

SEMINAR ON THESAURUS (1975). Paper CF

EXPLORATION AND PRODUCTION THESAURUS: ITS HELPFULNESS IN INDEXING IN
A LIBRARY AND INFORMATION SYSTEM

D K GUPTA and G N TRIPATHI

Oil and Natural Gas Commission, Institute of Petroleum Exploration Library, Dehra Dun

Exploration and Production (E & P) thesaurus published by the Information Service Department, University of Tulsa, is one of the most comprehensive thesaurus in its coverage on the exploration, development and production of crude oil and natural gas. The thesaurus is in continuous use in the Library of the Oil & Natural Gas Commission for indexing the literature and in the development of special schemes of classification for local use. The paper indicates the usefulness of the thesaurus for indexing the entries covered in the various bibliographies and for the preparation of a special scheme of classification for Exploration Geophysics.

1 INTRODUCTION

The Exploration and Production (E & P) thesaurus covers almost all the aspects/subject areas related to petroleum exploration and production: geology; geochemistry; geophysics; drilling; well logging; well completion and servicing; production of oil and gas; reservoir engineering and recovery techniques; pipelining; transportation and storage; supplemental technology (computer science and technology, economics, corrosion and instrumentation, etc). This is one of the comprehensive thesauri available for the petroleum industry. The first edition of this thesaurus was published in 1965, which was updated later in 1967 and 1968. The latest edition is due in 1975. We have studied the 1968 edition of the thesaurus. The thesaurus is supplemented with a good outline of hierarchical classification, fitted with an alphabetical index. The thesaurus has been compiled by the experts group of the Information Service Department of the University of Tulsa, which is a cooperative venture consisting of about 40 leading oil companies of the United States and other countries. This department has been generating a variety of information services, Petroleum Abstracts being the main for its subscribers and contributor-cooperative members. The Department consists of various expert sections: Administration (4 persons); Word Processing (5 persons); Exploration (4 persons: 2 geologists, 1 Geophysicist, and 1 Exploration editor); Engineering (5 persons: 1 Managing editor, remaining Petroleum Engineers); Software group (3 persons). The meetings of the Subscribers' Advi-

sory Committee are held occasionally to discuss the various aspects to improve upon the services,

E & P thesaurus is a special scheme of words and phrases, which are the concepts related to the petroleum exploration, development and production. To be precise, it is a controlled vocabulary for the subjects mentioned above, which is purposely constructed with the generic or hierarchical relationships. The relationship is established either on the basis of generic terms or associated ideas. All the terms of like meanings or concepts have been directed to one term as a descriptor. The selection is based mainly on the basis of concept and popular usage.

Example EPEIROGENESIS
USE EPEIROGENY
EPEIROGENIC MOVEMENT
USE EPEIROGENY

The authors and users may use the words in their natural language as "Epeirogenesis" or "Eperogenic movement", but the words have been controlled by the descriptor "EPEIROGENY" which is a correct word and used popularly. Under this descriptor is shown

EPEIROGENY
UF EPEIROGENESIS
UF EPEIROGENIC MOVEMENT
BT TECTONICS
SA EPEIROGENIC UPLIFT

Here Epeirogenic uplift is the geologic structure a product of epeirogeny - process.

2 STRUCTURE OF AN ENTRY OF
E & P THESAURUS

The structure of an entry for a term in E & P thesaurus, which displays a variety of relationships, is as follows:

- A SEISMIC SPREAD
- B **FOR GEOPHONE CONFIGURATIONS
FOR SHOTPOINT CONFIGURATIONS
See (Array)
- C UF GEOPHONE PATTERN
UF GEOPHONE SPREAD
UF LONG SPREAD
- D NT CROSS SPREAD
NT HORIZONTAL SPREAD
NT IN LINE SPREAD
NT INTERFERENCE SPREAD
NT OFF LINE SPREAD
NT SPLIT SPREAD
NT SUBSURFACE SPREAD
NT SURFACE SPREAD
NT VERTICAL SPREAD
- E BT PATTERN
SA PATTERN SHOOTING
SA SEISMIC EXPLORATION

SEISMIC TRACE DIFFERENT
- G *SEISMIC TRACE DIFFERENTIATION
- H USE SEISMIC INTERPRETATION

21 Explanation

- (a) Descriptor/Term
- (b) "Scope Note" providing definition, explanation, limitation of usage ; seismic spread is explained and also its usage for geophone and shot point configurations/arrays.
- (c) "Used for" reference. It indicates that the terms 'Geophone Spread', 'Geophone Pattern' and 'Long Spread' are not used as descriptors. Entries under these words have been directed to use "SEISMIC SPREAD".
- (d) "Narrower Term" reference indicates a subordinate/part or narrower hierarchical relationship to the descriptor. Here, these are the various seismic spread techniques.
- (e) "Broader Term" reference indicates the next higher term in the hierarchy or the generic term (Father-son relationship). Here "Pattern" is a very general term which is used in many contexts, and one of them is "Seismic Spread".

