

## INFORMATION TECHNOLOGY IN RESEARCH

# CHAPTER 3 : Working Tools for Retrieving Information

### CONTENTS

1. Information Science and the Library
2. Information Processing Techniques and their Role in Retrieving Information

Librarian Gabriela COSTEA  
Prof.Ph.D.Eng. Elena HELEREA

# 1. Information Science and the Library

***Information Science*** is the „information provision science and practice, whatsoever the organization or the individual providing such services : *documentation centers, libraries, information analysis centers, information brokers*”.

The Institute of Information Scientists (England)

*Library* not only has been constituted as data storing institution, its purpose being of our days, more than anytime, to ***enhance and transmit information***;

- in other words to provide the solicitors with precise, pertinent data, in the shortest while.

This has been facilitated by the library computation. Due to the *huge storing capacity* and to the *higher speed* in retrieving information, the use of the computer has resulted in *sparing the time destined for investigation and the space for reading* – information may be placed anywhere and anytime, simultaneously, at the users' disposal.

Moreover, using ***integrated library software*** offers the *possibility for searching information concomitantly*, through *several access points* – combined search (ex: according to the title and to the subject, according to the author, to a word within title and edition etc.)

## **2. Information Processing Techniques and their Role in Retrieving Information**

This modulus sets out to describe the *methods of documentary analysis used within libraries* to the purpose of elaborating the catalogue – essential tool for the ultimate achievement of any library process : the users' access to information or, in other words, the fulfillment of the information need.

**Documentary analysis** – *basic common method in processing information as object of documentary research* – it stands for the node wherein there intersect not only the **(inter)mediating disciplines (information and documentation sciences : library science, bibliography, archive science)** and these ones in their turn with **communication sciences** (according to several expressions – *structural linguistics, semiotics and language philosophy*) – reunited by the preoccupation for language as the modality for *rendering information*.

According to the *nature* of the information and to the *name* of the specific information, there may be distinguished two levels of data : ***formal and of content***, which determine the type of analysis :

- a) ***description of bibliographical data and***
- b) ***classification/indexing.***

a) **Bibliographical description** beside *catalogue organizing*, are gathered within theoretical literature by the name **catalogation**, vaguely defined within the specialized dictionary („process of elaborating catalographic descriptions of the documents and their arrangement [...]”);

*indexing* may have a broader sense, including *classification*, and a more restrained sense, only referring to the indexing with terms of the natural language.

**CATALOGUING** supposes describing all **bibliographical data** of the document (*author, title, edition, publishing house, series, ISBN* etc.) to the purpose of its **registering and retrieving**.

**Cataloguing** is achieved on the basis of the international standards **ISBD (International Standard Bibliographic Description)** adequate to the type of document analyzed (current monograph, old monograph, printed musical publications, audio-visual material, serial publications, standards, patent specifications, electronic resources etc.).

The catalogue allows the user's *access* to the document on the basis of the bibliographical data (**alphabetical traditional catalogue** according to the *author/title*; **online catalogue** according to any of the bibliographical data).

**b) Content Analysis** is designated through the generic term *indexing* and is achieved through *documentary languages*

- with *hierarchical structure* (codified classification schemes – used in systematic indexing) or
- with *combinatory schemes* (terms selected from the natural language, authority lists or subject headings and thesauri – used within analytical or alphabetical indexing).

**INDEXING** is complementary to cataloguing; it is the operation describing the **content of a document**. Through indexing, the user is granted the access to the document on the basis of the subject dealt with.

**Documentary languages** have been rather recently dated: the 19<sup>th</sup> century. They are ***artificial languages***, eliminating the ambiguities of the natural language, due to *omonimiy and polysemy*. Documentary languages easily allow searching and retrieving information.

Documentary languages hold the special analysis tools for treating information.

**A. Documentary languages with hierarchical structure** (*systematic, classifying languages*) : used in order to represent the content of the document in a synthetic manner

Within this type of classification there belong : Dewey Decimal Classification, Universal Decimal Classification, Library of Congress Classification, Colon Classification.

**Universal Decimal Classification (UDC)** is a general scheme for arranging all fields of human knowledge. This scheme was created by two Belgian lawyers – Paul Otlet and Henry LaFontaine in **1895**

– to the purpose of systematically arranging the bibliography they had created „Repertoire bibliographique universel („universal bibliographical repertory”)

– which was meant to be a bibliography of all publications existing at that moment on the market.

In **1908**, UDC was likewise applied in **Romania**, for the periodicals from the Library of Academy and, since 1915, it extended upon the books.

