International Journal of Library and Information Studies
Vol. 5(3) Jul-Sep, 2015 *ISSN: 2231-4911*

Designing a Thesaurus in Nanotechnology: Issues and Concerns

Abhishek Sharma
Scientist IV (I)
CSIR-NPL, Dr. K.S. Krishnan Road
New Delhi – 110012
e-mail- sharmaa@nplindia.org

Abstract

Linguistic ambiguity and linguistic variations are two major problems associated with modern information retrieval systems. Handling the vocabulary problem by using a thesaurus is an ideal solution. This paper deals with the overview of methodologies suggested for the development of a thesaurus. The work further describes the approaches adopted for designing and development of a thesaurus in Nanotechnology.

Keywords: Controlled vocabulary, information retrieval system, Nanotechnology, thesaurus

1. Introduction

The term "information" is considered as the currency of the information world. In the present scholarly society, it is the information which allows one to have edge over others and significantly permits us to successfully adjust with the external existence conditions. In national development, information performs a similar role as defence forces perform for national defence and power. The ammunition for information is various information carriers in the form of books, periodicals, etc. and people with distinct requisite skills to organize and disseminate information (Spade, 2012). Thus, considering information as power, Lesko (n.d.) stated that "those who control information are the most powerful people on the planet". Consequently, timely delivery of desired information is essential. Depending upon the development of technologies and so their requirements give rise to various information organization, retrieval tools and methods like cataloguing, classification, indexing, controlled vocabularies, search engines etc. Among the existing information organization and retrieval tools, controlled vocabularies (CVs) stand out as prevailing tools. They perform an important role in information organization and retrieval activities. CVs are often essential for cataloguers in making use of consistent terms to represent the same concept, when more than one intellectual is performing on the same index. In the retrieval operations, CVs guide users to select correct terms for their searches (Harping, 2010). Further, role of CVs extended to the Web world where search engines are considered as popular information retrieval tools. Hedden (2008) mentioned that search results obtained with search engines can be improved by using controlled vocabularies for assigning keywords for documents to be indexed on the Web.

2. Thesaurus as Controlled Vocabulary Tool

Among the existing controlled vocabulary tools thesaurus is considered as prominent aid in knowledge organization and retrieval system. The view has been unanimously accepted by number of scholars (Gopinath, 1985; Kumbhar, 2005; Pinto, 2008). Various minds have discussed the concept 'thesaurus' with distinct perspective. Aitchison and Clarke (2004, p.6) cited Shorter Oxford Dictionary to state that the word 'thesaurus' first appeared in English in 1736. In the perspective of information retrieval, thesaurus was first used in the year 1957

(Vickery, 1960, p.181). Among the existing controlled vocabulary tools, thesaurus emerged as an aid to indexers that adds to their vocabulary (Roberts 1984). Bernier (1968) describes thesaurus as an external guide which directs the user from known concepts to those required. Further, expanding the scope of thesaurus from users' point of view, Chen et al (1993) viewed thesaurus as a concept space, which represents various domain concepts along with their semantic relations and thus significantly enhance the understanding for a domain. Hence, for effective information processing and retrieval, thesaurus serves as helping tool for both practitioners and information seekers.

3. Approaches for Thesaurus Construction

Thesaurus is a controlled vocabulary tool that plays a vital role in indexing and retrieval operations of any knowledge organization system (Aitchison, Gilchrist & Bawden, 2000). To develop a knowledge organization system there are two fundamental approaches (Anderson & Perez-Carballo, 2001):

- Manual intellectual analysis
- Computational/ automatic algorithmic analysis

Further, for construction of controlled vocabulary Lancaster (1972) suggested following steps:

Table1
Comparison between Manual and Automated Approach for Thesaurus Construction

