THE ARBOR INFORMATION SYSTEM FOR CLASSICAL ARCHAËOLOGY AND HISTORY OF ART

1. Theorical 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 critique (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 until now still comes the reference knowledge, that is the knowledge about the previous publications on the topic at hand and about the 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, which is 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 content 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, firstly owing to the broad scattering of the objects, and secondly of the publications. If we try to include the computer in archaeology as a scientific process we will first of all assign it the role of vehicle of object knowledge. This can first of all be coded only textually as it is difficult for the processing of pictorial (or even better spatial) object information to be based on the digitization of the nowadays conventional recording procedures which are two-dimensional in their results, such as photography and drawing, but should be based on three-dimensional techniques, such as e.g. stereophotogrammetry, holography and tomography. It is only when this stage has been reached that the computer is useful for enriching the archaeological method knowledge. The first steps in this direction are already taking place (Kampffmeyer, Rupprecht, Witteyer

1986; KAMPFFMEYER *et al.* 1987; MAZZOLA, KRÖMKER, HOFFMANN 1986).

Normally, the textual description of an archaeological or art historical object uses a terminological inventory for differentiating the individual terminological recording levels of the objects. This terminological inventory is deduced in part from old text sources (historical authors, inscriptions), in part it has also established itself in the scientific world only by means of long-lasting and uncontradicted use. Thus, communication can take place on its basis. It applies to a large numer of objects, particularly for archaeology of the Mediterranean region and in European art history, that not only their morphological inventory — and thus the describing 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 hierarchical relation to one another — reflecting the division of the object into various parts. This is illustrated briefly in the Greek relief of the Korallion of the Kerameikos Cemetery in Athens (Fig. 1) which was made 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 yeld is, as we can only expect, a tree structure (Fig. 2) on the describing terms (EISNER 1984; EISNER 1984/85; EISNER 1988). 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 which deviate from one another. As the same objects of complex structure are extremely rare, in research comparability plays a decisive role 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, the hierarchical or the network, cf. DATE 1977) are invariant after the moment of their definition and therefore cannot be used in such a 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 be-

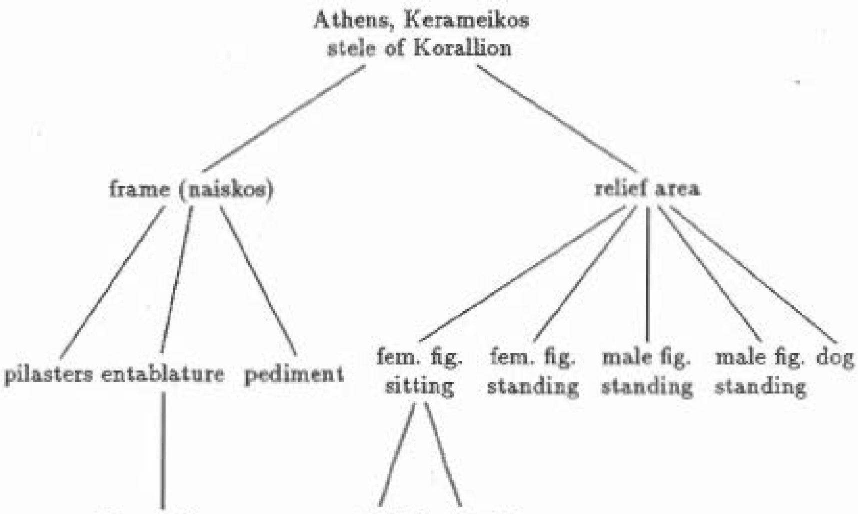


Fig. 1 — The Korallion stele (A. CONZE, *Die attischen Grabreliefs*, Berlin 1893-1922, pl. 98).
Fig. 2 — Tree structure describing the Korallion stele.

comes representable in every model. It is with some right that he can expect a user-friendly interface, and not a solution (such as in the form of a quantity of data relations) which presupposes a considerable analytical ability as by means of computer science, which the person working primarily as a computer specialist has, but hardly the art scientist possesses.

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 an 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 objects 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" which expresses the degree of direct or indirect dependence on the header. The limitlessness of the extent also applies to the individual document parts.

