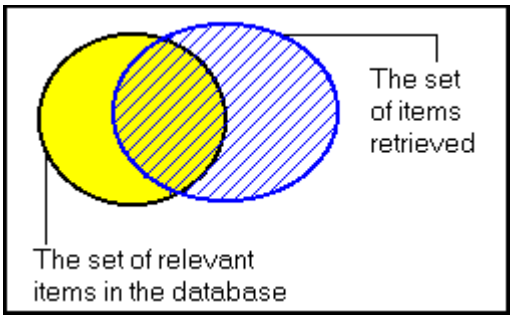
Measuring Search Effectiveness

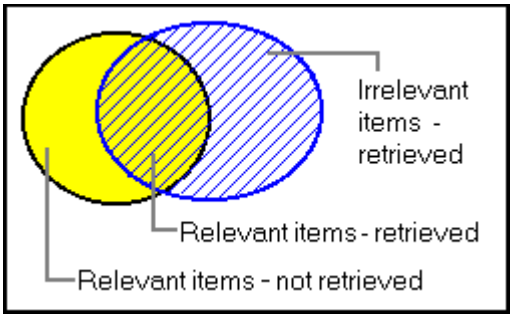
After finishing a search the nagging question in every searcher's mind is: "*Have I found the most relevant material or am I missing important items?*"

In addition every searcher hopes they don't retrieve "*a lot of junk*".

Unfortunately getting "*everything*" while avoiding "*junk*" is difficult, if not impossible, to accomplish. However, it is possible to measure how well a search performed with respect to these two parameters.

PRECISION & RECALL





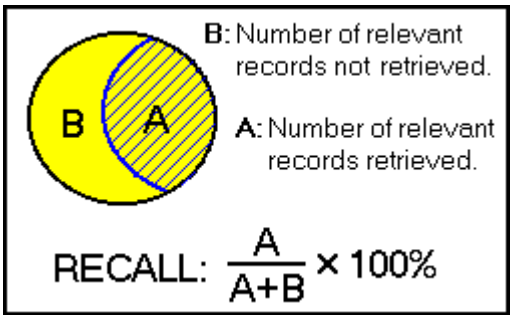
Precision and **recall** are the basic measures used in evaluating search strategies.

As shown in the first two figures on the left, these measures assume:

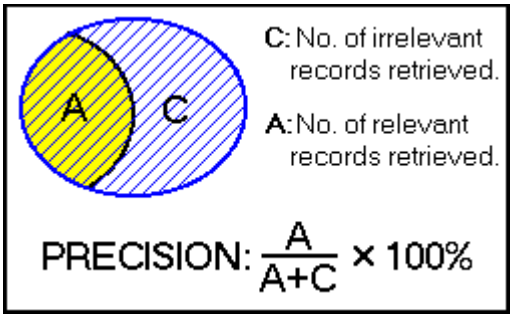
There is a set of records in the database which is relevant to the search topic

Records are assumed to be either relevant or irrelevant *(these measures do not allow for degrees of relevancy)*.

The actual retrieval set may not perfectly match the set of relevant records.



RECALL is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.



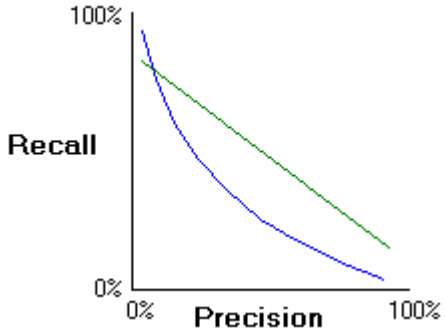
PRECISION is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

RECALL AND PRECISION ARE INVERSELY RELATED:

As recall ↑ precision ↓

conversely:

As recall ↓ precision ↑



In the graph above, the two lines may represent the performance of different search systems. While the exact slope of the curve may vary between systems, the general inverse relationship between recall and precision remains.

Why is there an inverse relationship?

