

## **D'ALEMBERT**

# **DEVELOPMENT OF A DIGITAL LIBRARY FOR HYPERMEDIA ENCYCLOPAEDIA**

### **A Project Outline**

**Peter Stockinger**  
**(MSH – ESCOM 1995)**

Maison des Sciences de l'Homme (MSH)  
Equipe Sémiotique Cognitive et Nouveaux Médias (ESCoM)  
54, Bd. Raspail – 75006 Paris - France

## Contents

1) INTRODUCTION.....	3
2) THE GENERAL CONTEXT.....	4
3) FROM THE EDITORIAL PRODUCT TO THE NOTION "ENCYCLOPAEDIA".....	6
4) THE IMPACT OF THE ENCYCLOPAEDIC COMPONENT FOR THE NEW IT.....	8
5) THE ENCYCLOPAEDIC COMPONENT AND THE CONCEPTUAL LEVEL OF IT-SERVICES.....	9
6) THEORETICAL AND METHODOLOGICAL ISSUES.....	13
6.1) SEMIOTIC AND COGNITIVE DESCRIPTION OF (TRADITIONAL) ENCYCLOPAEDIC DOCUMENTS.....	13
6.2) MODELIZATION, IMPLEMENTATION, AND PROTOTYPING.....	20

## **1) INTRODUCTION**

The D'Alembert project pursues the following principal objectives:

- the systematic semiotic and cognitive analyses of paper-printed general and specialized encyclopaedia for the elaboration of rich *conceptual models* that are used for the building, maintenance, and reuse of electronic hypermedia encyclopaedia;
- the development of a *library of generic structures* for the specification of encyclopaedic modules (that present across different medias a particular knowledge objet), the specification of narratively and/or argumentatively motivated links between encyclopaedic modules (or features of encyclopaedic modules), the specification of verbal and/or visual representation modalities of a (piece of) knowledge objet in a encyclopaedic module, the specification of a multimedia and multilingual interface design as well as of the H/S-interaction modalities;

- the development of *prototypes of electronic encyclopaedia* that are based on the quoted library;
- specification of a *streamlined methodology* for the conception and authoring of electronic encyclopaedia.

## **2) The General Context**

The conception and production of an impressive quantity of more or less important, more or less generally known and acknowledged encyclopaedia, is a specific feature of the Western Culture which may be explained by the exponential growth of general as well as especially specialized knowledge in the last three centuries.

The production of all sort of encyclopaedia has become an overall important activity in this century and, more particularly, in the last three decades because of the growing general necessity to organize a huge amount of knowledge and to conceive targetted accesses to relevant knowledge items within a structured whole of knowledge.

One of the principal purposes of encyclopaedia in general is the *organisation, viz. the classification* of knowledge in the various fields of human activities encompassing general cultural and historical as well as more specific, technical or again scientific fields.

Another principal purpose is, classically speaking and following the spirit of Diderot and D'Alembert, *the education and the enlightenment* of man. This humanistic purpose survives more or less in modern general encyclopaedia, at least in these ones that aim at a global classification of historical, cultural, and scientific knowledge.

But it is evident, also, that this classical educative objective cedes another, highly pragmatically oriented one which is the *communication and dissemination of knowledge relevant for a user* with respect to his/her activities, epistemic ("intellectual") state, and professional purposes. This second objective dominates clearly in encyclopedic works that aim at the organization and classification of knowledge domains peculiar to technical or technological, economic, and commercial constraints.

### **3) From the editorial product to the notion "encyclopaedia"**

Encyclopaedia are not at all limited to traditional cultural and educative domains and puposes. They are produced, too, for all those activities that have intrinsically to deal not only with more or less important amounts of knowledge but also, rather regularly, with knowledge belonging not necessarily to the delimited "competence" of one specific type of activity as well as with emerging, new forms of knowledge that may have an impact on existing activity routines.

With respect to different studies demonstrating that in a wide range of professional activities between 40 and 60% of the working time are consacrated to information research, documentation but also to the further education and formation related to a professional activity, it is obvious that the appropriate conception and production of encyclopaedias constitue *one* of the milestones in the redefinition of work and of its economic value.

