Information-Retrieval Systems: Systems Analysis of Problems of Quality Management

Eduard Sukiasyan, Russian State Library, Russia

Abstract: The structure, functions and quality parameters of Information Retrieval Systems (IRS), factors which effect the indices of the quality of indexing terms: depth, specificity and consistency are discussed with the help of the systems approach. Different factors, such as the technological features of Information Retrieval Languages (IRL), demands on the training and qualifications of the indexer, technology and organization of the indexing process are taken into account. The systems character of quality reveals itself in the course of a discussion of quantitative characteristics of the aggregate quality index, which represents the product of the separate indices contributing to it. IRS quality management is determined by the completeness of our understanding of the systems properties of IRS.

1. Introductory Comments

We do not intend to present a definition here of the quality of information-retrieval systems (IRS). For our further discussion it will be sufficient to proceed from the conventional definition of quality as of a certain aggregate of properties of the object which are necessary and sufficient for the solution of the stated problems. Hence it follows that the problem of quality can be stated only in the case when the problems and functions of the object are established, determined and fixed in advance. If in the process of development of a concrete IRS these or other problems were not stated, then the appropriate characteristics of the IRS will bear no relationship to the evaluation of its quality. In other words, in accordance with the extent of comfort or other complex indicators, IRS can be ranked by classes, but this will not concern the quality of a concrete IRS at all.

Quality usually pertains to countless categories. We must realize that quality is a complex phenomenon which still retains its abstract character, since in respect to the majority of technological processes and intellectual actions it is difficult to provide measuring qualities. However, quality can be measured with the help of the qualimetric methodology. Qualimetric determination (evaluation and measuring) of quality presupposes the conducting of a preliminary analysis of the object, its functions and the demands imposed on it. We see our task in the theoretical analysis of the problem of quality of IRS and would like to stress right away that within the framework of the given report we will dwell only on the most general points, deliberately ignoring details. Our work needs to be further continued, concretized, and expanded.

We shall speak of quality management of IRS, taking library catalogs as a particular case. In order to separate the indexing process from the retrieval process, we
shall not touch here on the technical aspects of the realization of IRS. The point is that in the retrieval process another aspect of the IRS - its effectiveness - comes into play. It goes without saying that the quality of an IRS is connected with its effectiveness, first of all with its functional effectiveness. However, the management of retrieval processes has its own specificities, which do not concern our theme.

For the analysis and solution of problems connected with the functioning of artificial, man-made systems, the methodology of systems analysis can be used. In essence it is made up of the breakdown of the system into elements, the identification of the problems and functions of each element, the establishment of structural-functional connections and the dependences existing between the elements. In the systems approach, the most important thing is the understanding of the totality of an IRS as of a system which constantly functions within the framework of a more sophisticated system: a library or information center. From the quality of the IRS we pass on to its effectiveness, and further on to the effectiveness of the library.

An IRS is a controlled system. As is generally known, management should ensure the normal functioning of the given object. In this way we come to an understanding of the close relationship existing between the norm and quality. In order to attain the qualitative state of the IRS which we call normal, the administration should manage quality.

In the next part of my paper I shall discuss the structure, functions and parameters of the quality of an IRS, and the factors which influence the quality indices.

2. IRS: Systems Analysis of the Quality of its Elements

In the „narrow“ meaning of the word, an IRS (catalogues, files, records) is the result of indexing, a process which is realized by the indexer. On the systems plane, the main elements of this process are: 1) the document as the object of indexing and 2) the information-retrieval language (IRL). To our thinking the problem of the quality of the document is outside the framework of our analysis, despite the fact that the quality of any IRS naturally depends on what kind of documents are fed into the information-retrieval file (are they adequate for the demands of readers and users, were they correctly selected by the selection officer, etc.). However, the management of stock development, in other words of the qualitative parameters of the flow of objects to be indexed and fed into the IRS, belongs to another, adjacent management system.

