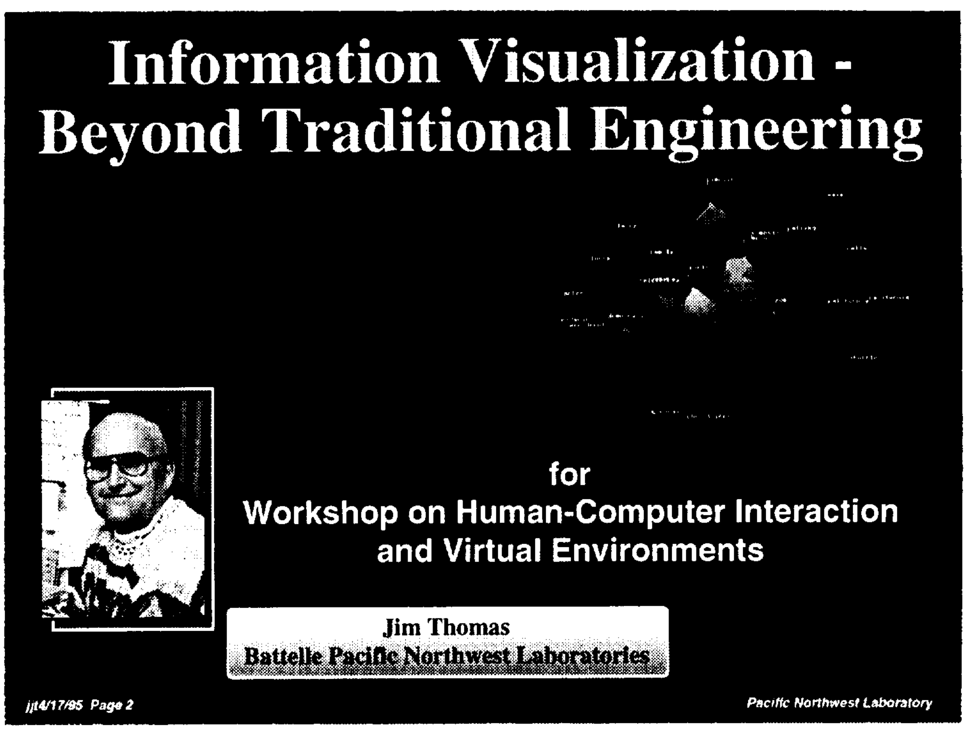
INFORMATION VISUALIZATION - BEYOND TRADITIONAL ENGINEERING

James J. Thomas
Technology Manager
Applied Physics Center
Battelle Pacific Northwest Laboratories
Richland, WA 99352

I would like to thank the Workshop committee for the opportunity to address this audience at this important workshop on Human-Computer Interaction and Virtual Environments.

So far you have heard a lot about the traditional devices and interactions within virtual environments. This presentation will address a different aspect of the human-computer interface; specifically the *human-information* interface.

This interface will be dominated by an emerging technology called Information Visualization. Information goes beyond the traditional views of computer graphics, CADS and enables new approaches for engineering.



TAKE HOME MESSAGE FOR HCI AND COMPLEX INFORMATION SPACES

The take home message is simple.

Information visualization is the visual and interaction technologies supporting analysis of all forms of information including text, images, diagrams, procedures, marketing materials and product quality information ....

Information visualization will add value by providing the holistic approach to engineering quality products.

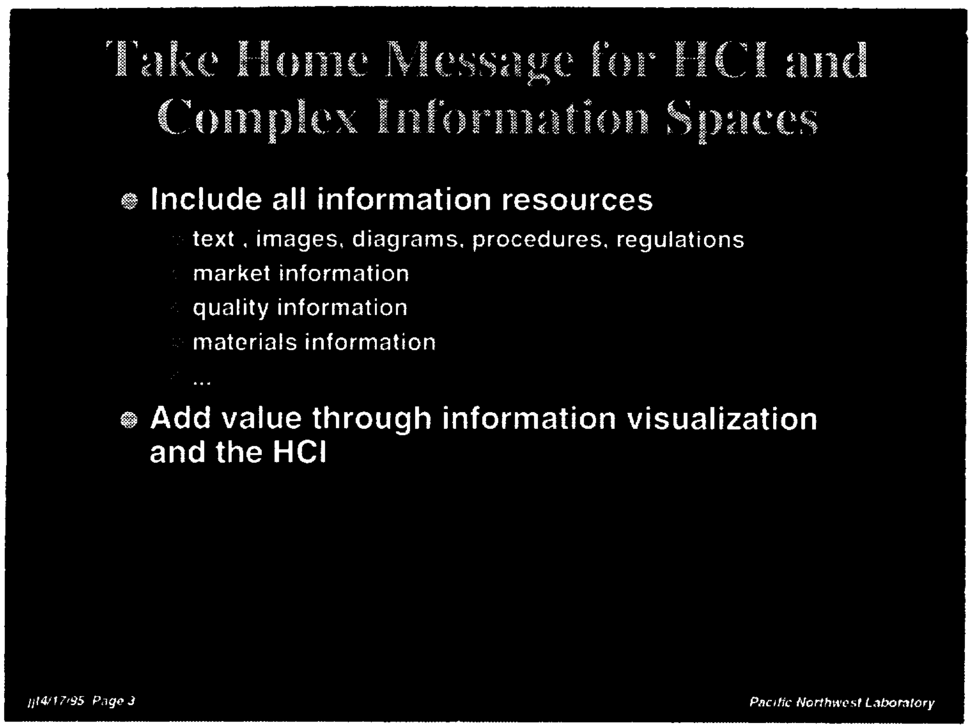


Figure 1

TAKE HOME MESSAGE FOR HCI AND COMPLEX INFORMATION SPACES

However, to enable this fundamental change in the engineering process, our R&D must focus on all forms of information spaces.

Government and industry alike must change their traditional views of scientific visualization. Information visualization will become an integral part of the entire product cycle.

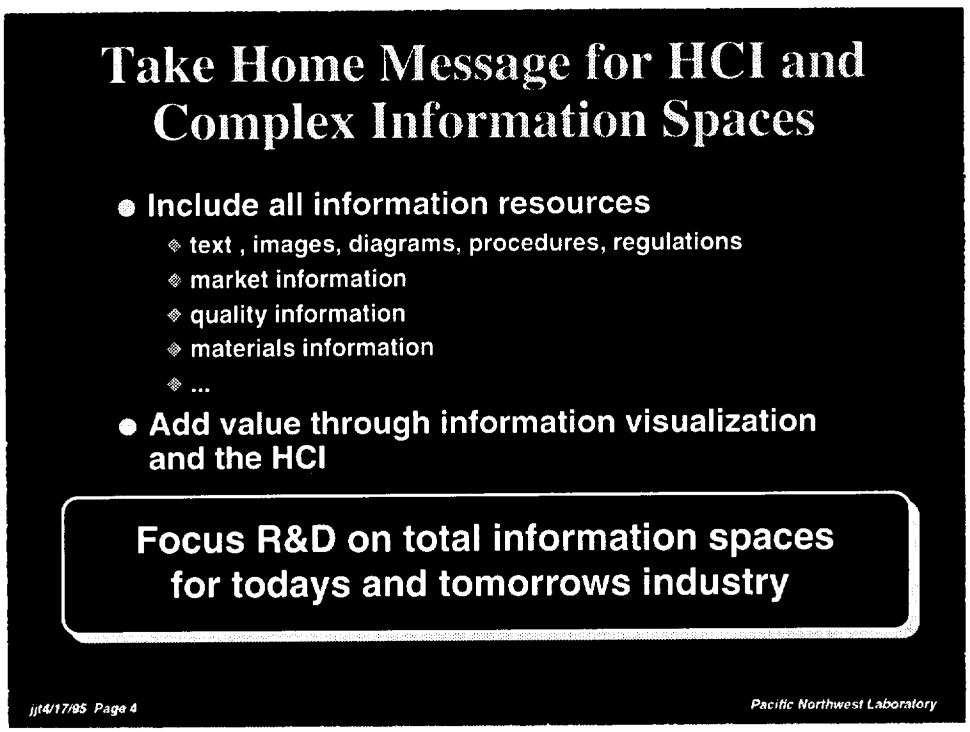


Figure 2

OUTLINE

I would like to organize this talk in two sections:

First, let's look at the progress to date over the last two decades.

Then, let's look at emerging technology for information visualization; specifically, technology to visualize masses of text. This is a fundamentally new approach and is right at the core of information visualization.

Before we start, I want to clarify what is meant by information visualization.

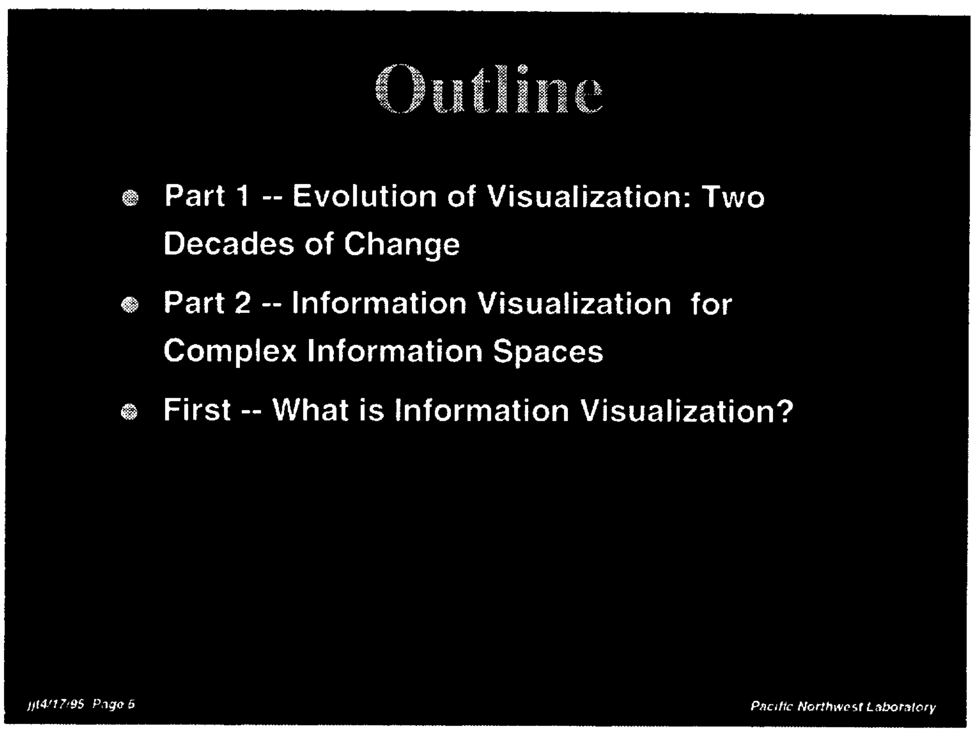


Figure 3

WHAT IS INFORMATION VISUALIZATION?

