AN OVERVIEW OF THE PROTOTYPE INFORMATION DISSEMINATION SYSTEM (PRIDES)

Virginia Cevasco

Logicon, Operating Systems Division

2100 Washington Blvd.

Arlington, VA 22204

gcevasco@logicon.com

(703) 486-3500

540

INTRODUCTION

The Prototype Information Dissemination System (PRIDES) is a TIPSTER technology insertion project sponsored by the Office of Research and Development (ORD). PRIDES applies a portion of the TIPSTER detection architecture and several TIPSTER components to the problem of timely dissemination of Foreign Broadcast Information Service (FBIS) articles. When PRIDES begins operation in July 1996, it will provide one of the first production tests of the TIPSTER architecture.

PRIDES FUNCTIONS

FBIS collects, translates, and disseminates selected foreign media content, including newspaper and magazine articles and television and radio broadcasts. In addition, FBIS analysts publish analyses of trends and patterns across articles. This information is available in hard copy on a daily basis, or on CD ROM on a quarterly basis. The hard copy format is timely, but difficult to work with. The CD ROMs are easy to search and process, but not timely. PRIDES seeks to provide both timely dissemination, customized to a user's particular interests, and comprehensive retrospective search support.

A user describes his dissemination need in an interest profile. Then, as new articles are received daily, each is compared to the interest profile. If the document scores above the threshold for the profile, it is added to the user's mail folder. Then the user can browse all the articles relevant to their profile. Each user may have any number of profiles covering different topics.

PRIDES provides a robust, easy to user retrospective search capability against a corpus of FBIS articles accumulated since May, 1995. The user may write queries in natural language or Boolean syntax. The user may also query by example, selecting articles which contain the sort of information they are interested in and allowing the system to build a query to locate similar articles. The user may also search within certain fields identified by the FBIS users as particularly content rich. These fields include date fields, which support date ranging. Search results are stored in hit folders.

The user may create private collections, called save folders, storing articles from other folders. Save folders allow the user to collect articles from both dissemination results and query results for an open ended amount of time. Save Folders can also be downloaded to the user's local disk for additional processing.

The user can list their mail, hit, and save folders, and then open any folder to see a list of the folder contents. This list includes a one line summary for each article, containing the article headline, relevance score, and date. The user has a variety of options for sorting and segmenting the folder contents list. Each headline is linked to its article. The article is displayed in its entirety. Again, the user can customize the display format to suit their personal work style.

PRIDES provides a World Wide Web interface, suitable for deployment on the Internet or an intelligence community intranet. This provides PRIDES to the maximum possible user community, while simultaneously eliminating the need for PRIDES-specific client side software. Any forms-compatible web browser can be used to access PRIDES.

PRIDES SYSTEM ARCHITECTURE

To fulfill the requirements for the PRIDES pilot system and simultaneously lay the foundation for the future system, the PRIDES architecture is comprised of three layers. Each layer has a unique set

of responsibilities, and communicates only with its adjacent layer(s) via a well-defined API. This three-layered architecture offers plug-and-play design. Software can be inserted into any layer with minimal impact on the other layers. This architecture promotes PRIDES evolution, because older user interfaces can stay in operation while new user interfaces are gradually tested and fielded to replace them, and new versions, or even different types, of TIPSTER-compliant search engines and routing engines can be tested without changes to the user interface.

The PRIDES User Interface Layer (PUI) is responsible for creating and managing the screen displays that comprise the PRIDES user interface. The User Interface is implemented using a World Wide Web Browser, a Web Server, and the Hypertext Markup Language (HTML) to provide custom screens. In this way PRIDES supports the most modern Internet access technology. PUI calls the PRIDES Application Layer (PA) to service requests and collects and formats that data for easy use. PA is responsible for performing any PRIDES-specific activity that is not provided by our TIPSTER components. PA is also responsible for maintaining and validating access privileges and collecting, storing and analyzing MIS data. PA provides an external message interface to the incoming FBIS documents. PA interfaces with the TIPSTER Data Access (TDA) layer to store, index, search and retrieve PRIDES data via API calls.

