**Saracevic, T. (2008). Effects of inconsistent relevance judgments on information retrieval test results: a historical perspective. *Library Trends,* 56(4), 763-784.**

Keyword: Information retrieval, history of information retrieval

Information retrieval systems first came into existence after the explosion of technical and scientific information that became available following World War II

A solution at the time was a device called a memex, which could be used to store all of an individual’s books, records and communications and could be searched with speed and flexibility. While the memex was never built, the idea of the need for a technological solution to control the information explosion did catch on. It was a motivation for agencies such as the National Science foundation to begin and support information retrieval development and testing.

While information retrieval systems and techniques have drastically changed over the years, the basic definition remains the same. The concept of “specification for search” through the development of algorithms is what differentiated information retrieval from other related methods and systems that preceded it. A second major difference is the concept of that relevance is the basic criteria of the information being retrieved. Retrieval of relevant information or information objects became and still is the primary objective of IR systems. Relevance is the parameter used to evaluate performance of information retrieval systems. As relevance is an outcome of human judgment, however, it is often difficult to use it as an objective, consistent measure.

To this day, graphing of precision-recall figures is an established way to demonstrate and compare performance, and improving on the inverse relation is a major goal of most procedures in IR tests. However, the key issue is still obtaining acceptable human relevance judgments that can then be used as a standard for calculating recall and precision. Information retrieval tests should be as close as possible to the actual, real-life situation in order to have real-life validity. As this is often very difficult to achieve, various

simulation methods have been developed. One method is to have the person who posed

the question make the judgment A second is to have a topic specialist judge. A third

method is to use an information professional, while the fourth method is to simply use a

bystander.

On the historical side, it is quite interesting, if not amazing, to note that the basic

methodological principles and model for testing laid down a half century ago are still

governing IR testing today. IR testing is like a river that became broader and deeper but

never changed its course.