

# Lawrence Berkeley National Laboratory

## Lawrence Berkeley National Laboratory

### **Title**

COMPUTER-AIDED VISUALIZATION OF DATABASE STRUCTURAL RELATIONSHIPS

### **Permalink**

<https://escholarship.org/uc/item/1f63p4x5>

### **Author**

Cahn, D.F.

### **Publication Date**

1980-04-01

Peer reviewed



# Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

## Employee & Information Services Division

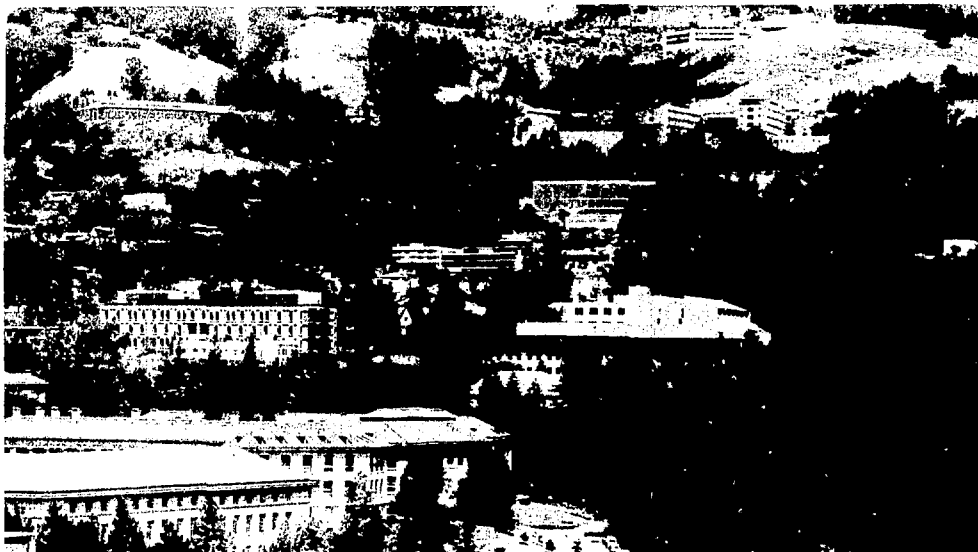
To be presented at the 43rd Annual Meeting of the American  
Society for Information Sciences, Anaheim, CA, October 5-10, 1980

COMPUTER-AIDED VISUALIZATION OF DATABASE STRUCTURAL RELATIONSHIPS

David F. Cahn

April 1980

**MASTER**



Submitted for  
ASIS Annual Conference,  
Anaheim, October 1980

LBL-10437

COMPUTER-AIDED VISUALIZATION  
OF DATABASE STRUCTURAL RELATIONSHIPS

David F. Cahn  
Information Methodology Research Project  
Lawrence Berkeley Laboratory  
Berkeley, California 49720

April 1980

DISCLAIMER

This document is prepared for the U. S. Department of Energy, Technical Information Center, under Contract W-7405-ENG-48. It contains information that is not to be distributed outside the Department of Energy.

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

COMPUTER-AIDED VISUALIZATION  
OF DATABASE STRUCTURAL RELATIONSHIPS

David F. Cahn

Information Methodology Research Project  
Lawrence Berkeley Laboratory  
Berkeley, California 94720

Abstract. Interactive computer graphic displays can be extremely useful in augmenting understandability of data structures. In complexly inter-related domains such as bibliographic thesauri and energy information systems, node and link displays represent one such tool. In this paper, we present examples of data structure representations we have found useful in these domains and discuss some of their generalizable components.

INTRODUCTION

Visualization of data element relationships is a key factor both in substantive development and in utilization of an information system. With the exponential link density increase that accompanies system growth, interactive computer aids for this visualization become increasingly attractive.

We have undertaken development of a set of computer driven display tools aimed at representation of data element relationships. The essence of the approach has been to display a limited conceptual neighborhood and allow the user to move this display window through data element space under interactive control. In the thesaurus domain, this involves display of a centroid term and its nearest conceptual relations as a single frame. In the numerical/fact domain, our early concentration has been on petroleum supply, and supplier/customer trees have proven to be of interest.

