Michael Eisner
Integrated Publication and Information Systems Institute (IPSI), Gesellschaft für Mathematik und Datenverarbeitung (GMD), Darmstadt, West Germany

The ARBOR Information System for Classical Archaeology and History of Art

Eisner, M.: **The ARBOR information system for classical archaeology and history of art.**
Int.Classif. 16(1989)No.3, p. 141-145, 9 refs.

Both domains divide scientific knowledge into object and method knowledge, with the former meaning knowledge of the organization of the individual real and normally complex research objects and the latter knowledge of the ways how to compare these objects. Object knowledge progresses stepwise from the object as a whole to its parts, subparts, etc. and can be visualized as an object-specific tree structure. ARBOR consists of a formal language able to represent textual object knowledge in a computer readable way. A PC-based implementation allows retrieval on the basis of ARBOR-coded object descriptions in different tree-structure-specifie query-modes. Author

1. Theoretical Background

Archaeological knowledge can be formally divided into object and method knowledge. The former consists of the knowledge of the concrete nature of the individual research objects, such as buildings, sculptures or pictures and is based on analysis. The latter means the knowledge about how to evaluate the object knowledge with the help of interdisciplinary methods, e.g. chronology, typology, stilistics, hermeneutics, statistics or text source criticism (as philology and history) and leads to historical knowledge as the synthesis. Object knowledge is based on individual observation and method knowledge on comparison. In addition to the factual archaeological knowledge described so far there also is reference knowledge, i.e. knowledge about previous publications on the topic at hand and about research history.

Archaeological knowledge is usually transferred over major spatial and temporal distances by means of printed publications consisting of text and illustrations. Usually a descriptive part, described as a "catalogue" if it covers several objects, serves for imparting (descriptive) object knowledge, while the (comparative) methodological knowledge is normally recorded in a "treatise". Mixed forms of these two also exist.

Viewed abstractly, method knowledge and its results form the specific contents of archaeology as a historical discipline, while object knowledge first of all presents quantitative and logistic problems. Not only that the number of objects found and more or less well published is very large and still increasing constantly, in addition, the acquisition of information about the objects is difficult, owing firstly to the broad scattering of the objects, and secondly to that of the publications. If we try to include the computer in archaeology as a scientific instrument we will first of all assign it the role of vehicle of object knowledge. This knowledge can initially only be coded textually, as it is difficult to base the processing of pictorial (or even better spatial) object information on the digitization of the now conventional recording procedures, which, like photogroaphy and drawing, produce two-dimensional results, but should be based on three-dimensional techniques, such as stereophotogrammetry, holography and tomography. It is only when this stage has been reached that the computer is useful for enriching archaeological method knowledge. First steps in this direction are already being taken (7) (8) (9).

Fig.1: The Korallion stele

Normally, the textual description of an archaeological or art historical object uses a terminological inventory for differentiating between the individual terminological recording levels of the objects. This terminological inventory is deduced in part from old text sources (historical authors, inscriptions), while in part it has also become established in the scientific world for no other reason than prolonged and uncontradicted use, providing a usable basis for communication. It can be said of a large number of objects, particularly in archaeology of the Mediterranean region and in European art history, that not only their morphological inventory – and thus the descriptive terminological inventory – is highly differentiated but that they also often carry representations which, in their turn, are arranged in a more or less complex fashion. We only need to think of mediaeval cathedral buildings as the structures containing altars and other pictures and reliefs. In the textual description of such objects we use a list of scientific terms which are in a hierarchically related to one another – reflecting the division of the object into various parts. This is illustrated briefly in the Greek stele of Korallion of the Kerameikos Cemetery in Athens (Fig.1, acc. (1)) which was produced around or soon after the middle of the fourth century B.C. First it has an architectonically formed frame, called a "naiskos", of lateral pilasters and a entablature with a pediment, with the entablature bearing the inscription. The relief area shows a woman sitting on a stool with her feet on a footstool. Behind her, half covered, another woman stands, further to the right two men. Behind the legs of the sitting woman we can see the head of a dog. The transformation of the pictorial representation into a division into various parts formulated in technical language yields, not surprisingly, a tree structure (Fig.2) of the descriptive terms (3) (4) (5). In this context it appears most important to point out that the characteristic descriptive tree structure proves to be individually, i.e. dynamically, formed for each object. The same objects produce the same description trees, while more or less different ones produce trees deviating from one another. As identical objects of complex structure are extremely rare, comparability plays a decisive role in research at the detail level. It must also be maintained in the textual description of the object.

