**A Parallel Computing Architecture for Information Processing: Visualizing, Indexing, and Mining**

Xiannong Meng, Wendy A. Lawrence-Fowler, Richard H. Fowler, Zhixiang Chen, Richard K. Fox  
Department of Computer Science  
The University of Texas - Pan American  
Edinburg, TX 78539-2999  
e-mail: wfowler@panam.edu

**Abstract:** In this paper we define DaVIME (Data Visualization, Indexing and Mining Engine), a software architecture that performs data visualization, indexing and mining in an integrated environment. DaVIME presents a unified view of information to the users. When a user issues an information service request, DaVIME calls the appropriate software module to provide the requested service. One or more components may be called into action depending on the request. DaVIME is an open, extensible architecture that allows researchers and developers to add software modules incrementally.

DaVIME (Data Visualization, Indexing and Mining Engine), provides an information services architecture including data visualization, indexing and mining. The core of this research is a suite of tools to provide internet-based information services. Our design goal is to have an open, extensible architecture that provides a comprehensive and unified view of an information service for users and allows incremental development for researchers. The system is divided into four major components, the User Interface Coordinator (UIC), the Data Resource Coordinator (DRC), the Computing Resource Coordinator (CRC), and various Extensible SoftBots (ESB). An overview is given in [Fig.1].

![FIGURE 1. DaVIME system architecture.](image)

Users request information services through the UIC. This module accepts, refines and then sends service requests to appropriate ESB softbots for processing. The ESB is an expandable collection of information processing modules. Current ESB modules include:
♦ Document Explorer - analyzes textual information in large document collections, calculates "distances" (i.e. similarity) among documents, and presents the results visually such that related documents are linked closely [Fowler, Fowler & Williams 1997].

♦ Dynamic User-created Searchable Index Engine (DUSIE) - extends hierarchical indexing schemes to allow users to build personalized context-based searchable indices [Lawrence-Fowler, Williams, Fowler & Meng 1998]. Users can organize Web indices into a hierarchy of their own choice as well as add annotation to the indices [Fowler, Fowler, Williams, Palacios & Palacios 1997].

♦ ParaCrawler - uses a novice ranking and indexing algorithm to collect, cull, and present accurate search information from the Web [Fox 1997; Fox, Ochoa & Paredes 1996]. The search uses personalized information from users and retrieves search results in parallel [Meng & Chen submitted].

♦ Gis2web - allows users to access GIS data from the Web by interacting with regular GIS software such as ArcView and ArcInfo [Meng, Fowler & Rieken 1997].

The DRC handles data exchange among ESB components. The final component, CRC, provides access to any available computing resources (CPU cycles & storage) from intranets and the Internet. The CRC will register resource requests from internal DaVIME components, work with brokers, find available hosts on the network that are willing to share resources, and request services from those hosts.

References


Acknowledgements

This project has been supported in part by a NASA grant NAG 9-842 a U.S. Department of Education grant P120A50059, and a NSF Alliance for Minority Participation Award.