FIGURING OUT WHAT THE USER WANTS - 8TEP8 TOWARD AN AUTOMATIC YELLOW PAGES ASSISTANT

Anatole Gershoum*

Schlumberger-Doll Research
RidgeAeld, Ct. 0*077

ABSTRACT

An experimental system, AYPA, for automatic Yellow Pages assistance is described. The system, which operates in the domain of automobiles, automobile parts, and related objects, reads the user's request in simple English, analyzes it and represents it in terms of the system's conceptual primitives. From this, the system tries to figure out the intent of the request and formulate a Yellow Pages query. It paraphrases the request back to the user in English and searches its data base for the relevant Yellow Pages categories. The system serves as a research vehicle for experiments with its various components and user interfaces.

L INTRODUCTION

This paper describes an experimental system, AYPA, for automatic Yellow Pages assistance. The system, which operates in the domain of automobiles, automobile parts, and related objects, reads the user's request in simple English, analyzes it and represents it in terms of the system's conceptual primitives. From this, the system attempts to infer the intent of the request and formulate a Yellow Pages query. It paraphrases the request back to the user in English and searches its data base for the relevant Yellow Pages categories.

We view this work as a research in natural language access to databases [1], [2], (3). In order to provide natural language access to a database we have to create an extensive structure, external to the database, which can serve as a bridge between the concepts used by people in talking about the items in the database and the database organization.

The program aims at performing the following type of interaction with the user:

USER: My windshield is broken, help.

AYPA: I understand that you want to replace the windshield in your car. Where are you located?

USER. Hamilton and Stuyvesant in Springfield.

AYPA: Mere are some establishments near you that might be able to help you:

1. Al Dubin Auto & Plate Glass Works: 1171 Stuyvesant Av.: 373-2555
2. Allstate Glass Co.: 400 Hamilton Blvd.: 757-5036

Let us analyze the kinds of "intelligence" AYPA must have in order to produce the above behavior. First, it has to extract from the user's request that the user has a car, and that the windshield in this car is broken. Second, AYPA must infer that the user wants to replace the windshield. We call this kind of processing *request analysis.* Request analysis requires: (1) an ability to extract information out of natural language input and express it in some language independent representational system; (2) knowledge about objects and actions and their relationship (e.g., the knowledge that windshields are car parts made out of glass, that glass parts are not repaired but replaced, etc.); (3) knowledge about the intentions of Yellow Pages users (e.g., if the user informs the system that he has a damaged item, he probably wants to repair it), and finally, (4) the inferential capabilities to reason about the user's intentions.

At the second stage of request processing (which we call *request answering)* the system must be able to fin the relevant entries in the Yellow Pages and select some that are in the vicinity of the user.

II THE KNOWLEDGE BA8E

AYPA uses the New York Telephone Manhattan Yellow Pages data base in which entries for individual businesses are grouped into categories such as: "Automobile Dealers.-New Cars", or "Glass-Automobile, Plate, Window, Etc.-Dealers." Each entry contains the name of the business establishment, its address and its phone number. Each category has an identifying number (e.g, AV480 is "Automobile Dealers-New Cars".)

AYPA's semantic and pragmatic knowledge is stored in the form of a hierarchical associative network which we call YPNET. A fragment of this net is shown on Fig. 1.

The core of the net consists of nodes which represent generic concepts such as *CAR, 'PERSON, •REPAIR, etc., and links which define the relationship

* The work described in this document was performed while the author was at Bell Laboratories, Murray Hill, N.J. 07974.

