

CONCEPTUAL MODELLING OF BUILDINGS
CIB W74 + W78 Seminar. October 1988. Lund, Sweden

**F A R T E C : A PROJECT TO SIMPLIFY THE ACCESS TO THE
TECHNICAL DOCUMENTS BY USING MODERN
COMPUTER TECHNOLOGIES**

Marc BOURDEAU

CENTRE SCIENTIFIQUE et TECHNIQUE du BATIMENT (CSTB)
Etablissement de SOPHIA ANTIPOLIS
B.P. 21
06561 VALBONNE CEDEX, FRANCE

October 10, 1988

KEYWORDS

Technical rules - Information retrieval system - Information packaging -
digitization

ABSTRACT

The CENTRE SCIENTIFIQUE et TECHNIQUE du BATIMENT has launched a project, named FARTEC, whose objective is to put building technical rules at professionals' disposal by use of modern computerized tools.

This project involves several topics : full digitization of the documents, packaging of the information, elaboration of retrieval systems and dialogue with professionals. Experimentation of advanced techniques has been carried out in 1988.



1. INTRODUCTION

During the last few years, it has become more and more urgent to put technical rules at building professionals' disposal by modern means.

Faced with this demand, the Authorities and the CENTRE SCIENTIFIQUE et TECHNIQUE du BATIMENT have prepared, in 1987, a national project named **F A R T E C**.

The CSTB is the manager of this project that started in 1988.

2. The OBJECTIVE of FARTEC

The main objective is to prepare the short-term achievement of complementary and coherent computer systems, meant to facilitate the access to building technical rules for professionals.

These systems will use modern means of information storage and manipulation, based on the following principles :

- full digitization of the documents : texts, tables, formulas, graphs, schemata, ...
- use compatible with computer and/or telematic devices within professionals' reach.

The organization and structuration of the information will be conceived so as to allow, in particular, its integration to the information services proposed to professionals in data banks. Finally, the **F A R T E C** project aims at allowing generalized and automated accesses to technical rules, from ordinary and user-friendly working stations.

That implies, in particular, a right perception of the professional demand related to the use of these informations.

Besides, considering the near european prospect, this project gives to France the opportunity of defining and settling mechanisms that could be exported or suitable for harmonized data inside EEC.

3. The DOCUMENTS CONCERNED

The technical rules are recorded in various documents, such as in particular :

- the laws, decrees and orders that constitute the main part of the regulations, strictly speaking, and that define the requirements to be respected ;
- the rules of art, concerning the design and the achievement of the works ;
- the standards, certifications or technical assessments, related to the products.

This scattering makes intricate the use of the whole : regarding the set of provisions applying to a given work, it is frequent to be obliged to refer to several documents, each of them taking into account only a particular requirement.

This complexity can get worse, due to the fact that this bulk of information has been progressively elaborated along several decades. With the passing years, the spirit of the texts, and the texts themselves, have evolved slowly, taking into account technological progress but also changes of mind. Successive updates have permitted to modernize some of the oldest documents but not suppressed all the lack of homogeneity.

4. WORK in PROGRESS

The **F A R T E C** project will last 3 years.

We present below the outline of the work begun in 1988, under the general management of the CSTB's Computer Science and Building department.

4.1. DIGITIZATION of the DOCUMENTS

That consists in creating and organizing a base of digital files corresponding to the documents concerned by **F A R T E C**.

An inventory of the documents, that are gathered today in a collection distributed by the CSTB, has led to the following estimations :

• Regulations	1 400 pages
• Unified Codes of Practice.....	3 600 pages
• Calculation rules	1 800 pages
• Examples of solutions	400 pages
• Technical Assessments	1 300 pages
• Standards	<u>5 500 pages</u>
TOTAL	14 000 pages

It is likely that the wider whole of documents to take into account in the framework of **F A R T E C** will reach about 20 000 pages.

In order to constitute the **F A R T E C** files, several approaches are used, according to the various origins of the documents :

- recovery of existing files, inherited from the photocomposition of the documents published by the French gazette (laws, decrees, orders, ...), or produced by AFNOR (standards) ;
- keyboarding of texts and optical scanning of schemata, drawings or graphs ;
- optical scanning of the full documents, and character recognition for the textual parts.

Besides, the various forms of information to take into account (textual, graphical, numerical, ...) must be distinguished. The file format corresponding to each form and the management tools for file mixing, are to be defined.

A short-term objective (end of year 1988) is to elaborate a first version of this file base, as complete as possible, and to develop the tools necessary for the management of the various types of information.

1.2. INFORMATION PACKAGING

The digitization of the documents must be accompanied by the preparation of the automated accesses to the information, whether it is in prospect of simply consulting the texts or, in a more ambitious way, coupling to bureautics, calculation or drawing software.

In the absence of an automatic process for analyzing the meaning of the information, a system for locating and describing the information units (also called provisions), has to be conceived.

