New Jersey Institute of Technology

Digital Commons @ NJIT

Spring 5-31-1999

Development of a context-specific search engine, an executive information system, and a novel www ready external cost model

Amit V. Revankar
New Jersey Institute of Technology

Follow this and additional works at: https://digitalcommons.njit.edu/theses

Part of the Electrical and Electronics Commons

Recommended Citation

ABSTRACT

DEVELOPMENT OF A CONTEXT-SPECIFIC SEARCH ENGINE, AN EXECUTIVE INFORMATION SYSTEM, AND A NOVEL WWW READY EXTERNAL COST MODEL

by
Amit V. Revankar

NJPIES is associated with Information Ecology and Sustainability, a holistic approach to environmental data collection, compilation, integration and provision that puts people, not technology, at the center of the environmental information world.

The first main goal of this project was to develop an algorithm and associated computer-based tool that could perform a lifecycle cost analysis for a model system. The application developed solved the primary problem associated with the lifecycle cost analysis of a product: it accounted for *all* costs (e.g., environmental costs such as ecological costs and health costs associated with emissions) of the activity. A lifecycle cost analysis attempts to identify, measure, and quantify the social costs of human activities such as manufacturing that are not considered with traditional accounting systems. The application developed will quantify, monetize, and rank the damage or external costs to the environment of certain types of emissions. We developed a preliminary algorithm and software and implemented it at two plants: load assembly pack operation at Iowa Army Ammunition Plant (IAAAP) and Armtec, a manufacturer of combustible cartridge cases.

The second main goal of this project is to act as a credible information-clearing house in pollution prevention (P2) and related environmental matters, and to educate the public and keep them aware of facts taking place in the environmental/manufacturing

world. Intelligent search engines have been built to access these huge databases in human readable format and correlate the data to various reports providing information on the environmentally hazardous chemicals, releases, and facilities in different regions.

The third main goal is the enhancement of EnviroDaemon with a hierarchical information search interface. This project describes some approaches that locate information according to syntactic criteria, augmented by pragmatic aspects like the utilization of information in a certain context. The main emphasis of this project lies in the treatment of structured knowledge, where essential aspects about the topic of interest are encoded not only by the individual items, but also by their relationships among each other. Benefits of this approach are enhanced precision and approximate search in an already focused, context specific search engine for the environment.

DEVELOPMENT OF A CONTEXT-SPECIFIC SEARCH ENGINE, AN
EXECUTIVE INFORMATION SYSTEM, AND A NOVEL WWW READY
EXTERNAL COST MODEL

by
Amit V. Revankar

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
In Partial Fulfilment of the Requirement for the Degree of
Master of Science in Electrical Engineering

Department of Electrical and Computer Engineering

May 1999

Blank Page

APPROVAL PAGE

DEVELOPMENT OF A CONTEXT-SPECIFIC SEARCH ENGINE, AN EXECUTIVE INFORMATION SYSTEM, AND A NOVEL WWW READY EXTERNAL COST MODEL

Amit Vasant Revankar

Dr. Sirin Tekinay, Dissertation Advisor   Date
Assistant Professor of Electrical and Computer Engineering, NJIT

Dr. Franz J. Kurfess, Committee Member   Date
Assistant Professor of Computer and Information Science, NJIT

Dr. Marcus J. Healey, Committee Member   Date
Director, New Jersey Program for Information Ecology and Sustainability,
NJIT

BIOGRAPHICAL SKETCH

Author: Amit Vasant Revankar

Degree: Master of Science

Date: May 1999

Undergraduate and Graduate Education

- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 1999.

- Bachelor of Engineering in Electronics and Communication,
  Karnatak University, Karnataka, India, 1996.

Major: Electrical Engineering

To my Dad, mom, my brother's Sachin and Avinash, my sister Apoorva, my fiancée
Mrudula and all those who love me.

ACKNOWLEDGMENT

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1   Environmental Lifecycle Cost Analysis Model (ELCAM)

The objective of this dissertation was to solve the primary problem associated with the lifecycle cost analysis of a product or series of unit processes which is accounting for *all* costs (e.g., environmental costs such as ecological costs and health costs associated with emissions) of the activity. A lifecycle cost analysis attempts to identify, measure, and quantify the social costs of human activities such as manufacturing that are not considered with traditional accounting systems.

The ultimate goal of lifecycle cost analysis is to develop a system that will quantify, monetize, and rank the damage or costs to the environment of certain types of emissions. We developed a preliminary algorithm and software and implemented it at two plants those manufacture/assemble components associated with the IAAAP and Armtec. Cumulative prioritized results (based on sulfur dioxide emission costs and associated human health and ecological risk components) for external costs ($/year) per tank round (adapter, cap, combustible cartridge case, and load, assembly, pack only) for Armtec and IAAAP together rank Xylene, Resin Component A, Ethyl Acetate, Resin Component B, and Toluene in descending order.

Future work insofar as algorithm development entails: a) identification of alternate external cost estimation proxies other than sulfur dioxide [SO_2] and lead [Pb] for to-be-defined groupings of chemicals (e.g., halogenated, inorganic, volatile) employed in the product lifecycle; b) perception attribute estimation (e.g., using depleted uranium as a test case); c) synthesis of external lifecycle cost analysis module with one

1

for internal cost estimation; d) incorporation of more scientific rigor through implementation of mass balances; e) the ability to account for step increases in production; and f) genercize model with regard to different products, points along the business process chain, and diverse manufacturing and service entities.

Broad required features for the envisioned environmental lifecycle cost analysis model are as follows: a downloadable PC-based ELCAM software that will reside on the end-users' hard drives and will permit easy uploading via e-mail/ file transfer protocols (FTP) of end results of bill of material and/ or internal cost data investigations; a WebTop (i.e., WWW-based) master program which is able to function exactly like the PC version and is also capable of integrating final input from end-users; open architecture for desktop and WebTop applications so as to not limit its end-user usage now or in foreseeable future; provisions for extensibility and scalability so that different lifecycle steps and industries as well as more complex products can theoretically be examined with this approach in the future; transparency in approach, design, and implementation; and openness (able to work across different operating systems and hardware). Work also entailed general functionality in the areas of: a) automation (data collection, result calculation, process modelling); b) security (internal costs, bill of materials information entry); c) transparency (user-friendliness within a component and across the lifecycle of a product); d) graphics (GUI for end-users, appropriate tables and figures, secure interface development for contractor entry of bill of materials and internal cost data); and e) predictive capabilities (first generation "what-if" capabilities for alternate chemicals).

1.2 NJPIES Metrics

The main goal of NJPIES is to act as a credible information-clearing house in pollution prevention (P2) and related environmental matters, and to educate the public and keep them aware of facts taking place in the P2 world. We have valuable information obtained from various sources about various companies in New Jersey and their P2 strategy. This information is in the form of DEQ, TRI, etc. databases. We have an ambitious plan of creating a "Data Warehousing Project" from this application. Once these databases are available on the Internet, a whole source of information can be accessed by an end user from anywhere in the world.

My job in NJPIES was to develop the metrics. Which is a powerful way of displaying data for the end-user to minimise the waste. It basically refers to efforts aimed at reducing the flow of waste from manufacturing practices and accomplishing it with a margin of profit. This application takes publicly available, non-confidential State databases - adds value in terms of Organization and insight, and helps the user to analyze the data in a better and efficient manner based on which he/she can answer policy questions. It is going to be helpful to managerial level people in making executive decisions. For a common man, it will answer questions like how much toluene was released into the air by manufacturing companies in his County, City and the State as a whole. Anyone can do comparative studies of all the industries in New Jersey. The application also graphically studies the trends of Releases, emissions, Use and NPO over years. It also gives a graphical representation of distribution of companies in the State. Some of the question, which can be answered in context to NJPIES Metrics and NJEnviroDatabases Development, are as follows:

WHO is the "user": There could be quite a large variety of users.

Different Companies in New Jersey can access the database to gain knowledge about it's own statistical data, studying trends of production/inventory/chemicals used/ TRI data, and a whole bunch of information. That company can also see what other companies in it's own field of production, or other fields are doing. He/she can also find information about a specific chemical and a specific facility.

Examples of queries:

- ◆ List of all companies in any city(s) you select.

- ◆ List of all branches of one company.

- ◆ List of all chemicals and their details used by that company, including the year wise Data; or any specific year.

- ◆ Direct links will be provided to all relevant sites regarding P2.

1.3 ENVIRODAEMON

Scientists and engineers have long harboured grand hopes for immediate, distributed network access to the entire science and technology literature. These hopes are well on their way to being realised as a result of the steady improvement in the computing and communications infrastructure and the popularisation of the Internet. The size of the organization able to perform search has decreased as groups of laypeople and scientists can now search "digital libraries" without the aid of trained reference librarians. Similarly, the document being sought has changed: from a citation with descriptive headers, to an abstract, to complete multi-media contents including text, audio, video, and

animation. Often cited reasons for the rise in digital information are associated with the ideas of preservation of the contents of physical, paper-based texts, the convenience associated with maintaining, searching and retrieving electronic text, and the lowered cost of acquiring and maintaining bits as opposed to atoms (Lesk, 1997).

A library collects, maintains, and indexes information for purposes of search, retrieval, and display. Since the advent of on-line information retrieval more than 30 years ago the mechanism for retrieval has been syntactic: a user specifies a keyword-based query and all the documents containing those words are returned. In the mid-1980s to the present there has existed document search on local area networks and the Internet. This type of search is best characterized as structure-based because it involves dealing with structure and classification of complete documents and not just keywords. Shortly after the turn of the century, information retrieval will migrate from keyword-based syntactic searches to concept-driven semantic searches across gigantic, distributed collections (Schatz, 1997).

My role in this project was to develop the front-end interface to EnviroDaemon search engine, where the user gives the structure of his search. The search structure would be hierarchical. The interface also has many more facilities discussed in detail later in this paper. We also compare and contrast tools for Internet search in a context-specific area: the environment. We provide context in terms of the history of information retrieval and the evolution of its research focus, describe the importance and rate of growth of the World Wide Web (WWW) and current Web-based generic search engines, outline the need and tools employed to search the Internet, demonstrate how to measure

effectiveness in terms of document retrieval, and compare and contrast our context-based

search engine (njpies.njit.edu) to generic search engines.

CHAPTER 2

ENVIRONMENTAL LIFECYCLE COST ANALYSIS MODEL

2.1 Algorithm

2.1.1 Objective

A main objective of this study was to prioritize the overall lifecycle costs of toxic and hazardous materials used in the manufacture of tank ammunition.

