EVALUATION OF INDEXING SYSTEM
Contents

- Introduction.
- Purpose.
- Evaluation criteria.
- Evaluation criteria and its system parameters.
- Recall and Precession.
- Factor affecting recall and precision.
- Evaluation of different types of index files.
- Glossary.
INTRODUCTION

- Evaluation of an information retrieval system essentially means measuring the performance of the system, success or failure, in terms of its retrieval efficiency (ease of approach, speed and accuracy), and its internal operating efficiency, cost effectiveness and cost benefit to the managers of the system.

- In other words, we evaluate a system in order to ascertain the level of its performance or its value.
An indexing system is a sub-system of an information retrieval system and hence, its performance is directly linked with its overall performance of the entire information retrieval system.

An information retrieval system can be evaluated by considering the following two issues:

I) How well efficiently the system is satisfying its objectives, that is, how well it is satisfying the demands placed upon it, and

ii) Whether the system justifies its existence.
INTRODUCTION

- Indexing system is evaluated in order to ascertain the level of its performance or value in terms of its effectiveness and efficiency.

- By effectiveness we mean the level up to which the given system attains its stated objectives.

- In an information retrieval system, the effectiveness may be a measure of how far it can retrieve relevant information withholding non-relevant information.

- By efficiency we mean how economically the system is achieving its objective.
INTRODUCTION

- By efficiency we mean how economically the system is achieving its objective.

- It may be necessary that the cost factors are to be calculated indirectly, such as response time (i.e. that is time taken by the system to retrieve the information), user effort (i.e. the amount of time and effort required by a user to interact with the system and analyze the output retrieved in order to get the required information), the cost involved, and so on.
INTRODUCTION

- Evaluation is generally done for computer-based systems. It can also be possible in case of manual system (Card catalogues) and other types of bibliographical products and services, including printed indexes.
The purpose of evaluation in general is to judge the efficiency of a system in relation to retrieval, identify the shortcomings, if any, rectify them and improve upon the system. Evaluation studies have one or more of the following purposes:

i) To show at what level of performance the system is now operating;

ii) To compare the performance of two or more systems against a standard or norm;
iii) To determine whether and how well goals or performance expectations are being fulfilled;

iv) To identify the possible sources of system failure or inefficiency with a view to raising the level of performance at some future date;

v) To justify the system’s existence by analyzing the costs and benefits;

vi) To explore techniques for increasing performance effectiveness;
PURPOSE (Contd...)

vii) To establish a foundation of further research on the reasons for the relative success of alternative techniques; and

viii) To improve the means employed for attaining the objectives or to redefine the goals in view of research findings.
The major criteria by which any type of information retrieval system are evaluated may be as follows:

i) **Cost**: This refers to the charges at which the service is provided which should be reasonable in relation to the benefits associated with it. Besides cost includes the effort involved in the use of the system, in terms of the time spent.

ii) **Response time**: the average time needed to obtain a response from the system.
EVALUATING CRITERIA (Contd..)

iii) Quality: This refers to Coverage, Recall, precision, Novelty, Accuracy of data.

iv) Cost effectiveness ($cost / quality$): This refers to cost per relevant item and Cost per new relevant item (Novelty-cost ratio).
EVALUATING CRITERIA (Contd..)

C.W. Clever don [1962] identified six criteria for the evaluation of an information retrieval system. They are:

I) **Recall**: It refers to the ability of the system to retrieve all the relevant items;

ii) **Precision**: It refers to the ability of the system to retrieve only those items that are relevant;

iii) **Time lag**: It refers to the time gap between the submission of a request by the user and his receipt of the search results.
EVALUATING CRITERIA (Contd..)

iv) **User Effort**: It refers to the intellectual as well as physical effort required from the user in obtaining answers to the search requests. The effort is measured by the amount of time user spends in conducting the search or negotiating his enquiry with the system. Sometimes, response time may be good, but user effort may be poor.
v) **Form of presentation** of the search output, which affects the user’s ability to make use of the retrieved items, and

vi) **Coverage of the collection:** It refers to the extent to which the system includes relevant matter. It is a measure of the completeness of the collection.
Salton and McGill identifies the various parameters for evaluation of an IRS:

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>System parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall and precession</td>
<td>1) <strong>Indexing exhaustively</strong></td>
</tr>
<tr>
<td></td>
<td>Recall tends to increase the exhaustively of indexing terms.</td>
</tr>
<tr>
<td></td>
<td>2) <strong>Term specificity</strong></td>
</tr>
<tr>
<td></td>
<td>Precision increases with the specificity of the index terms.</td>
</tr>
</tbody>
</table>
EVALUATION CRITERIA AND ITS SYSTEM PARAMETERS. (CONTD...)

