GEOINFO IV - SECOND CIRCULAR

4th INTERNATIONAL CONFERENCE ON GEOSCIENCE INFORMATION

PRELIMINARY PROGRAM
and
REGISTRATION BULLETIN

June 24-29, 1990
Ottawa, Ontario, CANADA

GEOINFO IV

Jacques GRAVESTEJIN (COGEODOC), Roberto POTENZA (COGEODOC,CNR), Oreste SIGNORE (CNR-CNUCE), Bruno TESTA (CNR)

THE MULTILINGUAL THESAURUS OF GEOSCIENCES: STRUCTURE AND MANAGEMENT TOOLS

GENERAL REMARKS

The MULTILINGUAL THESAURUS OF GEOSCIENCES is the core vocabulary common to thesauri and lexicons used by the various database producers in eight countries: Federal Republic of Germany (BGR), France(BRGM/CNRS), Italy (CNR), Spain (ITGE), USA (AGI), Chechoslovakia (GEOFOND) and Finland (Geol. Survey).

The MT is being prepared by the Commission on Geological Documentation (COGEODOC) of the IUGS (International Union of Geological Sciences) through a Joint Working Group (WGMT), sponsored by IUGS and the International Council of Scientific and Technical Information (ICSTI).

The objective of the contributors is the development of a geological vocabulary which can practically be used as a common terminological reference by different documentation centers.

The MT offers to the national bodies in charge of geoscience information the opportunity of creating national databases, compatible with the major international documentation systems, using one or more languages in accordance with their own terminological needs.

In the multilingual environment of the international networks the MT can play the role of a switching mechanism for the automatic translation of indexing terms in exchange procedures, making easier to share the informations among different national centers.

The first issue of the Multilingual Thesaurus was published by Pergamon in 1988 in six languages: English, French, German, Italian, Russian and Spanish, and immediately the WGMT started the main tasks for the following five years' period:

-updating and expansion of the first MT database for each of the specific Geoscientific subfields of the Thesaurus

- expansion of the systematics of Paleontology, Stratigraphy, Mineralogy and of sedimentary, igneous and metamorphic rocks

The provided expansion will almost duplicate the content of the MT, presently summing up to approximately 5000 entries, as an almost equivalent amount of complementary terms will be added, after a selection based on the frequency of use in existing databases, indexing practice and linguistic criteria.

STRUCTURE

The logical structure of the MT is of a mixed type, as the terms are jerachically grouped into a limited number of themes (36), but the keywords , divided in Descriptors and Non descriptors, are connected by a network of USE relations:

The Descriptors are self-standing terms, which can be used without restrictions for indexing: each one is related to only one theme ("Broader Term" relation) and, possibly, to an indefinite number of Non-descriptors as preferred term ("Use For" relation);

The Non-descriptors are also related to one theme, but must also be related to one, or to a couple of Descriptors, whose combined significance matches at best the meaning of the term (Use, And relations).

Each entry is defined by a KEY, chosen mainly among the entries of the A.G.I. Glossary of Geology, and is related to one of the 36 main themes of the Thesaurus.
An unique identification label is used for recalling purposes of the management programs.
The key is followed by the translations in the eight languages of the MT.
The record is closed by a string of characters defining the documentary role of the term in each language (D for Descriptors, N for Non-descriptors, A for Adjectives, + for general terms, - for terms having no translation in one language).

In the 1988 issue, the relations are restricted to the indication of the Descriptors recommended to replace the Non-descriptors (one or two). A more complex relational network is provided in the next developments.

INFORMATIC TOOLS

At present the management of the Multilingual Thesaurus is on care of the Group of Informatic Geology of the Italian National Council of Research (CNR); tools for the automatic treatment of the data are also prepared or tested by the CNUCE-CNR.

The MT database includes the six languages of the printed issue 1988, and also the Czech and Finnish versions. Studies about the possible joining of new languages (Portuguese and Arabic are the first candidates) are underway.

The first management tools for treating the MT were cumbersome home-made FORTRAN programs, of very poor flexibility and surely not user-friendly at all.

More skilled systems were soon adopted, and an extended use of PCs was provided, expecially for local use and updating. Mainframe systems were however still used, mainly at the CNUCE-CNR in Pisa, for structure development and testing.

