

Labelling on Academic Library Websites

Tanja Svarre ^[0000-0002-5468-0406]

Aalborg University, 9000 Aalborg, Denmark
tanjasj@hum.aau.dk

Abstract. This paper studies labels on Danish academic library websites. Labels are one amongst several elements that can support user interaction with library websites and their related content and thus add to a reduction of the vocabulary problem. A total of 2075 labels used on the websites of 21 academic libraries with special obligations were analysed using a combination of content analysis and clustering analysis. The findings show not only large variety in the use of labels amongst the libraries but also a large concordance of labels used across library domains and purposes. A cluster analysis of the labels reveals that some libraries with similar purposes and functions also tend to be similar in their use of labels, which indicates a shared terminology within domains, sectors and purposes. The findings add to our understanding of the characteristics and variety of recent labelling across libraries in the academic library sector.

Keywords: Academic libraries, Websites, Labelling

1 Introduction

Library websites increasingly serve as the point of contact between the library and its users [1]. An academic library website represents the online portal to the many resources, both digital and analogue, that are offered to students, researchers and other users [2, 3]. Being able to locate relevant information on research library websites is crucial for students, researchers and other users when approaching the library website. Thus, the library website is the gateway to the online information and resources that are necessary when acting within the academic world [4, 5].

Labelling is one of several elements that ensure the usability of academic library websites [3]. It has previously been shown how terminology and labelling on library websites challenge usability [3, 6–8]. The Danish National Statistics Office lists 22 Danish research libraries with special obligations. Most are connected to higher education institutions, such as universities and university colleges. The remaining libraries are associated with national museums for history and various aspects of the arts, except for one, which serves as the national library of Denmark and a university library for several universities across the country. The aim of this paper is to study the use of labels on Danish academic library websites. It sheds light on the characteristics of academic library labels and on how libraries differ in their use of labels in terms of communicating with their users.

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). This volume is published and copyrighted by its editors. IRCDL 2021, February 18-19, 2021, Padua, Italy.

2 Theory

Labelling is one of four systems within information architecture outlined by Morville, Rosenfeld and Arango [9]. In web environments, labelling is used to represent underlying information and guide users towards relevant content. Labelling is considered a representation of knowledge organising systems [9]. Thus, labelling follows the characteristics of knowledge organising systems as being controlled or free [10], simple or complex [11] and narrow or broad [12].

Users can experience a variety of challenges when interacting with an information system. One of these is known as the vocabulary problem [13, 14]. The vocabulary problem addresses how the same content can be described and identified in an endless number of ways, depending on the user's viewpoint. Several empirical studies of library websites illustrate this phenomenon. For example, Dougan and Fulton [15], in their usability study of an academic library website, found issues both with the specificity of terms and general problems in understanding the website terminology. General confusion in terms of terminology was also identified in Tidal's [16] study of an academic library website. Knowledge organising systems play an important role in reducing the vocabulary problem and explicating the content that vocabulary represents, but, clearly, they should be developed with an understanding of user terminology to support users in their interaction with the websites and the content.

Morville et al. [9] distinguish between four types of labels: 1) contextual links, 2) headings, 3) navigation system choices and 4) index terms. In this paper, the focus is on navigation system elements, as the analysis is based on the labels extracted from the websites under study.

3 Methodology

The Danish National Statistics Office lists 22 libraries as research libraries with special responsibilities. We based the selection of libraries on the 2019 statistics. One library, the Danish School of Media and Journalism Library, could not be crawled by the web crawler due to technical issues. This left 21 libraries for the empirical part of this paper. The libraries with abbreviations are listed in Table 1. To ensure the most extensive versions of the websites, we used the Danish version for the current analysis. Most of the websites offer English counterparts, but they usually contain limited information compared to the Danish equivalents.

We collected the data on 3 September 2020. We used an open access web menu crawler (<http://webscompare.com/>) to crawl the selected websites. The crawler scrapes category labels using Xpath [17] on the basis of one or several web addresses entered into the interface. The output is a csv file with a flat, alphabetical list of the labels found on the websites. After the data collection, the data was cleaned and prepared for data analysis. The scraper is not capable of handling Danish special letters (Æ, Ø, Å), which are commonly used across the websites in this study, and it replaces the special letter with a blank space. The blanks were therefore manually replaced with the correct special letters.