2. Implementation

The actually existing implementation requires an IBM PC or a compatible computer using PC-(MS-)DOS. (For more detailed information about this version and the theorical background of ARBOR see Eisner 1989). Two programs have been made. ASU (Arbor-Set-Up) and ART (Arbor-ReTrieval). ASU reads in an ASCII data file made 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 complete width, thus it begins at the front left. Single hyphens introduce subdocuments of the first order (that is those which depend directly on the header), two hyphens those of the second order (which do not depend directly on the header but on a subdocument of the first order) etc., while dependencies of the fourteenth order presently 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 contex-

```
*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
```

DESCRIPTOR	CONTEXTOR
Athens	135.0.0.0.0.0.0.0.0.0.0
necropolis	135.0.0.0.0.0.0.0.0.0.0
Kerameikos	135.0.0.0.0.0.0.0.0.0.0
stele	135.0.0.0.0.0.0.0.0.0.0
Korallion	135.0.0.0.0.0.0.0.0.0.0
relief	135.0.0.0.0.0.0.0.0.0.0
frame	135.1.0.0.0.0.0.0.0.0.0
naiskos	135.1.0.0.0.0.0.0.0.0.0
pilaster	135.1.1.0.0.0.0.0.0.0.0
entablature	135.1.2.0.0.0.0.0.0.0.0
inscription	135.1.2.1.0.0.0.0.0.0.0
pediment	135.1.3.0.0.0.0.0.0.0.0
area	135.2.0.0.0.0.0.0.0.0.0
figure	135.2.1.0.0.0.0.0.0.0.0
female	135.2.1.0.0.0.0.0.0.0.0
sitting	135.2.1.0.0.0.0.0.0.0.0
stool	135.2.1.1.0.0.0.0.0.0.0
footstool	135.2.1.2.0.0.0.0.0.0.0
figure	135.2.2.0.0.0.0.0.0.0.0
female	135.2.2.0.0.0.0.0.0.0.0
standing	135.2.2.0.0.0.0.0.0.0.0
figure	135.2.3.0.0.0.0.0.0.0.0
male	135.2.3.0.0.0.0.0.0.0.0
standing	135.2.3.0.0.0.0.0.0.0.0
figure	135.2.4.0.0.0.0.0.0.0.0
male	135.2.4.0.0.0.0.0.0.0.0
standing	135.2.4.0.0.0.0.0.0.0.0
dog	135.2.5.0.0.0.0.0.0.0.0

Fig. 3 — ARBOR version of the Korallion stele description.
Fig. 4 — Table of ARBOR descriptors and contextors (we assume that the current document is the 135th in the respective ARBOR file's sequence).

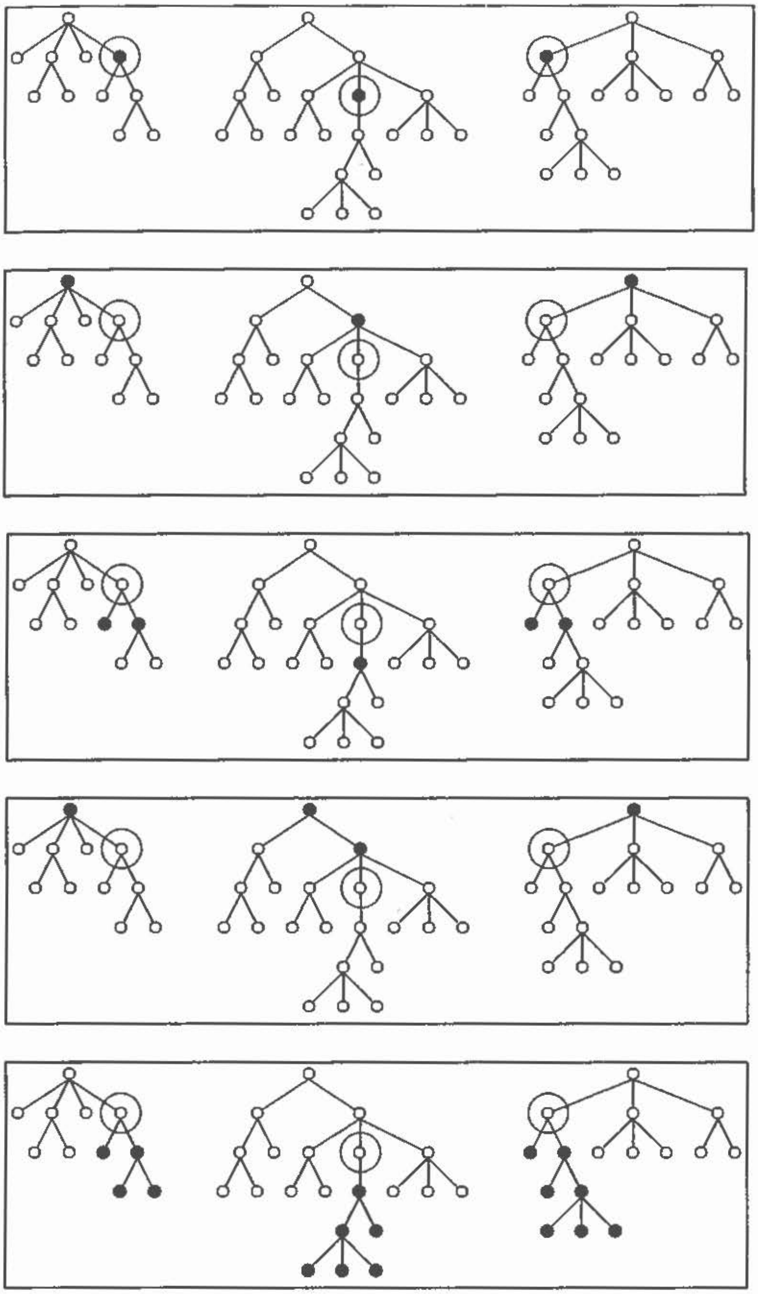
tor are set up (Fig. 4) which contains the serial numbers of the respective ARBOR document in the data file and a field of fourteen bytes with the description of the path of the description tree. Here all successors dependent on the same predecessor in the tree structure or the root or document number 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. Contextor 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 retrieval program ART permits in several steps the query for one or more (alternative) descriptors. Here, first a primary hitlist comes into being giving the number of hits. This can then be restricted again and again and this according to eleven different search modes which permit the search — with differing weightings — in super-or subordinated contexts or in the same document part, in the header, in neighbouring 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. The latter consist of a domain name (as identifier), a separator and one value (able to represent exact data, f.i. "Lenght = 15.3") or two values (giving a data range, f.i. "Height = 8.0..9.0"). The 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 (f.i. 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 numeric data allows to ask in the same way, i.e. for exact values or for ranges. In the last case all documents with named numerical descriptors fitting completely to the searched interval 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 works 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 informations which describe archaeological objects are very



