

INDEX OF KEYWORDS : Index Terminorum (ITER)

The index of keywords (*nomina appellativa* and *nomina propria*) that is being prepared by Termini will be a cumulation of special lists that cover one knowledge domain or 'ontology' each. Termini started extracting keywords from works of the Camena 'Thesaurus eruditionis', adding (broadly speaking:) 'semantic' data offered by these works, as far as automatic data mining is able to capture them. The data scheme we devised for this stage of our work comprises 24 fields (see below). Later on, more fields will be added, especially for including morphological data (often just the IDN of the corresponding Morpheus entry?), for mapping the keyword (or rather the entity it means or refers to) onto the most current systems of indexation (the German National Library authority lists of personal names and subject headings) and classification (the Dewey Decimal Classification), and possibly for including more types of lexical and semantic data.

Most fields of the basic data structure allow for more than one entry, but LEM, IDN and URL for one only. The maximum space available for each field (e.g. 100 keys) may seem to be too narrowly limited in some cases. These limits have been fixed for data extracted from one source only. A data set constituted by cumulated data, however, will often require more space.

24 fields (arranged by section, cf. ISO 12620:1999 'Computer applications in terminology')

IDN = IDentification Number (attributed automatically according to the sequence of insertion in the Termini database). - 7 keys (= limit 10 Mio. entries)

LEM = LEMma. The keyword may consist of several words that form a fixed expression designating one entity (as is usually the case with names). In order to avoid ambiguity a domain or class name (e.g. <zooology>, <instrument>) may be added to the keyword. - 70 keys

URL = URL of the original page image exhibiting the keyword. This data implicitly contains the title of the source. - 90 keys

BIB = BIBliographical reference: An author or work cited as a source. The exact position in the source cited should be recorded if mentioned. We may enter the form offered by the present 'Thesaurus' resource, as we will establish a concordance of the short forms that occur in our resources and the full forms as contained in current authority lists. - 200 keys

VAR = VARiant form of non-trivial nature, i.e. a variant that belongs to 'la langue' rather than to 'la parole' (or rather: 'la graphie'). 'Paleographic' (e.g. coelum and celum instead of caelum) and 'didactic' variants (e.g. rubîgo or rubîgo instead of rubigo) are not recorded here. - 100 keys

ABB = ABBreviation. An Abbreviation that in our source occurs as a lemma is entered in this field, too, with the full form entered in the lemma field. - 25 keys

SYN = SYNonym. - 100 keys

MET = METasemy, i.e. a figure of speech replacing the 'proper' word of the lemma field by a word or phrase that literally designates something else (*translatio vel immutatio verborum*). - 100 keys

PHR = PHRase: The lemma and one or several words that use to accompany it, such as a formulaic epitheton, a verb or an idiomatic turn. - 200 keys

LAT = LATin. This field is needed only when the lemma itself is not Latin. - 100 keys

GRE = GREek (ancient as well as Byzantine Greek). - 100 keys

ENG = ENGLISH. - 100 keys

GER = GERman. - 100 keys

FRE = FREnch. - 100 keys

ITA = ITAlian. - 100 keys

SPA = SPAnish. - 100 keys

DUT = DUTch. - 100 keys

LAN = LANguage not listed above or not identified. - 50 keys

DEF = DEFinition or description or paraphrase as offered by the source. - 120 keys

HEA = HEAding as contained in the source. - 100 keys

BRO = BROader term. It is often contained in a definition or heading offered by the source text. It should preferably be a close 'parent' of the lemmatic entity (e.g. LEM = dog, BRO = mammal), not a distant general classification (e.g. LEM = dog, BRO = animal). An alternative hierarchy may be represented by an added entry (e.g. LEM = dog, BRO = mammal ; BRO = domestic animal). - 70 keys

SIT = SITe, SITuation: The place, region, institution or complex entity the object designated by the keyword is situated in. - 70 keys

DAT = DATe(s): Life-time, period, date of an event. - 30 keys

VID = VIDE, i.e. the designation of another segment of the source work (e.g. another lexicon article). This field is for provisional (temporary) use only. The reference may exhibit a synonym, a lexical variant, a foreign language equivalent, a broader term or a complex entity that includes the entity designated by the keyword. This data can be used for enriching the data set of the keyword referred to and/or of the present keyword. - 70 keys

P.S.: This conception is not based on epistemological, semiotic or logical analysis, but on experience in library subject retrieval of old as well as recent books. The categories are in part determined by their potential use in formulating thematic (rather than textual surface) queries, and by the opportunities and limits of data extraction from our resources. Therefore, 'mixed' categories are not avoided.

W. Schibel, 2006