Step	Explanation	Manual System	Automated System
Identification of subject area to be covered	Defining the subject field, establishing the subject boundaries in which a thesaurus has to be constructed	Approach where intellectual efforts are required	Approach where intellectual efforts are required
Selection of suitable terms to describe the area	Collection of adequate number of domain concepts to be incorporated in thesaurus	Manual selection	Computational selection and preparation of ranked list by using statistical techniques
Making decision about the exact form of terms to appear in the vocabulary	Deciding acceptable grammatical forms, singular & plural forms, variant spellings, abbreviations, acronyms	Thesaurus construction and development standards are referred	Thesaurus construction and development standards are referred
Organization of terms in useful pattern	Establishing meaningful relationships between the domain concepts	Manual approach where intellectual efforts are required	Based on statistical and syntactical ground
Display of terms in helpful manner	Arrangement of terms in alphabetical order	Manual technique such as Card system are employed	Performed with the help of computer

Thus, if we analyse both the approaches individually, then it could be inferred that on semantic ground, manual procedures are preferred in contrast to automatic methods Chen & Thiel (2004), whereas pace in thesaurus development, flexibility in deriving domain concepts from author's text, and ease in editing and maintenance of thesaurus could be achieved with automatic approach Surace (1970).

Though, both the approaches for thesaurus construction have their pros and cons, therefore as suggested in number of scholarly works (Devadason, n.d.; Raizada, Ramachandran, Satyanarayana, & Pal, 1975; Surace, 1970), the present work of designing a thesaurus on Nanotechnology focused on semi-automatic approach where besides intellectual efforts computer application was utilized in the following areas of thesaurus construction:

- Gathering of indexing terms on the basis of frequency statistics.
- Sorting and counting to produce co-occurrence tables.
- Making a consistent alphabetical display.
- Representing thesaurus in desired format.
- Addition, deletion and alteration of terms at any point of time.
- Offer non-linear searching when printed mode of Nanotechnology thesaurus was transformed into its electronic version with the help of nMap software which was designed especially for the research work of the author.

4. Constructing a Nanotechnology Thesaurus

Before initiating the work on building a new controlled vocabulary tool, it is essential to establish either such tool exists for the domain or not. The thought was supported by Lancaster (1985). To investigate the existence of work, a check in the current and back issues of "Knowledge Organization", a quarterly publication of International Society for Knowledge Organization (ISKO) as was suggested by Aitchison, Gilchrist and Bawden (2000) was undertaken. With the analysis of publication, it was established that work related with the scope of this study was not listed in the publication. Further, coverage of Nanotechnology concepts in various sources like IET INSPEC Thesaurus (2010), Library of Congress Subject Headings (32nd Ed. 2010), DDC 23rd edition, series of vocabulary documents developed by the International Organization for Standardization (ISO), and publicly available Glossary of Nanotechnology Terms developed by The Institute of Nanotechnology were also explored. Analysis of these sources revealed very limited coverage of Nanotechnology aspects. With the aid of publicly available Web resources, it was further explored that the scholarly works available on Internet was confined to Bio medical field and not with Physical Sciences.

4.1 Sources used for the Study

Construction of controlled vocabulary tool involves scholarly as well as clerical operations. Tasks of identifying, describing important domain concepts, defining descriptors, linking descriptors with non-descriptors and arrangement of concepts, fall under the category of intellectual efforts. All intellectual operations require consultation of recorded knowledge structures and discussion with subject specialists. The phase which involves collecting the appropriate concepts that can sufficiently describe the domain is specified as significant step, as it provides a base on which castle of controlled vocabulary will be created. According to Aitchison, Gilchrist and Bawden (2000) task of terms assembling can be performed by consulting terminological sources in standard form, literature scanning, and inviting advices from users and subject experts. For extraction of candidate terms following sources were consulted.

