

## UNIT – V

# TOPIC: Text Search Algorithms

Text Search Algorithms Three classical text retrieval techniques have been defined for organizing items in a textual database, for rapidly identifying the relevant items and

For eliminating items that do not satisfy the search.

The techniques are

- 1) Full text scanning (streaming)
- 2) Word inversion
- 3) Multiattribute retrieval

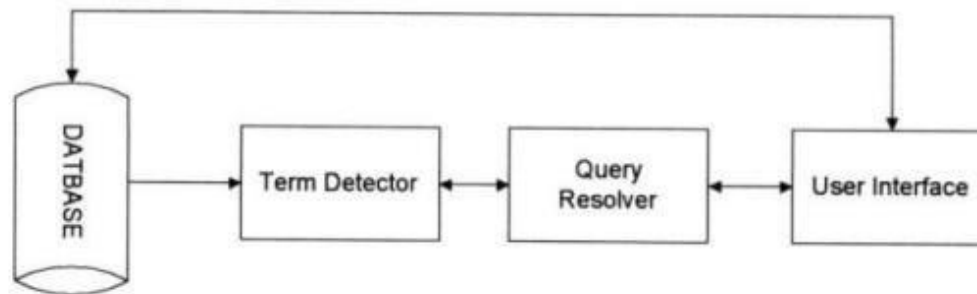


Figure 9.1 Text Streaming Architecture

## Software Text Search Algorithms:

In software streaming techniques, the item to be searched is read into memory and then the algorithm is applied. There are four major algorithms associated with software text search:

- 1) The brute force approach
- 2) Knuth-Morris-Pratt
- 3) Boyer-Moore, Shift-OR algorithm

## Hardware Text Search Systems

Software text search is applicable to many circumstances but has encountered restrictions on the ability to handle many search terms simultaneously against the same text and limits due to I/O speeds.

One approach that off loaded the resource intensive searching from the main processors was to have a specialized hardware machine to perform the searches and pass the results to the main computer which supported the user interface and retrieval of hits. Since the searcher is hardware based, scalability is achieved by increasing the number of hardware search devices

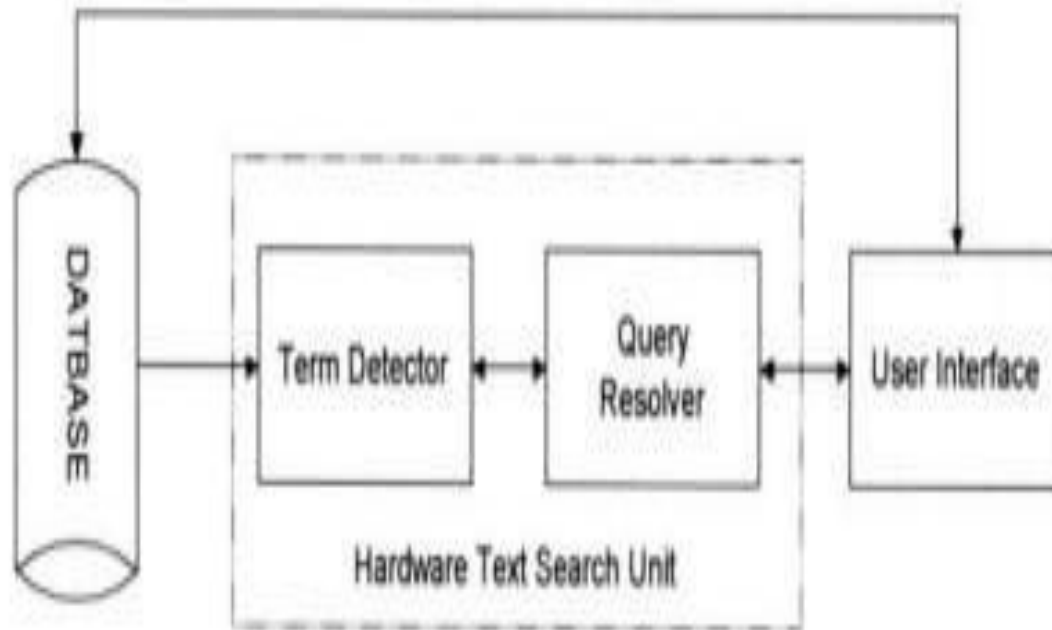


Figure 9.9 Hardware Text Search Unit

## Spoken Language Audio Retrieval:

Just as a user may wish to search the archives of a large text collection, the ability to search the content of audio

