Designing Thesaurus with Multilevel Hierarchy Systems

Partha Sarathi Mandal, Sukumar Mandal

[1]PhD Research Scholar, Department of Library and Information Science The University of Burdwan, Burdwan, 713104.
[2]Assistant Professor, Department of Library and Information Science The University of Burdwan, Burdwan, 713104.

ARTICLE INFO

ABSTRACT

Information retrieval systems are designed to retrieve information from a database quickly and accurately. A multilevel hierarchy is a recursive process that consists of logically related levels with relation and is represented in a linear sequence. They are intricate among the parts of a whole, statistically or dynamically. This paper explores the significance of incorporating multilevel hierarchy systems in designing a thesaurus. It aims to enhance the efficiency and usability of the linguistic resources. Thesauri are crucial in information retrieval, establishing semantic relationships between words. Traditional thesauri have employed a single-level hierarchy, but recent advancements in information organization and retrieval call for more sophisticated structures to accommodate the complexities of language and user needs. This article presents the methodologies and benefits of designing thesauri with multilevel hierarchy systems, which not only enhance the efficiency and usability of linguistic resources but also significantly improve the accuracy and speed of information retrieval.

Introduction

With the advancement of new technologies and the growing demand of the information society, libraries are trying to introduce innovative solutions to develop their organizational efficiency and user satisfaction. Integrating multilevel hierarchy with MultiTes, a thesaurus construction tool, is one of the advancements among various solutions. Thesaurus is essential for information retrieval, linguistics, and knowledge representation. Traditionally, it has used a single-level hierarchical structure to organize terms based on broader and narrower relationships. However, combining hierarchical structures with MultiTes technology optimizes library information management and empowers librarians and users with a sophisticated system. It can create a hierarchical structure of related terms in library and information science for information retrieval systems. This tool allows us to create a thesaurus organized into multiple levels. This study shows a 2-way or multilevel hierarchy within different terms or concepts in library science, which is significant as it allows for more nuanced and comprehensive information retrieval. It also emphasizes the pervasiveness of the information, accessibility, and user interaction within and outside the libraries.

Literature Review:

This paper presents a comprehensive review of related topics chronologically. Eid (2022) suggests a proposal for controlling the vocabulary system for Arabic languages on

the web, designed based on ISO26162:2012 for handling facets and textual content. Malik and Mandal (2022) include the development of a multilingual integrated framework for users to retrieve information on various terms and facets easily. Sunny and Angadi (2018) show the functionality of digital thesaurus. Multiple tools are used to create a digital thesaurus that can be accessible to the users. Mandal, S (2016) constructs thesaurus through templates for library management and retrieval. According to Stellato et al. (2015), devoted users easily separate the conceptualization and editing of terminology. Editors may use SKOS and integrate concepts as required. Mandal, S (2015) configures MultiTes to build an independent thesaurus from HTML (Hyper Text Markup Language) pages. Lana Soglasnova and Mary Hanson (2015) evaluate terminology's social responsiveness and relevance for constructing a community clinic library. Louise F Spiteri (2005) examines the hierarchies and categories of inter-term relationships for the thesaurus construction. Marianne, Lykke Nielsen (2004) presents various approaches to constructing thesaurus and their problems in collection, organization, and formation. All these papers discuss data integration, enrichment, and discoverability. Dykstra, M. (1989) defines thesaurus as a structured controlled vocabulary. It helps the indexer analyze the documents.

Objectives

The objectives of this paper are the following:

- To show the data input system in the establishment of a thesaurus using the MultiTes;

- To integrate various authority data;

- To access different hierarchy relationships;

- To demonstrate the technique as a prototype for publishing thesaurus gateways in the browser.

Methodology

This study is conducted to explore the practical application of the multilevel hierarchy system in the library. It is carried out with a systematic review. A literature search was done using the Web of Science (WoS), Scopus, LISA, and LISTA databases. For this study, MultiTes, a practical vocabulary control tool, is used. This tool, available at http://www.multites.com, was chosen for its support of both Ubuntu and Windows environments, as well as its support for both monolingual and multilingual thesauri with multilevel hierarchy. The constructions of the thesaurus and design

issues are demonstrated in the following two sections:

- Construction of thesaurus using MultiTes: This section details the step-by-step process of constructing a thesaurus using MultiTes, a practical vocabulary control tool.

- Architectural Designing for Multilevel Hierarchy.

This paper also demonstrates the various hierarchical relationships of the terms and their facets.

Construction of thesaurus: The thesaurus is built based on "Geographic Locations." It is made with the following steps:

- Create first thesaurus: At first, create a thesaurus with a file name;

- Collecting Terms: The relevant terms of the subject domain are included in thesaurus. Raw terms or vocabulary are imputed here.

- Modification of Terms: The terms are modified to standardize. Ambiguous terms are removed to ensure consistency with the subject domain.

- Establishing Relations: In this step, the terms are related to semantic relations between the terms, such as synonymy, antonym, and hierarchical relations. These are posted with various hierarchy relationships like 'Broader Term' (BT), 'Narrower Term' (NT), and 'Related Term' (RT).

