

NATURAL LANGUAGE DATA BASE ACCESS WITH PEARL

Wendy Lehnert and Steve Shwartz

Department of Computer Science at Yale University, New Haven, Ct., and Cognitive Systems, Inc. of New Haven, Ct.

Data bases access is typically approached as a programming problem rather than a user interface problem. We consequently see situations where only programmers can access information in data bases, in spite of the fact that the people who need this information are typically non-programmers. The most desirable way to extend data base access capabilities to naive user populations is by means of a natural language interface or "front-end" system that can process requests stated in the user's natural language. PEARL is a domain-independent system for natural language analysis that can be customized to handle domain-specific queries to existing data bases. PEARL analyzes English input with expectation-driven parsing techniques and question answering strategies developed by Artificial Intelligence researchers (Lehnert, /1978/, Schank and Riesbeck /1976/, Riesbeck /1975/, Gershman /1979/, Birnbaum and Selfridge /1979/, and produces code in whatever query language is currently being used to access a specific data base. PEARL can therefore be adapted to any data base that has a query language access module.

In any realistic application of natural language analysis, we can expect users to make mistakes, forget to specify necessary parameters, and use words that are not in the system's vocabulary. PEARL is designed to handle these inevitable occurrences with interactive user dialogues that

allow the user to augment, verify and alter the system's interpretation of an input request. PEARL accepts requests in unrestricted English: no constraints are placed on the user to stay within a restricted vocabulary or artificial syntax.

The following examples of PEARL in action are taken from a version of PEARL that has been customized to handle requests made by geologists to a data base containing information about oil wells. The geologists in this user population are oil explorationists with access to a graphics package that can generate maps on the basis of information in the database. The map requests given below are highly domain-specific and therefore contain many technical terms used by oil explorationists.

Yale TOPS-20 Command processor 4(560)-1

*(explorer)

EXPLORER Version 02 9/23/81

READY

*I want a map of all wells drilled before May 1, 1980 but
*since April 30, 1970 that show the tcg formation below 2000 ,
*are themselves deeper than 5000 , and are operated by Emitt.
*I only want to see wildcat wells where the operator reported
*a drilling problem, and that have mechanical logs, blew out,
*and were drilled by Lexaco within the area defined by
*latitude 30 deg 20 min 30 sec to 31:20:30 and 80-81. Make
*the scale 1" = 300 meters.

User requests a POSTED map
LATITUDE 30.34167 - 31.34167
LONGITUDE 80 - 81
output device: PAPER
filters: DRILLING DEPTH > 5000
COMPLETION DATE > = 4/30/1970
COMPLETION DATE < 5/1/1980
LEXACO ORIGINAL OPERATOR
BLOWOUT WELL
MECHANICAL LOG FILE WELL
DRILLING PROBLEM
WILDCAT WELL

EMITT CURRENT OPERATOR
PENETRATION OF TCG WHERE TOP DEPTH > = 2000'
scale of 1" = 984.252

READY

*Contur a third order trend difference on the tcg.

By CONTUR, do you mean CONTOUR (Y or N) *y

By ORDER, do you mean ORDER (Y or N) *y

DIFFERENCE not understood

Would you like to propose a synonym or correction? (Y or N) *y

Enter word: *residual

A map region was not specified.

Do you want the same geographic region as before? (Y or N) *y

Please specify contour interval.

*100 feet

A scale was not specified.

Would you like the scale from the previous request? (Y or N)

Previous scale = 984.252

*y

User requests a STRUCT map
LATITUDE 30.34167 - 31.34167
LONGITUDE 80 - 81
output device: PAPER
statistics:

TREND-EDITING
3RD-ORDER RESIDUAL

on TCG

CI = 100'

scale of 1" = 984.252'

PEARL uses conceptually-driven techniques for language analysis which must be customized for specific domains and task orientations. To customize PEARL for a new data base application, we must conduct a thorough study of typical language use for that application, so that PEARL will be able to handle technical or idiosyncratic constructions and vocabulary items. Once PEARL is capable of processing a realistic spect-

rum of possible language use, PEARL automatically adjusts to the expertise of individual users. Expert users are allowed to state their requests in as concise a manner as desired, while inexperienced users are carefully queried about possible ambiguities and insufficient specifications. This inherent flexibility is a natural by-product of any competent language processing system, and a necessary feature for effective and satisfying man/machine communication.

References:

- Birnbaum, L. and Selfridge, M. (1979). Problems in Conceptual Analysis of Natural Language. Yale University, Computer Science Dept., Research Report No. 168. Also in R. Schank and C. Riesbeck, (eds.), Inside Computer Understanding, Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.
- Gershman, A.V. (1979). Knowledge-Based Parsing. Ph.D. Thesis. Yale University, Computer Science Dept., Research Report No. 156.
- Lehnert, W. (1978). The Process of Question Answering. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Riesbeck, C. (1975). Conceptual analysis. In R.C.Schank, (ed.) Conceptual Information Processing. North Holland, Amsterdam.
- Riesbeck, C.K. and Schank, R.C. (1976). Comprehension by computer: Expectation-based analysis of sentences in context. Yale University, Computer Science Dept., Research Report No. 78. Also in W.J.M. Levelt and G.B.Flores d'Arcais (eds.), Studies in the Perception of Language. John Wiley and Sons, Ltd., Chichester, England, 1979.