



Taxonomy Development of PTAR Local Content Collections: A critical analysis of the expert-assigned and ICSH subject headings

A'dillah Mustafa¹, Eza Eliana A. Wahid², Noredayu Rosli³, Mawar Noor Abdullah³

¹ School of Information Science, College of Computing, Informatics & Mathematics, 40150 Shah Alam, Malaysia

² Tun Abdul Razak Library, Universiti Teknologi MARA, 40450 Shah Alam, Malaysia

³ Tun Dr Ismail Library, Universiti Teknologi MARA Johor, Segamat, Malaysia

adillah973@uitm.edu.my, eza080@uitm.edu.my, noredayu.9615@uitm.edu.my, mawar146@uitm.edu.my
Tel :012-2196358

Abstract

Developing a taxonomy for local content collection is a massive challenge for librarians which involves understanding the complexity of a particular subject classification. This paper aims to report the analysis of the taxonomy developed for the Sports Science & Recreational compared to Library of Congress Subject Headings (LCSH). The analysis employs subject analysis as the methodology for comparing expert-assigned keywords with LCSH. The analysis found 6 divisions in the local content compared to only 2 divisions in LCSH, and 364 subjects compared to only 98 in LCSH. The percentage of terms categories includes exact terms used (14%), similar terms used (11%), and not used terms (97%).

Keywords: Taxonomy, Sports science, Library classification, Academic Library

eISSN: 2398-4287 © 2024. The Authors. Published for AMER and cE-Bs by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers and cE-Bs (Centre for Environment-Behaviour Studies), College of Built Environment, Universiti Teknologi MARA, Malaysia.
DOI: <https://doi.org/10.21834/e-bpj.v9iS118.5473>

1.0 Introduction

Developing a taxonomy for local content collection is a massive challenge for librarians which involves understanding the complexity of a particular subject classification. Subject classification and knowledge organization in Library and Information Science are built upon a structured taxonomy known as Ontology. The development of taxonomy is affected by the massive production of literature and the rapid development of new discoveries of knowledge. Sports Science and Recreational are among the innovative subjects in recent world knowledge development. Hedden (2022) defines taxonomy as a type of knowledge-organizing system in which concepts are connected hierarchically. Taxonomy development according to Abukhader (2019) is the basis of knowledge classification. The importance of taxonomy development in research has been long acknowledged in identifying subject classification. According to Nickerson, Varshney & Muntermann (2013), taxonomy helps researchers understand the classification of objects and analyze complex domains, and also in understanding the divergence in previous research (Sabherwal & King, 1995). Problems in subject classification in specific subjects depend on the natural and local behavior of the content (Prashasti, Rajesh, Vivek, David & Fernando, 2020). Research reporting difficulties in taxonomy development including in information systems (Nickerson, Varshney & Muntermann, 2013), biology, and knowledge management (Abukhader, 2019). Concerning sports science, De Pauw, Roelands et. al (2013) highlight the utmost importance problems to standardizing the data according to the classified subject groups. Sport science & recreational subject classification in UiTM Local Content Hub is very much localized in the Malaysian context. Therefore the classification of the domain is

eISSN: 2398-4287 © 2024. The Authors. Published for AMER and cE-Bs by e-International Publishing House, Ltd., UK. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). Peer-review under responsibility of AMER (Association of Malaysian Environment-Behaviour Researchers and cE-Bs (Centre for Environment-Behaviour Studies), College of Built Environment, Universiti Teknologi MARA, Malaysia.
DOI: <https://doi.org/10.21834/e-bpj.v9iS118.5473>

unstructured and lacks uniformity compared to the established standard subject headings used in Library of Congress Subject Headings. The usefulness of controlled vocabulary has been debated over the years. Olson & Boll (2001) acknowledge these issues as becoming imperative with the popularity of online tools such as Google and the use of keywords as users' primary search strategy. However, several research debated on whether controlled vocabularies, such as Library of Congress Subject Headings (LCSH) are applicable to local content collections (Allyson, 1989 & Arievidtch, 2020). Thus, this study is intended to confirm the usability of the expert-assigned keywords with a comparison to LCSH. Thus, this study will provide a comprehensive analysis of sports science & recreational taxonomy including the knowledge domain and classification research, and subject classification works. This paper aims to report the analysis of expert-assigned keywords with a comparison to Library of Congress Subject Headings.

