GESCAN -- A study of its capabilities

Alexander M. MacEachern
Department of Mathematics and Computing Sciences
St. Francis Xavier University
Antigonish, Nova Scotia B2G 1C

The GESCAN system is an information management system which allows users to create and manipulate textual databases. Specifically, users can store, retrieve, update, copy and delete data in a textual database. The GESCAN system is designed to be used on a VAX/VMS minicomputer in conjunction with a special purpose processor called a Text Array Processor (TAP). The TAP is utilized by the GESCAN system to perform high-speed text searches in the retrieval of records from a database. The TAP uses four customized parallel comparators on LSI chips. Each chip contains a set of cells which can be cascaded or stacked vertically to handle many search terms or queries. GESCAN supports interactive processing on a timeshared basis allowing multiple users to search a database at the same time. Furthermore, the system runs user searches as a background process which allows the user to run multiple searches on multiple databases while simultaneously reviewing previously retrieved records. The majority of the GESCAN system commands are executed from the VAX/VMS $ prompt which allows the user to work on an entirely different application while the GESCAN system is performing a search on a large database.

The major structures of the GESCAN system are databases, queries, and hitfiles. A database is a collection of "documents". A document may consist of one or more fields, each of varying length. Each field is of type "text". The database consists of a single-entity set in which all documents have the same format. A query is a structure that the GESCAN system uses to selectively retrieve documents from the database. The query is created by the user and specifies the conditions that must be satisfied by a document before it will be extracted from the database. The conditions specify the strings of characters that must occur in given fields. The conditions allow exact matches, wildcard specifications, substrings and phrases to be specified for desired fields within the documents. These conditions form a group of terms and four such term-groups are permitted. The query may also specify logical conditions between the groups.All selected documents are placed in a "Hit file".

The user can then browse the "Hit file" using the HBROWSE subsystem. In this subsystem, it is possible to view the selected documents, perform modifications, print documents and/or create subsets of these documents.

All of the operations mentioned in the preceding paragraphs can be performed from a terminal. A user can generate the database, create and execute queries and use the browse system. To be efficient, the user should have a good working knowledge of each of the subsystems. Many potential users may not know, and should not have to learn, al of the details of each part of the systems. Through a set of provided GESCAN library routines it is possible to build a customized user interface to GESCAN.

In this study, the capabilities and some of the weaknesses of GESCAN will be discussed through the use of a sample session, and a simple user interface will be illustrated.

GESCAN does provide a means of searching large amounts of "raw text". Also, since the TAP is an independent processor the system can perform searches without overwhelming the central processor of the host system. It is relatively easy to learn to use the system, and the query conditions possible are reasonably powerfully.