

GENERATING NOUN PHRASES TO IDENTIFY NODES IN A SEMANTIC NETWORK

George E. Heidorn

Computer Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

The Problem

In knowledge-based systems, such as those for doing automatic programming, e.g. <Heidorn 1974, 1976>, information is often stored in the form of a semantic network. In order to present portions of this information to a user, such as when answering a question, it must be converted into natural language form. Since the semantic network shows relationships among objects, part of what is required is to be able to generate phrases to uniquely identify these objects. In many cases they can be identified simply by a canned phrase, possibly a proper noun (e.g. "John"). But when the objects have more complex names (e.g. "the extended price discount percentage field in a customer master record"), a canned phrase is inappropriate, primarily because a node may not have just one identifying phrase that can be used to refer to it in all contexts, especially if the network itself may change during the dialogue.

Related Research

Little has been written about generating noun phrases to uniquely identify objects in an internal model. Some papers, such as <Simmons and Slocum 1972> and <Heidorn 1974>, put more emphasis on paragraph and clause structures, using fairly simple noun phrases. Other papers, such as <Slocum 1973> and <Shapiro 1975>, are concerned with the syntax of noun phrases, such as determining the ordering of modifiers, given a set of relations to be expressed. <Shapiro 1975> states, "At this time, we have no theory for determining the number or which adjectives and relative clauses to generate" (p. 59). Those papers that have dealt with determining what modifiers are required, such as <Winograd 1972>, have done so only for a particular micro-world, rather than offering a general method.

A Solution

<Heidorn 1977> provides a detailed description of a fairly general method for constructing identifying noun phrases for nodes in a semantic network, which is briefly described here. This method complements the one described in <Heidorn 1976> for performing the inverse process, namely that of noun phrase resolution, i.e. determining the referent node(s) of a given noun phrase. The central notion in both cases is that a noun phrase is

considered to consist of a head and zero or more modifiers, where the head refers to a class of objects (e.g. "files"), and the modifiers restrict the reference to a subclass of those (e.g. "customer files").

In the network, some nodes have a single word associated with them (e.g. "files"), and other nodes do not (e.g. the node representing "customer files"). However, nodes without associated words do have superset links to their parent nodes, as well as links to other nodes which distinguish them from their sibling nodes. In generating an identifying noun phrase for such a node N, the head is the identifying phrase of the parent of N, and the modifiers are the identifying phrases of the nodes that N is connected to that distinguish it from its siblings. The identifying phrases for the head and modifiers may themselves have to be constructed in the same manner.

The processing for the above is done in two steps. The first (semantic) creates a sub-network which is an appropriately modified copy of that portion of the network required for the unique identification of the given node, considering the current contents of the network. (This step is language-independent.) The second (syntactic and morphological) converts this sub-network into an English string. <Heidorn 1977> includes a detailed flow chart of the semantic processing algorithm and examples of all three levels of processing.

References

1. G.E. Heidorn, "English as a Very High Level Language for Simulation Programming," SIGPLAN NOTICES 9, 4 (Apr. 1974), 91-100.
2. G.E. Heidorn, "Supporting a Computer Directed Natural Language Dialogue for Automatic Business Programming," COLING 76 PREPRINTS, 6th Int. Conf. on Comp. Ling., Ottawa, June 1976.
3. G.E. Heidorn, "A Method for Generating Noun Phrases to Identify Nodes in a Semantic Network," Res. Report, IBM T. J. Watson Research Center, Yorktown Heights, N.Y., Aug. 1977.
4. S.C. Shapiro, "Generation as Parsing from a Network into a Linear String," AMER. J. OF COMP. LING., Microfiche 33, 1975.
5. R.F. Simmons and J. Slocum, "Generating English Discourse from Semantic Networks," COMM. ACM 15, 10 (Oct. 1972), 891-905.
6. J. Slocum, "Question Answering via Canonical Verbs and Semantic Models: Generating English from the Model," Tech. Report NL 13, Comp. Sci. Dept., Univ. of Texas, Austin, Jan. 1973.
7. T. Winograd, UNDERSTANDING NATURAL LANGUAGE, Academic Press, New York, 1972.