Elvira de Andrade Oliveira
Brasilian Institute of Bibliography and Documentation (IBBD), Rio de Janeiro

# An Automatic Index in Natural Language for UDC Editions

Oliveira, E. de Andrade: An **automatic** index **in natural** language **for UDC editions.**
In: Intern. Classificat. 2 (1975) No. 2, p. 89–93

Description of the procedures for and the advantages of the automatic creation of index entries from the class descriptions in the schedules of the UDC. Examples show the addings to or alterations of the class descriptions in order that the index entries may be truly informative about the concept contained in a "classification number" (UDC-notation). In the annex sample pages are given of the computer printouts listing the cards, the entries in the schedule and the entries in the index, all relating to one sample page of UDC 532.614–533.42.
(I. C.)

## 0. Introduction

The problems of manual preparation of the indexes of UDC editions led us to study their automatic preparation.

A System was created, named ELVCDU3, which is based on the processing of natural language existing in the schedules in order to obtain an automatic preparation of the alphabetic index.

The System can work in any language and it requires only the inclusion of the translation of some grammatically important words in the programme of the System.

As the UDC schedules have not been written with their automatic transformation in an index in mind, a revision will have to be accomplished for small alterations in the text so that it may be adapted to the System. For an example of the existing UDC texts in the schedules see the page 73 of (2) in 7.1.

## 1. Description of the System

### 1.1 Format for data input

The text of a schedule is key-punched in cards, according to the following format:

The classification number is punched starting at column 1. The text related to this number is punched, preceded by two blank columns, and it can be continued in a new card, following the same format. The notes are punched following the same disposition they have in the schedule. For this purpose, the same classification number is punched in a new card, but three blank spaces are left before the punching of the note. See the printout of the listing of the cards in 7.2.

In each card, column 80 is reserved for punching information related to the type of schedule that will be processed; in the future, we will be able to have the UDC full edition on magnetic tapes for the processing of each complete edition, of the medium or abridged edition and, if necessary, of the whole full edition.

### 1.2 Processing

#### *1.2.1 Index creation*

The punched cards are recorded on magnetic tape. The System separates each sentence of the text and analyses each one of its elements (words or symbols); from this analysis, the elements considered grammatically important create new sentences when necessary. Each sentence so created is rotated in order to allow each significant word of the sentence to stand out in the index.

A vocabulary of non-significant words in the programme allows these words to be eliminated and thus not to be considered as main entries in the index; they appear merely as logical linking elements in the sentences.

As a general rule, one period or two spaces indicate the end of the sentence. Its elements are distributed for the sentence formation:

532.68 Motions due to capillary forces[1]
Sentences for the index:
Motions
due to capillary forces 532.68
Capillary
forces, Motions due to 532.68

Several guidelines were established for the formation of new sentences, according to the order of the elements in the original sentence:

1. The elements comma (,) and "AND", even when followed by a preposition or by an article, lead to the formation of new sentences:
532.78 Crystallisation of liquids and of solutions
Sentences for the index:
Crystallisation
of liquids 532.78
of solutions 532.78
Liquids
crystallisation of 532.78
Solutions
crystallisation of 532.78
533.24 Liquefaction and solidification of gases
Sentences for the index:
Liquefaction
of gases 533.24
Solidification
of gases 533.24
Gases
Liquefaction of 533.24
Solidification of 533.24

1 Examples were taken from the Abridged English Edition (1) and the English Full Edition (2)

As an exception, these two elements constitute one sentence when followed by a possessive:

532.1 Universe and its structure

Sentences for the index:

Universe
and its structure 523.1

Structure
Universe and its 523.1

77.03 Photographic records and their interpretation

Sentences for the index:

Photographic
records and their interpretation 77.03

Records
and their interpretation 77.03

Interpretation
Photographic records and their 77.03

2. The element "OR" interrupts the sentence, and the preceding elements will begin the next sentence to be created:

331.845 Use of leisure or spare time

Sentences for the index:

Leisure
Use of 331.845

Spare
time, Use of 331.845

Time
Use of spare 331.845

3. The elements "OTHER" and "ANOTHER" are considered the end of a sentence, and the text that follows will be considered a new sentence:

