Is User Satisfaction a Hobgoblin?

This paper is in response to William S. Cooper: "On Selecting a Measure of Retrieval Effectiveness."* Whereas Cooper considers (subjective) satisfaction of a user as the utility of an individual search (and then computes system utility as the average search utility), this paper argues that improvement in the task performance of the user is a much more appropriate measure of utility. From this it is shown that recall, while unimportant or even harmful in many search situations, is of vital importance in others. This is in contrast to Cooper's view that recall by itself is not a meaningful measure of system performance at all.

Dagobert Soergel
College of Library and Information Services
University of Maryland College Park, Maryland
20742

• Introduction

In two papers (1, 2) and in a subsequent exchange of letters to the editor (5), Cooper mounts a refreshing challenge to the uncritical adoption or continuation of *recall* and *precision* as measures for the performance of a reference storage and retrieval system. He argues that any performance measure must be derived from the ultimate objective of a reference storage and retrieval system, or to say it in another way, from the benefits to be derived from such a system. This point is extremely well taken and only the difficulties of implementing this principle can explain why, in spite of the obvious, the measures of recall and precision have been and are still used so often without much further reflection.

The crux of the matter is, of course, the definition of a system objective. Cooper concentrates completely on the individual user at the point of his or her interaction with or reaction to the system, and he defines subjective user satisfaction as the system objective. From this Cooper deduces, not always quite correctly, the irrelevance of recall as a performance measure.

This paper suggests a different approach. Instead of leaving the user at the point where he or she receives the information from the reference storage and retrieval system, we should follow him or her to see what the impact is from the information received with the aid of the

reference storage and retrieval system, how it actually affects the user's performance in his or her task (solving a problem, making a decision, etc.). If we do this, evaluation of recall as a retrieval measure becomes a less clear-cut affair.

• A Pointed Summary of Cooper's Approach

The central assumption in Cooper's approach is as follows: the ultimate objective of a document or reference storage and retrieval system is user satisfaction, as subjectively assessed by the user. The satisfaction of a user with the result of a given search, called search utility, can be determined as a sum of document utilities over all documents that the user examines. The utility of an individual document is determined by asking the user: "in effect, how many dollars his contact with that first document was worth to him, where any positive or negative amount is permissible" (p. 90). [Quotations are from Reference *(1)* unless specified otherwise.] A negative answer would correspond to the question: "How much would you have been willing to pay to *avoid* that experience?" (p. 90—emphasis in original).

Cooper's approach to the determination of utility is highlighted by the following quote:

> As explained, the term 'utilities' is used here as a cover term for anything about a docu-

ment that the user values. If a user values a witty style as much as informative content, who are we to say that he is wrong? Systems should be designed so as to optimize the satisfaction of their users, not the information scientists (p. 92).

From this approach, Cooper concludes that the traditional recall measure should be discarded since it takes into consideration relevant documents that were not retrieved and, therefore, were never examined by the user and, as a result, could not possibly contribute to the search utility as assessed by the user. It is worthwhile to quote Cooper's entire line of reasoning on this point:

> Objection 16. *A fair measure ought to take account of the amount of useful material not brought to the user's attention. It is for this reason that conventional evaluation procedures usually penalize a system as a large portion of the relevant documents are not retrieved.*
>
> The involvement of unexamined documents in a performance formula has long been taken for granted as a perfectly natural thing; but if one stops to ponder the situation, it begins to appear peculiar. Why should the status of a document which the user has not examined have any influence on a user-oriented performance rating? A document which the system user has not been shown in any form, to which he has devoted not the slightest particle of time or attention during his use of the system output, and of whose very existence he may be unaware, does that user neither harm nor good in his search.
>
> It is puzzling, to say the least, that the characteristics of that document have conventionally been accorded great significance in evaluating what the system has done for the user. How could the properties of that document possibly matter? Nor does it make it any less puzzling to add that the property these documents should be tested for is their potential *relevance*. The burden of proof should therefore be on those who make this bizarre claim to show why properties of documents that have not been brought to the user's attention should be taken into account in measuring the benefits received from the system (p. 95-emphasis in the original).

In the same vein in another article, Cooper states:

> The trouble is that historically, the unexamined documents have been given a highly prominent role in evaluation - specifically in the denominator of the still popular recall performance measure. I have still to hear a well reasoned argument for why the properties of unretrieved documents, of whose very existence the user is normally unaware, should be accorded such tremendous importance in evaluating how much utility he has experienced as a result of using the system [(3) p. 210].