<table>
<thead>
<tr>
<th>Evaluating criteria</th>
<th>System parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall and precession.</td>
<td>4) <strong>Query formulation</strong></td>
</tr>
<tr>
<td></td>
<td>Ability to formulate an accurate search request.</td>
</tr>
<tr>
<td></td>
<td>5) <strong>Search strategy</strong></td>
</tr>
<tr>
<td></td>
<td>Ability of the user or intermediary to formulate an adequate search strategy.</td>
</tr>
<tr>
<td>Evaluation criteria</td>
<td>System parameters</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Response time</td>
<td>1) Organization of stored documents.</td>
</tr>
<tr>
<td></td>
<td>2) Type of query.</td>
</tr>
<tr>
<td></td>
<td>3) Location of information centre.</td>
</tr>
<tr>
<td></td>
<td>4) Frequency of receiving users’ queries.</td>
</tr>
<tr>
<td></td>
<td>5) Size of the collection.</td>
</tr>
</tbody>
</table>
## EVALUATION CRITERIA AND ITS SYSTEM PARAMETERS. (CONT'D...)

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>System parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User effort</strong></td>
<td>1) Accessibility of the system.</td>
</tr>
<tr>
<td></td>
<td>2) Availability of guidance by system personnel.</td>
</tr>
<tr>
<td></td>
<td>3) Volume of retrieved items.</td>
</tr>
<tr>
<td></td>
<td>4) Facilities for interaction with the system.</td>
</tr>
<tr>
<td><strong>Form of presentation</strong></td>
<td>1) Type of display device.</td>
</tr>
<tr>
<td></td>
<td>2) Nature of output – bibliographic reference, abstract, or full text.</td>
</tr>
<tr>
<td>Evaluation criteria</td>
<td>System parameters</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Collection coverage.</td>
<td>1) Type of input device and type of size of storage device.</td>
</tr>
<tr>
<td></td>
<td>2) Depth of subject analysis.</td>
</tr>
<tr>
<td></td>
<td>3) Nature of users’ demands.</td>
</tr>
<tr>
<td></td>
<td>4) Nature of core subject area.</td>
</tr>
<tr>
<td></td>
<td>5) Physical forms of documents.</td>
</tr>
</tbody>
</table>
On the basis of all above parameters as suggested by the various studies, the evaluation expert has suggested two basic criteria for evaluation of an IRS i.e. recall and procedure. The most important parameter used for evaluation of an indexing system is recall and procedure:

- **Recall:** Also called sensitivity, it is the ratio between the relevant article retrieved from a system out of the total collection of a system or in other words the ability of a system to retrieve the relevant items out of the total collection.
RECALL AND PRECESSION. (Contd..)

- It can be calculated mathematically by the given formula:
  Recall = Total no. of items retrieved / total collection × 100

E.g. In a library, 40 items are relevant, retrieved out of 200 items
20/200×100 = 20, therefore recall = 20%
Precision: precision relates to the ability of an indexing system not to retrieve irrelevant items. Precision ratio is nothing but the proportion of retrieved items that are relevant.

By it, we measure how precisely an indexing system functions.

It is quite obvious that when the system retrieves items that are relevant to a given query it also retrieves some documents that are not relevant.
RECALL AND PRECESSION. (Contd..)

The precision ratio is sometimes referred to as a relevance ratio.

*Example:*
If a system retrieves 50 documents as answer to a query as output and only 8 of them are relevant to that particular query, the precision ratio is

\[
(8 / 50) \times 100 = 16\%.
\]
FACTOR AFFECTING RECALL AND PRECISION.

- Two performance criteria, i.e. recall and precision are influenced by the following factors:
  a) Indexing policy: exhaustively and specificity;
  b) Requests that imperfectly represent information needs; and
  c) Search strategy.