(f) "See also" reference indicates all possible synonymous terms/related terms or an alternate concept. Users of the thesaurus are directed to see the other terms depicting the similar concept.

(g) Abbreviated descriptor spelled out/expanded.

(h) "Use" reference indicates a preferred or a popular usage term/a descriptor.

3 THESAURUS UPDATING

The thesaurus is updated through the supplements which are issued at regular intervals of four months with all the new descriptors to be added to the thesaurus. The new descriptors are added on the basis of articles in periodicals and other documents which are included in Petroleum Abstracts. Thus, the thesaurus at the end of the users is kept always uptodate and provides good means for indexing.

4 HIERARCHICAL STRUCTURE

One of the two additional sections added to the main body of the thesaurus is the section of hierarchy. This provides a compact list of the descriptors in their structured form. The thesaurus is constructed on the concept or faceted basis. The ten principle concepts/facet headings are given below in the alphabetical sequence.

COMMON ATTRIBUTE	OPERATING CONDI-
EARTH & SPACE	TION
CONCEPTS	ORGANISM
ECONOMIC FACTOR	PHENOMENON
EQUIPMENT +	PROCESS
MATERIAL	PROPERTY

Example

- PHENOMENON
- ... WAVE
- MECHANICAL WAVE
- ELASTIC WAVE
- SEISMIC WAVE
- SURFACE WAVE
- LONG WAVE
- RAYLEIGH WAVE

An alphabetical index to the hierarchy is provided as an additional section of the thesaurus.

5 ANALYSIS AND EVALUATION

The E & P thesaurus is continuously in

use in the ONGC/IPE Library for indexing of documents. This specific purpose thesaurus is the only one in the field of petroleum exploration and production covering almost all the main subjects and related subjects very comprehensively. The succeeding paragraphs present some comments on the practical use of this thesaurus in indexing the documents.

51 Coverage

The E & P thesaurus (3rd edition) contains as many as 5,348 words. The details are given below:

1)	Total number of words used in the E & P thesaurus	5, 348
2)	Total number of word-groups which are the synonyms to the descriptors	3, 520
3)	Actual descriptors	1, 828
4)	Synonymy ratio	2. 92

These figures indicate the versatile structure of the E & P thesaurus. This thesaurus has been found to be helpful to authors in the choice of technical terms while writing papers and to searchers of information in formulating their queries to the information system. Besides this, this thesaurus functions as a good authority list for control of vocabulary.

52 Intellectual Structure

The thesaurus has been constructed by the experts in the field of petroleum exploration and production. Many of these experts are subject specialists working in the Information Department who are controlling the storage, retrieval of information and providing various information services on all related and pertinent subjects connected with petroleum exploration and production. The same expert group has been engaged in controlling the language, establishing the hierarchy and promotion, deletions and additions of words,

The arrangement of the main body of the thesaurus is alphabetical but not strictly based on dictionary arrangement. Single word and double word descriptors appear before the plurals and compound words, e. g. REFRACTIVE WAVE appears earlier than REFRACTION (SEISMIC), but such cases are not many.

53 Descriptors Section

1 Descriptors are mostly selected after checking API subject authority list to maintain the compatibility.

2 The descriptors are retained in combination with the other words which are commonly used and are helpful in explaining the concept such as: GAS COMPRESSIBILITY. COMPRESSIBILITY DRILL COLLAR. COLLAR. The words 'collar' and 'Compressibility' are used as separate words as well as in combination with DRILL-collar and GAS-compressibility for their individual meanings and usage.

3 As far as possible, nouns are used instead of adjectives, i.e.

ELECTRICAL
USE ELECTRICITY

4 In the case of synonyms, one which is supposed to be popular and standard is retained and others are referred to as:

EARTH MAGNETISM
USE GEOMAGNETISM

5 Inversions are very few and given only where they could not be avoided to retain the concept. Wherever inversions are used, they are put within parenthesis. The descriptor or uninverted term is referred to, e. g.

CYCLIC (GAS)
USE GAS CYCLING
GRADIENT (MAGNETIC)
USE MAGNETIC GRADIENT

Here, the 'cycling' and 'gradient' are the terms which are used by the authors and users very frequently and for maintaining the vocabulary control, indexer uses only the standard terms but for convenience and guidance of users, USE reference is made.

b Most of the descriptors are given in the singular form.