• The Fallacy of Cooper's Approach

Keeping in mind that strong rhetoric often hides a weak point, we shall now proceed to examine Cooper's approach. Firstly, Cooper's approach is demeaning to the user by imputing to him the attitude "What I don't know doesn't bother me." But often the user wants an assurance that the search was complete, that recall is high; and whatever information he or she can get on that point will enter his or her assessment of the search utility. Secondly, and more importantly, in many situations the user might be perfectly satisfied with the results of a search, but a more complete search (higher recall) would have resulted in better task performance. One can even imagine a situation where the user reacts negatively to the large number of documents retrieved, which results in lower subjective satisfaction or search utility, even though he or she needs the documents for proper task performance.

This brings us to the central point of this communication. The ultimate objective of any information storage and retrieval system is not subjective satisfaction as expressed by the user, but improved task performance/ problem-solving/decision-making by the user. For too long an uncritical attempt "to make the user happy" has been prevalent in the library and information services profession. What is needed instead is an attempt to make the user successful. (This is an important theme in the discussion on whether or not librarianship is a profession. In this discussion the librarian who takes the user's wishes and executes them literally—and thereby gets dependent on the user—is contrasted to the lawyer or doctor who assesses the user's situation and then uses his independent professional judgment as to what actions to take and/or recommend.)

The emphasis on user's task performance has important implications for the "utility" of a document. A user

may be highly satisfied by reading a document, even though reading the document has nothing to do with the task at hand. However, it is questionable whether reading the document will improve the user's task performance. Rather, the contrary will usually be true since the user is wasting his or her time. (The user's supervisor would be far from satisfied with the retrieval of this document.) On the other hand, Cooper includes among the "penalties" that lead to low satisfaction "any unpleasantness occasioned by the content of the document" (p. 90). Considering that in many cases unpleasant information is a very important ingredient in decision-making, this again seems rather unjustified.

It follows from this discussion that the information professional (whether the manager of an information system or a reference librarian/information specialist) must be thoroughly familiar with: 1) the task of the users of this system; 2) the information needed and the information available to contribute to this task; and 3) the information storage and retrieval techniques that can be used to make this information accessible. Based on this knowledge, the information professional must make his or her own independent professional judgment as to what services to provide (on a macro-level) and as to what references or documents (in a reference or document storage and retrieval system) to submit in answering a specific request.

The discussion so far was intentionally one-sided to make a point. In reality, determination of information needs and of services to be provided should be a joint enterprise of the user and the information professional. As is well-known, a dialogue must take place in which information needs are clarified and agreement is reached as to what services would best meet the information needs. This view is very similar to the view expressed by Bates *[(4)* p. 273]. However, there is a difference in emphasis. Bates states: "The librarian knows the resources; the user knows best what he wants."

Whereas there is certainly a degree of polarity in the relationship between information professional and user, we still stress the point that the information professional is not just the recipient of a communication on the information needs of the user but is, or should be, an active participant in the determination of these needs. [Line's (5) definitions of need, want, and demand are also relevant in this context]. As Bates points out, the relationship between an information professional and his or her client is no different from the relationship between, say, a doctor and his or her patient, or a lawyer and his or her client. In these cases, too, agreement of professional and client is needed for treatment or legal action to take place.

It is believed that this discussion clearly demonstrates the fallacy of Cooper's approach. However, the follow-ing hypothetical examples should serve to clarify the matter further.

Example 1: A physician has considered a certain drug for the treatment of a patient. In order to find out more about the exact effect of the drug for the disease of the patient, he requests a literature search. All of the documents retrieved report favorably about the results of the drug considered. The physician is very satisfied with the search results and proceeds to prescribe the drug. Unfortunately, another document was not retrieved and not examined by the physician. It reports on a case very similar to the patient's in which the drug had very serious side effects which outweighed its benefits. No harm done by missing *that* document?