- Toxic and Hazardous Substances of Concern (T&HSC) = "Hazmats"

- Accounts for external costs (i.e., public costs such human health and ecological risk costs)

- IAAAP and Armtec plants were places where we tested out algorithm and software

- "Costs" measured in dollars ($)

2.1.2 Rationale

Lifecycle assessments attempt to identify, measure and quantify the social costs of human activities such as manufacturing that are not considered with traditional accounting systems. The ultimate goal of this LCA is to develop a system that will quantify, monetize, and rank the damage or costs to the environment of certain types of emissions. The primary problem associated with a LCA of any manufacturing process is accounting for *all non-monetary* costs (e.g., environmental costs such as ecological damage due to pollution in an ecosystem and health costs such as increased incidence of asthma

associated with air emissions) of the activity[1]. Private costs, any monetary out of pocket expense such as material costs and wages, while often proprietary, are easily identifiable and quantifiable. As such, accounting for the private costs in a LCA, when data is available, is a straightforward exercise.

The scoring algorithm is based on measuring external costs of weapons manufacturing, but can be applied to other areas. What we have developed is a very preliminary estimation of external costs, in part by compiling data on the toxicity of chemicals with society's willingness to pay to reduce pollution. What is needed is a method of quantifying the costs associated with the perceived risks of hazardous emissions. Because the costs occur outside traditional markets (you cannot buy or sell credits that will increase or decrease the probability that you contract a certain disease), non-market valuation methods are needed to identify, measure, then quantify the social costs associated with potentially toxic chemical emissions (e.g., depleted uranium).

The numbers from the algorithm, the external costs, are the estimated costs of the damage to human health and the environment that are not paid during the production of weapons. A corollary is that if a non-smoking individual contracts lung cancer from breathing second hand smoke, this is an external cost, because it is not included in the price of cigarettes.

In order to compare external costs with private costs, the two must be measured in common units. The way most items are measured is in dollars per unit. The algorithm

---

[1] An external cost is a cost that is imposed on an agent who is not part of the activity that generates that cost. External cost is often used synonomously with environmental cost. External cost is used in this report due to the fact that it is broader because it incorporates environmental costs and all other non-monetary costs. Social cost and external cost will be used interchangeably.

measures external costs in dollars per round of ammunition manufactured. It is likely that the accountants at munitions facilities also measure production in dollars per round. If they do not, all that needs to be done is to divide the total cost of production by the number of rounds produced during a given time period to obtain the $/round value for internal or private costs.

Thus LCA attempts to identify, measure, and quantify the social costs of human activities such as manufacturing that are not presently considered. Techniques have been and are being developed, that quantifies, sometimes monetize, and rank the damage or costs to the environment of certain types of emissions.

2.1.3 Benefits to User

A salient benefit to the end-user is the identification of large contributions to overall lifecycle environmental costs.

– Focus on "high-payoff" T&HSC changes

– Design for lowest environmental cost materials

Another benefit is the assimilation of a vast array of product lifecycle and environmental information

– Computer-based decision tool

– Environmental Lifecycle simulation

A later section of this report shows preliminary results of methods that have been used to identify, quantify, monetize and rank the external costs associated with use and emissions of materials used in the production of the M829A2 training round at the Iowa Army Ammunition Plant.

2.1.4 How the Model Works

There exist two broad ways to estimate external costs associated with human activity. The first method is to measure all impacts of all material emissions at all levels. The second way to estimate external costs associated with human activity is to measure impacts of some materials on some levels based on existing data and use this information to approximate what other materials do to other areas. This is what we have been doing to date. As can be seen from the algorithm, many categories of external costs are nothing but broad approximations, which need to be improved.

Based upon USEPA data sets for risk- Waste Minimization Prioritization Tool [WMPT] (Beta Test Version 1.0), EPA530-R-97-019, June 1997.

- – Chemical toxicity data

- – Sulfur Dioxide [SO_2] used as basis at one end of range (average market value for emissions: $80/ton which is used as a proxy for the external or social cost) (Chicago Board of Trade, CBOT, regularly sells/ buys emission credits for this type of emission)

- – Lead [Pb] used as basis at other end of range (literature value for costs associated with required regulatory controls for airborne lead: $1600/lb. which is used as a proxy for the external or social cost) (Ackerman, 1997)

Toxicity data for other chemicals is compared to SO_2 or lead, which is not measured directly, but inferred.

2.1.5 Attributes Employed

External (public or social) and internal (private) costs are currently, or will be, estimated

by trying to quantify and monetize the following attributes and sub-attributes.

- Regulatory Concerns (mostly private)

- Worker Health & Safety (public)

- Management Costs (private)

- Ecological Costs (public)

- Process & Product Details (private)

- Perception (public)

2.1.6 Algorithm Employed - Pure Chemical

Units used:

Final Quantity used (ml/round)

Total external cost ($/round)

Specific gravity (g/ml)

For Pure Chemical using SO_2 as basis -

The score for a pure chemical is calculated using six attributes:

/*Implemented in the application using SQL */

Regulatory Concerns: (sum of Agency listed (1 unit/each list T&HSC found on) in

WMPT) * $100

Worker Health and Safety: (Human Exposure Potential Risk Unit [RU] from WMPT) * $0.02

Management Costs: (sum (0.25 units for each sub-attribute listed) * $100

Ecological Cost: (Ecological Risk Potential from WMPT) * $0.02

Process and product cost: (sum (0.25 units for each of the sub-attributes)) * $1000

Perception: (Human Health Risk Potential from WMPT) * $0.02

Procedure:

Import the toxicity data from WMPT for SO_2 or Pb.

Calculate the cost values for each attribute as described above.

Sum the attributes to get the External Cost.

Then calculate, Quantity used (g/round) = Quantity used (ml/round)(from some facility or point along the product lifecycle chain) * Specific gravity (g/ml)

The Total External Cost is calculated as follows:

Final External Costs (SO_2) = External cost * Quantity used * Specific gravity

($/round) = ($0.02/lb•[SO_2]-RU) * (ml/round) * (g/ml) * (lb/454g) * Risk Units for Chemical in Question (RU)

or for Pb

Final External Costs (Pb) = External cost * Quantity used * Specific gravity

(\$/round) = (\$123/lb•[Pb]-RU) * (ml/round) * (g/ml) * (lb/454g) * Risk Units for Chemical in Question (RU)

Constants: \$0.02/lb•[SO_2]-RU; \$123/lb•[Pb]-RU

Variables: ml/round for each chemical used; Risk Units for Chemical in Question (RU); *N.B. If enduser seeks final unit of \$/year, one has to input rounds/year as a variable in the algorithm. Any other product (e.g., adapters, caps, combustible cartridge cases) can be used in the place of rounds/year.*

Conversion Factor = (lb/454g)

The quantity used for each pure chemical also includes the percentage of that chemical used in the mixture.

2.1.7 Algorithm Employed - Mixture

For the Mixture using SO_2 as basis -

Regulatory Concerns: (sum of Agency listed (1 unit/each list T&HSC found on) in WMPT for all the chemicals in the mixture) * \$100

Worker Health and Safety: (sum (Human exposure potential for all the chemicals in the mixture)) * \$0.02

Management Costs: (sum (0.25 units for each sub attribute listed for all the chemicals in the mixture)) * \$100

Ecological Cost: (sum (Ecological Risk Potential for all the chemicals in the mixture)) * $0.02

Process and product details: (sum (0.25 units for each of the sub-attributes for all the chemicals in the mixture)) * $1000

Perception: (sum (Human health risk potential for all the chemicals in the mixture)) * $0.02

Procedure:

Import the toxicity data from WMPT.

Calculate the cost values for each attribute as described above.

Sum the attributes to get the External Cost.

Then calculate, quantity used (g/round) = Quantity used (ml/round)(IAAAP or Armtec) * specific gravity (g/ml)

Salient difference between pure chemicals and mixtures:

Quantity used is multiplied by the specific gravity for each chemical according to the percentage used.

Example:

X1 (for one chemical in the mixture) = Quantity used (ml/round) * specific

gravity (g/ml) * Percentage of that 1 chemical in mixture according to Material Safety

Data Sheets (MSDS) sheets.

the Sum (X) = X1 + X2 +... , where Sum (X) is the Final Quantity used

Final External Costs (SO_2)= External cost * Quantity used * Specific Gravity * percentage

($/round) = ($0.02/lb•[SO_2]-RU) * (ml/round) * (g/ml) * (lb/454g) * Risk Units for Chemical in Question (RU) * percentage of chemical in mixture

($/round) = ($0.02/•[SO_2]-RU) * (ml/round) * (g/ml) * (lb/454g) * RU * percentage of chemical in mixture

or

Final External Costs (Pb) = External cost * Quantity used * Specific gravity

($/round) = ($123/lb•[Pb]-RU) * (ml/round) * (g/ml) * (lb/454g) * Risk Units for Chemical in Question (RU) * percentage of chemical in mixture

Constants: $0.02/lb•[$SO_2$]-RU; $123/lb•[Pb]-RU

Variables: ml/round for each chemical used; Risk Units for Chemical in Question (RU)

Conversion Factor: (lb/454g)

2.1.8 Preliminary Estimation of External Costs

Described in this paper is a method of quantifying, monetizing, and then ranking most external costs associated with emissions of materials used in the production of M829A2 training round at IAAAP, Armtec, or some other facility. This method can be applied to emissions of any chemicals from any manufacturing source. Some categories of external cost estimation are likely close to the actual true social cost and can be justified based on current existing research. Other categories are only meant to provide a broad guide to what actual social costs are and will be refined in the future.

