

# Subject Indexing: The Challenge of LGBTQI Literature

Jenny Bergenmar<sup>1</sup>[0000-0002-6761-3544] and Koraljka Golub<sup>2</sup>[0000-0003-4167-4777]

<sup>1</sup> Department of Literature, History of Ideas and Religion, University of Gothenburg, Sweden

<sup>2</sup> Department of Cultural Sciences, Linnæus University, Sweden

firstname.lastname@lir.gu.se

firstname.lastname@lnu.se

**Abstract.** Despite a growing number of digital LGBTQI (lesbian, gay, bisexual, transsexual, queer, intersex) history archives, and research-driven digital LGBTQI initiatives, queer perspectives have not been prominent in the digital humanities. Furthermore, investigations of LGBTQI in literary history is hampered by the fact that, to date, there are no broad scholarly inventories of such literature. Due to the absence of exhaustive bibliographies, scholars need to perform time-consuming, human reading of individual works and imprecise searches in order to locate LGBTQI motifs and themes. Research on subject indexing has also revealed that controlled vocabularies in use are too general to describe LGBTQI themes, motifs, and characters in a relevant manner. The purpose of this paper is to discuss how LGBTQI literature can be made more searchable, and more visible through the development of a quality-controlled subject specific database (QUEERLIT database) in which specialized subject indexing is applied. Methodological challenges pertaining to indexing of queer literary texts with implicit LGBTQI motifs are discussed, as well as theoretical considerations raised when assigning certain contemporary subjects to historical texts.

**Keywords:** LGBTQI, Queer, Subject Indexing, Literature, Digital Humanities.

## 1 Introduction

While gender has certainly been an important topic for many projects employing computational methods on large text corpuses, sexuality has not been investigated to the same extent. This is most likely due to methodological challenges pertaining to representation of non-normative sexuality: to perform computational analysis on a phenomenon that is often not manifestly represented is problematic. As Hyttinen phrases it in a discussion on queer readings of early 20th century prose: “There is no formal or morphological feature that could be claimed as always signaling the queer” [17]. Moreover, ‘queer’ is often presented as a mode of reading, rather than an aspect of the text. While machine reading of large text corpuses might be inadequate for the reasons given by Hyttinen, developing subject indexing for LGBTQI texts is a more promising method. This paper discusses theoretical and methodological challenges connected to subject

indexing of LGBTQI literature (poetry, prose, drama) within a digital humanities framework, departing from the project QUEERLIT.

The purpose of the project is to develop methods for identification and indexing of Swedish LGBTQI literature, in order to enable further research within LGBTQI studies. A quality-controlled subject-specific bibliographic database is planned, which would enable research on the distribution and development of LGBTQI motifs and themes through time and within different genres. It would facilitate studies of such motifs or themes in different periods and could show larger patterns, for example regarding specific concepts, metaphors, and images used to describe LGBTQI, as well as their presence in, or exclusion from, the canon. The database could thus be used to trace the different, and changing understandings of LGBTQI over a long time span.

Better methods for identifying and indexing LGBTQI literature are important for several reasons. New research in comparative literature, gender and queer studies is currently impeded by the lack of methods for systematic identification of LGBTQI themes, motifs and characters. Internationally, there are a number of LGBTQI archives, but they usually focus on the LGBTQI community and movement, and normally do not include fictional texts to any large extent. Due to the absence of exhaustive bibliographies, researchers are forced to time-consuming, human reading of individual works and imprecise searches in order to locate LGBTQI motifs and themes. This constitutes a significant loss of information on ideas surrounding LGBTQI lives in history. Fictional works provide an insight into opinions and imagination of contemporaries which cannot be captured from factual records. Popular culture in general affects how LGBTQI is perceived by the general public [18]. Literatures, in particular novels, also function as influential opinion formers with the potential to generate perceptions that span over generations [3]. Disfranchised groups, like LGBTQI people in history, often leave a lack of historical sources behind, especially narrations told from their own perspective [13]. From this viewpoint, fiction is a useful source for understanding of LGBTQI experiences, and also for identifying forms of prejudice, oppression, and violence (both symbolic and actual) that have affected this group.