2.0 Literature Review

Libraries face a similar problem when it comes to classifying their documents and resources (Keck, Couton, & Altermatt, 2023). Most libraries find it difficult to distinguish the level of taxonomy and classification. Ensuring the success of taxonomy development requires both taxonomy expertise and in-depth knowledge of the knowledge/content. Therefore, it is crucial to include subject matter specialists and librarians in the team to help in identifying the taxonomy for the content. This review of literature will provide an overview of the taxonomy development through the basic lenses of definitions and differences in the three related elements which are Ontology, Taxonomy, and Classification.

2.1 Defining ontology, taxonomy, and classification

A general understanding of ontology, taxonomy, and classification liaises on the complexity level. Merriam-Webster Dictionaries (2019) classified ontology as the most basic and unstructured concept, followed by classification as a systematic arrangement. While, taxonomy is derived from the Greek word *taxis*, which means organization or order, and *nomos*, which means law or science. The term taxonomy is used in two aspects; first, the specific meaning, which is to refer to a hierarchical classification or categorization system, and second the broader sense, meaning to refer to any method of arranging knowledge concepts (Hedden, 2022). In library and information science, classification is described as hierarchically organizing knowledge by classifying concepts and topics concerning one another (Chatterjee, 2021). Classifying is also an investigative approach that involves sorting objects or events into groups or categories. The systematic arrangement of things based on certain similarities or differences allows us to better understand relationships and connections between things.

2.2 Differentiating classification and taxonomy

Naveen (2018) explains classification as the arrangement of organisms according to a set of principles and can be numerous, whereas taxonomy is one defined and the most respected classification system. In understanding the classification system and taxonomy development, librarians must possess the ability to differentiate the nature of each process. Table 2 highlights the differences between both processes.

Table 1: Difference between taxonomy and classification system

Classification System	Taxonomy
1. Based on codes and notation.	1. Did not apply any codes.
2. Systematic arrangements in groups or categories according to established criteria.	2. A form of knowledge organization system in which concepts are linked to one another in a hierarchical order.
3. Limitation in the expansion of numbers – only within the structure.	3. Can be expanded and adapted without any limitations or restrictions.
4. The process of distributing things into classes or categories of the same type.	4. Deals mainly with the description, identification, nomenclature, and classification of organisms.
5. It is created to be browsed from top to bottom in a hierarchical manner.	5. It is designed to be browsed, searched, or may not be completely presented to users.

Source: Hedden, (2022).

2.3 Taxonomy Development

The development of a taxonomy involves determining the characteristics of the objects of interest. The choice of characteristics in a taxonomy is a central problem in taxonomy development. The characteristics could be based on a theory but in reality, any 'theory' is often implicit (Aldenderfer & Blashfield, 1984). It is envisaged that taxonomies will change over time to incorporate new information. Taxonomy development helps to organize content and make connections between people and the information they need. Usman, Britto, Borstler & Hendros (2017) emphasize that developing taxonomies in software engineering can be expanded and must be done in a more systematic way to structure and better understand its area. As for the health and well-being subject, taxonomy development is continuing to evolve and update. Lee (2022) describes the need to adopt a consistent taxonomy to develop communication between and within the related sector. The study summarized the processes and outcomes of a collaborative, intersectoral, and interdisciplinary project to develop a shared terminology, taxonomy, and ontology for this area.