Example 2: A scientist in the Bureau of Foods receives a petition from a food manufacturer to permit the use of a certain chemical as a food additive. The manufacturer submits the results of various tests that show that the chemical is safe for human intake up to a certain tolerance level. Naturally, the scientist does not want to rely solely on these tests, and therefore orders a literature search to retrieve documents that can aid him in his decision on the petition. A number of documents are found that deal with animal experiments involving the chemical in question and also with the use of the chemical as a food additive in other countries. None of the documents report harmful effects and so the petition is approved. Unfortunately, two documents reporting on harmful effects of the substance and ten documents on harmful effects of a nearly related substance were not retrieved and not examined by the scientist reviewing the petition. No harm done?

Example 3: The Atomic Energy Commission must prepare an environmental impact statement as part of the procedure of approving a nuclear power plant. In order to assess the environmental impact properly, it is obviously important to have complete information about the geological, hydrological, metereological, etc., characteristics of the environment of the power plant and also complete information on general questions, such as the effect of higher water temperature on the ecology of a river. The effects of missing a document that reports about a ground water current. from the proposed site of the power plant to a drinking water reservoir could have devastating effects. This example is particularly interesting in that the Atomic Energy Commission was not interested in environmental information in the first place and had to be compelled by law to prepare environmental impact statements. With some overstatement to emphasize the point, we could say that in this case: the less information the user has, the more satisfied he is, but also the worse his task performance becomes.

Example 4: A scientist in a data analysis center has the task of determining the best value for the strength of a certain type of steel so that this value can be entered into the National Reference Data System. The scientist must consider all the measurements of

this particular constant, taking into account the quality of the measuring method used. Again, high recall is important since missing one document that reports on a high quality measurement of the constant in question might lead to a less precise figure cited in the system.

Example 5: A search is to be made in connection with a patent examination. Obviously, any omission could cause tremendous waste of labor and final frustration for the searcher. (This last example leads to the case where Cooper himself acknowledges the importance of recall in what he calls "existence search"; this type of search can be exemplified best by the doctoral student who is conducting a literature search to see whether his or her dissertation topic is already used. Failure to identify work on the topic already done or in progress has very serious consequences.)

As shown by these examples, there are many search situations where high recall is of utmost importance for the task performance of the user. However, there are also many situations where a few relevant documents will give the user all of the information he or she needs to solve the problem at hand, and where reading more documents would only be a waste of time. In these cases, high recall does more harm than good. This statement is true even if one disregards the fact that, for the individual search at hand, high recall usually leads to lower precision, so that not only would there be more *relevant* (but nevertheless unnecessary) documents to be read, but also more *irrelevant* documents.

The situation is complicated by the fact that recall requirements sometimes depend on the collection (Are there documents that contain the same information?); or even on the search outcome (One survey document retrieved can make 100 other documents superfluous, whereas these 100 documents are needed if the survey document is *not* retrieved.) In no case can the requirement of the user be reduced to a statement like: "The user needs 15 documents." Furthermore, many requests now treated as requiring low recall might in fact be requests not just for a few documents, but for a *few good*

documents. If the quality is made an integral component of the search request, the figures for recall and precision might look quite different.

The consequence of these considerations is very simple and known to every good search analyst. The recall requirements for a search must be determined from the task which the search is to support, and this determination cannot be left to the user alone, but must be made jointly by the user and the search analyst.

• Conclusion

The selection of suitable performance measures for reference storage and retrieval systems is more complicated than William Cooper has indicated. To design systems that will maximize an appropriate measure is even more difficult.

References

1. **Cooper, W.S.** 1973. *"On Selecting a Measure of Retrieval Effectiveness." Journal of the American Society for Information Science.* 1973; 24 (2): 87-100.
2. **Cooper, W.S.** 1973. *"On Selecting a Measure of Retrieval Effectiveness: Part II, Implementation of the Philosophy." Journal of the American Society for Information Science.* 1973: 24 (6): 413-424.
3. **Robertson,** S.E. 1974. *"In Defense of Relevance"* (letter to the editor). **Cooper, W.S.** *(authors reply). Journal of the American Society for Information Science.* 1974: 25 (3): 208-209.
4. **Bates, M.J.; Cuadra, C.A.** 1974. *Implications of the Conference."* In: **Cuadra, C.A.; Bates, M.J.** (eds.) *Library and Information Service Needs of the Nation.* Washington, D C , NCLIS, 1974; 265-279.
5. **Line,** M.B. 1974. Draft Definitions: *Information and Library Needs, Wants, Demands and Uses. AS LIB Proceedings. 1974;26(2):87.*