Extending Dynamic Queries to Handle Uncertain Data

Katherine Malan*, Gary Marsden†

*Department of Computer Science, University of Pretoria, Pretoria, 0002, South Africa
†Department of Computer Science, University of Cape Town, Private Bag, Rondebosch, 7707, South Africa

ABSTRACT

Dynamic querying is a technique which has been used successfully to enable novice users to gain access to and insight into data in databases. Some multimedia archives (such as archives of African art) contain data which have vague locations in time and space, that is, although there is some idea of when and where the entity originated, the precise information is unknown. This uncertainty creates problems with the display and querying of the data and so the data is generally not accessible to novice users. In this study we extend dynamic querying techniques to work with African art data with uncertain origins in time and space. We present methods for storing, visualising and querying such uncertain data within the framework of dynamic querying. Results of user tests indicate that our approach was clear to users and that users could successfully perform simple queries using the visual query tools. A similar approach of extending dynamic querying techniques could apply to other domains with any one-dimensional attribute data with probabilistic uncertainty. In this way we show how it is possible for novice users to query large databases with complex uncertain attributes.

KEYWORDS: dynamic queries, uncertain data, information visualisation, African art

1 INTRODUCTION

Dynamic queries combine the display and visual querying of data in a way that is accessible to novice users. In this study we applied these techniques to develop a front-end to an archive of African art. To use dynamic queries in this application, we had to develop new ways of working with uncertainty in data. In this section we explain dynamic queries, their use and benefits. We then describe the context of our study (African art) and the problems associated with data of this nature. Finally, we give an overview of the remainder of the paper.

1.1 Related work on dynamic queries

Ahlberg *et al.* [1] 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 [2].

The FilmFinder system [3] is one of the most well known applications that use dynamic queries. In this application films are displayed in a time/popularity space (Figure 1). Sliders can be used to formulate queries and the results are shown on the data display. For example, in Figure 1 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 to 276 min. The data display 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 [4]).

There have been many other applications that employ dynamic queries and the benefits of using this approach have been well established [1, 5, 6, 7, 2, 8]. Some of the benefits over traditional query systems are: it is quicker to express queries and novices learn to use the system quickly [8]; improvements in user performance and user enthusiasm [2]; and assisting users in finding trends and exception conditions [2, 6, 8]. Drawbacks of using dynamic queries include: the data needs to have some form of ordering and it is difficult to express more complex Boolean queries [8]. Some recent extensions to dynamic queries include: the creation of tangible query interfaces based on dynamic query techniques [9], dynamic queries over mobile objects [10] and new dynamic query tools, such as timeboxes (rectangular query regions drawn directly on 2D time series displays) [11] and box-shaped regions for querying 3D spaces [12].

Our aim was to use dynamic querying techniques to give novice users access to an archive of African art. In the following sections we describe the context of African art and the particular problems associated with data of this nature.

Email: Katherine Malan `kmalan@cs.up.ac.za`, Gary Marsden `gaz@cs.uct.ac.za`

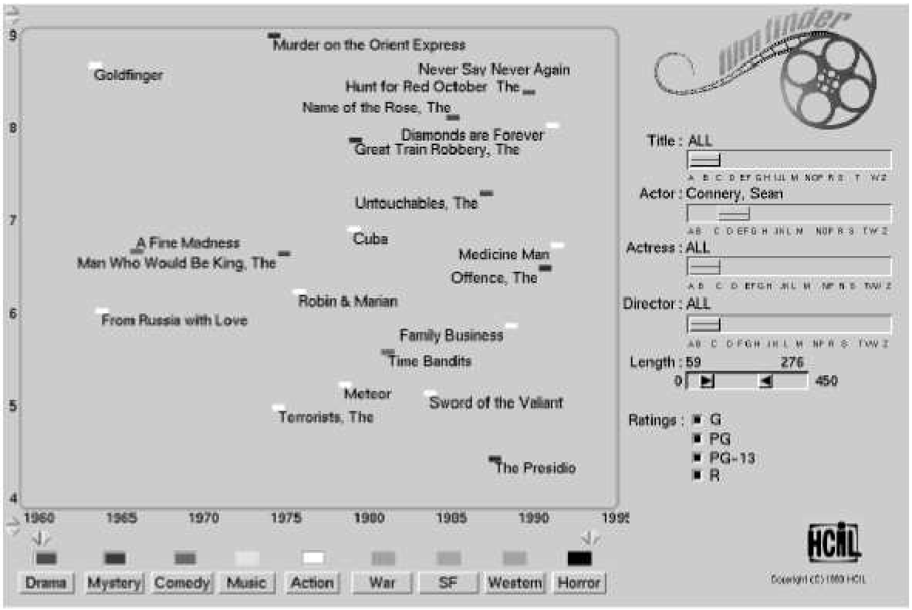

Figure 1: The FilmFinder system of Ahlberg *et al.* [3], showing how the selection has been narrowed down using dynamic query techniques

1.2 The context of African art

Africa's creative wealth of arts and culture is often under-acknowledged, both outside and inside the continent, partly because sufficient knowledge is not readily available. Data on traditional African art and culture, including physical artefacts, are dispersed throughout the world. Some is in Africa, owned by various 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 little known. Although a select few African artists and musicians have made a name for themselves, many talented artists are showcased neither outside, nor within the continent.

CAMA (Contemporary African Music & Arts Archive)[13] is an organisation that promotes African art and culture. One of its aims is to establish a continental multimedia database for African arts and culture and to make this archive accessible globally. The aim of our project was to develop query and visualisation tools for this archive. Typical users would include artists, historians or academics, most probably with limited expertise in databases. The front-end should therefore be accessible to novice users and use metaphors familiar to such users.