We hardly need to mention that the classical data acquisition structures in the field of database models (like the relational model, the hierarchical model or the network, cf. (2)) are invariant after the moment of their definition and therefore cannot be used in the afore described manner. It is only a poor consolation in an age in which the archaeologist or art historian would like to buy a microcomputer and use it as an aid in his work – for instance for setting up a textual object knowledge bank – that, by admitting pointer fields, finally everything becomes representable in every model. It is with some right that he can expect a user-friendly interface rather than a solution (e.g. in the form of a quantity of data relations) presupposing a considerable analytical ability in applying computer science, an ability which computer specialists possess, while art scientists usually do not.

For the representation of hierarchically structured object descriptions, therefore, another way was selected than the conversion into firm data acquisition structures, namely a formal language appropriately called ARBOR. It consists of a text which differentiates between (later retrievable) "descriptors" and (later non-retrievable) "commentaries". The two language elements can be mixed as desired. A respective marking serves for recognizing the descriptors. An ARBOR text is divided into "documents", with each individual document containing the description of a single research object. The extent of a document is limitless and may contain as many descriptors as desired so that even complex objects can be described. If the division into various parts so requires, document parts can be made accordingly. Data concerning the object as a whole form the beginning so that this document part is described as the "header". Data on the individual part quantities of the object each form a "subdocument", which is introduced by a "contextor" expressing the degree of direct or indirect dependence on the header. The individual document parts are likewise of unlimited extent.

2. Implementation

Implementation as actually applied requires an IBM PC or a compatible computer using PC-(MS-)DOS. (For more detailed information about this version and the theoretical background of ARBOR see (6)). Two programs have been realized. ASU (Arbor-Set-Up) and ART (Arbor-ReTrieval). ASU reads in an ASCII data file (created by means of an editor) with the ARBOR text, which normally consists of a quantity of documents separated by empty lines. The sign "*" (asterisk) presently serves as descriptor marking and the sign "-" (hyphen) as contextor element (Fig.3). The header fills the type area, if possible, throughout its entire width, thus beginning at the front left. Single hyphens introduce subdocuments of the first order (i.e.,those depending directly on the header), two hyphens those of the second order (i.e. not depending directly on the header but on a subdocument of the first order), etc., while dependencies of the fourteenth order currently form the limit. ASU produces first a direct access data file of the ARBOR text which during retrieval is used for showing the documents found on the screen. In addition, a table of the individual descriptor and an internal representation of the contextor are set up

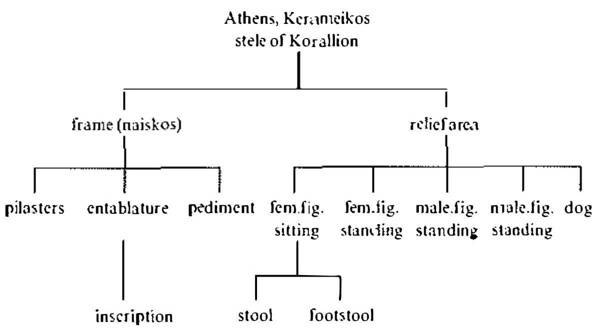


Fig.2: Tree structure describing the Korallion stele

```
*Athens, *necropolis of *Kerameikos. *stele of *Korallion, with *relief
-*frame (*naiskos)
--one *pilaster at each side
--*entablature
---*inscription
--*pediment
-relief *area
--human *figure, *female, *sitting
---*stool
---*footstool
--human *figure, *female, *standing on the left
--human *figure, *male, *standing in the center
--human *figure, *male, *standing on the right
--*dog
```