Information visualization (IV) is not just visualizing the numbers, engineering diagrams, etc. IV goes way beyond traditional scientific visualization based on physical properties. IV specifically must visualize text, documents, sound, images and video in such a way that the human can rapidly interact with and understand the content structure of information entities. These entities are not based on math, physics or chemistry, but rather on concepts; yes, often fuzzy, often incomplete, and often related to other concepts in many dimensions.

It is indeed a high-dimensional fuzzy set of information entities that we deal with every day of our lives.

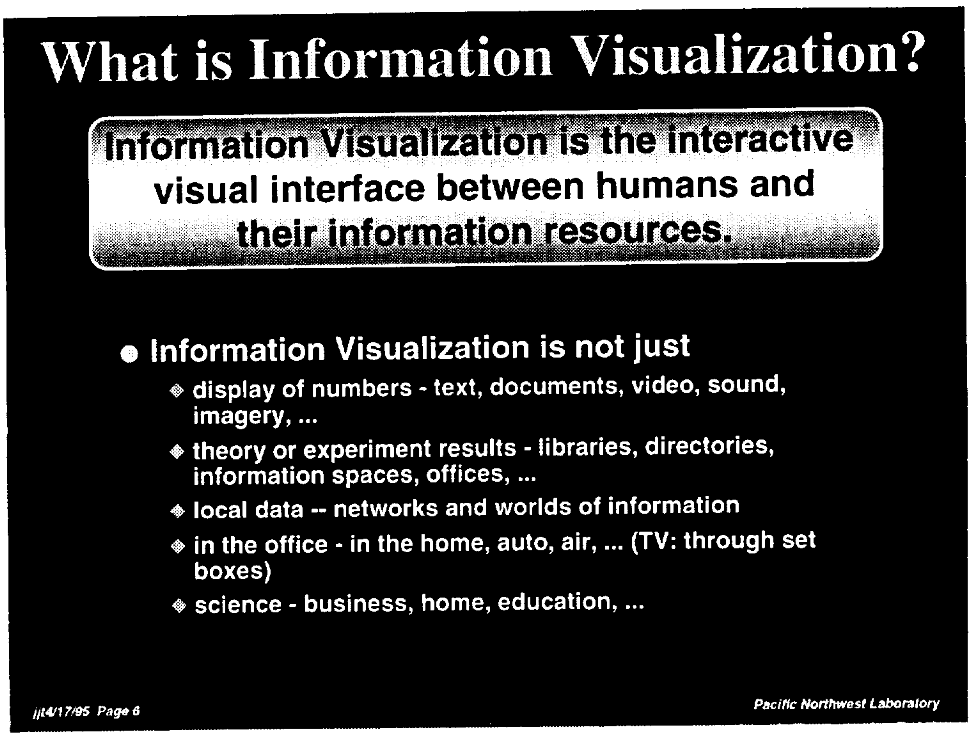


Figure 4

WHAT IS INFORMATION VISUALIZATION?

My colleague, Jim Wise, a cognitive psychologist, says that a primary goal of IV is to provide the presentation and interactions that match our trained perceptual capabilities. These capabilities are developed in our childhood and perfected as we mature both in the home and office settings. These capabilities also allow us to visualize and understand masses of information simultaneously. Look at our desks as only a part of the total information space in the office.

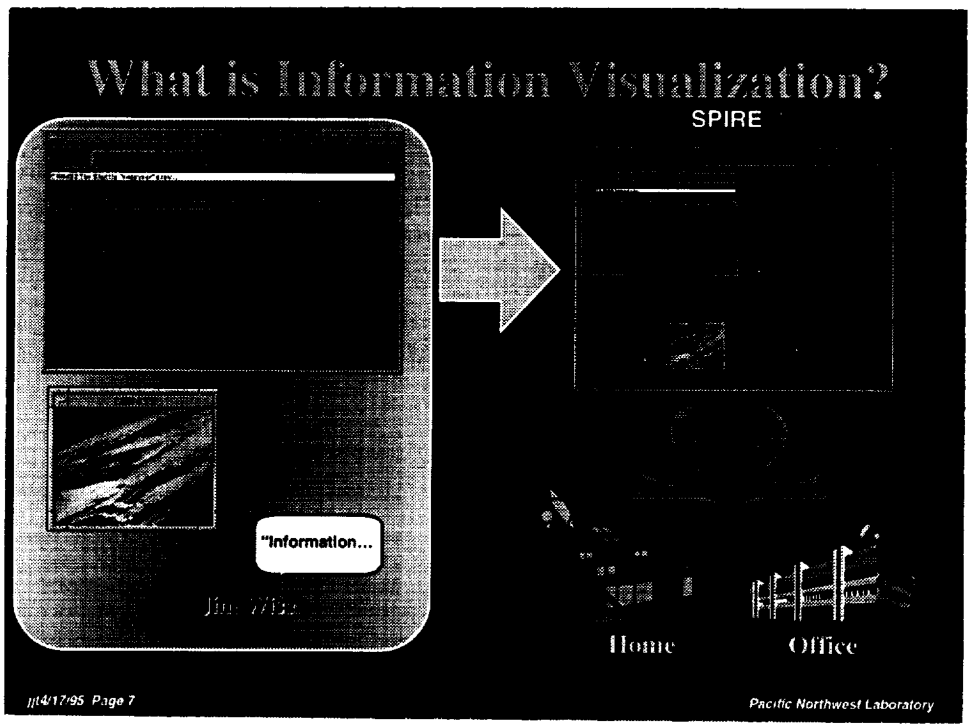


Figure 5

PART 1 - TWO DECADES OF CHANGE

Now let's look at what has happened in the last two decades. I think that you will be as surprised as I was when I spent some time reviewing our progress.

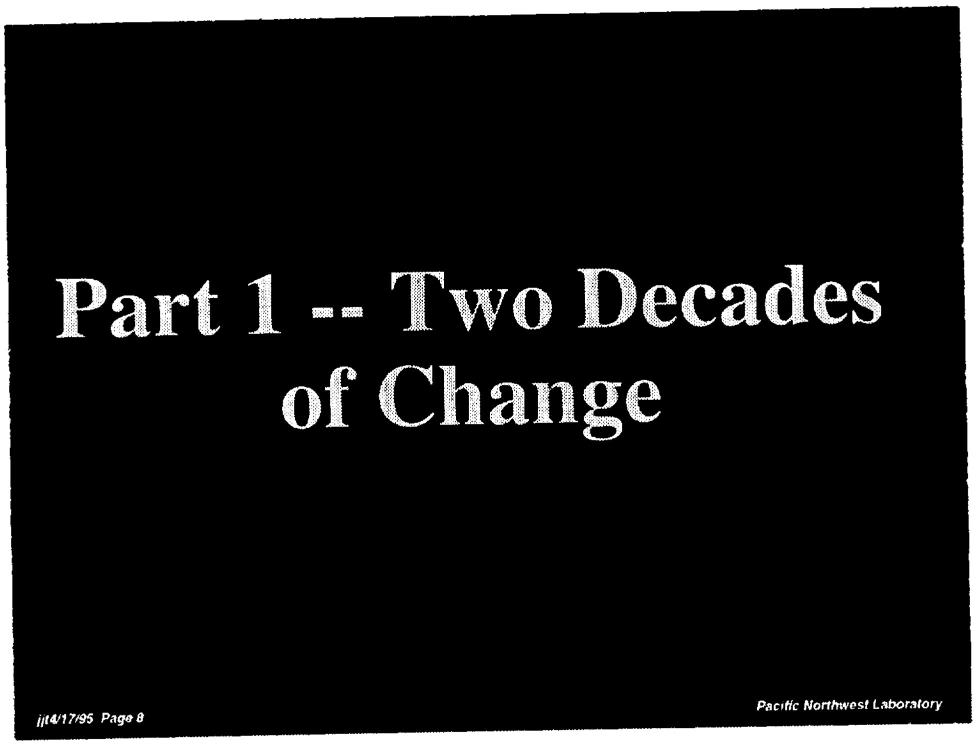


Figure 6

We are being driven by a dramatic change in our society. This was discussed by Alvin and Heidi Toffler in their book called War and Anti-War. If you have not read the book, I highly recommend it. It will enable you to think in a different perspective.

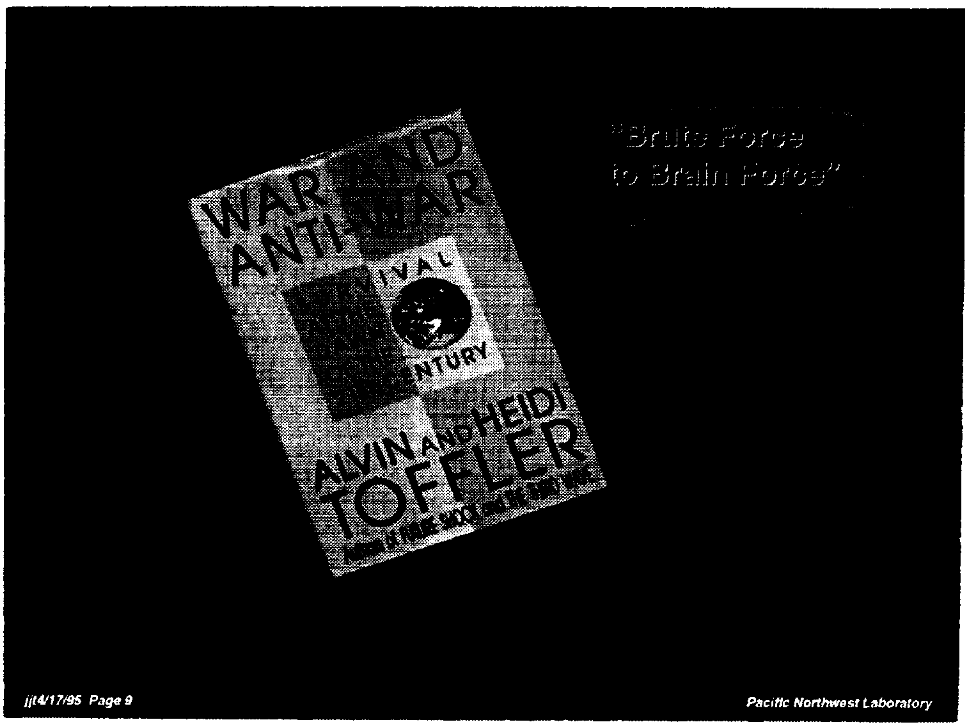


Figure 7

Leaving the geo-centric era behind, we are indeed entering the geo-information era.