- Quick data entry – Here, one can enter raw data quickly into the thesaurus with ctrl-Q

- New Term: New terms are also directly here by pressing ctrl-N

- View More Options – One can add more relationships here

- Thesaurus Display Format: In the final step, the display formats chosen for the results are displayed. Here, one can print various reports of Top terms, alphabetical, hierarchical, 2-way hierarchical, and multilevel hierarchical. It also generates HTML and XML files.

Architectural Designing for Multilevel Hierarchy: From the 'Report Generation' interface, one can modify as required from the tab menu like 'Term

Selection,' 'Display Items,' and 'Select by Relationship, Select.' From 'Term Selection,' one can get the specific term result declaring the first term and last term with various filtrations like preferred, non-preferred, Top, Orphans, and terms with flags. It also filters the terms with status and date. From 'Display Items,' one can suggest various relationships like: 'SC' for 'Subject Category,' 'SN' for 'Scope Note,' 'UF' for 'Used For,' 'Use' for 'Use,' 'NT' for 'Narrower Term,' 'BT' for 'Broader Term,' 'RT' for 'Related Term,' 'ENG' for 'English,' 'STA' for 'Status,' 'INP' for 'Input,' 'APP' for 'Approval Date,' 'UPD' for 'Last Update Date,' 'NVD' for 'Not Valid Date,' 'TNR' for 'Term Number,' 'FLG' for 'Flag' and 'Personal Note' for 'Personal Note.' From 'Selection by Relationship' and 'Select by Categories,' one can filter the terms with relationship and category, respectively

Fig. 1: Different Output Format for the Search Result

Fig.1 shows different interfaces for getting various output formats for the search result. These options may be 'Screen,' 'Printer,' 'text File,' 'Delimited Text,' 'HTML,' 'XML,' 'SKOS/RDF,' 'Custom HTML,' 'RTF,' 'CVS,' 'Cartesian, 'CVS+Extra Delim' and 'CVS column.'

Demonstrate Hierarchy in Library System:

Three hierarchical relationships exist in constructing thesaurus. These are the following:

- Genus and species, e.g. dog / Bulldog,

- Concept and instance, e.g. River / Ganga River

- Whole and part, e.g. Body / Arm.

Generally, the whole-part relationship is seen in associative relationships, but there are four exceptional cases where the part implies the whole. These are the following –

- Systems and organs, for example Central Nervous System - brain – white matter

- Geographical location, for example: India – West Bengal – PurbaBardhaman

- Discipline or Subject, e.g. Science - Physics – Classical Mechanics - Kinetics

- Hierarchical social structure, e.g., Hinduism – Brahmins.

Fig. 2: Hierarchical Term Generation using MultiTes

Fig.2 shows the hierarchical terms generation. The raw terms are displayed in an alphabetic order. Selecting HTML output files, one can get a lot of index files. The primary file name is 'repgen'. One can customize this file according to the requirements.

Fig. 3: 2-way Term Hierarchy

Fig. 3 explores the 2-way term hierarchy. Placing the cursor on the particular term, one can click on "Term Hierarchy" and choose '2-way' from the drop-down list to see the result of the 2-way hierarchy. One can set the limit and level of hierarchy here. It may be 'BT/NT' or 'All.'

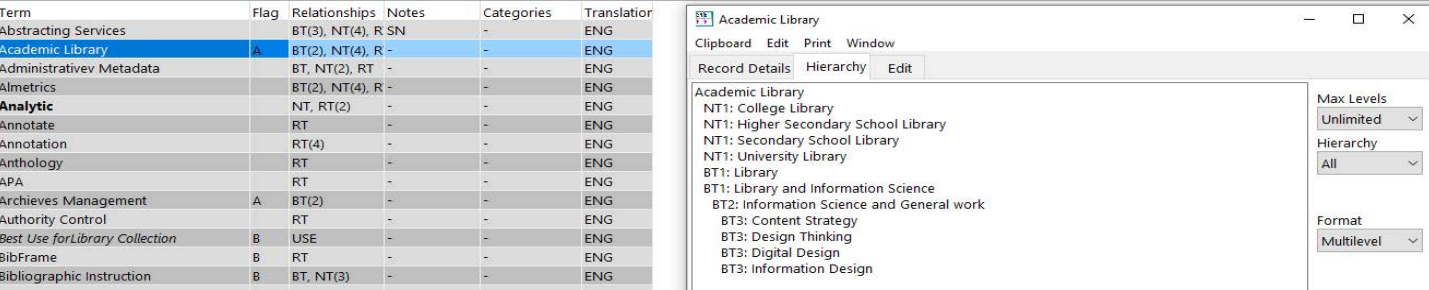| Term	Flag	Relationships	Notes	Categories	Translation
Abstracting Services		BT(3), NT(4), R	SN	-	ENG
Academic Library	A	BT(2), NT(4), R	-	-	ENG
Administrativev Metadata		BT, NT(2), RT	-	-	ENG
Almetrics		BT(2), NT(4), R	-	-	ENG
Analytic		NT, RT(2)		-	ENG
Annotate		RT	-	-	ENG
Annotation		RT(4)	-	-	ENG
Anthology		RT	-	-	ENG
APA		RT	-	-	ENG
Archieves Management	A	BT(2)	-	-	ENG
Authority Control		RT	-	-	ENG
Best Use forLibrary Collection	B	USE	-	-	ENG
BibFrame	B	RT	-	-	ENG
Bibliographic Instruction	B	BT, NT(3)	-	-	ENG