The data analysis consisted of two elements. First, we carried out a quantitative content analysis of the web categories [18, 19], analysing per library the number of categories, the average length of categories and the average number of terms in the categories. Subsequently, we used the text analysis functionality of the analysis software NVivo (version 12 Pro) (<https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>) for analyses at the term level. Specifically, we used the word frequency functionality and analysed the 1000 most frequent terms in the dataset. We used grouping with synonyms to consider similar terms as one. The analysis was carried out using the Danish language, and the most frequent categories were subsequently translated into English for reporting in the current paper. Subsequently, we carried out a cluster analysis on the basis of word similarity to identify similar libraries.

Table 1. Included libraries

Abbreviation	Library name	Type of library
AARCH	Aarhus School of Architecture Library	UNI
ABSAL	University College Absalon Library	UC
ARBMUS	The Workers' Museum Library	OTHER
AUB	Aalborg University Library	UNI
CBS	Copenhagen Business School Library	UNI
CINEMA	Danish Film Institute Library	OTHER
DESMUS	Design Museum Denmark Library	OTHER
DIIS	Danish Institute for International Studies Library	OTHER
DKDM	The Royal Danish Academy of Music Library	UNI
DST	Statistics Denmark Library	OTHER
DTU	Technical University of Denmark Library	UNI
FB	Royal Danish Defence Academy Library	OTHER
KADK	The Royal Danish Academy of Fine Arts Library	UNI
KB	The Danish Royal Library	UNI/OTHER
PHB	Copenhagen University College Library	UC
POLAR	Polar Library	UNI
SDUB	University of Southern Denmark Library	UNI
UCLB	UCL University College Library	UC
UCNB	UCN University College Library	UC
UCSB	University College South Denmark Library	UC
VIAB	VIA University College Library	UC

Legend: The libraries are divided into three categories by purpose: UNI (serving a university or university-like institution), UC (serving a university college) and OTHER (serving other types of institutions like museums or organisations).

4 Results

The libraries included for analysis serve different purposes (Table 1). Six libraries qualify as university college libraries, eight are located at universities or university-like institutions, and six libraries serve different organisations or museums. The Royal Library has a role both as the national library of Denmark and the university library for several Danish universities, resulting in two labels in the table: university library and other.

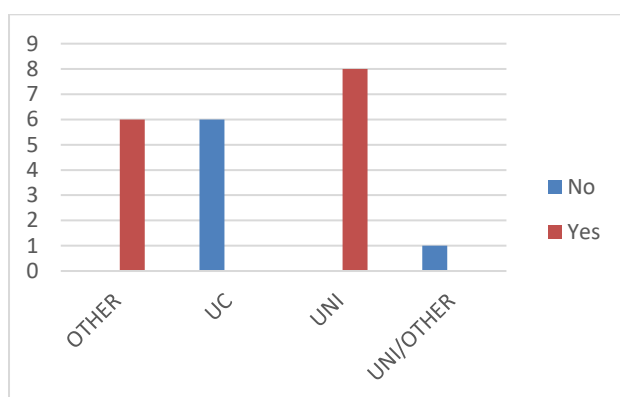


Figure 1. Number of libraries incorporated in the host institution website
Yes=incorporated; No=not incorporated

In many of the library cases, their organisational relations appear from the placement of the library website in the web structure. All the UNI and OTHER libraries (14 of the 21 libraries) are incorporated into their host institution websites. The remaining seven libraries, meaning all the UC libraries and the Danish Royal Library, have independent websites with no institutional connection to their host organisation (Figure 1).

4.1 Term Distribution

Crawling the 21 research libraries, we found the distribution of categories that appears in Table 2. The table shows a large variation in the number of labels on the library websites and an average of 98.76 labels per website. A standard deviation of 84.97 illustrates the large variation between the libraries.