The ability to deal with masses of information will be a key part of competitiveness within and between our societies.

Talk about quotes.

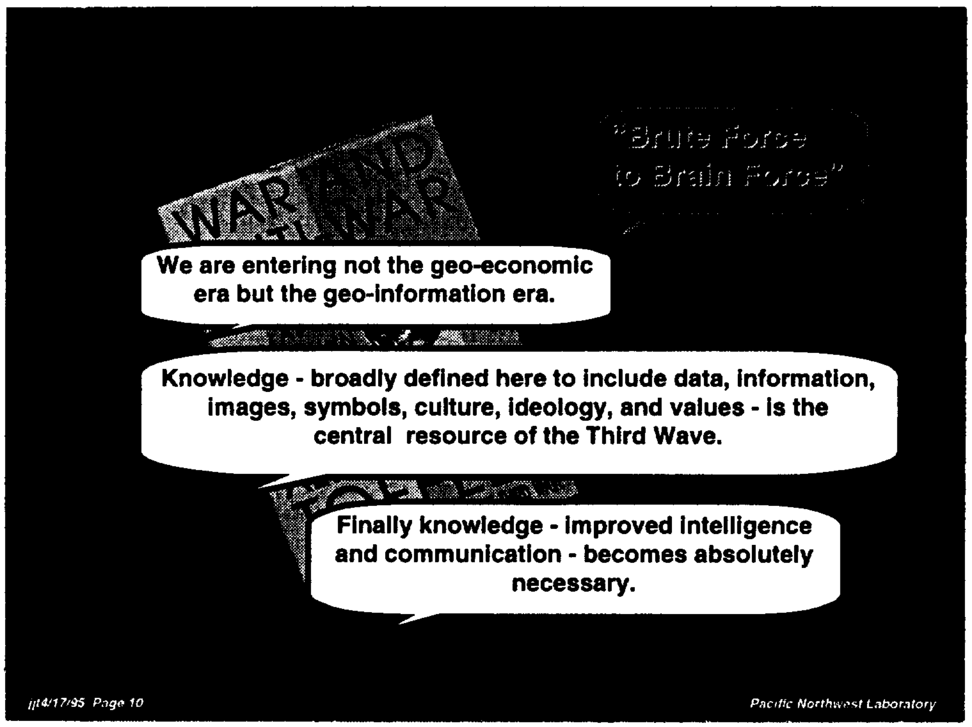


Figure 8

TWO DECADES OF CHANGE

Now that we have seen the motivations, where are we?

In the 1970's we invented scientific visualization that was driven by initial needs for CAD for the automobile and aircraft industries. Some of you in the audience are pioneers in this arena. Then technology was driven by the needs and $$$ within the entertainment industry. This was followed by a continuing push to enable the use of multimedia sources of information. However, the primary source of information, TEXT, was not addressed. Until we address this key form of information we will see little change. Some may not agree.

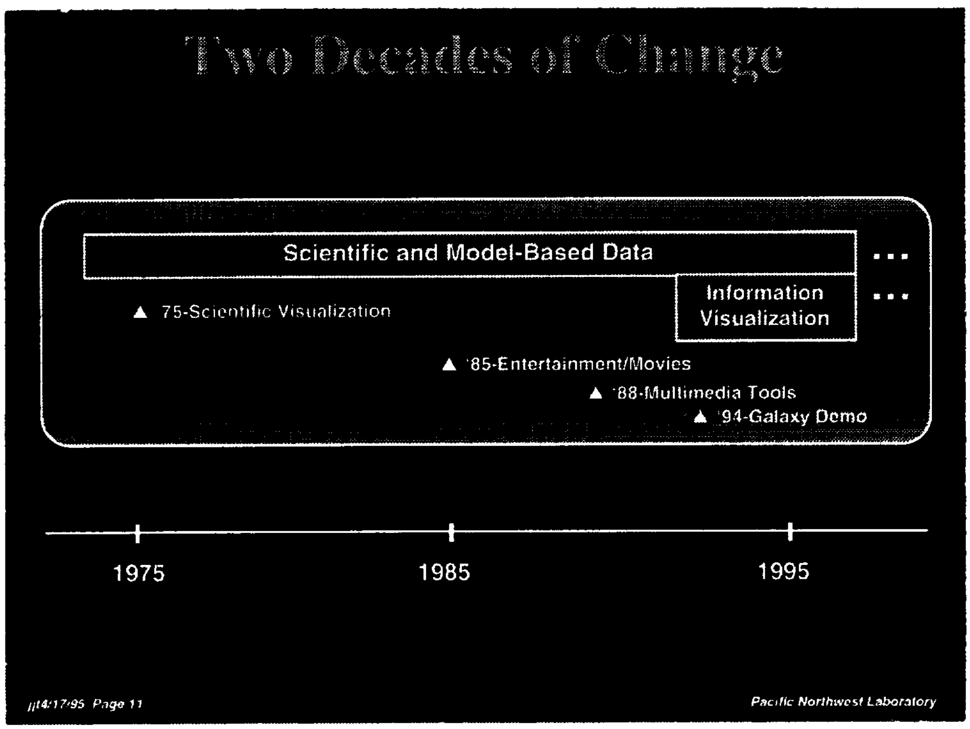


Figure 9

VISUALIZATION TECHNOLOGY EVOLUTION

Let's look at the changes in some visualization technology from the early 1980's through 1995. These are segments of film from contributors at the SIGGRAPH Film and Video show held each year. They are available and referenced from the Video Review below. These pieces are major contributions, and credits to the authors are provided.

Please note that the initial sequences took "CRAY" hours per frame to generate. Today these segments are almost always computed in real time; some are. Surprisingly, we see little change except for quality and speed.

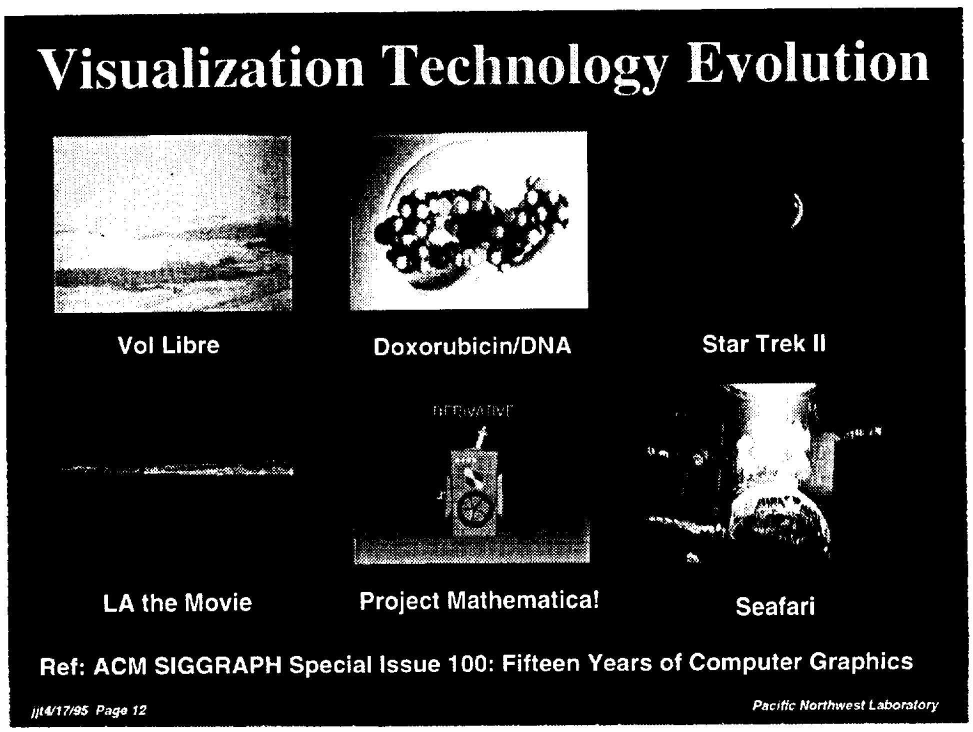


Figure 10

LAST TWENTY YEARS - OBSERVATIONS

You have already seen current technology in engineering visualization. What can we conclude?

- Faster and cheaper by two orders of magnitude - quality is definitely increased.
- It is being used effectively by scientists and engineers.
- The entertainment community has been the dominate driver.
- Many applications are being driven by the technology versus the application needs.
- We can deal with much larger volumes of data.

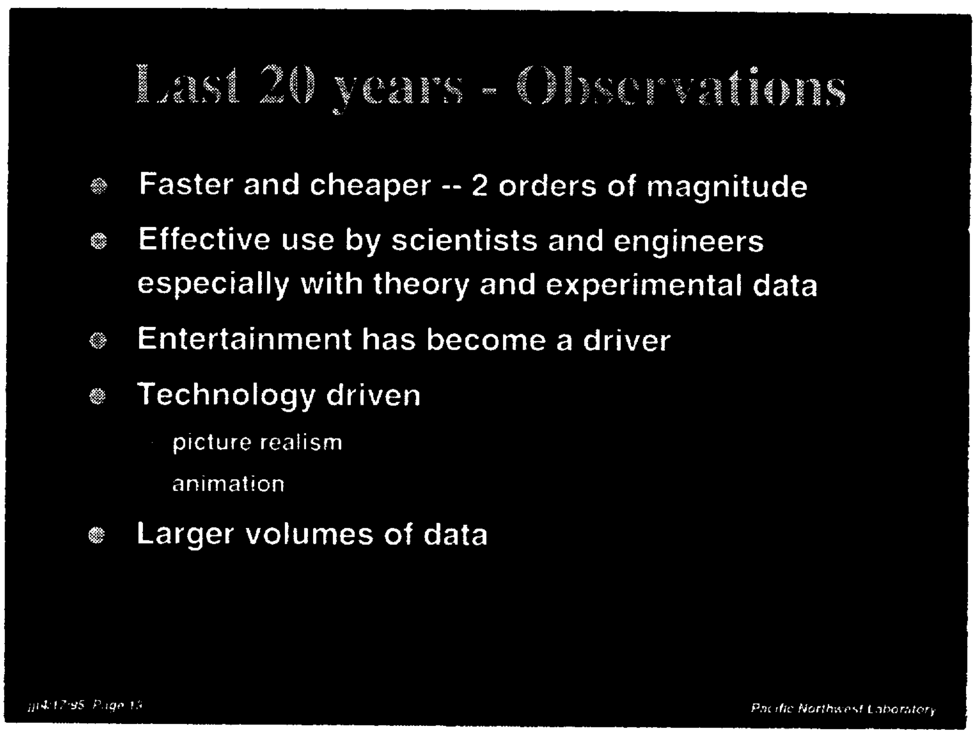


Figure 11

LAST TWENTY YEARS - OBSERVATIONS

- Interaction technologies are essentially the same with the possible exception of some of the VR interaction devices. It is windows; point and click, and WIMP.