1.3 The problem statement

Our aim was to develop a front-end to the CAMA archive suitable for novice users. 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 such as data volumes and areas of coverage. It was therefore important for the application to provide an overview of the data and visual feedback on the results of queries. The final aim of the front-end system was to allow users to generate a virtual gallery containing only those artefacts that they wished to explore (for more information on this aspect of the project see [14]). The tool therefore needed to help users select a subset of artefacts based on sensible criteria from a potentially huge and ever growing archive.

Our original plan was use a map of Africa as a basis for the data display. This task posed several complications. The archive contains a mixture of traditional and contemporary art, which results in a wide range of types of data with different degrees of uncertainty. Some artefacts originate from a particular point on the map (such as a particular village). On the other hand, most traditional artefacts 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').

A further complication 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 has similar degrees of uncertainty to the geographic origin. Some artefacts are known to be made in a particular year (such as the artwork 'Illusions of Permanence', created by Bonita Alice in July, 1998); whereas in the case of most traditional artefacts, the origin is a period. These periods can span anything from a few years to a few centuries. For example, one artefact could be dated '1390–1352 BC', another dated 'late 19th Century' and yet another '6th–4th Century BC'.

Although some artefacts have definite origins in time and space, this does not make them any more important than those with uncertain origins. Some data elements which have uncertain origins are extremely important from a cultural perspective and it is therefore important that uncertain items be treated with the same level of importance as items which have definite origins.

Our aim was to develop creative ways of working simultaneously with both certain and uncertain data within the context of a visualisation and query tool for the archive.

1.4 Overview of the paper

In Sections 2, 3 and 4 we discuss our proposed approaches to storing, visualising and querying the data in a way that extends dynamic queries to work with uncertain data. In Section 5 we give the details of how we evaluated our proposed visualisation and query techniques. We finally conclude (Section 6) and provide some ideas for further work (Section 7).

2 STORING UNCERTAIN DATA

2.1 The nature of the uncertainty

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'. This form of uncertainty is a *range of uncertainty* (rather than *statistical* uncertainty or uncertainty originating from *error*)[15] and is most suitably modelled using *probability* rather than fuzzy sets[16]. By nature, each object in the archive 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' is equally likely to have originated in 1800 as 1899.

2.2 Data model to store uncertainty

The overall guiding principle we used when modelling the data was to store as much detail as was available. Later, this could be aggregated to less detail, as required. A simplified version of our proposed data model is shown in Figure 2. 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 to support comparative queries (e.g. -1352 and -1345 for the corresponding date qualifier: '18th Dynasty, Amarna period'). 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.* When the date is precise (i.e., known to within a year), the *from* and *to* fields store the same date.

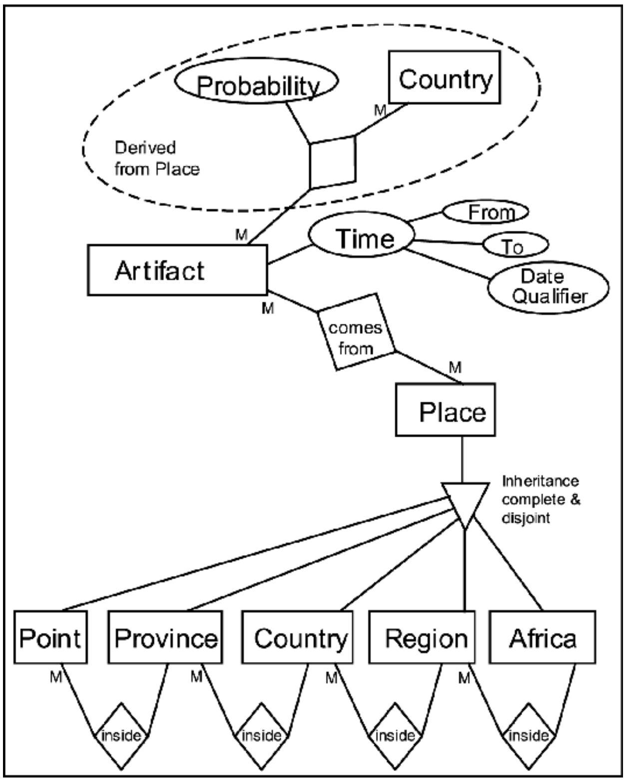

Figure 2: A simplified version of the entity-relationship diagram for the CAMA archive.

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. 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 (e.g. Ghana/Ivory Coast). For these cases, we derive and store a probability for the artefact and each country concerned. We do this because querying by geographic origin is most often done by country.

3 VISUALISING UNCERTAIN DATA

In this section we start by summarising the techniques others have used to visualise uncertain data. In Section 3.2 we explain work done in a previous study that attempted to visualise uncertain data on a geographic map, while in Section 3.3 we give the details of our proposed time-based visualisation of uncertain data.

3.1 Related work on 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. Gershon [17] 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).

A number of studies have been done in visualising uncertainty in scientific applications [15, 18, 19, 20]. For example, the use of uncertainty glyphs, animation and envelopes for visualising uncertainty in fluid flow [18]; the use of animation for visualising uncertainty in sea surface temperature data [19]; and the use of sonification for visualising geometric and fluid flow uncertainty [20].

In the case of scientific visualisation (as with the examples mentioned above), the aim is to visualise some underlying physical phenomenon. In our application, there is no such physical entity on which to base the information display. The term 'information visualisation' refers to the broader case where there is no underlying physical phenomenon – the aim is simply to visualise the data itself. The absence of a physical phenomenon (such as the human body, the earth or molecules) means that there is no obvious spatial mapping for the display [4]. Our challenge was to find an appropriate space in which to map the data with the associated uncertainty. We attempted two approaches: a geographic map-based visualisation and a time-based visualisation. These are discussed in the following two sections.

3.2 Map-based visualisation

Our initial plan was to use a geographic map to drive the interface of the system. This work formed part of a previous study [21] and is described briefly here.

