

structing query formulations capable of finding a given number of documents, and the found documents will be the best (according to the chosen criteria) for the user. This feature is provided by the selection algorithm for query formulation, which is a part of the algorithm for constructing query formulations.

Returning to those requirements that we are going to take into account, we note that the requirement for the first task provides for a more "narrow" search; that is, in this case the output should be relatively small. As for the requirements for the second task, they provide for a "broader" search; that is, output resulting from such a search would be considerable. So, both requirements may contain either a user-specified desired number of documents in the output or, when this number is not specified by a user, some predetermined number specified by the system's developers as a standard parameter. Such a number for the requirement to the first task may be 10, whereas for the second task the number may be 100. The existing numbers are specified for the automatic zone selection algorithm following which the algorithm for the automatic construction of query formulations will generate two query formulations: one will form output consisting of the 10 best documents and the other will form output consisting of the best 100 documents. Note especially that the requirements in the described case are automatically taken into account.

12.5

Future Developments in Considering IN Components

The direction of research that focuses on improving the quality of the satisfaction of IN by means of the IR system represents a wide area for future studies. This section will briefly list only basic problems, whose solution will contribute to the development of improved IR systems. First we note that it will be necessary to continue the study of IN components—their properties and characteristics. This is clear because more profound insight into IN will not only introduce new elements into the IR system structure but will also change the meaning of some presently known approaches used in the systems.

Of interest are developments in the field of creating lists (tables) of tasks and requirements. To achieve better results, the creation of lists should be based on sufficiently effective procedures or algorithms. Undoubtedly, the development of these means will contribute to the success of the discussed approach.

Another important challenge is the development of methods (primarily, automatic) that take into account the effect of complementary IN components on the results of information retrieval. The optimal solution seems to be the development of an algorithm that takes into account each requirement on the list. Obviously, this is yet another direction in the development of the design of IR systems.

It should be noted that we have not explicitly dealt with a very promising direction—the development of user-friendly interfaces. However, both the orientation of this book toward helping the user to best express his or her IN and the cited methods that allow the system to take into account the complementary IN components are undoubtedly steps along this path. Moreover, the selection of methods mentioned could be an important element of future interfaces or at least could considerably stimulate their development. For this reason, future studies directed toward using complementary IN components in the process of user-system dialogue (the feedback process), in addition to their own value, are also important in the development of user-friendly interfaces.

In summarizing our discussion of the considered direction in developing IR systems, we should emphasize that the approach described is aimed at improving the quality of IN satisfaction. Moreover, this approach illustrates an important point, mentioned before, namely, satisfaction of IN is improved when its properties and characteristics are taken into account. Therefore, we took a closer look at the notion of IN. As a result, IN is presented as an object composed of different elements—not only of a thematic "plane" that the user is interested in (the thematic component) but also of other "planes," for example, the user's goal (the goal component). This observation led to the development of the IR system's structure, which included a list of tasks and requirements for them. Again we should point out that this approach can be used successfully with existing IR systems.

12.6

Conclusion

A future development of IR systems is of interest to everybody who works in the field. In this chapter we considered only some directions in this development, and the only reason we discussed AI as one of the directions is because it has attracted a lot of attention in recent years. In spite of a lack of any serious results in using AI in information retrieval (and we are not optimistic about any progress in the near future), attempts to use ideas from AI allow us to better understand the problems that arise in information retrieval, enhance our understanding of the language, and stimulate the creative activity of researchers.

The most common discussions of information retrieval in the AI literature papers dealing with free text search which, as was mentioned earlier, is directly related to NLP. In Chapter 6 we showed that in these papers the descriptor dictionary is constructed from all the terms appearing in the analyzed document. Such dictionaries are called uncontrollable. Although some authors assert that they use natural language, this is no more advanced or deep than using natural language when constructing a controllable descriptor dictionary. In