- Today's visualization is largely based on the physical properties of materials and designs.

- There is little change in how we deal with information analysis. We can even simulate the old methods somewhat faster on the computer. With the changing data volumes, it is hard to determine any real progress.

- We mostly analyze information of known and complete structure.

- There is little-to-no technology for the inclusion of the total information spaces for analysis.

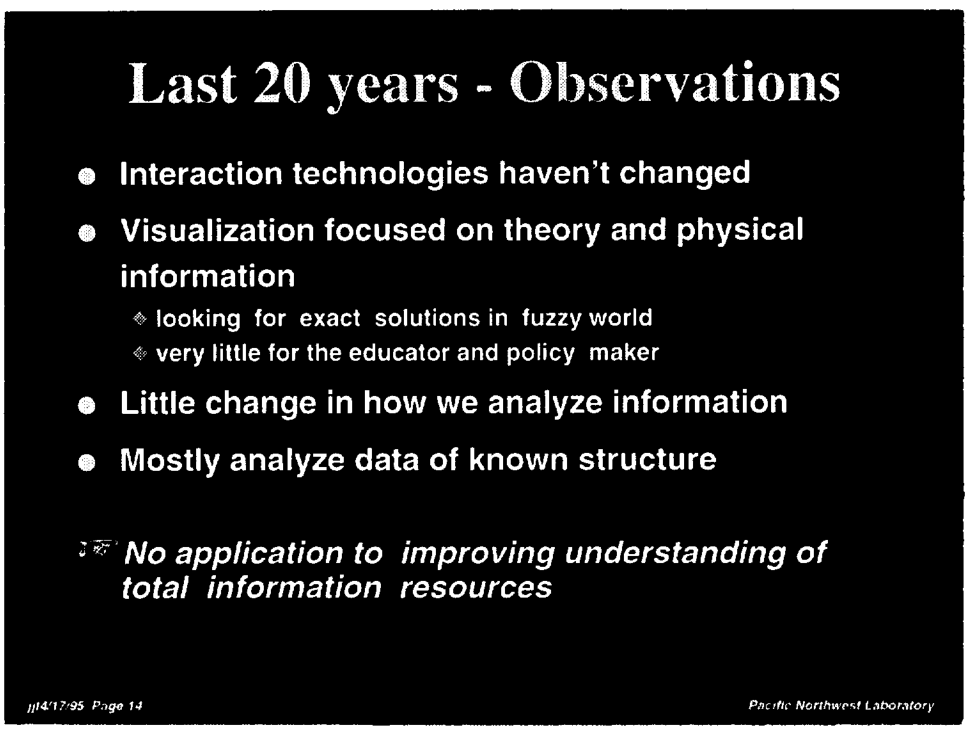


Figure 12

PART 2 - INFORMATION VISUALIZATION

Information overload is about to happen. You may say that you are swamped today. Consider what will happen when you have 100-1000 times the information available to you on any topic of your interest. That is the issue to be addressed by IV.

There are some leading researchers in the field attempting to address this.

PNL's approach is to conduct research while delivering today's technology, getting direct feedback from information analysts using a systems engineering methodology. This is the subject of another talk. This talk will focus on the resulting technology.

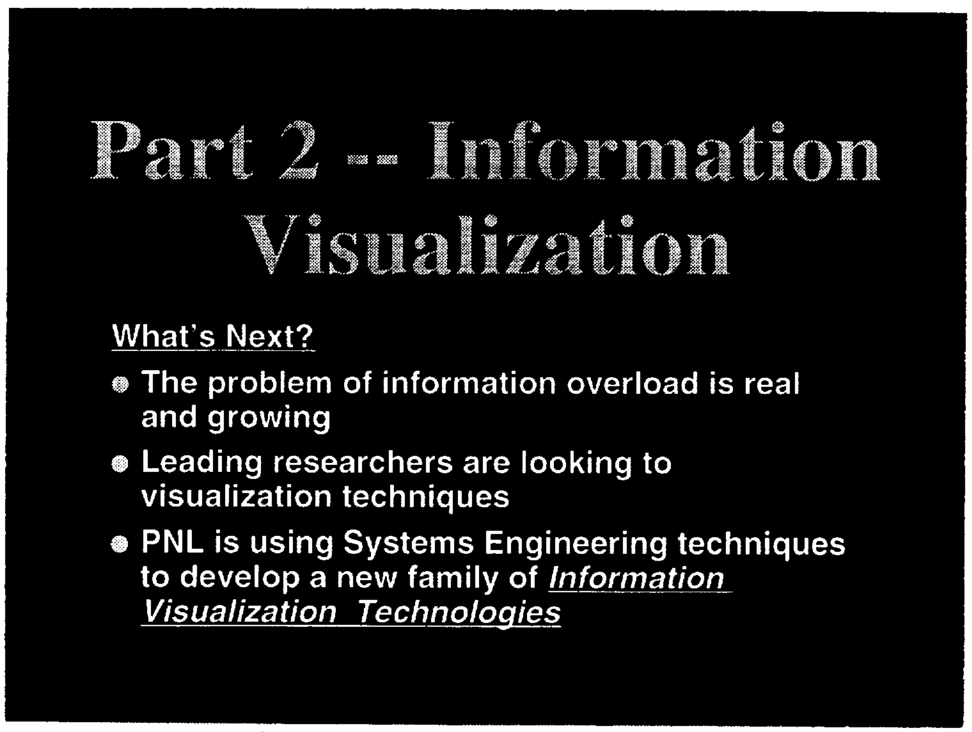


Figure 13

INFORMATION OVERLOAD

Is this your office?

Play video - illustrating a typical information analyst.

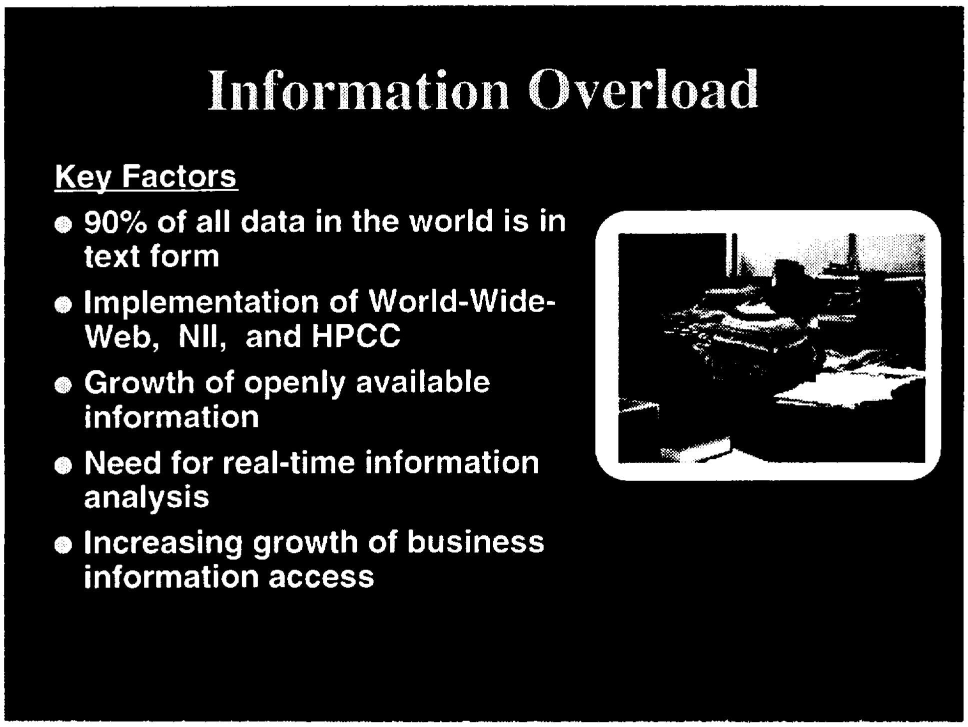


Figure 14

PURPOSE OF INFORMATION VISUALIZATION

The purpose of IV is to:

> **Provide an enhanced method of analysis that enables discovery, understanding and presentation through the use of computer graphics and the interactive interface between the human and their information resources.**

Note: Discovery, understanding and presentation are the three fundamental activities of an information analyst.

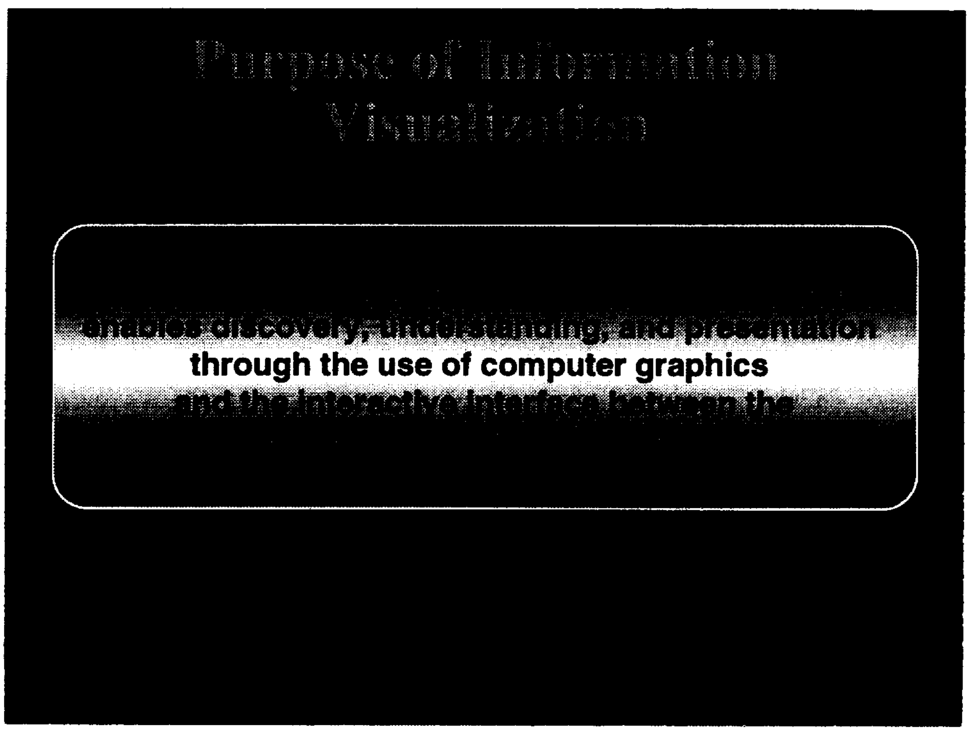


Figure 15

RESEARCH IN INFORMATION VISUALIZATION

Examples of significant contributions are from:

ALTA Analytics - Netmap. Their software provides that each phrase is related to other phrases through a vector map, which is illustrated. This works well for small numbers of single-dimensional information spaces. They have some nice interaction tools that help with the more complex information spaces.