661.31 Potash and other potassium compounds

Sentences for the index:

Potash 661.31

Potassium
compounds 661.31

Compounds
Potassium 661.31

4. The word or sentence that precedes the colon (:) is added to the words or sentences after the colon, in parentheses, as an explanation:

347.811 Aircraft: purchase, construction, registration

Sentences for the index:

Aircraft 347.811

Purchase (Aircraft) 347.811

Construction (Aircraft) 347.811

Registration (Aircraft) 347.811

5. The word or sentence in parentheses must have a complete meaning to allow its individual retrieval in the index:

532-2 Visible light rays (spectrum)

Sentences for the index:

Visible
light rays 532-2

Light
rays, Visible 535-2

Rays
Visible light 535-2

Spectrum 535-2

These guidelines were established through the analysis of the linguistic structure of the UDC schedules.

*1.2.2 Index listing* (see also 7.4)

After the alphabetic ordering, the index is listed in two columns with each significant word of the sentence standing out.

At the top of each column, the explanation (cont.) is added to the word that stands out, when it is a continuation of the preceding column.

When there is more than one classification number for one sentence, all of them are cited between commas. The cuts made in long sentences are pointed out by the symbols (<=).

*1.2.3 Schedule listing* (see also 7.3)

The schedule is listed in one column and will later be listed in two columns.

Classification numbers having up to 3 digits stand out by being centralised in the line together with their text.

*1.2.4 Updating and corrections*

The System allows inclusion, elimination and substitution of cards in the file on magnetic tape, thus permitting the continuous updating of the file.

## 2. Equipment and language

Programmes were written in PL1 language and processed in an IBM 370 mod. 165 computer.

## 3. Utilisation of the System

The system was used in the processing of the following UDC Editions:

1) UDC Medium Edition, in Portuguese
Classification numbers: 35,759
Sentences in the index: 154,386
Number of index pages: 1,446
Number of schedule pages: 781
Processing time (CPU):
Index creation: 36 minutes
Alphabetic ordering: 57 seconds
Index listing: 11 minutes
Schedule listing: 90 seconds

2) UDC Full Edition, 32 Politics, in Portuguese

## 4. Tests made with the System

Tests were made with other languages, such as English, French, German and Spanish, with positive results and without changes in programmes.

## 5. Advantages

With this System, after a speed revision of the text, the only thing that must be done is the punching of the input cards, since the preparation of the index and the schedules are done by the computer, and for publication it is sufficient to reproduce by offset the computer listings. Thus, we have eliminated the slowness of manual preparation of the cards to be alphabetized for the index, their typing and typographic composing of the index as well as of the schedules.

532.614 Surface and boundary layer energy
.2 Surface energy
.3 Boundary layer energy
615 Surface energy of dissolved substances

532.62 Properties of thin liquid films, etc. Soap bubbles. Minimum surfaces. Plateau's figures

.623 Soap bubbles

532.63 Capillarity in the strict sense

For electro-capillarity see 537.361
.631 Capillary pressure
.632 Capillary energy
.635 Surface tenacity and rigidity

532.64 Angle of contact. Wetting angle

.08 Measurement of the angle of contact

532.65 Free surface under the action of capillary forces

532.66 Capillary tubes

532.68 Motions due to capillary forces

.685 Moisture movement in porous bodies, e.g. in porous stones

532.69 Various properties depending on surface tension

.691 Surface activity in general
.692 Power of forming thin films. Oiliness
.694 Properties of gas-liquid systems
.1 Foam formation
.4 Fluffiness
.695 Behaviour relating to formation of emulsions
.1 Power of forming emulsions
.2 Power of being emulsified
.696 Behaviour relating to solids
.1 Wetting
.2 Washing

532.7 KINETIC THEORY OF LIQUIDS. OSMOSIS. DISSOLUTION. SOLUTION

532.71 Osmosis

.712 Osmotic pressure

.713 Osmotic energy. Heat of solution

532.72 Diffusion

532.73 Solution. Phenomena of solution

-1 ___Of gases in liquids
-2 ___Of liquids in liquids
-3 ___Of solid bodies in liquids
.731 Velocity of solution
.733 Mutual influence of dissolved substances on the solubility
.739 Other phenomena
.2 Solubility. Solvent power