of applications. An assortment of techniques have been developed to support the automated recognition of speech (Waibel and Lee 1990).

These have applicability for a range of application areas such as speaker verification, transcription, and command and control.

# BBN'S Rough and Ready:

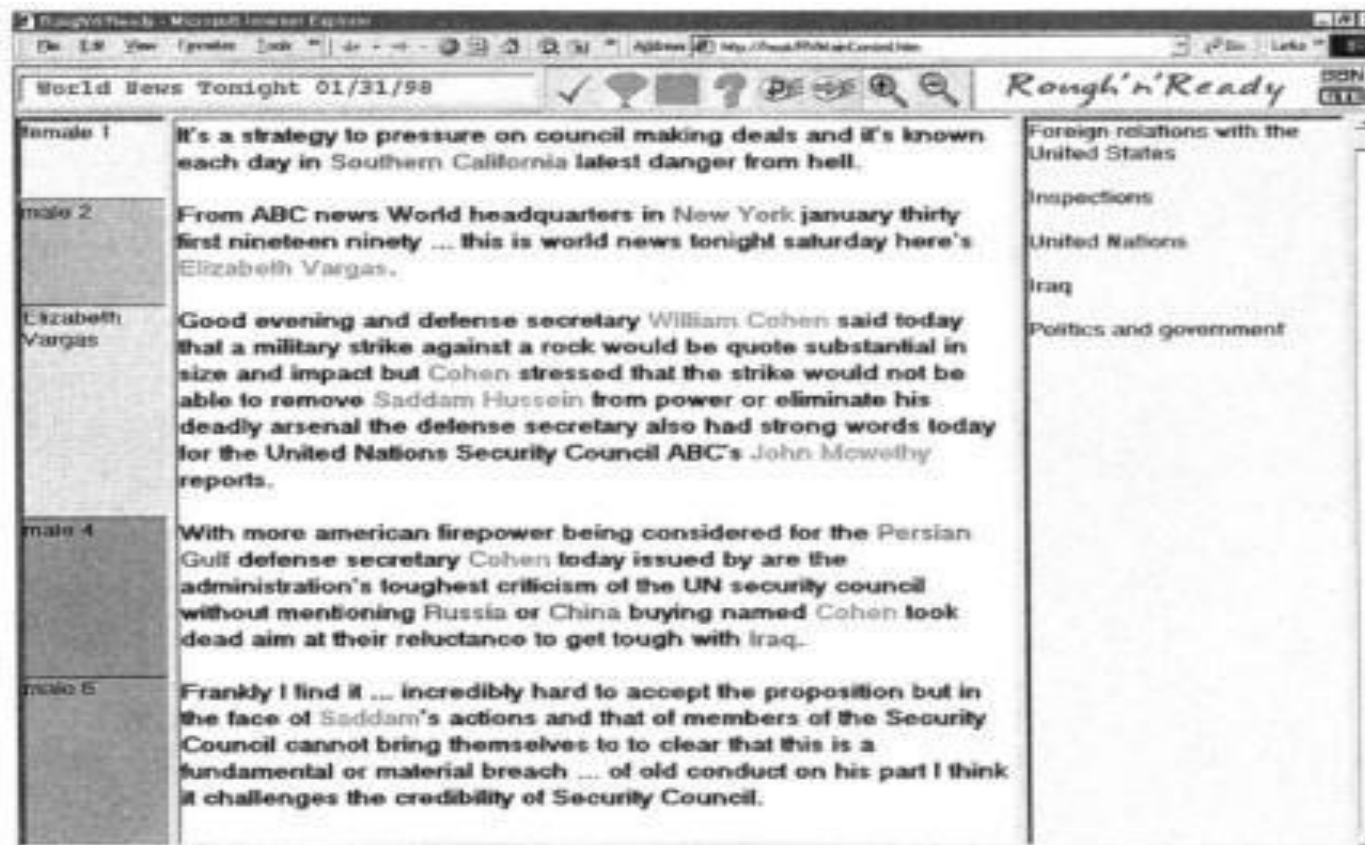


Figure 10.1. BBN's Rough and Ready

# Non-Speech Audio Retrieval

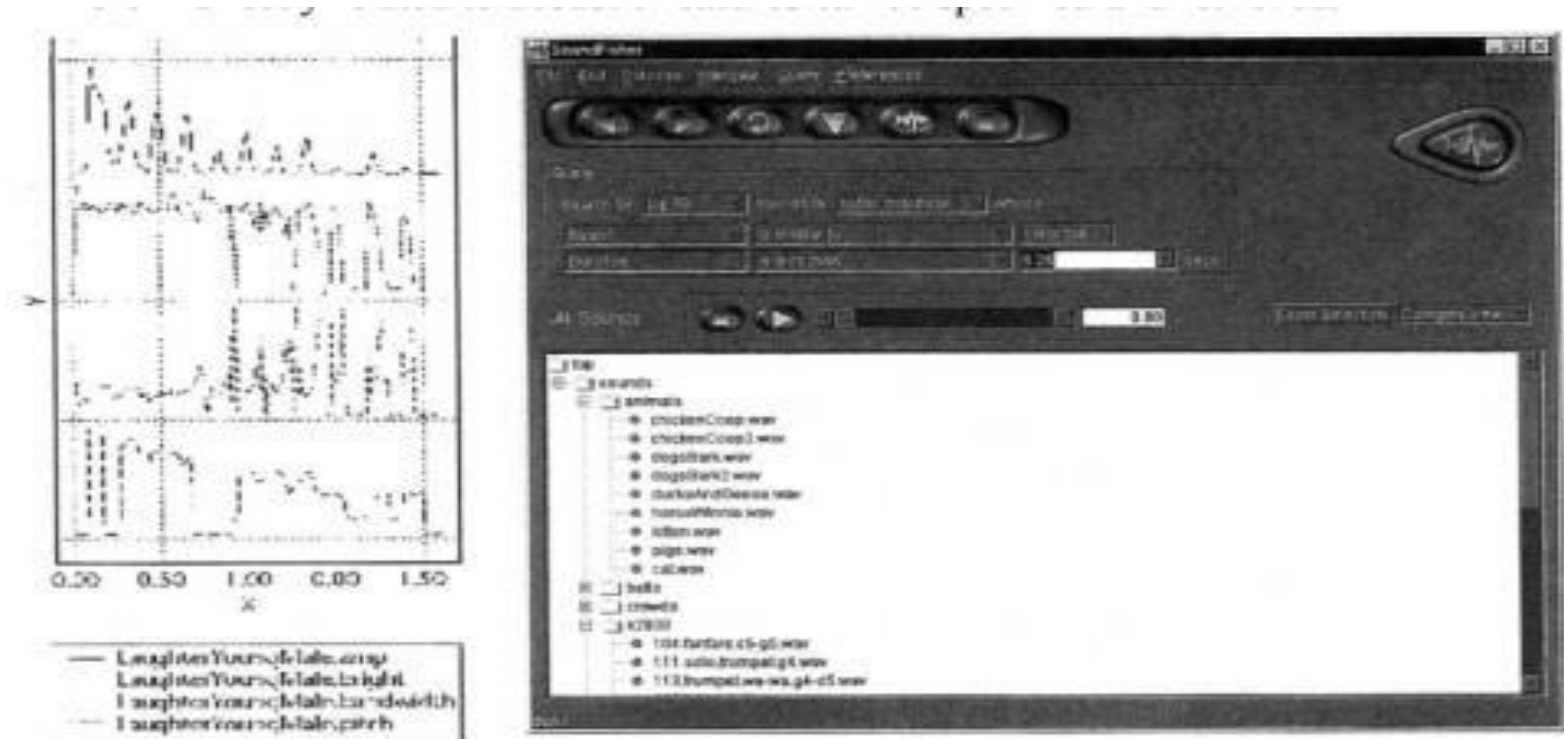


Figure 10.2a. Analysis of Male Laughter. Figure 10.2b. Content based access to audio.

## Graph Retrieval

Another important media class is graphics, to include tables and charts (e.g., column, bar, line, pie, scatter). Graphs are constructed from more primitive data elements such as points, lines, and labels. An innovative example of a graph retrieval system is Sagebook (Chuah, Roth, and Kerpedjiev 1997) created at Carnegie Mellon University (see [www.cs.cmu.edu/Groups/sage/sage.html](http://www.cs.cmu.edu/Groups/sage/sage.html)). SageBook, enables both search and customization of stored data graphics. Just as we may require an audio query during audio retrieval, Sagebook supports data graphic query