## BIBLIOGRAPHIC THESAURUS STRUCTURES

Past attempts at graphical display of thesaurus relationships (1,2,3) have relied on printed media for ultimate communication to the user. Because of the difficulty and expense associated with printed presentations of this sort there has been a tendency to encompass great volumes of information on each page, and the resulting complication has drastically limited user acceptance. By contrast, judicious use of a computer and video terminal allows display of limited descriptor neighborhoods one at a time, with user-controlled motion among the neighborhoods substituted for simultaneous display of a larger and more complicated content structure.

Figure 1 illustrates a paragraph from a typical thesaurus. As indicated in the figure legend, the actual display highlights different term groups (BT1, NT1, etc.) with color, and this continuous reference, coupled with consistent screen position, capitalization, etc., enhances understandability as a succession of paragraphs is displayed. The central or main term (MT) of the paragraph is displayed at screen center, and hierarchical depth increases from broader terms (BT) at the left through narrower terms (NT) at the right. Two levels of breadth (BT1, BT2) and depth (NT1, NT2) are displayed in a given screenload, with nearest neighbors (BT1, NT1) linked to the main term by vectors. Nonhierarchical related terms (RT) are displayed in column with the main term, since their unspecified depth is assumed to be the same.

After specification of an initial entry point into the thesaurus (which begets the first screenload), the user can investigate descriptors available for use in his search set by pursuing conceptually-linked relations. He does this by sequentially designating (via light pen or keyboard) one of the visible BT, NT, or RT's as the next main term, and then receiving a

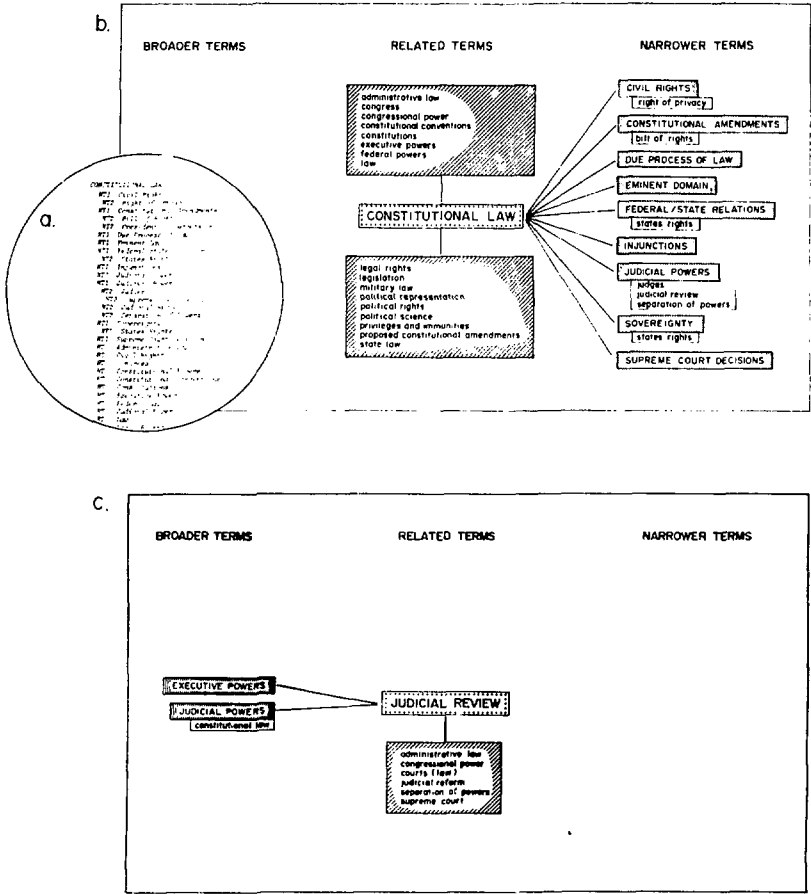


Figure 1. Typical thesaurus paragraphs in traditional tabular-alphabetic (a) and computer graphic (b) presentations. Shading has been substituted for the color-keying used in the actual computer graphic displays, but the clarity improvement is still apparent. The paragraph's main term (NT) appears at screen center, broader terms up to two levels distant (BT1, BT2) in the left column, narrower terms (NT1, NT2) in the right column, and related terms (RT) in column with the main term. The hierarchical terms are grouped in BT1/NT1 units, and the units are linked to the main term with vectors. Each presentation screenload is restricted to five hierarchical levels. In (c), a grandchild of the main term from (h) has been selected by the user, and the thesaurus neighborhood is displayed from this term's perspective. The original main term appears as a BT2, or grandparent.

