

**DATABASE TOMOGRAPHY FOR COMMERCIAL APPLICATION**

**Dr. Ronald N. Kostoff**  
Director, Technical Assessment  
Office Of Naval Research  
800 N. Quincy St.  
Arlington, VA 22217

**Mr. Henry J. Eberhart**  
Aerospace Engineer  
Naval Air Warfare Center Weapons Division, Code C02422  
1 Administration Circle  
China Lake, CA 93555-6001

*The views expressed in this paper are solely those of the authors and do not represent the views of the Department of the Navy.*

**ABSTRACT**

Database Tomography is a revolutionary method for extracting themes and their relationships from text. The algorithms employed begin with word frequency and word proximity analysis and build upon these results. When the word 'database' is used, think of medical or police records, patents, journals, or papers, etc. (any text information that can be computer stored). Database Tomography features a full text, user interactive technique enabling the user to identify areas of interest, establish relationships, and map trends for a deeper understanding of an area of interest. Database Tomography concepts and applications have been reported in journals and presented at conferences. One important feature of the Database Tomography algorithm is that it can be used on a database of any size, and will facilitate the users ability to understand the volume of content therein. While employing the process to identify research opportunities it became obvious that this promising technology has potential applications for business, science, engineering, law, and academe. Examples include evaluating marketing trends, strategies, relationships and associations. Also, the Database Tomography process would be a powerful component in the areas of competitive intelligence, national security intelligence and patent analysis. User interests and involvement cannot be overemphasized.

**INTRODUCTION**

The purpose of this paper is to introduce the Database Tomography process to the commercial community and to other departments of the government. Database Tomography is a revolutionary method for extracting themes and their relationships from text. The algorithms employed start with word frequency and word proximity analysis and build upon these results. This process was developed under the sponsorship of the Office of Naval Research for the purpose of exploring trends in naval research and also identifying promising research opportunities. It was while employing the process for research opportunities that it became obvious that this promising technology has potential applications to the commercial sector and to other government departments. For example, it could be used for evaluating marketing trends, strategies, relationships and associations. In addition, we believe that the Database Tomography process would be a powerful component in the areas of competitive intelligence, national security intelligence, and patent analysis.

Database Tomography is a full text, user interactive technique that enables the user to identify areas of interest, establish relationships, and map trends so as to provide the user with a deeper understanding of his area of interest. The reader is referred to the bibliography section of this paper for examples of Database Tomography concepts and applications that have been reported in journals and presented at conferences. User interests shape the way in which the process is used, and this user involvement cannot be overemphasized.

Generalized and discipline-specific computer databases have become a major resource for researchers and analysts in business, science, engineering, law, and academe. Database size can make analysis and interpretation difficult. One important feature of the Database Tomography algorithm is that it can be used on a database of any size and can also facilitate the user's ability to understand the volume of content therein. When the word 'database' is used, the reader is encouraged to think of journals, papers, memos, reports, medical or police records, and patents—in other words, any text information that can be stored on a computer.

An application to a Former Soviet Union (FSU) text database is shown. This text describes a broad spectrum of FSU science (35 reports generated by the Foreign Applied Sciences Assessment Center (FASAC)). The algorithm extracts words and word phrases which are repeated throughout this large database. It allows the user to create a taxonomy of pervasive research thrusts from this extracted data. The algorithm then extracts words and phrases which occur **physically close** to the pervasive research thrusts throughout the text. It allows the user to determine interconnectivity among research thrusts, as well as determine research sub-thrusts strongly related to the pervasive thrusts.

The focus of the present study was to identify **technical thrusts and their interrelationships**. The raw data obtained by the extraction algorithms allowed the user to relate **technical thrusts to institutions, journals, people, geographical locations, and other categories**. The methodology can be applied to any text database, consisting of published papers, reports, and memos.

### Background

About a decade ago, the U.S. Federal Government established the Foreign Applied Sciences Assessment Center (FASAC) under the operation of the Science Applications International Corporation (SAIC). The purpose of FASAC was to increase awareness of new foreign technologies with military, economic, or political importance. The emphasis was placed on "exploratory research" (Department of Defense 6.1/6.2 equivalent) in the FSU. This work seeks to translate fundamental research into new technology.

One of the main products of FASAC is reports on different areas of "exploratory research." FASAC assembles panels of expert consultants from academia, industry, and government. Each panel provides a written assessment of the status and potential impacts of foreign applied science in selected areas. Periodically, an Integration Report is generated that describes the trends in foreign research, including pervasive issues which affect research capabilities. By early 1992, there were about 40 reports on different aspects of FSU applied science.

