

Adapting Homosaurus for LGBTQI Fiction in Sweden

Koraljka Golub, Linnaeus University
Siska Humlesjö, University of Gothenburg
Olov Kriström, University of Gothenburg

While full-text information retrieval may work well for certain search tasks in some contexts, it does little to support subject searching of fiction. Looking for themes or topics of fiction is challenging in systems relying exclusively on full-text retrieval methods because works of fiction are characterized by language that is often on purpose metaphorical and themes are hidden in between the lines. The problem is further exacerbated in LGBTQI fiction (lesbian, gay, bisexual, transgender and intersex) where LGBTQI themes are not only subtly present but they can also be expressed through contested historical language.

On the other hand, searching for fiction with controlled vocabularies entails different problems. Historically, fiction has been poorly indexed because subject headings or thesauri as well as indexing guidelines have oftentimes used nonfiction as a model. Even when works of fiction have been indexed, the catalogers have relied on terminology not always suited to meet the needs of those looking for LGBTQI fiction, whether it has to do with outdated terms or lack of specificity. As an illustration of the problem with lack of specificity, SAO (Svenska Ämnesord, the major controlled vocabulary in Swedish libraries) lists only one narrower term to the subject heading "Transpersoner" ("Transgender people") - "Judiska transpersoner" ("Jewish transgender people").

In an attempt to address the challenges that LGBTQI literary fiction is in general very sparingly thematically described and that SAO LGBTQI terms are broad and lacking in acuity, the Swedish Queerlit project (<https://www.gu.se/en/research/queerlit-data-base>), 2021-2023, aims to create a sub-database of the Swedish union catalog Libris (<https://libris.kb.se>) with Swedish LGBTQI fiction, and create a subject-specific thesaurus for indexing the literature. In addition to the joint LIBRIS interface, the Queerlit sub-database is to be searchable through a separate one supporting advanced search functionalities based on the dedicated LGBTQI thesaurus also developed as part of the project.

The thesaurus is largely based on the English-language Homosaurus Vocabulary (<https://homosaurus.org/v3>) used by the Digital Transgender Archive (<https://www.digitaltransgenderarchive.net>) (among others), maintained by an independent editorial board. The Homosaurus was first published, in English and Dutch, in 1997 as a standalone thesaurus for use at the International Homo/Lesbian Information Center and Archive (now IHLIA LGBTI Heritage) in Amsterdam, The Netherlands (Zwaaf 2020). Revisions in the

2010's aimed at removing general, non-LGBTQ related terms, making the thesaurus more inclusive, primarily regarding queer and trans* subjects, and restructuring the thesaurus as linked data (Cifor and Rawson 2022).

As a thesaurus, the Homosaurus is developed to cover a range of varied information resources. Thus, apart from translating terms to Swedish and adjusting them to reflect the Scandinavian context, the Queerlit thesaurus also adapts the Homosaurus specifically to fiction. At the start of the Queerlit project, the Homosaurus comprised 1186 terms. Of these, first selected were those that were applied to works of fiction in the catalogs of IHLIA LGBTI Heritage and the Digital Transgender Archive. This selection resulted in 755 terms in the Queerlit thesaurus with exact or close matches in the Homosaurus.

Additional 67 terms without corresponding terms in Homosaurus were added. The move of refocusing from literary warrants in a specific collection, to prioritizing an inclusive structure of potential applicability was carried out in the development of the Queerlit thesaurus, not least by including intersex variants throughout for all the terms already represented with lesbian, gay, bi, trans and queer variants in the Homosaurus. In addition terms adjusted to the Swedish context, such as adjusted legal terms and terms for the indigenous Sami people were added and terms to reflect themes in fiction, such as characters whose gender isn't known to the reader and recurring symbolic themes e.g. mirrors and rainbows.

In comparison to SAO, the resulting Queerlit thesaurus gives indexers a much richer vocabulary to draw on when describing the literature. Returning to the example of SAO's term "Transpersoner", with its one narrower term, the Queerlit thesaurus gives 22 options of narrower terms for the same subject heading. A further addition in comparison to the Homosaurus was the inclusion in the Queerlit thesaurus of scope notes for all terms, giving definitions of terms and providing use guidelines for the indexers as well as searchers using those terms.

The thesaurus and the search functionalities are further informed by needs of potential users, including subject experts, librarians as well as the general public. So far completed is the first survey in which end-user needs were investigated via a web-based questionnaire comprising 40 closed and open questions of which 15 required. The survey was sent out to relevant mailing lists and advertised in associated social media channels in the last week of November 2021. On the survey closing date, 20 December 2021, 108 responses were received.

The survey showed that they are avid readers who tend to find information about what to read via social media (76.2%) as well as via their friends and acquaintances (72.3%); while for a copy of the book most turn to a library (71.3%) or a bookstore (68.3%). Their satisfaction when searching online is for the majority neither good nor bad (65.3%), and open replies demonstrate that online searching is suboptimal, listing problems mentioned earlier: relying on automatic searching or general subject terms often does not work for LGBTQI fiction.

When presented with options for improved searching, most respondents (88 or 87.1%) prioritized using related terms for broader search in order to get more results (higher recall).

Many at the same time (71 or 70.3%) also wished for search functionalities that would support them in retrieving fewer and more specific results based on narrower terms when too many results are retrieved (higher precision). More than half (60 or 59.4%) would also appreciate more user-friendly interfaces with the option to browse by hierarchically arranged subjects (as you would browse library shelves in a physical library), a feature that is lacking in many online search services today. Similarly, about half of the respondents (48 or 47.5%) would also appreciate the option for word sense disambiguation that would allow getting only those results related to the specific meaning of the search term they are looking for. All these results point to the need to implement these features in the Queerlit search interface which is currently under development.

References

Marika Cifor & K.J. Rawson. (2022). Mediating Queer and Trans Pasts: The Homosaurus as Queer Information Activism, *Information, Communication & Society*, DOI: 10.1080/1369118X.2022.2072753

Katrina Zwaaf. (2020). The Homosaurus <http://homosaurus.org>. *Technical Services Quarterly*, 37:2, 207-208, DOI: 10.1080/07317131.2020.1728140