combined only by the operator AND, and he gave an example of such a query formulation:

$$+T_1 \wedge +T_2 \cdots \wedge +T_n. \tag{1}$$

(Note that $+T$ symbolizes descriptors from pertinent documents.) In another case, for "expanding the previous search" he suggested that all descriptors from the same documents be combined only by the operator OR when the query formulation appears as follows:

$$+T_1 \vee +T_2 \cdots \vee +T_n. \tag{2}$$

Checking his approach experimentally the author wrote, "Eight efforts to use query formulation (1), consisting of terms only from pertinent documents in a conjunctive search, ended in a failure. Whenever query formulation (1) was used anywhere in the iterative search process the next search usually retrieved zero documents." In another case he said, "In 7 attempts query formulation (2) retrieved a mean of 861% more documents than their previous queries."

Such results, understandably, do not suit even the author himself. In an attempt to improve them, Rickman therefore suggested using negatively assessed documents. In other words, for such descriptors use should be made of the BUTNOT (~) operator, "which will be equivalent to a set difference operator." To this end he set out to construct a query formulation in the following manner. To the query formulation for which the previous search was performed, Rickman added query formulation (1) either with the aid of the AND operator or query formulation (2) with the aid of the OR operator, and in both cases the BUTNOT operator helps "subtract" all the descriptors from the negatively assessed documents. Because this approach also failed, the author tried to use only nonpertinent documents by "subtracting" all their descriptors from the previous query formulation. Though this attempt was just as unsuccessful, the result nevertheless was not as negative as before. On this score Rickman wrote this conclusion:

> The refining query formulation was basically a negative feedback process and reduced recall more than desirable. A searcher could manually outperform the automatic technique since he would not be restricted by any fixed format for his reformulated query formulation. Having to choose any fixed format for an automatic technique appears to be a distinct disadvantage. The technique as implemented was relatively unstable as to the number of documents retrieved. (Rickman, 1972)

The work of Rickman is interesting, but not in terms of a practical method for the automatic construction of query formulations. What is important is that this is the first attempt to use only nonpertinent documents (nonrelevance feedback). It is also sufficiently illustrative of complexities encountered in the problem of automating query indexing.

In 1971, we published a somewhat improved version of the M-algorithm (Voiskunskii & Frants, 1971). It provided for the ranking of subrequests consisting of two and more descriptors. To this end, all subrequests that appeared in a given zone were assigned a weight corresponding to the number of the documents of the marked set having descriptors from the "weighed" subrequests in their document profiles. This helped us avoid including in query formulation those subrequests whose descriptors never appeared jointly in the marked documents (such subrequests weighed zero). This certainly enhanced the algorithm's performance (which was very important in the early 1970s) without practically affecting the search quality.

Even more substantial improvements were proposed in papers published later that decade (Voiskunskii & Frants, 1974a; 1974b). In suggesting an improved algorithm, we partitioned the descriptors into zones, based not on a descriptor's occurrence frequency in the document profiles of the collection of documents (as was done in the M-algorithm) but from its search significance (inverse documents frequency) calculated with the use of the formula presented earlier, which in the 1974 works looked as follows:

$$\Psi_i = \frac{r_i \cdot N}{n \cdot R_i}. \tag{3}$$

where

r_i = occurrence frequency of the i-th descriptor in document profiles of the marked set of documents;
n = number of documents in the marked set;
R_i = occurrence frequency of the i-th descriptor in the document profiles of the entire collection; and
N = number of documents in the collection.

We give this formula once again because it is in this form that it has been used in various publications for more than 20 years. In constructing an algorithm published in 1974, we tested several propositions discussed in the paper. For example, the following proposition looked rather attractive:

> A set of descriptors from the combined set of descriptors from the document profiles of marked documents may form a subrequest, provided the relative occurrence frequency of this set in the documents of the marked set exceeds several times an occurrence frequency of the same set in the documents of the entire collection.

Still the proposition was not used for constructing an algorithm, though the method for calculating the significance for a search of descriptor combinations (the method of calculating Ψ for descriptor combinations) seems very attractive. The point is that during the work of the algorithm, an extremely large number of search operations has to be completed to calculate the frequency of the joint occurrence of descriptors. We noted in this context that if proceeding from the frequency or relative frequency of the occurrence of descriptors in the collec-