Construction of Tests of Information-Retrieval Systems

The Internet is utilized in constructing test collections.

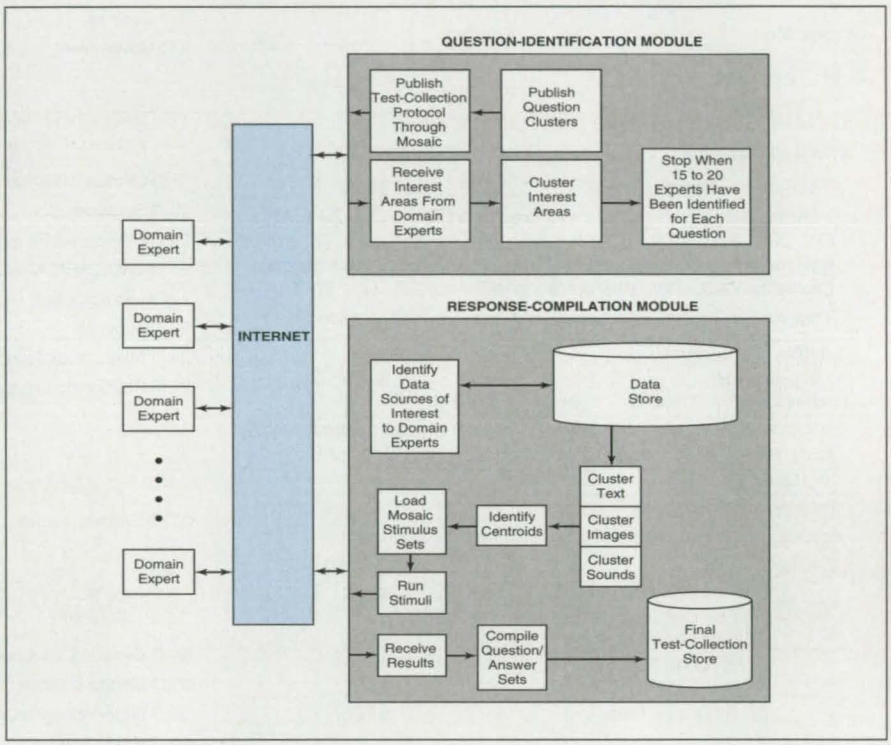
Lyndon B. Johnson Space Center, Houston, Texas

A method of measuring the performances of computerized multimedia (text, images, and sound) information-retrieval systems involves utilization of the Internet in constructing test collections with the help of experts in various domains of specialized knowledge. "Test collection" denotes a collection of sets of questions and answers, devised by the domain experts, that are representative of (1) questions asked of information-retrieval systems within their domains and (2) the correct and/or relevant (in their expert opinions) answers that information-retrieval systems should give to those questions.

In measuring the performance of an information-retrieval system by this method, one seeks to determine to what degrees the system approaches optimality according to the following three criteria:

- Retrieval of the minimum number of items considered by the experts to be irrelevant (this is the traditional precision criterion);
- Retrieval of the maximum number of items considered by the experts to be relevant to greater or lesser degrees (this is the traditional recall criterion); and
- Correlation between the ordering of the retrieved items and the ordering preferred by the experts.

An essential component of the method is a theory for constructing a test collection that is truly representative in the sense that it is robust in the face of variability of judgments among experts. The theory is based partly on techniques of psychometric scaling and stochastic transitivity as applied to the judgments of the experts.



The **General Procedures for Construction of Test Collections** are implemented in modules. The test collections are constructed from the judgments of experts with whom the test-collection constructor communicates through the Internet.

Heretofore, the acquisition of test collections from the experts has been impeded by the geographic dispersion of the experts. The Internet now makes it possible to largely overcome this impediment. In particular, its Mosaic presentation feature accommodates pairwise multimedia presentations of question and answer data to the experts and enables the experts to send their judgments back by electronic mail (see figure).

This work was done by Mark E. Rorvig of **Johnson Space Center**. *For further information,* ***write in 96*** *on the TSP Request Card. MSC-22542*

Spreadsheet Fourier Analysis of Repetitive Data Signals

The technique can be used to determine spectral contents of complex data signals.