2.2 System Architecture

Table 1. JAM/Web & MS Access 97 Components

Product	H/W	OS
MS Access 97	Dell	Windows NT 4.0
JAM/WEB Application Server	Dell	Windows NT 4.0
JAM/WEB Builder	Dell	Windows NT 4.0
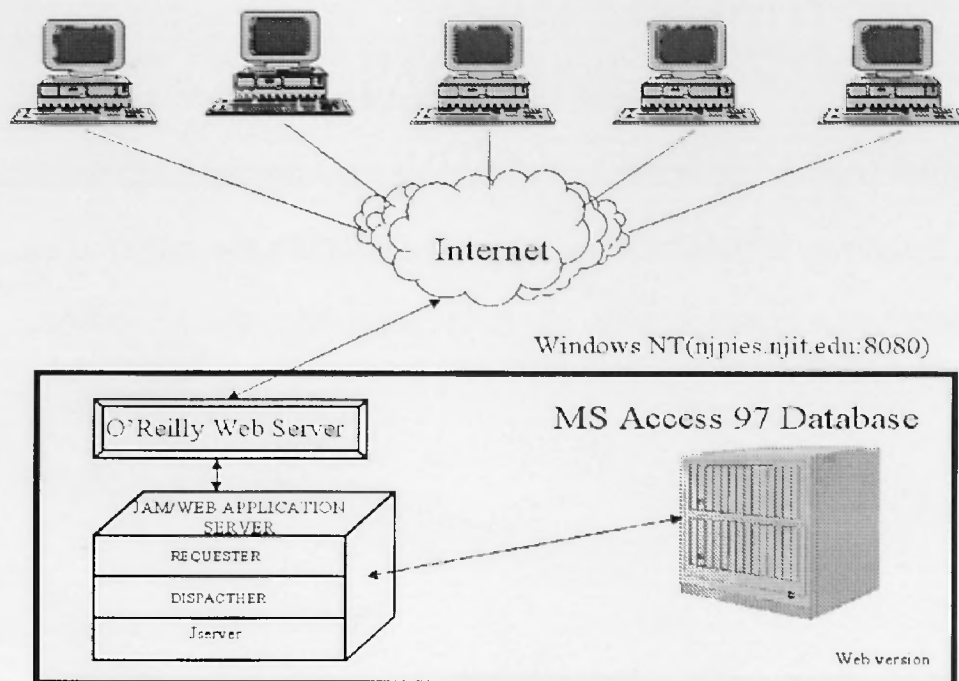
Figure 1. Web Version Environment
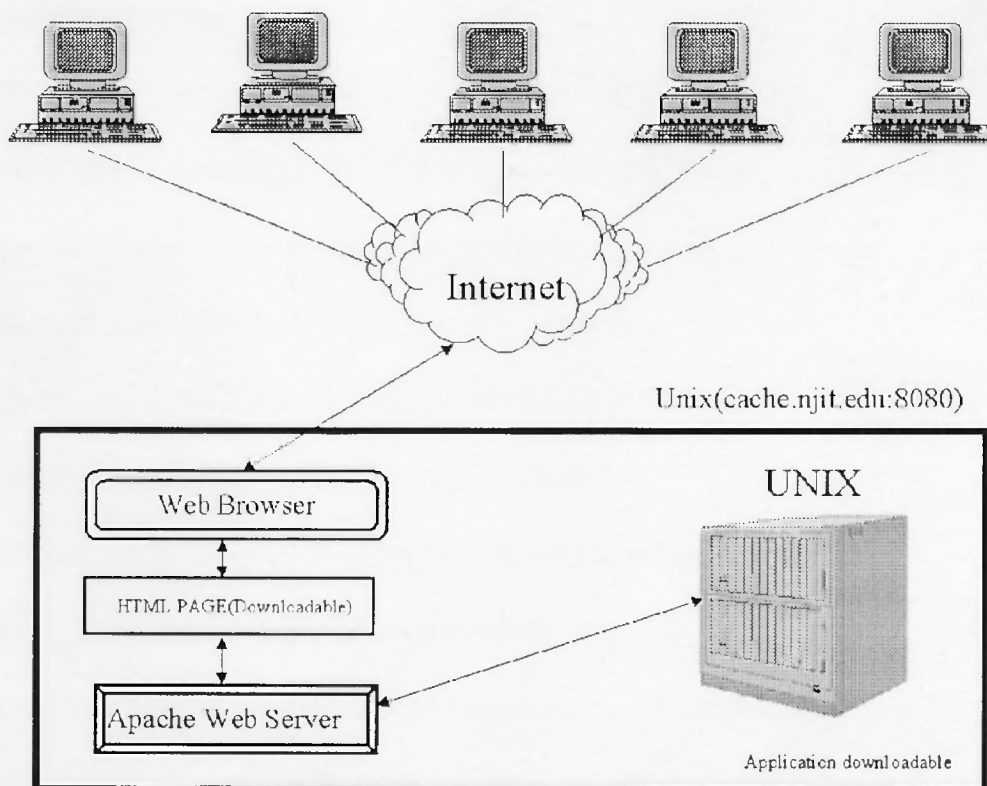
Figure 2. Downloadable Version Environment

JAM/WEB consists of the following elements:

- MS Access 97 development software with database

- JAM/WEB Application Server – application processing software that resides on the same machine with HTTP Server and processes JAM/WEB applications.

- JAM/Web Builder – development software which enables to design Web applications

- Apache HTTP Server 1.3.0

- O'Reilly's Web server

2.2.1 MS Access 97

Features in Microsoft Access 97

Microsoft Access 97 offers many new and improved features to help you create powerful database applications.

- New objects, properties, methods, and other language elements

- Accessing the Internet or an intranet from your application

- Creating custom objects with class modules

- Customizing menus and toolbars in your application

- Removing source code from your application

- Replicating only a specified part of a database

- Working with version 3.5 of the Microsoft Jet database engine

- Using new features in the Module window

- Using the Object Browser as a reference for objects and their members

- Using DAO to access ODBC databases without loading the Microsoft Jet database engine

- Creating a tabbed dialog box or multiple-page form with the tab control

- Setting references programmatically

- Using the enhanced Debug window

- Improving compilation performance

2.2.2 O'Reilly's Web Server

The Web site server....

- is a powerful, flexible, HTTP/1.1 compliant web server

- implements the latest protocols, including HTTP/1.1

- is highly configurable and extensible with third-party modules

- runs on Windows NT 4.0 without modification

- is actively being developed

- encourages user feedback through new ideas, bug reports and patches

Information on the latest version of Web site can be found on the O'Reilly's web server

The URL for the web server is http://website.oreilly.com. This will list the current

release, any more recent beta-test release, together with details of mirror web and

anonymous ftp sites.

2.3 Simulation

The final software exist in two different forms: a downloadable PC-based ELCAM

software that will reside on the end-users' hard drives and will permit easy uploading via

e-mail/ file transfer protocols (FTP) of end results of bill of material and/or internal cost

data investigations; a WebTop (i.e., WWW-based) master program which is able to

function exactly like the PC version and is also capable of integrating final input from

end-users.

2.3.1 Automation

There are a large number of necessary calculations for a) mixtures; b) within a single product; c) at a facility; and d) at various points along the entire lifecycle of a complex, multi-component product such as the M829A2. Similarly, there will be a large, and increasing stream, of data that will need to be performed *in situ* for contractors and then mailed/ FTP to a Web-based interface. Process modeling will also be performed by end-users and will require mass balance checking on a unit process level. All these requirements point toward the need for automation in the program whenever and wherever it a) appears possible; and b) makes the most sense.

A graphical "Process Modeling Unit" either as a separate tool or as integrated with the costing tool is actively under consideration for development in either late year 2 or year 3 of this project - money permitting.

The projected tools share common properties. Our initial thoughts are to design a "FAMILY OF PRODUCTS" strategy, and incrementally add new functionality for successive tools. As part of this strategy, tool modules to be constructed will be selected, groups of the modules will be selected to be integrated, and their graphical versions and/or Web versions would be implemented. Expectations are to develop tools with graphical representations for processes within which costing analysis is enabled.

The following building blocks will be used:

Existing Modules:

- Environmental Lifecycle External Costing Model (ELECAM)
- Internal Costing (ECAM)

Potential Future Modules:

- Graphical Process Modeling for Manufacturing Phase and other phases (Design, Transport/ Storage, Demilitarization.)

Presentation Media:

- MS Access 97 applications

- Graphical User Interfaces

- Web Applications

One of our strengths is in the integration of the modules with GUIs for presentation and user ease, as well as WWW deployment. The modules has been implemented as, for example, an MS Access database application first, and then converted to sophisticated MS Windows applications with GUIs. Later the same end-user "feel" has been provided in a WWW version where a central consistent database could be shared among users at different locations with a variety of access rights.

2.3.2 Security

It is clear that a) the preponderance of the necessary data needs to come from end-users such as contractors; b) this data needs to be elicited from them with their continuing interest and active involvement; c) there will be a continuing need to stress end-user application friendliness; and d) the absolute, unequivocal protection of their proprietary interests. In order to fulfill these needs, a first generation security mechanism needs to be elucidated so that we may acquire the appropriate information without any hint of compromising their proprietary information and not taking up unnecessary contractor time. This can be initially accomplished by the provision of a downloadable (password protected, but this application is now available from http://cache.njit.edu:8080/esdm)

software that will reside on the hard drive of the end-user. Provisions will be made for uploading of summary information via e-mail/ FTP to WWW-based version of this application. So long as there is agreement on the final information that will be uploaded, and the application is sufficiently user-friendly, we will be able to test out this first generation security capability as we extend down the product lifecycle chain.

2.3.3 Transparency

Transparency refers to the need to lay bare what is needed to appropriate end-users without exposing implementation or software architecture specifics. This approach applies equally to end-users and developers in that both parties should have transparent access to the data they require. Developers will need to leave appropriate online documentation for subsequent application developers.

2.3.4 Openness

Open architecture refers to a system in which the specifications are made public in order to encourage 3rd party vendors to develop ancillary products. In the case of this application, it refers to the ability to work across different operating systems and hardware so as to ensure continued longevity of the product.

2.3.5 Graphics

GUI for end-users, appropriate tables and figures, secure interface development for contractor entry of bill of materials and internal cost data.

A key element is to respond to the perceived expectations, including:

Graphical representation of the processes;

Integration among existing tools and the ones being developed by others; Porting some of the tools to World Wide Web to run inside browsers; And genericizing the outcome to different fields and different phases.

2.3.6 Predictive Capabilities

A first generation "what-if" capability we are exploring is the ability to a) view initial results; b) ask "what-if" one exchanges one new chemical for one that is too high on the priority list; and c) immediately view the results.

2.3.7 Master-Detail

A master-detail form is a one-to-many form where one part of screen is a master and the other part shows the details depending on the input to a master part of a form by the user. In our case, master form consists of a search criteria where a user will be asked to enter a chemical name OR cas number, based on his selection a user will see multiple records on the next screen which we will call as detail form.

We choose master-detail strategy, as user will be able to see search criteria as well as the result part being populated with multiple records fetched from the back-end database.

For example, see following screen figures in the next section.

