ing so, they are attempting in some cases to reduce the full text to an acceptable size, whereas in other cases their goal is to divide it into a number of independent parts of a size acceptable to the system. They are also pinning some hope on studies of coherent text structure as a direction of research in NLP. Within this direction researchers are trying to determine a comparative significance, which is statistically stable, of different parts of the text. Although some recommendations are now available, the reduction problem has not yet been solved.

Sometimes researchers argue that the indexing language should be natural language. By this they are usually referring to what is more frequently called a free text search (see Chapter 6). Recall that in this case the descriptors employed are words of natural language, and the lexical and morphological analyses certain bear a close relationship to NLP.

As shown earlier, many researchers associate IR development with successes in NLP. However, of ever-growing interest in recent years are attempts to use other directions in AI, such as a knowledge-based design, expert systems, and neural networks. First, let us take a closer look at the object of each of these directions, and after that we will show how researchers intend to use them for developing IR systems.

It would not be an exaggeration to say that to date the most popular (if not predominant) direction of AI research is a knowledge-based approach. Indeed, an overwhelming majority of approaches to the development of all kinds of systems assumes to take advantage of knowledge stored in the system itself. Because implemented AI systems use a computer, this means that knowledge must be introduced and stored in the computer; but it also means that there must be formal methods for using such knowledge. As a matter of fact, when speaking of a knowledge base we mean a certain organization in the computer of both knowledge and the means of using it. In AI we may think of knowledge as a collection of related facts, procedures, models, and heuristics that can be used in problem solving or inference systems. Knowledge may be regarded as information in context, as information organized so that it can be readily applied to solving problems, enhancing perception, and expanding learning. Knowledge varies in both content and appearance. It may, for example, be specific, general, exact, fuzzy, procedural, or declarative.

Today AI can already offer a number of implementation ideas for using knowledge that are not based on formal methods for the recognition of meaning expressed in natural language. One of these ideas, which is perhaps the most useful for information retrieval, is the idea of creating a knowledge base in the form of a semantic network (see, for example, Feigenbaum, Barr, & Cohen, 1982; and Quillan, 1968). Salton and McGill, for example, mention the applicability of a semantic network (though they do not provide any approaches to illustrate how it could be applied in information retrieval) and explain the idea of a semantic network in their book *Introduction to Modern Information Retrieval* (Salton & McGill, 1983). We will discuss semantic networks in more detail.

Semantic networks were first developed in order to represent the meanings of English sentences in terms of objects and relationships between them. This network was designed as a psychological model of human associative memory. The neural interconnections of the brain are clearly arranged in some type of network (apparently one with a highly complex structure), and the rough similarity between the artificial semantic nets and the natural brain helped to encourage the development of semantic nets. All semantic networks are constructed from two basic components:

- *Nodes,* representing objects, concepts, or situations, are indicated by boxes or circles.
- *Arcs,* representing relationships between nodes, are indicated by arrows connecting the nodes.

The following examples illustrate how knowledge may be represented and used in semantic networks. Let us assume that we wish to represent the fact that "Bob is a taxi driver." We can create two nodes—one to represent "Bob" and the other to represent "taxi driver"—and connect them with a very useful arc or link called an *isa* (Figure 12.1).

Clearly, this is very simple example. However, it illustrates how the method of writing (and storing) knowledge is used for creating the knowledge base. The idea of using semantic networks for information retrieval purposes may be described as follows.

Every node in the semantic net represents a text describing a concept, idea, situation, proof, and so on. Such text may be, for example, a segment of a book, an article, or a report. All meaningful relations are established between segments, not only between those of one document but also between the segments of any document entered into the system. Theoretically, such a semantic net may include existing ideas or situations that are contained in the documents entered into the system. The knowledge base built on such a net will become a very developed model of human knowledge and, theoretically, will be able to give an exhaustive answer to a query.

In principle, such knowledge may basically be created in the same fashion as is used to create an expert system. A group of experts is invited, and any document coming into the system is read by each expert. After completing this step, the experts jointly decide to enter into the knowledge base only text containing an original idea, situation, and so forth. In doing this, the experts establish in their own judgment all meaningful relationships between a newly created

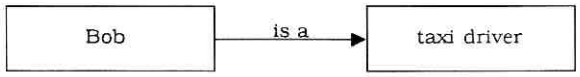

Figure 12.1
Semantic network representation of the statement "Bob is a taxi driver."