allow for the variation of sizes over wide range, and a single-descriptor set is as valuable as descriptor sets of any other size. Consider the following example. Assume that from a given search request an IR system has generated two descriptor sets, S_1 and S_2, to perform the retrieval. Assume also that the set S_1 contains only one descriptor, a ($S_1 = \{a\}$), and the set S_2 contains three descriptors, b, c, and d ($S_2 = \{b, c, d\}$). Then, using the disjunctive normal form, the query formulation can be written as follows:

$$a \lor (b \land c \land d).$$

It is clear that both sets "have equal rights"; that is, a document matching any of them is formally considered as required by the user and is included in the selection. In essence, this corresponds to the feature of the natural language mentioned previously: each set from the query formulation accounts for a different lexical expression of the same meaning. It is worth stressing that alternative sets constructed are unique; that is, in one query formulation there should be no descriptor sets being subsets of other sets.

As mentioned earlier, the Boolean criterion is utilized in a majority of practically functioning IR systems. Therefore, in Chapter 7 we will consider in greater detail the methods of constructing query formulations for systems using this criterion.

Thus one can conclude that the main direction in the IR system development based on the approach to meaning representation offered by Taube includes the query representation in the form of unordered descriptor sets and the use of Boolean criterion. However, in parallel, investigators have been attempting to change the semantical component of IRL. The use of weights seems to be the most interesting approach in this direction. In this approach, the developers considered the quality of the meaning representation in the form of unordered descriptor sets insufficient. They have proposed to characterize descriptors by the extent of their importance in the text, that is, by their weights. The more important descriptor, the larger its weight. Thus two sets containing the same descriptors can represent the contents of the corresponding texts in different ways, and the search results can be different too. Similar to other directions in IR development, investigators have concentrated on the representation (here, with the aid of weights) of search requests. Obviously, such an addition to the meaning representation should be accounted for by the selection criterion. Next we discuss several, better-known criteria of this kind.

It seems that one of the first weighting criteria was developed by IBM almost simultaneously with the partial matching criterion previously described above (Ofer, 1964). When constructing the query formulation (building the set of descriptors representing the search request), the user evaluates each descriptor of the request and assigns to each descriptor a weight coefficient. A special standard scale of points is used for this purpose. This scale contains both positive and negative values of importance. Besides, the user can specify a certain mandatory numerical value measured in conditional units (points), which a document should reach to be selected. A document is considered as selected (found) if the sum of weight coefficients of the query formulation descriptors coinciding with document profile descriptors is no less than the user-specified value. Let us illustrate this criterion by the following example.

Assume that an IR system uses the 18-point weight scale with 9 positive and 9 negative points (this particular scale has been used by IBM). Assume now that the user has selected the descriptor set, $W = \{a, b, c, d, e, f\}$, for the query formulation and has assigned the following values to the descriptor weights: $a = 7$, $b = -4$, $c = 2$, $d = 9$, $e = 1$, and $f = 6$. Assume also that the user has specified the threshold value, $P = 16$, for documents to be selected. Then the document is considered as found if its document profile contains such descriptors of the query formulation that the sum of their weights is no less than 16 (Kraft, 1963).

It is worth noting that deeper analysis of this criterion shows its similarity with the Boolean selection criterion. Indeed, in the preceding example the document is selected if its document profile contains descriptors (a) and (d) and does not contain (b), or if it contains descriptors (a), (d), and (f) together with (b). Other adequate document profiles may include (a, c, e, f) without (b); (d, e, f), again without (b); and so forth. This can be written as $(a \land d \land \sim b) \lor (a \land d \land f \land b) \lor (a \land c \land e \land f \land \sim b) \lor (d \land e \land f \land \sim b) \ldots$ where $\sim$ denotes the Boolean logical NOT. This expression differs from previous Boolean expressions by the presence of this particular operator.

Such a similarity with the Boolean criterion has led some investigators to assert that all weighting criteria are just Boolean criteria represented in a distinct form. This is, however, not true. There are weighting criteria that cannot be used (represented) in the Boolean form, and the retrieval cannot be classified as Boolean. The most well-known criterion among these is that developed by Salton (1971) and used in the SMART system. We will now consider this criterion in more detail.

Although in the majority of approaches only search requests were indexed using weights, a number of methods included weighting of both requests and documents. As an example, consider the weighting criterion used in the IR system developed by the U.S. Department of Interior, Bureau of Reclamation (Hilf, 1963). In this system documents are represented by 15 to 25 descriptors each. Descriptors which represent the main contents of the document most adequately (in the opinion of the person who performs indexing) are marked by a star (no more than four descriptors will be included in this category). The user also marks no more than four descriptors, which are most informative in the search request. The developers of this system used the following selection criterion: either (1) at least one descriptor from the document profile coincides with at least one descriptor from the query formulation and at least one of these is marked with a star or (2) at least three descriptors from the document profile coincide with those from the query formulation (marking in this case is ignored).

Earlier we have mentioned the weighting criterion used in the SMART