Much of this relationship has to do with language. If the goal of a search is comprehensive retrieval, then the searcher must include synonyms, related terms, broad or general terms, etc. for each concept. He may decide to combine terms using Boolean rather than proximity operators. In addition, some secondary concepts may be omitted.

As a consequence of these decisions, precision will suffer. Because synonyms may not be exact synonyms the probability of retrieving irrelevant material increases. Broader terms may result in the retrieval of material which does not discuss the narrower search topic. Using Boolean operators rather than proximity operators may increase the probability that the terms won't be in context Unfortunately, if the searcher doesn't use these techniques he won't achieve high recall!

Other problems with Precision and Recall:

As noted earlier, records must be considered either relevant or irrelevant when calculating precision and recall. Obviously records can exist which are marginally relevant or somewhat irrelevant. Others may be very relevant and others completely irrelevant. This problem is complicated by individual perception: what is relevant to one person may not be relevant to another.

Measuring recall is difficult because it is often difficult to know how many relevant records exist in a database. Often recall is estimated by identifying a pool of relevant records and then determining what proportion of the pool the search retrieved. There are several ways of creating a pool of relevant records: one method is to use all the relevant records found from different searches, another is to manually scan several journals to identify a set of relevant papers.

Precision and Recall are useful measures despite their limitations:

As abstract ideas, recall and precision are invaluable to the experienced searcher. Knowing the goal of the search -- to find everything on a topic, just a few relevant papers, or something in-between -- determines what strategies the searcher will use. There are a variety of search techniques which may be used to effect the level recall and precision. A good searcher must be adept at using them. Many of these techniques are discussed in the section on search strategies.

Calculating Recall and Precision Scores

If it helps you understand Recall and Precision, try the following two exercises.

Exercise 1 (calculations)**:**

Problem:

Assume the following:

- A database contains **80** records on a particular topic
- A search was conducted on that topic and **60** records were retrieved.
- Of the 60 records retrieved, **45** were relevant.

Calculate the **precision** and **recall** scores for the search.

Solution:

Using the designations above:

- A = The number of relevant records retrieved,
- B = The number of relevant records not retrieved, and
- C = The number of irrelevant records retrieved.

In this example A = 45, B = 35 *(80-45)* and C = 15 *(60-45)*.

Recall = (45 / (45 + 35)) * 100% => 45/80 * 100% = **56%**
Precision = (45 / (45 + 15)) * 100% => 45/60 * 100% = **75%**

Exercise 2 (judging relevancy)**:**

Read each of the following titles. Decide if the paper described by the title would be relevant for a student paper on *the intestinal absorption (bioavailability) of calcium*. If the title is relevant click put a **Y** in the first column, if it's not relevant, put an **N** in the column.

RELEVANT? (y / n)	**ARTICLE TITLE**
	Calcium does not inhibit iron absorption or alter iron status in infant piglets adapted to a high calcium diet.
	Calcium absorption from small soft-boned fish.
	Bioavailability and pharmacokinetic characteristics of two monofluorophosphate preparations with calcium supplement.
	Effects of milk and milk components on calcium, magnesium, and trace element absorption during infancy.
	Bioavailability of calcium and magnesium.
	The inhibitory effect of dietary calcium on iron bioavailability: a cause for concern?
	Iron absorption from the whole diet: comparison of the effect of two different distributions of daily calcium intake.
	Strontium as a marker for intestinal calcium absorption: the stimulatory effect of calcitriol.
	Calcium bioavailability and parathyroid hormone acute changes after oral intake of dairy and nondairy products in healthy volunteers.
	Intestinal absorption of calcium from foodstuffs as compared to a pharmaceutical preparation.
	The bioavailability of dietary calcium.
	Searching for the determinants of intestinal calcium absorption

Calculate the precision score for the search that retrieved these titles:

Relevancy = _____

Precision = _____

Now ask someone you know to evaluate the titles. Calculate his scores. Did your judgments about relevancy and the resulting scores agree?