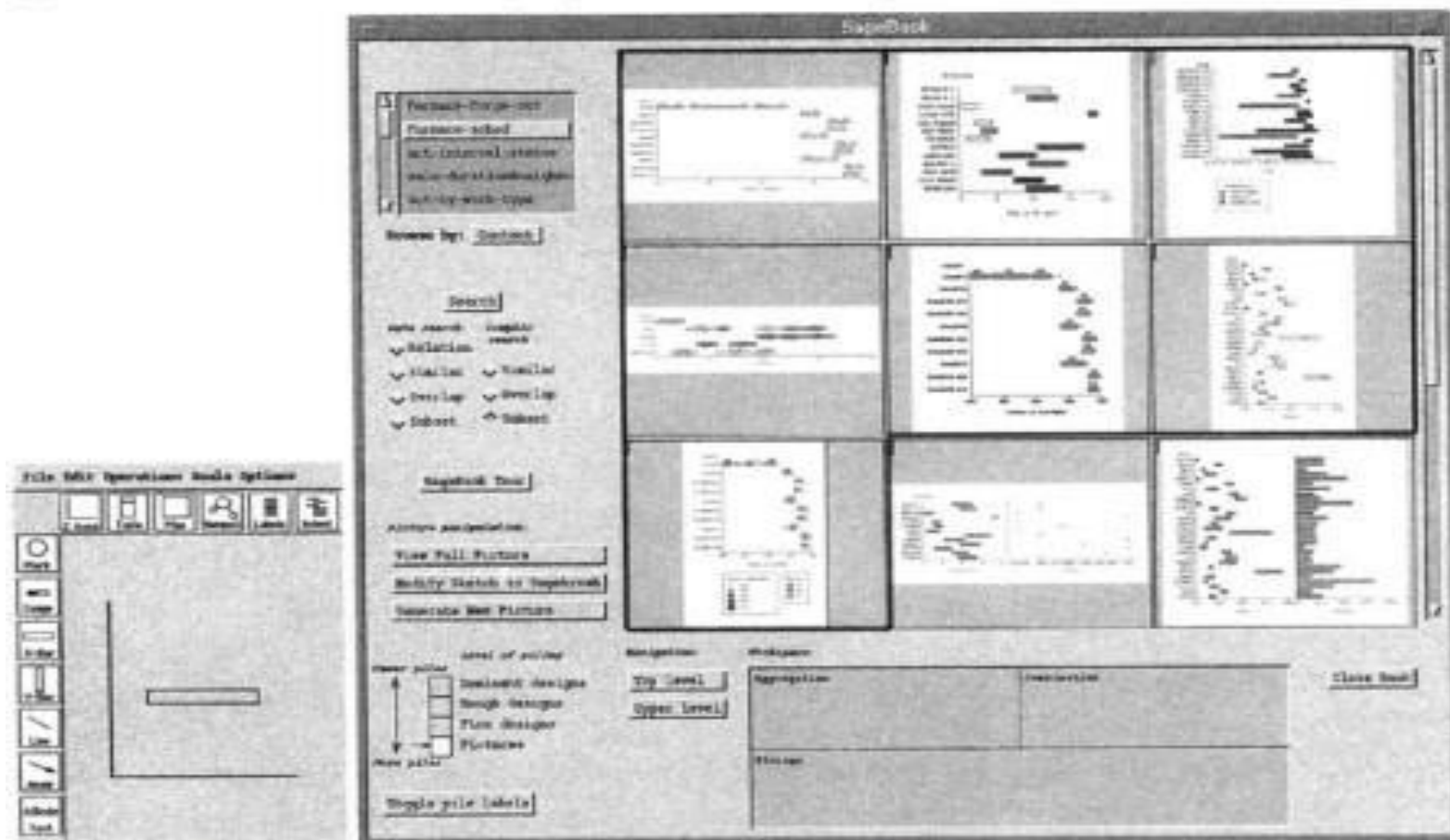


Figure 10.3. SageBrush Query Interface and SageBook display of retrieved relevant graphics

