rection essentially depends on successes achieved in such areas as linguistics and artificial intelligence (AI). That is why it seems useful to present here several well-known ideas in the development of the semantical components of IRL. But first let us see what resources have been considered by the developers using the representation of texts in the form of an unordered set of descriptors, which represents an indirect approach to the development of the semantical component (it was often done unconsciously). In addition to allowing us to present ideas used in many existing systems, it will also open the possibility of evaluating the effects of changing the semantical components of IRL in an explicit (conscious) form.

The main efforts in developing Taube's approach have been concentrated on the selection criterion. Recall that the following approach was proposed initially: a document is considered selected when the set of descriptors representing a given search request is contained in the set of descriptors representing this document. Obviously, the representation of document and request contents in the form of unordered sets of descriptors is quite sufficient for such a selection criterion.

Despite the usefulness of the underlying idea of the method—it allowed information retrieval to be formalized—retrieval results obtained with this criterion were frequently inadequate. One of the main problems was that the number of descriptors representing search requests varied over wide range. To illustrate the negative consequences of such a variation, consider the following two real search requests submitted to an IR system with a collection of documents in the area of information science (the IR system was developed in VINITI, Moscow, in the early 1970s):

1. Study of the information need (IN).
2. Automatic methods for correcting query formulations in a descriptor documentary IR system using the selection criterion in the Boolean form.

Considering the descriptor dictionary of this system and the fact that each of these requests will be represented as a set of descriptors with conditional equivalence classes containing words from the requests, the query formulations will be as follows. The first query formulation will be represented by the set containing one descriptor, "information need." The second one will be represented by the set of nine descriptors, namely: "automation," "method," "correction" (this word belongs to the same conditional equivalence class that contains the term "feedback"), "query formulation," "descriptor," "document," "IR system," "selection criterion," and "Boolean logic." If the system uses the selection criterion mentioned here, the retrieval based on the single descriptor, "information need" (first query), will, as a rule, provide a large number of documents. Among these will be a considerable number of noise documents, because all documents using the term "information need" in discussing some other ideas (not only those concerning the study of such a need) will be retrieved.

When considering the second search request, one finds the opposite situation. If we try to find documents with representations containing all nine of the descriptors listed, we will most likely find very few documents; that is, a number of documents important for the user will not be found. Moreover, in many systems the average number of descriptors in sets representing the contents of the documents (in document profiles) does not exceed twelve (this average is typically based on the system's design). This implies that the retrieval to the second query may give no results at all. In other words, there may be a document with the profile containing eight descriptors all coinciding with those out of the nine descriptors listed, but the document will not be selected.

Developers of IR systems very early noticed all of the negative factors resulting from the use of the previously mentioned selection criterion. This stimulated an intensive search for solutions capable of improving retrieval results. Eventually, these efforts concentrated on the following directions:

1. Modification of the selection criterion.
2. Modification of the rules for constructing query formulations.
3. Modification of the explicit ("realized") semantical component of IRL.
4. Various combinations of items 1, 2, and 3.

Actually, these directions correspond to ways of improving IR systems in general, and not only systems based on Taube's approach to the meaning representation. However, because we are considering these particular systems we will be interested in the first two directions (separately and jointly).

As we have noted, the efforts of many investigators were aimed at finding more effective search criteria. In this area the first idea was to use a partial (and not the complete) match of the descriptor set representing the request with that representing the document. At least, this appeared to be a remedy for many of the negative situations just illustrated (with a query with nine descriptors). Various approaches were proposed. For instance, in the IR system developed by the IBM Corporation the following selection criterion was used (at the end of the 1950s and the beginning of the 1960s) (Tritschler, 1962). The degree to which the document profile with the query formulation coincide (i.e., the degree to which the descriptor set representing the document matched that representing the request) is denoted by G. This value was measured in percent. The total number of descriptors in the document profile was denoted by d, and i was the number of document profile descriptors coinciding with those of the query formulation. The formula for G read as follows:

$$G = \frac{i}{d}100\%.$$