Table 2. Number of labels, average length of labels and average number of label terms

	Mean	Minimum	Maximum	Standard deviation
Number of labels	98.76	29	378	84.97
Length of labels (characters)	16.66	2	43	8.57
Number of terms	2.25	1	9	1.37

We used the independent sample T-test to test if the variation was related to whether the library website is incorporated into the host organisation’s website or has its own website. With no significant difference identified for the number of labels, the average length of labels or the average number of terms in labels, this does not appear to be the case.

4.2 Term Frequency

We used the term frequency functionality in NVivo to identify high-frequency terms in the data set. The results of the analysis appear in Figure 2. “Search” and variations of “Library” are the most frequent terms along with other library-related terms like “Archive”, “Books”, “Materials” and “Journals”. Another category of high-frequency terms relates to the library as a service function. This category is exemplified by terms like “Way” (representing street names and wayfinding in Danish), “Contact”, “Booking”, and “Opening”, inviting users to use the library and library services like assistance from a trained librarian. The organisational attachment of many of the libraries is also reflected in the most frequent terms in labels. Here we see the university abbreviations (“SDU”, “DTU” and “CBS”) and “Research”, “Education” and “Student”.

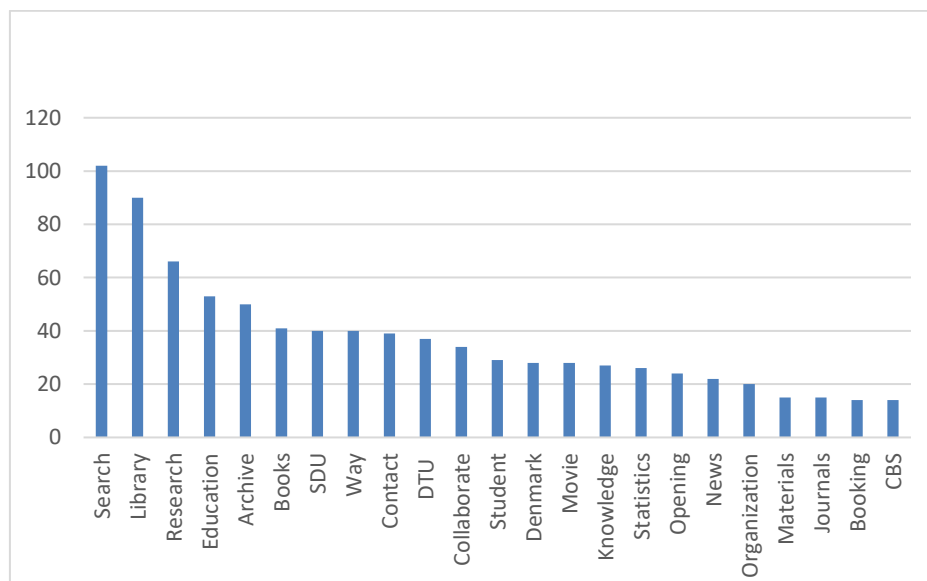


Figure 2. Most frequent terms in labels

Furthermore, we analysed the number of libraries in which the high-frequency terms occurred (Figure 3). This figure, to some extent, changes the impression of the most frequent terms. As would be expected, the institution-specific terms only appear in the related libraries, whereas general library-specific terms are used more generally across the included libraries, with “Library”, “Search” and “Contact” as the most used terms.