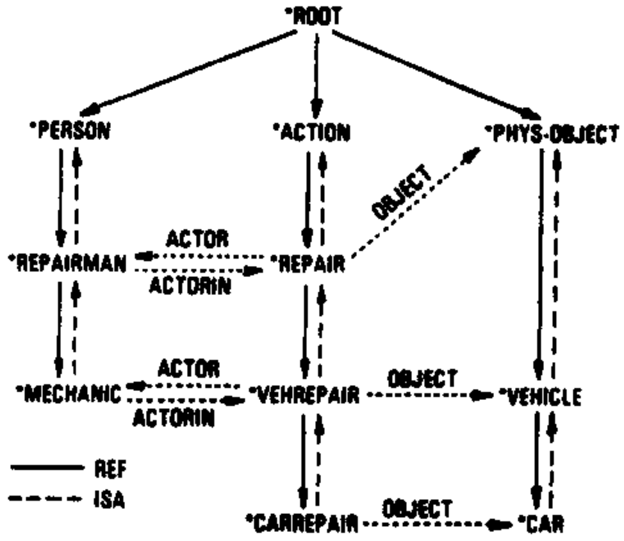

Fig. 1

among concepts in AYPA's world. There are two major kinds of links in YPNET: vertical (hierarchical) and horizontal (associative). Vertical links represent the hierarchical organization of the nodes in the net. The vertical links are the usual ones: ISA (is a) which goes from a more specific concept to its generalizations, and REF (refinement) link which goes from a general concept to its refinements (4).

Every node in YPNET is considered to be a frame with various attributes. For example, 'REPAIR has attributes ACTOR and OBJECT. The ACTOR of •REPAIR must be a 'PERSON and is called a repairman with a corresponding node 'REPAIRMAN. Horizontal links in YPNET link frames and their attributes.

Information contained in the network is used by AYPA in four ways:

- Retrieval of an item from the net (e.g., ACTOR of 'REPAIR).
- Verification of a predicate (e.g., "is a mechanic a person?", i.e., (ISA 'MECHANIC 'PERSON) = True or False).
- Finding a concept which relates two given concepts (e.g., 'CARREPAIR relates 'CAR and 'REPAIR). This is used heavily in the language analysts.
- Finding the nearest concept with a given property. For example, finding the nearest concept which corresponds to a Yellow Pages category.

III REQUEST ANALYSIS

Request analysis begins with a process which takes a natural language sentence and produces an expression representing its meaning. We call it *Conceptual Analysis* 15]. We use a modified version *of* Conceptual Dependency (CD) notation I6| as our representational system, using primitives from AYPA's knowledge base.

The conceptual analyzer used in AYPA is written in the general framework described in (7). It is a dictionary-based production-like system. This is essentially a rule-based system with rules dynamically assembled from information contained in the dictionary, syntactic frames, and most importantly, YPNET, which is the repository of the system's semantic and pragmatic knowledge.

A. Immediate Inferences

Consider the following example:

My windshield is broken. I have a 1979 Ford.

Having analyzed a sentence, AYPA activates all the inference rules indexed by the head of the conceptual expression. There are two rules stored under the concept 'DAMAGE which underlies such verbs as "to brc«k", "to dent" "to smash", etc. Their content can be paraphrased in English as:

Rule Ia: IF I) the physical state value of an object is equal to -1 (damaged) and
2) the user's goal has not been asserted

THEN assert the user's goal to repair the object.

Rule Ib: IF I) the physical state value of an object is equal to -2 (destroyed) and
2) the user's goal has not been asserted

THEN assert the user's goal to replace the object.

Thus, AYPA tentatively asserts that the user wants to repair his broken glass-autopart. (This goal will later be changed to "replace".) If a different goal is explicitly mentioned later, it will replace this one.

B. Goal Finding end Sharpening.

When all of the user's request has been read, AYPA goes to the next stage which we call goal *finding.* At this stage, the system checks if the user's goal has been asserted at the input stage and if it hasn't, AYPA tries to infer it using its general rules for inferring the user's goal. There are several such rules and here is one example:

Rule 2: IF 1) there is a physical object in the short term memory not controlled by the user

THEN assert the user's goal to buy the object.

Controlled in the above rule means either owned. rented, or occupied. If the user simply says: "car", AYPA, using Rule 2, will assume that he wants to buy a car. If the user says: "My car", AYPA will not assume that he wants to buy a car since he already has it.