### CO-WORD ANALYSIS

Co-word analysis utilizes the proximity of words and their frequency of co-occurrence in some domain (sentence, paragraph, paper) to estimate the strength of their relationship. When applied to the literature in a technical field, co-word analysis allows a map of the relationship among technical themes to be constructed. A history of co-word analysis applied to research policy issues and co-word analysis origins in computational linguistics will be published in a Special Issue of *Competitive Intelligence Review* on technology. Over the past year, a **full text** word association technique (database tomography, a variant of co-word analysis) has been developed by the authors to allow rapid scanning of large text databases. The initial purpose of this development was to identify pervasive research thrusts (thrusts which transcend disciplines) from those large text databases which contain descriptions of many research programs or areas of research. Two applications have been reported:

1. Identification of pervasive research thrusts in a database describing promising research opportunities for the Navy. The database consisted of thirty reports produced by the National Academy of Sciences panels and Office of Naval Research (ONR) internal experts on 15 technical disciplines.
2. Identification of pervasive thrusts in the 7400 project Industrial R&D (IR&D) database. Applications to other large databases of (mainly) research program descriptions are ongoing.

The reported studies and the present study have used the following procedure:

First, the frequencies of appearance in the total text of all single words (for example, MATRIX), adjacent double words (METAL MATRIX), and adjacent triple words (METAL MATRIX COMPOSITES) are computed. The highest frequency technical content words are selected as the pervasive themes of the full database (for example, SHOCK WAVE, REMOTE SENSING, IMAGE PROCESSING).

Second, for each theme word, the frequencies of words within  $\pm 50$  words of the theme word for every occurrence in the full text are computed. A word frequency dictionary is constructed which shows the words closely related to the theme word. Numerical indices are employed to quantify the strength of this relationship. Both quantitative and qualitative analyses of each dictionary (hereafter called cluster) yield those subthemes closely related to the main cluster theme.

Third, threshold values are assigned to the numerical indices. These indices are used to filter out the most closely related words to the cluster theme (e.g., see Figure 1 for part of a typical filtered cluster from the FASAC study).

Cij	Ci	Ii	Eij	CLUSTER MEMBER
		$Cij/Ci$	$Cij^2/CiCj$	
022	0036	0.611	0.0359	THERMAL INFRARED
056	0323	0.173	0.0259	ICE
070	0522	0.134	0.0250	SATELLITE

CODE:  
 Cij IS CO-OCCURRENCE FREQUENCY, OR NUMBER OF TIMES CLUSTER MEMBER APPEARS WITHIN +/- 50 WORDS OF CLUSTER THEME IN TOTAL TEXT;  
 Ci IS ABSOLUTE OCCURRENCE FREQUENCY OF CLUSTER MEMBER;  
 Cj IS ABSOLUTE OCCURRENCE FREQUENCY OF CLUSTER THEME;  
 Ii, THE INCLUSION INDEX BASED ON CLUSTER MEMBER, IS RATIO OF Cij TO Ci; AND  
 Eij, THE EQUIVALENCE INDEX, IS PRODUCT OF INCLUSION INDEX BASED ON CLUSTER MEMBER Ii (Cij/Ci) AND INCLUSION INDEX BASED ON CLUSTER THEME Ij (Cij/Cj).

Figure 1. Remote sensing cluster - closely related words.

Subsets of closely related words are combined into one file. Words which are common to more than one subset (cluster overlaps) are identified. Megaclusters, or strings of overlapping clusters (based on a threshold of numbers of common words, or overlaps), are constructed. These show umbrella areas of related research.

The final results identify:

1. The pervasive themes of the database
2. The relationship among these themes
3. The relationship of supporting sub-thrust areas (both high and low frequency) to the high-frequency themes.

Numbers are limited in their ability to portray the conceptual relationships among themes and sub-themes. The qualitative analyses of the extracted data have been at least as important as the quantitative analyses. The richness and detail of the extracted data in the full text analysis allows an understanding of the theme interrelationships not heretofore possible with previous text abstraction techniques using index or key words.

## Application of Co-word Analysis to FASAC Database

The FSU is a major contributor to many areas of science and technology. FASAC reports help to document and provide insight to these contributions. There is present interest in preserving the basic science capability of the FSU. This task would benefit from improved understanding of the FSU science and technology capability.