2.4 Expert-assigned keywords or Expert-assigned keywords

These keywords are chosen based on the expert assigned's judgment and understanding of the document's content, and they can vary in specificity and relevance (Sheng, Gero & Ho, 2022). The selection of appropriate expert-assigned keywords is essential for effective information organization and retrieval, as they help users locate and access relevant documents amidst a vast sea of information. According to Theda, Sevim & Margaret (2012) and Gil-Leiya & Alonso-Arroya (2007), the relationship between taxonomy development and expert-assigned keywords lies in their shared goal of categorizing and organizing information to improve its accessibility and retrievability. Both taxonomy development and expert-assigned keywords are methods used to structure and label information, but they serve different purposes and are often used in distinct contexts. The relationship is described in Table 2.

Table 2: Relationship between experts assigned keyword and taxonomy

Contexts	Expert-Assigned keyword	Taxonomy
Purpose and Scope	Specific words or phrases chosen by the creator of a document (e.g., a researcher, expert-assigned, or content creator) to describe the content of that particular document. Their purpose is to improve the discoverability of that specific work within a broader collection or database.	Typically comprehensive, hierarchical systems are used to classify and categorize a wide range of items or concepts within a specific domain. Their purpose is to create a structured framework for organizing knowledge, and they often involve a predefined set of categories and subcategories.
Hierarchy vs. Flat Structure	Typically a flat list of terms or phrases without inherent hierarchical relationships. Each keyword is treated as having equal importance in describing the content.	Hierarchical in nature, with categories organized in a structured manner, often with broader categories at the top and increasingly specific subcategories beneath them. This hierarchy allows for a systematic classification of information.
Scope of Application	Specific to individual documents and are applied at the discretion of the document's creator. They reflect the expert-assigned's understanding of the document's content and context.	Generally designed for broader, institutional, or field-wide use. They are applied consistently across a range of materials or documents within a particular domain.
Controlled Vocabulary vs. Free Text	More flexible and may include free-text terms that reflect the expert-assigned's unique perspective, terminology, or focus.	Involve the creation of controlled vocabularies, where terms are pre-determined and standardized to ensure consistency in categorization.

3.0 Methodology

This analysis applied a subject analysis method as the research approach. The analysis involved a comparison of the Sports Science & Recreational Digital Collection taxonomy from the UiTM Local Content Hub with the Library of Congress Subject Headings online (Classification Web). The Sports science & recreational digital collection taxonomy is based on the expert's expert knowledge which is known as an Expert-assigned keyword. The subject analysis involved 3 steps as described in Table 3.

Table 3: Subject analysis steps

Step	Procedure	Output
1	Identification of taxonomy divisions in both Sports Science & Recreational Digital Collection in UiTM Local Content Hub and Sport and Recreation in Library of Congress Subject Headings.	Comparison of the number of divisions
2	Identification of the number of subjects in both Sports Science & Recreational Digital Collection in UiTM Local Content Hub and Sport and Recreation in Library of Congress Subject Headings.	Comparison of the number of subjects
3	Summarizing the categories of terms in both Sports Science & Recreational Digital Collection in UiTM Local Content Hub and Sport and Recreation in Library of Congress Subject Headings.	Percentage of the terms categories

In summary, the analysis involved a total of 8 taxonomies at the level of divisions and 46 taxonomies of the subject's classification level. The analysis will identify the similarities and differences in the classification of the taxonomy used in the digital collection hub compared to the Classification Web.

4.0 Findings

The analysis reported the following main discoveries. The first analysis on the comparison of the number of divisions found that there are six (6) divisions in UiTM Local Content Hub compared to only two (2) divisions in LCSH. The comparison of the number of divisions from both collections is identified in below Table 4.

