Which h-index? – A comparison of WoS, Scopus and Google Scholar

JUDIT BAR-ILAN

Department of Information Science, Bar-Ilan University, Ramat Gan (Israel)

This paper compares the h-indices of a list of highly-cited Israeli researchers based on citations counts retrieved from the Web of Science, Scopus and Google Scholar respectively. In several case the results obtained through Google Scholar are considerably different from the results based on the Web of Science and Scopus. Data cleansing is discussed extensively.

Introduction

Until just a few years ago, when citation information was needed the single most comprehensive source was the ISI Citation Indexes. Although the Citation Indexes were often criticized for various reasons, there was no other source to rely on. Data from the ISI Citation Indexes and the Journal Citation Reports are routinely used by promotion committees at universities all over the world. In this paper we refer to the Web version of Citation Indexes, i.e., to the Web of Science (WOS) (http://portal.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame).

Recently two alternatives to the ISI Citation Indexes have become available. One of them is Scopus (http://www.scopus.com/) developed by Elsevier and the other is the freely available Google Scholar (http://scholar.google.com/). Each of these has a
different collection policy which affects both the publications covered and the number of citations to the publications. How different are these citation databases? In this paper we tried to provide a partial answer by considering the h-indexes [Hirsch, 2005a, b] of a group of highly cited researchers based on each of the three citation databases.

**Literature review**

**Comparing the databases**

The Science Citation Index was first published in print in 1963 with citation data from 1961 [Garfield, 1963]. “The Web of Science provides seamless access to current and retrospective multidisciplinary information from approximately 8,700 of the most prestigious, high impact research journals in the world.” [Thomson Scientific, no date]. The references from all the indexed items are extracted and the cited reference interface lists all citation to works of an author regardless of whether the cited items are indexed by WOS or not.

Until very recently, the Web of Science was the only comprehensive database to provide citation data. However, in November 2004 the citation database scenery changed considerably at once by the launching of Scopus on November 3, 2004 [Reed Elsevier, 2004] and Google Scholar on November 18, 2004 [Payne, 2004]. Scopus provides full citation coverage from 1996 and onwards, and claims to be “the largest abstract and citation database of research literature and select results from the web. Scopus covers 27 million abstracts, 230 million references and 200 million web pages” [Scopus, no date]. Scopus provides citation data only for the items indexed by it.

Google, probably on purpose, does not provide any explicit information either about the number of records or about its time coverage. Google Scholar, unlike WOS and Scopus, is freely accessible. They index data from publishers only if the publisher is willing to provide at least the abstract of the paper freely [Google Scholar, 2005] – viewing the full text may be fee or subscription based. The data comes from other sources as well, like freely available full text from preprint servers or personal websites as well, thus in many cases the full text is freely available for all users. References are automatically extracted from the full text of the indexed items. In case the reference itself is not indexed by Google Scholar, only the number of citation to that item appear in the search results.

Google Scholar was and is received with mixed feelings. For example, Giles [2005] reports: “Although there are no detailed studies, many librarians report that faculty members and students are beginning to use the search engine; some suspect that Scholar will replace more established, and more costly, search tools. It already directs more online traffic to Nature websites than any other multidisciplinary science search engine.” [Giles, 2005: 554]. Librarians seem to be less enthusiastic than their clients: