**Statistical Function** \texttt{ris:hIndex}(D): integer

\{Input \( D \) is a sequence of pairs of document \( doc \) and document \( docCite \) which cites document \( doc \).\}

\begin{verbatim}
begin
  Count number \( N_p \) of documents including the first element \( doc \) of \( D \) without overlap ;
  val \( h := 0 \); repeat
    begin
      \( h := h + 1 \);
      Count number \( L \) of documents cited at least \( h \) documents utilising \( D \);
      Count number \( M \) of documents cited at most \( h \) documents utilising \( D \);
      if \((L \geq h \text{ and } (N_p - h) \leq M)\) then break ;
    end;
  return \( h \);
end;
\end{verbatim}

**Supplement 1.** An algorithm that implements the statistical function \texttt{ris:hIndex} introduced in Section 2.1, which computes \( h \) index with a sequence of pairs of the document \( doc \) and document \( docCite \) that cites document \( doc \).
FROM NAMED <http://omicspace.riken.jp/GRASQL/single/Mm/MEDLINE>
FROM NAMED <http://omicspace.riken.jp/GRASQL/relation/Mm/MEDLINE>
WHERE {
  ?x ripInference:hasEntity2 ?gene2 ;
  ripInference:hasEntity1 ?gene1 ;
  ripInference:hasPValue ?pInference .
  ?y ripSingleSearch:hasEntity ?gene1 ;
  ripSingleSearch:hasWord %keyword ;
}
EVALUATE ?p FOR ?gene1 ?gene2 {
  ?p = 1-(1-?pSingle)(1-?pInference)
}
EVALUATE ?pTotal FOR ?gene2 {
  ?pTotal = ris:multiPValue(?p)
}
ORDER BY ?pTotal ?p

In this query, there are two EVALUATE clauses that are evaluated sequentially in the order of their appearance. In the example, P value ?p for each pair (?entity1, ?entity2) is computed and then P value ?pTotal for each entity ?entity2 is computed. Finally, by evaluating the ORDER BY clause, the solutions of 4-tuples (?entity1, ?entity2, ?p, ?pTotal) are sorted by ?pTotal and ?p.
The function ris:multiPValue in the second EVALUATE clause is an implementation of Equation (2). Furthermore, ris:minPValue is an implementation of Equation (3), which does not appear in this paper.
Supplement 3. A GRASQL query that ranks mouse genes in the input list of mouse genes by computing the statistical significance between mouse genes and the keyword 'type 2 diabetes'.
Supplement 4. A GRASQL query that ranks GO terms associated with a set of mouse genes.

In this example, a $P$ value is computed for each hit GO term using a $2 \times 2$ contingency table that contains the number of mouse genes

a) associated with the GO term and included in the gene ID list
b) associated with the GO term but not included in the gene ID list
c) not associated with the GO term but included in the gene ID list and
d) not associated with the GO term and not included in the gene ID list.