narrowing mode 1
(searching in primary context)

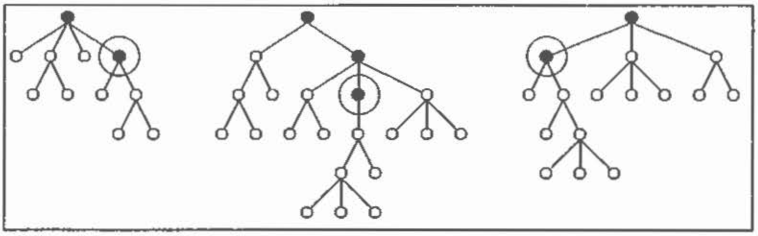
narrowing mode 2
(searching in next supercontext)

narrowing mode 3
(searching in next subcontexts)

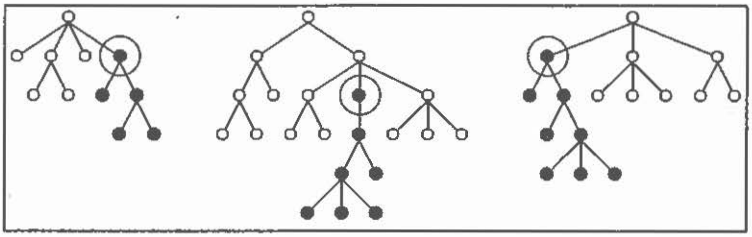
narrowing mode 4
(searching in supercontexts)

narrowing mode 5
(searching in subcontexts)

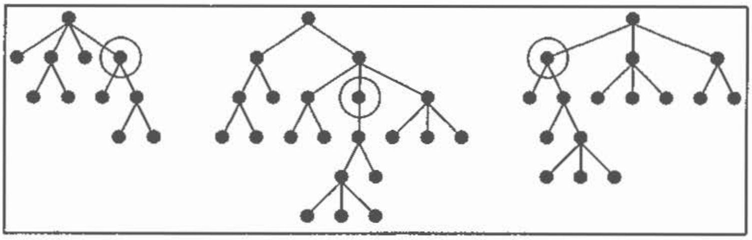
Fig. 5A — Hitlist narrowing modes 1 to 5 (every tree structure represents a document; nodes marked with a circle mean document parts containing primary hits; nodes marked black mean document parts in which retrieval for secondary hits takes place).