We wanted to see if it was feasible to place items on a map when the origins of the items are uncertain. 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 a system where items were plotted as a point on the map at some random location which fell within the country of origin for the item. To test this approach, we demonstrated the system to users from a range of academic disciplines (geographers, computer scientists, psychologists and artists) after which they were interviewed. Although users' said that they understood that each of the icons (representing artefacts) was associated with the whole country and not a specific point, this conflicted with comments made later on in the interviews. Some 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 users had the perception that the location of the icon represented a precise origin.

Preliminary feedback from users therefore indicates that it could be misleading to place icons on a map if the precise 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 finding.

3.3 Time-based visualisation

Given the potential problems with visualising the origin of uncertain items on a geographic map we decided to investigate an alternative time-based visualisation. 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.

The approach we developed for simultaneously displaying both certain and uncertain data is illustrated in Figure 3. On the main data display, the horizontal axis denotes time (a timeline 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). Each item in the archive is represented as a bar on the display. The certain items are plotted close to the timeline, whereas the more uncertain items are plotted further away from the timeline. The length of each bar is used to indicate the level of uncertainty. Items which are certain are depicted as points (short bars), 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 3.1), 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.

4 QUERYING TECHNIQUES

In this section we describe dynamic querying techniques that we developed around the timeline-based display.

4.1 Querying based on time

When querying the archive, users would typically be interested in selecting items within a particular time range, such as 'the 18^{th} century' or '1650–1700'. To

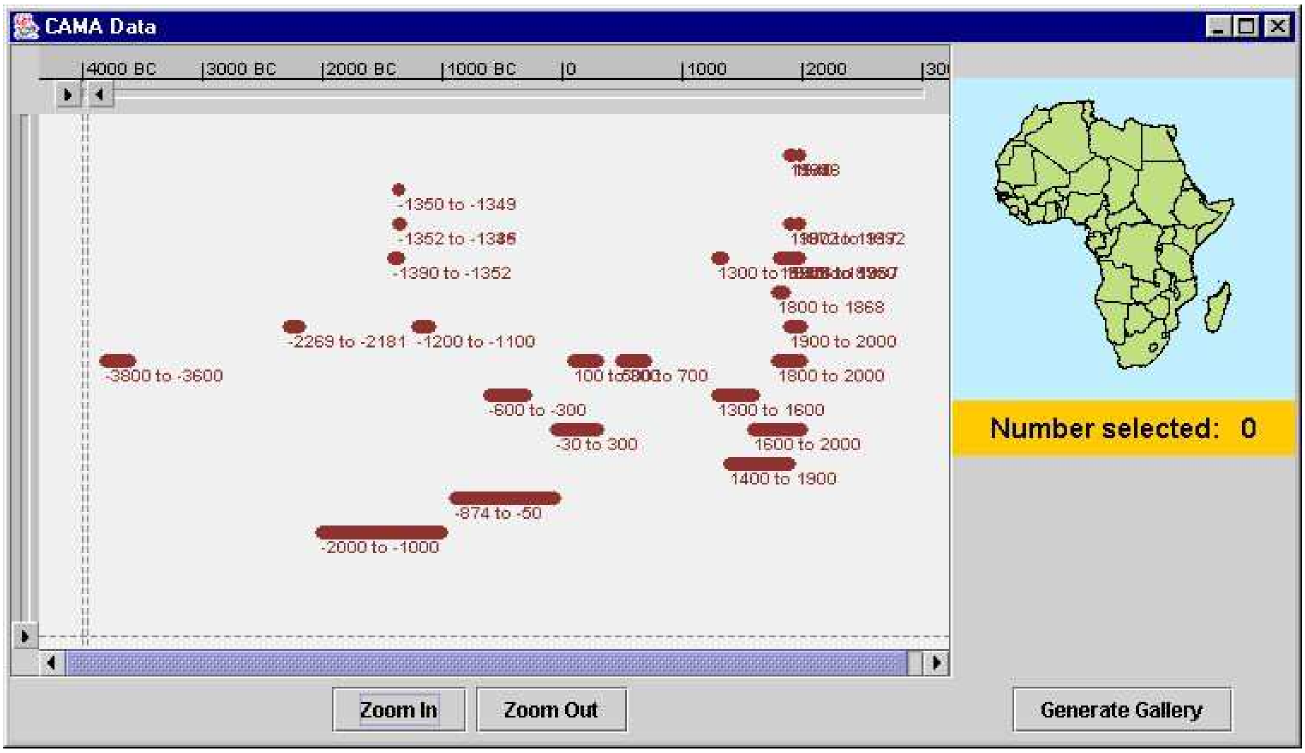

Figure 3: User interface of the prototype showing the main data display with timeline above the display.

support this form of querying in a visual way, we used double sliders on the timeline of the time map (Figure 4). Feedback is given in the form of vertical dashed lines. As the user drags one of the timeline buttons, the items which fall in the given range are selected.

An interesting question arises around the nature of ordering of time when uncertain items are involved. Does an item which is very uncertain in time (i.e. with a long uncertainty range) occur before an item which is very certain within the same range? For example, say one item (say item 1) has a time value of 1650 and another item (say item 2) has a time value of '17th Century' (i.e. it occurs somewhere in the range 1600–1700). Which of these items should come first in an ordering? Since we would like a query tool to return both certain and uncertain elements, we assume that item 2 is both *before* and *after* item 1 with respect to time. Items with uncertainty ranges which 'touch' either of the vertical dashed lines of the time slider are therefore selected. For example, in Figure 5 items 3 and 4 would be included in the selection as these items could possibly have originated from within the range of the query.

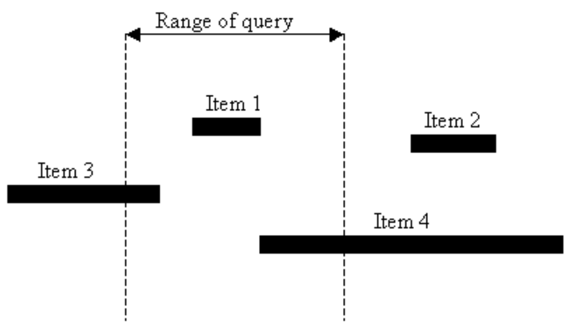

Figure 5: Given the range of a query, an item may fall completely inside the range (item 1), completely outside the range (item 2) or may fall partly inside the range (items 3 and 4). Items 1, 3 and 4 would be selected by this query.