The quality of indexing finds reflection in the form of indexing terms. In keeping with ISO Standard 5963, the indexing term may be expressed by classification numbers, subject headings, descriptors or keywords. Qualitative indices of indexing terms are known through numerous publications. The majority of specialists hold the opinion that here quality is determined by the following three indices: depth of indexing (the extent of fullness of reflection of the properties of terms in
the process of indexing), specificity of indexing (the extent of conformity of, on the one hand, the volume and content of the concepts expressed by the indexing term with, on the other hand, the volume and content of the concepts characterizing the content of the document) and the consistency of indexing (the extent of coincidence of the indexing terms attributed to documents that are adequate in content).

In order to note that the quality of the indexing term exerts an influence on three parameters, no in-depth analysis is necessary. They are: 1) IRL quality; 2) indexer’s work quality; 3) qualitative characteristics of the indexing process. Let us discuss these in detail.

2.1 IRL Quality

Many serious publications deal with the analysis of different types of IRL (classification systems, thesauri, etc.). An interesting thing here is the lack of discussion in the works: some authors recommend their own criteria and indices, while others supplement and develop their ideas. They well-nigh never argue, nor do they disclaim earlier stated opinions. Obviously from the standpoint of „pure theory“ everything is in order here, but in this connection I want to say a few words from the standpoint of the head of an indexing service. I think that our theoreticians do not analyze the technological aspects of the quality of an IRL thoroughly enough. Usually an IRL analysis is conducted on the level of a system, while the indexers usually do not deal with a system, but with a material carrier, e.g. the UDC classification tables of a certain edition. In such cases the qualitative parameters of a system depend in many respects on organization factors: the effectiveness of the issue and dissemination of extensions and corrections, possibilities of acquisition (by ordering through the library center) of every new edition or of the machine-readable version, etc.

Not all authors of IRL - far from it - are aware what mutually irreconcilable situations heads of indexing services will encounter. Part of the indexers hold the opinion that all extensions and corrections, even those of a most specific nature, should be immediately introduced into the IRL. The other group aims to retain the stability of the IRL for some time and agrees to introduce changes only in exceptional cases. It should be mentioned here that the editors of the DDC have managed to find a golden mean which satisfies all users.

2.2 Quality of the Indexer’s Work

The quality of the indexer’s work is determined by his/her training, qualification level and length of service.

The indexer’s training is an extremely original and sufficiently complex system of knowledge and skills. This training is achieved through formal courses and is supported by continuous everyday study. The demands imposed on the indexer increase year after year, for both the IRL and IRS become more and more sophisticated, and many of these systems are replaced by automated ones. The study of advertisements in which indexers are offered jobs, published e.g. in „American
Libraries" (Amer.Libr.Assoc.), allows me to formulate the demands imposed on
the training of indexers.

1. **General Education.** As a rule a sufficiently high level of general culture, a cer­
tain universality of knowledge, a well-read personality is needed, since all these
things will allow the indexer to view any theme as a part of the system of the
Universe of Knowledge.

2. **Professional training** is usually on the Master of Library and Information Sci­
ence (MLIS) level. We are aware that many specialists consider that the profes­
sional training of which we shall speak later is quite sufficient. However, it is
difficult to imagine that an indexer who does not know the „large system“ identi­
ified with a library or information body, will work efficiently. The indexer must
have a good idea not only of the immediate results of indexing, but of the ex­
tremely remote ones, too.

3. **Linguistic training** plays an extremely important role in upholding the quality of
indexing. In the contemporary research library the indexer as a rule works with
documents written in many different languages. Of course the consultation of ex­
perts in this or that language is not ruled out, but this slows down the normal flow
of the indexing process.
The experience of the department which I represent here shows that length of
service in practical work with documents written in different languages (our in­
dexing service processes documents in more than 30 languages of European coun­
tries) brings with it a special skill. The indexer, usually a professional in some
subject, is proficient in the terminology of his subject field and he independently
analyses the content of the document. When an indexer has worked for more than
10 years in the field, the knowledge of two foreign languages becomes a norm.
Any indexer can continue to learn foreign languages at courses organized by the
library.