The majority of related work on LGBTQI consists of projects relating to digitization of material in existing collections, or creating collections of new digital material. Different projects within the framework “LGBTQ Oral History Digital Collaboratory” in Canada are examples [4]. Explorations of the history and philosophy of technology from a queer perspective have also been made, for example by Keeling [20], Tsika [33], and Fancher [11]. If the boundaries of digital humanities as an academic field are drawn more narrowly, using the definition of digital humanities as employing data-intensive methods, LGBTQI is less likely to be included. This problem is connected to scale: how to quantify what is so illusively described in literary language or even left unsaid? Even though Mandell has argued, in the case of gender analysis, that “computation enables complexity” if modelled adequately [25], a queer text mining will not be attempted. In the QUEERLIT project LGBTQI literature is identified by way of previously collected data instead. This paper consists of following parts: a short presentation of the QUEERLIT project, a discussion of the possibilities and challenges connected to subject indexing of LGBTQI literature, and a theoretical reflection on the identification of LGBTQI texts.

## 2 The QUEERLIT Project

The project has four specific aims:

- 1) To develop a thesaurus for indexing LGBTQI literature and to map it to an existing LGBTQI thesaurus in English (Homosaurus), in collaboration with KvinnSam (Swedish National Resource Library for Gender Studies and university-wide research infrastructure at Gothenburg University) and LIBRIS (The Swedish National Union Catalogue).
- 2) To identify LGBTQI literature in collaboration with an advisory board consisting of experts in Swedish LGBTQI literature and a network of librarians.
- 3) To construct a subdatabase in LIBRIS containing bibliographic records of LGBTQI literature, in collaboration with the Swedish National Library and KvinnSam.
- 4) To make this database available to the general public through a separate interface allowing for more specialized searches than LIBRIS does, and to link the records to open datasets such as DBpedia (extracting structured content from various Wikimedia projects) and LibraryThing (a web community where books are cataloged and described by users) to retrieve content information about the titles or authors.

The project will develop a thesaurus for indexing LGBTQI literature, based on specialized subject headings developed by KvinnSam for KVINNSAM, a subdatabase in LIBRIS, which contains bibliographic references on gender research and women's history. Compared to a subject headings system, a thesaurus allows for better coordination of terms both in hierarchical and non-hierarchical relationships and creation of thesauri is a standardized process by ISO. The thesaurus will be used for indexing in the QUEERLIT subdatabase, but will also be available for institutions and international scholars working with similar material. As far as possible, the created thesaurus will also be mapped to Homosaurus and, if possible, to Swedish Subject Headings (SAO), in order to facilitate interoperability and access through other library catalogs in Sweden and abroad.

## 3 LGBTQ Perspectives on Classification and Indexing

Research on controlled vocabularies such as classification systems, subject headings and thesauri has revealed problems in relation to representation. Insufficient and misleading subject headings may not only make it hard to identify LGBTQI literature, but also work formatively. In her review of the changes in subject headings describing homosexuality, Adler argues that the practice to classify books with queer characters and storylines with the heading "sexual perversion" in libraries before the 1970s in itself had a pathologizing effect [1]. While the later disconnection of "homosexuality" from the larger category of "sexual perversion" was an important development, this term was still not representative of the everyday language use, where "gay" would be more accurate and less tainted with medical associations. Adler's overview also provides insights in the complicated processes of updating a subject headings system, requiring resources for re-indexing of any collection using the system, which is a known problem

and a common reason for slow updating of controlled vocabularies. Even today, research has shown that subject indexing and classification systems in use are inadequate for describing LGBTQI or other minority groups, or even women [e.g. 2, 5, 6, 8, 28, 32]. Golub concludes that “searching by ‘LGBTQ’ in Swedish Subject Headings will retrieve no results as the term does not exist, neither is it cross referenced from ‘HBTQ’, which exists” [14]. The findings in this study provides a strong argument for the development of a specialized and linked thesaurus, as proposed in the project.

The question of how we find the things that are left unsaid is highly relevant in relation to digital methods for text analysis. In spite of the recent confidence in the methodological possibilities of computational methods, certain areas are still left unexplored. Large amounts of digitized texts make text mining and automatic subject indexing and classification possible, something which has proven useful not least when it comes to older literature. However, LGBTQI themes, motifs and characters are often impossible to find through full text searches, and even through text mining. This is particularly true for literature written before the late 20th century. In older literature, LGBTQI motifs are not only described through a different nomenclature than that of today, but such themes are also often implicit. Due to these issues, automatic subject indexing and classification is not a viable option for LGBTQI literature. Furthermore, challenges of evaluating the usefulness of automatic subject indexing and classification solutions in operative information systems remain unaddressed [15].

