom In
Zoom Out
Number Tagged: 0
Number Selected: 1
Generate Gallery


CAMA Data
|4000 BC |3000 BC |2000 BC |1000 BC |0 |1000 |2000
i
C
S
Include Uncertain Items
Anubis mask
6th - 4th century BC
Zoom In
Zoom Out
Number Tagged: 0
Number Selected: 1
Generate Gallery

Appendix D

Publications

The publications which relate to this research are included in the following few pages. These are:

1. K. Malan, G. Marsden, and E. Blake. Visual query tools for uncertain spatio-temporal data. In *Proceedings of the ninth ACM international conference on Multimedia*, pages 522—524. ACM Press, 2001.

2. G. Marsden, K. Malan, and E. Blake. Using digital technology to access and store African art. In *CHI 02 extended abstracts on Human factors in computer systems*, pages 528—529. ACM Press, 2002.

3. G. Marsden, K. Malan, Z. Hendricks and J. Tangkuampien. Interfaces to Digital Collections of African Art. *CHI-SA 2001* (Electronic), 2001.

4. Z. Hendricks, J. Tangkuampien, and K. Malan. Virtual galleries: is 3d better? In *Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa*, pages 17—24. ACM Press, 2003.

Bibliography

[1] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of CHI'94, ACM Conference on Human Factors in Computing Systems*, pages 313–317, 1994.

[2] C. Ahlberg, C. Williamson, and B. Shneiderman. Dynamic queries for information exploration: an implementation and evaluation. In *Proc CHI'92: ACM Conference on Human Factors in Computing Systems*, pages 619–626, 1992.

[3] G. Bennette, L. Radloff, and K. Malan. African music query system. Technical Report CS02-09-00, University of Cape Town, 2002.

[4] I. Bretan, R. Nilsson, and K. S. Hammarstrm. V: A visual query language for a multimodal environment. In *Proceedings: CHI 94*, pages 145–146. ACM, 1994.

[5] K. Brodlie. Visualization techniques. In K.W.Brodlie, L.A.Carpenter, R.A.Earnshaw, J.R.Gallop, R.J.Hubbold, A.M.Mumford, C.D.Osland, and P.Quarendon, editors, *Scientific Visualization: Techniques and Applications*, chapter 3, pages 37–85. Springer-Verlag, Berlin Heidelberg, 1992.

[6] P. A. Burrough. Natural objects with indeterminate boundaries. In Burrough and Frank [7], chapter 1, pages 3–28. UCT Library: DS 910 285 GEOG.

[7] P. A. Burrough and A. U. Frank, editors. *Geographic Objects with Indeterminate Boundaries*. Taylor & Francis, London, 1996. UCT Library: DS 910 285 GEOG.

[8] CAMA. Contemporary African Music and Arts Archive. http://www.cama.org.za.

[9] S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. Morgan Kaufmann, San Fransico, USA, 1999.

[10] S. K. Card, J. D. Mackinlay, and B. Shneiderman. Information visualization. In *Readings in Information Visualization: Using Vision to Think* [9], chapter 1, pages 1–34.

[11] S. K. Card, J. D. Mackinlay, and B. Shneiderman. Using vision to think. In *Readings in Information Visualization: Using Vision to Think* [9], chapter 7, pages 579–581.

[12] S. K. Card and P. Pirolli. The cost-of-knowledge characteristic function: Display evaluation for direct-walk dynamic information visualizations. In *Proc. ACM Conf. Human Factors in Computing Systems, CHI*. ACM, 1994.

[13] T. Catarci and G. Santucci. Query by diagram: A graphical environment for querying databases. In *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, Minneapolis, Minnesota, May 24-27, 1994*, page 515, 1994.

[14] H. Couclelis. Towards an operational typology of geographic entities with ill-defined boundaries. In Burrough and Frank [7], chapter 3, pages 45–55. UCT Library: DS 910 285 GEOG.

[15] M. Derthick, J. Kolojejchick, and S. F. Roth. An interactive visual query environment for exploring data. In *ACM Symposium on User Interface Software and Technology*, pages 189–198, 1997.

[16] ESRI GIS and Mapping Software. ESRI shapefile technical description. www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

[17] P. Fisher. Boolean and fuzzy regions. In Burrough and Frank [7], chapter 6, pages 87–94. UCT Library: DS 910 285 GEOG.

[18] R. Fitzpatrick. Strategies for evaluating software usability. Department of Mathematics, Statistics and Computer Science, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland, 1999. http://www.comp.dit.ie/rfitzpatrick/papers/chi99_strategies.pdf.

[19] N. Gershon. Visualization of an imperfect world. *IEEE Computer Graphics and Applications*, pages 43–45, July/August 1998.

[20] N. D. Gershon. Visualization of fuzzy data using generalized animation. In *Proceedings: Visualization '92*, pages 268–273. IEEE, 1992.