I**. Printed Resources**

i. Encyclopedia
- Nalwa, HS (ed.) 2004, *Encyclopedia of Nanoscience and Nanotechnology*, American Scientific Publishers, California.

ii. Standards Terminologies for Nanotechnology
- ASTM Standard Terminology Relating to Nanotechnology (ASTM E2456-06)
- ISO/TS 80004-1 (Nanotechnologies-vocabulary Part1: core terms)
- ISO/TS 80004-3 (Nanotechnologies-vocabulary Part3: carbon nano objects)
- ISO/TS 80004-4 (Nanotechnologies-vocabulary Part4: nano structured materials)
- ISO/TS 80004-5 (Nanotechnologies-vocabulary Part5: bio/ nano interface)
- ISO/TS 80004-7 (Nanotechnologies-vocabulary Part7: medical, health and personal case applications)

- ISO/TS 27687 (Nanotechnologies-terminology and definitions for nano objects, nanoparticles, nano fibre and nano plate)
- British Standards Institution (BSI) Terminology

iii. Terminological sources in standard form:

- Thesauri **:** IET INSPEC Thesaurus

iv Conference Proceedings and Handbooks

In addition to monographs and journal articles, conference papers and technical reports were scanned.

- AdMet Conference Proceedings *(A joint venture of CSIR-NPL & Metrological Society of India)*

- Handbooks

i. Nalwa, HS (ed.) 2000. *Handbook of nanostructured materials and Nanotechnology*, Academic Press, USA.
ii. Whitehouse, DJ 2003. *Handbook of surface and nanometrology*, Institute of Physics, Bristol.

II. **Electronic Resources**

In addition to various Web sources, three databases namely "INSPEC", "SCOPUS", and "Web of Science" were retrieved:

- **INSPEC**

A product of Institution of Engineering and Technology (IET) that facilitate abstracted and indexed literature created specifically for research in Physics, Electrical Engineering, Electronics, and Computer Science. This database is updated weekly and provides coverage to over 13.6 million records from 1969 to the present.

- **SCOPUS**

Scopus is abstract and citation database containing both peer-reviewed research literature and quality Web sources in the fields of Science, Technology, Medicine, Social Sciences, Arts and Humanities. With over 21000 titles from 5000 publishers, it offers 53 million records to support the research need.

- **Web of Science**

A product of Thomson Reuters that connects publications and researchers through citations and controlled indexing in curated databases spanning every discipline. It includes 2.6 million records and back files dating back to 1898.

In addition to above listed three databases and text available on Internet, following three online resources were consulted for term collection and further, to add scope notes.

- **Glossary of Nanotechnology Terms**

Produced by Institute of Nanotechnology, Scotland to define the concepts which fall under the Nanotechnology domain. It covers around 200 concepts related with various subfields of Nanotechnology.

- **Dictionary of Nanotechnology**

Nanodic.com is an online, searchable and descriptive dictionary about Nanotechnology. This site classifies the terms related to Nanotechnology into 9 different categories: General - Nanomaterial - Nanoelectronic - Nanomedical - Nanobio & Bionano - Carbon nanostructure - Nanocharacterization - Nanofabrication - Molecular Nanotech.

- ***Nanotechnology Now***

Online gateway that provides consultation, technology monitoring, and in-depth analysis, as well as up-to-date news briefs and breaking developments in the Nanosciences. World's leading nanotech experts contribute to this gateway. Nanotechnology Glossary available

on this gateway is a work-in-progress which provides detailed explanation for Nanotechnology concepts.

III. **Human Resources**

Nanotechnology word map is created in consultation with subject specialists. Further, role of human resource is not limited to term selection but as defined by Lancaster (1972) committee approach was utilized during terms refining.

4.2 Data Cleaning and Evaluation of Candidate Terms

Domain concepts acquired from distinct sources were in different format. Like in INSPEC, concepts were separated by '-' whereas in SCOPUS and Web of Science ';' is used as a separator. Therefore, to get the concepts in desired format. i.e. single concept in a single cell of Excel sheet, manual as well as fundamental formulas of MS Excel were applied. This step of data cleaning and evaluation of candidate terms was performed to remove the incomplete word forms and, outside the scope terms.

Incomplete Word forms: This may appear due to the limitations associated with the representation style of databases, and application of MS Excel formulas which were applied to obtain concepts separately.

Eg: **air -** which may be air gap, air bearings etc
micro - which may be micromachines, microscopy, microscopes, micrometer screw drives etc

Out of the Scope Terms: Terms were defined as out of the scope when they did not fall under the scope of this study i.e. Physical Sciences.

