Anniversary Statement

Howard D. White

College of Computing and Informatics, Drexel University,
Philadelphia, PA, 19104, USA

whitehd@drexel.edu

Congratulations on the tenth anniversary of this workshop! As someone who has thought for decades about bringing bibliometrics to bear on information retrieval, let me cite three of my papers that express an idea worth emphasizing. White [1,2] and White et al. [3] are based on quantitative results, but they mainly present in extensive qualitative detail why the results are potentially useful. Results of a certain kind must be dependable—hence the algorithms and the statistical tables—but they should also have face validity whichever retrieval is examined. At least as a goal, they should be intelligible in the sense that they could be explained and justified to non-technical persons. In other words, *how does this bibliometrically-enhanced retrieval help?*

Much more broadly, perhaps the concluding two paragraphs of [4], a review of bibliometrics from more than 30 years ago, can still stand as a vision statement today:

> Bibliometric models are necessitated by one kind of information overload—the unceasing appearance of published works, of copies of texts. "Copy" and "copious" both come from *copia*, the Latin word for "abundance," preserved in *cornucopia*, the horn of plenty. For information science, the "horn" is the array of modern reproduction technologies and the "plenty" is copious copies, endlessly recopiable. Yet there is no commensurate gain in time to read. Against the plenitude of published works, human attention is always sharply limited—a scarce resource. Specialists may read all, or a large part, of a literature over time, and in a lifetime they may read and master (at least to their own satisfaction) several literatures. But no one can read or master more than a minute fraction of what has been published and thereby indefinitely preserved. A barrier always exists. Bibliometricians, like information scientists in general, operate on the other side of the reading barrier. They look for formally specifiable ways of describing literatures without necessarily reading them, or of minimizing and targeting reading. Increasingly, they look for ways of describing literatures algorithmically, through computer rather than human recognition of specified bibliographic features. Interestingly, this is bringing them closer to the retrievalists—a development interwoven through this review.
>
> Both the strengths and weakness of bibliometrics in algorithmic modeling derive from its bibliographic foundations. On the one hand, bib-

liographies supply very extensive bodies of data—extensive enough to register the full outlines of contemporary learned research at several levels of detail and with time a measured variable. Nothing else in the way of "knowledge representation" is comparable in breadth. On the other hand, bibliometrics can deal only with *explicit* data—occurrences and co-occurrences of words—in bibliographies and (to a lesser extent) in the texts of published works. Vast realms of *implicit* textual meaning elude it, as do the realms of human life not yet—and perhaps never to be—consigned to text. Nevertheless, even if limited, its possibilities seem well worth further exploration. Bibliometrics models literatures, yes; but its distinctive displays can also be thought of as modeling the structure of human interests. When viewed in this more psychological light, its implications go beyond information retrieval, to bear on learning, knowing, and creating. We may yet see it as part of a cognitive science that is only beginning to emerge.

References

1. White, H.D.: Bag of works retrieval: TF*IDF weighting of works co-cited with a seed. International Journal on Digital Libraries 19(2–3), 139–149 (2018), doi:10.1007/s00799-017-0217-7
2. White, H.D.: Pennants for Garfield: Bibliometrics and document retrieval. Scientometrics 114(2), 757–778 (2018), doi:10.1007/s11192-017-2610-9
3. White, H.D., Lin, X., Buzydlowski, J.W., Chen, C.: User-controlled mapping of significant literatures. Proceedings of the National Academy of Sciences 101(suppl 1), 5297–5302 (2004), doi:10.1073/pnas.0307630100
4. White, H.D., McCain, K.W.: Bibliometrics. Annual Review of Information Science and Technology 24, 119–186 (1989)