

Multilingual Thesaurus of Geosciences

Executive Summary

Joachim Gersemann, June 2004

Marketing of geoscience products and services is difficult as long as there is no additional information simplifying their access. Adding reliable and understandable metadata will increase their awareness and improve their value.

Information (e.g. metadata) = data + context. The quality of metadata can be improved by developing and using standards. We need clear definitions to avoid misunderstanding and to mitigate diversity of interpretation for text based metadata. Best practice can be achieved by providing multilingual thesauri.

The predecessor of CGI, CoGeoInfo, developed the Multilingual Thesaurus of geosciences (GRAVESTIJN et. al. 1995). This thesaurus with ca. 6000 terms could be the base for a modern web based indexing and searching tool.

As the CoGeoInfo working group disbanded, CGI took the initiative to establish a new working group. The first meeting for updating the current Version of the Multilingual Thesaurus took place in Burgdorf / Hannover (27.-28.October 2003). This meeting produced the following conclusions:

(1) Mission

To enable the global exchange of geoscience information with the help of a common multilingual core vocabulary by developing and expanding the Multilingual Thesaurus of Geosciences

(2) Fundamental declaration of the CGI:

The Multilingual Thesaurus of Geosciences (MTG) has to be:

- easily accessible (incl. on the Web).
- available free of charge;

(3) Considerations:

The participants recognize the importance of ensuring the compatibility of national information systems with commonly accepted international standards.

(4) Main objectives:

- Updating categories and terms
- Establishing a polyhierarchy
- Adding synonyms and related terms
- Installing the thesaurus as a web based tool
- Adding new languages

The working group decided to elaborate deliveries mainly at home and to meet once/twice a year for consultations. In the meantime all procedures should be coordinated by e-mail

Working group members

Marco Amanti (APAT, Italy), Marielle Arregros-Rouvreau (BRGM, France), Claude Beaupère (BRGM, France), Joachim Gersemann (BGR, Germany) [spokesman], Rachel Heaven (BGS, United Kingdom), Jan Jellema (TNO-NITG, Netherlands), Tomasz Mardal (PGI, Poland), Luca Olivetta (Ministry of the Environment, Italy), Maija Pennanen (GTK, Finland), Sharon Tahirkheli (AGI, USA); Caj Kortman (GTK, Finland)) [in advisory capacity]

Activities

The working group agreed to review the current version at first. This was finished in April 2004 and the results were distributed to all working group members for verification before the second meeting in Utrecht (1.-2. June 2004).

The review of the current version of the MTG revealed some severe shortcomings (Appendix 1). In particular the current relationships and hierarchy were regarded as unsatisfying.

To cope with this problem, the working group started a revision discussing scope and relevance of the 36 categories. In turn of this revision the categories were adapted to modern requirements (Appendix 2).

The next step will be to re-assign the ca. 6000 terms to the new categories. To achieve this we established four teams with 1500 terms each to overwork. At the same time the applicability of the new categories are checked. This will be finalised until the end of October, results will be presented at the next meeting in early November 2004.

An important factor will be the thesaurus management software. It should be able to maintain the database of terms and relationships via remote access by working group members. It also should produce a web application which can be easily used by everyone. At the Utrecht meeting two packages for developing multilingual thesauri were presented. They will be seriously checked during 2004.

Until now the project is supported by CGI and the national agencies of the working group members. A lot of work is done voluntarily.

Ref.:

GRAVESTEIJN, J., KORTMAN, C.; POTENZA, R. & RASSAM, G.N. (1995): Multilingual thesaurus of geosciences.- 645 p.; Medford: Information today.

Appendix 1

Multilingual Thesaurus of the Geosciences

General Synopsis of Review Comments

Sharon Tahirkheli

Overview

The following summary of comments is compiled from those sent by members of the Committee. I have tried to include as many of your comments as possible without being repetitive. Some questions raised by the review of the comments are summarized at the end.

Software

Most encountered some difficulties working with the software. Consideration of availability of standard Thesaurus software was suggested. Current software does not make an overview very easy and makes checking of correct assignment of category and verification of appropriate translation of primary terms difficult. Some issues: general selection or searching not very flexible or intuitive (examples of searches – exclusion of terms while searching, searching for multiple descriptors); documentation of updates, deletions and new entries not clear. Manual is primarily for updating and adding terms primarily.

Style

Hierarchies appear in the Multilingual even though these were not evident in the original. Non-discipline hierarchies have been introduced for the systematic lists in the original. Spot-checking the hierarchies show that they do not always match the hierarchies in the Multilingual (eg. Igneous rocks). Some hierarchies appear to be confused – repeating within the list of their own narrower terms. Other descriptors from the original Multilingual do not show up in any hierarchy.

Style variations exist between different national thesauri. For example: the English thesaurus follows the ANSI/NISO standard for the use of plural and singular. Other national thesauri use singular only. This preference was left to the national thesaurus builder. ANSI/NISO Z39.19 and ISO 2788 and 5964.

Diacriticals are not properly represented in many languages. Other character sets are absent – Cyrillic, etc.

Coverage

More terms would be needed – very specific terms – to cover completely the entire geoscience terminology. The current Multilingual includes less than 10,000 terms. The complete geoscience terminology must be at least 40,000 terms. For example: many Glossary of Geology terms are not included.