Eg: **Au films**
Why out of the scope: as the term represents Gold Films and the present study deals with thin films in general.

Cameras

Why out of the scope: as the term represents an optical instrument that is used to record images but the present study deals with optical devices that have nanoscale resolution.

5. Preferred Term Selection

Preferred term in a thesaurus is the focal point where related information about the concept is placed. As stated by Aitchison and Gilchrist (1987, p.12) preferred term which also is referred to as descriptor or keyword is used consistently during indexing of a concept. In contrast, non-preferred term defines the scope of preferred term. To perform this task for the present work, Hulme's Principle of Literary Warrant was used. According to this principle, frequently appearing concepts in literature may be considered as preferred terms. Here, appearance of a term in literature i.e. usage of the term justifies the inclusion of the term in controlled vocabulary, which is a prerequisite of the principle. Hulme (1950, p.447) described the principle as "the plotting of areas pre-existing in literature". Going further for indexing system, Dabney (2007, p.242) stated that index heading should be created for thoughts that are represented in literary work. On the same line, three criteria for inclusion of words in controlled vocabulary – literary warrant, user warrant and organizational warrant were suggested in American National Standard (ANSI/NISO) (2010).

In contrast to this, views of other scholars reject this quantitative approach. As mentioned by Lancaster (1972), and Ghose and Dhawle (1977) for term selection, when we adopt the

approach of frequency of terms in a collection of documents, then the terms with very high frequency are considered to have little relevance for the subject, as such terms are considered too general to be useful in describing the subject matter. Similarly, those that have been used very infrequently may represent concepts that have little relevance for the subject.
Thus, there is a debate between scholars- one who favour the word frequency as a base for inclusion in controlled vocabulary and the other limits the thought. Therefore, selecting any approach from the two mentioned approaches demands proper justification i.e., warrant for inclusion of term in the work.

Examining the suitability of views of various scholars in the present study reflected that omitting frequently as well as infrequently appearing terms resulted in neglecting core concepts that actually represent the needs of researchers working in Nanotechnology field. The situation that appeared during labeling the preferred terms guided approach all the terms as preferred terms that appeared more than once.

In addition, new concepts were also admitted as member terms to Nanotechnology thesaurus, as per the advice/interest of subject experts who would be the people to be served with this knowledge structure. Thus, in the present study concepts of 'literary warrant' and 'user warrant', were applied mutually.

Table2
Frequency of Occurrence of Terms in Literature

Terms	Frequency
Nanotechnology	366
Nanometrology	344
Atomic force microscopy	266
Calibration	157
Nanostructured materials	139
Nanoprobes	6
Quantum dot	5
Fullerene	4
Nanograting	3
Nanosensors	2

6. Semantic Relationship between the Concepts

Semantic relationship can be described as meaningful linking between two or more entities. Concept relationship suggested semantic map between concepts. Semantic linking between concepts is an intellectual work and usually performed by a group of subject experts and information science professionals. Efforts made in this direction helps the society by providing organized knowledge structure. This will further help human beings in making logical reasoning. The basic relationships described by Mazzocchi et al (2007) are Hierarchical, Associative and Equivalence.

Hierarchical Relationship: Pairing of terms under this are represented in their super-ordinate or sub-ordinate status. Notations used to display this category are BT (Broader Term) & NT (Narrow Term).

Eg: **annealing**
BT nano fabrication
NT ball milling annealing
laser annealing
thermal annealing

Equivalence Relationship: This category is used to denote the relationship between preferred and non-preferred terms. Notations uses are USE & UF (Used for).
As suggested by Gopinath (1985, p59-60), following term forms were covered under this category

i. **Variant spellings**
Eg: fiber laser & fibre laser

ii. **Abbreviations and Full name**
Eg: AFM & atomic force microscopes

iii. **Synonyms**
Eg**:** ballistic conduction & ballistic transport

Associative Relationship: This type of association is used to denotes the relationship between terms that are neither hierarchical nor equivalence. Notation used to represent this category is RT (Related Terms).