Xerox Parc's Cone Trees provides an excellent approach for IV of hierarchical information.

Ben Schneiderman's scatter plots, demonstrated by the FilmFinder, is an excellent example of interacting with medium-sized data sets. This provided many clues leading to the following work from PNL.

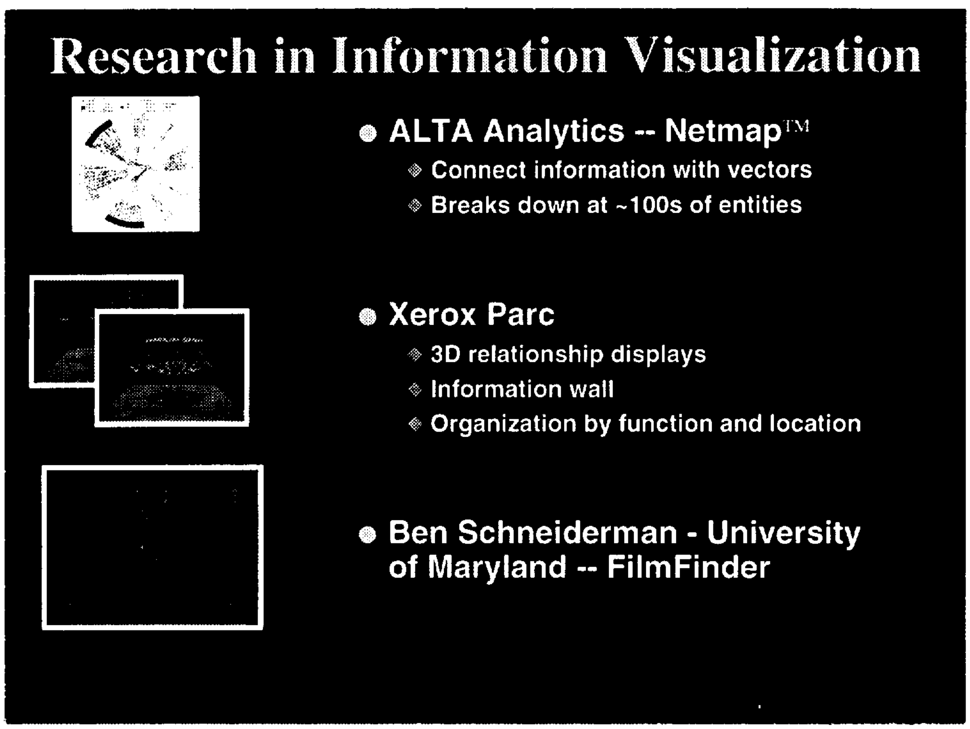


Figure 16

KEY TECHNOLOGIES FOR INFORMATION VISUALIZATION

There are at least ten fundamental technologies that support work in IV. These have been documented within a state-of-the-art report that is available if you are interested. Please contact me via e-mail (JJ_Thomas@pnl.gov) for a copy.

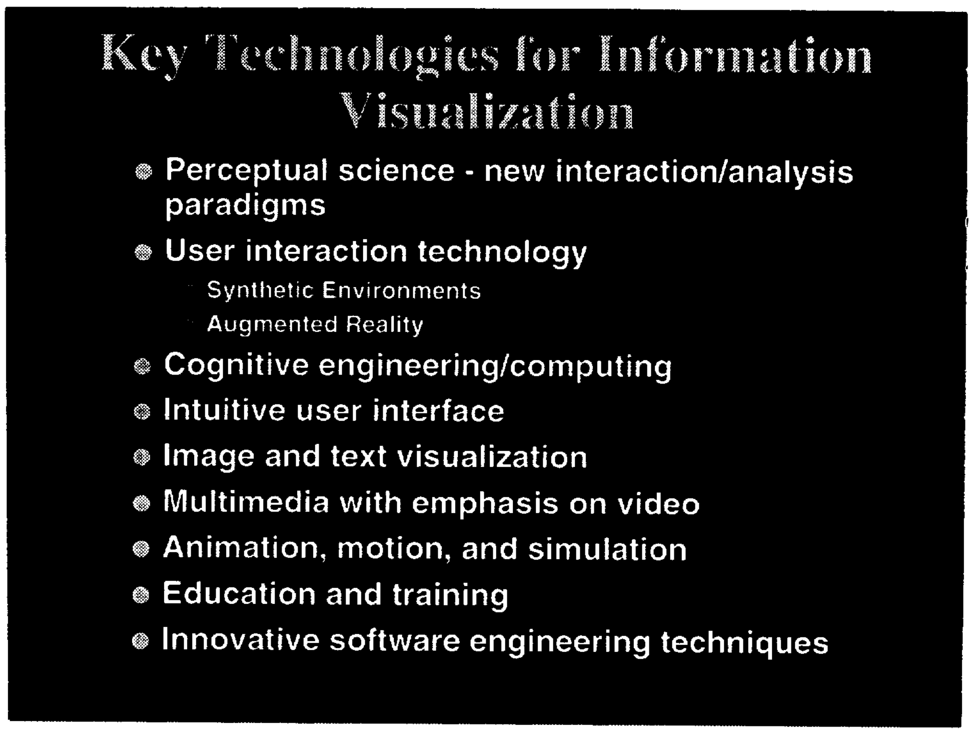


Figure 17

THE SPIRE™ SYSTEM

The work at PNL is now within a system called SPIRE - Spatial Paradigm for Information Retrieval and Exploration. The systems approach discussed below was followed.

We developed the concepts of Galaxies, received rapid feedback from users, and are now developing Themescapes within SPIRE. These will be illustrated.

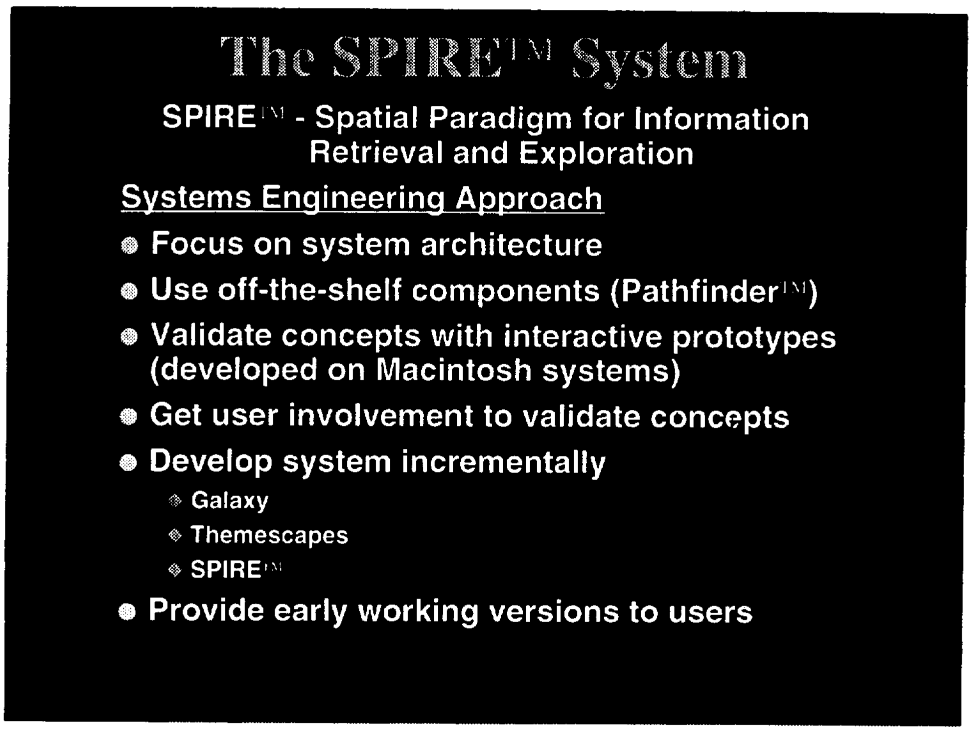


Figure 18

SPIRE™ SYSTEM ARCHITECTURE

This is the architecture for our technology. Note that the technology is designed to be included in other application suites. This will someday be a stand alone application, but currently must be associated with other analysis tools. Details of the architecture are contained within another paper. Please contact me if you would like more information.

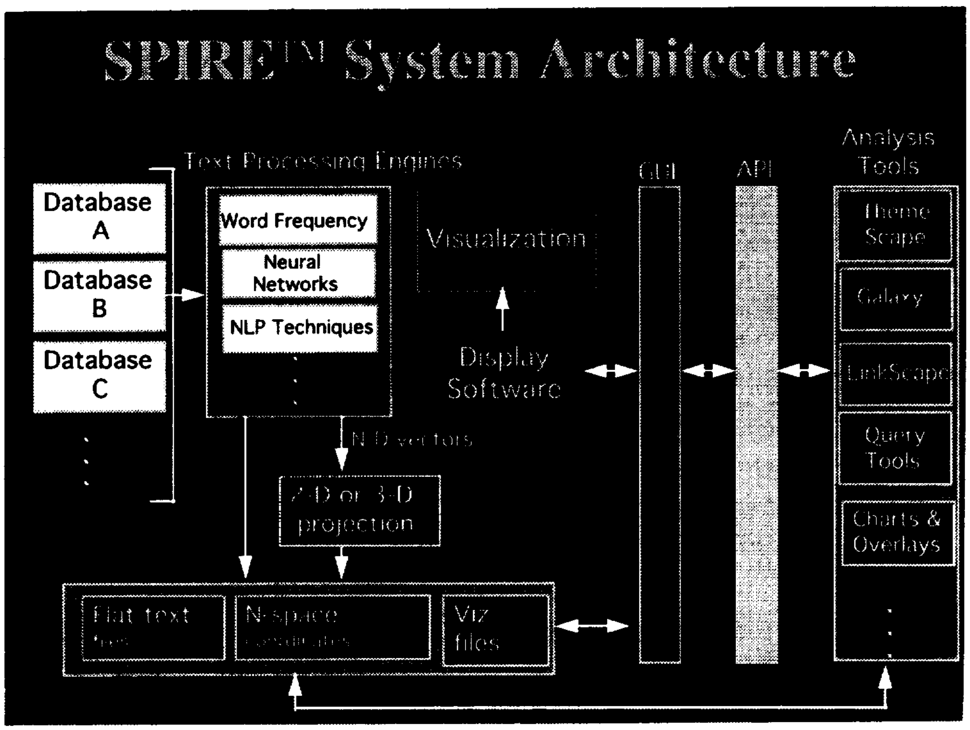


Figure 19

THE GALAXY PROTOTYPE

This is an example of the original information space from Galaxies.