After the goal-finding stage, the system has a conceptual expression representing the system's idea *of* the user's goal. This expression may still be too general by comparison with the user's real goal. For example, when the user says: "1 want a car" the system asserts that the user's goal is to control (buy or rent) a car. If the user says: "I need a garage" the system asserts that his goal is to control a garage. In reality, the first request usually means buying while the second renting. We could ask the user for clarifications at this point, but this may be annoying especially when the circumstances and the object strongly imply a preferred way of controlling it. Thus, we need another stage of request processing which we call *goal-sharpening*. This is done by activating the goal-sharpening rules stored under the conceptualization heads.

C. Goal-Elaboration.

Having found and sharpened the conceptual expression that represents the system's idea of the user's goal, AYPA goes into the last stage of request analysis which we call *goal elaboration.* At this stage, AYPA goes through all the concepts in the goal conceptualization and tries to fill the remaining gaps in them. Consider, for example, the following request:

My carburetor is broken. I have a 1979 Ford Fairmont.

The system decides that the user wants to repair a broken carburetor but, before the goal-elaboration stage, it did not associate the carburetor and the user's Fairmont. Since there were no other cars mentioned, AYPA assumes that the broken carburetor is a part of the user's Fairmont.

Goal elaboration ends the request analysis stage. Now the system expresses in English what it thinks the user's goal is and goes into the request answering stage

IV REQUEST ANSWERING

Answering a user's request may be a complicated process which requires both specific knowledge about information organization of the Yellow Pages, and "normal" pragmatic inferences, for example: "if you want to find an object part and do not find it, try to find the whole object." The Yellow Pages categories cover the world in a patchy and inconsistent way. For example, the Yellow Pages user who wants to repair a carburetor should know that the category associated with a car part is relevant to both buying and repairing of the part and look under "Carburetors". This information is implicit in the assumed model of interaction between the user and the Yellow Pages and must be made explicit if we want to use it. This is not limited to car parts. Implicit assumptions are made about nearly every Yellow Pages category. Businesses and professionals arc assumed to perform actions usually associated with them, physical objects are assumed to be repaired, bought, sold, or rented. We have made these assumptions explicit by attaching specific search rules to the concept nodes and to the user's goals.

V SOME FACTS AND CONJECTURES

The work described in this document was an experiment in trying to identify the knowledge needed for an intelligent Yellow Pages assistant. The program was used as a vehicle for experiments with various algorithms of conceptual analysis, associative network organization and searches, and user interfaces. The network consists of 85 concept nodes, 46 of which have Yellow Pages pointers. The total number of links is 223. The natural language part has a vocabulary of about 150 words. This is a long way from a full-scale Yellow Pages assistance system which would require 4500 categories. However, although we have experimented with a very small fraction of the Yellow Pages (about 1%), we have almost completely covered the area of auto buying and repairing. There is no reason to think that other parts of the Yellow Pages would give a significantly greater branching factor for the network.

The natural language part is by far the most difficult and critical in the whole project. However, its development crucially depends on the development of the associative network since the language analyzer takes most of its information from there. In fact, we have experimented with an interface which knows next to nothing about English syntax but combines the meanings of the words in the user's requests directly using the network. This produced an imperfect but very usable interface which we used for testing the network itself.

REFERENCES

[1] Hendrix G.G. (1977), "The LIFER Manual," Technical Note 138, Artificial Intelligence Center, SRI, Menlo Park, California.

[2] Waltz, D. L. (1978), "An English Language Question Answering System for a Large Relational Database." *CACM,* vol.21, July 1978, pp. 526-539.

[3] Ginsparg J. (1980), "Using the Natural Language Processor as a Data Base Interface," internal Bell Labs Technical Memo.

[4] Findler, N. V. (1979), *Associative Networks Representation and Use of Knowledge by Computers.* Academic Press, New York 1979.

15] Riesbeck, C. (1975), "Conseptual Analysis" in *Conceptual Information Processing,* ed. R. C. Schank. North-Holland, Amsterdam 1975.

[6] Schank R. C. (1975), *Conceptual Information Processing,* North-Holland, Amsterdam 1975.

[7] (iershman, A. V. (1980), "A Framework for Conceptual Analyzers," Bell Laboratories Technical Report 80-1274-13, Murray Hill N.J.