From the perspective of queer studies, a common point of critique regarding subject indexing is the seemingly neutral and objective status of the library record. It is not possible to “contain entire fields of knowledge or ways of being in accordance with universalizing systems and structures,” as Drabinski puts it [9]. Queer theory is thus useful for a critical reading of library catalogs and indexing systems, but how it might translate to a methodology is more unclear [23]. Drabinski suggests that since correction of subject headings will always be contingent and never final, one queer methodology in this case might be to invite users to read the catalogs and subject headings critically and to propose terms more corresponding to vernacular uses [9]. Olson has argued that technical solutions, as for example search interfaces that make visible related and broader terms, can produce more situated and contextualized results [29].

In terms of this project’s methodology, it means that existing subjects headings (in particular SAO) will be critically read with queer theory as a framework. Moreover, the project will make use of social tagging and the actual use of existing search terms in the determination of which subjects and terms should be included in the thesaurus. Olson [27] suggests that transaction logs from online catalogs might be analyzed to understand the terminology that is actually in use.

The thesaurus will be developed based on existing LGBTQI thesauri and subject headings systems. The most recent is the thesaurus of the International Gay and Lesbian Information Centre and Archive (IHLIA) in Amsterdam [19]. That thesaurus has been translated and developed into a linked data vocabulary ([homosaurus.org](http://homosaurus.org)) and is currently used both by IHLIA and The Digital Transgender Archive. QRAB, the Archives and Library of the Queer Movement in Sweden, is in the process of translating it to Swedish for the indexing the material in their archive. The Homosaurus will also function as a groundwork for Queerlit and the project will collaborate with QRAB in the translation of it. It will however also need further adaptation, since both IHLIA and

QRAB have a broader range of archival material, while Queerlit will be focused exclusively on literature, for the most part published in book format and to some extent in press and periodicals.

#### 4 Queer Theory and the Identification of LGBTQ Literature

Speaking of LGBTQ history is inevitably anachronistic. Although this project will make use of the term LGBTQ, the project is also invested in historical understandings. Previous research indicates that readings of literature both before and after the advent of a particular term can sometimes enlighten both the term itself and those groups that are addressed by it. Such example is given by Prosser's seminal study about trans literature, which argues that there was a transsexual narrative even before medical science defined the specific diagnosis of transsexualism in the 1950s [30]. This is not to say that the concepts 'queer' or 'LGBTQ' are historically transferable, but that they can fill strategic functions and enable discussions and investigations of historical material. Although themes, motifs, and characters represented in the material are very different and, furthermore, a product of a reading process that can never be neutral, there are tactical gains in bringing the material together under a common terminology with relevance for the present. By renaming these historical representations with an umbrella concept of today, it becomes possible to investigate how they relate to each other and to study the overall image that these representations provide [7].

Love [22] also points out that our need for recognition is inevitably controlled by our present-day identities and that those identifications will steer our understandings of the past no matter the anachronisms of that understanding. Holmqvist [16] draws on works by US-based trans studies scholars like Stryker [31] who have adopted understandings of trans as a movement and use it as a verb. By understanding LGBTQ as something which is done rather than something that one is (being lesbian/gay/bi/trans/intersex) it is possible to understand a wide array of motifs as within the LGBTQ spectrum. Rather than trying to define what characters may be described 'as' LGBTQ, the project will encompass themes and motifs which may be understood as doing gender transgressions and/or same-sexual activity. This allows for an understanding of LGBTQ that uses modern day terminology in order to investigate both the past and the present.

Although this can be seen as an important methodological choice, it does not solve the problem of how to identify which texts should be included in the QUEERLIT database. Even though it is to some extent possible to search by keywords and subject headings, the results produced in this way will be far from adequate. In the absence of text mining methods that are flexible and fine grained, enough to capture the changing representations of queer characters and topics, traditional humanistic methods are employed, that is, use of subject-specific bibliographies [24, 26], and the data previously collected by researchers in the field. Usually those kind of data are kept privately on hard-drives of individual researchers, but by making them part of an advisory board, this data can help to identify texts to be included in the database, and at the same time share research data that have not been made public due to lack of infrastructure.