Given the extreme importance of knowledge acquisition, knowledge control, and knowledge dissemination, it is obvious, too, that professionals interested in encyclopaedia are not only publishers, as it has traditionally been the case.

New professional groups who are, since more recently, interested in encyclopaedia are, too, societies and compagnies,

national and international administrations, consumer organizations or syndicates.

The important fact to retain is the following one: even if the denomination "encyclopaedia" points, typically, to opera such as the "Encyclopaedia Britannica", the "Encyclopaedia Einaudi", the "Encyclopaedia Universalis" or again the "Grosze Brockhaus" (for only to speak of some prominent modern general encyclopaedias), the *concept* itself of encyclopaedia, viz. the classification, organization, and the targetted access to a knowledge item, is an highly general one and common to every field or domain that produces a certain amount of knowledge that has to be organized, classified, memorized, and disseminated.

Besides, therefore, the known typical formes of general culture-oriented encyclopaedia that propose the sum of human knowledge, there are innumerable more specific encyclopaedia in science and technology, biology, medicine, ecology, the humanities, art, history, and philosophy created for educative, professional or again merely distractive ones.

The concept of encyclopaedia is, moreover, not restricted to the sole editorial product "encyclopaedia". Contrarily, at least as far as its classificational principle is concerned, it is embedded, too, in each (cultural, historical, geographical, ... ) guide, handbook or again in

each one of the innumerable catalogues which constitute one of the most important information sources in the actual economic and ecommercial activities.

#### **4) The impact of the encyclopaedic component for the new IT**

Taken in this sense, the encyclopaedia constitutes, too, a central and in general even essential component of almost all information and communication services developed with the help of the new IT: electronic documents and information kiosks, information research and exploration services, computer assisted learning systems, electronic publishing of (multimedia) documents, as well as, finally and more particularly, knowledge based systems.

The central place that an encyclopaedia occupies in the electronic information and communication services is simply due to the fact that these services have, in one way or another, to integrate or at least to refer to *structured knowledge or information modules* that constitute, together, more or less *expanded and specialized knowledge libraries* in order to be able to promote a certain type of requested services.

Finally, in pushing this conception of what an encyclopaedia is to its logical end, there arise a rather paradoxal conclusion. The

encyclopaedia stresses rather clearly the *conceptual problem* of the so called information highways: given the fact that the new communication and information technologies allow the storage, communication, and retrieval of tremendous quantities of information objects, the problem, nevertheless, remains by the means of which *conceptual or classificational grids* these quantities of information objects should be stored, accessed, and disseminated in such a way that an interested user could be satisfied.

### **5) The encyclopedic component and the conceptual level of IT-services**

Therefore, the conceptual level has clearly to be distinguished from the technological (i.e. physical and functional) level of information storage, communication, and retrieval .

At the conceptual level, following more particular questions arise:

- what are the generic structures, the generic regularities by the means of which *classificational grids* may be elaborated that underly a class of knowledge objects ?;

- what are the characteristic regularities by the means of which *knowledge modules* of an encyclopaedia may be *networked* ?;
- what are the appropriate interaction and communication modalities between an encyclopaedia and its potential and real users by the means of which the *encyclopaedic knowledge "universe"* may be *purposefully exploited* ?

The first question sticks to the problem of the elaboration of *highly modularized electronic librairies* where each module comprises a (piece of a) relevant knowledge object. With respect, for instance, to an object oriented approach, a module is constituted typically by a class of structured or configured features that refers to a particular objet. Understood in this sense, a class by itself possess a generic structure which may be used for the specification of whole set of particular objects. Classes of configured features defining, for instance, an object such as "political personnality", "economic crisis" or again "compagny" may be (at least partially) re-used for the specification of real political personnalities, economic crises and compagnies for the creation of electronic librairies concerning political and economic issues.

