Title
A KNOWLEDGE BASED SYSTEM FOR PROVIDING INTELLIGENT ACCESS TO A PETROLEUM USER DATABASE

Author
Rosenberg, Steven

Publication Date
1979

To be presented at the 6th International
Joint Conference on Artificial Intelligence,
Tokyo, Japan, August 20-24, 1979

A KNOWLEDGE BASED SYSTEM FOR PROVIDING INTELLIGENT ACCESS TO A PETROLEUM USER DATABASE

Steven Rosenberg

January 1979

Prepared for the U. S. Department of Energy
under Contract W-7405-ENG-48

LBL-8720

A KNOWLEDGE BASED SYSTEM FOR PROVIDING INTELLIGENT ACCESS TO A PETROLEUM USER DATABASE

Steven Rosenberg

Information Methodology Research Project
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

i d a d INFORMATION & DATA ANALYSIS DEPARTMENT

A KNOWLEDGE BASED SYSTEM FOR PROVIDING INTELLIGENT ACCESS TO A PETROLEUM USER DATABASE

Steven Rosenberg

Information Methodology Research Project
Lawrence Berkeley Laboratory
Berkeley, California 94720

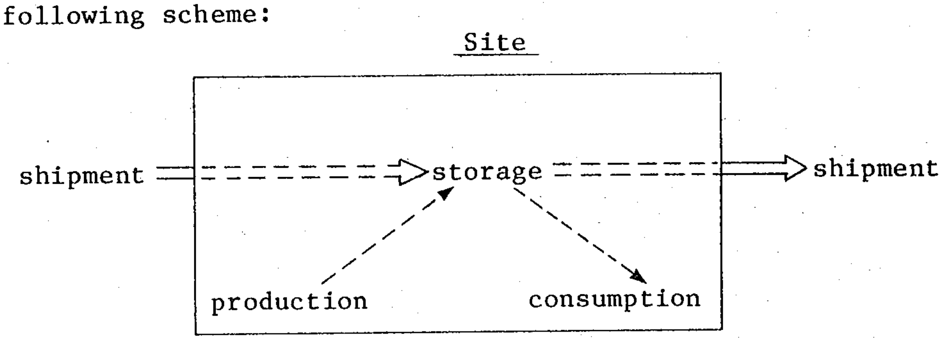
This paper describes the design of a knowledge based system for handling information requests concerning petroleum usage in the United States. The system is a testbed designed to demonstrate the usefulness of such an approach to the real world problem facing a user whose requests for information require sophisticated organization of the many reports on oil usage submitted weekly by producers, refiners, vendors, etc. The basic analysis of the petroleum industry from which this model is derived was developed in conjunction with two other members of our group, Krishnan and Cahn (1979).

The basic semantic system is constructed using FRL (Roberts and Goldstein, 1977). FRL is a sophisticated, higher level language specifically designed for the representation of knowledge in a variety of domains. It provides a hierarchically organized, frames-based semantics with inheritance and procedural attachments among other features.

The model organizes information around the following fundamental categories: site, area, and company. These represent the basic physical loci whose petroleum usage we wish to keep track of.

A site represents any actual physical location at which oil is handled, such as a port, tank farm, refinery or oil field. Areas, such as states, are considered to consist of a set of sites physically located within their boundaries, while companies are represented by ownership of a collection of sites in arbitrary locations.

Since many types of activities can occur at sites, and by extension, within states and companies, it is necessary to develop a simple cannonical model capable of representing within one basic type the different activities of refining, storage, extraction, and so on. Conceptually, we can think of any site as being some instantiation of the following scheme:



We can accurately characterize sites, and, by extension, states and companies which are physical and social collections of sites) in terms of this structure. A site, as a physical location, always represents oil in storage. Shipments coming in add to this store. Production, if it exists at a site, also adds to the store. Shipments from the sites reduce the storage. Some oil is locally consumed, either because of shipment costs, refining costs, leakage, or end-use by consumers in the local service area of that site.