# Imagery Retrieval:

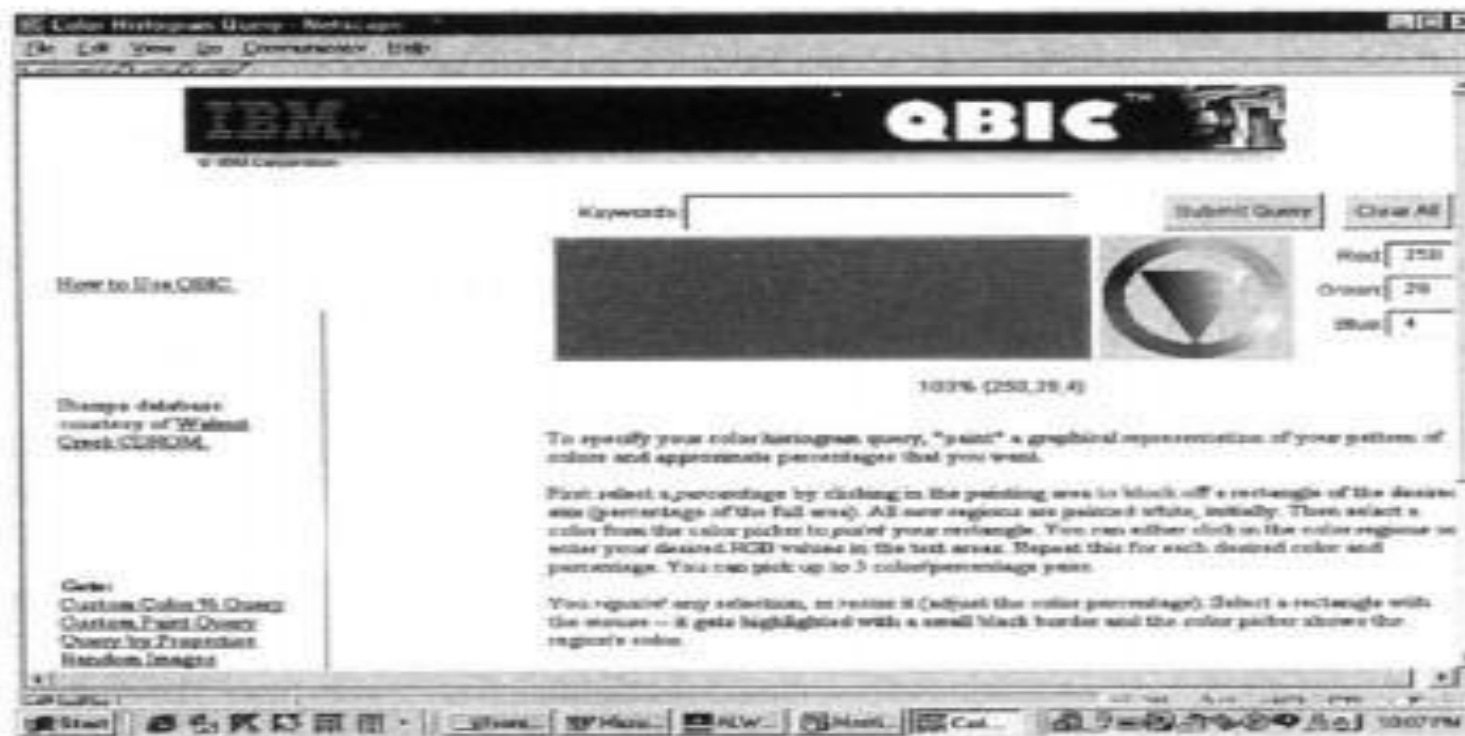


Figure 10.4a. QBIC Query by Color red