532.74 Molecular state. Association

532.77 Physical properties of solutions. Influences on the boiling point and freezing point and on the vapour pressure

The influence of the solute on the properties of the solvent is classified by using the colon, e.g. 532.77 : 532.612.4 Surface tension of solutions
.771 Dilute solutions
.772 Concentration with reference to the quantity of the different constituents
.773 Concentrated solutions

532.78 Crystallisation of liquids and of solutions

See also 548.5 Formation and growth of crystals
.781 Crystallisation of pure liquids
.782 Undercooled metastable liquids. Solid solutions
.785 Crystallisation of solutions
.787 Supersaturated solutions

# 533 Mechanics of Gases. Aerodynamics

See also
621.5 Compression and rarefaction of gases
621.6 Pumping and blowing machinery
661.9 Gases in chemical technology

Under 533 are to be classed only works referring exclusively to gases. Those referring to gases and liquids come preferably under 532. If necessary, subdivisions of 522 can be combined with those of 532 with the aid of the colon

533.1 PROPERTIES OF GASES

.11 Specific volume
.12 Density
For determination of density see 531.758
.13 Principle of Archimedes applied to gases. Correction of weight. Upthrust
See also 533.61 Theory of balloons
.15 Diffusion and penetration. Inflow
.16 Internal friction and viscosity
.17 Outflow

533.2 ELASTICITY. COMPRESSIBILITY. LIQUEFACTION. GAS MIXTURES

.21 Compressibility
.22 Approximate laws. Boyle's or Mariotte's law
.23 Limit of Boyle's law. Experiments of Pouillet, Dulong, Arago and Regnault
.24 Liquefaction and solidification. Liquid air
See also
661.9 Gases in chemical technology
621.56 Refrigeration
.27 Mixtures of gases. Properties
.275 Humidity. Hygrometry
See also 551.57 Aqueous vapour
.276 Katathermometers. Katathermometry

533.4 BAROMETERS

For manometers see 531.787
.41 Barometers containing liquids, mercury, etc.
.42 Corrections and reductions for liquid barometers

*7.1 Copy of the UDC page 73*

Since the text is in accordance with the System, corrections and updating of the text and the automatic processing of the index and schedules are easy to accomplish because all that is necessary is the punching of cards with the new text and their inclusion in the UDC file through an appropriate programme.

In the future, we will have the Full UDC Edition, in Portuguese, on computer file.
Another advantage of the System is the high processing speed.

## 6. References


1) International Federation for Documentation: *Universal Decimal Classification. Abridged English Edition*. 3rd ed. rev., 1961. London: British Standards Institution 1961. 245 p. = FID Publ. No. 289.

2) International Federation for Documentation: *Universal Decimal Classification. English Full Edition*. Fourth Intern. Ed., Vol. 2, Pt. 1, Classes 50/53. London: British Standards Institution 1943. 103 p.

3) Oliveira, Elvira de Andrade: *Automação dos índices das tabelas da Classificação Decimal Universal*. In: Ciencia da Informação, Rio de Janeiro, 2 (1973) No. 2, p. 139–168 (Dissertation to obtain the Master Degree in Librarianship and Documentation)


## 7. Annexes

The sample used is taken from the English Full Edition (2); the modifications of the text are indicated by lines and underlinings in the copy of the UDC page (7.1) as well as in the printout of the schedule (7.3).

