



INFORMATION

A detailed black and white illustration of a mechanical component, possibly a filter or a part of a machine. It features a large, curved metal housing with a fine mesh screen inside. Several circular ports or holes are visible on the housing. In the upper right corner, there is a circular pressure gauge with a needle pointing to approximately 120. The gauge has two scales: an outer scale with markings from 25 to 165 and an inner scale with markings from 20 to 85. The word "FILTERING" is prominently displayed in large, bold, sans-serif capital letters across the middle of the image, partially overlapping the mesh and the gauge.

FILTERING

Guest Editors

*Shashikanth Bhat
and David T. Long*

The promise of the information age entails making information available to people any time, any place and in any form. Realizing such a promise depends on innovations in areas that impact the creation of information services and their communication infrastructures. However, this realization can easily become a mixed blessing without methods to filter and control the potentially unlimited flux of information from sources to their receiving end-users.

Realistic deployment scenarios for information-filtering technologies have many differentiating characteristics. For example, the type of information (e.g., TV and radio programming, live news services, electronic mail), and the information transport architecture (e.g., broadcast, narrow-cast, point-to-point) are two of the characteristics which strongly affect the appropriate choice of filtering technology.

The success of many new information services that provide end users with access to diverse information sources is crucially dependent on the availability of effective filtering technology. This technology can be used by both the information sources and their end users, to route and control the delivery of information. For example, in the domain of entertainment, the individual information sources may use filters to target material to preferred end user groups, and individual end users may use filters to select the material of their choice out of all available sources.

The demand for information filtering technology is not new, however, and is not limited to new information services. Over a decade ago, Peter Denning's ACM President's Letter on "Electronic Junk" (*Commun. ACM*, March 1982, 163-165) focused on the implications of automatic document preparation systems and electronic mail, and on the quantity of information being received by end users. He pointed out that "The visibility of personal computers, individual workstations, and local area networks has focused most of the attention on *generating* information—the process of producing documents and disseminating them. It is now time to focus more attention on *receiving* information—the process of controlling and filtering information that reaches the persons who must use it."

In November 1991, Bellcore hosted a Workshop on High Performance Information Filtering in Morristown, N.J. Organized and sponsored by Bellcore in cooperation with ACM SIGOIS, the workshop was the first of its kind. The event brought together over one hundred researchers from major university and industrial research labs who share a strong interest in the creation of large-scale personalized information delivery systems.

The workshop covered all aspects of this emerging area including its relation to the established field of information retrieval (IR), a variety of methods for filtering, architectural concerns of high-speed filtering systems, and a variety of existing prototype applications, as well as requirements for future applications.

This special issue features five articles that represent the scope and content of that workshop. Each article represents a different aspect of the field and together they form a realistic view of the workshop. In addition, we present four sidebars depicting individual snapshots of an emerging filtering approach or applications.

Belkin and Croft ask and answer the question, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" The authors determine that information filtering is a well-defined process. By examining its foun-

and comparing it to the foundations of the IR enterprise, the authors find there is very little difference between filtering and retrieval at an abstract level. They conclude that the two enterprises have the same goal; namely they are both concerned with getting information to people who need it. However, the authors emphasize that IR research has ignored some aspects of the general problem which both IR and information filtering address, and that these aspects are precisely those which especially relevant to the specific contexts of filtering.

Loeb picks up where Belkin and Croft's article left off—examining some of the ways information-filtering models may extend IR models. More specifically, Loeb's article centers on "Architecting Personalized Delivery of Multimedia Information," providing both a mapping of the filtering application and usage scenarios, and a specific example of a novel filtering model and its implementation. The author provides an analysis of successful filtering applications in the context of the personalized multimedia music system.

In "Personalized Information Delivery: An Analysis of Information Filtering Methods," Foltz and Dumais present results of an experiment aimed at determining the effectiveness of four information-filtering methods in the domain of technical reports. The experiment was conducted over a six-month period with 34 users and over 150 new reports published each month. Overall, the authors conclude that filtering methods show promise for presenting personalized information.

In "Using Collaborative Filtering to Weave an Information Tapestry," Goldberg, Nichols, Oki, and Terry describe an experimental system that manages an in-coming stream of electronic documents, including email, newswire stories and NetNews articles. The system implements a novel mechanism for collaborative filtering in which users annotate documents before the documents are filtered. Because annotations are not available at the time a new document arrives, the system supports continuous queries that examine the entire database of documents and take into account newly introduced annotations during the filtering process.

In "The Datacycle Architecture" Bowen et al., present the operating principles of a fully implemented platform that supports very high-performance information filtering. Key to realizing the architecture is the on-the-fly data filtering operation, which supports both expanded information retrieval functionality and conflict resolution for management of changes to database contents. This article complements the others in this section by describing an application-independent platform that embeds enough of the application semantics to adequately meet high-performance requirements.

We believe these five articles together with the sidebars capture the excitement and quality of the work as reflected in the workshop. **C**