4.2 Querying based on geographic origin

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 (including all artefacts which could *possibly* originate 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 6, 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.

4.3 Combination queries

The time range selector buttons can be used in conjunction with the map query tool to perform a union query. We also included a horizontal uncertainty slider to allow the user to exclude the more uncertain elements from the query. In Figure 7 the time query tool has been used to select artefacts in the period 2000BC–1000AD; the map query tool has been used

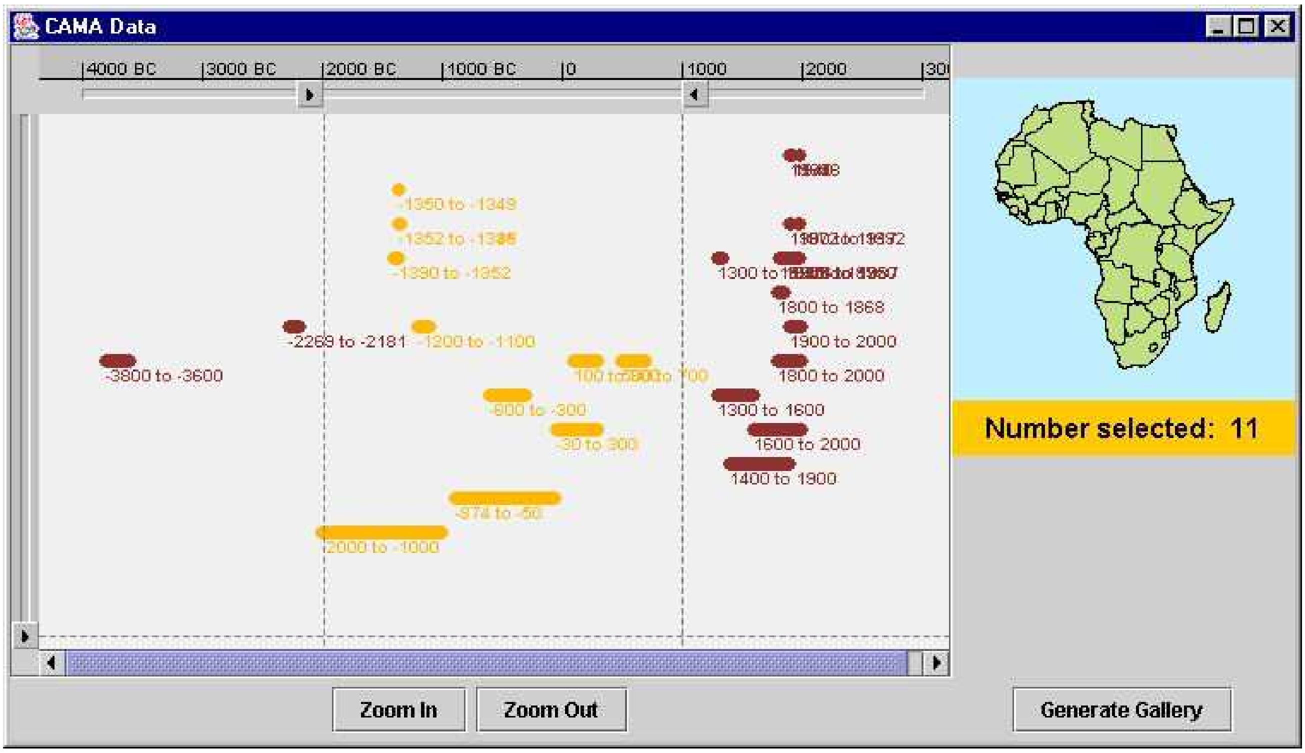

Figure 4: Double sliders on the timeline allowing the user to select items within a particular time range. The time range selectors have been used to select all artefacts originating in the period 2000BC to 1000AD.

