We will note also that these models allow for the construction of a large number of complex search characteristics.

The model we start with was suggested and described in detail in Voiskunskii (1980, 1983, 1987, 1992) and led to the construction of a large specter of diverse CSCs. Now, let us assume that the search is performed in the collection containing $N_0$ documents. We will also assume that the documents retrieved in the search process are graded "1" and the other documents of the collection are graded "0." Finally, we will assume that the documents in the collection have been analyzed by the user and grade "1" was assigned to those that, in the user's opinion, should be included in the output (i.e., are pertinent); the remaining documents are graded "0." We will illustrate this as follows (Figure 10.7).

Thus, we can see that, in accordance with the assumptions, as the result of the search based on a user's search request and the user's analysis of the search collection, two sets of evaluations are produced, namely, a set of evaluations of "output/nonoutput" documents and a set of evaluations of "pertinent/nonpertinent" documents. It is clear that if we had an instrument to evaluate how close (that is, how much alike) the produced sets of evaluations are, that instrument could have been employed to evaluate the quality of the output resulting from the search. This leads to an idea that forms the foundation for constructing the CSCs based on the model under discussion: to assign the role of complex search characteristics to instruments that make it possible to evaluate the closeness of the set of the grades "assigned" to collection documents after the search to the set of grades "assigned" to the collection documents by the user.

**a**

| Document | Document grade assigned as a result of the retrieval |
|---|---|
| $D_1$ | 1 |
| $D_2$ | 0 |
| $D_3$ | 0 |
| $D_4$ | 1 |
| • | • |
| • | • |
| • | • |
| • | • |
| • | • |
| $D_{N_0-1}$ | 0 |
| $D_{N_0}$ | 1 |

**b**

| Document | Document grade assigned by the user |
|---|---|
| $D_1$ | 1 |
| $D_2$ | 0 |
| $D_3$ | 1 |
| $D_4$ | 0 |
| • | • |
| • | • |
| • | • |
| • | • |
| • | • |
| $D_{N_0-1}$ | 1 |
| $D_{N_0}$ | 1 |

**Figure 10.7**
Comparison of "states" of the document collection. (a) The state of document collection after retrieval; (b) The state of document collection after documents are analyzed by the user.



We present this now in a more formalized fashion. Let us assume that $\nu_i$ is the correspondence coefficient of the $i$-th document of the search collection ($D_i$) to the query of a certain user, the coefficient calculated based on the search results, and $k_i$ is the correspondence coefficient of the same document to the same query calculated outside the search, for instance, by the user. Also assume that $k_i$ and $\nu_i$ are determined by a binary scale; that is, $k_i = 1$ if the $i$-th document of the search collection is pertinent, and $k_i = 0$ if it is not, and $\nu_i = 1$ if the $i$-th document of the search collection was found during search, and $\nu_i = 0$ otherwise. If $k_i$ and $\nu_i$ are determined for each document of the collection of $N_0$ documents, they can be used in forming the following vectors:

$$K = (k_1, k_2, \ldots, k_{N_0}) \text{ and } V = (\nu_1, \nu_2, \ldots, \nu_{N_0}).$$

Then the complex search characteristics are introduced as functions that enable one to evaluate the closeness of these vectors.

It would be most natural, in our opinion, to use the following functions for this purpose (see, for instance Kolmogorov & Fomin, 1968):

$$1.\ \cos \phi_{XY} = \frac{\sum_{i=1}^{N_0} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{N_0} (x_i)^2} \cdot \sqrt{\sum_{i=1}^{N_0} (y_i)^2}},$$

$$2.\ \rho(X, Y) = \sum_{i=1}^{N_0} |x_i - y_i|$$

$$3.\ \rho_1(X, Y) = \sqrt{\sum_{i=1}^{N_0} (x_i - y_i)^2},$$

where $X$ and $Y$ are vectors such that $X = (x_1, x_2, \ldots, x_{N_0})$ and $Y = (y_1, y_2, \ldots, y_{N_0})$; $\cos \phi_{XY}$ is the cosine of the angle between vectors $X$ and $Y$, $\rho(X, Y)$ and $\rho_1(X, Y)$ are the functions that specify the distance between the vectors. Before applying these functions to vectors $K$ and $V$, we will note that the number of 1s in vector $K$ is equal to the number of pertinent documents in the search collection, that is, to $C$. Consequently,

$$\sum_{i=1}^{N_0} (k_i)^2 = C.$$

The number of 1s in vector $V$ is equal to the number of documents in the output, that is, to $N$. Consequently,

$$\sum_{i=1}^{N_0} (\nu_i)^2 = N.$$