The core concept is that each dot is a document. Two dots that are close to each other are similar in content. If they are far apart, then they are dissimilar in content. The clusters indicate a group of documents within similar content.

The process to gain this visual is complex. First, we obtain or calculate a proximity measure for all the words within all documents. This is a high-dimensional space. Then, we project this high-dimensional space into a view space, as illustrated.

Note that the axis means nothing. Only proximity has meaning within this visual.

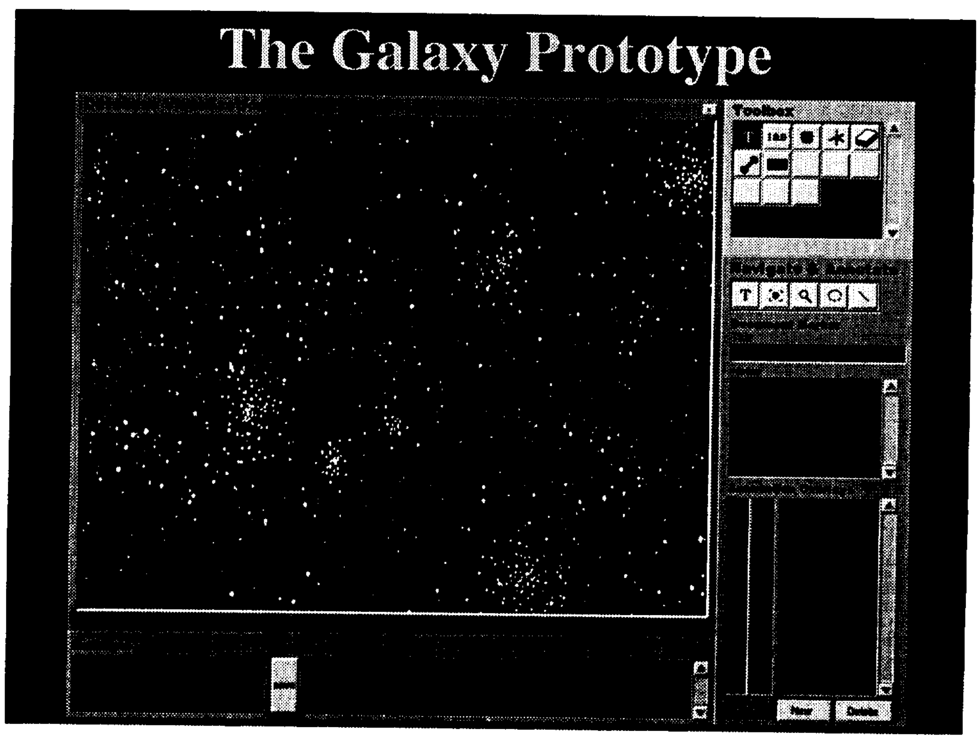


Figure 20

This is the first real analysis performed by our users. It provided the core understanding of about 600 documents so an analyst could rapidly find the right information. These were abstracts from an on-line service covering a topic. The analysis of the 600 documents was completed in less than an hour, with an understanding of what the state of this specific technology was, who was collaborating on its development, when it was developed, and what are the "close" technologies that illustrate an understanding of the issues.

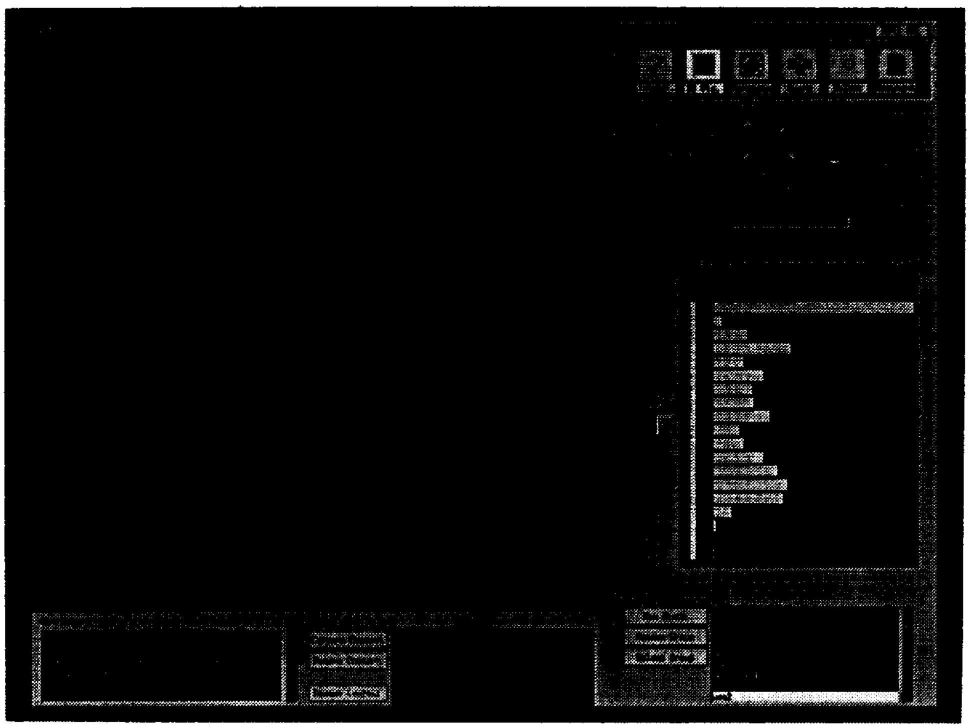


Figure 21

APPLICATION OF SPIRE™

The latest technology now allows for not only significant increase in the information space but direct visual understanding through technology now called Themescapes.

For a test data set we have selected the closed-caption from CNN. This is easily obtained through a small box connected to the television. It is important to note that this approach requires no knowledge of the topic or information space, no pre-formatting or keywording, and simply develops the proximity measures based on the content with the unstructured text. There is no human intervention in the process. This is not an AI based process.

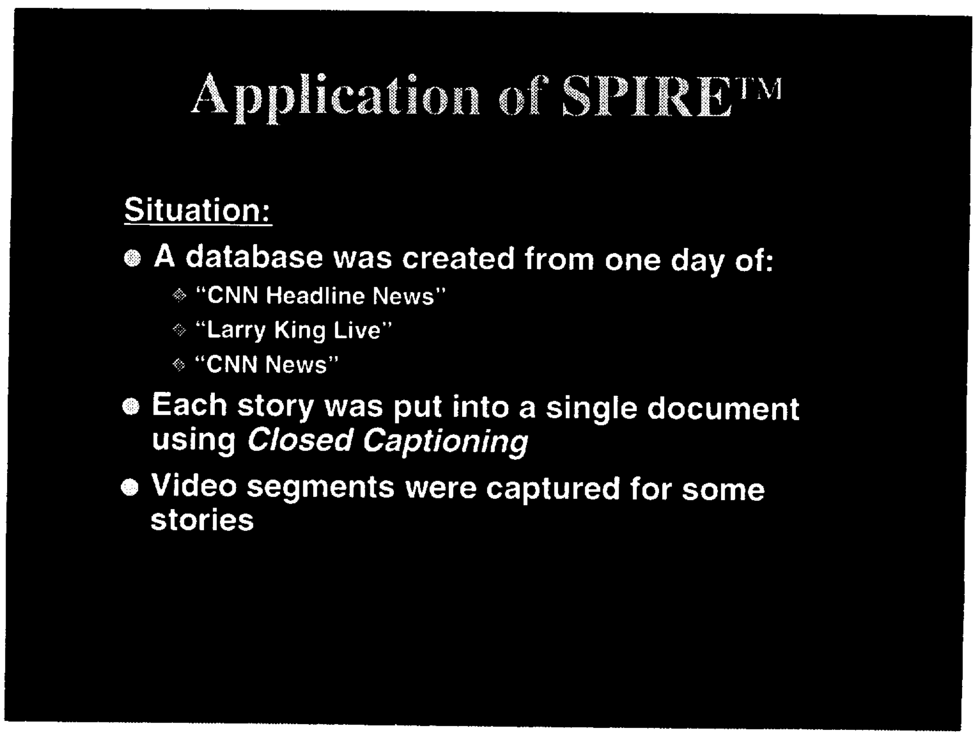


Figure 22

This is the essence of a week's programming at CNN. The large dot represents a cluster containing N dots. Each dot is a news article. The user can open and close clusters. The proximity of one cluster to another cluster is based on the information content closeness to all others. Then within a cluster the proximity of dots is based on how each dot relates to all other clusters as well as the documents within its parent cluster.

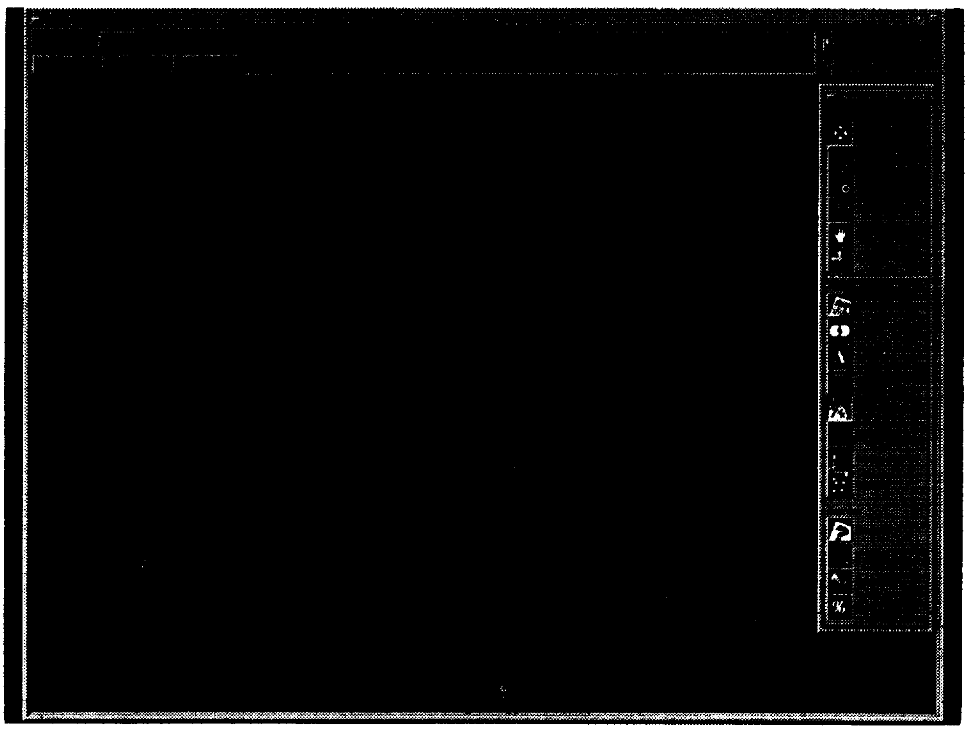


Figure 23

We have a time slicer that allows for time-based analysis from a minute-to-a-year intervals. This illustrates a five day span of time on the CNN channel, with the last two days being highlighted.