7.1 Copy of the UDC page 73
7.2 Listing of the cards
7.3 Schedule
7.4 Index

```
532.614  SURFACE ENERGY. BOUNDARY LAYER ENERGY
532.614.2  SURFACE ENERGY
532.614.3  BOUNDARY LAYER ENERGY
532.615  SURFACE ENERGY OF DISSOLVED SUBSTANCES
532.62  PROPERTIES OF THIN LIQUID FILMS, ETC.. SOAP BUBBLES. MINIMUM SURFACES.
532.62  PLATEAU'S FIGURES
532.623  SOAP BUBBLES
532.63  CAPILLARITY IN THE STRICT SENSE
532.63  FOR ELECTRO-CAPILLARITY SEE 537.361
532.631  CAPILLARY PRESSURE
532.632  CAPILLARY ENERGY
532.635  SURFACE TENACITY AND RIGIDITY
532.64  ANGLE OF CONTACT. WETTING ANGLE
532.64.08  MEASUREMENT OF THE ANGLE OF CONTACT
532.65  FREE SURFACE UNDER THE ACTION OF CAPILLARY FORCES
532.66  CAPILLARY TUBES
532.68  MOTIONS DUE TO CAPILLARY FORCES
532.685  MOISTURE MOVEMENT IN POROUS BODIES.  E.G. IN POROUS STONES
532.69  VARIOUS PROPERTIES DEPENDING ON SURFACE TENSION
532.691  SURFACE ACTIVITY IN GENERAL
532.692  POWER OF FORMING THIN FILMS. OILINESS
532.694  PROPERTIES OF GAS-LIQUID SYSTEMS
532.694.1  FOAM FORMATION
532.694.4  FLUFFINESS
532.695  BEHAVIOUR RELATING TO FORMATION OF EMULSIONS
532.695.1  POWER OF FORMING EMULSIONS
532.695.2  POWER OF BEING EMULSIFIED
532.696  BEHAVIODR RELATING TO SOLIDS
532.696.1  WETTING
532.696.2  WASHING
532.7  KINETIC THEORY OF LIQUIDS. OSMOSIS. DISSOLUTION. SOLUTION
532.71  OSMOSIS
532.712  OSMOTIC PRESSURE
532.713  OSMOTIC ENERGY. HEAT OF SOLUTION
532.72  DIFFUSION
532.73  SOLUTION, PHENOMENA OF SOLDTION
532.73-1  SOLUTION OF GASES IN LIQUIDS
532.73-2  SOLUTION OF LIQUIDS IN LIQUIDS
532.73-3  SOLUTION OF SOLID BODIES IN LIQUIDS
532.731  VELOCITY OF SOLUTION
532.733  MOTOAL INFLUENCE OF DISSOLVED SUBSTANCES ON TUE SOLUBILITY
532.739  OTHER PHENOMENA OF SOLUTION
532.739.2  SOLUBILITY. SOLVENT POWER
532.74  MOLECULAR STATE. ASSOCIATION
532.77  PHYSICAL PROPERTIES OF SOLUTIONS. INFLUENCES OR THE BOILING AND
532.77  FREEZING POINT OF SOLUTION. INFLUENCES ON THE VAPOUR PRESSURE OF
532.77  SOLUTION
532.77  THE INFLUENCE OF THE SOLUTE ON THE PROPERTIES OF THE SOLVENT IS
