descriptor languages. Obviously, specially formed words (descriptors) needed to be "memorized," that is, to be fixed in dictionaries specially prepared for this purpose in concrete IR systems. The latter circumstance was especially important because a certain word could be a descriptor in one system (and entered into the dictionary) and might not be a descriptor in another system. Such dictionaries were later called *descriptor dictionaries,* and the systems for which they were used were called *descriptor systems* or *controlled vocabulary systems*. The terminology presented here has been widely adopted, although many specialists do not use it. In some books devoted to IR systems, one would not find the word "descriptor" at all. Their authors use such words as "term" or "keyword." However, we note again that the terminology we use has been widely adopted and therefore can be regarded as universally accepted.

Already at the initial stages of IR systems creation, it became evident that (for the purposes of computer implementation) it was more convenient to use a certain compact code of standard length as a descriptor, rather than a word or a word combination of the natural language. A part of the alphabet, decimal numbers, or combinations of characters and digits were frequently used as descriptors. Even Mooers' technical implementation of IRL (on special cards) included descriptors represented by randomly chosen codes, each consisting of four groups of digits. The choice of the way to represent (to write) descriptors is not a great problem for developers. It is easily resolved in each concrete case. Eventually, their major concern is what "block of meaning" should an IRL word represent; that is, what meaning should be assigned to a descriptor? According to Taube's approach (discussed earlier), one can assume that the meaning of the text is contained in the words of this text, while the meaning of each idea in the text (a "block of meaning") is represented by a certain nonempty set of these words. Therefore developers decided not to supply each descriptor with an explanatory article but to use some set of words that would give the descriptor a certain collective meaning. In other words, the meaning of the descriptor equals the union of meanings of words from the word group attributed to the descriptor. Such a word group is frequently called the *conditional equivalence class*.

To illustrate, here are several fragments of the information science descriptor dictionary developed at Fordham University. This dictionary was created for educational purposes by students studying information retrieval systems in the Department of Computer and Information Science:

B073
- Librarianship
- Library system
- State library system
- Scientific library
- Science library
- Research library
- Technical library
- Science and technical library

C052
- Note
- Letter
- Memorandum
- Report
- Paper
- Article

K006
- Automatic information retrieval
- Bibliographic retrieval system
- Document retrieval system
- Information retrieval system
- Text retrieval system

Three descriptors appear in the preceding example, namely B073, C052, and K006. Obviously, these descriptors constitute a letter-and-digit code that is four symbols in length. Each descriptor has its conditional equivalence class, which should include at least one word of the natural language. Normally, *words and word combinations included in the conditional equivalence class are called keywords*. Note that the conditional equivalence class of the B073 descriptor contains eight keywords, which are words and word combinations of the natural language. Keywords of the C052 descriptor are represented only by words, and those of the K006 descriptor are represented only by word combinations. The meaning of the B073 descriptor consists of the meanings of all eight words making up the conditional equivalence class of this descriptor. Meanings of other descriptors are given in the similar way.

As a rule, descriptor dictionaries are created for collections of documents on the same topic. For instance, for the mathematical document collection, a mathematical descriptor dictionary is compiled, whereas the collection of medical documents requires a descriptor dictionary on medicine. Normally, depending on the subject, descriptor dictionaries contain about 1000 descriptors each. Probably, this can be explained by the fact that the 1000 most frequently used words of the natural language make up 80% of all texts.

The developers of descriptor dictionaries had to resolve many problems, some of which we list here:

1. Who should be engaged in the creation of the IRL for IR systems?
2. How many descriptors should the dictionary contain?
3. What meaning should be attributed to each concrete descriptor?
4. What are the criteria for including a specific word of the natural language in a certain conditional equivalence class?
5. Can this process be automated?