Obviously, different natural-language formulations (texts) expressing the same meaning do not imply the use of different semantical components. In all cases, one such component exists—the semantical component of the natural language. The same is true for IRL: the search request representation in the form of several unordered sets of descriptors instead of a single set (as well as representing request by different descriptors within the set or by sets of different sizes) does not imply the use of some distinct semantical component of IRL (we mean, distinct from Taube's semantical component). Therefore, when discussing the ways to improve IR systems using this component, we first assume the request representation in the form of unordered sets of descriptors. We say "first" because this particular approach has been the most popular for creating IR systems and, more importantly, is used in the majority of contemporary functioning systems.

But why has this approach turned out to be so popular? Mainly because the selection criterion using this approach accounts for a possibility of different expressions of meaning in natural-language texts (as was discussed earlier). Indeed, when writing documents, their authors use different words to describe the same phenomenon. This means that in an IR system, different documents describing the same phenomenon may be represented by a different sets of descriptors. Hence, the use of several descriptor sets representing the request (rather than one set) increases the probability of finding the required documents, because now not a single descriptor set mandatory for all required documents is used; instead, a collection of sets is used, each of which might be sufficient for successful search. Such a criterion is normally called the *selection criterion in the Boolean form* or the *Boolean selection criterion*. Now we consider this criterion in more detail.

First, note that the use of the term, "Boolean expression," for an unordered set of descriptors is due to the fact that both the form of its representation in information retrieval and the operations in which this set is used could be implemented within the framework of Boolean algebra. In essence, the Boolean criterion is simple, and it looks natural and reasonable. It could be illustrated by the example most people are familiar with, the lottery game.

Assume that a person in New Jersey is interested in participating in the state's lottery. By buying a ticket, a player formulates his request for a "win" in the lottery. This request is represented by the set of, say, six numbers. According to the rules of the New Jersey lottery, these numbers should be from 1 to 49 and there should be no repetitions in the set. Using formal mathematical symbols we write

$$\mathrm{U} = \{x\colon x \in \mathrm{N},\ 1 \le x \le 49\},$$

where U represents the set of numbers participating in the lottery and N represents the set of natural numbers. We also write

$$\mathrm{S} = \{a,\ b,\ c,\ d,\ e,\ f\},$$

where a, b, c, d, e, and $f \in \mathrm{U}$. S is the set of six numbers chosen by the player (hence, they are all distinct). $\mathrm{S} \subset \mathrm{U}$; that is, S is a subset of U. S will be called a query-set.

Note that the collection of possible subsets of U consisting of six elements includes about 20 million combinations. During a lottery, a single win-set is formed. This set, denoted by T, is formed from the set U and includes six different numbers. In other words, $\mathrm{T} \subset \mathrm{U}$. Obviously, finding T is a more rare event than finding a required document in an IR system (in addition, the collection is smaller and several documents are normally required). However, we are not interested in the probability of winning in the lottery, but in the selection (winning) criterion. The win-set will be selected from the collection if every number in the query-set appears in the win-set. In other words, all numbers, a, b, c, d, e, and f from S coincide with numbers from T, (i.e., $\mathrm{S} = \mathrm{T}$). To indicate that all numbers a, b, c, d, e, and f have to appear at the same time, the language of Boolean algebra could be used. Therefore, $a \wedge b \wedge c \wedge d \wedge e \wedge f$, where ($\wedge$) denotes the logical AND. To "win," one needs all elements (numbers) unified by AND to coincide with the numbers in the win-set.

One can obviously improve the win-retrieval results (increase the probability of winning) by buying several tickets instead of one. This means that the player will use

$$\mathrm{M} = \{\mathrm{S}_1,\ \mathrm{S}_2,\ \ldots,\ \mathrm{S}_n\},$$

where for all $1 \le i \le n$, $\mathrm{S}_i \subset \mathrm{U}$, $|\mathrm{S}_i| = 6$, $\mathrm{S}_i \ne \mathrm{S}_j$ for $i \ne j$; and n equals the number of tickets bought.

In this case, the retrieval will be performed according to several unordered sets of six numbers. For instance, if a person buys five tickets, she has five different sets, namely S_1, S_2, S_3, S_4, and S_5. It is clear that the retrieval will be successful if S_1 or S_2 or S_3 or S_4 or S_5 coincides with the win set. In Boolean algebra, such a condition is written as follows:

$$\mathrm{S}_1 \vee \mathrm{S}_2 \vee \mathrm{S}_3 \vee \mathrm{S}_4 \vee \mathrm{S}_5.$$

Here $\vee$ denotes the logical OR; that is, it prescribes the coincidence of any of five sets to be considered as a success. Now let us write this Boolean expression in more detail so that we represent each S_i in the form of the element set by means of the AND operator. This provides the following: $(a_1 \wedge b_1 \wedge c_1 \wedge d_1 \wedge e_1 \wedge f_1) \vee (a_2 \wedge b_2 \wedge c_2 \wedge d_2 \wedge e_2 \wedge f_2) \vee (a_3 \wedge b_3 \wedge c_3 \wedge d_3 \wedge e_3 \wedge f_3) \vee (a_4 \wedge b_4 \wedge c_4 \wedge d_4 \wedge e_4 \wedge f_4) \vee (a_5 \wedge b_5 \wedge c_5 \wedge d_5 \wedge e_5 \wedge f_5)$. This Boolean expression is in what is called the *disjunctive normal form*.

Thus the lottery game provides a clear illustration of the Boolean criterion used in the information retrieval process. The use of the Boolean criterion in IR systems is more flexible than that in the lottery game. For example, information retrieval does not require alternative sets to be of equal size. In other words, whereas in the lottery all sets are of equal size (six numbers), IR systems