have a value Ψ_i in the interval:

$$(L_3, L_2].$$

As we mentioned earlier, the algorithm will use the descriptors from the first interval to construct subrequests consisting of two descriptors, it will use the descriptors from the second interval to construct subrequests consisting of three descriptors, and so forth. The number of intervals (in our case 7) will depend on the average length of the document profiles in the IR system and has to be determined in advance by the developers of the IR system. For example, in an IR system where the average length of a document profile is 10 descriptors it is sufficient to use not more than 7 intervals which means that subrequests will contain at most 8 descriptors. On the other hand, in an IR system where the average length of a document profile is 80 descriptors, we might use 20 intervals and subrequests will contain at most 21 descriptors.

Now we will consider additional conditions used in constructing subrequests consisting of more than one descriptor. It is known that not all descriptors can simultaneously appear in one document. Therefore, when constructing subrequests we have to consider the simultaneous occurrence frequency of descriptors in the marked set.

Clearly it is possible to have some intervals of values j that do not contain any descriptors from the relevant neighborhood (for example, intervals 3 and 7 in Figure 7.6), and therefore we will not construct subrequests consisting of the number of descriptors corresponding to these intervals. In our example, the algorithm will not construct subrequests consisting of four or eight descriptors. It is also clear that in some intervals (in our example, intervals 4, 5, and 6) we will not have enough descriptors to construct subrequests containing descriptors from these intervals. For example, descriptor F from interval

$$(L_4, L_3]$$

has to be included in the subrequest consisting of five descriptors. Our algorithm deals with such a situation as follows.

When the algorithm constructs a subrequest consisting of more than two descriptors, it may use descriptors from intervals to the right of the interval under consideration (that is, intervals with larger L values). Thus, when constructing subrequests consisting of three descriptors, we will consider not interval

$$(L_2, L_1]$$

but rather interval

$$(L_2, L]$$

and we will require that each subrequest consisting of three descriptors contains at least one descriptor from interval

$$(L_2, L_1].$$

For subrequests consisting of four descriptors, we use interval

$$(L_3, L]$$

and require the presence of at least one descriptor from interval

$$(L_3, L_2].$$

The pragmatism of this approach can be seen from the following example. Say, interval $(L_1, L]$ contains only one descriptor. Then we cannot construct a subrequest consisting of two descriptors. But the usefulness of a descriptor whose Ψ_i value is very close to L is quite clear. In a case like this, we include the descriptor from the first interval into subrequests consisting of more than two descriptors.

It is important to consider the case when a subrequest constructed for a specific interval contains descriptors which themselves form a subrequest for another interval (to the right of the interval under consideration); that is, one subrequest properly contains another subrequest. The search using the larger subrequest can only find a subset of a set of documents that could be found using a smaller subrequest. Therefore, a subrequest that properly contains another subrequest we will call *extraneous*, and we will not include extraneous subrequests in our query formulation.

The algorithm begins its analysis of the transformed matrix by looking at the interval $(L_1, L]$; that is, it tries to find subrequests consisting of two descriptors. If interval $(L_1, L]$ contains more than one descriptor, then the algorithm analyzes all combinations of two descriptors from this interval. The set of two descriptors will constitute a subrequest only if the occurrence frequency of this set in the document profiles of the marked set is greater than some required value. For example, if we consider the search request represented by the transformed matrix in Figure 7.6 and require that any set consisting of two or more descriptors from interval $(L_1, L]$ should appear in at least five document profiles in the marked set, then we will not find any subrequest consisting of two descriptors.

After analyzing all possible combinations of two descriptors, the algorithm looks at the descriptors whose values are in the interval $(L_2, L]$ and using those descriptors constructs subrequests consisting of three descriptors. The set consisting of three descriptors will be considered a subrequest if (1) at least one of the descriptors from this set has Ψ_i value in the interval $(L_2, L_1]$, and (2) the occurrence frequency of this set in the documents of the marked set is greater than some required value. In our example, if the required frequency is 3, the algorithm will construct 6 subrequests consisting of 3 descriptors:

(1) $G \wedge L \wedge C$ (2) $B \wedge L \wedge C$ (3) $M \wedge B \wedge C$
(4) $B \wedge G \wedge C$ (5) $B \wedge G \wedge L$ (6) $M \wedge G \wedge C$