Application of full text co-word analysis to the FSU component of the FASAC database could provide a unique perspective on the FSU science and technology capability. This database has a different structure from the databases analyzed previously. FASAC contains topical area assessments, whereas, the other databases analyzed contain program, project, or promising opportunity descriptions. Full text co-word analysis is sufficiently powerful and flexible to be applicable to FASAC as well. (Unclassified FASAC reports were used.) The FASAC database has a moderate density of technical terms. Most are scientific, but there are many institute names, journal names, publishers, and people names. Determination of the relationship among only technical areas is more difficult than in some purely technically focused databases which were analyzed previously. However, the data allows analyses which go beyond purely technical relationships.

### MULTIWORD FREQUENCY ANALYSIS

The output of the multiword frequency analysis allows construction of a multilevel taxonomy of the full database. This taxonomy derives from the language and natural divisions of the database (analogous to a natural coordinate system of the database). Database entries are easily categorized. Other taxonomies are generated top-down and usually attempt to force-fit database subjects into pre-determined categories.

One advantage of the present full text approach over the index or key word approach is that many types of taxonomies can be generated, such as:

- Science
- Technology
- Institution
- Journal
- Person name

Within any one of these categories, such as science, many types of taxonomies can be developed. An example of one science taxonomy of the FASAC database will be shown.

Based on the high frequency single, adjacent double and triple words, the following high level taxonomy was generated. The capitalized words are *sample* high frequency words from the multiword frequency analyses:

- Information:
  - DATA
  - IMAGE PROCESSING
  - STATISTICAL PATTERN RECOGNITION
- Physics:
  - LASER
  - SHOCK WAVE
  - CHARGED PARTICLE ACCELERATORS
- Environment:
  - OCEAN
  - SEA SURFACE
  - INTERNAL GRAVITY WAVES

- **Materials:**
  - **MATERIALS**
  - **THIN FILM**
  - **METAL MATRIX COMPOSITES**

Caution must be exercised in relating the above taxonomy based on FASAC to the actual taxonomy of all of FSU science. The FASAC reports represent selected areas of FSU science. We do not know how representative all the FASAC reports are of total FSU science. The FASAC reports tend to reflect the open FSU literature. We do not know how well this open literature represents all of FSU science, including classified work and other unreported work.

The above taxonomy reflects frequency of word usage. It represents the numbers of words written about technical areas in the FASAC reports. Dollars spent on these areas, or other measures of FSU priorities, were not taken into account. The taxonomy could be skewed relative to FSU importance attached to these areas. Nevertheless, the above taxonomy does offer insight into areas of FSU science of interest to the U.S.

### Megaclusters

Clusters which had three or more overlaps (three or more common members) were combined to form strings of related clusters, or megaclusters. The following megaclusters were obtained:

- Ionospheric Heating/Modification
- Image/Optical Processing
- Air-Sea Interface
- Low Observable
- Explosive Combustion
- Particle Beams
- Automatic/Remote Control
- Frequency Standards
- Radar Cross Section

Of the 60 cluster themes that were used to compute overlaps, 52 were in one of the nine megaclusters above. Most of the eight remaining themes could be subsumed under the nine megaclusters.

The science discipline taxonomy for the FASAC database was derived from the multiword frequency analysis. It was defined as Information, Physics, Environment, and Materials. In terms of the megaclusters:

- Information would encompass:
  - **IMAGE/OPTICAL PROCESSING**
  - **AUTOMATIC/REMOTE CONTROL**
- Physics would encompass:
  - **IONOSPHERIC HEATING/MODIFICATION**
  - **PARTICLE BEAMS**
  - **FREQUENCY STANDARDS**
  - **RADAR CROSS SECTION**
- Environment would encompass:
  - **AIR-SEA INTERFACE**
- Materials would encompass:
  - **EXPLOSIVE COMBUSTION**
  - **LOW OBSERVABLE**

Categorizing the database with the megacluster subcategories allows a re-interpretation of the FASAC database. FASAC is a compendium of those aspects of FSU science of interest to the U.S. for strategic and military purposes rather than a microcosm of all of FSU science.

