Pacific Northwest
NATIONAL LABORATORY

Starlight Research

Starlight research began in the early 1990's and represents the first attempt to marry a variety of different types of "conventional" (and novel) information visualization capabilities into a single, integrated information system capable of supporting a wide range of analytical functions. Further, Starlight visualization tools employ a common XML-based information model capable of effectively capturing multiple types of relationships that may exist among information of disparate kinds. Together, these features enable the concurrent visual analysis of a wide variety of information types. The result is a system capable of both accelerating and improving comprehension of the contents of large, complex information collections.

Motivation: Why visualize information?

Consider an arbitrary set of "information objects," for example, a collection of web pages or database records, or perhaps a group of related email messages. What makes such a collection useful and valuable? We argue that it is potentially valuable because it can be used to help solve problems and, further, that its value for problem solving lies in one or both of two places:

- Within individual items (i.e., taken in isolation)
- In the relationships among the items

Deriving value of the first sort is an *information retrieval* problem: strictly a matter of finding and examining the item or items that have a certain property. Deriving value of the second type is an information analysis problem. Human cognitive analysis is largely a matter of comparison: comparing various properties of items with one-another, and comparing such properties with prior knowledge. As the volume and complexity of information increases, however, human ability to make these kinds of comparisons mentally degrades rapidly. Visualization technologies can effectively reverse this trend by capturing relationships in a kind of external, graphical, "memory" where they can be more easily compared and evaluated.
Visualization is a potentially powerful tool for information analysis because it enables humans to make rapid, efficient, and effective comparisons.
Note: A good rule of thumb to use when evaluating visualization designs is to ask yourself two questions: 1) What information does this design let me compare?, and 2) How easy is it to make the comparison?

Making it Work: How is information visualized?

In practice, enabling "visual" analysis of information is a two-step process. First, relationships among information objects (as well as the information itself) must be explicitly captured in a computer-manipulable form. Once this is achieved, interactive graphical representations of the relationships can be generated for analytical purposes.

Relationships are captured in a digital construct generically referred to as an *information model*. A model intended to support information visualization should be comprehensive (i.e., support many different relationship types), flexible (in order to support many different information types), and, above all, human-oriented. By this, we mean that, ideally, the model will capture relationships in a form that mimics the way humans naturally relate information.

The Starlight Information Model is our attempt to effectively meet these criteria. The Starlight Model is comprehensive, capable of accommodating a wide variety of relationship types, including discrete property (i.e., field/value pair) co-occurrences, free-text similarity, temporal relationships, parent-child associations, network relationships, and spatial (e.g., geospatial) relationships. Because the model is designed to capture relationships among XML objects, it can flexibly accommodate the full range of information types expressible in XML (i.e., almost any type of digital information).

Finally, the model is human-oriented, explicitly designed for capturing and manipulating the types of relationships humans need to understand in order to solve complex, multifaceted, real-world problems.

Relationship Type:	General Similarity	Explicit Reference	Field/Value Co-occurrence	Parent/Child	Spatial	Temporal
Model Type:	*Vector-space*	*Network*	*Multidimensional Index*	*Hierarchical*	*Spatial*	*Ordinal Index*
Examples:	*Reports, articles, DB records*	*References & citations, hyperlinks*	*DB records, document metadata*	*File paths, taxonomies, IP addresses*	*Geolocations, CAD models*	*Event descriptions*

Components of the Starlight Information Model

Once relationships have been explicitly captured, Starlight can generate graphical representations of various aspects of the model that enable the underlying relationships to be visually interpreted. Importantly, Starlight visualizations are interoperable, enabling viewers to interactively move among multiple representations of the same information in order to uncover correlations that may span multiple relationship types.

For example, email messages can be related to one another in a number of different ways. There may be topological relationships among the senders and recipients. There may be conceptual similarities among the message contents, or temporal correlations among the messages.

Different email messages may even mention different places that are, in fact, physically near one-another: a spatial correlation. We are working to develop an information model capable of seamlessly accommodating all of these relationship types, and visualization tools to enable users to quickly understand the potentially complex interdependencies among them.

To illustrate the potential power of this approach, consider again an arbitrary collection of email messages. A Starlight user may choose to graphically depict such “email spaces” in any of a number of different ways, depending on the problem he or she is trying to solve at any given moment. An analyst may initially wish to view the collection as a network diagram in which the emails are portrayed as edges connecting nodes that represent senders and recipients. This method enables the viewer to identify important topological relationships among individuals based on “who sent what to whom.”

Once a particular subset of email had been identified based on its network topology, an analyst might switch to a "conceptual" representation of the same information that summarizes the concepts described in the items of interest. Following that, the user could switch the display to another alternate representation that spatially groups the items according to author or recipient.

In this way, even extremely complex and multifaceted relationships that exist in the collection can be quickly and easily characterized and assimilated.