Visualization of co-readership patterns from an online reference management system

Peter Kraker [a,*], Christian Schlögl [b], Kris Jack [c], Stefanie Lindstaedt [a]

[a] Know-Center, Inffeldgasse 13, 8010 Graz, Austria
[b] University of Graz, Universitätsstraße 15, 8010 Graz, Austria
[c] Mendeley, 144a Clerkenwell Road, EC2 RF London, UK

A B S T R A C T

In this paper, we analyze the adequacy and applicability of readership statistics recorded in social reference management systems for creating knowledge domain visualizations. First, we investigate the distribution of subject areas in user libraries of educational technology researchers on Mendeley. The results show that around 69% of the publications in an average user library can be attributed to a single subject area. Then, we use co-readership patterns to map the field of educational technology. The resulting visualization prototype, based on the most read publications in this field on Mendeley, reveals 13 topic areas of educational technology research. The visualization is a recent representation of the field: 80% of the publications included were published within ten years of data collection. The characteristics of the readers, however, introduce certain biases to the visualization. Knowledge domain visualizations based on readership statistics are therefore multifaceted and timely, but it is important that the characteristics of the underlying sample are made transparent.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In recent scientometric literature, usage data is being discussed as a valuable alternative to citations. With the advent of e-journals, digital libraries, and web-based archives, click and download data have been suggested as a potential alternative to citations (Kurtz et al., 2005; Rowlands & Nicholas, 2007). Compared to citation data, usage data has the advantage of being available earlier, shortly after a paper has been published. In many instances, usage statistics are also easier to obtain and collect (Bollen, Sompel, Smith, & Luce, 2005; Brody, Harnad, & Carr, 2006; Haustein & Siebenlist, 2011). Furthermore, usage statistics allow for an analysis of publications and research outputs that do not receive citations or for which citations are not tracked (Priem & Hemminger, 2010).

Another type of usage data besides clicks and downloads is created in social reference management systems like BibSonomy[1] and Mendeley.[2] These systems enable users to store their references in a personal library and share them
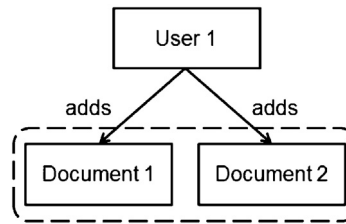

Fig. 1. Co-readership of two documents is established when at least one user has added the two documents to his or her user library.
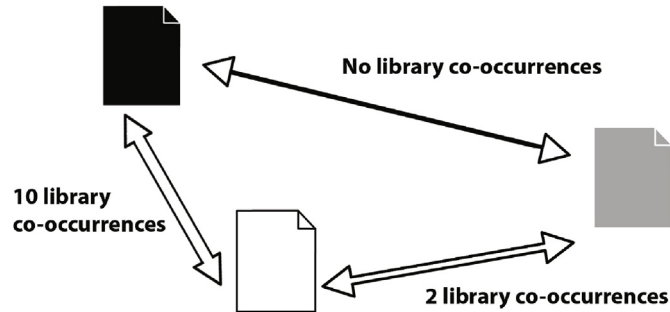

Fig. 2. Relationships between documents in a field based on co-readership. Co-occurrence in user libraries is employed as a measure of subject similarity.

with other people. The number of times an article has been added to user libraries is commonly referred to as the number of readers, or in short readership.[3]

Readership statistics have been of high scientometric interest in recent years. It has been shown that readership statistics provide a good coverage of top publications (Bar-Ilan et al., 2012), and that there is a medium correlation between readership data and citations (Schlögl et al., 2013) and a medium to high correlation between the impact factor and journal readership (Kraker, Körner, Jack, & Granitzer, 2012). Furthermore, Jiang, He, and Ni (2011) employ readership statistics from CiteULike to form clusters based on the occurrence and co-occurrence of articles in user libraries. They also correlate these clusters with ISI subject categories, and find them as effective as citation-based clusters when removing journals that cannot be found in CiteULike.

Therefore, we consider co-readership as a measure of subject similarity. Co-readership relation between two documents is established when at least one user has added the two documents to his or her user library (see Fig. 1). We assume that the more often the same two documents have been added to user libraries, the more likely they are of the same or a similar subject. The topical relationship established by co-readership can then be exploited for visualizations by clustering those papers that have high co-readership numbers (see Fig. 2). To the best of our knowledge, this measure has not been exploited before for knowledge domain visualization.

In this study, we first investigate the distribution of subject areas in user libraries of educational technology researchers on Mendeley. Then, we employ co-readership patterns for knowledge domain visualization to explore the field of educational technology. Educational technology is multi-disciplinary and highly dynamic in nature, as it is influenced by changes in pedagogical concepts and emerging technologies (Siemens & Tittenberger, 2009), as well as social change (Czerniewicz, 2010). Therefore, it seemed to be especially appropriate for this kind of analysis.

2. Related work

Traditionally, knowledge domain visualizations are based on citations. Small (1973) and Marshakova (1973) proposed co-citation as a measure of subject similarity and co-occurrence of ideas (see Fig. 3, left side, for a graphical representation of the relationship). This relationship can be employed to cluster documents, authors, or journals from a certain field and to map them in a two-dimensional space. Co-citation analysis has been used to map many fields, for instance information management (Schlögl, 2001, p. 48), hypertext (Chen & Carr, 1999), and educational technology (Chen & Lien, 2011) to name just a few. Furthermore, co-citation analysis has also been used to map out all of science (Boyack, Klavans, & Börner, 2005; Small, 1999).

---

[3] Initially, the term readership might seem a bit misleading, because the addition of an article to a user library does not guarantee that the article has actually been read by said user. Nevertheless, researchers need to make a second decision after downloading an article before they add it to their user libraries. Furthermore, the term is already well established among researchers (see e.g. Bar-Ilan et al., 2012; Haustein & Larivière, 2014; Thelwall & Maflahi, 2014; Zahedi, Costas, & Wouters, 2014); thus we use it in our research for reasons of consistency and to avoid neologisms.