

## Remembering Don Swanson: Link to Bibliometric-enhanced Information Retrieval\*

Aparna Basu

Formerly at National Institute of Science Technology and Development Studies  
(CSIR-NISTADS), New Delhi and South Asian University,  
New Delhi, India

[aparnabasu.dr@gmail.com](mailto:aparnabasu.dr@gmail.com)

Don R. Swanson (Oct. 10, 1924 – Nov. 18, 2012) was an American information scientist, known for his work in literature-based discovery in the biomedical sciences. Swanson received his B.S. degree in Physics at Caltech, in 1945, followed by an M.A., and then a PhD in Theoretical Physics from the University of California at Berkeley in 1952. He worked as a physicist at various laboratories until 1963, when he was made a professor and served as dean of the Graduate School of Library Science at the University of Chicago until 1972 and again from 1977–79 and 1987–89. In 2000, he was awarded the ASIST Award of Merit, the highest honor of the society, for his “lifetime achievements in research and scholarship.” (Wikipedia). He developed a system called Arrowsmith to obtain meaningful links between MEDLINE articles in the area called ‘Literature-based Discovery.’

The simple basis of Swanson’s idea of what he called ‘undiscovered public knowledge’ was as follows. He saw that science and the scientific literature were divided into many specialties—partly as a result of a spontaneous response of scientists to the exponential growth of science. Specialties further divided into sub-specialties, each with its own literature. An unintended consequence of specialization was fragmentation. Each specialty would have its own community that would largely restrict itself to its own literature for study, and hence also for citation. As the total literature grew, the number of specialties (but not in general the size of each), increased [4,10]. It was not difficult to foresee that at some point one could come up with complementary literatures which have some idea or knowledge element in common,—a logical connection—but which do not cite each other and are not co-cited. Such non-interactive literatures would be difficult to connect using usual methods of information retrieval. Swanson suggested that even in the 1950’s it was clear that conceptual problems of greater subtlety were involved in Information retrieval. Unfortunately, there has been little interest in these problems in spite of the rapid development and growth of online services since that time He called this the problem of ‘undiscovered public knowledge’ [8].

---

\* A companion video is hosted at <https://youtu.be/tEPyL-x-R1o>.

The focus here is not accessing a certain body of literature with given keywords, but looking at the relatedness between two bodies of literature. For example, if certain environmental conditions A are known in the Biomedical literature to cause certain symptoms B, and some disease C in the Medical literature has symptoms B then it may be inferred that environmental conditions A could be the cause of disease C. The significance of the ‘information explosion’ is therefore not just in the growth of literature but in the combinatorial explosion of “unnoticed and unrecognized” mutual connections. If these knowledge fragments from two (+) complementary and non-interactive literatures are brought together, one may hope to gain useful information.

With the recent introduction of Bibliometric-enhanced Information Retrieval or BIR [1,5], the stage seems to be set for addressing more complex problems in IR including the problem of ‘undiscovered public knowledge’ suggested by Swanson. The two conditions, complementarity and noninteraction describe a model structure that shows how useful information can remain undiscovered even though its components consist of public knowledge [9,10]. The motive for bringing together the literatures, is to call attention to an apparently unnoticed association that may be worth investigating. The scientific literature is an abstract world of human-created “objective” knowledge. It is open to exploration and discovery, for it can contain territory that is subjectively unknown to anyone [6].

Neil Smalheiser, Swanson’s long term collaborator writes of “problems and issues which were inherent in Don’s thoughts during his life, but which have not yet been fully taken up and studied systematically” [7]. Some work done in this direction explores the indirect associations between terms as direct associations tend to proliferate. The proposed indirect association measures extend traditional association measures to quantify indirect rather than direct associations while preserving valuable statistical properties [3]. Other papers need more investigation, see, e.g., [2]. This abstract may be regarded as a call for a practical research program in this direction, to look for useful knowledge, as it were, “hidden in plain sight.”

## References

1. Cabanac, G., Frommholz, I., Mayr, P.: Bibliometric-Enhanced Information Retrieval: 10th Anniversary Workshop Edition. In: Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F. (eds.) ECIR’20: Proceedings of the 42nd European Conference on Information Retrieval. LNCS, vol. 12036, p. to appear. Springer (2020), doi:[10.1007/978-3-030-45442-5\\_85](https://doi.org/10.1007/978-3-030-45442-5_85)
2. Chen, C., Song, M.: Visualizing a field of research: A methodology of systematic scientometric reviews. PLOS ONE 14(10), e0223994, doi:[10.1371/journal.pone.0223994](https://doi.org/10.1371/journal.pone.0223994)
3. Henry, S., McInnes, B.T.: Indirect association and ranking hypotheses for literature based discovery. BMC Bioinformatics 20(1), 425 (2019), doi:[10.1186/s12859-019-2989-9](https://doi.org/10.1186/s12859-019-2989-9)

4. Kochen, M.: On natural information systems: Pragmatic aspects of information retrieval. *Methods of Information in Medicine* 2(4), 143–147 (1963), doi:[10.1055/s-0038-1636217](https://doi.org/10.1055/s-0038-1636217)
5. Mayr, P., Scharnhorst, A.: Scientometrics and information retrieval: Weak-links revitalized. *Scientometrics* 102(3), 2193–2199 (2015), doi:[10.1007/s11192-014-1484-3](https://doi.org/10.1007/s11192-014-1484-3)
6. Popper, K.R.: *Objective Knowledge: An Evolutionary Approach*. Oxford University Press, New York, NY (1972)
7. Smalheiser, N.R.: Rediscovering Don Swanson: The past, present and future of literature-based discovery. *Journal of Data and Information Science* 2(4), 43–64 (2017), doi:[10.1515/jdis-2017-0019](https://doi.org/10.1515/jdis-2017-0019)
8. Swanson, D.R.: Undiscovered public knowledge. *The Library Quarterly* 56(2), 103–118 (1986), doi:[10.1086/601720](https://doi.org/10.1086/601720)
9. Swanson, D.R.: Two medical literatures that are logically but not bibliographically connected. *Journal of the American Society for Information Science* 38(4), 228–233 (1987), doi:[dq8phz](https://doi.org/dq8phz)
10. Swanson, D.R.: Integrative mechanisms in the growth of knowledge: A legacy of Manfred Kochen. *Information Processing & Management* 26(1), 9–16 (1990), doi:[10.1016/0306-4573\(90\)90005-m](https://doi.org/10.1016/0306-4573(90)90005-m)