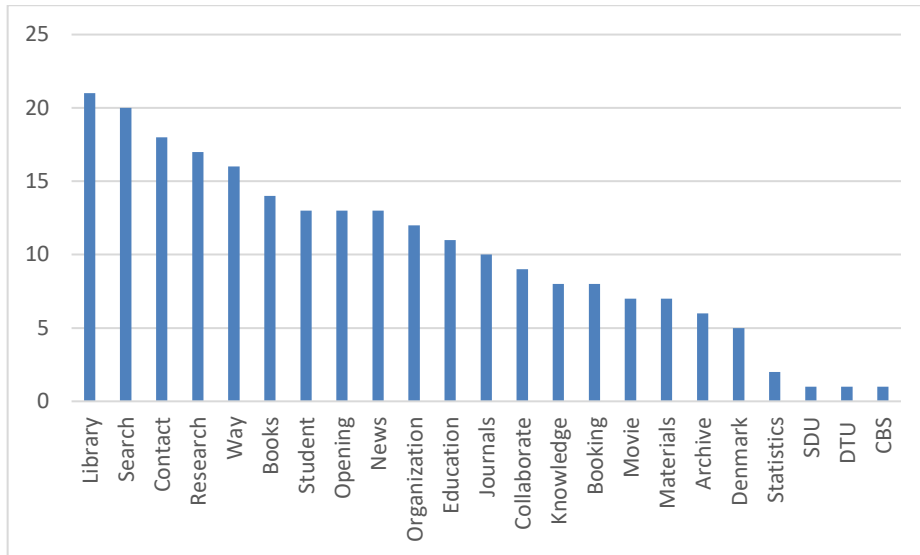


Figure 3. Number of libraries using the most frequent terms in labels

4.3 Cluster Analysis

We used the clustering functionality in NVivo to analyse the similarity between the selected libraries. The results of the analysis appear in Figure 4, which illustrates the two main clusters that evolved from the analysis.

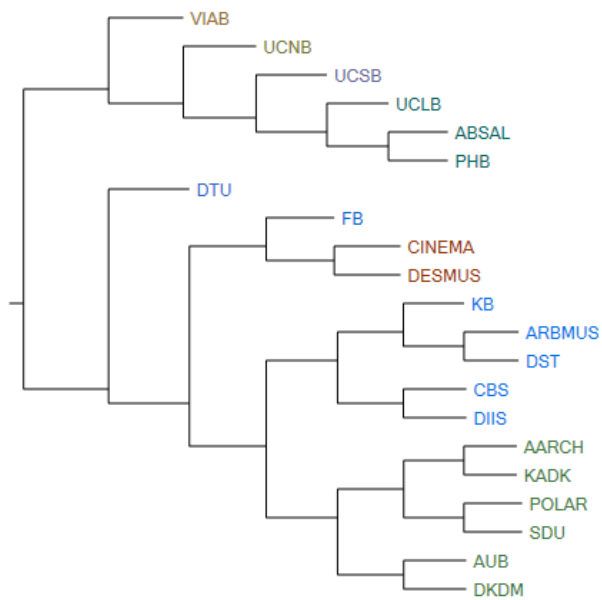


Figure 4. Cluster analysis based on word frequency

The upper cluster consists of all the university college libraries in the population. It is interesting to identify how the purpose and the target group of the library actually influences the choice of terminology at these libraries. The lower and larger cluster consists of a combination of university libraries and the category “other”. Here, the picture is a bit more muddled than the upper cluster of the figure, but still some observations can be made. For instance, the Danish Film Institute (CINEMA) and the Design Museum (DESMUS) libraries are so similar that they end up in the same cluster. If the next level of the cluster is considered, they are also connected to another museum, the Royal Danish Defence Academy Library (FB). Likewise, the Aarhus School of Architecture Library (AARCH) and the Royal Danish Academy of Fine Arts Library (KADK), which amongst others are connected to architecture education in Copenhagen, are also so similar that they share a cluster in the figure.

5 Discussion and Concluding Remarks

Our analysis shows that the academic libraries with special obligations in Denmark serve various institutions and purposes. They represent a large variation in their use of labels, both in numbers and variety. Previous research has identified challenges with library jargon on academic library websites [e.g. 16]. Considering the most frequent labels used in the current study, the deliberate use of terminology does not appear to be prevalent. However, users should be involved in further studies to obtain a deeper understanding of this issue.

The cluster analysis revealed that academic libraries with similar purposes also tend to be similar in their use of labelling. Independent sample T-tests did not reveal that this can be explained by whether the libraries are incorporated into their host institutions’ websites and thereby have institution labels as part of their pool of labels. Instead, it seems that the specific use of labels is similar, for instance, between university college libraries, between some museum libraries and between the two Danish schools of architecture. The findings indicate that the libraries in their labelling draw on a shared terminology within their domains, which is a step towards reducing the vocabulary problem in information interaction. Further studies with users within the specific domains can further elaborate on how they understand and experience the vocabulary problem within their domains.

