Quantifying Accuracy of Systematic Literature Review

The Case of Gerontology Literature Review

Shahryar Eivazzadeh

Peter Anderberg

Tobias C. Larsson

Johan S. Berglund

ISG16, 28 September 2016, Nice, France

BLEKINGE TEKNISKA HÖGSKOLA · BTH ·

The problem:
How accurate and inclusive an SLR is in finding the relevant literature?

How to compare performance of two sets of keywords (closely related)?

What other keywords might be included?

How many snowballing rounds to go?

Any indicator on how far the SLR is from capturing all the relevant papers?

Definitions:

Definition: A set of papers are *closed set* if they only have references (cite) each other

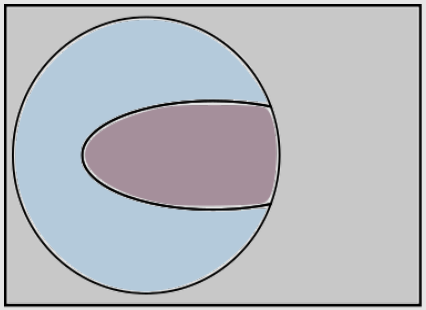
Definition: An SLR is *locally complete* if it reaches a closed set of papers

Definition: The *degree of incompleteness* of an SLR is the number of references in the papers captured in the last stage of SLR that are not available in the pool of papers captured earlier, divided by the whole number of captured papers in all stages

Definition: When sensivity of a keyword search string is 1, it means it can capture all qualified papers and snowballing does not add any new one

Definition: When specificity of a keyword search string is 1, it does not suggest any unqualified paper

Metaphor

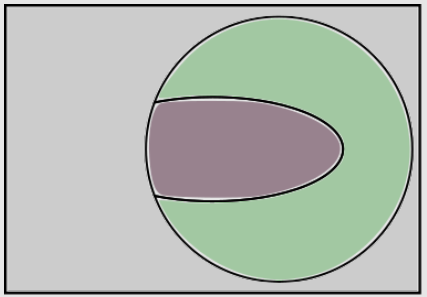


all indexed papers

selected papers
by the keywords

rightly select papers
by the keywords

Keyword Phase

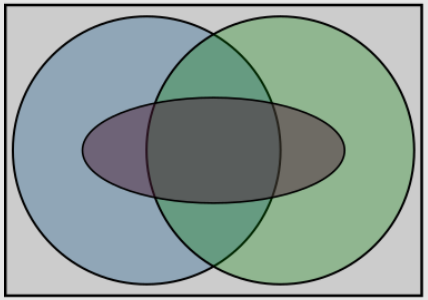


all indexed papers

selected papers
by snowballing

rightly select papers
by snowballing

Snowballing Phase



all indexed papers

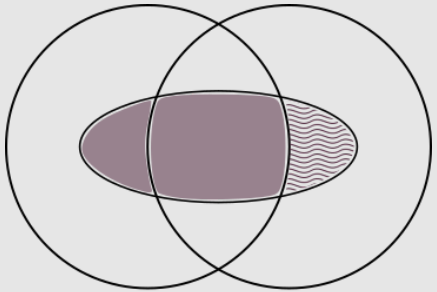
selected papers
by keywords

selected papers
by snowballing

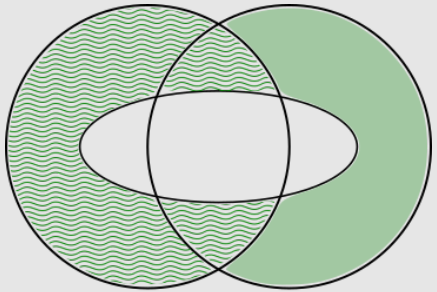
correctly select papers either
by keywords or snowballing