The second question sticks to the problem of how a specific module (or a specified feature in a given module) is related to,

interacts with, another module (another specified feature). A prominent type of links constituted by *causal links* that contextualize a given module M with respect to other modules that constitute either consequences or causes of M. A module specifying a political personality may contain a feature specifying the political vision of this personality which, itself, interacts with these modules where the consequences or the causes of such a vision (in general or more particularly to the historical context of the given political personality) are thematized.

The third question sticks to several, strongly related problems. Given the rather high probability that an electronic library is constituted by a high number of interrelated modules, the user must have access to one or a subset of these modules not only by the traditional means of index but also by the means of content-driven, bottom-up technics. Such a technic, for instance, is constituted by a generic query frame (or class) which will be successively specified by the requests of the user and which has to match with the corresponding classes that constitute the classificational grid of a set of existing modules.

Another problem concerns the fact that even a rather fine-grained set of types of links or not necessarily sufficiently appropriate for the peculiar structure of a given domain of knowledge. So, in order to give the user the possibility to explore and to take into account

highly domain-constrained knowledge specificities, a particular effort consists in the specification and definition of so called heuristic rules concerning the particular structural and dynamic "behaviour" of a knowledge domain. For instance, it is not always enough to give a user the possibility only to compare two or more companies; there may exist information or knowledge needs from the side of the user to compare and evaluate companies with respect to their relative position concerning "investment efforts", "market expansion" or again "geographical presence". In order to cope with such information requirements, an electronic library has to be provided with a sort of, at least simple knowledge base in which such heuristic rules are deposited.

A third, more commonly recognized problem is constituted by the fact that the "navigational liberty" of a user in a modularized and highly interlinked electronic library must be controlled and monitored by the system in order to support the user in his/her information or knowledge acquisition tasks.

It is obvious that these three fundamental questions concerning the conceptual level of encyclopaedia cannot be solved in a purely apodictic way; it cannot be solved, neither, by purely formal means.

They stresses, first of all, *empirical issues*, viz. the necessity of the identification and description of the characteristic features and regularities that underly and orient the classificational activities employed in already existing (traditional) encyclopaedia or again in (traditional) documents such as the already quoted catalogues, handbooks, and guides which incorporate a more or less extensive encyclopaedic component.

## **6) Theoretical and methodological issues**

### **6.1) Semiotic and cognitive description of (traditional) encyclopaedic documents**

An explicit purpose of the D'Alembert project is not only to contribute to the electronic publishing of encyclopaedia but also, and even more fundamentally, to the cognitive or conceptual aspects of building, maintaining, and reusing electronic encyclopaedia (viz. electronic knowledge libraries).

Given this purpose, the D'Alembert attaches an high importance of the systematic description of the particularities of encyclopaedia or again encyclopedic documents.

One of the fundamental purposes of an encyclopaedic article is, as it has been already developed above, to provide a user not only

with sufficiently characteristic information but also with information that the user could understand in order to grasp the particularity of an knowledge object.

In order to guarantee this quality of which an encyclopedic article has to witness, there are traditionally specification technics by the means of which the generic structures (or again the types of structure) of encyclopedic articles are created and organized.

Among these technics, one of the most prominent is constituted by the targetted interviewing of experts who decide which types of features must be taken into consideration for a qualitatively good presentation and description of a knowledge object such as a technical product, a natural object or again an abstract one.

In the area of knowledge elicitation for knowledge based systems, different methods and procedures have been developed for a structured way of how to elicit knowledge and how to organize it in object classes. These methods, basically based on interviewing, testing, probing, and so on, can be fruitfully transposed and implemented in the context of this project.

In technical domains, there exist already different conceptual standards (standards of quality such as those proposed by the French organism AFNOR).

Nevertheless, even if these specification technics are indispensable for the creation of encyclopedic articles (or modules), they are not always sufficient for this task. They should to be completed by another type of technics based on *comparative - cognitive or semiotic - studies* showing how traditional encyclopaedia, handbooks, guides or again catalogues are conceived and structured.

The relevant point concerning these studies is the following one: encyclopaedia in the narrow or in the broad sense not only specify the characteristic features of a knowledge object, but they employ, too, strategies of how to describe and to explain them, of how to introduce them following a given plan structure that underly an encyclopaedic article, or again of how to distribute, between several expression (i.e. verbal and/or visual) codes, the charge of the representation of the several problematized aspects of a knowledge objet.

