

Effectively and Efficiently Locating Research Evidence: The Pearl Harvesting Information Retrieval Framework

Robert Sandieson,
University of Western
Ontario,
London, Ontario, Canada

Modern Technologies Modern Potentials Modern Problems

- 1. Modern digital technologies allow for the storage of vast amounts of information, readily accessible
- 2. Poor organization of existing knowledge makes it difficult to satisfy search intents
- 3. The development of systematic frameworks for developing knowledge seeking pathways could assist in harnessing the wealth of knowledge available through databases

Difficulties with online bibliographic databases

- **Tremendous amount of information becoming available** (Adair & Vohra, 2003; Curan & Adams, 1999; Duff, 2001; Speight, 1999)
- **Most users trying to be self sufficient in their searches, yet report feeling overloaded with the amount of information available** (Drabenstott, 2003; Haglund & Olsson, 2008; Jankowska, 2004; Rudner, 2000; Wessel, Tannery, & Epstein, 2006)

Studies on people doing information searches

- limited use of the Boolean search function (Ivanitskaya, O'Boyle, & Casey, 2006; Rudner, 2000)
- spent little time during a specific search,
 - performed searches for others even though they were unqualified (Rudner, 2000)
- performed a limited number of queries per search,
 - made limited use of database thesauri (Hertzberg & Rudner, 1999);
- reported a lack of time to do extensive information searches (Wessel, Tannery, & Epstein, 2006)
- had difficulty with subject searches and did not know how to explore the range of keywords applicable for particular subjects (Drabenstott, 2003)
- Academics
 - citing fewer articles,
 - referencing more recent literature
 - referencing fewer journals (Evans, 2008).

Even researchers experience difficulties searching

- two large reviews of the literature on formative educational assessment
 - published a year apart produced a total 332 articles.
 - only 9 citations overlapped between the two studies (Black & Wiliam, 1998)
- Minor and Raju (2004) replicated a prior meta-analysis published in the Journal of Applied Psychology
 - found 57% more studies than original study, using same inclusion criteria
 - data reanalysis resulted in completely different conclusions than the original study
- Sandieson (2006) found that 42% of relevant articles missing in a published review of the literature on the topic of special education mathematics.
- Arendt (2007) reviewed 38 meta-analyses contained in APA journals published in 2004.
 - Only twenty-three of the studies provided complete lists of search terms that were used in database searches.
 - Of these, she found extra search terms that could have been used from the database controlled vocabulary (i.e., thesauri) for 78% of the articles.
 - 57% of articles had terms that were not included in the database thesauri.

Currently Recommended Search Strategies

- Citation tracking
 - Backward: use reference list from known articles (also called ancestral search)
 - Forward: find studies that cited known article, i.e., use Social Science Citation Index, Google Scholar and PsycINFO

Currently Recommended Search Strategies (con't)

- **HOWEVER**
 - Assumes that citations are linked
 - They may not be, as authors are biased in selecting other authors to cite

Currently Recommended Search Strategies (con't)

- Large scale browsing or hand searching
 - 1. Survey of indexes of known relevant journals, looking for potentially relevant studies
 - 2. Use very general keyword, e.g., Special Education, OR Disabilities
- **HOWEVER**
 - No methodological prescription as what journals to include in search, or how many journals to use
 - Resulting in inconsistencies of searches found between articles
 - Labourious
 - Takes a lot of time and cost to have people go through sometimes extremely large number of citations
 - Terms must be understood before hand or will overlook articles,
 - e.g., global developmental delay

Keyword/Keyphrase searching

- Words or phrases are entered into database search fields
- Words derived intuitively, consultation with experts in field or through reference to database predefined vocabularies, i.e., thesauri

Keyword/Keyphrase Searching (con't)

- Pearl Growing method
 - Find a relevant article
 - Locate it in a database
 - Investigate any relevant keywords in title, abstract, & descriptors (and identifiers in PsycINFO)
 - Use the keywords found to do a search
 - Investigate resulting relevant articles for any new relevant keywords/keyphrases
 - Continue search process with new keywords until cannot find more new keywords/keyphrases
 - => a list of relevant keywords/keyphrases = Synonym Ring