Table 4: Comparison of the number of divisions

UiTM Local Content Hub	Divisions	Library of Congress Subject Headings	Divisions
Sports Science & Recreational	Health and Fitness Martial Arts Outdoor recreation Sport Management Sport Science Sport Tourism	Sport and Recreation.	Sports Recreation
No of divisions	6		2

The second analysis on the comparison of the number of subjects found that there are 364 subjects in UiTM Local Content Hub compared to only 98 subjects in LCSH. The comparison of the number of subjects from both collections is identified in below Table 5.

Table 5: Comparison of the number of divisions

UiTM Local Content Hub	Divisions	No of Subjects	Library of Congress Subject Headings	Divisions	No of Subjects
Sports Science & Recreational	Health and Fitness	186	Sport and Recreation.	Sports	81
	Martial Arts	27		Recreation	17
	Outdoor recreation	38			
	Sport Management	156			
	Sport Science	210			
	Sport Tourism	18			
No of subjects		364			98

The analysis of the number of terms revealed that 164 terms from the Sports Science & Recreational Digital Collection subjects were analyzed, and three different categories were identified which are Exact terms used to refer to the same subject terms used in the LC classification web, Similar terms used refer to partial or similarity in the terms used, and finally Not used terms referring to terms not used in classification web. Below Table 6 highlights the number of exact terms, Table 7 number of similar terms, and Table 8 number of Not use terms.

Table 6: Number of Exact terms used

Division	Expert-Assigned Keywords and Terms used in LCSH
Sports Science	14
Sports Management	2
Health & Fitness	5
Outdoor Recreation	3

Table 7: Number of Similar terms used

Division	Expert-Assigned Keywords and Terms used in LCSH
Sports Science	4
Sports Management	1
Health & Fitness	3
Sport Tourism	1
Outdoor Recreation	1
Martial Art	1

Table 8: Number of Not used terms

Division	Expert-Assigned Keywords and Terms used in LCSH
Sports Science	31
Sports Management	47
Health & Fitness	14
Sport Tourism	12
Outdoor Recreation	1
Martial Art	21

5.0 Discussion

The taxonomy of the Sports Science & Recreational Digital Collection in the UiTM Local Content Hub shows that the terms were identified by the expert or expert-assigned keywords and are collective in research in other fields. The findings revealed a huge difference in the number of divisions. Omair & Alturki (2020) emphasized that the unstructured terms proposed by the professionals in system development are acceptable terms for their study. An important phase of taxonomy development highlighted by Poser, Wiethof & Bitter (2022) is the evaluation of division and subjects, using the illustrative scenario as the evaluation technique enables them to access the coherence of internal structure. Nevertheless, the differences in the number of divisions in the subject are also highlighted by Poser, Wiethof & Bitter (2022) to vary based on the subject development and new knowledge categorization.

The Exact term used in the study explained that the expert-assigned have the same understanding of the subject matter. The identification and selection of terms were influenced by the expert-assigned/user's prior knowledge of the subject. The exact terms were also reported by Milne (2010) in records management practice and Abukader (2019) in his study on library and information science which found a significant number of terms that are valid and usable for library and information science even though the percentage is at an average of 17%. This was reported to be influenced by the prior knowledge or subject background of the expert-assigned. The same understanding can be viewed in this study result which gained 14% exact terms used in UiTM Sports Science & Recreational Digital Collection.

The expert-assigned keywords or terms having similarity with the established classification were also reported by Valderrama-Zurián et al., (2021) in the cannabis research. They reported the majority of the terms are similar at 75% due to classification redundancy. Early Wang, Chaudhry & Khoo (2007) reported the challenge in the taxonomy is having redundancy in the categories, comprehension level,

subject coverage, and hierarchical structure of the classification terms. A recent study reported the approach to undertaking redundancy is to apply procedural realignment of a past study by incorporating taxonomy-building elements for taxonomy creation (Ahmad, Allen & Jennifer, 2022). However, a contrast finding reported in this study which only 7% are similar terms, indicates local keywords influence the keywords selection (Olson & Boll, 2001). The most common situation in subject analysis is Not Used Terms. This study found that 79 % of terms are not in use in LCSH. Researchers from various fields reported the same situation and the main reason was a distinction between expert or professional indexer knowledge and subject field knowledge.

