Experts at IBM reasoned that appropriate values of G for different systems varied within 15% with a lower limit no less than 25%. For example, in a system where the degree of the coincidence $G \geq 30\%$ is assumed sufficient and the size of the document profile is twelve descriptors, any document with the document profile containing at least four (any) descriptors coinciding with those from the set representing the request (from the query formulation) is considered as fitting the selection criterion. If we now return to the second request of the preceding example (with query formulations containing nine descriptors), we will see that the use of such (less stringent) selection criterion (the document profile is not required to contain all nine descriptors—any four will suffice) reduces the number of documents that are present in the system but are not found (specialists would say that "losses are reduced") and, as a rule, improves the retrieval results.

Thus, the introduction of the partial match criterion reduces losses. But what about noise? How can one reduce it? Noise reduction is what the second direction mainly seeks to achieve. Indeed, if one constructs the descriptor set representing a search request of only those descriptors with conditional equivalence classes including words of the query, one would frequently encounter situations similar to that produced by the first request in the example (where the query formulation contains only one descriptor). This means that the criterion developed by IBM (again assume $G \geq 30\%$) will yield only those documents with document profiles containing three or fewer descriptors. It is clear that the system may contain no such documents and hence no documents will be selected. Yes, noise is reduced (in some cases down to zero), but one can easily imagine the user's reaction. Therefore, experts (responsible for constructing query formulations in the IR systems) proposed in some cases to include in the query formulations descriptors that were not directly derived from the search request but were somehow related to the request. This will increase the size of the query formulation above some reasonable threshold. Such an approach will reduce noise; for example, the query formulation corresponding to the first request will be augmented by other descriptors that are required to be present in the document profile and hence the set of selected documents will be smaller. This approach was combined with the partial match criterion allowed, in many cases, for the attainment of an acceptable retrieval quality. It was done because the query formulation already contains more than one descriptor, and the value $G \geq 30\%$ in such a query formulation means the use of more representative—that is, more representative than one consisting of three or fewer descriptors—document profiles (which is more realistic in functioning IR systems). The search in the collection with more realistic document profiles assumes the use of at least two descriptors in the query formulation, which are used together during search. The analysis of the following IBM criterion better explains several features of information retrieval.

One can easily see that for $G \geq 30\%$, any document with a profile containing twenty-seven descriptors can be found only with a query formulation

containing at least nine descriptors all coinciding with those in the twenty-seven descriptors of the document profile. Obviously, such query formulations are not so common, and hence the search will result in many losses. In other words, this feature of the criterion means that the more aspects are discussed in the document, the less the probability of finding it. Mainly this concerns large manuscripts, textbooks, and so on. Also the requirement to match the large number of descriptors in the case of multiaspect documents implies that the value of an aspect detailed in a book represented by thirty descriptors is considerably less than the value of the same aspect in a short report represented by six descriptors. Thus, given the search request, the probability of yielding the report and "loosing" the book is rather high. More detailed analysis of this criterion shows that, for instance, the set of three descriptors representing a certain request generates seven subsets, which will be used in the retrieval process. During such a retrieval, one subset (only one) containing three descriptors will be used to search among documents represented by sets containing seven, eight, nine, and ten descriptors, subsets (three) containing two descriptors (among sets containing four, five, and six descriptors), and those (also three subsets) containing one descriptor (among sets containing three and fewer descriptors).

To a certain extent, this analysis characterizes the main directions in the efforts to create IR systems. The consideration of early attempts to improve retrieval results shows that developers concentrated mainly on descriptor sets representing search requests. Virtually all modifications have been proposed for these particular sets leaving aside sets representing documents. This partly comes from the fact the number of search requests is much smaller than the number of documents, so it is more convenient to change hundreds of requests than thousands of documents. However, another important factor is the understanding that each request is unique and, as a rule, does not exactly describe (reflect) the IN it presents to the system, while documents are written for a variety of users, and hence it is not efficient to adapt them for a single user.

One of the most important ideas contributing to the creation of successfully functioning IR systems was to represent a query formulation not in the form of a single unordered set of descriptors but in the form of a set of such sets. This idea was based on the understanding of several features and properties of natural languages and the representation of the same meaning by different words. As an example, consider two real search requests given to the IR system on information science: (1) "Algorithmization of the construction of query formulations" and (2) "Automatic methods of search requests translation from natural language to IRL." One can easily see that these requests reflect similar IN but include not only different words but also a different number of words. Moreover, after translating them to IRL, one may obtain different descriptor sets and, as a consequence, different retrieval results. A similar effect on the search results may be due to another well-known factor—the possibility of expressing meaning by a different numbers of sentences.