Fig. 4: Multilevel Hierarchy

Fig 4 exposes a multilevel hierarchy of the terms. Placing the cursor on the particular term, one can click on the "Term Hierarchy" and choose 'Multilevel' from the drop-down list to see the result of the multilevel hierarchy list. Multilevel hierarchies facilitate collaboration between libraries globally, allowing for the standardization of categorization and improved resource sharing.

Fig. 5: Platform for Federated Search

Fig 5 suggests the designed search interface for searching the terms from the thesaurus. Google, VIAF (Virtual International Authority File), and LOC (Library of Congress) are integrated to search various terms and concepts from a single user interface. This authority data search interface provides detailed information. The integration of augmented reality (AR) and virtual reality (VR) provides better insights as users navigate the multilevel hierarchy. It also facilitates global collaboration and resource sharing.

Benefits of Multilevel Hierarchy Systems:

Multilevel hierarchy systems have several advantages. These are the following:

- Resource Management—This system helps librarians manage digital collections effectively and efficiently. It ensures high recall and precision of the resources.

- Information Retrieval – This system enhances the precision of the information retrieval system. It is user-friendly, so one can search for information, reducing information overload.

- Resource Discovery – This system helps in efficient resource discovery.

- User satisfaction – User interface design in multilevel hierarchy systems provides navigation and accessibility for diverse user groups.

Conclusion

In conclusion, the multilevel hierarchy introduces a day break in library management, elevating accessibility, efficiency, and user expectations. Interacting with the web of information, libraries can meet the demands of diverse and dynamic users' needs. The strategic implementation of multitier tessellation enhances data organization and facilitates user interactions. Both librarians and users can access required information or knowledge at their fingertips. This emerging technology not only faces the instant challenges in the age of information exploration but also positions them to thrive in the future.

References

Dykstra, M. (1989). Subject analysis and thesauri: A background. *Art Documentation: Journal of the Art Libraries Society of North America*, 8(4), 173–174. https://www.jstor.org/stable/27948126

Eid, S. (2022). A proposed Arabic monolingual web-based terminology management system. *Journal of Library Metadata*, 22(1–2), 47–61. https://doi.org/10.1080/19386389.2022.2042168

Soglasnova, L., & Hanson, M. (2015). Socially responsive design

and evaluation of a workers' compensation thesaurus for a community organization with selective application of cognitive work analysis: A case study. *Cataloging & Classification Quarterly*, *53*(8), 905–926. https://doi.org/10.1080/01639374.2015.1044632

Spiteri, L. F. (2005). Word association testing and thesaurus construction: A pilot study. *Cataloging & Classification Quarterly*, *40*(1), 55–78. https://doi.org/10.1300/J104v40n01_05

Malik, S., & Mandal, S. (2022). Sparql and skos enabled multilingual polyhierarchical thesaurus: Designing a framework. *SRELS Journal of Information Management*, 239–246. https://doi.org/10.17821/srels/2022/v59i4/169656

Mandal, S. Developing Thesaurus Construction through Tematres for the College Libraries under the University of Burdwan. International Journal of English Language, Literature and Humanities. 2016, 4 (6). 302 -316. www.ijellh.com

Mandal, S. MultiTes: A Knowledge Organization Thesaurus Construction Tool for College Libraries under the University of Burdwan. International Research Journal of Interdisciplinary & Multidisciplinary, 2015, 63 – 79. http://www.irjims.com

Nielsen, M. L. (2004). Thesaurus construction: Key issues and selected readings. *Cataloging & Classification Quarterly*, *37*(3–4), 57–74. https://doi.org/10.1300/J104v37n03_05

Myntti, J., & Neatrour, A. (2015). Use existing data first: Reconcile metadata before creating new controlled vocabularies. *Journal of Library Metadata*, *15*(3–4), 191–207. https://doi.org/10.1080/19386389.2015.1099989

Stellato, A., Rajbhandari, S., Turbati, A., Fiorelli, M., Caracciolo, C., Lorenzetti, T., Keizer, J., & Pazienza, M. T. (2015). Vocbench: A web application for collaborative development of multilingual thesauri. In F. Gandon, M. Sabou, H. Sack, C. d'Amato, P. Cudré-Mauroux, & A. Zimmermann (Eds.), *The Semantic Web. Latest Advances and New Domains* (Vol. 9088, pp. 38–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-18818-8_3

Sunny, S. K., & Angadi, M. (2018). Evaluating the effectiveness of thesauri in digital information retrieval systems. *The Electronic Library*, *36*(1), 55–70. https://doi.org/10.1108/EL-02-2017-0033