For example, many classes of materials were researched and developed in the FSU. Yet the materials subcategory in the FASAC analysis focuses on FSU capabilities in energetic materials (explosives and propellants) and coatings to reduce radar cross sections. Both classes are important from a military viewpoint. The main environmental focus is air-sea interface. There is little mention of the terrestrial environment. The primary information category focus is on image and optical processing, and the secondary information category focus is on remote control. We could conclude that the FASAC concern was FSU capability in sensing the ocean for ship and submarine activity, and remotely processing and interpreting this information.

The secondary environmental focus of FASAC was on the ionosphere. Specifically, it was on FSU capabilities for modifying the ionosphere through high power radio wave heating and exploiting its use as a communication medium. One focus of the physics category was particle beams. These could have dual applications of high energy directed weapons and heaters for magnetically confined plasmas and inertial fusion targets.

Cluster Theme/Member Relationships

The final display, Figure 2, shows high technical content words from one of the smallest of the 60 clusters. The selection cutoff criterion was an Equivalence Index (see Figure 1 for definition) greater than or equal to 0.001. A simple division of word categories into quadrants based on Inclusion Index values was used to display the relationships of the cluster members to the cluster theme and to each other.

In Figure 2, the underlined topic, ATMOS OCEANIC PHYS, is the cluster theme. The cluster members are segregated into quadrants headed by their values of Inclusion Indices.  $I_j$  is the ratio of  $C_{ij}$  to  $C_j$ , and is the Inclusion Index based on the theme word.  $I_i$  is the ratio of  $C_{ij}$  to  $C_i$ , and is the Inclusion Index based on the cluster member. The dividing points between high and low  $I_j$  and  $I_i$  are the middle of the "knee" of the distribution functions of numbers of cluster members vs. values of  $I_j$  and  $I_i$ . All cluster members with  $I_j$  greater than or equal to 0.1 were defined as having high  $I_j$ . All cluster members with  $I_i$  greater than or equal to 0.5 were defined as having high  $I_i$ .

<u>ATMOS OCEANIC PHYS</u> CLUSTER - HIGH TECHNICAL CONTENT WORDS		
HIGH $I_j$ HIGH $I_i$		
HIGH $I_j$ LOW $I_i$	LOW $I_j$ HIGH $I_i$	
SEA INTERNAL WAVE ACOUSTIC SCATTERING RADAR SEA SURFACE ATMOSPHERE	RADIOACOUSTIC SOUNDING ACOUSTIC SOUNDING THEORY OF WIND MODELING OF SURFACE WIND WAVES ATMOS INFRASOUND AND INTERNAL ATTENUATION OF SOUND THEORY OF WAVE	
	LOW $I_j$ LOW $I_i$	
	WIND WAVES SOUND PROPAGATION OCEAN SURFACE GRAVITY WAVES STRATIFIED FLUID	SHEAR FLOW TURBULENT SATELLITE INTERNAL GRAVITY WAVES SOUND WAVES
		PROCESSING OF RADAR WAVE PROPAGATION WIND VELOCITY POINT SOURCE

Figure 2. High technical content words of final display.

A high value of  $I_j$  means that, whenever the theme word appears in the text, there is a high probability that the cluster member will appear within  $\pm 50$  words of the theme word. A high value of  $I_i$  means that, whenever the cluster member appears in the text, there is a high probability that the theme word will appear within  $\pm 50$  words of the cluster member.

Thus, words located in the upper quadrant (high  $I_j$  high  $I_i$ ) are coupled very strongly to the theme word. Whenever the theme word appears, there is a high probability that the cluster member will be physically close. Whenever the cluster member appears, there is a high probability that the theme word will be physically close. Whenever either word appears in the text, the other will be physically close.

Consider words located in the left quadrant (high  $I_j$  low  $I_i$ ). Whenever the cluster member appears in the text, there is a low probability that it will be physically close to the theme word. Whenever the theme word appears in the text, there is a high probability that it will be physically close to the cluster member. This type of situation occurs when the frequency of occurrence of the cluster member  $C_i$  is substantially larger than the frequency of occurrence of the theme word  $C_j$ , and the cluster member and the theme word have some related meaning.

Single words have absolute frequencies of an order of magnitude higher than double words. Thus, the words in the left quadrant are typically high frequency single words. They are related to the theme word but much broader in meaning than the theme word. A small fraction of the time that these broad single words appear, the more narrowly defined double word theme will appear physically close. However, whenever the narrowly defined double word theme appears, the broader related single word cluster member will appear. The words in the left quadrant can also be viewed as a higher level taxonomy of technical disciplines related to the theme ATMOS OCEANIC PHYS.