6.0 Conclusion and Recommendation

In conclusion, the subjects used in Sports Science & Recreational Digital Collection in UiTM Local Content Hub are very minimal compared to the structured classification of subjects in the classification web. Therefore, it is important to understand information retrieval from a global perspective. Libraries must at first facilitate the information retrieval process by using and promoting a standard and uniform subject classification and at the same time empower the expert in assigning keywords and subjects. Thus, significant exposure to standard subject classification such as the classification web is to be offered to UiTM researchers and experts. Besides, the UiTM Local Content Hub may also venture into the expansion of the thesauri and taxonomy tree. This was in line with the emerging literature in recent knowledge development and creation. A standard format or template is introduced to researchers to propose new terms or subjects. At the same time, a verification element should be enforced to control the variation of terms and subjects. Thus this study recommends a centralized effort from PTAR headquarters and branches with respective faculty to coordinate the taxonomy terms assigned by the expert. UiTM Local Content Hub content is an excellent effort in gathering the multi-content of UiTM intellectual property while promoting library functions in development and creation. Future research may focus on the compassion of the subject in Sports Science & Recreational with other local content from international libraries especially neighboring countries such as Indonesia and Singapore. This will provide a better view of the taxonomy development in local in-house collections.

Acknowledgment

This article would not have been possible without the support from the management of the Perpustakaan Tun Abdul Razak, Universiti Teknologi MARA, (UiTM) Shah Alam support in both content resources and finances. A special thanks to Pusat Pengajian Sains Maklumat, Kolej Pengajian Pengkomputeran, Informatik & Media for being very supportive in providing us the opportunity and funding to complete this study under the Faculty of Information Management, UiTM, and the Research Management Institute(RMI), UiTM, to conduct this study using Dana Dalam Fakulti (DDF) 600-FPM (PT.5/1).

References

- Abukhader, S. (2019), "A taxonomy of the expected roles of librarians towards knowledge management: An eight-layer perspective for practice", *Library Management*, Vol. 40 No. 1/2, pp. 34-44. <https://doi.org/10.1108/LM-08-2017-0081>
- Ahmad F., Allen, J & Jennifer, K. (2022). *Belonging: Applying an Integrative Framework to Welcoming Communities*. 48. Metropolis Canada Conference 2022, Vancouver, British Columbia, Canada.
- Aldenderfer M, Blashfield R. (1984). *Cluster analysis*. Beverly Hills, Sage Univ Pap.
- Allyson Carlyle,(1989) "Matching LCSH and User Vocabulary in the Library Catalog," *Cataloging & Classification Quarterly*. 10, no. 1/2: 37–63.
- Arievitch, I.M. (2020). The vision of Developmental Teaching and Learning and Bloom's Taxonomy of educational objectives. *Learning, Culture and Social Interaction*, 25, 100274. <https://doi.org/10.1016/j.lcsi.2019.01.007>
- Chatterjee, S., & Samanta, M. (2021). *Conceptualizing Subject Classification*. Munich Personal RePEc Archive. <https://mpra.ub.uni-muenchen.de/109489/>
- De Pauw, K., Roelands, B., Cheung, S. S., de Geus, B., Rietjens, G., & Meeusen, R. (2013). Guidelines to classify subject groups in sport-science research. *International journal of sports physiology and performance*, 8(2), 111–122. <https://doi.org/10.1123/ijspp.8.2.111>
- Gil-Leiva Isidoro and Alonso-Arroyo Adolfo. (2007). Keywords given by expert-assigned of scientific articles in database descriptors. *Journal of the American society for information science and technology*, 58, 8, 1175–1187.
- Hedden, H. (2022). *Accidental Taxonomist*. New Jersey: Information Today.
- Keck, F., Couton, M., & Altermatt, F. (2023). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Molecular Ecology Resources*, 23, 742–755. <https://doi.org/10.1111/1755-0998.13746>
- Olson, H. O. and Boll, J. J.,(2001), *Subject Analysis in Online Catalogs*, 2nd ed. (Englewood, Colo.: Libraries Unlimited 2001): 40; *Library of Congress Subject Headings*.
- Lee, E.Y., de Lannoy, L., Li, L. et al. Play, learn, and teach outdoors - network (PLaTO-Net): terminology, taxonomy, and ontology. (2022). *International Journal of Behavioral Nutrition and Physical Activity*, 19(66). <https://doi.org/10.1186/s12966-022-01294-0>
- Merriam-Webster, Incorporated. (2019). *The Merriam-Webster Dictionary*. Springfield, Massachusetts.