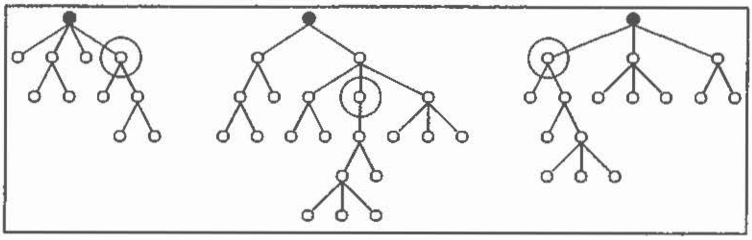
narrowing mode 6
(searching in primary context and supercontexts)



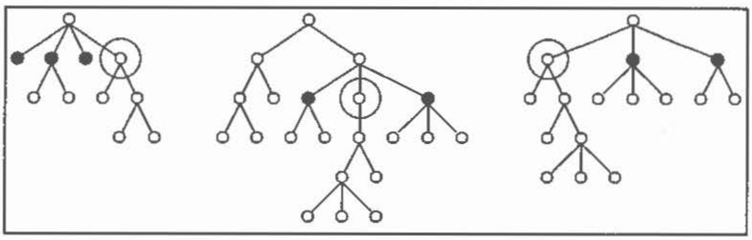
narrowing mode 7
(searching in primary context and subcontexts)



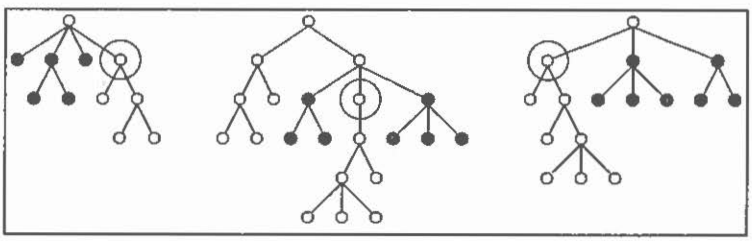
narrowing mode 8
(searching in the same document)



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).

difficult to verbalize, f.i. the typical artistic or workmanlike aspects normally denominated as "style". In these cases a medium allowing the synchronous vizualisation of retrieval results would be very convenient, e.g. a pictures-managing-device producing presentations of digitized object images. The first step in this direction will be to port ARBOR software from the PC-(MS-)DOS-world to a more powerful system environment like f.i. UNIX. Actually this is put into practice.

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

MICHAEL EISNER
Integrated Publication and Information System Institute
Gesellschaft für Mathematik und Datenverarbeitung Darmstadt

BIBLIOGRAPHY

DATE C. J 1977, *An Introduction to Database Systems*, Reading (Massachusetts) (2nd edition), 51-67.

EISNER M. 1984, *Zur Rolle von Datenbanken als Hilfsmittel Instrument kunsthistorischen und archäologischen Fachwissens*, in L. CORTI (ed.), *Automatic Processing of Art History Data and Documents. Pisa, Scuola Normale Superiore, September 24-27, 1984*, Firenze, 325-329.

EISNER M. 1984/85, *Zur Bedeutung der Datenverabeitung als der Archäologie unter besonderer Berücksichtigung von Datenbanken*, « Acta Praehistorica et Archaeologica », 16/17, 278-285.

EISNER M. 1988, *Zukunftsperspektiven archäologischer Wissensdarstellung auf dem Computer*, in *Bathron. Beiträge zur Architektur und verwandten Künsten. Für Heinrich Drerup zu seinem 80. Geburtstag*, Saarbrücken, 115-120.

EISNER M. 1989, *ARBOR. Eine Sprache zur Beschreibung und ein Programmpaket zur Verarbeitung hierarchischer Datenobjekte der klassischen Archäologie und Kunstgeschichte*, Gesellschaft für Mathematik und Datenverarbeitung. GMD-Studien, n. 159, Sankt Augustin.

KAMPFFMEYER U., RUPPRECHT G., WITTEYER M. 1986, *ARCOS: Ein Computer zeichnet römische Keramik. Die Ergebnisse der Testuntersuchungen mit dem ARCOS 1 im Landesamt für Denkmalspflege, Abtg. Archäologische Denkmalspflege, in Mainz*, « Mainzer Zeitschrift », 81, 191-200.

KAMPFFMEYER U. *et al.* 1987, *Untersuchungen zur rechnergestützten Klassifikation der Form von Keramik*, Frankfurt (Main), 79-104.

MAZZOLA G., KRÖMKER D., HOFFMANN R. 1986, *Rasterbild — Bildraster. Anwendungen der graphischen Datenverarbeitung zur geometrischen Analyse eines Meisterwerkes der Renaissance: Raffaels 'Schule von Athen'*, Berlin (West).

ABSTRACT

Both domains divide scientific knowledge into object and method knowledge. The first one means the knowledge of the organization of the single real and normally complex research objects and the latter the 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 the retrieval on ARBOR-coded objects descriptions in different tree-structure-specific query-modes.