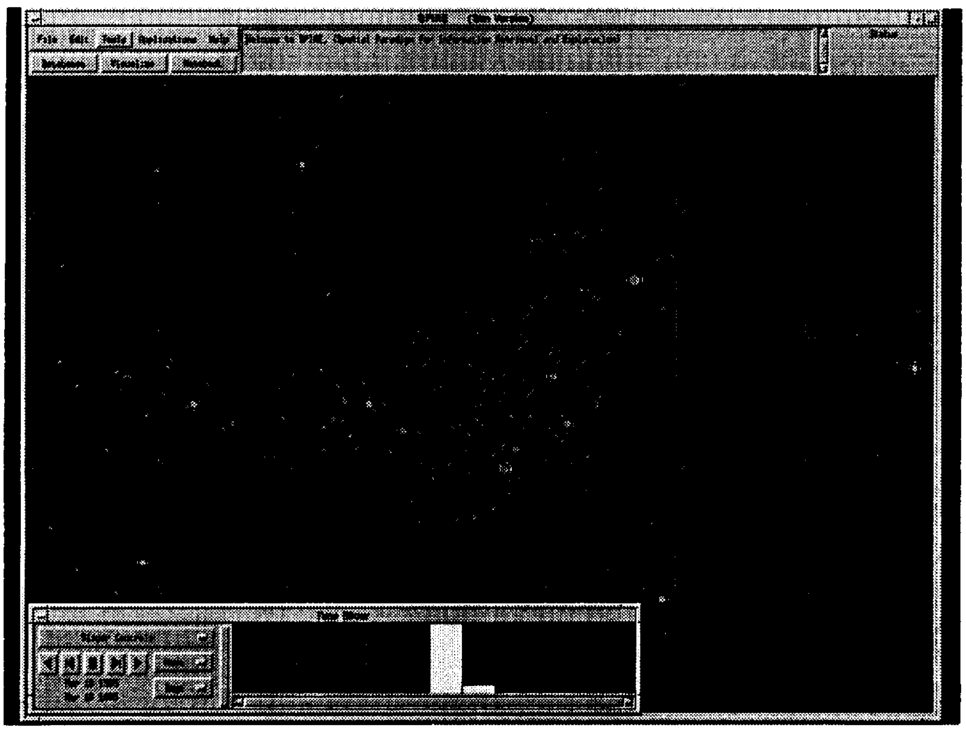


Figure 24

The lower left tool illustrates theme probing. Looking at a top view of the document clusters, one can select points, determine the primary words and make of the theme. Again, these are automatically selected based on the content and discovered relationships within the document space. The middle left tool illustrates the selection tool for subsetting, union and intersections of clusters and selected groupings via full text searches.

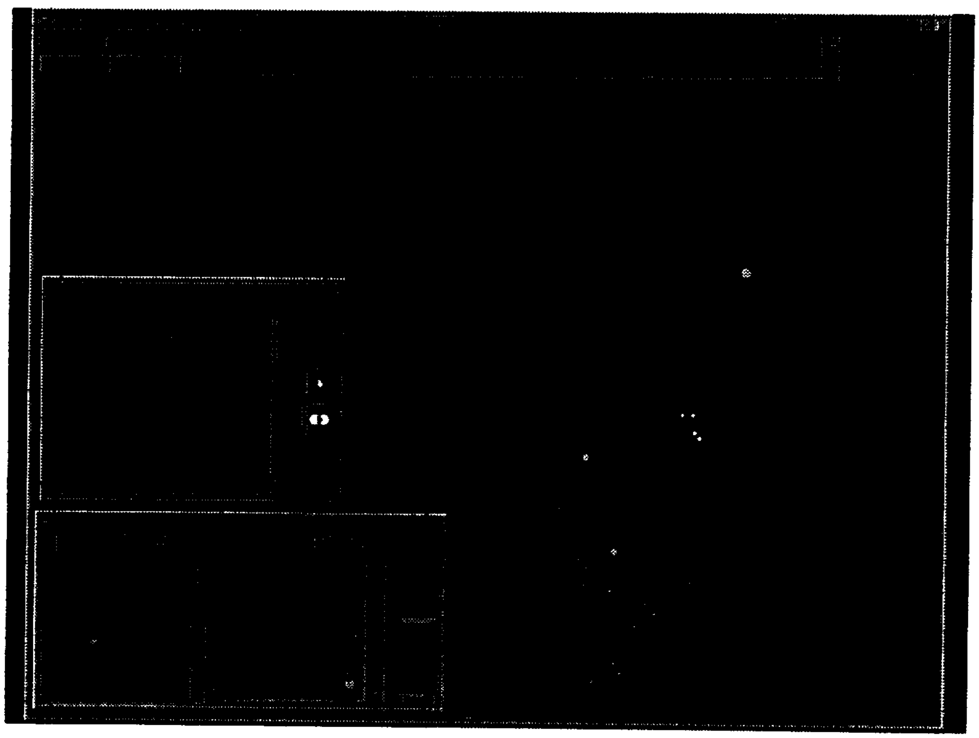


Figure 25

This visual illustrates a new concept called Themescapes. Note the basic structure similarity to the document space. This provides a landscape based on the thematic infrastructure contained within the document space. The basic principle of proximity holds. Note the themes on the right middle. These are the MCI and AT&T commercials.

The two California peaks in the lower left indicate two topics closely related. One is the weather news during the recent heavy rainstorms and the other are features about the damage.

Note the three OJ clusters. Each have a different thematic structure. They are close to the Sports clusters. The Larry King clusters are on the upper left and are separated by major topics.

Not only does the proximity contain information but the scape of the terrain is information rich.

Figure 26

This visual helps you see the shape of the themescape. The sharp slopes between the Sports and Simpson topics indicate that they are close yet there are some fundamental differences, as you might expect.

Also, note that the Shuttle is close to California as it was landing that week in California. To take full advantage of this visual, one needs a rich suite of interaction tools. A few of these tools will be available during the first release of this software.

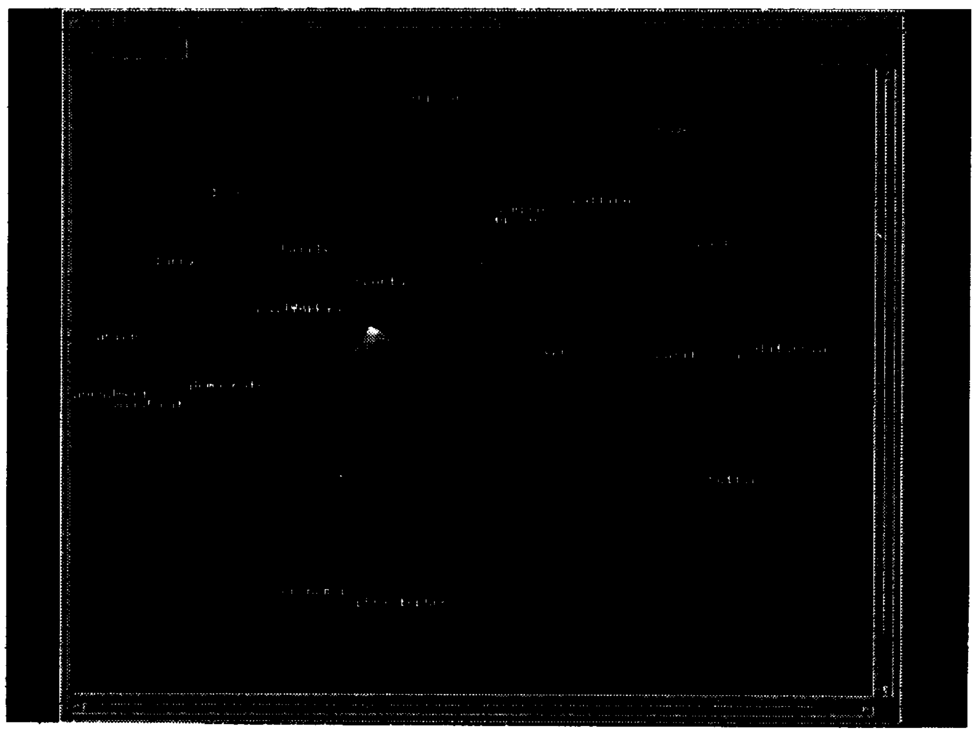


Figure 27

One can combine and reshape all of these visuals so that the information analyst can see each one simultaneously. This is important for complex information space analysis.

This also illustrates the result of a search for all documents dealing with California.

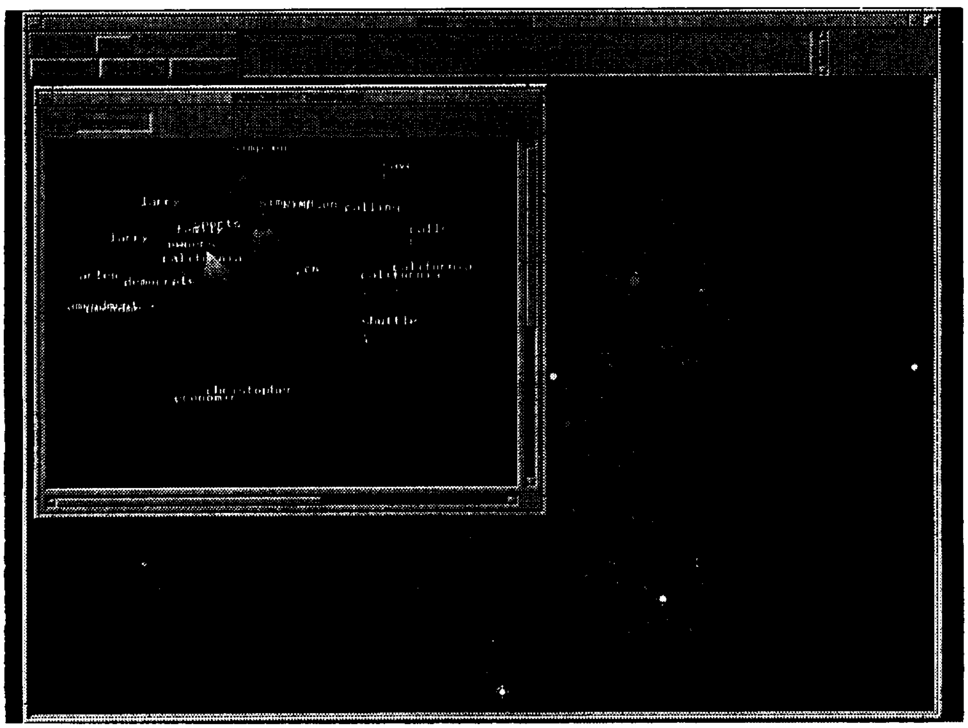


Figure 28

A video is played at this point to illustrate the dynamics of the analysis environment.