## 5 Conclusion

The challenges pertaining to finding fictional texts with LGBTQI themes, motifs, or characters have much in common with the attempts to recover the history of other marginalized groups. As Earhart has pointed out, many of digital projects promoting non-canonical, minority perspectives on literature have been small, grassroots initiatives [10]. There is also a need to recover not only that which has not been included in the canon, but also early digital projects carried out by activists and individual researchers [12]. In the case of LGBTQI literature, Magnusson's previously mentioned LGBTQ database from 2008 [26] is one example of such work in need of a sustainable solution. Due to the situation that LGBTQI heritage to a large extent has been excluded by archives and digitization projects, it is no surprise that many of the efforts to bring LGBTQI into digital humanities consist of archiving, collecting and making material available in different databases and digital archives. However, these are seldom devoted to literary heritage.

When undertaking a project to make literary LGBTQI heritage searchable, there are some methodological issues to address: the first concerns criteria for inclusion, the second naming them with subjects that are understandable today, even though they are not historically transferrable. There is an epistemological conflict between queer theory and the knowledge construction implied by subject indexing. Gay and lesbian studies emerged from a recognition that those identities were lacking in the writing of history and scholars in this field had an ambition to recover those histories. In the area of library indexing, this typically entailed a correction of misleading subject headings and an inclusion of more words corresponding to actual experiences.

From the perspective of queer theory, on the other hand, such a correction would necessarily produce new universalizing terms. Framed in this manner, the difference between gay and lesbian studies and queer studies is essentially a conflict between a modernist approach, striving towards an expansion of knowledge using rational and scientific methods, and a postmodernist approach, underlining the fluidity of identity categories and engaging in the deconstruction of unified stories about subjects [21]. There is an interesting paradox at work here: while there certainly might be risks of social reproduction in the subject indexing, with the inadequately developed subject indexing of LGBTQI, the representation of queer identities in history will remain an area inaccessible for research. QUEERLIT will seek to combine elements of these two approaches: while a new, specialized subject indexing system will be developed, older subject headings will be read critically from a queer perspective.

## References

1. Adler, M. "Let's not homosexualize the Library Stacks." *Liberating Gays in the Library Catalog*. *Journal of the History of Sexuality*, 24(3), 478–507 (2015).
2. Bates, J & Rowley, J.: Social reproduction and exclusion in subject indexing: A comparison of public library OPACs and LibraryThing folksonomy. *Journal of Documentation*, 67(3), 421–448 (2011).
3. Björklund, J.: *Lesbianism in Swedish Literature: An Ambiguous Affair*. New York: Palgrave Macmillan (2014).