References

1. Guay, S., Rudin, L., Reynolds, S.: Testing, testing: a usability case study at University of Toronto Scarborough Library. *Library Management*. 40, 88–97 (2019). <https://doi.org/10.1108/LM-10-2017-0107>.
2. Million, A.J.: Help Needed: Best Practices, Collaborative Advantage, and Library Websites. *International Information & Library review*. 50, 312–318 (2018). <https://doi.org/10.1080/10572317.2018.1526851>.

3. Silvis, I.M., Bothma, T.J.D., de Beer, K.J.W.: Evaluating the usability of the information architecture of academic library websites. *Library Hi Tech.* 37, 566–590 (2019). <https://doi.org/10.1108/LHT-07-2017-0151>.
4. Kim, Y.-M.: Users' perceptions of university library websites: A unifying view. *Library & Information Science Research.* 33, 63–72 (2011).
5. Hu, C.-P., Hu, Y., Yan, W.: An empirical study of factors influencing user perception of university digital libraries in China. *Library & Information Science Research.* 36, 225–233 (2014).
6. Pant, A.: Usability evaluation of an academic library website: Experience with the Central Science Library, University of Delhi. *The Electronic Library.* 33, 896–915 (2015). <https://doi.org/10.1108/EL-04-2014-0067>.
7. Gillis, R.: "Watch Your Language!": Word Choice in Library Website Usability. *Partnership : the Canadian Journal of Library and Information Practice and Research.* 12, (2017). <https://doi.org/10.21083/partnership.v12i1.3918>.
8. Kous, K., Pušnik, M., Heričko, M., Polančič, G.: Usability evaluation of a library website with different end user groups. *Journal of Librarianship and Information Science.* 52, 75–90 (2020). <https://doi.org/10.1177/0961000618773133>.
9. Louis Rosenfeld, Peter Morville, Jorge Arango: *Information architecture, For the Web and Beyond.* O'reilly Media, Inc., Sebastopol (2015).
10. Dubois, C.P.R.: Free text versus controlled vocabulary. *Online Review.* 11, 243–253 (1987).
11. Zeng, M.: Knowledge organization systems (KOS). *Knowledge Organization.* 35, 160–182 (2008).
12. Soergel, D.: Indexing and retrieval performance: The logical evidence. *Journal of the American Society for Information Science.* 45, 589–599 (1994). [https://doi.org/10.1002/\(SICI\)1097-4571\(199409\)45:8<589::AID-ASI14>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-4571(199409)45:8<589::AID-ASI14>3.0.CO;2-E).
13. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The Vocabulary Problem in Human-system Communication. *Commun. ACM.* 30, 964–971 (1987). <https://doi.org/10.1145/32206.32212>.
14. Hearst, M.: *Search User Interfaces.* Cambridge University Press, Cambridge (2009).
15. Dougan, K., Fulton, C.: Side by Side: What a Comparative Usability Study Told Us About a Web Site Redesign. *Journal of Web Librarianship.* 3, 217–237 (2009). <https://doi.org/10.1080/19322900903113407>.
16. Tidal, J.: Creating a user-centered library homepage: a case study. *OCLC Systems & Services: International digital library perspectives.* 28, 90–100 (2012). <https://doi.org/10.1108/10650751211236631>.
17. Luthfiyanto, A., Kusumo, D.S.: Extraction of Website Navigation Label Using A Multiple Web Crawler: A Case Study on 14 University Websites in Indonesia. In: *2020 International Conference on Data Science and Its Applications (ICoDSA).* pp. 1–7. IEEE, New York (2020).
18. Krippendorff, K.: *Content Analysis: An Introduction to Its Methodology.* SAGE, Thousand Oaks (2004).
19. White, M.D., Marsh, E.E.: Content Analysis: A Flexible Methodology. *Library Trends.* 55, 22–45 (2006).