

Critical about clustering of tags: An intersectional perspective on folksonomies

Folksonomies and social tagging systems have been proposed as a possible solution to counter the marginalising tendencies of large universal classification systems. Deodato [4] suggests that tagging provides an opportunity for librarians to release some of the traditional control and empower users, marginalised or not, to create their own structures of organising knowledge. Relating to the perspective of Spivak, Olson and Fox [9] associate self-representation with "such user-centred technologies as social tagging [that] may exist as a means for singular users to identify with multiple realms, creating a dynamic and user-organized social classification system."

In spite of the open participation and use of ad hoc keywords, as a whole, folksonomies tend to represent majority views as any other KO systems (e.g. [3]). For instance, Huvila [6] proposes as a conclusion of a study of aesthetic representation in Flickr folksonomy that tag clusters can be used as candidate representations of predominant 'viewpoints'. The related tags of a certain tag form a tentative representation of the meaning of that particular tag. For instance, if adjective 'cute' is related to tags girls, cats and dogs, these categories may be seen as tentatively cute in the context of the particular folksonomy.

The aim of our presentation is to discuss the implications of critical theory and the concept of intersectionality to the interpretation and use of tag clusters as representations of viewpoints. In spite of the wealth of empirical research on folksonomies and clustering of tags together with examples of critical perspectives focussing on marginalisation in the context of folksonomies (e.g. [1]), no comprehensive critical approach to tag clusters has been presented.

Our argument of the practicability of intersectionality as a framework of understanding the nature of tag clusters is based on the review of earlier literature together with a grounded theory based empirical analysis of a set of Flickr tags similar to those used by Huvila [6]. The tags and tag clusters were retrieved from Flickr using the basic search function and built-in clustering functionality. A total 112 clusters were analysed.

Theories on intersectionality have in the last decades become more and more prominent in analyses of situations and processes relating to identity categories such as the ones Staunæs refers to as the "classical

background categories of gender, ethnicity, race, age, sexuality and class"[10]. More recent studies often include for example religion and functionality, but which background categories are relevant to study in a specific case are not seen as given beforehand. One of the central aspects of intersectionality in relation to multiple viewpoints in tagging is the questioning of ideas about exclusive categories. These are seen as intersecting, fluid and socially constructed. Intersectional theories also bring forth the importance of that people from different intersectional categories should be able to express themselves. Self-expression is seen as important both in terms of their identities and opinions in general, and in relation to identity categories. For example, one woman cannot speak for all women, because of the existence of other identity categories that maybe are taken for granted (especially if they represent majorities) and hidden.

On the basis of the earlier literature it is clear that the pluralistic potential of folksonomies has major limitations. As a whole, folksonomies have a tendency to emphasise majority viewpoints even if they allow pluralism on the level of individuals and groups. We suggest, however, that a critical reading of tag clusters may provide a practicable framework for alleviating some of the theoretical and practical shortcomings. We agree with the conclusions of Huvila [6] that tag clusters are related to viewpoints but emphasise at the same time that the relation of these two notions requires further elaboration.

Firstly, we suggest that there are analytically usable parallels between tag clusters and intersectional categories in the sense of suggested by an intersectional perspective. From an intersectional point of view, clusters combine categories, point to different directions within the tagging system and mix different levels of representation including perspectives (and opinions), form and content. The practical benefit of an automatic identification of tag clusters is that it is possible to expose new marginalised categories without prior knowledge of their existence. The empirical analysis of the tag clusters in Flickr data gives evidence of the plausibility of the proposal. The clustering of tags related to adjectives like *graceful* exposes three major categories of gracefulness (birds, dance and women). Anecdotal experiments with other types of tags such as Finnish (two clusters related to design/architecture and music) and Swedish (four categories: flag/food/ikea/blue/yellow, music, women and cars) reveal similarly intelligible clusters. According to the analysed material, the possibilities to adjust categories could probably add to the 'intersectional possibilities' of folksonomies. The analysis of tag

clusters might be refined by transparency of trust relations in KO systems [7] and by using certain explicit user-produced expressions of identity and viewpoint.