- The effectiveness of an indexing system is governed by the indexing policy-exhaustively and specificity.
EVALUATION STEPS AND EVALUATION STAGES

- There are a number of distinct steps involved in the conduct of an evaluation programme. An information service can also be evaluated at various stages in its development. The major steps involved are:

1. Defining the scope of the evaluation
2. Designing the evaluation programme
EVALUATION STEPS AND EVALUATION STAGES. (CONTD...)

3. Execution of the evaluation
4. Analysis and interpretation of the results
5. Modifying the system or service on the basis of the evaluation results.
EVALUATION OF DIFFERENT TYPES OF INDEX FILES

- Evaluation of various kinds of index system can be undertaken, such as subject catalogues of library collections, Current awareness services, SDI, Printed indexes etc.

- The methodology does not vary but each situation would warrant different designs.
EVALUATION METHODOLOGY

- An evaluation study is conducted to determine the level of performance of the given system and also to identify those factors that are the reasons for weaknesses of the system.

- Lancaster identifies five major steps involved in the evaluation of an information retrieval system, which are discussed below:
  
a) Defining the scope of evaluation: The first step of an evaluation study involves the designing of a precise set of questions in conformity with the evaluation objectives in order to learn capabilities and weakness of the system.
The questions are usually concerned with one or more of the following:

i) Overall performance level of the system;
ii) Coverage and processing;
iii) Indexing;
iv) Index language;
v) Searching; and
vi) Input procedure and computer processing.
b) Designing the evaluation programme: This step involves the preparation of a detailed plan of action concerning the identification of parameters and procedure for collection of data needed to answer the question (s) set in the definition of scope.

c) Execution of the evaluation: It is the stage in which data (performance results like recall and precision ratios, etc.) are continuously derived in a way prescribed in the design stage.
In most cases, a repeated number of observations are required to avoid sampling error and bias.

Derived data are manipulated and reduced to a form suitable for interpretation and analysis so that it can answer or contribute to answering the questions set in the definition of scope.

The execution of the evaluation is obviously the most time-consuming step in an evaluation study.
d) Analysis and interpretation of results: The success of an evaluation programme rests upon the method of interpretation of results and its accuracy. This stage should begin before the execution stage is completed.

Performance results or data collected on different parameters during the execution of the evaluation programme are analyzed and interpreted in this stage.

Once the data have been manipulated suitably, the evaluator gets a set of results that is to be interpreted in the light of the set of objectives.
Appropriate statistical techniques are applied to the analysis and interpretation of the performance results.

The evaluator might need to conduct failure analysis so as to justify the results and also to suggest improvements.

Lancaster mentions that the joint use of performance figures and failure analysis should answer most of the questions identified in the objectives of the evaluation.
e) Modifying the system in the light of the evaluation results: Finally, the retrieval system is modified, if necessary, in the light of the results of the evaluation study.
CONCLUSION

- In the foregoing rather cursory account, an attempt is made to focus attention on some of the important areas in the evaluation of an indexing system.

- Essentially an evaluation study is a management aid and regarded as a diagnostic activity, intended primarily to identify weakness or limitations in an existing system and to suggest ways in which the situation can improve in the future. Indeed it is an investment for the future.
GLOSSARY

- **Cost-benefit**: It is the evaluation of the worth of the system by relating the cost of providing the service to the benefits of having the service available.

- **Cost-effectiveness**: It is the evaluation in terms of how to satisfy user requirements in the most efficient and economical way.

- **Efficiency**: Being able to achieve a goal with minimum resource.

- **Effective**: Being able to achieve a set of goals.
GLOSSARY

- **Exhaustively**: It is the measure of the extent to which all the distinct topics discussed in the documents are considered.

- **Evaluation**: It refers to the act of measuring the performance of the system in terms of its retrieval efficiency (ease of approach, speed and accuracy) to the users, and its internal operating efficiency, cost effectiveness and cost benefit to the managers of the system in order to ascertain the level of its performance or its value.

- **Coverage**: It is a measure of the completeness of a collection.
GLOSSARY

- **Novelty**: refers to the newness of information supplied by a service and is of obvious importance in the evaluation of current awareness services since presumably, a good current awareness service will bring documents to the attention of users they learn of them by other means.

- **Specificity**: It is measure of the degree of preciseness of the subject to express the thought content of the document.
References:


- Class Notes.

- Ignou notes.