[21] A. Globus. Validation, verification and evaluation of visualization software: Position statement. Technical Report RNR-91-028, NASA Advanced Supercomputing Division, 1991. http://www.nas.nasa.gov/Research/Reports/Techreports/1991/PDF/rnr-91-028.pdf.

[22] A. Globus and S. Uselton. Evaluation of visualization software. *Computer Graphics*, May 1995.

[23] M. Hascoet-Zizi, C. Ahlberg, R. Korfhage, C. Plaisant, M. Chalmers, and R. Rao. Where is information visualization technology going? In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pages 75–77. ACM Press, 1996.

[24] Z. Hendricks, J. Tangkuampien, and K. Malan. Virtual galleries: is 3d better? In *Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa*, pages 17–24. ACM Press, 2003.

[25] L. Hesselink, F.Post, and J. van Wijk. Research issues in vector and tensor field visualization. *IEEE Computer Graphics and Applications*, 14(2):76–79, Mar. 1994.

[26] M. Y. Ivory and M. A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, 33(4):470–516, Dec. 2001.

[27] S. Jones. Dynamic query result previews for a digital library. In *Proceedings: Digital Libraries 98*, pages 291–292. ACM, 1998.

[28] S. Jones. Graphical query specification and dynamic result previews for a digital library. In *Proceedings: Symposium on User Interface Software and Technology*, pages 143–151. ACM, 1998.

[29] JShape. Java GIS. http://skyscraper.fortunecity.com/redmond/829/jshape2.htm.

[30] S. K. Lodha, A. Pang, R. E. Sheehan, and C. M. Wittenbrink. Uflow: Visualizing uncertainty in fluid flow. In *Proceedings: Visualization '96*, pages 249–254. IEEE, 1996.

[31] S. K. Lodha, C. M. Wilson, and R. E. Sheehan. Listen: Sounding uncertainty visualization. In *Proceedings: Visualization '96*, pages 189–195. IEEE, 1996.

[32] A. MacEachren, editor. *How Maps Work*. The Guilford Press, New York, 1995.

[33] G. Marsden, K. Malan, and E. Blake. Using digital technology to access and store african art. In *CHI '02 extended abstracts on Human factors in computer systems*, pages 528–529. ACM Press, 2002.

[34] B. Mirel. Complex queries in information visualizations: Distributing instruction across documentation and interfaces. In *Proceedings on the 17th annual international conference on Computer Documentation*, pages 1–8, 1999.

[35] V. Ndaba, M. Seotsanyana, K. Malan, and I. Webb. Positioning of non-specific data. Technical Report CS98-12-00, University of Cape Town, 1998.

[36] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. Technical report, University of California, 1996.

[37] C. Plaisant and V. Jain. Dynamaps: Dynamic queries on a health statistics atlas. In *Proceedings of CHI'94, ACM Conference on Human Factors in Computing Systems*, pages 439–440. ACM, Apr. 1994.

[38] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: Visualizing personal histories. In *CHI'96*, pages 221–227, 1996.

[39] C. Plaisant, B. Shneiderman, K. Doan, and T. Bruns. Interface and data architecture for query preview in networked information systems. In *ACM Transactions on Information Systems*, volume 17, pages 320–341. ACM, July 1999.

[40] B. Shneiderman. Dynamic queries: for visual information seeking. In *IEEE Software*, volume 11, pages 70–77. IEEE, Nov. 1994.

[41] Z. Stojanovic, S. Djordjevic-Kajan, and D. Stojanovic. Visual query and analysis tool of the object-relational gis framework. In *Proceedings of the ninth international conference on Information and knowledge management*, pages 328–335. ACM Press, 2000.

[42] H. Thimbleby, editor. *User Interface Design*. Addison-Wesley, 1990.

[43] E. L. Usery. A conceptual framework and fuzzy set implementation for geographic features. In Burrough and Frank [7], chapter 5, pages 71–85. UCT Library: DS 910 285 GEOG.

[44] C. Ware. *Information Visualization: Perception for Design*. Academic Press, USA, 2000.

[45] I. Webb, K. Malan, D. Cook, E. Blake, J. Turest-Swartz, and M. Wiese. Technical vision for cama, 1998. CAMA, Department of Computer Science, UCT.

[46] C. Williamson and B. Shneiderman. The dynamic homefinder: Evaluating dynamic queries in a real-estate information exploration system. In *ACM SIGIR*, pages 338–346. ACM, June 1992.

[47] D. Young and B. Shneiderman. A graphical filter/flow representation of boolean queries: A prototype implementation and evaluation. *Journal of the American Society for Information Science*, 44(6):327–339, July 1993.

[48] M. M. Zloof. Query by example. In *Proceedings of the AFIPS National Computer Conference*, volume 44, pages 431–438, 1975.