4. **Professional training** appropriate to the problems of indexed subject fields is
obligatory in many instances on the Bachelor of Arts or Master of Arts level. The
indexer should understand without difficulty the contents of documents which he
processes, see new problems and know the main scientific centers and the most
prominent researchers.

5. **Special training** is usually realized on the job, and this guarantees the success of
the indexer’s performance in the given organization. It must impart, first of all,
knowledge of the IRL used by the given organization, then knowledge of the tech­
technique and technology of the IRS itself, including the equipment and software. The
practical mastery of indexing technology: the gaining of skill in the analysis of
documents, the proper use of reference aids, etc. is also important.

The level of the indexer’s qualification is indicative of his ability to creatively
exploit his education and „know how“ for attaining a high quality level in index-
ing. A highly qualified indexer works not only quickly and efficiently, he also becomes a consultant for his colleagues. Unfortunately, the length of service and the level of one’s education do not always characterize the real qualification of a worker. The potential possibilities of a worker may reveal themselves not immediately, but only in the course of time. Here we see the influence of such factors as age, sex, family status and previous experience in life. Much depends on the way the indexing process is organized and even on the extent of the department head’s exactingness.

2.3 The Indexing Process: Quality Characteristics
The general principles, content and sequence of operations in the indexing process are set out in ISO Standard 5963 „Documentation - Methods for examining documents, determining their subjects and selecting indexing terms“. In our country, on the basis of this ISO Standard the national standard GOST 7.59-90 „Indexing of documents. General requirements for classifying and subject indexing“ was developed. Somewhat later a similar standard for coordinating indexing was worked out.

Indexing represents a system of consecutively proceeding, interconnected operations. The omission of any element of this system can lead to a violation of the norm, a qualitative violation, first of all. The ISO Standard was supplemented by the „Flowchart of the indexing operation using a thesaurus“. We have developed a graphic algorithm for the process of classifying. Along with the consistency of the main operations, this algorithm also shows such things as: the connections with the reference system (encyclopedias, guide-books, vocabularies, atlases, thesauri, etc.), the possibilities of turning to experts with questions concerning specific fields or rare languages, the necessity for a more precise analysis of the indexer’s preliminary decision with the help of different aids, consultations and files, etc.

In our opinion a special indexing operation should consist in carrying out a verification of the preliminary decision by the indexer himself with the help of catalogues: if the decision is correct, the document will be reflected next to other documents of similar content. Nevertheless, the indexer may be subjective and reflect in his decision his personal position, a thing which is by no means always permissible in indexing. That is why we have introduced the obligatory editing of the decision. The work of the editor, who is as a rule appointed from a group of the most qualified indexers, is called upon to ensure the depth, specificity and consistency of indexing terms. In our rather large indexing service there are about 70 indexers in different subject fields. Editing is accomplished by a special subdivision. Here the indexing term is checked and corrected not only from the standpoint of the rules of indexing, but also from the standpoint of formal criteria. In accordance with the national standard, editing has become a mandatory operation in any case, even when the indexing is performed by only one person who then has to do the editing himself. In this case a temporal interval becomes obligatory; one should edit one’s own decisions only after a short break.
In order that the indexing be of good quality, it is necessary to correctly organize
the process of distribution of documents among indexers and regulate the move­
ment of the document (it frequently happens that one document has to be analyzed
by a number of specialists). Many parameters, such as conditions of labor, the
placing of personnel, the socio-psychological climate in the group and so on should
be taken into account. If we had only indexing in view, all this would be superflu­
ous, but we are speaking of quality, management of quality, and a systems ap­
proach to the problem of management. Here there are no superfluous elements or
secondary details.

3. Quality in the Systems Meaning: A Product and not a Sum

A specific property of quality" is that it pertains only to one characteristic, to one
category of an object. „Too much“ quality in one thing does not compensate for
shortcomings in other things: quality in its quantitative expression is not summed
up, it does not divide and does not form arithmetic means. We can express the
quality of separate elements of the object, for example of the indexing term, in the
form of quantitative indices of depth, specificity and consistency. The indices will
be expressed in the form of a decimal fraction. The norm (standard of quality) will
be score 1, which the decimal fraction is to reach.