- Milne, C. (2010), Developing information architecture through records management classification techniques, *Aslib Proceedings*, Vol. 62 No. 4/5, pp. 366-386. <https://doi.org/10.1108/00012531011074636>
- Naveen, D. (2018). Difference between taxonomy and classification. Difference Between. Retrieved January 26, 2023, from <https://www.differencebetween.com/difference-between-taxonomy-and-vs-classification/>
- Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336-359.
- Omar, B. & Alturki, A. (2020). An Improved Method for Taxonomy Development in Information Systems. *International Journal of Advanced Computer Science and Applications*. 11. 10.14569/IJACSA.2020.0110470.
- Poser, Mathis., Wiethof, Christina., and Bittner, Eva A. C. (2022), Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions (2022). 30th European Conference on Information Systems (ECIS 2022), May 2022, Timisoara, Romania. Research Papers. 65. https://aisel.aisnet.org/ecis2022_rp/65
- Prashasti S., Rajesh, P., Vivek, K. S., David, P., Vivek, S., and Fernando, P., (2020). Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions. *J. Intell. Fuzzy Syst.* 39, 2 (2020), 2471–2476. <https://doi.org/10.3233/JIFS-179906>
- Sabherwal, R., & King, W. R. (1995). An empirical taxonomy of the decision-making processes concerning strategic applications of information systems. *Journal of Management Information Systems*, 11(4), 177–214. <https://doi.org/10.1080/07421222.1995.11518064>
- Sheng J., Gero, Z., Ho J, C. PubMed Expert-assigned Keyword Extraction (PubMedAKE) Benchmark (2022). Proceedings of the ... ACM International Conference on Information & Knowledge Management. ACM International Conference on Information and Knowledge Management. 2022 Oct;2022:4470-4474. DOI: 10.1145/3511808.3557675. PMID: 36382341; PMCID: PMC9652778.
- Theda, Schwing., Sevim, McCutcheon., & Margaret, B. M., (2012) Uniqueness Matters: Expert-assigned-Supplied Keywords and LCSH in the Library Catalog, *Cataloging & Classification Quarterly*, 50:8, 903-928, DOI: 10.1080/01639374.2012.703164
- Usman, M., Britto, R., Börstler, J., & Mendes, E. (2017). Taxonomies in software engineering: a systematic mapping study and a revised taxonomy development method. *Information and Software Technology*, 85, 43-59.
- Valderrama-Zurián, J. C., García-Zorita, C., Marugán-Lázaro, S., & Sanz-Casado, E. (2021). Comparison of MeSH terms and KeyWords Plus terms for more accurate classification in medical research fields. A case study in cannabis research. *Information Processing and Management*, 58(5). <https://doi.org/10.1016/j.ipm.2021.102658>
- Wang, Z., Sattar Chaudhry, A. and Khoo, C.S.G. (2008), "Using classification schemes and thesauri to build an organizational taxonomy for organizing content and aiding navigation", *Journal of Documentation*, Vol. 64 No. 6, pp. 842-876. <https://doi.org/10.1108/00220410810912424>