Fig.3: ARBOR version of the Korallion stele description

DESCRIPTOR	CONTEXTOR
Athens	135.0.0.0.0.0.0.0.0.0.0.0
necropolis	135.0.0.0.0.0.0.0.0.0.0.0
Kerameikos	135.0.0.0.0.0.0.0.0.0.0.0
stele	135.0.0.0.0.0.0.0.0.0.0.0
Korallion	135.0.0.0.0.0.0.0.0.0.0.0
relief	135.0.0.0.0.0.0.0.0.0.0.0
frame	135.1.0.0.0.0.0.0.0.0.0.0
naiskos	135.1.0.0.0.0.0.0.0.0.0.0
pilaster	135.1.1.0.0.0.0.0.0.0.0.0
entablature	135.1.2.0.0.0.0.0.0.0.0.0
inscription	135.1.2.1.0.0.0.0.0.0.0.0
pediment	135.1.3.0.0.0.0.0.0.0.0.0
area	135.2.0.0.0.0.0.0.0.0.0.0
figure	135.2.1.0.0.0.0.0.0.0.0.0
female	135.2.1.0.0.0.0.0.0.0.0.0
sitting	135.2.1.0.0.0.0.0.0.0.0.0
stool	135.2.1.1.0.0.0.0.0.0.0.0
footstool	135.2.1.2.0.0.0.0.0.0.0.0
figure	135.2.2.0.0.0.0.0.0.0.0.0
female	135.2.2.0.0.0.0.0.0.0.0.0
standing	135.2.2.0.0.0.0.0.0.0.0.0
figure	135.2.3.0.0.0.0.0.0.0.0.0
male	135.2.3.0.0.0.0.0.0.0.0.0
standing	135.2.3.0.0.0.0.0.0.0.0.0
figure	135.2.4.0.0.0.0.0.0.0.0.0
male	135.2.4.0.0.0.0.0.0.0.0.0
standing	135.2.4.0.0.0.0.0.0.0.0.0
dog	135.2.5.0.0.0.0.0.0.0.0.0

Fig.4: Table of ARBOR descriptors and contextors (we assume that the current document is the 135th in the respective ARBOR file's sequence)

(Fig.4) which contain the serial numbers of the respective ARBOR document in the data file and a field of fourteen bytes with the description of the path of the description tree. Here all successors dependent on the same predecessor in the tree structure or the root or document number are given a number from 1 to 255, while the field addresses on 0 symbolize unoccupied or non-existent nodes. In this kind of path description, the contextor of a hierarchically subordinated descriptor can always be recognized by the fact that it contains the contextor of a hierarchically superordinated descriptor. Conversely, hierarchically superordinated contextors are contained in subordinated ones. Contextors of descriptors describing the same node in the tree structure are the same. The elements of the table, descriptor and contextor are managed in the same index-sequential (ISAM) data file.

The ART retrieval program permits in several steps the querying for one or more (alternative) descriptors. Here, first a primary hitlist is produced showing the number of hits. This number can then be narrowed down again and again according to eleven different search modes which permit searching – with differing weightings – in super– or subordinated contexts or in the same document part, in the header, in neighboring contexts (and possibly also in their successors) as well as, finally, completely independent of the hierarchical structure of the document (Fig.5). Documents with hits can be displayed or printed out at every retrieval stage.