These - and a variety of other - strategies are not at all purely superficial rhetoric devices but contribute fundamentally to the quality of knowledge dissemination in encyclopaedias.

Comparative semiotic and cognitive analysis should propose concret and explicit results concerning:

- the structuring of encyclopedic articles,
- the linking of a set of encyclopedic articles,
- the distribution between verbal and visual expression codes by the means of which a particular object of knowledge is presented
- the acquisition strategies of relevant knowledge objects that are developed in encyclopaedia.

As far as the *structuring of encyclopedic articles* is concerned, following two parameters have to be elucidated:

- what are the characteristic features that intervene in the specification of an encyclopedic article developing a particular object of knowledge ?
- what are the typical variations of these features if a particular object of knowledge is described from different point of views ?

Concerning the problem of *linking of different encyclopaedic articles*, the cognitive or semiotic studies should unveil a set of recurrently used link types by the means of which, in traditional documents, the author guides the reader through a "book".

Even a rather loose analysis and comparison of different paths between encyclopaedic articles show that there exist a variety of highly motivated links which haven't yet been really well recognized. The real problem is not so much to create as many links as possible between different modules of an electronic library; it is constituted, contrarily, by the choice of a set of motivated links that facilitate not the navigational liberty of the user but his/her search and exploration of relevant knowledge objects.

Indeed, the problem of the physical support of a document should not be confused with the problem of its content. There exists, actually, a kind of "electronic wishful thinking" believing that an electronic document such as a hypertext constitutes already by itself a decisive amelioration with respect to "traditional" documents in the sense that the first one doesn't possess the limitation that are intrinsic to the second one. From a physical point of view, this affirmation is obvious, but from a conceptual point of view it is really not necessarily true: an electronic hypertext and, more generally, an electronic (digital) library that should constitute a substantial amelioration with respect to information, knowledge acquisition, and education purposes must, at least, possess the same qualitative level that possess the already existing encyclopaedia printed on a traditional physical support.

Given this context, semiotic and cognitive analyses of recurrently used link types possess, therefore, a high relevance.

Among these recurrently used types of links there are, for instance specialisation links, causal links, chronological links, classificational and taxinomic links as well as functional ones.

But there exist, too, more complexe ones which possess a *narrative and argumentative (evaluative, explanative and persuasive) character* and which have to do with the controversial nature that may exist between two (or more) encyclopedic articles (such as, for instance, between two articles that develop a particular knowledge object from two different point of views).

Concerning the *distribution between verbal and visual expression codes*, the cognitive and semiotic studies should explicit not only different types (or "styles") of verbal and visual language employed in encyclopedic opera but also more precisely the different functions that they fulfill in such works.

It is, for instance, recognized that visual codes such as pictures, schemas, diagrammes, flow chartes, or again figurative graphics can be systematically employed in order to show, from different points of view, specific properties of a particular knowledge object.

It is generally recognized, too, that verbal and visual means, together - or better: in cooperation - fulfill particular information functions that optimize the knowledge acquisition process.

Concerning, finally, the *acquisition strategies* of how a user exploits for his/her purposes the classified knowledge contained in an encyclopaedia or an encyclopaedic document, it deals more particularly with the interactive aspects between, on the one hand, the user and the document, and, on the other hand, with the author (the "system") and the user.

Semiotic and cognitive analyses of (traditional) encyclopedic documents should unveil particular "scenarios" by the means of which the author, user, and document interacts in order to fulfill a communication purpose.

Such scenarios are not only provided by particular (author-centered) help and (user-centered) help request strategies but also by intertextual references identifying external knowledge resources that may complete a given knowledge or information need.

## 6.2) Modelization, Implementation, and Prototyping

The basic functional specification of an electronic encyclopedic library dissociates:

- a *set of modules* that contain (pieces of) a particular knowledge object presented following different expression (verbal or visual) expression modalities;
- a set of (*hypermedia*) *links* by the means of which modules are connected;
- a (*multimedia*) *H/S-interface* defining the H/S-interaction modalities and, more particularly, the modalities of the knowledge or information access (i.e. query and navigation).