We extend this to the frames representing sites, states and companies by giving them each the following set of slots (= features) in common, in addition to other slots they might have. This preserves an essential conceptual identity in the way we treat sites, states and companies. (a) bought; (b) sold; (c) consumption; (d) production; (e) carryover.

(a) The bought slot contains cumulative information concerning the volume of oil shipped to that site (or company, or state). (b) The sold slot contains the cumulative amount of oil sent from that site (or company, or state). (c) Since these two slots contain aggregate information, summed over many shipments, they do not tell us how much oil is available for shipment or consumption at a site at any one instant. This is dynamically calculated through the use of the carryover slot. This has each new shipment or production added to it, and each consumption or ship-out subtracted. It contains the actual storage amount available at a site at any given time. (d) The consumption slot represents how much oil is locally consumed. (e) The production slot, how much is locally produced.

The site frame contains additional slots, among which are the source-from slot, which contains a list of shipments which have gone to that site, and source-to, which is a list of shipments from that site, allowing examination of the raw data, if needed. Ownership and locative relations among the sites, areas and companies are also encoded into the frames.

Sites are owned by companies (which may in turn be subsidiaries of other companies) and located in areas. It follows that information contained in frames for particular companies and states will be aggregated over the sites they own or contain. Sites represent the smallest physical grain in our system beyond the raw data.

Generic frames represent the semantics in our model. The instances of these generic frames are used to represent actual data. This data is entered in particular instances of generic frames, which inherit the semantic properties of the generic type. These instances represent the grain of our model. Besides the physical grain, there exists a time grain. Frame instances will have a particular date specified. We have chosen to use a grain of a month. The basic data elements for sites, areas and companies will consist of month by month instances of these generic types, showing the cumulative data for that month. Other frames provide chronological aggregation over the physical grain, by having a cumulative time span of a season or a year. Thus, we might ask about oil consumption in New England last winter.

The basic transactional unit in our model is a <u>shipment</u>. Raw data is entered as shipments between sites. A shipment has the following structure:

Shipment
- <u>price</u>
- <u>volume</u>
- <u>status</u>
- <u>bought</u>
- <u>sold</u>
- <u>date</u>

This structure is sufficient to specify all transactions which occur in our model. Each shipment specifies the <u>price</u> and <u>volume</u> of the oil involved; where it came from (<u>bought</u>), and where it is going (<u>sold</u>). The status slot specifies which class this shipment falls into. (Shipments are marked as to whether they are imports, production, destined for export, are merely being transhipped (say from one state to another), or being shipped to the final consumers.)

Sites form the physical locus for all transfers of petroleum. All transactions between sites involve shipments of petroleum from one site to another, the adjustment of the cumulative statistics maintained for each site, and changes in the available carryover of each site.

Production and consumption are viewed as <u>within site</u> transactions that are also instantiated as shipments. To produce oil at a site representing an oil field, for example, and send it to another site, involves two shipments. The first shipment is from the production at the site to the site's storage, where it gets added to the site's carryover, and the second is from that site to another site.

Sites and shipments serve as the interface between the continuous aspects of production, consumption and transportation and the discrete model. The two cannonical entities of shipments and sites provide a simple basic semantic structure for the entire system. These provide the two elements necessary for a transaction.

There is only type of transaction possible, shipping oil, and this occurs only within or between sites. All other entities in the model represent aggregation of data. Aggregation occurs for the owner (company) of each site involved in a shipment, and for the location (state) of each site. This can in turn trigger further aggregation, as when a company is a subsidiary of another, or when an area is part of a larger geographical grouping (such as the New England states). All transactions are annotated by linking the elements involved (one shipment and either one or two sites). The shipment is added to the site frames' source slots. Thus, the raw data for all subsequent calculations and resulting changes is recorded at this level and can be examined if necessary. Simple information requests can often require sophisticated processes. Default procedures, caveats and validation processes are a few of the mechanisms used to provide this.

References

Krishnan, V.V. and Cahn, D.F. "An Aggregated Vectorial Model of Petroleum Flow for the United States," LBL-8874, January 1979.

Roberts, R.B. and Goldstein, I.P. "The FRL Manual," AI Memo 409, MIT, June 1977.

e