To define the categories of terms that can be incorporated under this class, Chowdhury (1999, p.130) has listed various categories. For this work following categories were defined:

i. **A process or operation and its agent or instrument**
Eg: interferometry & interferometers
nano fabrication & focused ion beam

ii. **Action and the products of the action**
Eg: emission spectra & emission spectroscopy
scanning probe lithography & nano scale features

iii. **Concepts and its properties**
Eg: nano sensors - sensitivity (property of)
nano structures - dielectric properties (property of)

iv. **A technique similar to some other technique**
Eg: ferromagnetic resonance is similar to electron magnetic resonance, and also somewhat similar to nuclear magnetic resonance

7. Standards and Guidelines

Standards and guidelines are developed by the community to label best practices, clarify principles and resolve conflicts. Like other fields, standards contribute significantly to terminology work. They help the practitioners in making decisions at various stages of thesaurus construction (Krooks and Lancaster, 1993). Ample efforts have been made in this direction:

ANSI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

The standard emerged with the overwhelming efforts of the working groups of National Information Standards Organization (NISO). It is meant for monolingual thesaurus and offers guidelines for display, construction, evaluation, maintenance and management of controlled vocabulary tools. The development of proposed standards involves rigorous peer review from NISO voting members. Other interested parties are also invited to participate in

reviewing process. With the completion of peer review process, standard is submitted to the American National Standards Institute (ANSI) for its final approval. Once approved and verified, NISO standards get the status of American National Standards (NISO, 2010). All the standards listed under NISO projects are openly available (Milstead 1998).

ISO 25964: Information and documentation - Thesauri and interoperability with other vocabularies

ISO 25964 is an international standard issued by the International Organization for Standardization. This standard is the revised, updated version of its precursors i.e. ISO 2788, jointly produced by UNESCO and ISO in 1972-73 (UNESCO 1976). Key difference between the two generations is of the move from conventional printed thesaurus world to a networked scenario that focuses on interoperability.

It is available in following two parts and can be purchased individually either from ISO or from any of its member nations.

Part 1: Thesauri for information retrieval
Part 2: Interoperability with other vocabularies

In addition to the international standard ISO 25964 and national standard ANSI/NISO Z39.19 (United States) there exists other national standards for thesaurus construction like BS 5723 (1987, British), AFNOR NFZ 47-100 (1981, French), and DIN 1463 (1987-1993, German) Morville and Rosenfeld (2006).

As the present work deals with the construction of monolingual thesaurus for Nanotechnology domain, therefore, as suggested in the contribution by Aitchison, Gilchrist and Bawden (2000), the US Standard ANSI/ NISO Z39.19-2005 which is available freely in public domain was consulted. Guidelines covered in a book by Aitchison, Gilchrist and Bawden were also followed in addition to ANSI/ NISO Z39.19-2005. Further, to represent the structure of entries, IET INSPEC Thesaurus was approached. These listed standards and guidelines helped the researcher in taking various decisions regarding the form of terms.

I. **Singular and Plural Forms**

Use of plurals: Names of objects or concepts that are subject to the question “How many?” but not “How much”? should normally be expressed as plurals.

Eg: errors
gratings
interferometers
microscopes
standards

- **Use of singular:** Names of materials or substances that are subject to the question “How much”? but not “How many”? should be expressed in the singular

Eg: activities, properties should be expressed in the singular
activities: measurement, calibration, error analysis
properties: precision, reproducibility

II. **Noun Form**

Efforts were made to represent the concepts in noun form

Eg: buckypapers
calibration
capacitance
fullerene

- **Proper Noun Form:** Use of institution names, organization names, personal names, manufactures etc., are not recommended. However, the cases where personal name forms an essential part of the compound terms, forced their inclusion.

Eg: Raman nano metrology

- **Prepositional Noun Form:** While listing the indexing terms, use of preposition alone was excluded. Efforts were made to remove the prepositions by changing the text segment.

Eg: effect of shape CONVERTED TO shape effect
thickness of films CONVERTED TO film thickness

III. **Adjective Forms**

According to the standards, adjectives are not suitable to denote indexing terms. But the cases where adjectives form the important part of compound terms like size, dimension etc., the usage of adjectives may be acceptable.

