**Course: Advanced Bioinformatics**

**Module title: IR Evaluation Measures**

**Module no. : 170**

The evaluation of an information retrieval system is the process of assessing how well a system meets the information needs of its users. Traditional evaluation metrics, designed for Boolean retrieval or top-k retrieval, include precision and recall. Many more measures for evaluating the performance of information retrieval systems have also been proposed. In general, measurement considers a collection of documents to be searched and a search query. All common measures described here assume a ground truth notion of relevancy: every document is known to be either relevant or non-relevant to a particular query. In practice, queries may be ill-posed and there may be different shades of relevancy.

Virtually all modern evaluation metrics (e.g., mean average precision, discounted cumulative gain) are designed for ranked retrieval without any explicit rank cutoff, taking into account the relative order of the documents retrieved by the search engines and giving more weight to documents returned at higher ranks.[citation needed]

The mathematical symbols used in the formulas below mean:

The mathematical symbols used in the formulas below mean:

X \cap Y - [Intersection](https://en.wikipedia.org/wiki/Intersection_(set_theory)) - in this case, specifying the documents in both sets X and Y

| X | - [Cardinality](https://en.wikipedia.org/wiki/Cardinality) - in this case, the number of documents in set X

\int - [Integral](https://en.wikipedia.org/wiki/Integral)

\sum - [Summation](https://en.wikipedia.org/wiki/Summation)

\Delta - [Symmetric difference](https://en.wikipedia.org/wiki/Symmetric_difference)

Precision:

Precision is the fraction of the documents retrieved that are relevant to the user's information need.

 \mbox{precision}=\frac{|\{\mbox{relevant documents}\}\cap\{\mbox{retrieved documents}\}|}{|\{\mbox{retrieved documents}\}|} 

In binary classification, precision is analogous to positive predictive value. Precision takes all retrieved documents into account. It can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system.

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

In binary classification, recall is often called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query.

\mbox{recall}=\frac{|\{\mbox{relevant documents}\}\cap\{\mbox{retrieved documents}\}|}{|\{\mbox{relevant documents}\}|} 

It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore, recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by computing the precision.

Note that the meaning and usage of "precision" in the field of Information Retrieval differs from the definition of accuracy and precision within other branches of science and statistics.

Fall-out

The proportion of non-relevant documents that are retrieved, out of all non-relevant documents available:

 \mbox{fall-out}=\frac{|\{\mbox{non-relevant documents}\}\cap\{\mbox{retrieved documents}\}|}{|\{\mbox{non-relevant documents}\}|} 

In binary classification, fall-out is closely related to specificity and is equal to (1-\mbox{specificity}). It can be looked at as the probability that a non-relevant document is retrieved by the query.

It is trivial to achieve fall-out of 0% by returning zero documents in response to any query.