

EQUIVALENCE RELATIONS IN INFORMATION RETRIEVAL THESAURUS

M. A. Gopinath, *Documentation Research and Training Centre,
Indian Statistical Institute, 31 Church Street, Bangalore 560 001*

Differentiates between linguistic equivalence of terms from that of information retrieval situations. Presents a set of seven varieties of equivalence relations on the basis of the "used for" relations found in textile and energy thesauri. Some criteria for preferred terms from a set of equivalent terms are listed.

1 Introduction

An information retrieval thesaurus is a helpful aid in indexing and searching information retrieval system. It improves efficiency by effecting optimum recall and precision to searchers' queries. Such a thesaurus arranges a set of entries in an alphabetical sequence. It is enriched by a systematic presentation of the terms in a categorised sequence.

2 Entries in a Thesaurus

Each entry in a information retrieval thesaurus generally consists of terms in the following kinds of relations: (a) Equivalence; (b) Hierarchical; (c) Associative. Here is an example:

	NEUTRON FLUX
EQUIVALENCE	UF Flux (Neutron) UF Neutron economy UF Neutron flux density
HIERARCHICAL	BT Radiation flux NT Adjoint flux
ASSOCIATIVE	RT Flux synthesis RT Hetrogeneous effects

RT Neutron age
RT Neutron fluence
RT Neutrons

Studies on hierarchical relations and associative relations have been discussed in 1975 DRTC Annual Seminar (5). In this paper, we focus our study on Equivalence Relations.

3 Equivalence Relations

Equivalence relation is generally an aid for an information searcher when he approaches the system with a term which is likely to have synonyms, antonyms and homonyms. There are a few linguistic equivalent terms for an idea in a natural language. But in an information retrieval language, the cluster of equivalence is a wide spectrum of terms. In the linguistic thesaurus terms denoting one and the same idea but with slight linguistic variance are considered to be synonyms. But in an information retrieval thesaurus terms denoting ideas which are likely to be considered by a user as equivalent to the ideas which are likely to be considered by a user as equivalent to the idea of his search are also included. This is a kind of threshold range of ideas. Therefore, the choice of equivalent terms in such a thesaurus calls for guidelines and relational matrix.

4 Guidelines for Relations

There are several guidelines for choice of equivalent terms for inclusion in an information retrieval thesaurus (1-6). However, they generally confine to the conventional synonym and antonym relations. In order to find out the actual choice of equivalent terms, we have examined two thesauri, namely

- 1) The textile technology thesaurus (8)
- 2) ERDA subject indexing and retrieval thesaurus (7)

The examination of the "Used for (UF)" relations in these two thesauri have helped us to develop a typology of equivalence relations. This is presented in the succeeding sections of the paper.

EQUIVALENCE RELATIONS

5 Typology of Equivalence Relations

The following table presents seven types of equivalence relations. These have been found to be incident in the two thesauri studied. The table presents definition and examples from two thesauri for each type of equivalence relations. The threshold range of equivalence has, in its ambit,

- (a) terms that represent viewpoints of the same property continuum;
- (b) terms having significant overlap; and
- (c) specific concepts subsumed under broader terms, for example, terms which connote highly specific ideas for the needs of retrieval language.

Sl. No.	Type of Relations	Example from	
		Textile Technology	Energy
1	SYNONYMS are separate words whose meaning—both denotation and connotation—is identified so that one can always be substituted for the other without change in the effect of the sentence.	1. Acoustics	Air cleaning
		UF Sound 2. Combustion UF Burning	Air purification Semiconductor detector UF Semiconductor counters.
2	NEAR SYNONYMS are separate words whose meanings lie in contiguity in a threshold range of associations—so that one denotation reminds the incidence of other terms.	1. Air conditioning UF Aircooling	Aviation fuels UF aircraft fuels
3	ANTONYMS are separate words whose meanings imply direct opposite or other extreme. It is a contrary meaning.	1. Corrosion UF Non-corrosion	Storage life UF Market life.