2.3.8 Application Design

Below Figure shows exactly how the application goes from one screen to another screen based on the search criteria. All the screens mentioned in the rectangular blocks above have been covered in detail in following Sections.
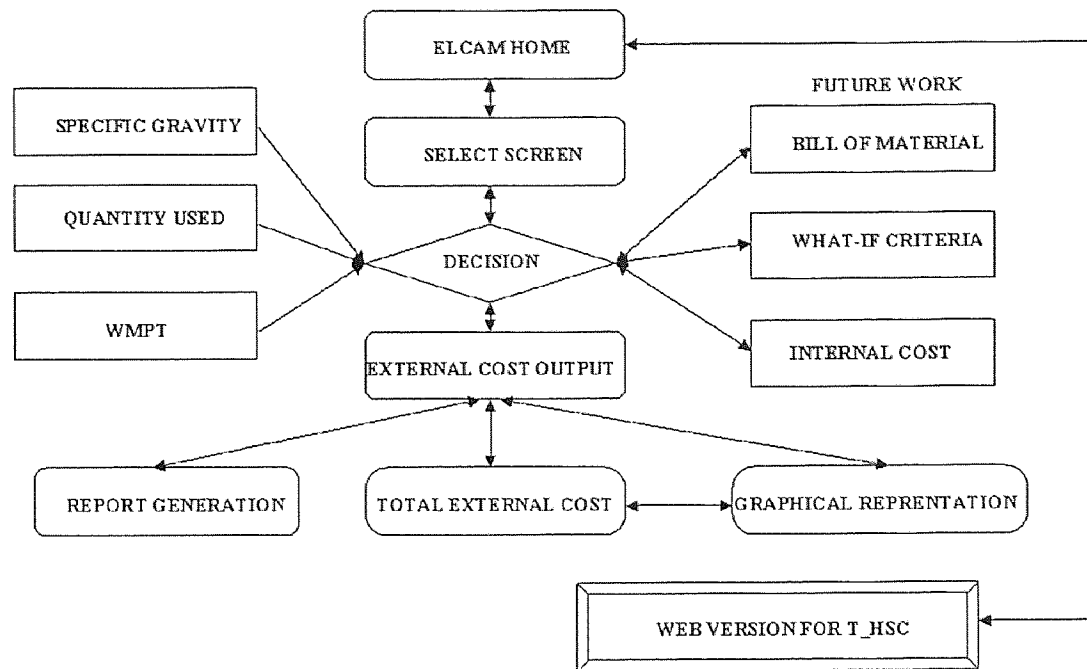
Figure 3. Application flow diagram

2.3.9 System Design Overview

NJPIES was designed with consistent page layout that conforms to NJPIES applications

standards, which reduces the amount of mental efforts the user needs to make. Every

screen was equipped with navigation buttons that allow the user to navigate the

application with ease.

2.3.9.1 System Organizational Chart: NJPIES home page (http://njpies.njit.edu) listed

program mission statement and provides links to the member products and services page.

The Products related to my thesis are NJ EnviroDatabase, Environmental Lifecycle cost analysis Model, and EnviroDaemon. I have emphasized on the ELCAM.

2.3.9.2 Select a Chemical and Select the Facility: Allows the user to view the data related to chemical and facility where the chemical is used. The user must select the chemical name and facility name which is abbreviated. Then submit the selection. After submit the data for regulatory cost, ecological cost, public and process details, perception, management cost, worker health and safety and quantity used will be displayed on the same screen.

The user also has five different options mentioned below:

- Final reports

- Chart view

- Graphical view

- Values for Sulphur Dioxide

- Help

Each buttons has its own significance.

There is also another option for the user to input data into the database. The user has to use the feedback button on the top of home screen.

User Output Specification: The user output for a Web transactional, database-oriented application is delivered by the Web browser display. After the user specifies the selection criteria, through the web server the requested screen will be sent to JAM Application

server. The selection result HTML generated by JAM Application server will be sent back to Web server, and back to the browser that requested it.

Identification of Output Data: The data return by the database will be displayed in a tabular format on the browser. For each selection, the optimal output is the chemical detail record screen that displays different information as described before.

Destination of Output Data: The output data will be delivered to the requested browser. The load on the server is reduced by implementing the JAM Application features, which allocates server processes and database connection on a per-server basis, not a per-user basis, and re-uses them for each request.

Output Medium and/or Device: The browser will send the user input to the Web sever, which passes the screen name to the JAM Application server using the CGI interface. The requester accepts the CGI request from the Web server and pass it to a Jserver process, and wait for and transmit a response. The Jserver is responsible to for all application –specific processing.

The HTML generated by JAM Application sever will then be submitted through the Web server to the requested browser.

Output Format/Syntax: The output format is displayed as HTML in the browser that requested the selection.

Screens for search by chemical name or facility:

The search criteria for this screen are by Chemical name and Facility.

The user has two options:

      i) Select the Chemical Name and Facility

      ii) View the Reports

    When the user selects one of the options, he will be taken to the detail screen,

which then follows a flow of screens. The flow gives the end-user the detail information

of external cost and related graphical represented outputs.

Screen Flow 1: Following flow will be observed when user selects search criteria "Select

chemical", he will see the following screen as shown in figure

Figure 4. Search by Chemical Name and Facility (Master Form)

Figure 5. Search By Chemical Name and Facility (Detail Form)
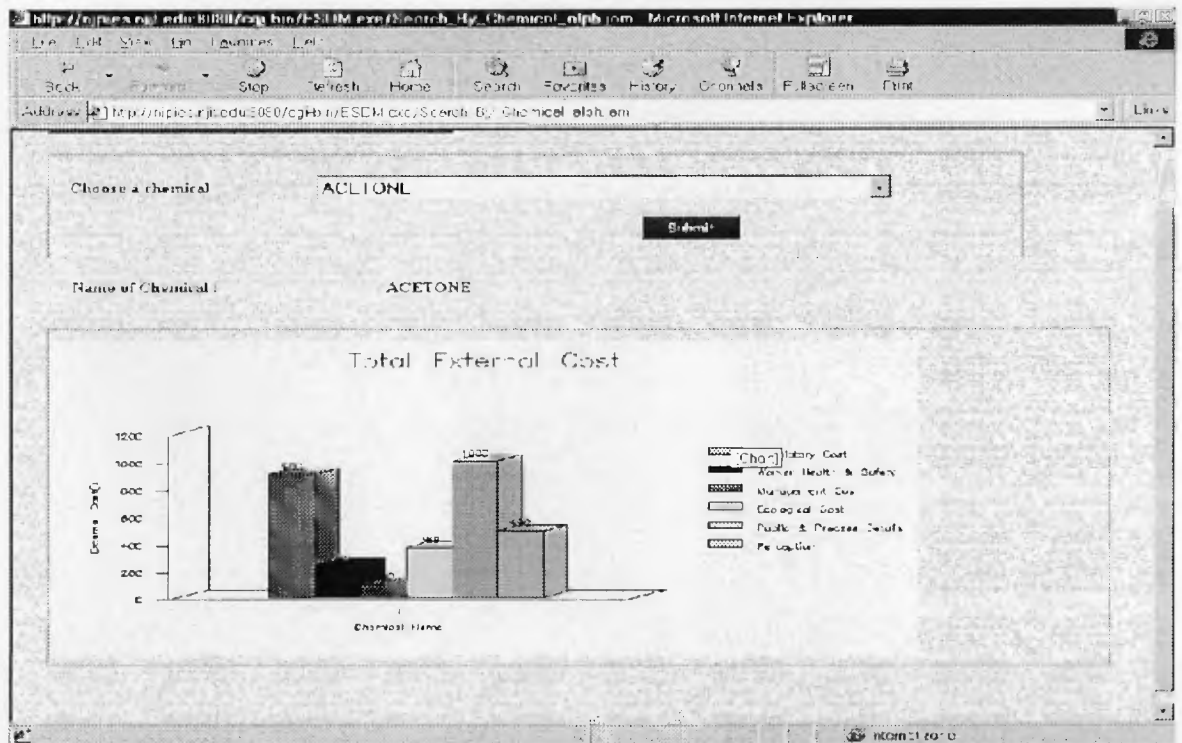
Figure 6. Graph for the Selected Chemical Name and Facility
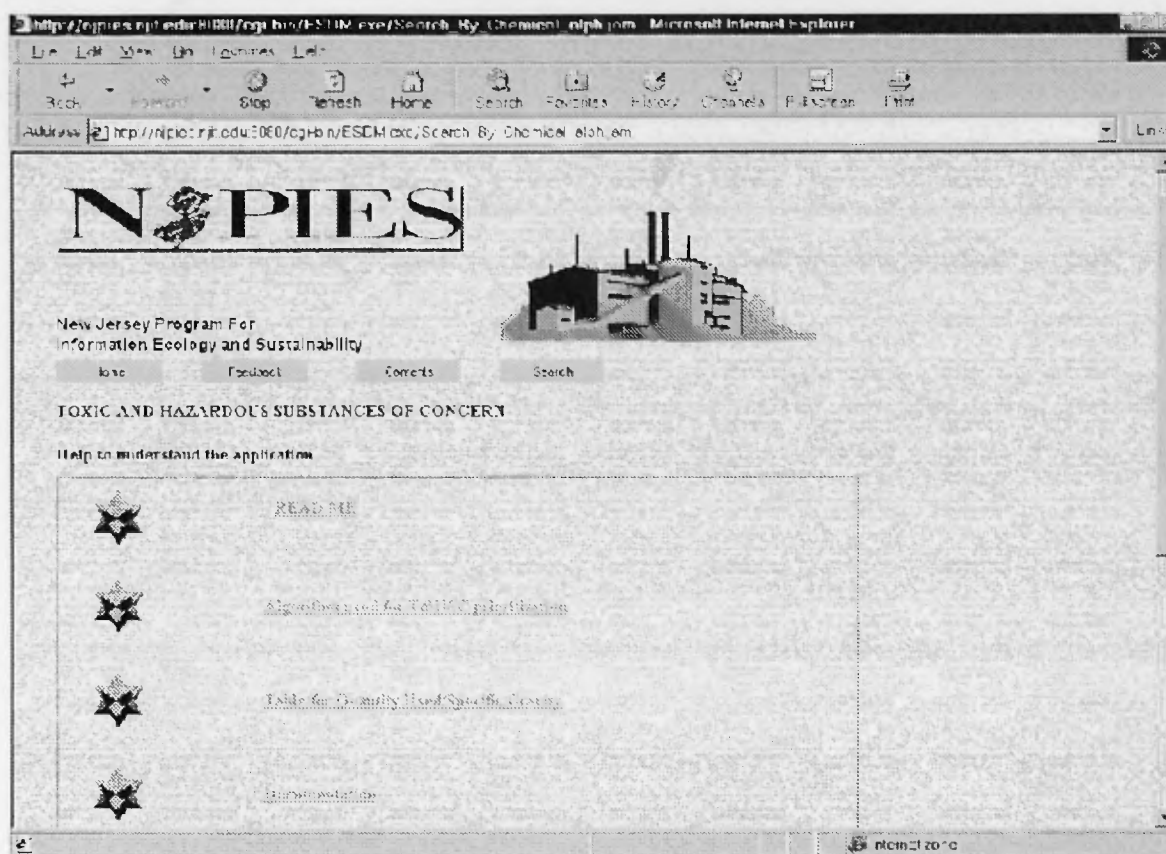
Figure 7. Help to understand the Application