Consider words located in the right quadrant (low  $I_j$  high  $I_i$ ). Whenever the cluster member appears in the text, there is a high probability that it will be physically close to the theme word. Whenever the theme word appears in the text, there is a low probability that it will be physically close to the cluster member. This type of situation occurs when the frequency of occurrence of the cluster member  $C_i$  is substantially smaller than the frequency of occurrence of the theme word  $C_j$ , and the cluster member and the theme word have some related meaning. Thus, the words in the right quadrant tend to be low frequency double and triple words, related to the theme word but very narrowly defined.

A large fraction of the time that these very narrow double and triple words appear, the relatively broader double word theme will appear physically close. However, a small fraction of the time that the relatively broad double word theme appears, the more narrow double and triple word cluster member will appear. This quadrant grouping has the potential for identifying "needle-in-a-haystack" type thrusts which occur infrequently but strongly support the theme when they do occur. One of many advantages of full text over key or index words is this illustrated ability to retain low frequency but highly important words, since the key word approach ignores the low frequency words.

The words in the bottom quadrant (low  $I_j$  low  $I_i$ ) are the remainder of the culled words. They relate to and support the theme, but do not have the strong inclusions based on theme or cluster member occurrence of the members of the other quadrants. The upper quadrant typically contains very few or no words. The left quadrant contains very broad words related to the theme. The right quadrant contains extremely narrow words related to the theme. The bottom quadrant contains words related to the theme of the same level of specificity as the theme (on average).

Figure 2, ATMOS OCEANIC PHYS, has a null upper quadrant (typical of the majority of clusters for the threshold values of Equivalence index chosen). The left quadrant, the broad taxonomy of related areas, appears to describe two major thrusts:

1. Underwater related (SEA, INTERNAL WAVE, ACOUSTIC, SCATTERING) focusing on sound propagation through the sea.
2. Atmosphere related (ATMOSPHERE, RADAR, SEA SURFACE, SCATTERING) focusing on radar propagation through the atmosphere.

The thrusts have a common juncture at the sea surface, where both acoustic and radar scattering occur on different sides.

The right quadrant focuses on very specific subareas related primarily to acoustics. These include acoustics applied to the atmosphere (RADIOACOUSTIC SOUNDING), and other aspects of atmospheric science (THEORY OF WIND).

The bottom quadrant provides the most balanced view of the two thrusts. It expands on the underwater propagation medium (STRATIFIED FLUID, SHEAR FLOW, INTERNAL GRAVITY WAVES), the radar platform issues (SATELLITE, PROCESSING OF RADAR), and the ocean surface issues (WIND WAVES, TURBULENT, OCEAN SURFACE). The integrated picture presented by the three quadrants is the use of radar from a space platform to view the ocean surface, and the research problems arising from the wind and undersea flows governing the conditions and structure of the ocean surface and impacting the interpretation of the radar images.

## CONCLUSIONS

Based on the results and interpretation of the multiword frequency analysis and the co-word analysis, the FASAC database used in this study is a compendium of those aspects of FSU science of interest to the U.S. for strategic and military purposes. The microlevel analysis of selected theme clusters, showing how the cluster members related to each theme, reinforced this conclusion and provided more detail about those aspects of each theme on which FASAC concentrated.

A wealth of information resulted from the FASAC output, and only a small fraction of that information was presented and analyzed in this paper. The analysis was restricted to technical themes and their relationships. Raw data was available for relating technical themes to non-technical themes such as institutions, scientists, journals, and geographical regions.

In the future, full text co-word analysis could be used to obtain a more representative structure of FSU (or any other country's) science. If a large number of randomly selected published FSU scientific papers were entered into a database, then a multiword frequency analysis and co-word analysis could be performed on this text database.

Assume that a paper represents about \$100K worth of effort. A 10,000 paper database would represent \$1B worth of effort, and would offer a very representative sample of FSU science output. The 10,000 paper database could be analyzed on an existing advanced desktop computer. The critical path would be assembling this database, not analyzing it.

Full text co-word analysis is in its formative stages. Much development remains to be done to understand the breadth of analyses which can be performed and the breadth of applications which can be covered. It is hoped that the initial techniques and results reported in this study will motivate and stimulate other organizations and researchers to develop and apply the general technique of full text co-word analysis on a much broader scale.

## ACKNOWLEDGMENT

The authors wish to note the contribution of Mr. David Miles in creating the computer-based algorithm, *TEXT SLICER*.