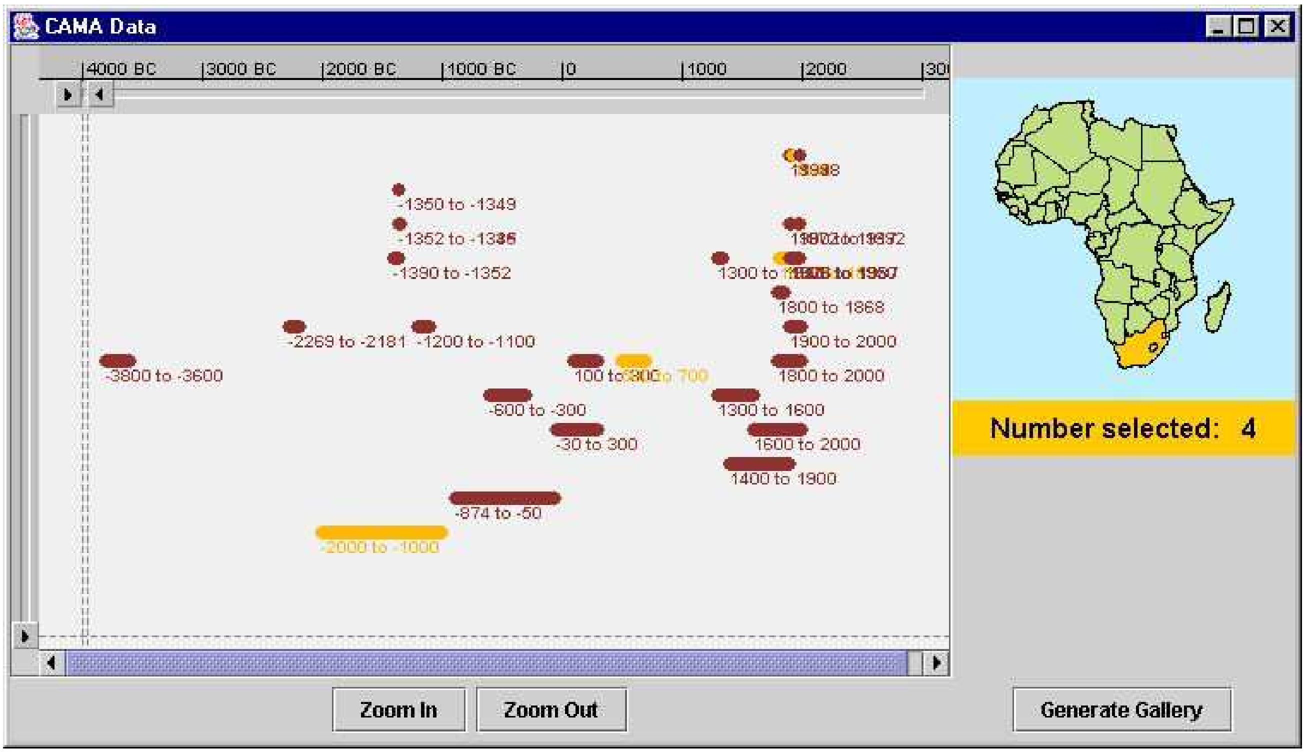

Figure 6: On the map, South Africa is highlighted in yellow. Artefacts originating from South Africa are highlighted in yellow on the display and the number selected is displayed below the map (on a yellow background)

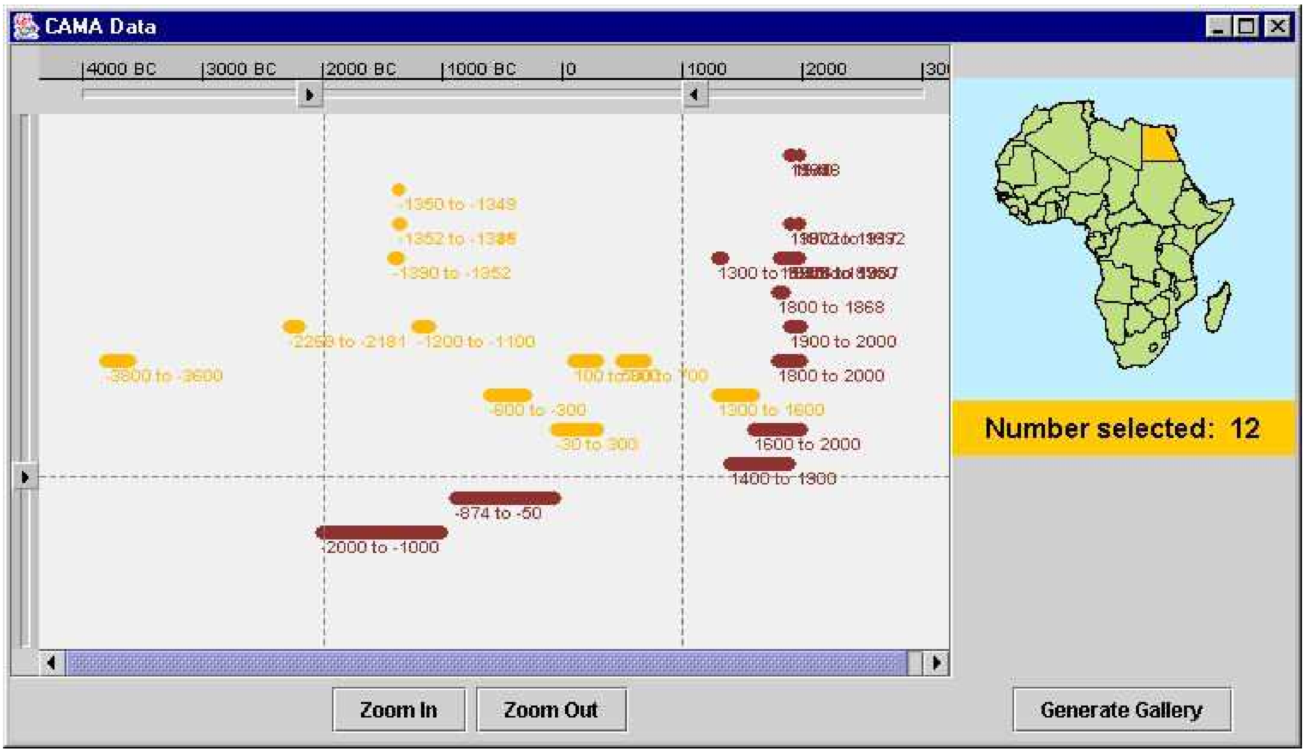

Figure 7: Selected artefacts include those originating in the period 2000BC to 1000AD as well as those originating from Egypt (highlighted in yellow on the map). The lower two artefacts have been excluded using the uncertainty slider.

to add artefacts originating from Egypt and the uncertainty slider has reduced the total number of selected items by two. In this way the user can generate a virtual gallery containing a manageable number of artefacts of interest for detailed browsing.

4.4 Probing

Clicking on a single bar on the display results in a pop-up window depicting an image of that particular artefact. Limited textual information associated with the artefact (including the time and geographic origin) is also displayed in the window (Figure 8). 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.

4.5 Querying on other dimensions

In this study we focused on two important dimensions of querying African art: time and location. There are, however, many other attributes on which to base visual queries, for example the type of art or name of artist. We believe that the proposed time-based framework can be easily extended to include additional tools for querying on other attributes. For example, the type of art could be handled in a similar way to the film genre of the Ahlberg *et al.* [3] FilmFinder system (see buttons at the bottom of Figure 1). In addition, the 'name of artist' attribute could be added as a query widget in a similar way to the Actor attribute (see panel on the right in Figure 1). We will not discuss these extensions further in this paper.