The aggregate quality of the object is expressed by the multiplication of indices. It
is therefore clear that the smaller one of them will be in terms of quality, the smaller
will also be the aggregated index. This situation is shown in the following table in
which the letters a, b and c stand for the quantitative expressions of the indices of
depth, specificity and consistency respectively. In this case we conditionally pro­
ceed from the assumption of the equivalence of all these indices.

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>a * b * c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0,9</td>
<td>0,9</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0,9</td>
<td>0,9</td>
</tr>
<tr>
<td>4</td>
<td>0,9</td>
<td>0,9</td>
<td>0,9</td>
</tr>
<tr>
<td>5</td>
<td>0,9</td>
<td>0,8</td>
<td>0,8</td>
</tr>
<tr>
<td>6</td>
<td>0,9</td>
<td>0,8</td>
<td>0,8</td>
</tr>
<tr>
<td>7</td>
<td>0,8</td>
<td>0,8</td>
<td>0,8</td>
</tr>
<tr>
<td>8</td>
<td>0,8</td>
<td>0,7</td>
<td>0,7</td>
</tr>
<tr>
<td>9</td>
<td>0,8</td>
<td>0,7</td>
<td>0,7</td>
</tr>
<tr>
<td>10</td>
<td>0,7</td>
<td>0,7</td>
<td>0,7</td>
</tr>
</tbody>
</table>

The laws of mathematics are inexorable! However, the score 0,7 is not so poor at
all: in 100 indexing terms only 30 do not meet the demands imposed on one of the
indices. Usually such a norm is considered permissible. But look what happens
when there are three such indices: the score 0,343 cannot, of course, be considered
the „norm“.
The table shows that an aggregate index over 0,7 can be ensured only if the three components are not all lower than 0,89 - which figure means that in 100 documents, 89 of them should have no remarks of the editor.

Another important regularity is connected with the quantity of parameters according to which the score is estimated - when the number of parameters increases the problems connected with the assurance of quality will also greatly increase. Thus, the conditionally „permissible“ aggregate score 0,7 with the number of indices equalling seven can appear only if each of them is not less than 0,9.

4. IRS Quality Management: Desires and Possibilities

The general functions of management are widely known. In the management of an indexing service such components of management as planning, accounting, control, selection and placing of personnel, etc. fully retain their importance. Here we are interested in those aspects of management which are directly connected with the management of IRS quality.

Let us assume that we are concerned with an indexing service which works in the mode described and within a functioning IRS. In this case the problems of operative management have a more concrete character: they consist first of all in an analysis of the functioning of the system, the diagnostics of disharmonies of different kinds, of violations and mistakes occurring in the normal regime. It is precisely here that we must see the entire multi-stage system of feedforward and feedback, reasons and consequences, conditions and dependences. It is only after we find the unit which produced the mistake or which led to an error on a higher step of the system that we can take this or that management decision.

The modern IRS is the result of a man-machine interface. When it is a question of technology, no problems of diagnostics usually appear, and if they do appear they are solved with the help of appropriate measuring instruments or appliances. It is much easier to manage a machine than a person because with a machine we can estimate, formalize and give an algorithmic presentation of any management action. However, when we deal with a human being (and up till now only a live person operates as an indexer), it is difficult to speak about algorithms. In this case management depends on whether the head of the service will be able to identify the factors responsible for a drop of the quality of indexing and locate the unit in the system with the help of which the qualitative balance can again be restored. Here, too, we can find good analogies in the search for defects in the scheme of a radio-electronic apparatus. In this scheme, as we all know, thin wires connect hundreds of small details which function within the framework of a well-regulated system. A similar system creates the IRS and forms its qualitative parameters. It is evident that the effectiveness of IRS quality management is determined by the study of the entire problem on the systems level, by the completeness of our understanding of the elements of the system, of its general structure, the functions of each subsystem and of the system connections and relationships existing within the framework of the system.