Eg: **brittle** fracture (mechanical property)
low energy electron diffraction (energy level)
nano structures (size)
one dimensional nano structures (dimensional property)

IV. **Abbreviations**

To facilitate universal understanding of concept, use of abbreviated forms is not recommended in standards. In the present work, cross references have been made from the abbreviated forms. This will also help in maintaining uniformity throughout the work.

Eg: AFM
USE **atomic force microscopes**
CARS
USE **coherent anti stokes Raman spectroscopy**

V. **Capitalization**

It is recommended in the guideline that predominantly lowercase characters be used for terms in controlled vocabularies. Capitals should be used only for the initial letter(s) of proper names.

Eg: Bosch process

VI. **Parentheses**

To eliminate problems in filing and searching, use of parentheses should be avoided in controlled vocabularies. In the present work, parentheses were used rarely to enclose qualifiers.

Eg: surface treatment (semiconductor technology)

VII. **Spelling – Authorities**

British English was used rather than American English. To include both spellings, references were made by employing the prefix 'USE' for directing the users from American spelling to British spelling.

Eg: fiber optics
USE **fibre optics**
characterization
USE **characterisation**
chemical vapor deposition
USE **chemical vapour deposition**

VIII. **Punctuation Marks**

To eliminate problems in filing and searching, use of hyphens is avoided and the same is replaced by one character space between words

Eg: x-ray as x ray

Further, during literature scanning task, it was observed that different authors used the prefix 'nano' in distinct format. For example: literature contains the concepts nanoobject in various form like nanoobject, nano-object, nano object. Thus, to maintain uniformity throughout the work, a single space is inserted after the prefix 'nano'.

IX. **Synonyms**

In case of interchangeable words, subject experts were consulted to find the word which is in common usage. Accordingly, the most commonly used word was admitted as preferred term and references were made from non-preferred terms to preferred terms.

Eg: ballistic conduction
USE **ballistic transport**

X. **Scope note**

Usually, scope note for a preferred term is added to define its scope. In the present work, majority of the scope notes denotes the application of concept for Nanotechnology, as covered in literature, whereas, some terms which are identified by the experts as core concepts are defined to explain the concept.

Eg: **precision**
SN precision is the ability of a measurement to be consistently reproduced

self assembly
SN using self assembly for the fabrication of nano scale electronic and photonic devices

XI. **Arrangement**

For arrangement of terms in alphabetical order there are two basic systems Word- by- Word and Letter-by-Letter.

Word- by- Word: In Word- by- Word arrangement, the filing principle is called 'nothing before something'. It keeps all the terms together that begins with same word.

Eg: nano Raman spectroscopy
nano SQUID
nanoscale surface metrology
nanoscience

nanotransistors
nanotubes
nanowires

Letter-by–Letter: Letter-by–Letter approach is usually preferred in dictionaries. According to Wellisch (1999), this type of organisation may be applied for the continuation of an existing arrangement, or where it is essential to have various spellings of the same heading together. In this system, word spaces and all other punctuation marks are ignored.

Eg: nano Raman spectroscopy
nanoscale surface metrology
nanoscience
nano SQUID
nanotransistors
nanotubes
nanowires

Taking into account the pattern followed in INSPEC thesaurus, word-by-word alphabetization system was followed in the compilation of Nanotechnology thesaurus. The same pattern of arrangement was suggested in the The Chigaco Manual of Style (2003) which stated that in a system with many open compound terms starting with the same word, the word-by-word system may be easier for users.

Conclusion

Research in interdisciplinary areas like Nanotechnology is among the rapidly growing fields in Science and technology wherein there is no single approach for research. Scholars working in different discipline are using their own terminology to represent the same concept. Thus, it becomes difficult to ensure access to all relevant literature. Some relief has been achieved with the terminology standardization initiated by organizations like ISO, ASTM etc. However, no study has been traced where efforts have been made to define the semantic relationships between the Nanotechnology concepts whose scope is limited to Physical Sciences. Therefore, by designing Nanotechnology thesaurus, the present study aimed at representing and expressing semantic knowledge mapping for Nanotechnology concepts within the defined scope.

