# *Identifying topics of interest of Mendeley users using the text mining and overlay visualization functionality of VOSviewer*

Zohreh Zahedi and Nees Jan van Eck

*z.zahedi.2@cwts.leidenuniv.nl; ecknjpvan@cwts.leidenuniv.nl*
Centre for Science & Technology Studies (CWTS), Leiden University, P.O. Box 905, 2300 AX Leiden, (The Netherlands)

This poster presents the results of a study in which we have analysed the topics of interest of Mendeley users (i.e. Students, PhDs, PostDocs, Researchers, Professors, Librarians, Lecturers & other Professionals) using text mining and visualization techniques. Beside analyzing topics of interest of Mendeley users, we have also identified fields of science for which readership information can be an interesting source of information complementary to citation information. For this purpose, we have used WoS citation data and Mendeley readership data for a set of 980,698 WoS publications (articles and reviews) with a DOI from 2011[1].The VOSviewer software tool (Van Eck & Waltman, 2010) was used to create so-called overlay visualizations. These visualizations show additional information on top of a base map. Two types of base maps were used. A base map containing the 250 WoS subject categories was used to analyze differences in readership activity across research fields and to analyze differences in interest between types of users. Base maps containing terms extracted from titles and abstracts using the text mining functionality of VOSviewer (Van Eck & Waltman, 2011) were used to analyze differences in readership activity within research fields.

**Preliminary findings :**

*General overview of readership vs citation density of WoS publications across research fields*

The readership density of the WoS 2011 publications across the 250 WoS subject categories normalized by the number of citations per field revealed that there is a higher density of Mendeley readerships over citations for some fields of science. For example, some fields within social sciences (like management, business, psychology), humanities, neurosciences, computer sciences and biology have an above average readership activity compared to other fields. In contrast fields such as clinical medicine, natural sciences, and engineering exhibits relatively more citation density than readerships. This may show the potential value of readership counts for the fields with a low citation density.

*Term maps visualizing readership activity of Mendeley users within research field*

The term maps created for the above mentioned fields (social sciences and neurosciences) were used to explore in more detail the topics of interest of Mendeley users within these fields. Due to space limitation, the term maps of other fields within social sciences (such as management, business, psychology) are not presented here. According to figure 1, some interesting differences in readership activity are visible within both social sciences and neurosciences fields respectively. In social sciences most attention seems to be given to cognitive psychology, marketing and innovation while least attention seems to be given to the topics such as politic, law, philosophy and theology (figure 1 left). In neurosciences, some topics such as imaging (FMRI or MRI), neuron, drug addiction, brain activity accumulated the highest readerships while the least amount of readerships accumulated by terms such as protein, vitro, toxicology and cell death (figure 1 right).

[1] .Compared to the previous study, in this study we have accessed to the full readerships data per academic status of Mendeley users.

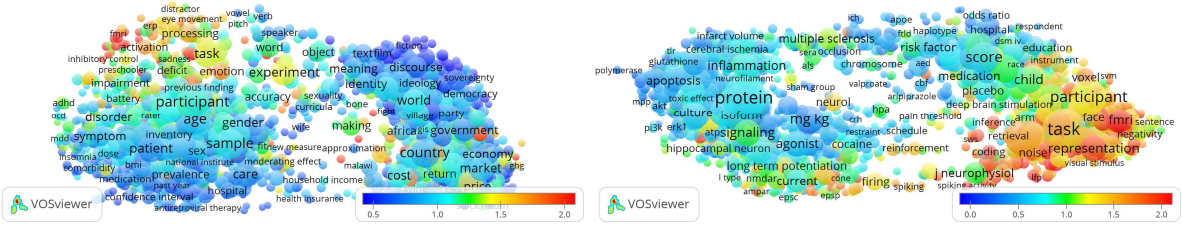


**Figure 1. Term maps of readership activity of Mendeley users within the social sciences (left) and neurosciences (right). The size and the color of a term indicate, respectively, the number of publications in which the term occurs and the average number of readers of these publications (where blue represents a low number of readers, green a normal number of readers, and red a high number of readers)**