The TDA consists of a set of TIPSTER-compliant search engines and database management software. The Document Manager software stores and indexes documents coming into PRIDES. The PRIDES Document Manager is fully TIPSTER compliant and available for use in other TIPSTER systems. The routing engine and Document Manager process user profiles and route these incoming documents to mail folders. Similarly, the search engine and the Document Manager process user queries and build hit folders. TDA satisfies requests for retrieval of a PRIDES document, given either an internal or external document ID.

To fulfill the PRIDES requirements, Logicon has selected technology products that adhere to the TIPSTER architecture, that are consistent with an open design, and that can be scaled up to accommodate larger volumes of input and more users. Where custom software was necessary for the PRIDES system, it was designed within the layered architecture approach described above, in order to guarantee maximum flexibility, scalability and extensibility.

For the PRIDES detection engines, PRIDES uses the ACSIOM products INQUERY and InRoute. INQUERY is an information retrieval system based upon a Bayesian inference network model of information retrieval. Inference networks are ideally suited for the uncertainties encountered when matching a person's statement of an information need with a document expressed in natural language. In addition to using inference networks, INQUERY incorporates several different methods of combining evidence, enabling a rich query language in which to express information needs. In accordance with PRIDES requirements, the INQUERY product was modified to:
- Optimize retrieval algorithms.
- Optimize concurrency control to allow frequent updates of the document index.
- Upgrade the API for robustness in an integration environment.
- Develop natural-language to query-language transformations.
- Extend the API to support "query by example" and "cancel search."
- Collect resource consumption data.

InRoute is an inference network system tailored for document filtering. Both InRoute and INQUERY return identical scores for a given document/query pair. While INQUERY is optimized for quickly searching one or more multi-gigabyte document collections, InRoute is optimized for quickly comparing a steady stream of documents to a large number of profiles. In accordance with PRIDES requirements, the InRoute product was modified to:
- Convert SGML in FBIS articles into TIPSTER annotations.
- Develop algorithms for incremental relevance feedback to replace the existing batch-oriented feedback.
- Upgrade the API for robustness in an integration environment.
- Collect resource consumption data.

Standard COTS Web Server products provided the capabilities needed to define the PRIDES user interface. A Web Server package, augmented by a set of PRIDES-specific Common Gateway Interfaces (CGIs), communicates with the client via Hypertext Transport Protocol (HTTP). A CGI is a standalone script or program invoked by a Web Server to provide services beyond those included in its suite. In the case of PRIDES, CGIs provide access to all PRIDES services and data, subject to user access privileges. PRIDES end users use a Web Browser to communicate with the Web Server. The Web Browser software may be Mosaic,

Netscape Navigator, or any other browser which can process HTML forms. No special PRIDES software is needed in the end user's workstation. These web browsers also provide a user-friendly interface to the other protocols of the Internet, such as File Transfer Protocol (FTP) and Network News Transfer Protocol (NNTP), and allow printing of text and graphics on the user's local printer.

PRIDES DEVELOPMENT AND PILOT OPERATIONS

PRIDES was designed and developed by Logicon and ACSIOM from June 1995 through April 1996. After installation at the customer's site and an acceptance test period, PRIDES will begin serving production users in July 1996.

PRIDES is a pilot effort which will serve users operationally for six months, between July 1996 and January 1997. Volunteer pilot users will be selected from among the FBIS analysts and consumers. During pilot operations, an extensive evaluation program will gather quantitative and qualitative data about how users work with PRIDES. The analysis of this data will attempt to evaluate the user acceptance of the new features in PRIDES, such as the Internet delivery mechanism, relevance ranking, and automatic query refinement. The results of the evaluation effort will provide input to the requirements of the final FBIS softcopy dissemination system.