ARBOR knows not only "textual", but also "named numerical" descriptors, which consist of a domain name (as identifier), a separator and either one value (able to represent exact data, e.g. "length = 15.3") or two values (giving a data range, e.g. "height = 8.0..9.0"). These values can be either of integer or of real type. The representation of numerical ranges has special importance in a science in which inexact data are very common (e.g. the assumed dating of an object in the period between 450 B.C. and 425 B.C. may be described as "ChronDate = -450..-425"). The retrieval of numerical data can be retrieved by asking in the same way, i.e. for exact values or for ranges. In the latter case all documents with named numerical descriptors completely fitting the interval searched will be considered as hits. Retrieval of textual descriptors allows right-side truncation. After setting up a primary hitlist it is possible to exclude documents with certain textual or numerical descriptors.

Normally the vocabulary of an ARBOR database should be controlled by a thesaurus. Actually the latter acts only as a list of allowed descriptors. In the future abstraction hierarchies will be possible in order to find documents by searching for more generic terms in relation to the (textual) descriptors used in the single ARBOR documents. Some of the items of information describing archaeological objects are very difficult to verbalize, e.g. the typical artistic or workmanlike aspects normally denominated as "style". In these cases a medium allowing the synchronous visualisation of retrieval results would be very convenient, e.g. a picture-managing device producing presentations of digitized object images. The first step in this direction will be to transport ARBOR software from the PC-(MS-)DOS-world to a more powerful system environment like such as UNIX. Currently this is being put into practice.

As ARBOR can be considered the beginning of an "intelligent" picture archive manager it was integrated in 1988 into the PAVE project (Publication and Visualisation Environment) of GMD's IPSI department at Darmstadt.

References

(1) Conze, A.: Die attischen Grabreliefs. Berlin: 1893-1922. pl.98

(2) Date, C.J.: An introduction to database systems. Reading, MA: 1977 (2nd ed.), p.51-67

(3) Eisner, M.: Zur Rolle von Datenbanken als Instrument kunsthistorischen und archäologischen Fachwissens. In: Corti, L.(Ed.): Automatic processing of art history data and documents. Pisa: Scuola Normale Superiore, Sept.24-27, 1984. Firenze 1984. p.325-329

(4) Eisner, M.: Zur Bedeutung der Datenverarbeitung als Hilfsmittel der Archäologie unter besonderer Berücksichtigung von Datenbanken. In: Acta Praehistorica et Archaeologica 16/17, 1984/1985, p.278-285

