file. Moreover, in some cases the main content of the document is beyond the thematic orientation of the system (and, consequently, is beyond the spheres of interest of the users), and only side aspects of this document are of interest. That is true, but nevertheless many researchers reason that one should not use these situations as a base, not because such situations are not essential but because researchers do not have a choice. This is due mainly to the following reasons. First, when a document is being indexed, it is really not clear which descriptor will be most appropriate for the purposes of retrieval. Second, it is not clear whether a descriptor introduced into the document profile will be as good for future system users as for present ones. Third, it is not clear how the retrieval "usefulness" of a descriptor will depend on the query (that is, it is not clear which query, or set of queries, should be used to determine the usefulness of each descriptor included in the document profile). Certainly, this is only a partial list of reasons. But they are sufficient for researchers to reject the orientation of choosing and using descriptors that could be most appropriate for the purposes of a search. This does not mean that researchers are not interested in the quality of the search. In essence, researchers try to find the most promising search descriptors implicitly by choosing descriptors that are most important for the representation of the document's meaning. They assume that these descriptors will allow a good quality of search. Although its shortcomings are obvious, this approach is universally accepted, and at present no reasonable alternatives exist. It is quite evident that the practical implementation of this approach is characterized by its own complications. Indeed, it is not easy to find the most important descriptors in the document text. However, in this case one does not need to consider system users and their queries. In the framework of this approach, the attention is concentrated on the document.

We would like to emphasize here that, when speaking about the quality of indexing, researchers imply the quality of the actual search performed by the system. However, the obtained output and its quality do not entirely depend on the indexing of documents. The document indexing subsystem is only one element of the IR system. All other elements also influence the output. In this sense it is worth saying a few words about well-known experiments that were conducted to evaluate the quality of indexing methods. Although the indexing process itself is aimed at selecting the most important descriptors in the text (from a certain point of view), it is evaluated according not to the descriptor importance but to the quality of the retrieval performed by the entire system. By evaluating the output results, we can draw conclusions about advantages of one indexing method over another. Such conclusions should be treated with caution, however, because they only mean that given a certain composition of subsystems, the given indexing method appears to be better. They do no mean that if, for instance, the query formulation construction method is changed, the same document indexing method will again be the best. Thus, with the development of other elements of the system, many conclusions based on a number of well-known experiments would have to be reevaluated.

In summary, we note that indexing is the process of translating texts from natural language to IRL, and within the framework of the Boolean search criterion, the translation results consist of an unordered set of descriptors. As mentioned earlier, IR system developers, by aiming at quality information retrieval, are oriented toward introducing the most important descriptors (from a certain point of view) into the document profile during document indexing. But how can the importance of a descriptor meaning be determined automatically (algorithmically)? This is quite difficult to accomplish, even when attempted by a human intellect. However, we will consider this point next.

6.3 Main Directions in Automatic Indexing

When speaking about the importance of a descriptor, we mentioned that different points of view on the subject exist. We now discuss these points. Two approaches are used to determine the importance of a descriptor in terms of meaning. Those who adhere to the first approach assume that a descriptor in a document profile should most completely and accurately represent the very meaning of the document being indexed. Moreover, they frequently state that the descriptor should represent the main meaning, or the main theme, of this document. It is clear that the term "the main theme of the document" is rather subjective; for instance, the author and the user of the document may have different opinions on it. However, when developing automatic indexing methods, researchers do not use real opinions (sensible evaluations), but various formal criteria. This implies that any term contained in the document may be judged as reflecting its main meaning providing that it satisfies certain criteria (for instance, it reaches a certain threshold value present by the IR system developer). In other words, the vocabulary content in such systems is not fixed; it depends on the vocabulary content of the documents being indexed.

There is no universally accepted opinion about how well one can determine the importance of terms contained in the document using formal criteria. This is the subject of ongoing discussions that deal with a far broader scientific area than that covered by IR system development and even information science itself. Nevertheless, the absence of consensus on this question does not stop the efforts to use such formal procedures to develop automatic indexing methods. The majority of the literature about indexing is devoted to this particular investigative line.

It is worth noting that the entire investigative line, aimed at determining terms that most appropriately represent the meaning of documents being indexed, is of an experimental and theoretical nature rather than a matter of practice. Maybe that is why many works mainly focus on the problem of determining the importance of terms in a given document, while the information retrieval problem is somewhat of a second priority. In these works, indexing and