7 Words are bound to form concepts, but a concept is not bound to another concept, e. g.

WELL GAS LIFT TEST
USE GAS LIFT (WELL)
PLUS WELL TESTING

SEISMIC CASE HISTORY
USE SEISMIC EXPLORATION
PLUS CASE HISTORY

Such examples are many in the thesaurus where the correct usage is directed with a complete concept of the term in natural language use.

6 FACETED STRUCTURE

Hierarchies have been established on the basis of various facets/categories of concepts.

61 Common Attributes

Hierarchy 1H - 4H are described under this category. The terms/phrases described here are of most general nature; generic terms, subject categories are given under this heading. No principle is used in the choice of these terms or to arrange them in a particular sequence. Direction, industry, location, physical form, quality, quantity, shape, size and technology, the main sub-headings, are supposed to be of common nature to describe the various concepts. Under technology, the subject categories are described in the following manner:

Acoustics	Dynamics	History
Archaeology	Ecology	Hydraulics
Biology	Electronics	Kinetics
Botany	Engineering	Metallurgy
Chemistry	Geodesy	Meteorology
Cosmology	Geography	Morphology
Cryopedology	Geophysics	Oceanography
Crystallography		Optics
		Palynology
		Photogrammetry
		Physics
		Rheology
		Taxonomy
		Thermodynamics
		Zoology

These subject categories are further subdivided into sub-categories in a hierarchy.

62 Space and Earth Concepts

Hierarchy 4H - 11H. Under this heading, all the general concepts with appropriate hierarchical order related to space and earth and its environment, structures, etc are described. But no geographical schedule for cities/towns or countries is provided, as it is out of scope of the thesaurus.

63 Economic Factors

Hierarchy 4H - 11H. All the concepts related to cost, legal matters and economic considerations involved in the exploration and production of petroleum are categorized under this heading.

64 Equipment

A hierarchic schedule (13H - 25H1

of the equipment is given with all possible hierarchic relations following the Whole-Organ principle .

64 Material Concept

Material Concept {Hierarchy 25H - 37H} is scheduled by composition, by function, by phase, by information. The schedule is well structured and adopts the Whole-Organ principle in the arrangement of material concepts,

66 Operating Condition

Operating concepts (Hierarchy 37H - 40H) with their appropriate hierarchical structures such as Acceleration, Accuracy, Air-Fuel ratio, Adiabatic, Application, Blow out, Delay, Drainage, Drilling problems, Failures, Damage, Frequency, Overlap, Mobility, Parameter, Pattern program, Rate, Result, Safety, have been listed,

67 Organism

A complete hierarchical list of organisms (40H - 42H) is given under this category.

68 Phenomenon (4ZH - 49H)

Phenomenon concepts (Hierarchy 42H - 49H) are result/process oriented; products of certain actions ,

691 Process

The hierarchy (49H - 62H) of process such as business operations, chemical processes, mathematical calculations, physical operations, testing, etc are the words/phrases denoting the concepts of actions processes involved in operations.

692 Property (62H - 65H)

The hierarchy of property (62H - 65H) consists of composition of matter, physical properties and structures of the matter, etc.

693 Annotations

The hierarchical list in its structure with various concepts, resembles concepts, such as Basic Subjects, Common Isolates, Personality, Matter, Energy and Space Isolates. The terms under these categories are listed in alphabetical sequence.

7 USEFULNESS OF THESAURUS

The E & P thesaurus is found to be a very useful tool for vocabulary control in indexing the petroleum literature and developing the depth classification schedules for technical processing of documents. The thesaurus, as it is evident from the synonym ratio 2.92, has a good vocabulary control; its arrangement and search strategy is also well designed to avoid many inconveniences and provide quick answer. The thesaurus, as it is expected, improves recall value in information retrieval. It optimises the Recall-Precision ratio. Any term given in the thesaurus directs, as NT "Narrow Term", BT "Broader Term", SA "See also", etc. to the other terms which are related to the main term in the hierarchy or to the concept pertinent only to the exploration and production of petroleum. The following example would illustrate this:

Example:

```

MAGNETIC FIELD
  NT GEOMAGNETIC FIELD
  BT MAGNETISM
  SA HALL EFFECT
  SA INDUCED MAGNETISA-
    TION
  SA MAGNETIC FORCE
  SA MAGNETIC INDUCTION
  SA MAGNETIC PERMEA-
    BILITY
  SA MAGNETIC PROPERTY
  SA MAGNETIC SUSCEPTI-
    BILITY
  SA POTENTIAL FIELD

```

"Magnetic field" is described as the concept with the basic concept "MAGNETISM" and precise concept in context to Petroleum E & P - "Geomagnetism" and closely related concepts only in the same context are given with 'SA' direction.

71 Use of Thesaurus for Indexing

Thesaurus is used in the library as a good source of guidance for technical processing. It is used for clarification in identifying connotation of terms while classifying the documents in the ONGC/IPE Library which uses UDC.