# Video Retrieval

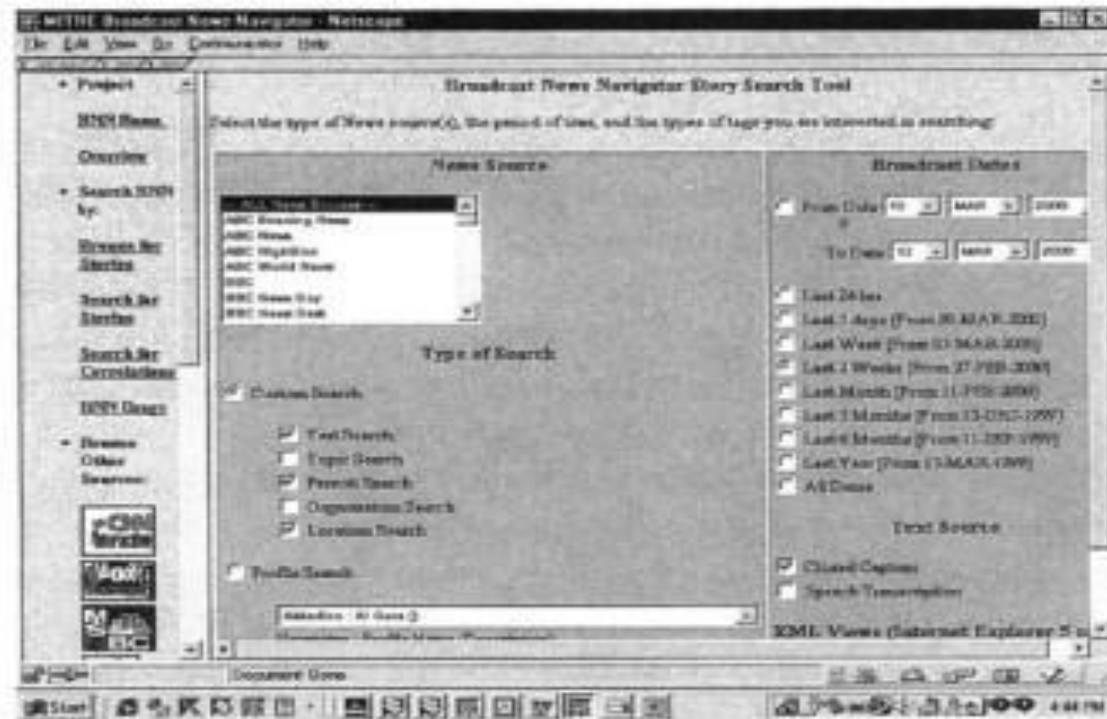


Figure 10.5a. Initial Query Page

THANK YOU