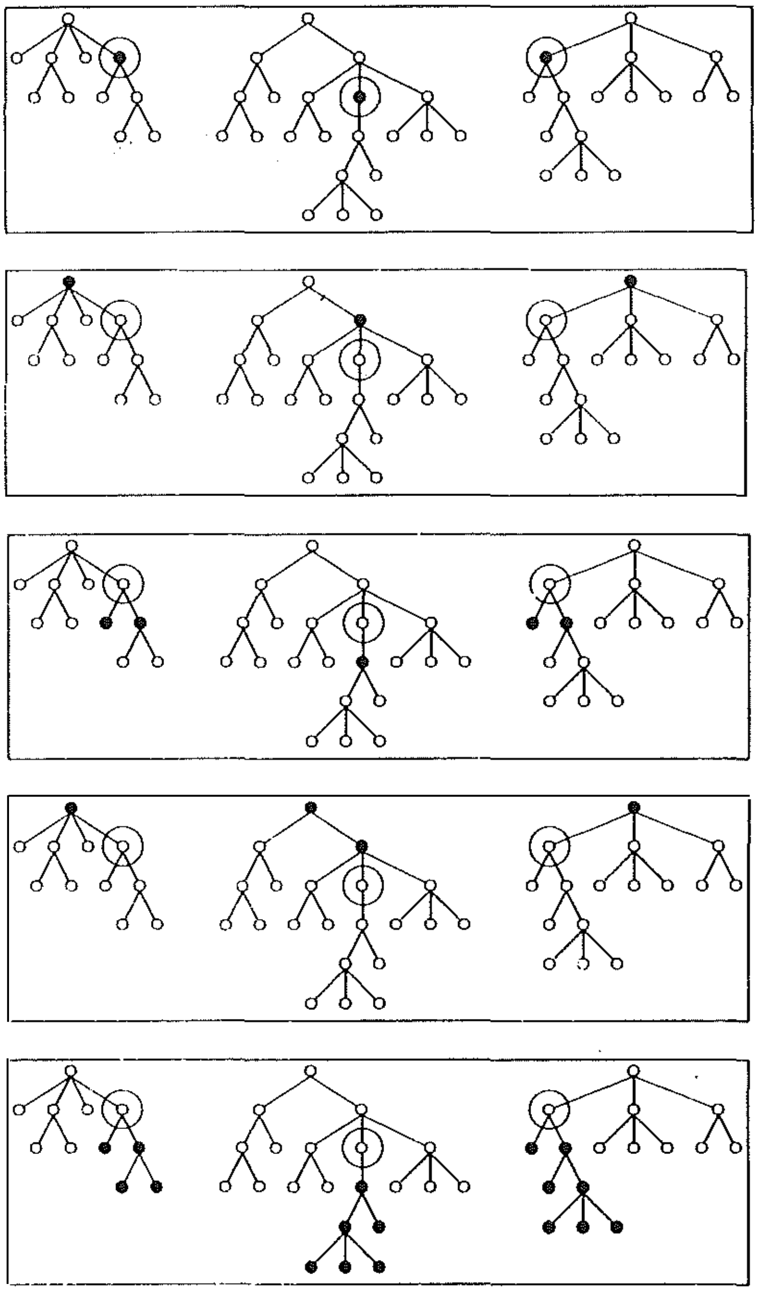
(5) Eisner, M.: Zukunftsperspektiven archäologischer Wissensdarstellung auf dem Computer. In: Bathron. Beiträge zur Architektur und verwandten Künsten. Für Heinrich Drerup zu seinem 80. Geburtstag. Saarbrücken 1988, p.115-120

(6) Eisner, M.: ARBOR. Eine Sprache zur Beschreibung und ein Programmpaket zur Verarbeitung hierarchischer Datenobjekte der klassischen Archäologie und Kunstgeschichte. Gesellschaft für Mathematik und Datenverarbeitung. St.Augustin 1989. GMD-Studien Nr.159

(7) Kampffmeyer, U., Rupprecht, G., Witteyer, M.: ARCOS: Ein Computer zeichnet römische Keramik. Die Ergebnisse der Testuntersuchungen mit dem ARCOS 1 im Landesamt für Denkmalspflege, Abt.Archäologische Denkmalspflege, in Mainz. Mainzer Zeitschrift 81(1986)p.191-200

(8) Kampffmeyer, U.et al: Untersuchungen zur rechnergestützten Klassifikation der Form von Keramik. Frankfurt (Main) 1987. p.79-104

(9) Mazzola, G., Krömker, D., Hoffmann, R.: Rasterbild -- Bildraster. Anwendungen der Graphischen Datenverarbeitung zur geometrischen Analyse eines Meisterwerkes der Renaissance: Raffaels "Schule von Athen". Berlin(West) 1986.



narrowing mode 1
(searching in primary context)