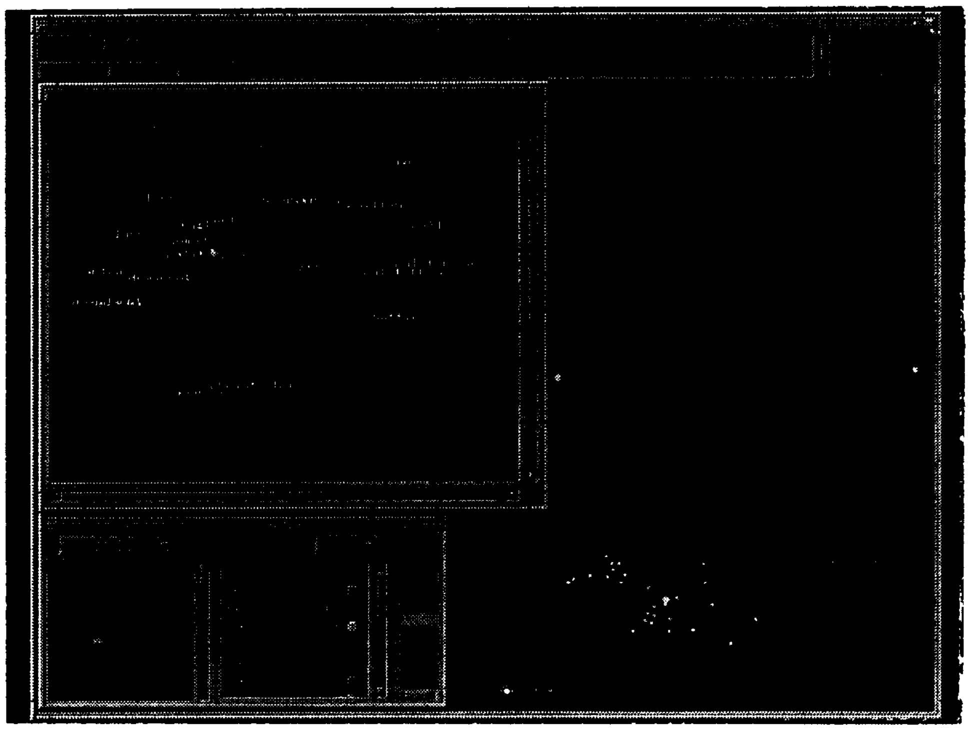


Figure 29

Also, some selected video segments were captured and can be directly played on systems that have support for video and audio.

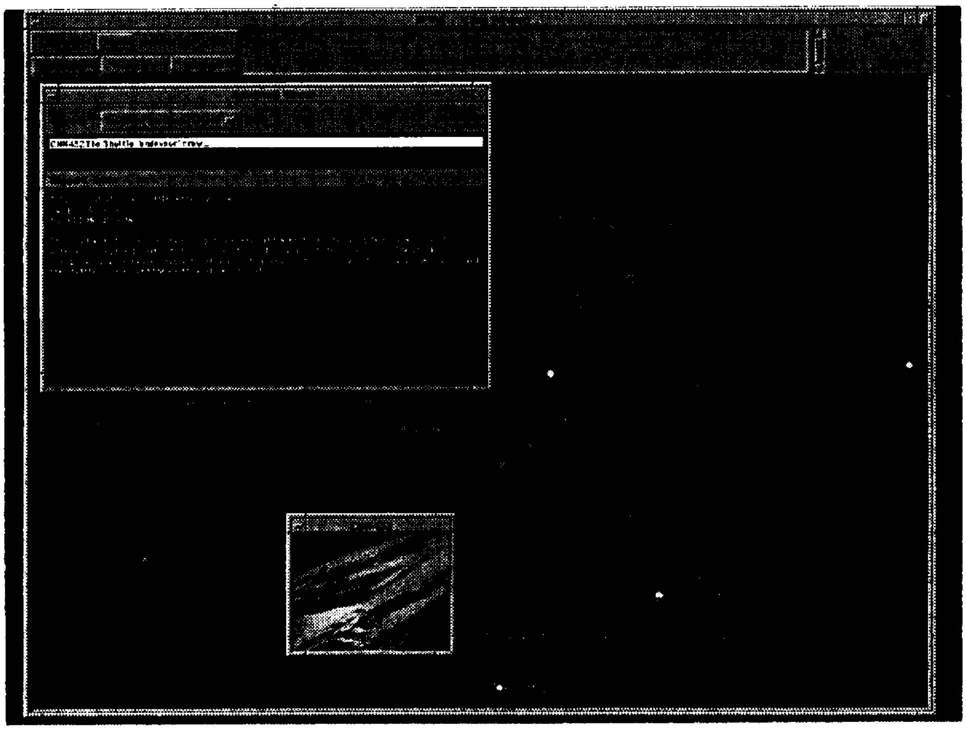


Figure 30

SPIRE™ DEVELOPMENT - KEY CONCEPTS

In summary, you have seen innovative Information Visualization that provides an analyst a new method for information analysis. This is based on a high-dimensional visual informational space that maps the information content into a spatially interactive analysis environment.

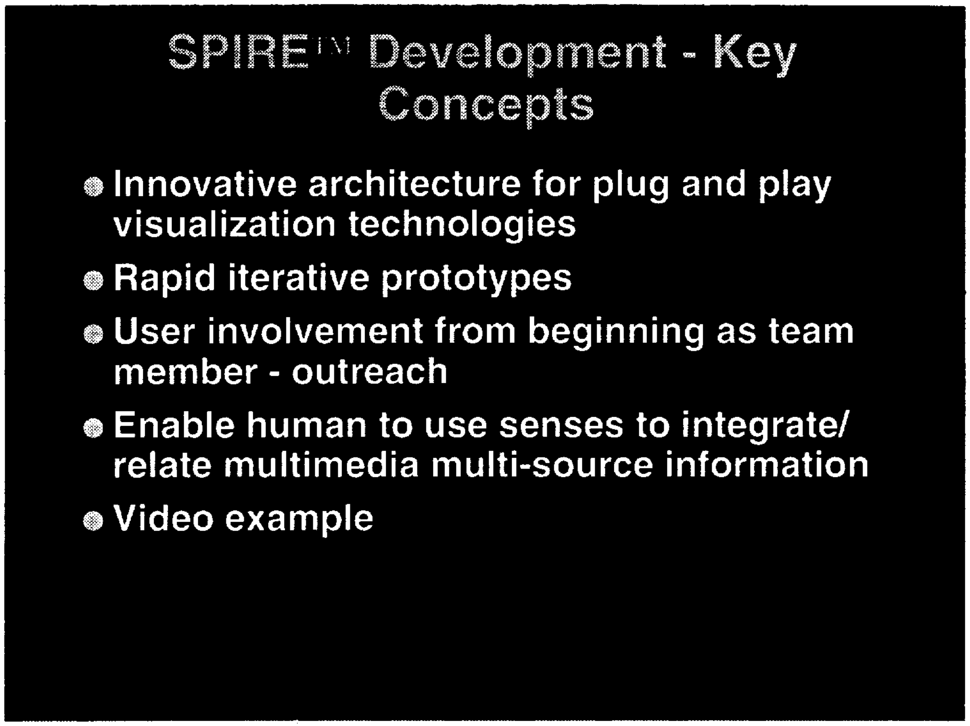


Figure 31

TAKE HOME MESSAGE FOR HCI AND COMPLEX INFORMATION SPACES

In conclusion, the key take home message is: add value by displaying all of the information by the use of high-dimensional analysis of the content spaces. The mind easily adapts to discovery of the process within these spaces.

To achieve this we need to consider all of the information spaces in selecting our fundable research tasks in this area.

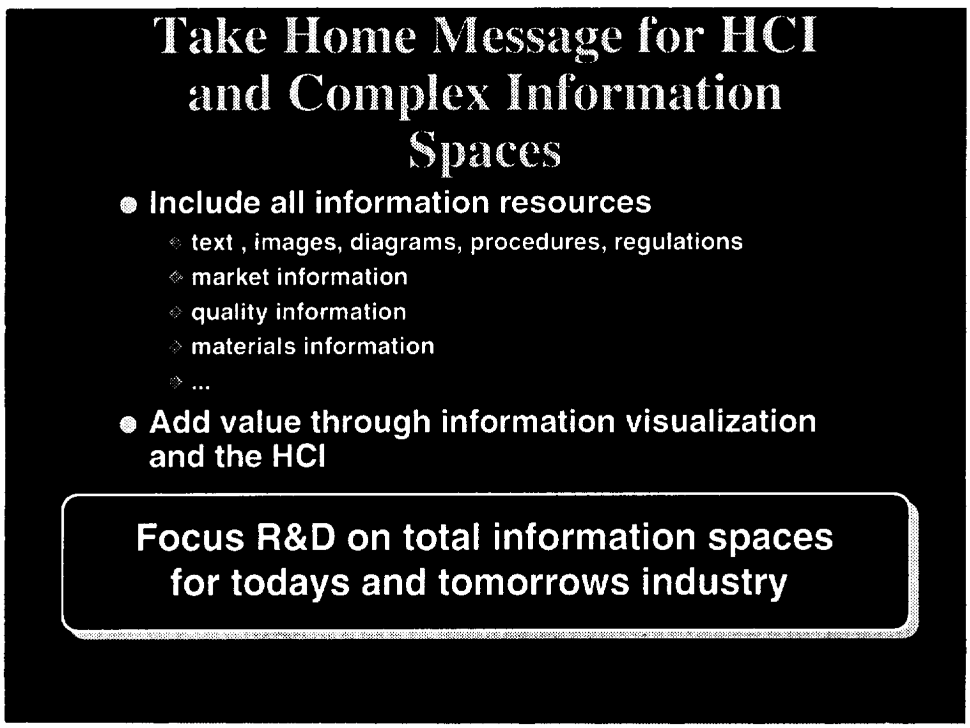


Figure 32