2.3.10 Reports

This section tells about the final results generated by the application. Every facility has its own report generated which helps to analysis the chemical ranking in terms of there external cost and the internal cost of the chemical. The figures below will give the exact idea about ranking.
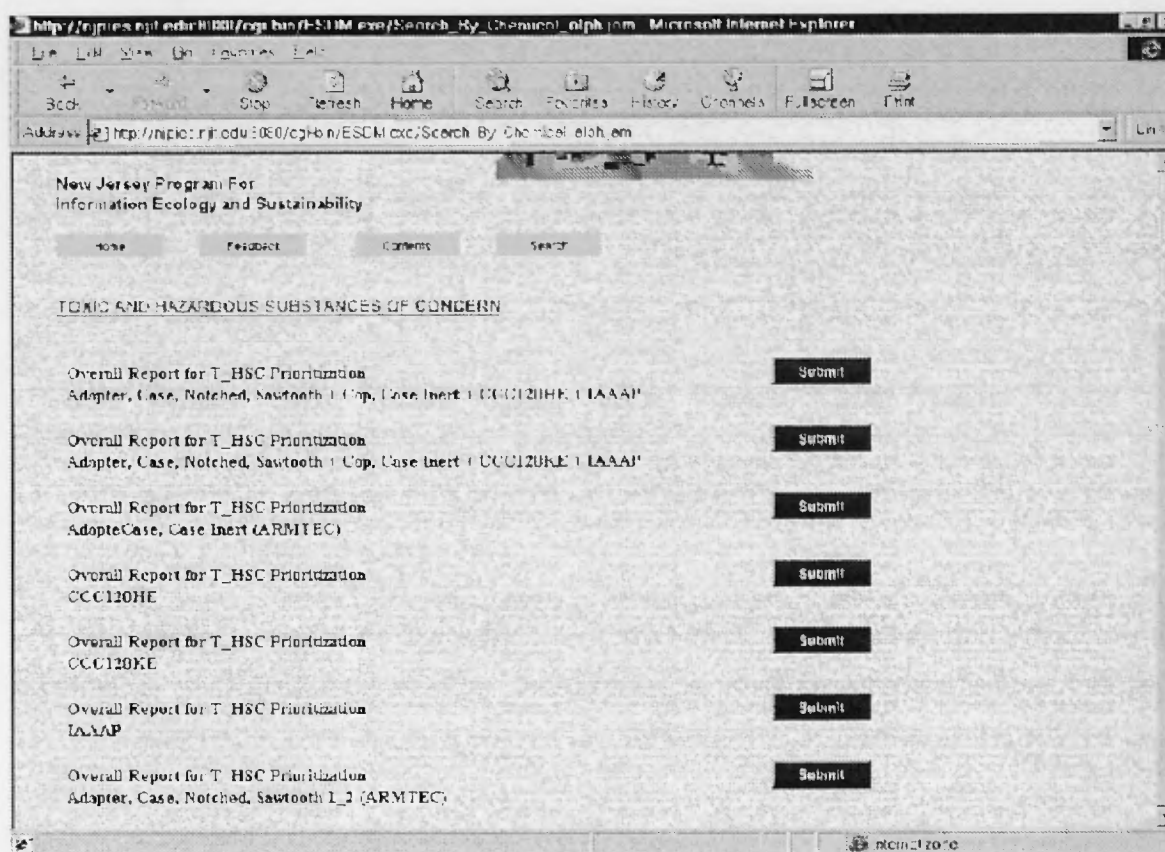
Figure 8. View reports for six different facilities

Mentioned below are the six different facilities shown in the figure 8.

- Overall Report for T_HSC Prioritization (Adapter, Case, Notched, Sawtooth + Cap, Case Inert + CCC120HE + IAAAP)

- Overall Report for T_HSC Prioritization (Adapter, Case, Notched, Sawtooth + Cap, Case Inert + CCC120KE + IAAAP)

- Overall Report for T_HSC Prioritization (Adapter, Case, Case Inert (ARMTEC))

- Overall Report for T_HSC Prioritization (CCC120HE)

- Overall Report for T_HSC Prioritization (CCC120KE)

- Overall Report for T_HSC Prioritization (IAAAP)

- Overall Report for T_HSC Prioritization (Adapter, Case, Notched, Sawtooth 1_2 (ARMTEC)).

Figure 9. View reports for IAAAP

2.4 Summary

The overarching concept for this project was the fulfillment of a need to incorporate environmental considerations into the lifecycle of products. In order to bring life to this concept, a means to perform an environmental lifecycle cost analysis and simulation of a typical, multi-component product (M829A2 tank ammunition round) was developed. The primary problem associated with the lifecycle cost analysis of a product or series of unit processes is accounting for *all* costs (e.g., environmental costs such as ecological costs and health costs associated with emissions) of the activity. A lifecycle cost analysis attempts to identify, measure, and quantify the social costs of human activities such as

manufacturing that are not considered with traditional accounting systems. The ultimate goal of lifecycle cost analysis is to develop a system that will quantify, monetize, and rank the damage or costs to the environment of certain types of emissions.

The broad future goals of this project are to a) further develop this algorithm and associated computer-based tool; b) continue to perform lifecycle cost analyses (internal and external costs) using the developed algorithm and software at representative facilities (Radford, VA propellant manufacture facility, depleted uranium penetrant manufacture facility) along the business process chain: design, manufacture, transport/ storage, use, and post-use; and c) identify future work needed to continue to develop the algorithm and model so that it can be more generic and applicable to the environmental lifecycles of different products from diverse industries.

Future work insofar as algorithm development entails: a) identification of alternate external cost estimation proxies other than sulfur dioxide [SO_2] and lead [Pb] for to-be-defined groupings of chemicals (e.g., halogenated, inorganic, volatile) employed in the product lifecycle; b) perception attribute estimation (e.g., using depleted uranium as a test case); c) synthesis of external lifecycle cost analysis module with one for internal cost estimation; d) incorporation of more scientific rigor through implementation of mass balances; e) the ability to account for step increases in production; and f) genercize model with regard to different products, points along the business process chain, and diverse manufacturing and service entities.

Broad required features for the envisioned environmental lifecycle cost analysis model are as follows: a downloadable PC-based ELCAM software that will reside on the end-users' hard drives and will permit easy uploading via e-mail/ file transfer protocols

(FTP) of end results of bill of material and/ or internal cost data investigations; a WebTop (i.e., WWW-based) master program which is able to function exactly like the PC version and is also capable of integrating final input from end-users; open architecture for desktop and WebTop applications so as to not limit its end-user usage now or in foreseeable future; provisions for extensibility and scalability so that different lifecycle steps and industries as well as more complex products can theoretically be examined with this approach in the future; transparency in approach, design, and implementation; and openness (able to work across different operating systems and hardware).

Envisioned future work for the software entails addition of general functionality in the areas of: a) automation (data collection, result calculation, process modeling); b) security (internal costs, bill of materials information entry); c) transparency (user-friendliness within a component and across the lifecycle of a product); d) graphics (GUI for end-users, appropriate tables and figures, secure interface development for contractor entry of bill of materials and internal cost data); and e) predictive capabilities (first generation "what-if" capabilities for alternate chemicals).

- Product to date reflects current thinking and will evolve

- Need to continue to develop the rationale and scientific basis for the algorithm

- IAAAP and Armtec data gave us a good start but work needs to be continued elsewhere.

- Need end-user feedback

  – most important characteristics

  – format of data and analysis

    – breakthrough ideas allied to this.

CHAPTER   3

NJPIES MARICS

3.1 Objective

The main goal of using metrics in NJPIES is to minimize the waste. This refers to efforts aimed at reducing the flow of wastes from manufacturing practices and accomplishing it with a margin of profit.

Somewhat synonymous terms for Waste Minimization:

- Waste Reduction (a 'no-no' in classic P2)

- Clean Technologies

- Pollution Prevention

- Low and non-waste technologies

The other goal is to answer some of the questions like:

- Do companies release or transfer off-site the same chemicals that they acquire as inputs?

- What companies submitted data that is significantly out of balance? (Can we assume that there is no room for improvement in the way they manage their chemical flows?)

- What are the top chemicals in terms of total releases? Air releases? Water? Land?

- What are the top industries (SIC) in terms of total releases? Air releases? Water? Land?

- Which facilities are the top emitters of each of the top chemicals?

- What are the totals- for the state and each country- for a specific chemical of all inputs, outputs and inventory?

3.2 Environmental Decision Support Tool

We are awash in data from sundry studies on the environment, annual submissions from manufacturing and service industries and demographic and physiographic data. As the quantity and types of data have increased, quality issues have arisen. In general, state-level environmental data are not appropriately organized into a data warehouse. Little value is added interdepartmentally; the data is not commercially available in usable form; it is generally not all employed in co-ordinated decision-making, policy and planning; and the public is ill informed.

Numerous barriers exist. Policy and planning is typically carried out in government agencies or other institutions using state and federal data regarding the growth of the economy, productivity, and demographic projections. Environmental data is characteristically collected by program or medium with few interconnections between the two. Data "cleaning" is often minimal; coordination and the development of content provide information to select isolated decision-makers. There is little, if any, simulation computing that serves to integrate and communicate the results of dynamic data manipulations. Environmental data is used for decision making from a few selected databases. No real time, integrated model and simulation exists which blends environmental, demographic and physiographic data can be used by one and all: commercial entities in the private sector, decision makers in government, the legislature, and lay people.

Wise decision-making is dependent on the quality and availability of relevant knowledge. Therefore, NJPIES thought of building this tool which now and in the future, will provide NJ communities with the best knowledge available to help them manage their natural resources. Ultimately, information and the relationships between pieces of information form the basis of effective and equitable decision-making (Sarokin & Schulkin, 1991).

3.3   System Architecture

3.3.1 Overall System Architecture

Currently NJPIES System architecture related to NJEnviroDB consists of two parts. One is runtime environment and the other is development environment.

Following Figures and Table describe those architecture.

Table 2. JAM/Web Components

Product	H/W	OS
JAM/WEB Application Server	Sun Ultra 10	Solaris 2.6
Oracle 8.0.3	Sun Ultra 10	Solaris 2.6
JAM/WEB Builder	Dell	Windows NT 4.0

JAM/WEB consists of the following elements:

- JAM/WEB Application Server – application processing software that resides on the same machine with HTTP Server and processes JAM/WEB applications.

- JAM/Web Builder – development software which enables to design Web applications

- Apache HTTP Server 1.3.0

3.3.2   Apache HTTP Server 1.3.0

The Apache httpd server ...