At the same time, it is necessary to consider the possible limitations of perceiving tag clusters as functional categories. First, the co-occurrence of certain tags is not necessarily an indication of the existence of a viewpoint. Secondly, the automatic construction of clusters (i.e. clustering algorithm) has a major influence on the sections that become visible and the ones that are marginalised. In the essence, the clustering algorithm may become a new poorly understood force of marginalisation without critical analysis of the results and use of other complementary methods of exposing otherwise hidden categorisations. Thirdly, in contrast to the urge of Feinberg [5] of more transparency, the heterogeneity of the participants of the tagging communities and the relative anonymity of individuals can make it difficult to judge what viewpoints are represented and how. Even if the tags and user profiles would be transparent similarly to the clustering methods, the motivations of tagging are not necessarily intelligible without further elaboration. Finally, the attempts to improve the usability of clusters by refining categorisations are also controversial. Revisions based on an explicit attachment to a group identity or trust relation could reduce the diversity of represented viewpoints and imply a risk of ghettoisation of viewpoints. All attempts to refine clusters may also be criticised of being evaluative. The concept of “noise” [2], the perceived existence of ‘wrong’ tags or suboptimal clusters and the strive for refining clusters are related to acts of marginalisation comparable to an implementation of a hegemonic classification system. Simultaneously with exposing certain viewpoints, they can marginalise others and reduce the possibility for serendipitous discovery of potentially meaningful categories of information. Olson proposes and demonstrates the applicability of an idea of serendipitous consciousnessraising as an antidote to ghettoisation of certain topics or perspectives [8]. In her view, serendipitous searches and systems give the possibility to put forward and raise awareness of marginalised viewpoints. Not only to find something that proves to be valuable but also to be exposed to other viewpoints and be shown that they exist can be meaningful.

As a conclusion we suggest that there is a general need for a critical perspective to categorisations such as tag clusters and the consequences of the acts of categorisation in the context of folksonomies. Secondly, we suggest that explicit measures similar to Olson’s

consciousnessraising are needed to counter the marginalising tendencies of tagging systems.

References

- [1] Melissa Adler. Transcending Library Catalogs: A Comparative Study of Controlled Terms in Library of Congress Subject Headings and User-Generated Tags in LibraryThing for Transgender Books. *Journal of Web Librarianship*, 3(4):309–331, 2009.
- [2] Nicholas Auray. Folksonomy: the New Way to Serendipity. *Communications and Strategies*, (65):67–89, 2007.
- [3] danah boyd. questions of classification. *Many-to-Many*, 30 Jan 2005.
- [4] John Deodato. *Critical theory for library and information science : exploring the social from across the disciplines*, chapter Deconstructing the Library with Jaques Derrida, pages 75–87. Libraries Unlimited, Santa Barbara, CA, 2010.
- [5] Melanie Feinberg. Hidden bias to responsible bias: an approach to information systems based on Haraway’s situated knowledges. In *Proceedings of the COLIS6 conference - Featuring the Future*, 2007.
- [6] Isto Huvila. Aesthetic judgments in folksonomies as a criteria for organising knowledge. In Claudio Gnoli and Fulvio Mazzocchi, editors, *Paradigms and conceptual systems in knowledge organization. Proceedings of the 11th International ISKO Conference 23-26 February, Rome, Italy*, pages 308–315. Ergon Verlag, 2010.
- [7] Jens-Erik Mai. Classification in a social world: bias and trust. *Journal of Documentation*, 66(5):627–642, 2010.
- [8] Hope A. Olson. *The power to name : locating the limits of subject representation in libraries*. Kluwer, Dordrecht and Boston, 2002.
- [9] Hope A. Olson and Melodie J. Fox. *Critical theory for library and information science : exploring the social from across the disciplines*, chapter Gayatri Chakravorty Spivak: Deconstructionist, Marxist, Feminist, Postcolonialist, pages 295–309. Libraries Unlimited, Santa Barbara, CA, 2010.

[10] Dorte Staunes. Where have all the subjects gone? bringing together the concepts of intersectionality and subjectification. *NORA*, 11:101–110, 2003.