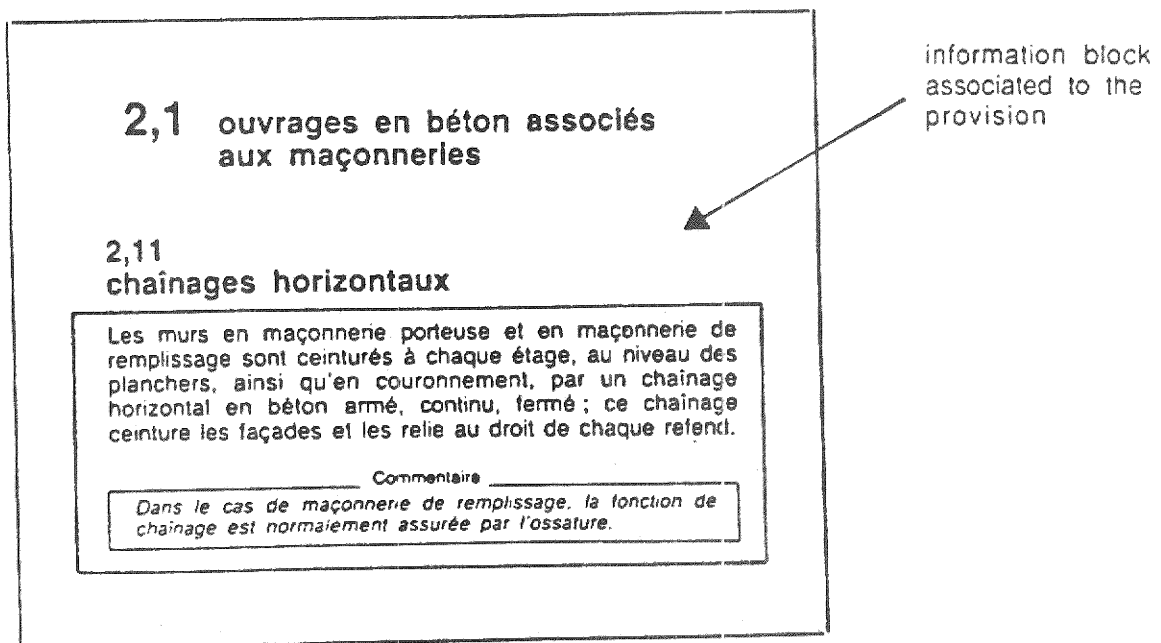
The descriptions, associated to the information units, will constitute the keys enabling access to the technical provisions.

They will use the notion of "context" so as to precise the high-level information that characterizes the document, or the part of document, in which the provision is included : purpose of the building (dwelling building, public-receiving building, ...), technology (concrete, wood, ...), concern (construction rules, calculation rules, ...), etc.

The description sheets will contain all the synthetic information linked to the meaning of the provisions : in particular, the "object" which the provision deals with, and the "wording", i.e. the main idea that can be extracted from it. In some cases, it will be necessary, in addition, to precise the "sub-context" of validity of the provision : type of work, location, ...

This information about the "objects", the "wordings" and the "sub-contexts" of the provisions will be expressed in free language by the experts responsible for the analysis of the documents.

Example of an information unit (extracted from the Unified Code of Practice n°20.1 entitled : "Walls in masonry of small elements")



- { CONTEXT : construction rules concerning masonries of small elements
- { OBJECT : horizontal reinforced concrete chaining
- { WORDING : horizontal reinforced concrete chaining is compulsory

4.3. DIALOGUE WITH PROFESSIONALS

This dialogue has started at two levels :

- An interprofessional Direction Committee, placed under the aegis of the Department of Construction, defines the technical and sectorial priorities, and plans the strategic directions related to the tools and information services aimed at.
- Collaborations between the CSTB and the service producers have been set up, in order to define specifications for an information packaging (type, structure, format, ...) suitable for a future supply of the corresponding files, with a view to integrating them into the services in question : data banks (ARIANE, INFOBAT, ...), software (bureautics, technical calculations, CAD, ...). These collaborations are placed under the patronage of the main professional organizations concerned with : FNB (firms), CAPEB (craftmen), COPREC (inspection offices).

4.4. EXPERIMENTATION of NEW TOOLS

The **F A R T E C** processes imply the use of modern computerized means for storing and retrieving digitized information. In order to prepare the choices suitable for both the types of information and the services considered, the characteristics, the possibilities and the limitations of the various potential techniques should be evaluated.

Besides, for the impact of **F A R T E C**, it is important to be able to periodically present attractive models, illustrating the advantages of advanced technologies, to the professionals, institutions, administrations or information suppliers themselves.

The main experiments carried out on this account in 1988 are the following :

- demonstrative models allowing the access from micro-computers (Macintosh) to the information (texts, drawings, ...) contained in documents such as the Unified Codes of Practice. One of these models implements a retrieval procedure based on a description of building works ; it leads to the display of the selected information units after the user has checked that the wordings associated to these provisions fit in with the subject of the search.

Another model, developed under HYPERCARD (APPLE), implements the consultation, through the contents, of the various chapters of about twenty Unified Codes of Practice. This second model has been the subject of an experimental production of a CD-ROM.

- Two experiments have been carried out, in order to evaluate on-line full-text query systems, using the Minitel network. The models have been built by companies that have developed specific software. In both cases, the user expresses his information search by associating "freely" chosen words or expressions. In reply, the systems indicate the passages of the documents that contain the elements of the query. The advantage of these software is double : they save from indexing the documents and give a very little constraining use. Therefore, we must check, as exactly as possible, whether they fit the professional needs of text querying : that forms the subject of the evaluations under way.

CONCLUSION

The **F A R T E C** project is very ambitious. The complete mastery of a corpus constituted by 20 000 pages of technical information, amassed in 40 years, implies important work :

- construction of the file base ;
- packaging of the information ;
- elaboration of retrieval systems.

Probably, a long time will be necessary to reach the stage of a satisfactory semantic mastery ; but, from now on, collaborations have been set up with specialized teams in order to approach this ambition by promising ways : expert interpretations based on Artificial Intelligence mechanisms, logical representation of the content of the documents, statistical and neuronal analysis, ...