- is a powerful, flexible, HTTP/1.1 compliant web server

- implements the latest protocols, including HTTP/1.1 (RFC2068)

- is highly configurable and extensible with third-party modules

- can be customised by writing 'modules' using the Apache module API

- provides full source code and comes with an unrestrictive license

- runs on most versions of Unix without modification

- is actively being developed

- encourages user feedback through new ideas, bug reports and patches

Information on the latest version of Apache can be found on the Apache web server at http://www.apache.org/. This will list the current release, any more recent beta-test release, together with details of mirror web and anonymous ftp sites.

3.4   Description of Database used for Metrics

3.4.1. TRI Database

TRI is unique in that it marks the first time that the public has direct access to detailed information about releases of toxic chemicals in their communities. TRI offers an opportunity for citizens to increase their knowledge of chemical usage in their area and to use this knowledge to affect community environmental policy and change.

3.4.1.1. TRI Information:

The TRI database includes information on.

- What chemicals were released into the local environment during the preceding year?

- How much of each chemical went into the air, water, and land in a particular year.

- How much of the chemicals were transported away from the reporting facility for disposal, treatment, recycling, or energy recovery.

- How chemical wastes were treated at the reporting facility?

- The efficiency of waste treatment.

- Pollution prevention and chemical recycling activities.

TRI provides the first comprehensive overview of toxic chemical pollution from manufacturing facilities in the United States. However, the law does not cover toxic chemicals that reach the environment from non-industrial sources, such as dry cleaners or auto service stations. Reported releases are annual estimates. The amounts reported could have been released evenly over the course of the year or, possibly, in a single large burst. Though the TRI database is a starting point for assessing possible health effects resulting from industrial chemical use, the user cannot ascertain levels of exposure or risk without combining TRI information with information from other sources. Even though the TRI reporting base has its limitations, it provides communities with a springboard from which citizens can seek further vital information about toxic chemicals in their area.

EPA is currently examining other industry sectors for potential addition to the TRI reporting scheme and intends to issue a proposal to add appropriate industries in 1996. EPA is also investigating the possibility of collecting chemical use information

under TRI. This investigation is in the preliminary stages and EPA has not yet determined what, if any, data to collect.

A Public "Report Card"

TRI is a public "report card" for the industrial community, creating a powerful motivation for waste reduction. This annual accounting of the nation's management of industrial toxic chemical wastes is a valuable source of information for concerned individuals and communities. Citizens can use TRI to evaluate local facilities through comparisons...determine how toxic chemicals are used...and, with other information, evaluate potential health risks for their community. Organizations can use TRI information as a starting point for constructive dialogue with manufacturing businesses in the area.

3.4.1.2. Pollution Prevention and TRI: Following implementation of the Pollution Prevention (PPA) of 1990, TRI reporting has become even more comprehensive. Historically, government agencies and waste generators try to resolve environmental problems using "end-of-pipe" waste management practices, that is, treating or disposing of waste after it has been created. Pollution prevention strategies focus instead on avoiding creation of wastes by redesigning products, changing processes, substituting raw materials for less toxic substances and other techniques.

With passage of the PPA, Congress adopted as national policy an environmental hierarchy that establishes pollution prevention as the first choice among waste management practices. For waste that cannot be avoided at the source, recycling is

considered the next best option. A waste generator should turn to treatment or disposal only after source reduction and recycling have been considered.

Reporting requirements for TRI changed in 1991 as a result of the PPA. Prior to 1991, facilities were required to report toxic substances released into the environment and transferred offsite for treatment or disposal. Beginning in 1991, facilities were also required to indicate amounts of chemicals that are recycled, used for energy recovery, and treated on-site. (Energy recovery means burning the chemical so that resulting heat energy contributes to subsequent manufacturing operations.) These amounts must be reported for the past year and the current year, as well as projected amounts for the next two years. Furthermore, facilities must indicate source reduction activities that have been implemented.

These changes to TRI will highlight the importance of pollution prevention and encourage reporting facilities to develop and implement strategies for reducing waste. This new information will also help the public gauge industry's commitment to improving the nation's environment. By working together, businesses and neighboring communities can build on emerging pollution prevention practices for everyone's benefit.

3.4.2. DEQ-114 Database

3.4.2.1. DEQ-114 Information: New Jersey requires facilities reporting to TRI to submit additional information above and beyond the federal requirements. Collected on New Jersey's DEQ-114, the major items of data that supplement TRI is called throughput data. They include the amounts of the chemical brought on-site, produced on-site, consumed in the manufacturing process, and shipped off-site in or as a product. These

data, together with the amounts of releases and transfers, allow an analysis of the amount of the chemical going "through" the facility. They also permit calculations of the amount of use of the chemical, the waste generated, the amount released on-site, and the amount shipped off-site in product or in waste.

The specific reporting requirements have changed over the years, as have the list of chemicals to be reported and the type of facilities covered. Not all TRI chemicals were required to be reported to New Jersey until 1991. Also, some chemicals on the New Jersey list are not covered by the federal TRI list. These chemicals are not included here. The data cover the years 1988 through 1994. Changes in reporting requirements result in differences in the data from year to year. These are described in New Jersey DEQ-114 Data. Further, Database Structure lists and describes all data items.

3.4.2.2. New Jersey Pollution Prevention Planning: Beginning with the year 1994, pollution prevention planning was added to the DEQ-114 form reporting system. Only facilities in certain industry sectors report on their pollution prevention plans and progress in meeting their planning goals. The industries covered are Paper (SIC code 26), Chemicals (SIC code 28), Rubber and Miscellaneous Plastics Products (SIC code 30), Primary Metals (SIC code 33), and Fabricated Metal Products (SIC code 34).

3.5   Calculations

The calculations mentioned below are been used to calculate the different quantities required for NPO (Non Product Output).

NPO (DEQ)=(Is + Qbos + Qp + Recycled) – (Qc + Qsh + Ie)

NPO- Non Product Output

Level of Pollution

Release Q_R

Transfers Q_T

Waste $= Q_R + Q_T$

NPO(deq) $= [I*s + Q*bos + Q*p + RECYCLED \text{ if all reused}] - [Q*c+Q*sh+I*e]$

NPO(tri) $= [8.1-8.7]$

Level of Activity

Throughput = Quse = $[I*s + Q*bos + Q*p] - [I*e]$

Throughput = use = Quse = $[Q*c + Q*sh + NPO]$

P2 Metrics

(Qw)min   = (Qw/Quse)min *Quse

Min. Waste   Efficiency   Use

Other Calculations

Total inputs

Total outputs

Mass balance = inputs – outputs;

Waste = Release + Transfers

Efficiency = Waste/Use

NPO normalized

Throughput ratio = NPO/(NPO + Qc Qsh)

Use efficiency = (Qr + Qt + Qdest)/Quse * 1000

Waste intensity = (Qr + Qt)/use *1000

3.6 Application Design

Below Figure shows exactly how the application goes from one screen to another screen based on the search criteria. All the screens mentioned in the rectangular blocks above have been covered in detail in following Sections.

3.7 Search Criteria

The search criteria for this screen are either by Chemical name.

The user has four options as shown in application design above:

i)   Search by Chemical Name

ii)   Search by Facility

iii)   Search by County

iv)   Search by State
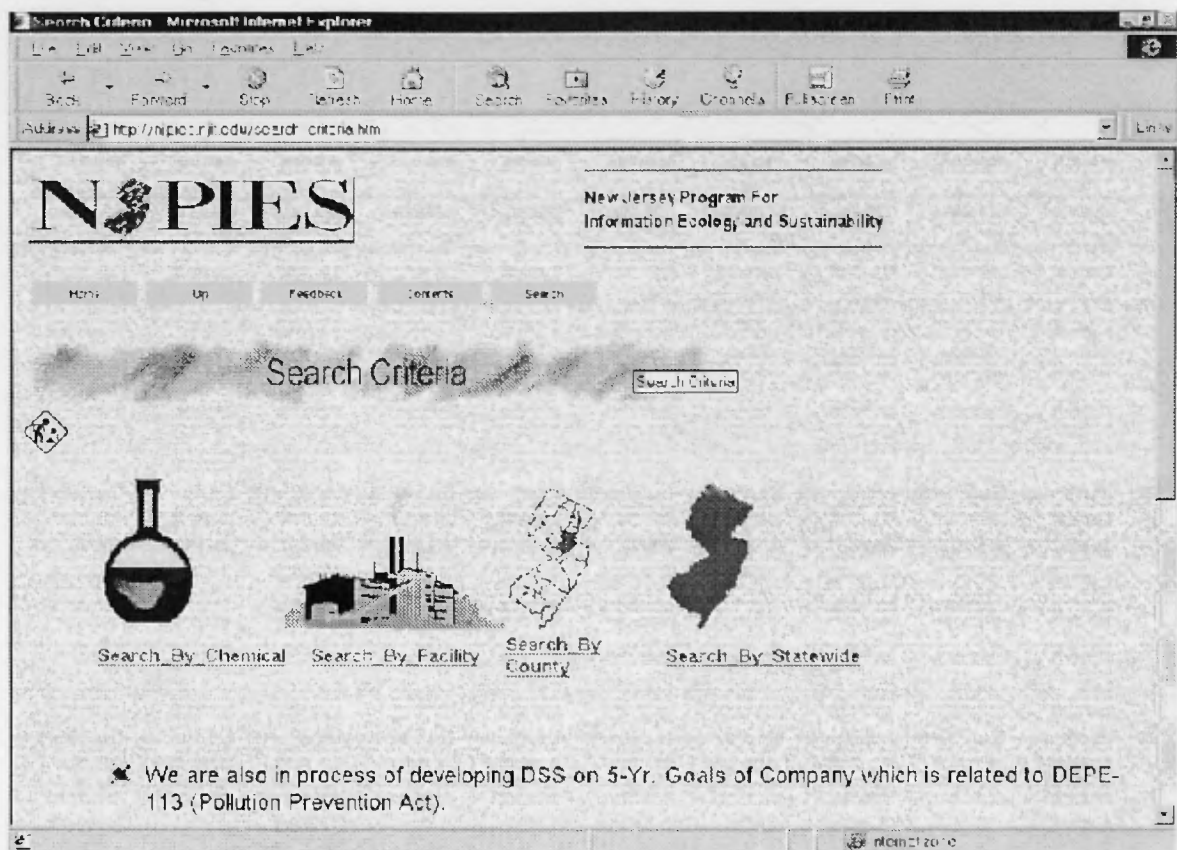
Figure 10. Search Criteria

When the user selects one of the options, he will be taken to the detail screen, which then follows a flow of screens. The flow gives the end-user the detail information of external cost and related graphical represented outputs.