The presently circulating version of the MT is implemented on the DbaseIII system; specific programs allow to manage it in a "relational" way. A fair portability of the compiled programs has been obtained, although the homogeneity of the PCs is still too scarce to completely ensure from failures in the installation.

Other informatic packages are being tested to improve the relational properties of the system and to widen its portability to other documentary systems.

Among the tools developed by the CNUCE - CNR are the following implementations:

SQL/DS version, on IBM mainframe, in view of the integration of the MT in a structure relating it to the other terminological tools managed by the CNR

PS2 versions for local management of the TM:

-MS-DOS environment with ORACLE/Windows ("C" language)

-OS2 environment with Query Manager/Presentation Manager ("C2" language)

An implementation of the MT in CDS-ISIS is also led by UNESCO, in view of the use of the MT for the development of a documentary project, related to the Pan African Network of Geological Information Systems (PANGIS).

This wide spectrum of tools and applications surely will improve the portability of the MT; this variety, together with the geographic dispersion of the contributors to the MT program, will however raise serious problems of consistency of the updatings and, consequently of homogeneity, among the versions exploited in the different places.

Table 1

LIST OF THEMES

PALS Paleontology-Systematics
SURF Geomorphology-Quaternary geology
PHCH Physical and chemical properties, processes
IGNE Petrology of igneous rocks
STRU Structural geology
PALE Paleontology
SEDI Sedimentology
ECON Economic geology
STRS Stratigraphy-Systematics
SOLI Solid Earth geophysics
GEOL General geology
TEST Textures-structures
MISC Miscellaneous
GEOH Hydrology
METH Methods
ENGI Engineering geology
MATH Mathematical geology
MINE Mineralogy
CHEE Elements
COMS Commodities
STRA Stratigraphy
APPL Applied geophysics
SEDS Sedimentary rocks-Systematics
SUSS Soils-Systematics
CHES Chemical compounds
MARI Marine geology
EXTR Extraterrestrial geology
ISOT Isotope geochemistry / Geochronology,absolute age
IGMS Metamorphic rocks-Systematics
IGNS Igneous rocks-Systematics
EXTS Meteorites,planets
MING Mineral groups
INST Instruments,equipments
ENVI Environment
MINI Mining
GEOC Geochemistry

THEME	D/N		IN MT	USED	NOT US.	FREQUENCY	MAX	
APPL	D	G	68	66	2	33444	4049	507
APPL	N	G	21	6	15	356	269	59
CHEE	D	S	101	93	8	28698	3395	309
CHES	D	S	65	62	3	32244	6139	520
CHES	N	S	12	7	5	222	136	32
COMS	D	S	71	67	4	31691	5461	473
COMS	N	S	27	7	20	810	628	116
ECON	+	G	1	0	1	0	0	0
ECON	A	G	3	0	3	0	0	0
ECON	D	G	133	126	7	62011	5568	492
ECON	N	G	51	8	43	455	380	57
ENGI	D	G	101	96	5	42030	4990	438
ENGI	N	G	17	7	10	1310	1062	187
ENVI	+	G	1	1	0	7	7	7
ENVI	D	G	33	33	0	28255	4751	856
ENVI	N	G	1	0	1	0	0	0
EXTR	A	G	2	0	2	0	0	0
EXTR	D	G	44	35	9	4155	520	119
EXTR	N	G	19	4	15	20	11	5
EXTS	D	S	20	20	0	7019	2365	351
EXTS	N	S	20	17	3	391	49	23
GEOC	D	G	22	19	3	12172	6932	641
GEOC	N	G	7	2	5	20	12	10
GEOH	A	G	1	0	1	0	0	0
GEOH	D	G	106	102	4	58200	6906	571
GEOH	N	G	43	9	34	1651	621	183
GEOL	+	G	15	6	9	41	16	7
GEOL	A	G	14	1	13	13	13	13
GEOL	D	G	89	83	6	57408	6537	692
GEOL	N	G	48	13	35	4243	2169	326
IGMS	D	S	34	31	3	21234	8474	685
IGMS	N	S	20	9	11	458	225	51
IGNE	+	G	1	1	0	45	45	45
IGNE	A	G	9	0	9	0	0	0
IGNE	D	G	160	143	17	56420	2969	395
IGNE	N	G	114	20	94	2701	1235	135
IGNS	D	S	35	35	0	49661	17333	1419
IGNS	N	S	8	6	2	315	119	52
INST	D	C	16	14	2	5380	3892	384
INST	N	C	21	6	15	73	41	12
ISOT	D	G	40	33	7	22307	7417	676
ISOT	N	G	20	3	17	7341	7326	2447
MARI	A	G	1	0	1	0	0	0
MARI	D	G	46	42	4	16274	2077	387
MARI	N	G	22	6	16	686	553	114
MATH	D	C	64	62	2	43381	7233	700
MATH	N	C	39	10	29	198	96	20
METH	+	C	1	0	1	0	0	0
METH	D	C	99	92	7	38704	5552	421
METH	N	C	19	7	12	1963	1823	280

