

2015

Which VIVO Harvester queries best populate the VIVO Database with the most accurate faculty member publications?

JD Eldredge

PJ Kroth

E Garbin

Follow this and additional works at: <https://digitalrepository.unm.edu/hslic-posters-presentations>

Recommended Citation

Eldredge JD, Kroth PJ, Garbin E. Which VIVO Harvester queries best populate the VIVO Database with the most accurate faculty member publications? Presidents Program. ASIS&T Annual Meeting. St. Louis, MO. November 6-10, 2015.'

This Book is brought to you for free and open access by the Research and Scholarship at UNM Digital Repository. It has been accepted for inclusion in Posters and Presentations by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Which VIVO Harvester Queries Best Populate the VIVO Database with Accurate Faculty Member Publications?

Jonathan D. Eldredge, Philip J. Kroth, Eric F. Garbin
THE UNIVERSITY OF NEW MEXICO, ALBUQUERQUE, NM

OBJECTIVE

The National Institutes of Health encourages its research partners to use VIVO™ software to enhance research investigator collaboration. VIVO Harvester was created to populate VIVO profiles with authors' article citations from PubMed™. Presently, typical VIVO Harvester queries often inaccurately and incompletely populate author citations. This study identified critical design elements for VIVO Harvester queries in PubMed for accurately identifying author citations.

METHODS

Action research approach to a fidelity study. The authors iteratively improved query design accuracy by comparing each iteration with a gold standard for 25 faculty members' citations for 2001-2011. The query design consisted of three sub-queries: 1) Author query; 2) Institutional affiliation; 3) Publication date. We applied this query design against the then 23 million citations in PubMed.

RESULTS

We increased the sensitivity rate of the PubMed queries via Harvester to 70% through the iterative search query modifications. This means that the improved query produced 70% of the total citations in the gold standard. The specificity rate remained at 100% meaning that almost all of the citations linked to the institution's faculty members were accurate according to the gold standard. Table 1 provides the calculations.

Table 1: Sensitivity and Specificity of New Query

	(+)	(-)	
Harvester (+)	443	19	462
Results (-)	190	7,416,638	7,416,828
	633	7,416,657	7,417,290
	Sensitivity	0.7	
	Specificity	1.0	

DISCUSSION

The new expert-based PubMed author query yields greater sensitivity while maintaining specificity. The improved query emerged from a collaboration between a librarian, an IT professional, and an informaticist. All collaborators had local knowledge of the faculty investigators at their institution.

The National Library of Medicine began to link all authors in PubMed with their institutional affiliation during 2014. This study covered citations for the years 2001-2011. We predict that our improved author query will yield even more accurate results as sensitivity rises above 70% with each passing year while not sacrificing accuracy in the specificity rate of 100%.

VIVO Harvester

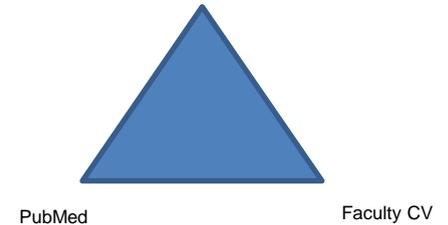


Figure 1. Triad of Ambiguity

Over the course of this study we identified a "Triad of Ambiguity" that surrounds these kinds of disambiguation challenges. The "Triad" consists of the lack of agreement between a faculty member's CV, a PubMed author search, and the results of a naive VIVO Harvester query. Figure 1 depicts the Triad of Ambiguity. This conceptualization aided us in understanding our challenges and might be adaptable to others' efforts to disambiguate authors' actual citations to publications.

ACKNOWLEDGMENTS

Clinical and Translational Sciences Center, National Institutes of Health award #UL1TR000041 and the Health Sciences Library and Informatics Center, University of New Mexico, Albuquerque, NM. UNM HRPO (IRB) # 12-511 exempt.