Both Phases

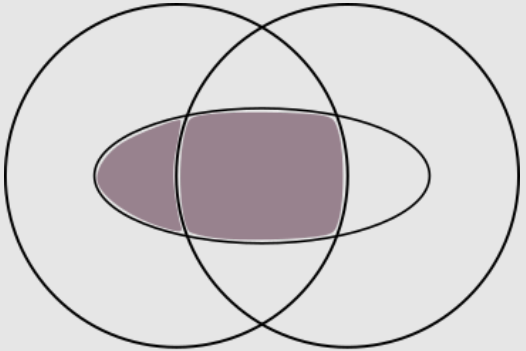
Keywords performance



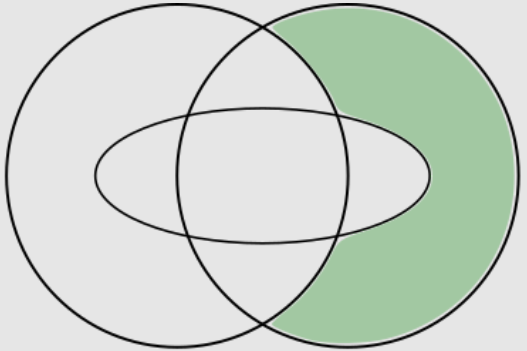
Sensitivity



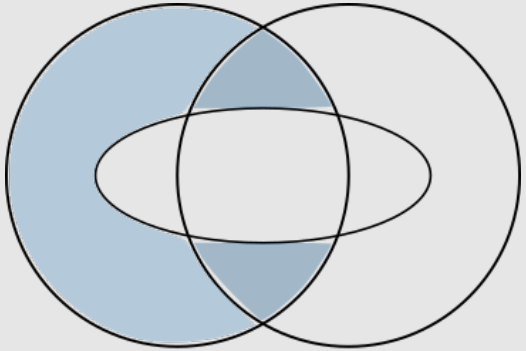
Limited Specificity



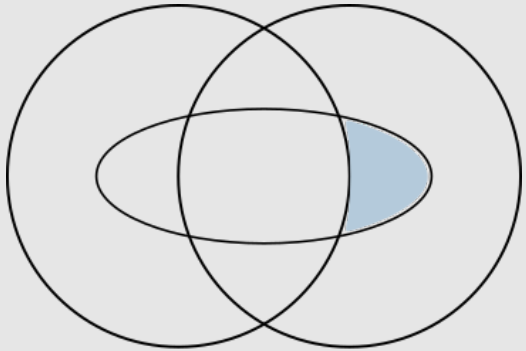
True Positives
by Keywords



Limited True Negatives
by Keywords



False Positives
by Keywords



False Negatives
by Keywords

Suggestions + Sample SLR on Prognosis of Dementia by Machine Learning

- To report sensivity, specificity and accuracy of keyword search string and (if feasible) the degree of incompleteness of the SLR

- To tune keywords for a better sensivity, specificity and accuracy (by analyzing snowball stage keywords) and to follow snowballing for less incompleteness

- Sample (numbers from:"

Prognosis of Dementia Employing Machine Learning and Microsimulation techniques: A systematic literature review, Ana Luiza Dallora Moraes, Shahryar Eivazzadeh, Emilia Mendes, Johan Berglund, Peter Anderberg, 2016

")

$TP_k = \#\ of\ qulified\ papers\ captured\ by\ the\ keywords = 37$
$TN_k = \#\ of\ unqulified\ papers\ captured\ by\ snowballing\ (but\ not\ the\ keywords) = 1158$
$FP_k = \#\ of\ unqulified\ papers\ captured\ by\ the\ keywords = 526$
$FN_k = \#\ of\ qulified\ papers\ captured\ by\ snowballing\ (but\ not\ the\ keywords) = 41$

$$Sensivity_k = \frac{TP_k}{TP_k + FN_k} = 0.47$$

$$Specificity_k = \frac{TN_k}{TN_k + FP_k} = 0.68$$

$$Accuracy_k = \frac{TP_k + TN_k}{TP_k + FN_k + TN_k + FP_K} = 0.678$$

$R_{nl} = \#\ of\ references\ in\ the\ last\ stage\ of\ the\ SLR\ not\ captured\ before$
$P_{all} = \#\ of\ all\ papers\ captured\ in\ all\ stages\ of\ the\ SLR$

$$Incompleteness = \frac{R_{nl}}{P_{all}}$$

Thanks!

sei@bth.se

BLEKINGE TEKNISKA HÖGSKOLA · BTH ·