532.77  CLASSIFIED BY USING THE COLON, E.G. 532.77:532.612.4 SURFACE TENSION
532.77  OF SOLUTIONS
532.771  DILUTE SOLUTIONS
532.772  CONCENTRATION WITH REFERENCE OF QUANTITY OF THE DIFFNNENT CONSTITDENTS
532.773  CONCENTRATED SOLUTIONS
532.78  CRYSTALLISATION OF LIQUIDS AND OF SOLUTIONS
532.78  SEE ALSO 548.5 FORMATION AND GROWTH OF CRYSTALS
532.781  CRYSTALLISATION OF PURE LIQUIDS
532.782  UNDERCOOLED METASTABLE LIQUIDS. SOLID SOLUTIONS
532.785  CRYSTALLISATION OF SOLDTIONS
532.787  SUPERSATUBATED SOLUTIONS
533  MECHANICS OF GASES. AERODYNAMICS
533  SEE ALSO
533  621.5  COMPRESSION AND RAREFACTION OF GASES
533  621.6  PUMPING AND BLOWING MACHINERY
533  661.9  GASES IN CHEMICAL TECHNOLOGY
533  UNDER 533 ARE TO BE CLASSED ONLY WORKS REFERRING EICLUSIVELY TO GASES.
533  THOSE REFERRING TO GASES AND LIQUIDS COME PREFERABLY UNDER 532. IF
533  NECESSARY, SUBDIVISIONS OF 522 CAN BE COMBIVED WITH THOSE OF 532 WITH
533  AID OF THE COLON
533.1  PROPERTIES OF GASES
533.11  SPECIFIC VOLUME OF GASES
533.12  DENSITY OF GASES
533.12  FOR DETERMINATION OF DENSITY SEE 531.758
533.13  PRINCIPLE OF ARCHIMEDES APPLIED TU GASES. CORRECTION OF WEIGHT OF
533.13  GASES. UPTHRUST
533.13  SEE ALSO  533.61  THEORY OF BALLOONS
533.15  DIFFUSION AND PENETRATION OF GASES. INFLOW
533.16  INTERNAL FRICTION AND VISCOSITY OF GASES
533.17  OUTFLOW
533.2  ELASTICITY, COMPRESSIBILITY AND LIQUEFACTION OF GASES. GAS MIXTURES
533.21  COMPRESSIBILITY OF GASES
533.22  APPROXIMATE LAWS. DOYLE'S AUD MARIOTTE'S LAW
533.23  LIMIT OF HOYLE'S LAW. EXPERIMENTS OF POUILLET, DULONG, ARAGO AND
533.23  REGNAULT
533.24  LIQUEFACTION AND SOLIDIFICATION OF GASES. LIQUID AIR
533.24  SEE ALSO
533.24  661.9  GASES IN CHEMICAL TECHNOLOGY
533.24  621.56  REFRIGERATION
533.27  MIXTORES AND PROPERTIES OF GASES
532.275  HUMIDITY. HYGROMETRY
532.275  SEE ALSO 551.57  AQUEOUS VAPOUR
533.276  KATATHERMO.ETERS. KATATHERMOMETRY
533.4  BAROMETERS
533.4  FOR MANOMETERS SEE 531.787
533.41  BAROMETERS CONTAINING LIQUIDS, MERCURY, ETC.
533.42  CORRECTIONS AND REDUCTIONS FOR LIQUID BAROMETERS
```

