

# Exploring Enhancements in the Research Literature Base through Techmining

Fernando Palop<sup>a</sup>, Alan L. Porter<sup>b</sup>, Blanca de-Miguel-Molina<sup>a</sup>, and Andrés Avalos<sup>c</sup>

<sup>a</sup> Business Department, Universitat Politècnica de Valencia, Spain

<sup>b</sup> TPAC, School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA

<sup>c</sup> Universidad Pontificia Bolivariana, Colombia

Email: fpalop@ingenio.upv.es

**Keywords:** Research literature sample, research field, TechMining, ISI WoS, SCOPUS

## Abstract

The authors explore ways to enhance the research literature samples used in literature analysis, scientometrics and TechMining studies. We compare result fusion from two literature leading sources: ISI WoS and Scopus. To this purpose, results of merging record samples on the same topics using the two mentioned sources are empirically tested. We look for the percentage of unique records that belong to one or both of the sources. We explore data fusion and record fusion using VantagePoint to build a richer dataset for further analyses than obtained from simple analysis of either WOS or Scopus searches. We investigate reasons for unique records, including: database coverage of various subject areas, degree of journal overlap, within-journal coverage comparison (and record fusion gains), and implications for basic analyses (e.g., trend fitting). We investigate diverse areas, probing several topics in more detail. Implications are drawn for various analytical sensitivities and strategies.

## *Searching for results differences working with two reference database sources*

The method we have followed includes 3 steps, which are explained through different cases and topics. First, we undertake searches in WoS and SCOPUS databases, using the same search terms in each case and we download the results. Second, we mine the articles records for each case with the software VantagePoint. Once results are merged, we compare them, in order to detect if there are differences in the results. Three types of divergences were found in our research. In some cases dissimilarities are in the source while in others cases the divergences are in the number of articles.

Type a) there are journals which are included both in WoS and Scopus, but we have found differences in the number of papers included in the retrieved results;

Type b) some journals are only included in WoS, and

Type c) some journals are only included in Scopus.

Third, one of the journals for each case is analysed to detect in a more detailed way the type a) divergence.

**Table 1 Results comparison upon sources and number articles for the six research results samples used**

	Sample topic					
	Artwork & restoration	Microalgae heterotrophic growth	Pulping Methods Banana Crops	Pulping methods waste oil palm	Pulping methods crop residue sugar cane	Chlorella or scenedesmus microalgae
Number of Journals shared by WoS and SCOPUS in the retrieved results of each research topic	23	91	1	3	1	6
Journals in WoS but not in Scopus	23	57	20	18	37	60
Journals in Scopus but not in WoS	31	32	5	3	5	11
<b>Total Journal Sources in the retrieved literature</b>	<b>77</b>	<b>180</b>	<b>26</b>	<b>24</b>	<b>43</b>	<b>77</b>
	<b>Num of differences among the papers number retrieved</b>					
Differences in the number of papers retrieved within the shared Journal (1)	43	87	1	9	3	3
Num of papers present in WoS Journals not contained in Scopus results	23	67	22	24	58	75
Num of papers present in Scopus Journals not contained in WoS results	46	35	10	3	8	17
<b>Total differences found in the number of papers</b>	<b>112</b>	<b>189</b>	<b>33</b>	<b>36</b>	<b>67</b>	<b>95</b>
<b>WoS total article records retrieved since 1960 (2)</b>	<b>258</b>	<b>368</b>	<b>24</b>	<b>36</b>	<b>60</b>	<b>84</b>
<b>SCOPUS total articles records retrieved</b>	<b>248</b>	<b>249</b>	<b>13</b>	<b>24</b>	<b>19</b>	<b>77</b>

(1) In this part of the research we only have established quantitative differences in the number of articles records retrieved. With help of TechMining we will be able to identify which percentage of these differences correspond to the same articles and so to which databases correspond certain contains not shared in the other.

(2) The time coverage from SCOPUS starts in 1960 so we limited the search to the period 1960 to 2013 to allow comparisons.

The two database sources comparison allows to disclose significant differences in the number of sources covered but as well what was less expected, in the number of documents retrieved from the same source. These differences depending on the database provider and your topic of interest may interfere with your research field when we are talking from more than two digits in number of Journals or papers covered.

We validated that some papers were not retrieved, particularly in the case of SCOPUS database, although the research terms were in the article record field. This introduces questions on the quality and performance of the retrieval software and at what point affects negatively the search results.

The use of Tech Mining allowed merging the results samples for each topic. The percentage won with the Tech Mining fusion in relation with the start just with WoS results vary from 13% to 21% depending of the topic. Just in one of the six cases analyzed this percentage arrives to 31%.

Finally this is an ongoing research. Our results still correspond to a reduced number of topics and knowledge areas. We would like that other researchers submit comparisons, running exactly the same searches or testing other topics with this approach and enlarge with equivalent analysis this research field. Our intention is to provide in the Conference paper the detailed step by step followed here.