Legend

TERM	FUNCTION	ACTUAL COLOR KEYING
Major term		white on red
Broader term		white on purple
Broader term 2		black on white
Narrower term 1		black on blue
Narrower term 2		blue on cyan
Related term		red on yellow

XBL 804-9423

neighborhood display centered on it. In Figure 1(c), 'judicial review' has been selected in this manner, and the screen focus has been appropriately shifted. 'Constitutional law', the original focus, appears as a BT2 when seen from the perspective of 'judicial review'; this is the reciprocal of the NT2 relationship 'judicial review' had to 'constitutional law' in Figure 1(b). Since the user is not artificially constrained by the page flipping normally required with an alphabetic listing, it is easy for him to explore the thesaurus' holdings in his area of interest; he can then be sure that the term he ultimately selects for his search set is the most appropriate available for his purposes.

From the thesaurus development standpoint, the computer driven displays serve not only to illustrate the existing terminology space, but also to point out its bounds, i.e., areas where new terminology may be desirable. When modifications are implemented, the displays are useful in verifying that the requisite term reciprocations have been performed correctly and have not produced unwanted side effects.

The crux of the developments reported here is that a reasonably minor viewpoint change, coupled with recent computer hardware advances, makes some fairly striking improvements in an otherwise difficult data visualization task. Display of the relationships among data elements in a purpose-keyed format, manipulated via interactive user control, improves comprehensibility and exposes both intended and artifactual term linkages that would otherwise be difficult to discern.

The thesaurus graphics system as outlined operates on the tabularly-formatted thesaurus files typical of the major on-line bibliographic search services, and is presently being developed as an additional production

feature for DOE's RECON system. Technical details and a more comprehensive discussion are available in an earlier paper (4).

#### NUMERICAL/FACT DATA STRUCTURES

As in the bibliographic domain, numerical/fact data systems are hard to comprehend without some mechanism for visualizing data element relationships. An example of an appropriate display, concerning petroleum supply dependencies, is presented in Figure 2. If a petroleum supply perturbation (such as an import cut-off) impinges on the U.S., it is important to be able to determine what facilities or geographic zones are dependent on the perturbed supply and in what proportion. This kind of information aids not only impact estimation for potential shortages, but assessment and evaluation of available courses of action.

Figure 2(a) depicts a hypothetical distribution tree for crude oil from its importation point at Linden, N. J. If Iranian imports are reduced and Linden is normally a prime recipient of Iranian oil, then its dependent customer family will suffer. Boston will be especially hard hit, since it imports Iranian petroleum directly as well as through the Linden terminal.

Converse to the supplier's customer tree, it might be useful to know where a particular zone gets its oil (i.e., to look 'upward' in the flow hierarchy). This is illustrated in Figure 2(b) from Boston's perspective, and it is apparent that previously invisible source chains are now exposed parallel to the one from Linden. (Looking downward in Figure 2(a), the same is true for customers parallel to Boston.)

The displays indicate dependency volumes and critical paths as well as the identity of the dependents. Historical normal flow volumes along the