*Readership activity of Mendeley users across research fields: PostDocs vs Students*

Figure 2 provides an overview of the readership activity of two different types of Mendeley users, namely PostDocs and Students. Due to space limitations, the visualizations of all other types of users are not shown here. The figure for readership activity of PostDocs and Students shows a different and quite opposite pattern. PostDocs have relatively more attention for natural sciences (physics and mathematics), neurosciences and biology and less attention for the social sciences, humanities and medicine (figure 2 left). In contrast to PostDocs, Students seem to have a relatively strong focus on social sciences, humanities, computer sciences and some engineering fields than other fileds (figure 2 right).

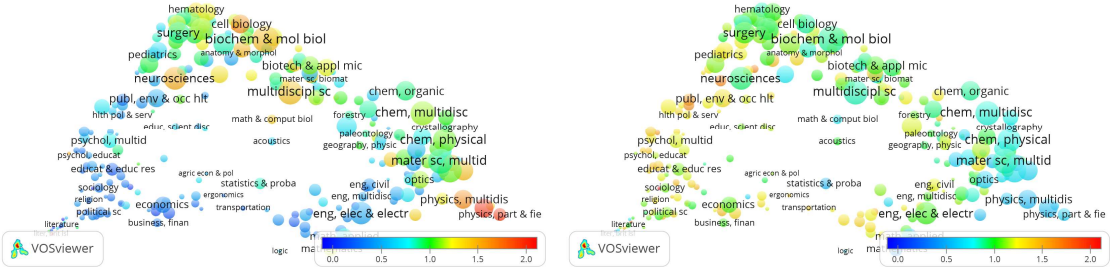


**Figure 2. Readership activity of Post docs (left) vs Students (right) normalized by the number of publications per field . The size and color of a subject category respectively indicates the number of publications and the readership activity of Mendeley users in the subject category ranging from blue to red (blue represents a low and red represents a high readership activity).**

## Conclusion and outlook:

Mendeley readership information is a rich source of altmetrics (Zahedi, Costas & Wouters, 2014), correlates the most with citations (Costas, Zahedi & Wouters, 2014) and has a stronger presence across some fields of science compared with citations and also in comparison with other social media metrics (Costas, Zahedi & Wouters, 2015). Our findings show that there are quite some disciplinary differences in terms of readership activity and in the topic interests within fields and among different users types. In general, the average readership activity is relatively highest in social sciences and humanities and relatively lowest in medicine, engineering and natural . This confirms the preliminary results found in a previous study (Zahedi & Van Eck, 2014). Mendeley is an interesting source of data for exploring the use of scientific publications by different users. Still more research is needed to understand the difference in the usage and topic interests of Mendeley users. In the full paper, we willl include all subject maps of users and the term maps of the subfields with highest readership density.

**Selected References:**

Costas, R., Zahedi, Z., & Wouters, P. (2015). The thematic orientation of publications mentioned on social media: large-scale disciplinary comparison of social media metrics with citations, Aslib Journal of Information Management, 67(3). http://dx.doi.org/10.1108/AJIM-12-2014-0173.

Van Eck, N.J., & Waltman, L. (2010). Software survey: VOS viewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.

Van Eck, N.J., & Waltman, L. (2011). Text mining and visualization using VOS viewer. ISSI Newsletter, 7(3), 50-54.

Van Eck, N.J., Waltman, L., Van Raan, A.F.J., Klautz, R.J.M., & Peul, W.C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. PLoS ONE, 8(4), e62395.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of “alternative metrics” in scientific publications. Scientometrics, 101 (2): 1491-1513. doi:10.1007/s11192-014-1264-0.

Zahedi, Zohreh & Van Eck, Ness (2014). Visualizing readership activity of Mendeley users using VOSviewer. Presented at the altmetrics14: expanding impacts and metrics An ACM Web Science Conference 2014 Workshop, 23-26 June, Indiana University, Indiana, USA.