Wider coverage of related disciplines needs to be added and the categories made less geo-specific. New areas of research should be included and current 'hot topics' should be given more emphasis, such as the environment and global change areas.

The categories are based on disciplines while users are looking for objects or features. More emphasis should be given to these types of descriptors.

Some Multilingual terms are obsolete and should be weeded. A current frequency analysis should be conducted.

Content

Some terms are mis-translated between languages. In some languages nouns are used while verbs or adjectives are used in others. Again, the use of national thesauri for various internal purposes, makes this type of variation common.

Some terms exist in current Multilingual that did not exist in the last version of the Multilingual distributed in 1999. These have not had their terms agreed upon by the various national groups.

Terms 5831 and above have a large number of category mis-assignments and some lack categories. In addition, some national thesauri have many more terms than are represented in the Multilingual.

Questions raised by Review:

If we want to use the MULTILINGUAL as the starting point, we will need to make some changes to the existing Thesaurus.

General

Software -- the current software gave some a difficult time for the process of review. There are limitations on searching (MS Access). Will these issues impact future use for updating and viewing? Can we have a more user-friendly interface?

Variation among national thesauri – not all thesauri follow the same standard for form and style of term. How do we handle the variations and what is the impact on the structure?

MULTILINGUAL – changes needed to original

Hierarchies that exist within the current Multilingual don't match the original and may not be the ones that we want. Review by subject experts needed?

Missing hierarchies need to be added to the Multilingual. Consideration needs to be given to the discipline versus 'object' needs of the user.

Translations from language to language show wide variability even for higher level descriptors. This is a reflection of the need for the original Thesaurus to serve a national purpose when it was constructed. Should these be made consistent?

Missing translations in various languages need to be remedied.

The extent of coverage desirable for the Thesaurus varies. How big is big enough?

Appendix 2

abbr.	old categories	proposed new categories	propos. new abbr.	Remarks	including
APPL	Applied geophysics			to Methods	
CHEE	Chemical elements	Chemical elements and isotopes	CHEI		Isotope chemistry
CHES	Inorganic chemical components			to Chemical compounds	
COMS	commodities and mineral deposits	Commodities, natural resources	COMR		solar energy, windpower
ECON	Economic geology			to Scientific disciplines	
ENGI	Engineering geology			to Methods	
ENVI	Environmental geology	Environmental concerns	ENVC		pollution, protection, conservation
EXTR	Extraterrestrial geology			to Scientific disciplines	
EXTS	Planets and meteorites-systematics	Extraterrestrial bodies	EXTB	(other materials to rock)	
GEOC	Geochemistry			to Scientific disciplines	
GEOH	Hydrology			to Scientific disciplines	
GEOL	General Geology			to Scientific disciplines	
IGMS	Metamorphic rocks			to Rocks and sediments	
IGNE	Petrology			to Scientific disciplines	
IGNS	Igneous rocks			to Rocks and sediments	
INST	Instruments and equipment	Instruments and equipment	INEQ		tools
ISOT	Isotope chemistry			to chemical elements	
MARI	Marine Geology			to Scientific disciplines	
MATH	Mathematical geology			to Methods	
METH	Methods	Methods	METO		Research and test methods (field and laboratory) Applied geophysics Engineering geology Mathematical geology
MINE	Mineralogy			to Scientific disciplines	
MINS	Minerals	Minerals	MINR		All mineral names
MISC	Miscellaneous	Miscellaneous	MISS		
ORGS	Organic compounds			to Chemical compounds	
PALE	Paleontology			to Scientific disciplines	
PALS	Paleontology systematics	Paleontology systematics	PALT		taxonomy
PHCH	Physical and chemical properties	Physical and chemical properties	PCPR		data concerning engineering concerns
SEDI	Sedimentology and sedimentary petrology			to Scientific disciplines	
SEDS	Sedimentary rocks			to Rocks and sediments	
SOLI	Solid earth geophysics			to Scientific disciplines	
STRA	Stratigraphy			to Scientific disciplines	
STRS	Stratigraphy systematics	Stratigraphy systematics and facies names	STFA		
STRU	Structural geology	Tectonic structures	STRT		Structural geology and tectonics
SURF	Geomorphology and Quaternary geology	Geomorphological features	SUGM		
SUSS	Soils systematics	Soil systematics	SOSS		pedology
TEST	Textures and structures	Textures and structures	TEXS		small scale features of rocks
		Chemical compounds	CHCP		Inorganic chemical components Organic compounds
		Sedimentary structures	SEST		
		Natural processes and phenomena	NAPR		global change, volcanic eruptions,
		Man made features	MMAD		highways, harbours, buildings, etc.

abbr.	old categories	proposed new categories	propos. new abbr.	Remarks	including
		Survey results	SURV		Result of field surveys; photographs; field reports
		Scientific disciplines	SCDC		Economic geology Engineering geology Extraterrestrial geology Geochemistry Hydrology General Geology Marine Geology Mineralogy Paleontology Petrology Sedimentology and sedimentary petrology Solid earth geophysics Stratigraphy
		IT concerns	ITCC		incl. geostatistics, numerical simulation, IT, databases
		Rocks and sediments	ROSD		Igneous rocks Metamorphic rocks Sedimentary rocks