In 1992, UDC turns into UDCC (Universal Decimal Classification Consortium) – consortium of editors headquartered at the Royal Library from Prague.

***Conventional Language UDC*** is based on *Arabic figures* from 0 to 9, dividing this way the entire body of knowledge in 10 general classes (class 4 was cast out in 1961) comprehended in the so-called principal Tables. These classes are :

- 0 Generalities
- 1 Philosophy. Psychology
- 2 Religion. Theology
- 3 Social Sciences
- 4 (Vacant at present)
- 5 Mathematics and Natural Sciences
- 6 Applied Sciences
- 7 Fine Arts. Applied Arts. Entertainment. Games. Sport
- 8 Language. Linguistics. Literature
- 9 Archaeology. Geography. Biography. History

Beside *generality* (includes all fields of knowledge) and *universality* (independence from the natural language), **UDC** has the property of being a ***decimal classification*** as it is based on the conventional distribution of human knowledge into ten great classes (0, 1, ..., 9) which, in their turn, are divided up to the desired level, creating thereby the possibility of extracting the particular from the general.

UDC is a ***systematic classification***, in which all human knowledge are grouped within a hierarchy, considering the relations among themselves. To put it otherwise, UDC analyzes the content of the subjects and of the notions, drawing together the similar ones and drawing apart the dissimilar ones.

Example (sub-classes of the class 6 – Applied Sciences) :

- 60 Generalities of Applied Sciences
- 61 Medical Sciences. Medicine
- 62 Engineering. Technique in general
- 63 Agriculture
- 64 Household Economy
- 65 Management and Industry Organization
- 66 Chemical Technology
- 67/68 Industries and Trades
- 69 Civil Engineering Materials

**UDC** is an ***aspectual classification***, in which a phenomenon may find its place in several classes.

For instance, for the concept “coal” there are associated several *UDC indices*, according to the aspect wherein it may be dealt with :

552.574	petrography aspect
553.94	aspect of economic geology
622.23	mining aspect
622.411.52	aspect of mine dust
624.131.27	aspect of soil mechanics
631.878	aspect of fertilizer

**UDC** is a *hierarchical classification*, which means that every sub-division may be further divided into its logical compounding elements. This conveys *complexity and flexibility*, with high absorption capacity of new terms and notions.

Example 1

373 Education of general culture

373.3 Primary (elementary) education

373.5 Secondary (intermediate) education

373.51 Organization of the secondary education

## Example 2

821.111-1 English poetry. Poems. Lyrics

821.111-2 English dramaturgy. Theatrical plays

821.111-3 English prose. Narrative prose

821.111-4 English literature. Essays

821.111-5 English literature. Discourses

821.111-6 English literature. Correspondence

821.111-7 Satire in English prose. Humor, epigrams, parodies

The existence of hierarchy allows easily *retrieving* the more general or more restricted concepts, *in the automated environment*, through simply deleting or adding a figure.

Here is as follows an example for such a search (the truncated use of the index UDC 811.111'36 according to the automated search *search*):

811.111'36

- 1 The Cambridge grammar of the English language / Rodney Huddleston and
- 2 Gramatica limbii engleze / Leon Levitchi. – București : Editura Teora,
- 3 Mastering English : an advanced grammar for non-native and native speak
- 4 Themes anglais pour toutes les grammaires : les milles problemes, dictionnaire

811.111'36(035)

- 6 The good grammar guide / Richard Palmer. – London ; New York : Routledge

811.111'36(075.8)

- 7 Gramatica limbii engleze : [pentru uzul studenților] / sub îndrumarea
- 8 The Random House guide to writing : [for students] / Sandra Schor and

811.111'366.5

- 9 Parsing and easy analysis. – 1936. – 56 p. : tab.

811.111'367.3(075.8)

- 10 The historical evolution of the impersonal sentence structure in English

811.111'367.33(075.8)

- 11 Complementation in English : a minimalist approach / Alexandra Cornilescu

811.111'367.333(075.8)

- 12 Verb complementation in English : [for students] / Gina Măciucă. – Suceava

811.111'367.4

- 13 Functional categories in English : Vol.1/ Ileana Baciucă. – Bucuresti

811.111'367.4(075.8)

- 14 Verb complementation in English : [for students] / Gina Măciucă. – Suceava

811.111'367.6

- 15 Functional categories in English / Ileana Baciucă . – București : Editura

## UDC

- is an **artificial language of classification**, with whose help there are *eliminated* many of the *ambiguities* of the natural language.
- Every **index** represents a **clearly defined concept** and not a word or an expression whose meaning(s) many vary according to the context.