5 EVALUATION

We wanted to evaluate two aspects of our tool:

- Firstly, we wanted to test if the visualisation of uncertainty ranges effectively conveyed information on the archive in an easy-to-understand way.
- Secondly, we wanted to see if users could perform basic querying tasks using the query tools.

These two aspects are discussed in Sections 5.1 and 5.2 respectively.

5.1 Evaluation of time-based visualisation

To test the validity of our time-based visualisation we performed two user tests. The first test was an inquiry-based evaluation[22] using questionnaires to obtain feedback from users. The same questions were posed to a different set of users during a second testing phase using a prototype.

5.1.1 Aims

We wanted to measure if the representation of items on a time-map using uncertainty ranges was valid. In particular we wanted to see whether the following would be clear to users:

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.

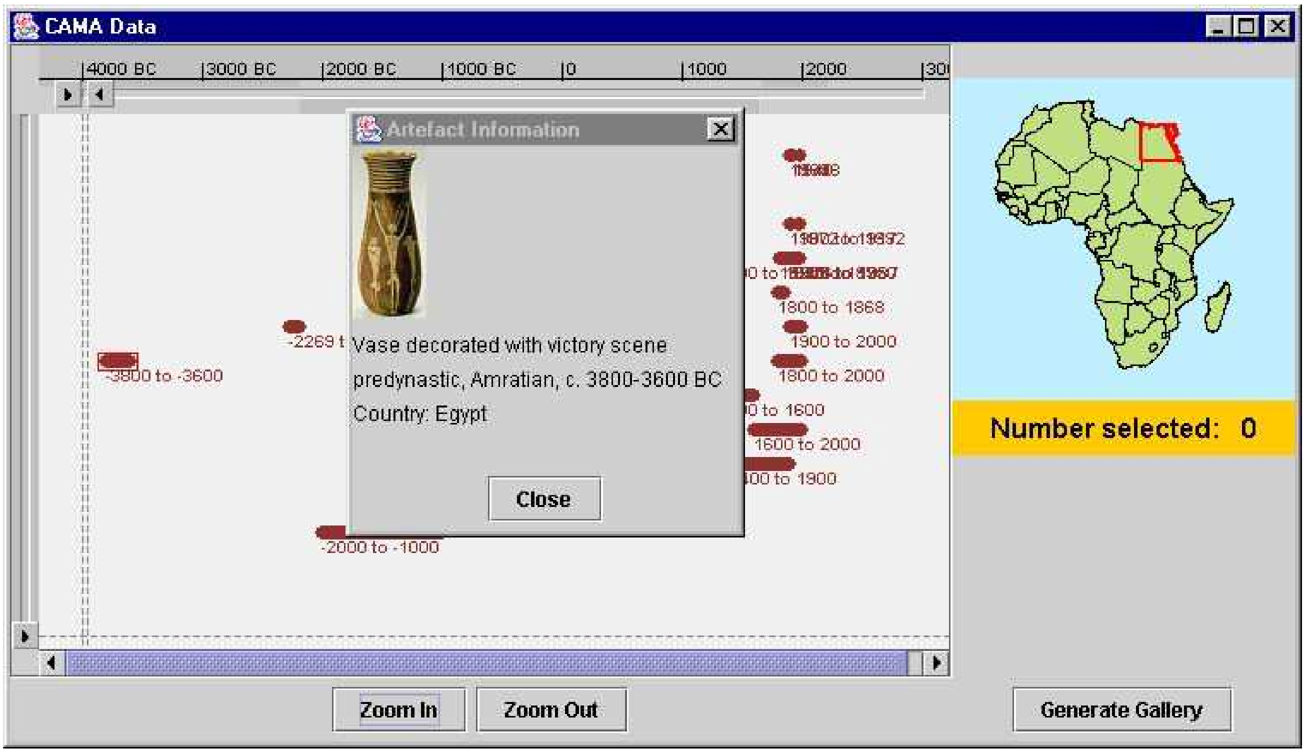

Figure 8: Popup window showing information on a particular artefact. The associated artefact is tagged with a red rectangle and the country of origin is outlined in red on the map.

3. A bar represents a time-frame within which that artwork could have been made (i.e. a range of uncertainty).

5.1.2 Subjects

For our initial questionnaire-based evaluation we chose subjects who could be seen as realistic users. Twelve people 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.

For the second phase prototype-based evaluation, 16 random people took part in the experiment (using a paired-user paradigm of usability testing, see Section 5.2.1). There was no overlap with the twelve people used in the initial testing phase.

5.1.3 Form of evaluation

In both testing phases, users where shown two screen dumps of the visualisation depicting the contents of two different (fictitious) archives of data (similar to Figure 3) – archive A and archive B. Archive A contained more artworks than archive B, whereas archive B contained fewer and older artworks than archive A.

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**

   This question tested 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**

   This question tested 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 were bars in the representation of Archive B that appeared earlier than 2000BC, whereas all the bars in Archive A occured 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 tested whether it was clear that a bar represents a range of uncertainty. The question was open-ended in order to elicit as many different interpretations as possible.

The subjects involved in the initial testing were given the following limited background information (in written form on the questionnaire):

- 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.

The subjects involved in the second phase prototype-based testing were given no background information except that the project related to an archive of African art. These subjects were, however, given time to explore the data by interacting with the prototype before answering the questions.

5.1.4 Results

In the initial questionnaire-based evaluation, 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. In the prototype-based evaluation all 8 user-pairs answered both question 1 and question 2 correctly.

From these results we deduce that it was clear to all user-pairs in the prototype-based evaluation that each artwork is represented by a bar and that the positioning of the bar is dependent on the time attribute of the artwork. Although the subjects in the initial questionnaire-based evaluation were given more background information, the results on these two questions were not as convincing. It seems that the second group of users developed a more accurate understanding of the data through interaction with the prototype.

The answers to question 3 (*What do you think it means if a bar in either one of the figures is longer than another bar*) have been grouped into categories and are shown in Tables 1 and 2.