Sl. No.	Type of Relations	Example from	
		Textile Technology	Energy
4	HOMONYMS (POLYSEME) are same form of words with different meanings.	1. Concentration (composition) UF Molarity 2. Apron (Drafting) UF Aprons (Drafting)	1. Switches UF Contacters 2. Synchrotrons UF Phasotrons
5	SEMEMES (Semantic relation) A kind of specific terms which substitutes in an apparent manner the denotations of generic terms	1. Accounting UF Cost accounting 2. Humidity UF Absolute	1. Switches UF Electric switches 2. Stimulation UF Growth stimulation
6	MORPHEMES (Synactical Relations) are terms which have equivalent effects due to close associations of the ideas and their frequency of co-occurrence in a sentence structure such as the following:		
	a. Cause and effect	Communication UF Liaison	Polymerisation UF Radiation Hardening
	b. Thing and property	Hand UF Clamminess	Specific heat UF Heat capacity
	c. Thing and its function	Antifoaming agents UF Foaming	PH value UF Acidity
	d. Process and product	Acetylation UF Acetylated collar	Atmospheric precipitation UF Fog
	e. Process and agent	Cushioning UF Cushion	Monitoring UF Monitoring Network

Sl. No.	Type of Relations	Example from	
		Textile Technology	Energy
f.	Process and device	Gamma rays UF Radiation Grafting	Subcritical Assembly UF Neutron multiple facility
g.	Process and type of processes	Degradation UF Oxidative degradation	Pauli spin operators UF Pauli matrices
h.	Process and steps in process	Bleaching UF After bleaching - colouring - solving	Metabolism UF Enzyme activity
7	WORD FORM RELATION		
1	Reversal	Analysis of variance UF Variance analysis	Electric fuses UF Fuses (Electric) Cooling Pond UF Pond (Cooling)
2	Verb/Noun Forms	Applying UF Application	Biochemistry UF Biochemical action
3	Multiworded/ single worded and vice-versa	Deformation UF Fabric defor- mation	Geochronology UF Age Estimation
4	Short forms	Cystalline structure UF Crystallinity Automatic doffers UF Auto doffers	Electric potential UF Voltage Rodents UF Kangaroo rats Larmer Precision UF Larmer Nuclear precision
5	Abbreviations	Carboxymethyl cellulose UF CMC	RNA UF Ribonucleic acid Liquid natural gas UF LNG
6	Spelling variations	Labour UF Labor	Colour UF Colour

6 Preferred Terms

Among the set of equivalent terms, one term is chosen for lead position and it is called preferred term. It is cross-referred by "User For" to non-preferred terms. The choice of preferred term is guided by the following criteria.

1. Current term rather than the superseded terms
2. Scientific terms when more familiar to users
3. Local term for a local system
4. Form most understood by the users of the system
5. The purpose of the preferred term concept is to satisfy the law "save the time of the reader" and its corollary 'save the time of the staff'.

7 Development of Equivalent Relations

The equivalence relations in an information retrieval thesaurus context is steadily developing one. The pan-disciplinary development of subjects generate variety of equivalent terms. This is evident from energy thesaurus where biological terms occur in fairly significant quantity. Drawing of relational typologies from time to time and assessing their utility in information retrieval contexts would improve the vocabulary control devices.

8 Bibliographical References

1. Aitchison (J) and Gilchrist (A). *Thesaurus Construction: A practical manual*. 1972.
2. ISC: 2788 *Guidelines for the establishment and development of monolingual thesauri*. 1974.
3. KENT (Allen). *Information analysis and retrieval*. 1971.
4. LANCASTER (FW). *Vocabulary control for information retrieval*. 1972.
5. NEELAMEGHAN (A). *Non-hierarchical associative relationships: Their types and computer generation of RT links*. (DRTC Annual Seminar 1975. Paper AA).
6. SOERGEL (D). *Indexing languages and thesauri. Construction and maintenance*. 1974.
7. UNESCO *Guidelines for the establishment and development of monolingual thesauri*. 1973.

Thesauri Referred for Example.

8. TECHNICAL INFORMATION (USERDA). ERDA Subject Indexing and Retrieval Thesaurus. 1975. NTIS. 1975.
INSTITUTE OF TEXTILE TECHNOLOGY (USA). Textile Technology: An information retrieval thesaurus. 1966.