References

1. American National Standards Institute & National Information Standards Organisation (ANSI/NISO) (2010), *Guidelines for the construction, format, and management of monolingual controlled vocabulary* (ANSI/NISO Z39.19-2005 (R2010)), National Information Standards Organisation, Maryland, USA. Available at:http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. (Accessed 04 December 2011).
2. Aitchison, J & Clarke, SD (2004), "The thesaurus: A historical viewpoint, with a look to the future", *Cataloging & Classification Quarterly*, Vol. 37 No. 3-4, pp. 5-21. doi: http://dx.doi.org/10.1300/J104v37n03_02
3. Aitchison, J & Gilchrist, A (1987), *Thesaurus construction: A practical manual*, 2nd ed, Aslib, London.
4. Aitchison, J, Gilchrist, A, & Bawden, D (2000), *Thesaurus construction and use: A practical manua*l, 4th ed, Europa, London.

5. Anderson, JD & Perez-Carballo, J (2001), "The nature of indexing: How humans and machines analyze messages and texts for retrieval: Part I: Research, and the nature of human indexing", *Information Processing and Management*, Vol. 37 No. 2, pp. 231-254. doi:10.1016/S0306-4573(00)00026-1.
6. Bernier, CL (1968), "Indexing and thesauri", *Special Libraries*, Vol. 59, pp. 98-103.
7. Chen, H, Lynch, KJ, Basu, K & Dorbin, T (1993), "Generating, integrating, and activating thesauri for concept-based document retrieval", *IEEE Expert*, Vol. 8 No. 2, pp. 25-34. doi: 10.1109/64.207426.
8. Chen, L & Thiel, U (2004), "Language modeling for effective construction of domain specific thesauri", In F. Meziane & E. Metais (Eds.), *Proceedings of the 9th International conference on applications of natural language to information*, Springer-Verlag, Berlin, Heidelberg, pp. 242-253. Available at: http://books.google.co.in/books?id=7FD94nXAOboC&printsec=frontcover#v=onepage&q&f=false. (Accessed 07 November 2014).
9. Chowdhury, GG (1999), *Introduction to modern information retrieval*, Library Association, London.
10. Dabney, D (2007), "The universe of thinkable thoughts: Literary warrant and West's key number system", *Law Library Journal*, Vol. 99 No. 2, pp. 229-247. Available at: http://www.aallnet.org/main-menu/Publications/llj/LLJ-Archives/Vol-99/pub_llj_v99n02/2007-14.pdf. (Accessed 10 July 2014).
11. Devasason, FJ (n.d.), "Generation of thesaurus in different languages: A computer based system", pp. 303-314. Available at: www.aclweb.org/anthology/C80-1046. (Accessed 12 November 2014).
12. Ghose, A & Dhawle, AS (1977), "Problems of thesaurus construction", *Journal of the American Society for Information Science*, Vol. 28 No. (4), pp. 211-217.
13. Gopinath, MA (1985), "Equivalence relations in information retrieval thesaurus", *SRELS Journal of Information Management*, Vol. 22 No. 1, pp. 57-63.
14. Harping, P (2010), *Introduction to controlled vocabularies: Terminology for art, architecture, and other cultural works*, Getty Research Institute, Los Angeles. Available at: http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/ (Accessed 25 July 2014).
15. Hedden, H (2008), "Controlled vocabularies, thesauri, and taxonomies", *The Indexer*, Vol. 26 No.1, pp. 33-34.
16. Hulme, EW (1950), *Principles of book classification*. London: Association of Assistant Librarians, pp. 444-449. Available at: http://www.iva.dk/bh/Core%20Concepts%20in%20LIS/Hulme_444-449.pdf (Accessed 10 July 2014).
17. Krooks, DA & Lancaster, FW (1993), "The evolution of guidelines for thesaurus construction", *Libri*, Vol. 43 No. 4, pp. 326-342.
18. Kumbhar, R (2005), "Speciator based faceted depth classification's application in thesaurus construction", *Annals of Library and Information Studies*, Vol. 52 No. 1, pp. 15-24. Available at: http://nopr.niscair.res.in/bitstream/123456789/3982/1/ALIS%2052(1)%2015-24.pdf (Accessed 01 November 2014).
19. Lancaster, FW (1972), *Vocabulary control for information retrieval*, Information Resource Press Washington, DC.
20. Lancaster, FW (1985), *Thesaurus construction and use: A condensed course*, UNESCO, Paris. Available at: http://unesdoc.unesco.org/images/0007/000703/070359eb.pdf (Accessed 07 May 2013).