Example

Concept	Natural language (semantic meaning)	UDC language
Magazine	Periodical publication	050
	Show	792.7
Power	Mathematics	511.132
	Car building	621.1.018.7
	Electrotechnique	621.3.016.2
	Law (State power)	342.5

**UDC** is a ***synthetical classification***, in which *compound notations* are built from simple, enumerative classes.

***Classification indices*** are systematically arranged. A ***UDC index*** has two components, which are :

***class description*** – a term which defines the concept in the framework of its hierarchical context, through its precise description in natural language;

***class index***– notation which stands for the code used in representing the class and which determines the place of that class within the classification scheme. Adding another figure implies a sub-division of the previous file. Often, further adding a figure indicates a sub-ordinated concept.

The figures are grouped in clusters of three, from left to right, separated by a point.

Example:

6	Applied Sciences
62	Engineering
622	Mining Industry
622.2	Mining Works
622.23	Excavations
622.233	Holing for explosions
622.233.4	Tools and hammer holing

**UDC notation** is based on *Arabic figures*, on the letters of the Latin alphabet and on a series of authorized UDC symbols.

Example :

025.45 CZU

621.313.33.073.1

7.036.5(520)

908(498-21Brasov)"1788/1850"

**UDC indices** are : I. ***principal*** (from 0 to 9 with their sub-divisions)  
***auxiliary*** (they are added to the main indices to express other facets of the subject, as physical form, as geographical zone, as chronological sub-divisions etc.)

II. ***simple*** (taken from a single place on the table)  
***compound*** ( created through synthesis, using several indices)

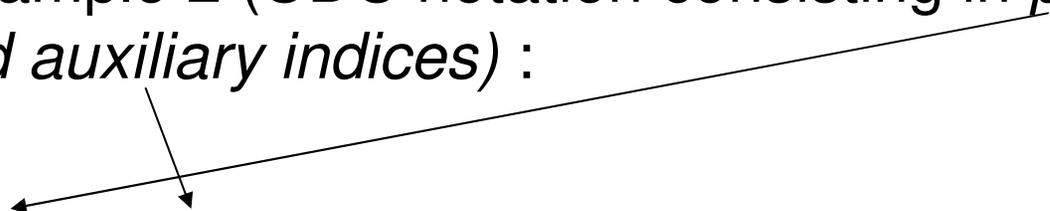
Example 1:

622 Mining Industry - simple index

(410) Great Britain - simple index

622(410) Mining Industry in Great Britain  
- compound index

Example 2 (UDC notation consisting in *principal indices*  
*and auxiliary indices*) :



630.8(075.8) Wood Industry – Handbook

**Main Index :**

- It is found in the main Table, according to the principle ruling the arrangement into hierarchy, from the general to the particular
- it results from the decimal division of the main classes
- it is attributed to a determined notion (concept)

Example :

6	Applied sciences. Medicine. Technique
62	Engineering
621	Car building
621.3	Electrotechnique
621.31	Electrical machines and apparatuses
621.313	Electrical machines
621.313.3	Machines with alternative current

**Auxiliary indices** are of two types :

➤ ***Common (general)***

➤ ***Special (analytical)***

## Common Auxiliary Indices:

- they are ***numerical indices***, hierarchically enumerated and arranged
- they may be applied to **all indices** within the ***main Table***
- they denote **general characteristics** (form, place, time, etc.)
- they may be found in the **Auxiliary Tables of UDC** (Table 1a, Table 1b, ... Table 1k)

## Special Auxiliary Indices :

- they are **locally** used for certain indices
- They use the following representations :

a) **-1/-9**

Example: 821.135.1-2 Romanian prose

b) **.01/.09**

Example: 334.02 Politics. Organization

7.02 Art. Technique of Art

c) ‘

Example: 329.17’23’12 Nationalist-Republican Parties

## Examples of UDC Classification Indices

1. **symbol “+”** (addition-plus)  
622 + 669 Mining and metallurgical industry  
(44 + 480) France and Spain
2. **symbol “/”** (consecutive extension-up to)  
611/612 Human anatomy and physiology  
(it includes all subdivisions of 612  
up to the last)
3. **symbol “:”** (simple relation - it does not specify the  
nature and the sense of the relation)  
17 : 7 may be 7 : 17 Ethics in relation to Art or the  
reverse  
or the reverse)

4. **symbol “::”** (double relation - the order is irreversible, which means the second concept is subordinated to the first)

77.044::355.4

War photographs

5. **symbol “[ ]”** (grouping)

