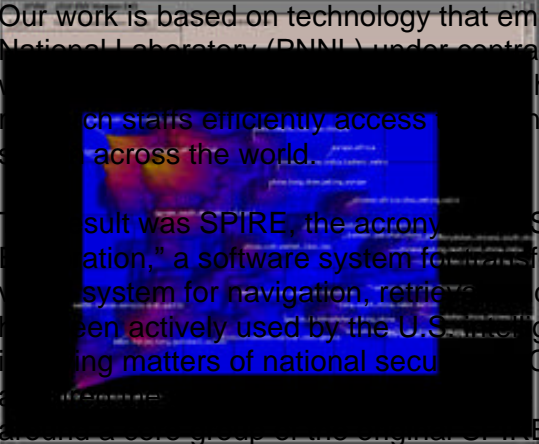# ThemeMedia Company Backgrounder and Technology Overview

ThemeMedia is developing software tools for *"content mapping"* — a process that graphically represents thousands of unstructured documents on a single computer screen for quick, focused navigation, retrieval, and insight.

Our work is based on technology that emanated from the Battelle Pacific Northwest National Laboratory (PNNL) under contract with the U.S. Department of Energy. PNNL were looking for a way to help intelligence and national security research staffs efficiently access thousands of publications, documents, and transcripts spanning across the world.

The result was SPIRE, the acronym for "Spatial Paradigm for Information Retrieval and Exploration," a software system for transforming text-based information retrieval into a visual system for navigation, retrieval, and analysis. Over the last three years, SPIRE has been actively used by the U.S. intelligence community for research and analysis involving matters of national security. In October 1996, the founders of ThemeMedia acquired the exclusive license to SPIRE technology and formed a company around a core group of the original SPIRE team.

Desperately seeking information

Today, the information search method of choice is based on Boolean logic, whereby a document must include one or more user-specified terms, or keywords, to make it eligible for consideration. Existing search engines, such as those offered by Yahoo, Excite, AltaVista, Lycos, and others, typically generate a list of hundreds or thousands of documents, with only limited ability to order them by relevance. Moreover, there is no common measure of relevance to help information seekers determine true value. What AltaVista considers relevant for a particular query, Lycos may relegate to a position of less importance farther down the list. Users are not only at the mercy of how each company defines relevance, they have no way of evaluating the methodology behind the retrieval process — no way of actually *seeing* the relationships among the documents listed.

The weakness of the Boolean search has to do with the user's role in two standard retrieval measures: *precision* and *recall*. Recall measures how well a search produces *all* the documents that fit the search criteria, while precision measures how successful the search is at eliminating irrelevant documents from that pool. If information seekers were capable of knowing exactly what they wanted and, then, how to ask for it, there wouldn't be a problem. But, understandably, it's extremely difficult for most of us to state our precise information needs to a database we can't see and have never explored. As a result, Boolean searches often return too much irrelevant information or not enough of what we really need.

Given the sheer size and number of databases now available, the sweeping diversity of information, and the lack of a common categorization scheme, it seems unlikely that Boolean-based search methods can effectively manage our ever increasing information retrieval needs.

Information Visualization and Relevance

As frustration with existing information retrieval methods mounts, the appeal of visualization technologies grows. Visually-based software tools, like those being developed by ThemeMedia, give users a quick way to actually *see* everything available to them from a given information set, with topics and documents grouped by degree of similarity and level of importance.
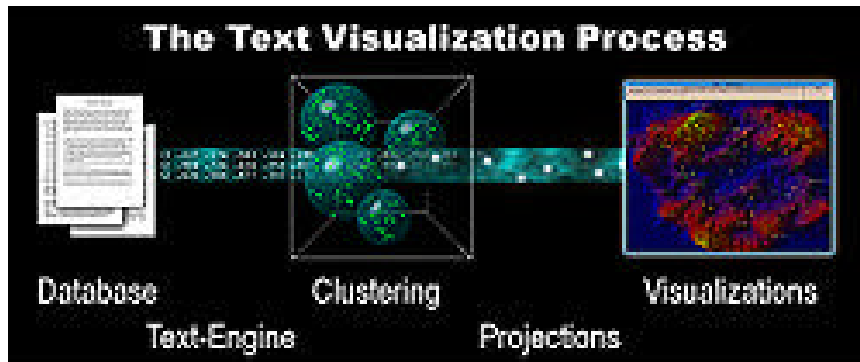
ThemeMedia's System for Information Discovery (SID) starts by capturing any number of documents into a database. By analyzing patterns of word usage and relationships between words, SID autonomously discovers salient themes, derives semantic distances between them to represent degrees of similarity, and transforms the results into vector representations arranged to reveal document relevance.

In this way, ThemeMedia's technology eliminates the typical precision/recall dilemma faced by information seekers — whether to retrieve *all* potentially relevant documents (recall) or only those that are unquestionably relevant (precision). By using ThemeMedia *"visual content maps"* to display information users can immediately *see* everything available, along with the relationships between content and the location of information. Our processes discriminate for the user by recalling information in ample detail through visualization. Consequently, the user is spared irrelevant information, while quickly and precisely navigating to all relevant documents.

ThemeMedia Software Applications

ThemeMedia is in the process of transforming SPIRE, an existing standalone application designed primarily for analysts that runs on an SGI workstation, into a *"new look"* three-tiered client/server application designed for the specialized needs of several business markets:

- Information Providers and Content Aggregators like Lexis/Nexus and Individual Inc.
- Publishers such as Ziff Davis and Knight-Ridder.
- Corporate Intranets and archives.

Our new product application will consist of three modules:

- NT or UNIX server software for capturing and organizing text documents.
- An editorial tool for creating and publishing customized content maps.

- Java-based client software used for navigating content maps and linking to documents.

SPIRE is available today for $5,000/seat that will be fully credited to the purchase of our new client/server application that will be released in Q1 '98. An early adopter/beta site program for our **"new look content mapping software"** will begin this November. ThemeMedia will provide additional information once a Confidentiality Agreement is signed.

For more information and details please contact:

Steve Ardire
Sr. Director Business Development
Error! Bookmark not defined.
PH: 425-602-3559