Keyword/Keyphrase searching

HOWEVER

- Keywords used in the scholarly field vary widely due to
 - History, philosophies, geography, culture, profession
- Database Thesauri as a sources for Keywords limited
 - Terms used in thesauri may not match with terms used in the scholarly field
- Using Pearl Growing assumes interconnectivity of articles' keywords, but this may not be true as with citation tracking

Keyword/Keyphrase Searching (con't)

- Some authors recommend **NOT** using keywords or keyphrases for comprehensive searching (Black & Wiliam, 1998; Raines, 2008)
 - Should only use hand searching of journal indexes

A Pickle?

- Searching for articles appears to be an Achilles' heel for comprehensive searching of information in databases

The Pearl Harvesting Information Retrieval Framework (PHIRF)

PHIRF

- Some authors suggest using a larger number of articles as a source for doing Pearl Growing (Schlosser, Wendt, Bhavnani, & Naila, (2006); White, 1994), but have not offered any further methodology
- PHIRF recognizes the importance of a wide sample of articles as a potential solution to gathering, i.e., harvesting, all the essential keywords on a topic
 - with a corresponding methodological framework

PHIRF

- **Basic PHIRF premise**
 - At any one time there are a finite set of terms that people use to represent ideas or topics
- The population of these terms is not initially known
- Therefore, a sampling procedure with extensive verification may yield a close approximation of the complete set of terms used to denote a particular topic or idea
- The collection of these terms can be used in the search fields of databases as keywords/keyphrases to produce information searches with the highest recall and precision

PHIRF Methodological Framework

Step One:

- Locate a sample of articles (i.e., Pearls) that represent the topic domain under consideration

PHIRF sampling

- sample sources for "Pearls":
 - A collection of articles found within in a collection of meta-analyses, e.g., Mostert (2003), as a source to find terms that denoted Developmental Disabilities/Intellectual Disabilities/Mental Retardation
 - Used 98 articles that were used in 6 meta-analyses

PHIRF keyword identification: Building search sensitivity

- Step 2, Keyword harvesting
 - Locate the chosen articles or “Pearls” in relevant databases, e.g., ERIC, PsycINFO
 - Analyze the bibliographic information, i.e., title, abstract, descriptors (and identifiers in PsychInfo) for possible relevant terms or phrases used by authors and librarian database coders
 - Document a complete list of potential terms that could be used for keyword searching

Refining the list: e.g., Mental Retardation

- Step 3: refining the list:
- retard* = mental retardation OR mentally retarded OR mild retardation OR severe retardation
- retard* produces 96%/98% relevant citations (ERIC/PsycINFO)

Example: "Disability" = low Intellectual functioning, i.e., intellectual disability or mental retardation

Possible terms in ERIC database

- Disabilities, Disabled, Severe disabilities, Multiple disabilities, Moderate disabilities, Moderate mental disabilities, Mild intellectual disabilities, Severe intellectual disabilities, Multiple disabilities, Developmental disabilities, Developmentally disabled, Learning disabilities
- + from PsycINFO: Substantial disabilities, Moderate intellectual disabilities, Mental disabilities

Refining the list: sensitivity plus precision

Truncation:

- Disabilities OR disabled => Disab*
- HOWEVER, the precision of Disab* is very low = 21% in ERIC and 15% in PsycINFO
- Specification:
 - specific versions of disab*:
 - mental disab* OR mild disab* OR moderate disab* OR severe disab* OR ...
 - => 85% relevant in ERIC and 87% in PsychInfo, without loss of any relevant articles in the search
- **Note:** do not use the Boolean NOT as a way of filtering articles

Validating the initial Synonym Ring

Step 4, Validation

- Because PHIRF uses a sampling method, it is likely that some terms may not have been located in the sample
- Cross checking with other sources can help verify if there are any outstanding terms