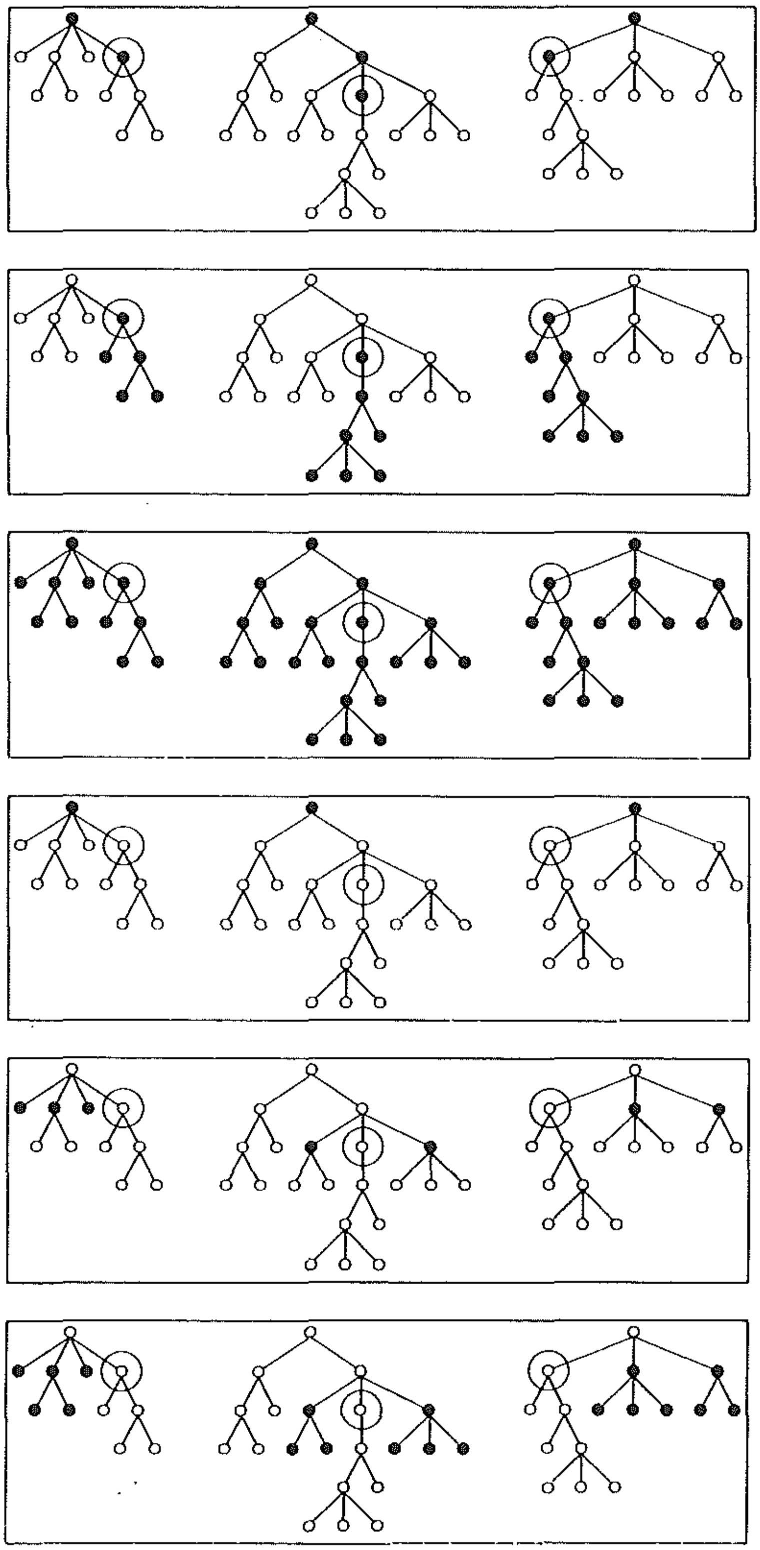
narrowing mode 2
(searching in next supercontext)

narrowing mode 3
(searching in next subcontexts)

narrowing mode 4
(searching in supercontexts)

narrowing mode 5
(searching in subcontexts)

Fig.5A: Hitlist narrowing modes 1 to 5 (every tree structure represents a document; nodes marked with a circle mean document parts containing primary hits; nodes marked black mean document parts in which retrieval for secondary hits takes place)



narrowing mode 6
(searching in primary context and supercontexts)

narrowing mode 7
(searching in primary context and subcontexts)

narrowing mode 8
(searching in the same document)

narrowing mode 9
(searching in root context)

narrowing mode 10
(searching in adjacent contexts)

narrowing mode 11
(searching in adjacent contexts and their subcontexts)

Fig.5B: Hitlist narrowing modes 6 to 11 (cf. fig. 5A)