## BIBLIOGRAPHY

Kostoff, R. N. "Database Tomography: Multidisciplinary Research Thrusts from Co-Word Analysis," Proceedings: *Portland International Conference on Management of Engineering and Technology*, 27-31 October 1991. More detailed paper available from author.

-----, "Research Impact Assessment," Presented at *Third International Conference on Management of Technology*, Miami, FL, 17-21 February 1992. Larger text available from author.



Kostoff, R. N. "Co-Word Analysis for the Evaluation of R&D," in *Assessing R&D Impacts: Method and Practice*, B. Bozeman and J. Melkers, eds. Kluwer Academic Publishers, Norwell, MA, 1993a.

----- . "Database Tomography: Origins and Applications," *Competitive Intelligence Review*, Special Issue on Technology, 1993b, to be published.

----- . "Database Tomography for Technical Intelligence," *Competitive Intelligence Review*, Volume 4, Number 1, Spring 1993c.

----- . "Database Tomography: Origins And Applications," *Competitive Intelligence Review*, Special Issue on Technology, Volume 5, Number 1, Spring 1994, Wiley & Sons.

NAS. "Reorientation of the Research Capability of the Former Soviet Union," Results of a Workshop on 3 March 1992, NAS, NAE, IOM, National Academy Press, Washington, D.C., 1992.

#### AUTHOR IDENTIFICATION

Ronald Neil Kostoff received a Ph.D. in Aerospace and Mechanical Sciences from Princeton University in 1967. At Bell Labs, he performed technical studies in support of the Office of Manned Space Flight, and economic and financial studies in support of AT&T Headquarters. At the U.S. Department of Energy, he managed the Nuclear Applied Technology Development Division, the Fusion Systems Studies Program, and the Advanced Technology Program. At the Office of Naval Research, he is Director of Technical Assessment, and his present interests revolve around improved methods to assess the impact of research.

Henry Joseph Eberhart holds a Masters degree in Mechanical Engineering from the University of Southern California and a Bachelors degree in Engineering from the University of California, Los Angeles. His entire career has been with the Navy at China Lake, California and has involved materials, stress analysis, and test and evaluation of air-to-surface missile systems. He is presently engaged with the planning functions at what is now the Naval Air Warfare Center Weapons Division China Lake.

33 B2  
5506  
p. 8

**AUTOMATED MAINFRAME DATA COLLECTION  
IN A NETWORK ENVIRONMENT**

**David L. Gross  
Computer Engineer  
Analex Space Systems, Inc.  
NASA Kennedy Space Center  
Post Office Box 21206  
Kennedy Space Center, FL 32815-0206  
Phone: 407/861-5716  
Fax: 407/861-5774**

**ABSTRACT**

The progress and direction of the computer industry have resulted in widespread use of dissimilar and incompatible mainframe data systems. Data collection from these multiple systems is a labor intensive task. In the past, data collection has been restricted to the efforts of personnel specially trained on each system. Information is one of the most important resources an organization has. Any improvement in an organization's ability to access and manage that information provides a competitive advantage. This problem of data collection is compounded at NASA sites by multi-center and contractor operations. The Centralized Automated Data Retrieval System (CADRS) is designed to provide a common interface that would permit data access, query, and retrieval from multiple contractor and NASA systems. The methods developed for CADRS have a strong commercial potential in that they would be applicable for any industry that needs inter-department, inter-company, or inter-agency data communications. The widespread use of multi-system data networks, that combine older legacy systems with newer decentralized networks, has made data retrieval a critical problem for information dependent industries. Implementing the technology discussed in this paper would reduce operational expenses and improve data collection on these composite data systems.

**INTRODUCTION**

The need to access and retrieve data from mainframe systems is a widespread labor intensive activity. A number of commercial products based on the client/server concept are available to solve this problem. In a client/server system the "client" portion of the applications reside on workstations or Local Area Networks (LAN) with the "server" portion running on larger machines (i.e. mainframes). Economically the cost of purchasing, installing, and maintaining such products on one or more systems can outweigh the savings in manhours. These systems do save time in data retrieval and system access but they require a significant initial investment in additional training, equipment, and software development tools. The cost and time required for data retrievals increase geometrically when multiple, usually dissimilar systems are integrated. Tying different systems together means connecting incompatible architectures, protocols and languages. This paper discusses a composite system that can perform many of the same retrieval functions of a client/server system but without the technical restrictions and financial overhead involved.