Screen Flow 1: Following flow will be observed when user selects search criteria "Select chemical", he will see the following screen as shown in figure

Figure 11 Search by Chemical Name (Master Form)

It's a Master-Detail screen, contains a search criteria chemicals and a year. The idea is to mine information about the facility using selected chemical for the years 1988 through 1994. The drop down menu is populated by the list of the chemicals from the database. Based on the search criteria entered by user the screen will be populated by the list of facilities using this chemical for selected year.

Figure 12. Search by Chemical Name (Detail Form)

Detail screen shown above is part of the same screen as the master shown in Figure 11. It will be populated with names of the facilities using selected chemical for selected year. Tri_id field forms the link which takes the user to the result screen which provides user with the detailed report about that particular facility to which Tri_id belongs.

Following screens shows the detailed report for the following search criteria:

- Name of Chemical (from Master form)

- Year (from Master form)

- Tri_id (from Detail form)

Figure 13. Search by Chemical Name for NPO's (Detail Form)

Figure 14. Search by Chemical Name for NPO's (Detail Form)

3.8   Summary

The theme behind this project was the fulfillment of a need to incorporate environmental decision support model that also takes care of Waste. This refers to efforts aimed at reducing the flow of wastes from manufacturing practices and accomplishing it with a margin of profit. Somewhat synonymous terms for Waste Minimization: Waste Reduction (a 'no-no' in classic P2), Clean Technologies, Pollution Prevention, Low and non-waste technologies

In order to bring life to this concept, a means to perform an environmental decision support analysis and simulation was developed. The primary problem associated with the model was to implement the NPO's. The decision support attempts to identify, measure, and quantify the NPO's for all the chemicals. The ultimate goal of decision support model is to develop a system that will give the throughput, efficiency, and waste intensity.

CHAPTER 4

ENVIRODAEMON

4.1 Objectives

Many search engines have their distinctive features: multi-domain, single domain, meta-search engines which front-end other search engines. However, none of them make attempts to exploit the explicit underlying hypertext tags. Our new search filtering tool, HIST, however, was able to exploit the full syntactic detail of any hypertext markup language and provide hierarchical query.

The essence of HIST is to permit end-users to specify in which parts of a document a keyword should appear: in a document title, in a section header, somewhere in a paragraph, or in a table. This allows more precise control over the search process; hence, results in much better selection of relevant material. Another salient feature of HIST is the idea of "fuzzy search." The documents returned by HIST do not have to be exact match. Users have control over how similar the target document should be in the hierarchical query. Furthermore, at the speed that Web technology standard is proceeding, an information retrieval tool must be able to meet the challenges of the next generation of the document standard on the WWW. Since HIST is based on the Document Type Definition (DFD) Model, it is suitable for the new emerging document exchange standard for the WWW - Extensible Markup Language (XML).

This new tool - HIST, does not replace the existing EnviroDaemon search engine. The work presented in this project enhances the EnviroDaemon for information searching on associated environment-related topics. For a general search on the Web, where the user has little knowledge about the nature of the document (in terms of document

49

structure), we still encourage the use of EnviroDaemon. On the other hand, when the situation permits, our new tool will be extremely useful when the target document structure is partially or completely known.

4.2 Methodology

System Architecture: An Overview

Our toolkit is composed of four modules: Extractor, Parser, Query Processor, and Tree Comparator. The Extractor retrieves the actual HTML pages from the Web to be used by the Tree Comparator. The Parser translates the HTML pages into hierarchical tree structures based on the HTML DTD. The Tree Comparator contains various approximate tree and string matching programs while the Query Processor handles queries and invokes the Tree Comparator when necessary. The toolkit interacts with web browsers, the EnviroDaemon search engine, and Web servers.
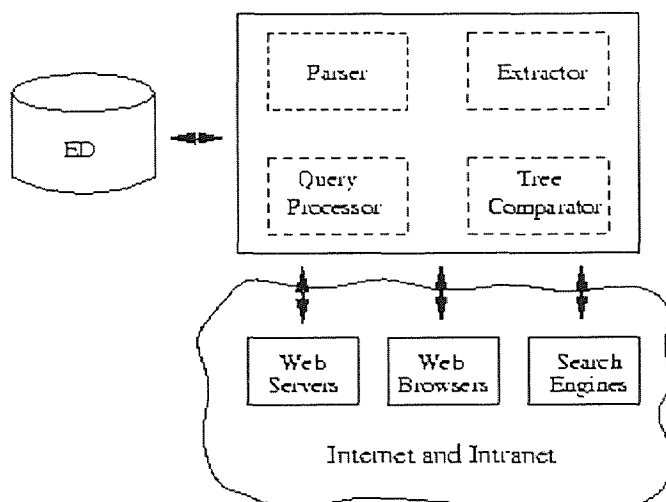

Figure 15. System Design.

4.3 Approach

Our approach is not meant to replace the existing WWW search facilities. In this sense, the work presented in this dissertation complements the tools that are available for Web searching. For "global" searches where the user has little knowledge about the information they want (in terms of possible locations, keywords etc.) and wants to query the entire WWW, we still advocate querying the entire WWW or a specific domain there in using a more domain-specific search engine. On the other hand, there are circumstances where the user has partial knowledge of the information required and could benefit from our approach. As many previous approaches have suggested, to analyze documents from the WWW, the first step is to derive a scheme that describes the HTML page. While specially designed schema provide a structural way of analyzing the hypertext document, much of the semantics associated with the document is lost after the transformation process. Our approach differs from others in that we do not design a special schema, but instead work with the DTD associated with hypertext documents.

With DTD in hand, we can parse each hypertext document using its own DTD, capitalizing on the fact that the DTD provides not only the grammatical information, but also semantic information. A document tree structure is illustrated in Figure 16. For HTML documents, we can parse it using HTML DTD 4.0. For article type SGML/XML documents, we can parse with article DTD. Figure 16. shows the parsed tree structure of a typical article type document.
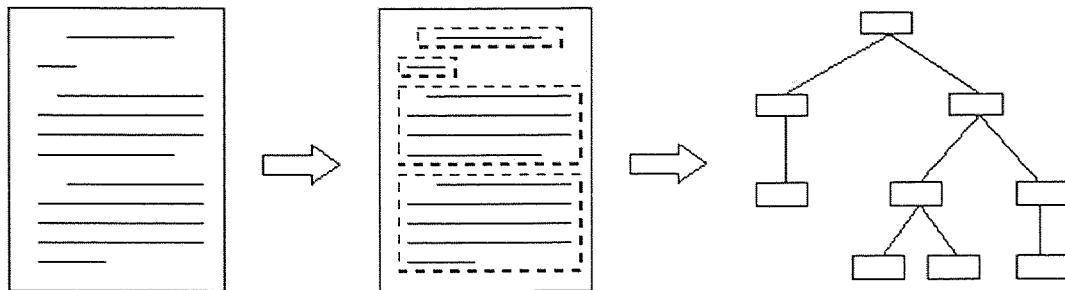
Figure 16. Parsed tree structure

4.4 Hierarchical Information Search

Consider the hierarchical query in Figure 17a. This query is to find the HTML pages containing the word "database" in an H1 header followed by a paragraph consisting of "object" followed by "relational." The "*" notation in the internal node of the query is a "variable length don't care" (VLDC) symbol, which represents an unspecified portion of a document (Zhang et al., 1994) as described below. The query may be issued when an individual intends to locate some HTML pages available on the Internet while conducting a database-related research. Here the user places an emphasis on "database" and is interested in only those HTML pages having the word in an H1 header, rather than in any other place of a document.

To process such a hierarchical query using EnviroDaemon, we take the conjunction of the keywords appearing in the query and invoke our EnviroDaemon search engine (Gchang) to find the HTML pages containing these keywords. The EnviroDaemon search engine returns a collection of candidate URLs, ranked based on their relevance to the keywords. Duplicate URLs are deleted and a document corresponding to each matching URL is then retrieved using a libwww Perl module, with

time-outs set to 20 seconds to account for a busy network or failed connections. Each retrieved HTML document is then transformed into a hierarchical tree structure based on DTD described earlier. The transformed tree structures then became candidate HTML trees that will be compared with hierarchical query trees. Figure 17b. shows an example tree for an HTML page. The tree is rooted, labeled and ordered (i.e., each node has a label and the order of siblings is important). An internal node represents an HTML tag and a leaf contains the associated text.

Our toolkit compares the query tree with each candidate HTML tree using the previously developed approximate tree matching (Zhang et al., 1994) in conjunction with regular expression matching on leaves when it is required. The URL of qualified pages is then returned. In comparing the query with an HTML tree, a VLDC can be matched, at no cost, with a path or portion of a path in the tree. The tree matching algorithm calculates the minimum edit distance between the query and the tree after implicitly computing an optimal substitution for the VLDCs in the query, allowing zero or more cuttings at nodes from the tree (Wang et al., 1994). Cutting at a node "n" means removing the subtree rooted at "n.
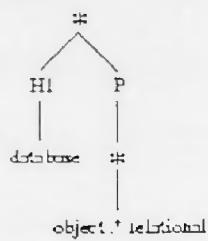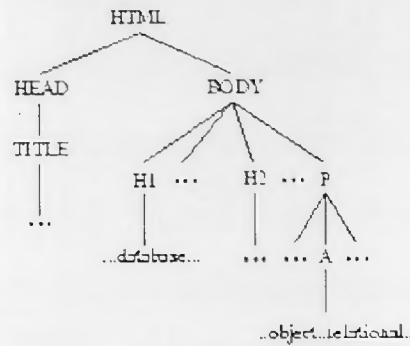
Figures 17a. Hierarchical query, 17b. Example tree for an HTML page
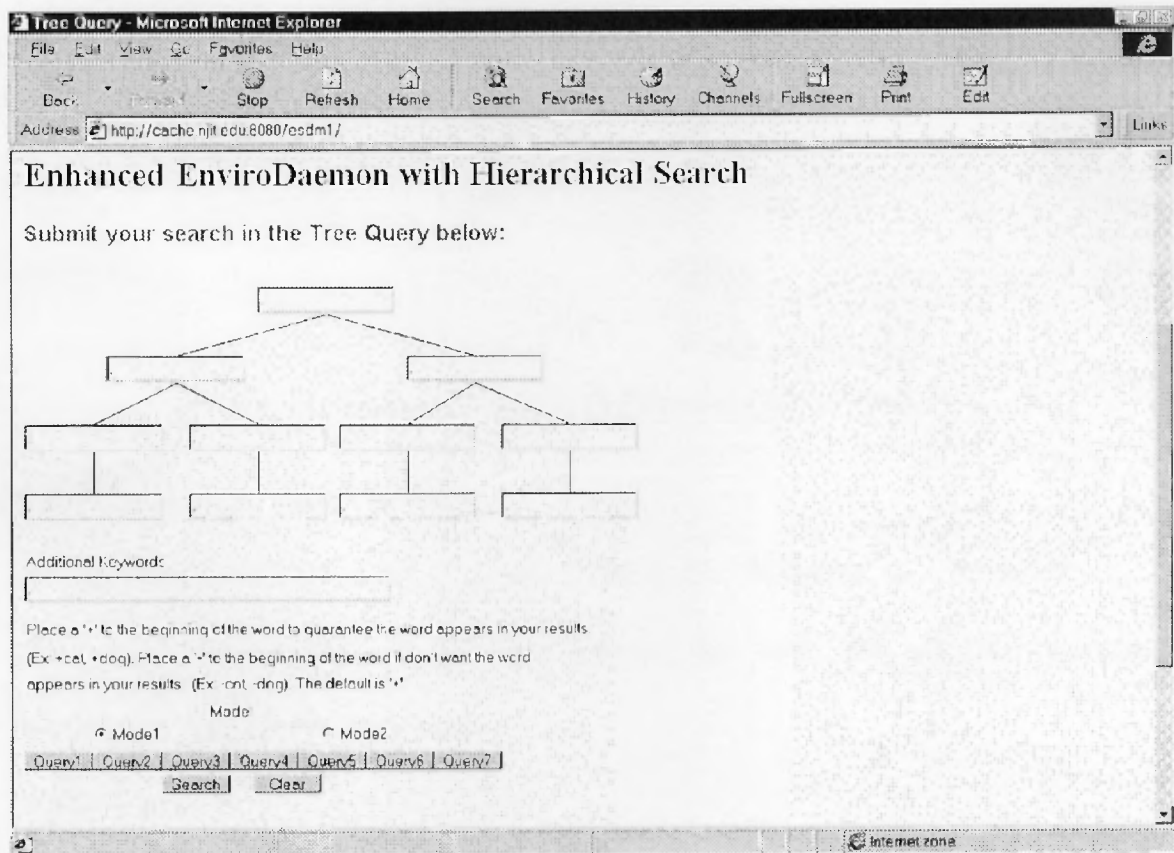
Figure 18. Front-end Interface