*7.2 Listing of the cards*

```
532.614    SURFACE ENERGY. BOUNDARY LAYER ENERGY
532.614.2    SURFACE ENERGY
532.614.3    BOUNDARY LAYER ENERGY
532.615    SURFACE ENERGY OF DISSOLVED SUBSTANCES
532.62    PROPERTIES OF THIN LIQUID FILMS, ETC. SOAP BUBBLES. MINIMUM SURFACES
          PLATEAU'S FIGURES
532.623    SOAP BUBBLES
532.63    CAPILLARITY IN THE STRICT SENSE
           FOR ELECTRO-CAPILLARITY SEE 537.361
532.631    CAPILLARY PRESSURE
532.632    CAPILLARY ENERGY
532.635    SURFACE TENACITY AND RIGIDITY
532.64    ANGLE OF CONTACT. WETTING ANGLE
532.64.08    MEASUREMENT OF THE ANGLE OF CONTACT
532.65    FREE SURFACE UNDER THE ACTION OF CAPILLARY FORCES
532.66    CAPILLARY TUBES
532.68    MOTIONS DUE TO CAPILLARY FORCES
532.685    MOISTURE MOVEMENT IN POROUS BODIES.  E.G. IN POROUS STONES
532.69    VARIOUS PROPERTIES DEPENDING ON SURFACE TENSION
532.691    SURFACE ACTIVITY IN GENERAL
532.692    POWER OF FORMING THIN FILMS. OILINESS
532.694    PROPERTIES OF GAS-LIQUID SYSTEMS
532.694.1    FOAM FORMATION
532.694.4    FLUFFINESS
532.695    BEHAVIOUR RELATING TO FORMATION OF EMULSIONS
532.695.1    POWER OF FORMING EMULSIONS
532.695.2    POWER OF BEING EMULSIFIED
532.696    BEHAVIOUR RELATING TO SOLIDS
532.696.1    WETTING
532.696.2    WASHING
532.7    KINETIC THEORY OF LIQUIDS. OSMOSIS. DISSOLUTION. SOLUTION
532.71    OSMOSIS
532.712    OSMOTIC PRESSURE
532.713    OSMOTIC ENERGY. HEAT OF SOLUTION
532.72    DIFFUSION
532.73    SOLUTION, PHENOMENA OF SOLUTION
532.73-1    SOLUTION OF GASES IN LIQUIDS
532.73-2    SOLUTION OF LIQUIDS IN LIQUIDS
532.73-3    SOLUTION OF SOLID BODIES IN LIQUIDS
532.731    VELOCITY OF SOLUTION
532.733    MUTUAL INFLUENCE OF DISSOLVED SUBSTANCES ON THE SOLUBILITY
532.739    OTHER PHENOMENA OF SOLUTION
532.739.2    SOLUBILITY. SOLVENT POWER
532.74    MOLECULAR STATE. ASSOCIATION
532.77    PHYSICAL PROPERTIES OF SOLUTIONS. INFLUENCES ON THE BOILING AND
          FREEZING POINT OF SOLUTION. INFLUENCES ON THE VAPOUR PRESSURE OF
          SOLUTION
           THE INFLUENCE OF THE SOLUTE ON THE PROPERTIES OF THE SOLVENT IS
          CLASSIFIED BY USING THE COLON, E.G. 532.77:532.612.4 SURFACE TENSION
          OF SOLUTIONS
532.771    DILUTE SOLUTIONS
532.772    CONCENTRATION WITH REFERENCE OF QUANTITY OF THE DIFFERENT CONSTITUENTS
532.773    CONCENTRATED SOLUTIONS
532.78    CRYSTALLISATION OF LIQUIDS AND OF SOLUTIONS
           SEE ALSO 548.5 FORMATION AND GROWTH OF CRYSTALS
532.781    CRYSTALLISATION OF PURE LIQUIDS
532.782    UNDERCOOLED METASTABLE LIQUIDS. SOLID SOLUTIONS
532.785    CRYSTALLISATION OF SOLUTIONS
532.787    SUPERSATURATED SOLUTIONS

                533    MECHANICS OF GASES. AERODYNAMICS
                                  SEE ALSO
                  621.5  COMPRESSION AND RAREFACTION OF GASES
                    621.6  PUMPING AND BLOWING MACHINERY
                      661.9  GASES IN CHEMICAL TECHNOLOGY
          UNDER 533 ARE TO BE CLASSED ONLY WORKS REFERRING EXCLUSIVELY TO GASES.
            THOSE REFERRING TO GASES AND LIQUIDS COME PREFERABLY UNDER 532. IF
        NECESSARY, SUBDIVISIONS OF 522 CAN BE COMBINED WITH THOSE OF 532 WITH THE
                                AID OF THE COLON

533.1    PROPERTIES OF GASES
533.11    SPECIFIC VOLUME OF GASES
533.12    DENSITY OF GASES
           FOR DETERMINATION OF DENSITY SEE 531.758
533.13    PRINCIPLE OF ARCHIMEDES APPLIED TO GASES. CORRECTION OF WEIGHT OF
          GASES. UPTHRUST
           SEE ALSO  533.61  THEORY OF BALLOONS
533.15    DIFFUSION AND PENETRATION OF GASES. INFLOW
533.16    INTERNAL FRICTION AND VISCOSITY OF GASES
533.17    OUTFLOW
533.2    ELASTICITY, COMPRESSIBILITY AND LIQUEFACTION OF GASES. GAS MIXTURES
533.21    COMPRESSIBILITY OF GASES
533.22    APPROXIMATE LAWS. BOYLE'S AND MARIOTTE'S LAW
533.23    LIMIT OF BOYLE'S LAW. EXPERIMENTS OF POUILLET, DULONG, ARAGO AND
          REGNAULT
533.24    LIQUEFACTION AND SOLIDIFICATION OF GASES. LIQUID AIR
           SEE ALSO
           661.9  GASES IN CHEMICAL TECHNOLOGY
           621.56  REFRIGERATION
533.27    MIXTURES AND PROPERTIES OF GASES
532.275    HUMIDITY. HYGROMETRY
           SEE ALSO 551.57  AQUEOUS VAPOUR
533.276    KATATHERMOMETERS. KATATHERMOMETRY
533.4    BAROMETERS
           FOR MANOMETERS SEE 531.787
533.41    BAROMETERS CONTAINING LIQUIDS, MERCURY, ETC.
533.42    CORRECTIONS AND REDUCTIONS FOR LIQUID BAROMETERS
```

*7.3 UDC Schedule*

*7.4 Index*