21. Lesko, M. (n.d.), "Nothing is secret anymore!: The confessions of a millionaire information broker". Available at: http://www.freestuffforentrepreneursontheinternet.com/nothingissecret.htm. (Accessed 11 April 2012).
22. Mazzocchi, F, Tiberi, M, De Santis, B & Plini, P (2007), "Relational semantics in thesauri: Some remarks at theoretical and practical levels", *Knowledge Organization*, Vol. 34 No. 4, pp. 197-214.
23. Milstead, JL (1998), "NISO Z39.19: standards for structure and organization of information retrieval thesauri", A paper presented at the workshop on the compilation, maintenance, and dissemination of taxonomic authority file in, Washington, DC, 22-23 June, 1998. Available at: http://researcharchive.calacademy.org/research/informatics/taf/proceedings/milsteadtaf.html. (Accessed 13 November 2014).
24. Morville, P & Rosenfeld, L (2006), *Information architecture for the world wide web: Designing large scale web sites*, 3rd ed, O'Reilly, Sebastopol, US.
25. Pinto, M (2008), "A user view of the factors affecting quality of thesauri in social science databases", *Library & Information Science Research*, Vol. 30, pp. 216-221. doi: 10.1016/j.lisr.2007.12.003.
26. Raizada, AS, Ramachandran, M, Satyanarayana, R & Pal, Y (1975), "Principles of thesaurus construction and Indian science abstracts indexing system", pp. A22-A28. Available at: http://drtc.isibang.ac.in/ldl/bitstream/handle/1849/467/Paper%20AD.pdf?sequence=1&isAllowed=y (Accessed 12 November 2014).
27. Roberts, N (1984), "The pre-history of the information retrieval thesaurus". *Journal of Documentation*, Vol. 40 No. 4, pp. 271-285. doi: http://dx.doi.org/10.1108/eb026769.
28. Spade, JM (2012), *Information as power: China's cyber power and America's national security*, U.S. Army War College, Pennsylvania. Available at: http://www2.gwu.edu/~nsarchiv/NSAEBB/NSAEBB424/docs/Cyber-072.pdf . (Accessed 27 October 2014).
29. Surace, CJ (1970), "The display of a thesaurus (Report No. P-4331)". Available at: files.eric.ed.gov/fulltext/ED039002.pdf. (Accessed 12 November 2014).
30. The Chicago manual of style: Indexes (2003), The University of Chicago, Chicago. Available at: http://www.press.uchicago.edu/Misc/Chicago/CHIIndexingComplete.pdf. (Accessed 16 July 2014).
31. United Nations Educational, Scientific and Cultural Organization (UNESCO) (1976), *UNISIST guidelines for the establishment and development of multilingual thesauri*. UNESCO, Paris. Available at: http://unesdoc.unesco.org/images/0002/000205/020561eb.pdf. (Accessed 04 November 2014).
32. Vickery, BC (1960), "Thesaurus – A new word in documentation", *Journal of Documentation*, Vol. 16 No. 4, pp. 181-189. doi: http://dx.doi.org/10.1108/eb026286 .
33. Wellisch, HH (1999), *Guidelines for alphabetical arrangement of letters and sorting of numerals and other symbols* (Report No. TR03-1999). Maryland, USA: National Information Standards Organization. Available at: http://www.niso.org/publications/tr/tr03.pdf. (Accessed 16 July 2014).