This basic functional specification constitutes, too, the general reference framework in the D'Alembert project.

On the basis of the results of the descriptive analysis of encyclopaedia and encyclopedic documents of basic generic structures will be elaborated by the means of which encyclopedic articles as well as libraries of encyclopedic articles can be specified and, further, developed.

Given the above quoted basic functional architecture, there are four basic types of such generic structures that have to be distinguished:

- types of generic structures that define canonically the the *characteristic features of a domain of knowledge*;
- types of generic structures that define canonically the *plans (rhetoric acts such as narratives)* following which the characteristic features of a particular knowledge object will be developed;
- types of generic structures that define the *(verbal or visual) representation* (viz. the so called "view" component in an object oriented object) of a particular knowledge object ;
- types of generic structures that define the *interaction modalities* between the user and the system as well as their (visual and/or verbal) representation.

Together, the generic structures constitute a kind of "*library*" that guides the processes of conception, specification, and development of a particular electronic library or encyclopaedia.

Generic structures, in general, are always domain-dependant, viz. they define particular configurations that are considered to be characteristic of a domain or field of knowledge.

In this sense, generic structures - especially these belonging to first of the three above quoted types - are not forcibly restricted to an encyclopedical use: they constitute, too, knowledge modules (i.e.

classes, in an object oriented approach) that specify, in the context of a knowledge based system, the structural dimension of a domain.

The cognitive or semiotic description of different types of relations that an author employs in order to guide his reader through a "book" (i.e. a book with at least a strong encyclopedic component) constitutes the input for the more formally oriented definition and modelization of types of links by the means of which encyclopedic articles, viz. information developed in different articles, will be linked.

The objective is to define conceptually richer link types than those that exist actually for current hypertext or hypermedia applications. The reason for this objective is that an electronic encyclopaedia (in a general - cultural - sense but also in a more specialized - technical - sense) should not contradict cognitively highly motivated information or knowledge acquisition procedures that are traditionally employed in the thematic structuring of a text or a discourse.

Due, especially, to the long tradition of rhetorics, modern approaches in text- and document analysis such as, for instance, formalism and structuralism, semiotics and cognitive studies, have developed more specific conceptual "grammars" showing how particular rhetoric acts or discourse-plans structure and orient not only

the communication of pieces of knowledge but also their reception and interpretation.

Among these acts, the most prominent ones are the *description*, the *narration*, the *explanation*, the *evaluation* as well as the *discussion* or again the (controversal) *debate*.

In elaborating more *simplified models* of such acts or plan structures, the D'Alembert project will contribute to the establishment of a library of conceptually rich and cognitively motivated links that can be usefully applied to the differentiated linking of encyclopedic articles and to the constitution of purpose oriented electronic libraries.

Furthermore, the linking of knowledge containing modules will be completed by highly domain-constrained heuristic rules involving formal technics pertaining to the field of knowledge based systems. The introduction of such heuristic rules makes it possible, as already mentioned, to propose a user a highly targetted contextual exploration of knowledge modules - exploration procedure which is impossible to propose otherwise.

The definition of generic structures concerning the presentation (i.e. the verbal and especially the visual) modalities of knowledge objects constitute another component of the library by the means of

which knowledge objects expressed by diagrams, images, icons, figurative graphics, video clips, ... can be more precisely and contextually more appropriatedly indexed and accessed.

They constitute, too, an authoring tool for the specification of the (verbal and visual) presentation of knowledge objects in an electronic encyclopaedia.

With respect to the elaboration of generic H/S-interaction modalities, it must be noted that a particular effort has to be accomplished:

- concerning the user-bounded actions, besides the already well known technics of (contextual) browsing, in the specification of more generic *bottom-up* or *content driven query* procedures;

- concerning the system-bounded actions, in the specification of *contextually appropriated help functions* by the means of which a user may be conducted purposefully through an electronic encyclopaedia.