Given two trees T_1 and T_2, the algorithm runs in time Q ($|T_1| \times |T_2| \times \min \{\text{depth}(T_1),$

leaves $(T_1)\} \times \min \{\text{depth}(T_2), \text{leaves}(T_2)\}$. Thus, for example, in matching the query in

Figure 17a. and the HTML tree in Figure 17b, the "*" at the root in Figure 17a. Would be matched with (or instantiated into) the nodes HTML and BODY in Figure 17b., and the "*" underneath P in Figure 17a. would be matched with the node A (i.e., the Anchor tag) in Figure 17b. The nodes H1, "database" and "object.*relational" in Figure 17a. would be matched with their corresponding nodes in Figure 17b. All the other nodes in Figure 17b. are cut.

4.5 System Architecture

Table 3. Development Components

Product	H/W	OS
JDK1.1.7	Dell	Windows NT 4.0
HTTP server	Solaris	UNIX
Symantec Café	Dell	Windows NT 4.0

4.5.1 Symantec Visual Café 2.5a

Symantec's Visual Cafe Development Edition for Windows is a complete Rapid Application Development (RAD) environment for Java that provides features a sophisticated set of high-level tools.

The environment:

Visual Café Development Edition takes Java development to a new level by incorporating three powerful development tools:

- Visual Cafe

- Database Development Edition Extension

- dbANYWHERE Server

You can assemble complete Java applets and applications from a library of standard and third-party objects without writing a lot of Java code. These Java programs can be database-aware. Visual Cafe Development Edition seamlessly integrates visual and source development of Java software, letting you switch effortlessly between visual and source views. Modifications that you make in the visual view are immediately reflected in the source code. And the source code automatically updates changes to the visual view.

4.6   Experimental Results

In order to test the effectiveness of HIST, we performed an experiment on hierarchical search query [Gchang]. Retrieval effectiveness is in terms of relevant items retrieved. Recall refers to the percentage of relevant items that are retrieved and precision refers to the percentage of items retrieved in a search that are relevant[i]. The query we evaluated is Figure 3a. The query has been tested with four index servers (AltaVista, Excite, Infoseek, and Lycos). There were several thousand URLs returned by each search engine. To restrict the test set to a manageable number, we only processed the first one hundred URLs returned by each search engine and proceeded with HIST. The precision with and without HIST is summarized in Table 4. The results indicate that most URLs returned by search engines do not have the hierarchical structure specified in the query, and HIST was able to eliminate all of them.

Table 4. Precision of HIST.

Search Engine	Without HIST	With HIST
AltaVista	6/100	6/6
Excite	9/100	9/9
Infoseek	3/100	3/3
Lycos	4/100	4/4

The approximate search results are summarized in Table 5. HTML document distance with respect to query structure, ranging from 0 to 4 was found; where distance 0 indicates an exact match. The column "ϕ" indicates the number of URLs that we were unable to retrieve due to a parsing error, and other network problems not uncommon on the Web.

Table 5. Approximate retrieval with distance

Search Engine	Dist. 0	Dist. 1	Dist. 2	Dist. 3	Dist. 4	ϕ
AltaVista	6%	45%	12%	13%	0%	24%
Excite	9%	39%	33%	9%	0%	10%
Infoseek	3%	43%	34%	8%	0%	12%
Lycos	4%	31%	37%	9%	1%	18%

Competitors

George:

ⁱ In order to evaluate any information retrieval system, performance is presented by calculating the recall (100a/a + c) and precision (100a/a + b) ratios.

	Relevant	Non-Relevant	
Retrieved	a	b	a + b
Not Retrieved	c	d	c + d
	a + c	b + d	a + b + c + d = N (Total Collection)

4.7 Summary

Information retrieval has evolved from searches of references, to abstracts, to entire documents. With the advent of the World Wide Web, scientists have their first harbinger that their grand vision will one day be realized: immediate, distributed access to the entire science and technology literature. Search on the Web involves search engines, latter day descendants of the library card catalog that promise to deal with full-text and a plethora of other files involving audio, video, and multi-media. With the indexable Web at approximately 320 million pages and growing, difficulties with locating relevant information have become apparent. The most prevalent form of location of information on a topic of interest currently relies on syntax-based search methods: syntactical patterns like keywords or strings of characters are presented to a search engine, and it returns all the matches in the available documents. Whereas this method is satisfactory for many purposes and relatively easy to implement, it has some inherent limitations that make it unsuitable for many tasks. Instead of looking for syntactical patterns, the user often is interested in the meaning associated with the keywords, or the location of a particular

word in a title or header. This paper describes some approaches that locate information according to syntactic criteria, augmented by pragmatic aspects like the utilization of information in a certain context. The main emphasis of this paper lies in the treatment of structured knowledge, where essential aspects about the topic of interest are encoded not only by the individual items, but by their relationships among each other. Examples for such structured knowledge are hypertext documents, diagrams, logical formulae, chemical formulae, or objects in images. Benefits of this approach are enhanced precision and approximate search in an already focused, context specific search engine for the environment.

APPENDIX-A   CALCULATIONS OF THE NPO

Non Product Output: NPO

Release Q_R

Transfers Q_T

Waste $Q_R + Q_T$

NPO (DEQ)=$(I_S + Qbos + Qp + Recycled) - (Qc + Qsh + Ie)$

NPO(tri)   =[8.1+8.2+8.3+8.4+8.5+8.6+8.7]

Throughput = Q_{use} = [Is + Q_{bos} + Qp] – [Ie]

Throughput = use = Quse = [Qc + Q_{sh} + NPO]

Mass balance = inputs – outputs;

Waste = Release + Transfers

Efficiency = Waste/Use

NPO normalized

Throughput ratio = NPO/(NPO + Qc + Qsh)

Use efficiency = (Qr + Qt + Qdest)/Quse * 1000

Waste intensity = (Qr + Qt)/Quse *1000

NPO Change

NPO Change = $NPO_y - NPO_{y-1}$

NPO Change(%) = $(NPO_y - NPO_{y-1})/NPO_{y-1} * 100$

REFERENCES

[1]   Ackerman, F., Why do we Recycle?, Island Press, ©1997, ISBN 1-55963-505-3

[2]   Anstine, J., A. Revankar, and M.J. Healey (in progress), WORKING PAPER, The Sustainable Green Manufacturing Project at NJIT: Measuring Social Costs and Perceived Risks Associated with Hazardous Material Emissions from Weapons Manufacturing

[3]   de Caluwe, N. (1997). Ecotools manual - A comprehensive review of Design for Environment tools., Design for Environment Research Group, Manchester Metropolitan University

[4]   Environmental Security Technology Certification Program (ESTCP) Validation Tasks, Environmental Cost Analysis Methodology (ECAM) Handbook, January 7, 1998, Contract No. DAAA21-93-C-0046, Task No. N.098, CDRLK No. A013, Prepared by National Defense Center for Environmental Excellence (NDCEE), Operated by Concurrent Technologies Corporation

[5]   Greene, W. (1993). Econometric Analysis. Second Edition. Macmillan Publishing Co. New York, New York.

[6]   Wenzel, H., Hauschild, M., and Wenzel, H. (1998). Environmental Assessment of Products, Volume 2. Scientific background, Chapman & Hall, ©1998, ISBN 0-412-80810-2

[7]   Portney, Paul. 1993. Public Policies for Environmental Protection. Resources for the Future. Washington D.C.

[8]   Waste Minimization Prioritization Tool (WMPT), Beta Test Version 1.0, User's Guide and SystemDocumentation, Draft, USEPA, EPA530-R-97-019, June 1997

[9]   Lynnwood Brown, Oracle Database Administration on Unix systems, Prentice Hall, New Jersey, 1997.

[10]   Peter G.W. Keen, "Information Systems and Organizational Change," Communications of the ACM, January 198 1, Volume 24.

[11]   Keen, P.G.W. and Scott Morton, M.S. Decision Support Systems: An Organizational Perspective. Addision-Wesley, Reading, Mass., 1978.

[12]   K. Zhang, D. Shasha, and J.T.L. Wang, Approximate tree matching in the presence of variable length don't cares. Journal of Algorithms, 16(1): 33-66, Jan 1994.

[13]   JAM/WEB (2.0) Developers' Guide, JYACC, Inc. New York, NY, 1995

[14]   Oracle Corporation : http://www.oracle.com, November 1998

[15]   US Environmental Protection Agency : http://www.epa.gov, August 1998

[17]   Apache HTTP Server Project : http://www.apache.org, August 1998

[18]   Microsoft Corporation : http://www.microsoft.com, August 1998

[19]   O'Rielly's Website: http://website.oreilly.com., July 1998