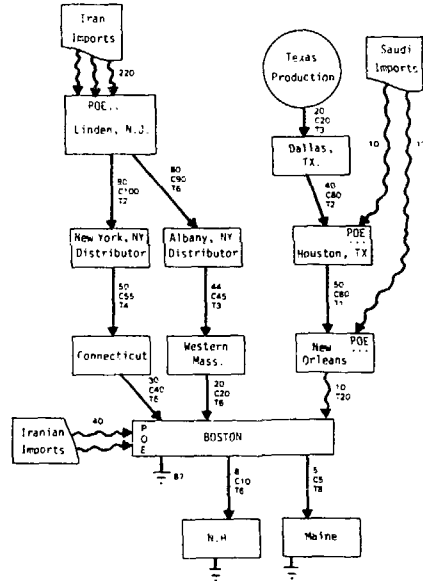
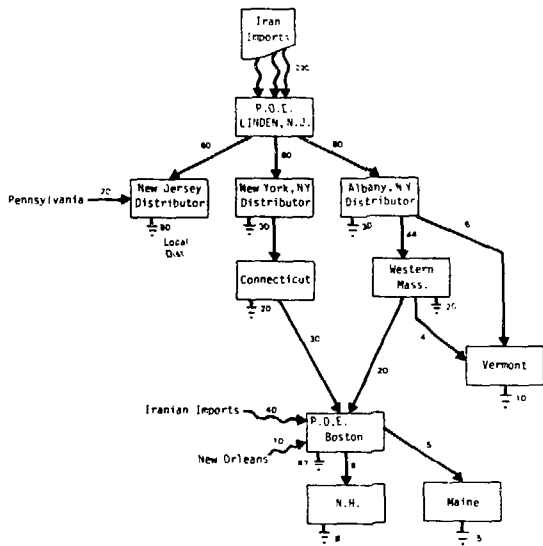


Figure 2. Computer graphics for numerical/fact data relationships: petroleum supply dependency nets, from the perspectives of (a) typical supplier (Linden, N.J.) and (b) typical recipient (Boston, Mass.). Since the database-driven displays can directly access only data familiarly linked to the node of interest, the upward- and downward-looking perspectives differ. Network segments shown cover a small portion of the U.S. net, as represented in the database; user control of node fixation and aggregation level obviate simultaneous display of the entire net. Numbers on the links indicate normal flow volumes, maxima, and transit times, and support critical path and temporal propagation analyses.

Symbol Key	
.....	use light trucks or rail
~~~~~	use ship or barge
—	use pipeline
⊥	local consumption
80	Historical normal volume carried in 1000 bbl/day
CS	Maximum flow rate (capacity)
TB	Transit time (days)

data paths can be summed around each node, yielding percentage or absolute dependency information. Similarly, the additional flow that can augment supply on one path when another is disrupted is limited by the difference between maximum capacity and normal flow at the weakest link in the path. For example, in Figure 3(b), if Boston wishes to compensate for loss of its direct Iranian imports, it cannot do so via its Western Massachusetts supplier since the link between them is already running at capacity. The display also indicates transit times associated with the connecting links, and thus assists in analysis of shortage and relief propagation cycles.

In the absence of displays such as those of Figure 2, and, particularly, in the absence of the ability to interact with them, a decisionmaker would be hard pressed to recognize critical relationships among the data. If, for example, he received a tabular listing of the crude oil flows among the zones, alphabetized by zone, rather than the graphical presentation, he would have considerable difficulty reconstructing even a single flow path, much less the entire net.

#### SUMMARY

Clearly, we have only scratched the surface of interactive data structure displays. The examples outlined are appropriate to specific kinds of data, and must interface both with other kinds of displays and with appropriate data management methods in order to operate effectively. In the numerical/fact area especially, the structural display described is intended to operate in conjunction with cartographic display maps and color-enhanced tabular matrix presentations, and maximum power of the system is dependent on the ensemble. Nonetheless, the structural displays described make a real

contribution to data system utility and have reasonable generality both in application and in the principles on which they are based.

#### ACKNOWLEDGMENT

This work was prepared for the U. S. Department of Energy, Technical Information Center, under Contract W-7405-ENG-48.

#### REFERENCES

- (1) Kolling, L., "The Role of Graphic Display of Concept Relationships in Indexing and Retrieval Vocabularies including a Thesaurus of Documentation Terms." EUR 2291.e, Center for Information and Documentation, Euratom, Brussels, 1965.
- (2) Rolling, L., "Euratom-Thesaurus, Part II: Terminology Charts used in Euratom's Nuclear Documentation System." EUR 500.e, Center for Information and Dissemination, Euratom, Brussels, 1967.
- (3) SPINES Thesaurus, Vol. III: Terminological Graphic Displays. Unesco Press, Paris, 1976.
- (4) Cahn, D. F., and Murano, C. V., "Interactive computer graphic displays for hierarchical data structures." Proc. Amer. Soc. Info. Sci. Mid-Year Meeting, Pittsburgh, 1980.