John F. Kennedy Space Center, Florida

A technique for frequency- and time-domain analysis of repetitive data signals involves the insertion of the signal representation into a commercial spreadsheet computer program, computing the Fourier series of the signal. By use of a macro in the spreadsheet, any number of harmonics of the fundamental frequency of the stream can be determined (see Figure 1). The technique can be used to determine the spectral contents of complex data signals, such as modulated unbalanced synchronous quaternary or binary phase-shift keying (see Figure 2),

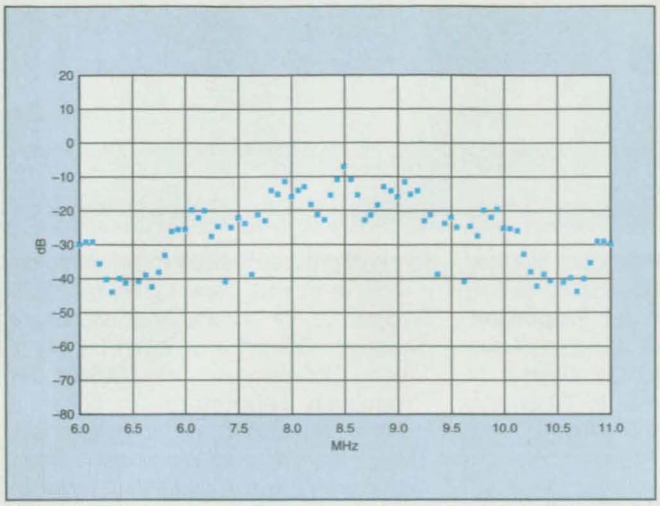


Figure 1. This **Computed Frequency Spectrum** of binary phase shift keyed 32-bit pattern, hexidecimal 59CDB300, modulated at 8.5 MHz and data rate of 2Mbps was derived via a macro in a commercial spreadsheet computer program.

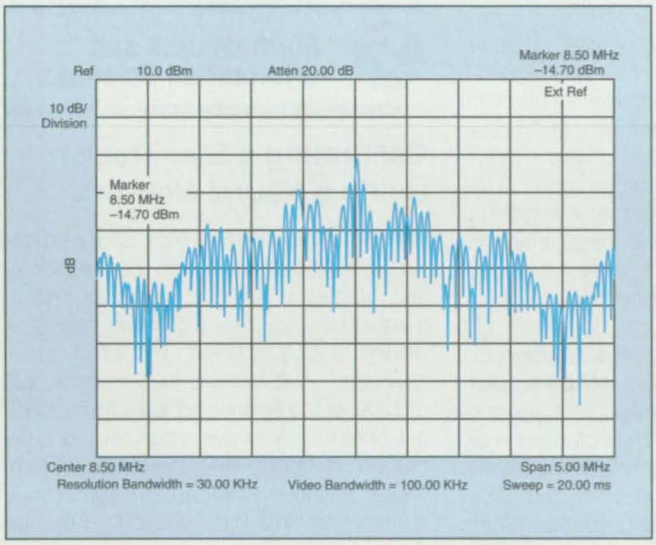


Figure 2. The **Actual Spectral Plot** of the data stream demonstrated good correlation between the computed approach in Figure 1 and the actual shown here.

baseband biphase or non-return-to-zero, level. A principal ad-vantage of this technique over techniques based on fast Fourier transforms (FFTs) is that the Fourier series represents the exact spectral contents as precisely as desired, whereas FFTs are only approximate. One use of this technique is to determine exact filters for a given code and to use the magnitude of those filters in an error-correction algorithm. By extension of this concept, the technique can be used to design coding algorithms coupled with filtering aimed at achieving transmission at higher data rates, less bandwidth, and/or lower error rates in an operating environment. The technique can also be used to aid instruction in frequency analysis and/or the scientific use of spreadsheets.

This work was done by Lewis Lineberger of **Kennedy Space Center.** *For further information,* ***write in 12*** *on the TSP Request Card.*

Inquiries concerning rights for the commercial use of this invention should be addressed to the Patent Counsel, Kennedy Space Center; (407) 867-2544. Refer to KSC-11812.