[622 + 669](485) Mining and metallurgic industry in Sweden

004 : [371.3 : 811.111] Computer in teaching English

6. **symbol “=”** (language of the document whose subject is described by UDC)  
 663.4(493)(075)=112.5 Beer industry in Belgium – manual – in German
7. (0/09) (documentary form of the document)  
 54(038) Chemistry dictionary grouped at chemistry  
 (038)54 Chemistry dictionary grouped at dictionaries
8. (1/9) (place-geographical positioning)  
 (498) Romania  
 (493-11) Eastern Belgium

9. **sign (=...)** : (ethnic grouping and nationality)  
78(=411.16) Jewish music
10. **sign “”** (time, period)  
821.111-1(73)”19” American poetry of the 20<sup>th</sup> century
11. -02 (general characteristics of properties)  
-03 (general characteristics of materials)  
-05 (general characteristics of persons and personal characteristics)

**B. Documentary languages with combinatory structure (*analytical languages or indexing languages*)** : used to the purpose of representing the content of the document and of the search requests in an analytical manner.

**The indexing languages** used so as to describe the content of the documents may be divided in two categories :

***free languages*** and ***controlled languages***.

**Free language** has been constituted following the indexing in natural language. It is the type of language used in computer-aided research, ensuring a better exhaustiveness to the detriment of precision. It is corroborated with the lists of *key-words*.

*Key words*: simple words, which accept any grammatical form (noun, verb, adjective, singular, plural, masculine, feminine) and any orthographical form. The lists of key-words are non-ordered collections.

Example :

***The key-words*** of the course herewith are : cataloguing, classification, classification language, UDC, indexing, pre-coordinated indexing, post-coordinated language

**Controlled language** (on the *terminological level*) has been construed before the document indexing, being represented by the ***authority lists (subject headings)*** and by the ***thesauri of descriptors***.

***Subject headings*** : *non-ordered collection of concepts* (words, phrases), used so as to uniquely represent (without synonymy and/or polysemy) the *content* of the documents.

***Thesauri of descriptors:*** *structured list of concepts;* used in order to uniquely represent the *content* of the document; *controlled language* (i.e. the concepts are expressed through words in a standardized grammatical form); it contains a limited number of terms connected through (semantic) relations of: equivalence, hierarchy or association.

Examples of controlled languages:

LCSH (Library of Congress Subject Headings) – Library of Congress

RAMEAU (Repertoire d’Autorite-Matiere Encyclopedique et Alphabetique Unifie) - Bibliotheque Nationale de France

**Coordinated languages** (on the level of syntax):

- A subject is no longer globally formulated (as in the case of the hierarchical classifications).
- It consists in a series of basic concepts, resorting to a combinatory language.

There are **two types** of coordinated indexing : *pre-* *post-coordinated*.

## **Pre-coordinated Indexing:**

*The concepts are ready-defined by the indexer and are represented through more or less complex terms.*

Example:

Both the *classification*, and the *RAMEAU language*, are ***pre-coordinated languages***, this way:

for the subject *History of the Cinema in Romania between 1950-1989*,

***UDC classification* is: 791.43(498)"1950/1989"**

***RAMEAU Indexing*** with subject headings:

**Cinema – Romania – 1950-1989**

### ***Post-Coordinated Indexing:***

The document is allotted simple terms, which are to be logically combined during the research moment.

Example: Suppose we have a literary work that tackles the issue *national sovereignty*.

From the point of view of the *Theory and Practice of International Law*:

***Decimal classification*** would be: **341.211.01**, including a few important aspects pertaining to law, which are: *International Law, Subjects of International Law, State Sovereignty and Theory of Law*;

***Indexing with Descriptors*** would materialize this way :

**Sovereignty**

**International Law**

**Subjects of International Law**

**International relations**

**Theory of Law**

From of the standpoint of the *Theory of State*:

***UDC Classification is :*** **321.01**

The translation in ***descriptors of the:*** **Theory of State**  
***classification*** **Sovereignty**

From the standpoint of the *Constitutional Law*:

***UDC Classification is:*** **342.3**

***The Indexing through Descriptors :*** **Constitutional Law**  
**Sovereignty**

If a material deals with this theme (sovereignty) in the framework of the integration within the European Union under all three aspects discussed above, then:

***Decimal classification:*** **321.01**

**341.211(4):341.17UE**

**342.3(4):341.17UE**

***Indexing through descriptors*** : National sovereignty

Theory of State

Constitutional Law

International Law

Supranational Institutions to the States

European Integration

European Union

European States

In order to retrieve these works within an automated catalogue (OPAC – Online Public Access Catalogue), the user (either reader or bibliographer) must proceed through *intersection* and *restriction* of the above enumerated terms.