MINE	A	G	1	0	1	0	0	0
MINE	D	G	62	57	5	23448	4056	411
MINE	N	G	39	4	35	41	23	10
MING	D	S	40	40	0	19321	3216	483
MINI	D	G	13	11	2	5967	1783	542
MINI	N	G	20	2	18	16	11	8
MISC	+	C	12	3	9	20	10	7
MISC	A	C	13	0	13	0	0	0
MISC	D	C	93	88	5	63989	7305	727
MISC	N	C	32	8	24	6087	5908	761
PALE	A	G	4	0	4	0	0	0
PALE	D	G	144	139	5	58737	5082	423
PALE	N	G	91	24	67	829	706	35
PALS	D	S	322	271	51	45628	4870	168
PALS	N	S	139	31	108	841	182	27
PHCH	+	C	3	0	3	0	0	0
PHCH	A	C	7	0	7	0	0	0
PHCH	D	C	219	211	8	80754	3504	383
PHCH	N	C	75	21	54	1700	1384	81
SEDI	A	G	2	0	2	0	0	0
SEDI	D	G	161	141	20	65091	7251	462
SEDI	N	G	41	10	31	1148	1053	115
SEDS	D	S	53	53	0	82118	31745	1549
SEDS	N	S	31	23	8	10581	9945	460
SOLI	D	G	124	119	5	65913	7231	554
SOLI	N	G	45	12	33	3438	2974	286
STRA	A	G	1	0	1	0	0	0
STRA	D	G	50	48	2	30664	6938	639
STRA	N	G	43	8	35	4800	4425	600
STRS	D	S	159	158	1	190418	18383	1205
STRS	N	S	15	7	8	222	65	32
STRU	D	G	166	160	6	84740	8124	530
STRU	N	G	105	25	80	3059	1948	122
SURF	+	G	1	1	0	4	4	4
SURF	A	G	8	3	5	1182	1157	394
SURF	D	G	255	240	15	78374	9382	327
SURF	N	G	120	29	91	1915	378	66
SUSS	D	S	31	28	3	2154	327	77
SUSS	N	S	50	10	40	129	50	13
TEST	A	C	7	0	7	0	0	0
TEST	D	C	69	64	5	13083	3089	210
TEST	N	C	78	8	70	281	141	35

Table 3

FIGURES ABOUT THE MULTILINGUAL THESAURUS

CONTENT

Descriptors	3308	67.9%
Non-descriptors	1454	29.9
Adjectives	73	1.5
General terms	35	.7
Total entries	4870	

CATEGORIES

Terms	Descriptors	Non-descriptors	Others
Geoscience	1817 (1693 used)	860 (190)	65 (13)
Systematic	931 (858)	322 (117)	--
Common	560 (531)	264 (60)	43 (3)

TERMS OF THE MT USED IN PASCAL-GEODE
(years 1983-1989)

	used in P-G		not used		
Descriptors	3082	63.3%	226	4.7	
Non-descriptors	369	7.6	1085	22.2	
Others	--	---	108	2.2	
Total entries	3451	70.9	1419	29.1	4870

USE OF MT TERMS IN PASCAL-GEODE

	Total	Geol.	Syst.	Common
Total Descriptors	1.561.087	805610	510186	1292
Mean	506.5			
Total Non-descriptors	58.300	34009	13969	---
Mean	158.0			
Total others (Adjectives,General, non translated)	1312	245291	10302	20
Mean	12			
Total frequence of MT terms	1.620.699			
Mean	467.5			