Validating the Initial Synonym Ring

- There is no set way of doing this, it's a matter of whatever is available

For the Intellectual Disabilities Synonym Ring

- Compare with database thesauri of ERIC and PsycINFO
 - The term cognitive impairment was a new term found in the thesaurus, so was added to the Synonym Ring
 - Many specific terms from Synonym Ring missing in the thesauri, or would need to know ahead of time where to look

Validating the Synonym Ring (con't)

Comparison with a list of 66 terms developed by Sandieson (1998)

- Present study showed that retard* is better than mental* retard*
- Specific versions of disab*/handicap* located in present investigation
- Terms in list not found through PHIRF:
 - cognitive disability, developmental handicap, and two other “delay” terms, mental and intellectual
- Other terms on list but citations found could be located through existing Synonym Ring

Validating Synonym Ring (con't)

Compared with more recent list of meta analyses in the field of intellectual disabilities to see what terms and search strategies used (in progress)

- Located 15 recent meta analyses on the topic of intellectual disabilities (since 2000)

Recent Meta Analyses in the field of IDD

Wide range of reporting of search strategies

- Browder, Spooner, Ahlgrim-Delzell, Harris, & Wakeman, (2008): 146 permutations of terms for IDD and mathematics, 6 hand searched journals
- Didden, Korzilius, van Oorsouw, Sturmey (2006): no terms specified, mentioned searched databases but not which ones
- Other meta analyses: 3–10 terms used; 8-46 journals; ancestral searches reported in some
- No new terms found for existing Synonym Ring

PHIRF Synonym Ring for Intellectual and Developmental Disabilities

- retard* OR mental* disab* OR mild disab* OR moderate* disab* OR severe* disab* OR profound* disab* OR multipl* disab* OR intellectual* disab* OR developmental* disab* OR substantial* disab* OR mild* handicap* OR moderate* handicap* OR severe* handicap* OR mental* handicap* OR multi* handicap* OR profound handicap* OR developmental* delay* OR delay* development OR mental* impair* OR intellectual* impair* OR cognitive impair*

Validating the Synonym Ring (con't)

- Did a search for meta-analyses using the same search criteria as Mostert (2003), using the PHIRF Synonym Ring for Intellectual Disabilities AND "meta analysis"
- Found 9 more articles (47%) than the 19 found by Mostert

Conclusions and Implications

- Today's global information societies
 - knowledge can expand rapidly, and the surge of information becoming available, it is important that people know how to "keep up" with the research
- Being able to access the existing literature is extremely important for researchers, scholars, students, professionals, and the public
 - Promotes informed democratic thinking, evidence based policy and practice
- Existing evidence shows that only a few people are doing high quality comprehensive searches

Conclusions and Implications

- Some researchers doing reviews and meta analyses favour hand searching
 - If enough journals are examined and documented this may prove to be a thorough procedure

HOWEVER

Hand searching is

- time consuming,
- may be prone to human error because of the volume of articles to search,
- is costly,
- lacks a methodology for choosing which journals to search,
- may miss articles where the relevant terminology is not known ahead of time

Conclusions and Implications(con't)

Keyword/Keyphrase Searching

- Problematic
 - Language is constantly evolving making it hard to track how it's used by various cultures and professions over time
- Although likely less than perfect, the PHIRF provides a methodological framework for harvesting relevant keywords on a specific topic

Conclusions and Implications(con't)

- Synonym Rings, when made publicly available
 - Address the problem of lack of uniformity of keywords/subject headings
 - Can be refreshed with new terms that are found
 - can be used by anyone (cut and paste) who has access to research databases without having to reinvent which terms to use
 - For those who do not have access to the financial or time resources to do large scale analysis from scratch
 - To facilitate transdisciplinary or multidisciplinary work
- Other topics that have been used: gifted education, mathematics learning disabilities, hearing impairment, attitudes and acceptance of disabilities

Limitations (at this point)

- Validation, here, proved sampling the way it is currently used is inadequate, so what might be better strategies for harvesting keywords/keyphrases
 - Maybe a hybrid of harvesting and pearl growing would be better
- There is still the question of knowing which databases are relevant to a particular query
 - Pearls might be used in the future to track which databases contain them

• *MERCI BEAUCOUP*



THE END