4. Brown, E. H.: LGBTQ + Digital Humanities, <https://www.elspethbrown.org/lgbtq-digital-humanities>, last accessed 2019/09/24.
5. Campbell, D. G.: Queer theory and the creation of contextual subject access tools for gay and lesbian communities. *Knowledge Organization*, 27(3), 122–31 (2001).
6. Campbell, D. G.: A Queer Eye for the Faceted Guy: How a Universal Classification Principle can be Applied to a Distinct Subculture. In: *Knowledge Organization and the Global Information Society: Proceedings of the 8th International Conference of the International Society for Knowledge Organization*. London, U.K., 13–16 July 2004, pp. 109–113. Wurtzburg: Ergon Verlag (2004).
7. Chess, S.: *Male-to-Female Crossdressing in Early Modern English Literature: Gender, Performance, and Queer Relations*, New York: Routledge (2016).
8. Christensen, B.: Interfiling intersex: How Dewey classifies intersex in theory and practice.” In: Greenblatt, E.: *Serving LGBTIQ library and archives users: essays on outreach, service, collections and access*, pp. 201–211, Jefferson, North Carolina (2011).
9. Drabinski, E.: Queering the Catalog. *Queer Theory and the Politics of Correction*. *Library Quarterly: Information, Community, Policy*, 83(2), 94–111 (2013).
10. Earhart, A. Can Information Be Unfettered? Race and the New Digital Humanities Canon, In: Gold, M. K.: *Debates in the Digital Humanities*, University of Minnesota Press, Minneapolis (2013), <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfd1e/section/cf0af04d-73e3-4738-98d9-74c1ae3534e5>, last accessed 2019/09/24.
11. Fancher, P. Embodying Turing’s Machine. *Queer embodied rhetorics in the history of digital computation*. *Rhetoric Review*, 37(1), 90–104 (2019) DOI: 10.1080/07350198.2018.1395268
12. Gallon, K.: “Black Digital Humanities”. In Gold, M. K. & L. F Klein: *Debates in the Digital Humanities*. Minneapolis and London: University of Minnesota Press (2016), [dhdebates.gc.cuny.edu/debates/text/55](https://dhdebates.gc.cuny.edu/debates/text/55), last accessed 2019/09/24.
13. García, M.T.: *Literature as History. Autobiography, Testimonio, and the Novel in the Chicano and Latino Experience*. Tucson: University of Arizona Press (2018).
14. Golub, K.: Potential and Challenges of Subject Access in Libraries Today on the Example of Swedish Libraries. *International Information & Library Review*, 48(3), 204–210 (2016)
15. Golub, K.: Automatic Subject Indexing of Text. In: *Encyclopedia of Knowledge Organization* (2017), <http://www.isko.org/cyclo/automatic>.
16. Holmqvist, S.: *Transformationer: 1800-talets svenska translitteratur genom Lasse-Maja, C.J.L. Almqvist och Aurora Ljungstedt*. Diss. Uppsala University, Gothenburg & Stockholm: Makadam (2017).
17. Hyttinen, E. Spectacles of Silence: The Queer in 1910’s Finnish Literature, *lambda nordica*, 21(3–4), 35–55 (2016).
18. Inness, S.A.: *The Lesbian Menace: Ideology, Identity, and the Representations of Lesbian Life*. Amherst: University of Massachusetts Press (1997).
19. Johnson, M.: A Hidden History of Queer Subject Access. In: Roberto, K.R.: *Radical Cataloging. Essays as the Front*, pp. 18–27, Jefferson & London: McFarland & Co (2008).
20. Keeling, K., *Queer OS*. *Cinema Journal* 53(2), 152–57 (2014).
21. Lovaas, K. E., Eila, J.P, and Yep, G.A.: Shifting ground(s). Surveying the contested terrain of LGBTQ studies and queer theory. *Journal of homosexuality* 52(1–2), 1–18 (2010).
22. Love, H.: *Feeling Backward. Loss and the politics of queer history*, Cambridge Massachusetts: Harvard UP (2007).
23. Love, H.: Queer Messes. *Women’s Studies Quarterly*, 44(3–4), 345–349 (2016).
24. Lönnlöf, S.: *Böcker bortom normen*, Lund: BTJ (2014).

25. Mandell, L.: Gender and cultural analytics. In: Gold, M.K & Klein, L.F.: Debates in the Digital Humanities 2019, <https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/5d9c1b63-7b60-42dd-8cda-bde837f638f4#ch01>, last accessed 2019/09/24.
26. Magnusson, J.: Bibliografi GLBTQ-litteratur 1800–1900, <http://www.janmagnusson.se/bibliografay1.html>, last accessed 2019/09/24.
27. Olson, H.: The Power to Name. Representation in Library Catalogs. *Signs*, 26(3), 639–668 (2001).
28. Olson, H. A. The Power to Name: Locating the Subject Representation in Libraries. Dordrecht; Boston, Mass.: Kluwer (2002).
29. Olson, H. A.: How we Construct Subjects. A Feminist Analysis, *Library Trends* 56(2), 509–541 (2007).
30. Prosser, J.: *Second Skins: The Body Narratives of Transsexuality*, New York: Columbia University Press (1998).
31. Stryker, S.: *Transgender History*, Berkeley, CA: Seal Press (2008).
32. De la Tierra, T.: Latina Lesbian Subject Headings. In: Roberto K.R.: *Radical Cataloguing. Essays at the Front*. Jefferson, pp. 94–102, N.C. & London: McFarlane (2008).
33. Tsika, N.: Compu Queer. Protocological Constraints, Algorithmic Steamlining, and the Search for Queer Methods Online. In: *WSQ: Women’s Studies Quarterly* 44(3–4), 111–130 (2016).