The first group of responses (which we have called 'Uncertain') are what we originally thought of as the only 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, archeology, art history, history. (None of the users in the second testing phase had this background and gave this category of answer.) These people have a deeper understanding of where the estimate of time comes from. When presented with a time period such as '6^{th} to 4^{th} 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:

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

We have called the third category of answer to question 3 'longer period'. These responses are true, but do not give us enough information to determine whether users understood what this longer time period actually meant.

The only answers which we can confidently say are incorrect are the ones that state: "The period is longer / more artworks" and "more artefacts". These users thought that a longer bar represents more artworks than a shorter bar.

Combining the results from the two evaluations, we can summarize as follows: 11/20 answers were correct (either 'uncertain' or 'cultural era'), 7/20 answers were unclear (we do not know if they are correct or not) and 2/20 answers were incorrect (although through verbal comments between the users we could see that the user-pair later changed their understanding to a correct answer).

We can therefore say that most of the users understood that a bar represents a time-frame within which that artwork could have been made. We are satisfied that these results confirm the validity of the visualisation. Alternative representations of ranges of uncertainty should be tested to determine if even better results can be achieved.

5.2 Evaluation of query mechanisms

The aim of this evaluation was to determine whether a person with no knowledge of the project and without help could correctly perform basic querying tasks using the prototype. For the tests, 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.

5.2.1 Method

We designed a set of five short tasks for testing the effectiveness of the visual querying techniques. Sixteen people took part in the evaluation. None of these

Category	**Answers to Question 3:** (What do you think it means if a bar in either of the figures is longer than another bar.) [The profession of the user is shown in square brackets]
Uncertain: correct answer (3 answers)	– *There is no specific determined date on the artwork. It was made between date X & date Y* [Businessman] – *Express uncertainty of when an artefact was produced* [Computer Science lecturer] – *The artworks have been done between or over that period of time, it's unsure* [Graphic designer]
Cultural era: correct answer (4 answers)	– *Longer time span (the length of cultures)* [African art lecturer] – *Continuation of a 'tradition' a trend in the chronological sense* [Archaeologist] – *That the culture lasted longer* [Art History lecturer] – *It gives an indication of the chronological period and possible contents* [Historian]
Longer period: unknown (4 answers)	– *Longer time period* [Bookkeeper] – *It implies that 'the bar' (section) covers a longer time period* [Communications lecturer] – *It was over a longer period* [Secretary] – *A longer time period* [Web designer]
Incorrect (1 answer)	– *The period is longer / more artworks* [Marketer]

Table 1: Answers to question 3 grouped into categories from the initial questionnaire-based evaluation

Category	**Answers to Question 3:** (What do you think it means if a bar in either of the figures is longer than another bar.) [Author comments are shown in square brackets]
Uncertain: correct answer (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: unknown (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 written answer was, however, not changed.]

Table 2: Answers to question 3 grouped into categories from the second phase prototype-based evaluation

people had any prior knowledge of the project. There was no overlap between the subjects used for this evaluation and any prior user testing.

For the evaluation we employed the paired-user paradigm of usability testing (also known as 'codiscovery learning') [23]. In this approach pairs of users perform a set of tasks together at a single workstation. Users are encouraged to talk to each other and in this way the observer is able to capture what the users are thinking without the akwardness associated with individual thinking-aloud approaches to usability testing.

Each user-pair was required to complete 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.

5.2.2 Task description and results

In this section we discuss each task and how it was answered by all eight subject pairs.

Task 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 user-pairs performed this task without hesitation and all wrote down the correct answer (Egypt). In the case of two of the user-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 user even 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.

Task 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 user-pairs 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 users 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 correct answer 4.

Some of the users 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 user 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.

Task 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 user-pairs 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 users (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.

Task 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 (the default position before any user manipulation is with both time selector buttons to the left of the oldest artefact in the archive as shown in Figures 3 and 6) .

All eight user-pairs 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 user-pairs immediately used the buttons on the timeline to select a range, the other pairs tried other options before noticing the timeline slider buttons. Three user-pairs 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. The visibility of the time query buttons therefore needs to be improved. It is worth noting, however, that once users found the buttons, they had no problems in using them to formulate the query.

Task 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 user-pairs 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 task).

All users were therefore able to correctly use the time query tool with the map query tool to formulate a simple union query.

5.3 Discussion

We evaluated two aspects of our work:

- the time-based visualisation of data;
- the visual query mechanisms (using the time and map query tools).

Results showed that all users were able to correctly answer simple questions about the contents of the archive after interacting with the data through the prototype. In addition, most users understood that a bar on the display represents a timeframe within which an artwork could have originated. Although we feel the results confirm the validity of our approach, alternative representations should be developed and tested to determine if better results can be achieved.

Results of the evaluation of query mechanisms showed the following:

- Users easily located artefacts based on their time attribute.
- Users easily used the map query tool to select and unselect items based on geographic origin.
- Although all users easily manipulated the time query tools to select items based on time, some users took a while to find the time query tool buttons. The visibility of the time query tool should therefore be improved.
- Users easily combined the time and map query tools to perform union queries.

6 CONCLUSION

In this study we extended the concept of dynamic queries to manipulate data with uncertain attributes in time and space. We have shown how uncertain time-based data can be visualised on a two-dimensional time-map using the additional dimension to display uncertainty information. We have also shown how a subset of the data can be selected using a combination of a double-slider time query tool and a map query tool. Both of these tools were designed to include items in the selection which *could possibly* fall into the specified time range and/or geographic area.

We applied these techniques to an archive of African art. This archive contains a mixture of contemporary art (generally with certain origins in time and space) and traditional art (generally with uncertain origins in time and space).