For the purpose of indexing, E& P thesaurus is a very helpful tool as it provides a good vocabulary and hierarchical control; modulation of links in a chain upto the fifth

order or even higher hierarchy is possible with the help of this thesaurus. The following example is quite illustrative:

```

ELECTRICITY
  PHYSICAL PROPERTY
    SURFACE PROPERTY
      ELECTRIC POTENTIAL
        ZETA POTENTIAL
          SELF POTENTIAL
            STATIC SPONTANEOUS
              POTENTIAL

```

"STATIC SPONTANEOUS POTENTIAL" is the micro concept of "ELECTRICITY". The concept is described here in a perfect hierarchy providing facility for chain indexing and permuted indexing upto the fifth order in depth. Here, "Static SP" is "self-potential" which is one kind of "Electric potential" and which is "Surface Property" in the "Physical Property" of •Electricity".

Each term of the above example is separately described in the thesaurus with appropriate directions and well delineated in the hierarchy schedule under 44H.

The thesaurus is extensively used while preparing the indexes to the bibliographies compiled by the ONGC/IPE library. It has been a very resourceful subject authority list for the indexes used in the ONGC/IPE Library.

72 Use of Thesaurus for Developing Schedules

The thesaurus is one of the good sources in providing relevant standard set of descriptors/terms. E & P thesaurus, which is in the alphabetic sequence, with the hierarchy schedule in faceted form and index to it, covering a variety of subjects related to petroleum exploration and production, is a good source in developing the special depth classification schedules for subjects in the field of Petroleum Exploration and Production. Construction of such schedules can be done easily by appropriate choice of terms from the thesaurus. NT, BT, and SA and other words expressing relation will provide assistance in the selection of other related words. These selected words, as they are described under a particular concept in the hierarchy schedule, will require only the re-examining of those concepts to identify the manifestation of the category and to determine the helpful sequence of the concepts. A depth schedule for a freely faceted model of Ranganathan can be easily developed

at least upto 70% perfection. A depth classification schedule developed at DRTC for "Exploration Geophysics" by D K Gupta has been examined and compared. The thesaurus has been found quite comprehensive in its coverage. 60% of all the words/terms described in Gupta's scheme are found in the thesaurus ; the arrangement of categories is also quite similar,

Thus, it is found that the thesaurus would be found quite useful in developing other schedules also. A project of this nature has been initiated in ONGC/IPE Library.

73 Use of Thesaurus in Mechanization

No attempts have been made so far to use the thesaurus for mechanized indexing work. But it is found that the thesaurus with all its merit can be used for developing computer-based information storage and retrieval system. KWIC/KWOC can be conveniently developed/obtained with the help of computers, with vocabulary control of this thesaurus.

8 CONCLUSION

E & P thesaurus is really a good medium for establishing a dialogue between authors (generators of ideas) and users of the knowledge. It is no doubt a unique tool in the hands of the Librarians/documentalists/information scientists engaged in the petroleum industry,

It covers almost all the subject areas of petroleum exploration, development, and production. This faceted scheme provides good hierarchical concepts, helpful in chain indexing, permuted indexing and in developing special schemes for depth classification. As there are no useful schedules available with CC, UDC and DC for the depth classification and cataloguing of the petroleum literature, this thesaurus provides the basis and ready-made assistance for minute classification and facility of extension of the existing UDC schedules. On account of its high synonym ratio,

it provides a good recall-precision ratio in information retrieval.

91 PROPOSITION

Thus, from the foregoing discussions on E & P thesaurus, the following propositions can be suggested:

1 A thesaurus should provide a good hierarchical schedule of concepts.

2 A faceted thesaurus is useful in developing depth schedules.

3 It is desirable that the designers of thesaurus should also provide a scheme of classification, based on sound principles.

92 BIBLIOGRAPHICAL REFERENCES

1 AITCHISON (J). Thesauri-facet: A multipurpose retrieval language tool (J Doc. 26; 1970; 188-203)

BLAGDEN (J). Structured thesauri, (Aslib Proc. 23 ; 1971 ; 139-43)

JONES (K). Basic structures for thesauri systems. (Aslib Proc. 23 ; 1971; 577-90)

MANDERLOOT (W G B), DOUGLAS (E M B), and SPICER (N). Thesaurus control - the selection, grouping and cross-referencing of terms for inclusion in a coordinate index list (J Amer Soc Inf Sc. 21 ; 1970:49-57)

ROLLING (L N). Compilation of thesauri for use in computer systems, (Inf Storage Retr. 6; 1970; 341-50).

THOMAS (P A). Some problems in compiling a thesaurus of transportation and planning terms. (Aslib Proc. 23; 1971 ; 595-606).