Results of user tests show that these techniques are reasonably effective. Based on the visualisation, users were able to correctly answer simple questions relating to the contents of the archive. A small percentage of users misinterpreted the depiction of a range of uncertainty. Alternative representations should be developed to determine if a better solution can be found. Furthermore, users were able to effectively manipulate the visual query tools to perform queries based on time and geographic origin. The visibility of the time query tool is an aspect which needs to be improved.

A similar approach to visualisation and querying could apply to other domains with any one-dimensional attribute data with probabilistic uncertainty. Applications need not be restricted to historical data. For example, a possible application could be a database of project proposals with estimated cost ranges (in monetary terms based on worst case and best case scenarios). A visualisation based on cost could display degrees of financial uncertainty for different proposals. In this way, by combining dynamic querying with techniques for displaying and querying one-dimensional probabilistic uncertain data, we have shown how it is possible for novice users to query large databases with complex uncertain attributes.

7 FURTHER WORK

One area of further work lies in finding effective ways of visualising uncertain two dimensional (2D) data to support visual queries. In our application, we have visualised one dimensional (1D) uncertain data on a 2D time map. Initial user feedback indicated that visualising uncertain 2D data on a 2D geographic map was a problem. Further studies are needed to confirm this. Ideas which could be considered include:

- Using a 3D 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 2D 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.

Further work could include investigating the possibility of splitting the contemporary data (which is normally the certain data) and the traditional data (which is normally uncertain) into separate query tools. Tests would need to determine whether anything is lost by approaching the problem this way.

Another area of further work is in extending the dynamic query approach to express a wider range of combinations of queries. Our tool only catered for union queries. An example query such as:

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

requires the use of the intersection and difference Boolean operators. A possible approach to supporting such queries could include the use of an operator toolbar (such as the one proposed by Mirel [24]). The use of such a toolbar would require the notion of tagging elements before selection.

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.

ACKNOWLEDGMENTS

Our thanks to John Turest-Swartz and the rest of the CAMA team.

REFERENCES

[1] 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*, pp. 619–626. 1992.

[2] B. Shneiderman. "Dynamic Queries: for visual information seeking". In *IEEE Software*, vol. 11, pp. 70–77. IEEE, Nov. 1994.

[3] 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*, pp. 313–317. 1994.

[4] S. K. Card, J. D. Mackinlay and B. Shneiderman. "Information Visualization". In S. K. Card, J. D. Mackinlay and B. Shneiderman (editors), *Readings in Information Visualization: Using Vision to Think*, chap. 1, pp. 1–34. Morgan Kaufmann, San Fransico, USA, 1999.

[5] 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*, pp. 189–198. 1997. URL `citeseer.nj.nec.com/derthick97interactive.html`.

[6] 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*, pp. 439–440. ACM, Apr. 1994.

[7] 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*, vol. 17, pp. 320–341. ACM, Jul. 1999.

[8] C. Williamson and B. Shneiderman. "The Dynamic HomeFinder: Evaluating Dynamic Queries in a Real-Estate Information Exploration System". In *ACM SIGIR*, pp. 338–346. ACM, Jun. 1992.

[9] B. Ullmer, H. Ishii and R. J. Jacob. "Tangible query interfaces: physically constrained tokens for manipulating database queries". In *Proceedings of the 9th IFIP international conference on humancomputer interaction (INTERACT 2003)*. Sep. 2003.

[10] I. Lazaridis, K. Porkaew and S. Mehrotra. "Dynamic Queries over Mobile Objects". In *Proceedings of 8th International Conference on Extending Database Technology*, pp. 269–286. Mar. 2002.

[11] H. Hochheiser and B. Shneiderman. "Dynamic query tools for time series data sets: timebox widgets for interactive exploration". *Information Visualization*, vol. 3, no. 1, pp. 1–18, 2004.

[12] D. Akers, A. Sherbondy, R. Mackenzie, R. Dougherty and B. Wandell. "Exploration of the Brain's White Matter Pathways with Dynamic Queries". In *VIS '04: Proceedings of the conference on Visualization '04*, pp. 377–384. IEEE Computer Society, Washington, DC, USA, 2004.

[13] CAMA. "Contemporary African Music and Arts Archive". URL `http://www.cama.org.za`.

[14] 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*, pp. 17–24. 2003.

[15] A. T. Pang, C. M. Wittenbrink and S. K. Lodha. "Approaches to Uncertainty Visualization". Tech. rep., University of California, 1996.

[16] P. Fisher. "Boolean and Fuzzy Regions". chap. 6, pp. 87–94.

[17] N. Gershon. "Visualization of an Imperfect World". *IEEE Computer Graphics and Applications*, pp. 43–45, July/August 1998.

[18] S. K. Lodha, A. Pang, R. E. Sheehan and C. M. Wittenbrink. "UFLOW: Visualizing Uncertainty in Fluid Flow". In *Proceedings: Visualization '96*, pp. 249–254. IEEE, 1996.

[19] N. D. Gershon. "Visualization of Fuzzy Data Using Generalized Animation". In *Proceedings: Visualization '92*, pp. 268–273. IEEE, 1992.

[20] S. K. Lodha, C. M. Wilson and R. E. Sheehan. "LISTEN: Sounding Uncertainty Visualization". In *Proceedings: Visualization '96*, pp. 189–195. IEEE, 1996.

[21] V. Ndaba, M. Seotsanyana, K. Malan and I. Webb. "Positioning of Non-Specific Data". Tech. Rep. CS98-12-00, University of Cape Town, 1998.

[22] M. Y. Ivory and M. A. Hearst. "The State of the Art in Automating Usability Evaluation of User Interfaces". *ACM Computing Surveys*, vol. 33, no. 4, pp. 470–516, Dec. 2001.

[23] D. Wildman. "Getting the most from paired-user testing". *Interactions*, vol. 2, no. 3, pp. 21–27, 1995.

[24] B. Mirel. "Complex Queries in Information Visualizations: Distributing Instruction Across Documentation and Interfaces". In *Proceedings on the 17th Annual International Conference on Computer Documentation*, pp. 1–8. 1999.