

[« Return to Contents page for this issue](#)

LINKS TO PEOPLE AND SITES REFERENCED IN THIS ARTICLE

E-mail the article's author, [Susan Trulove](#)

Webpage for [Naren Ramakrishnan](#)

[Computer Science Department](#)

[Advanced Research Computing at Virginia Tech](#)

Webpage for [Richard Helm](#)

[Virginia Tech Center for Genomics](#)

[Biochemistry Department](#)

[Office of the Vice President for Research](#)

LINKS TO THE OTHER ARTICLES IN THIS ISSUE OF VIRGINIA TECH RESEARCH MAGAZINE

[Battle of the bugs:](#) Researchers fight a tiny, tree-killing insect

[Students chase storms](#) and get class credit for the adventure

[E-textiles:](#) wearable, washable, potentially life-saving computers

[Storytelling software](#) helps scientists make connections

Exploring the connection between [cognition and emotion development](#) in children

Slow road to recovery in [war-torn Bosnia](#)

Recognition and thanks for the contributions of [five of our faculty members](#)

Storytelling software helps scientists make connections

By **Susan Trulove**

We are all familiar with search engines that provide a list of hits based upon the terms we enter. Researchers in computer science and biochemistry at Virginia Tech are developing a search capability that they call **Storytelling** to discover connections between information that appears, at the onset, dissimilar. The algorithm detects a sequence of related events or relationships to create a chain of concepts between specified start and end points.

Imagine, for instance, asking for a connection between the U.S. senators from Virginia and the cast of the TV series "Friends." These concepts seem unconnected because no U.S. senator from Virginia has acted in "Friends," says **Naren Ramakrishnan**, associate professor of computer science, "but we would like to think of intermediate concepts that would help bridge the gap."

He outlines one possible "story" as follows: "U.S. senators from Virginia" overlap with "Filmmakers who specialize in wartime stories," which include "People who have won multiple People's Choice awards" and are connected to "Cast of 'Friends.'"

The overlap between the first two concepts is the junior senator from Virginia, Jim Webb, who has produced such movies as "Rules of Engagement." The overlap between the second and third concepts includes such filmmakers as Rob Reiner, who has directed, for instance, "A Few Good Men," and also has won many awards for his work. The overlaps between the last two concepts include "Friends" cast members, such as Jennifer Aniston. "This example is a manually constructed story. The goal of the Storytelling project at Virginia Tech is to automate the discovery of such sequences of logical connections," says Ramakrishnan.

"Storytelling can be viewed as a carefully argued process of removing and adding participants, not unlike a real story," he says. "Knowing exactly which objects must be displaced, and in what order, helps expose the mechanics of complex relationships."

Storytelling has not been used to expose the connection between a senator and a sitcom, but rather to find connections when applied to the most challenging tasks – beginning with discovering hidden relationships in the biological literature.

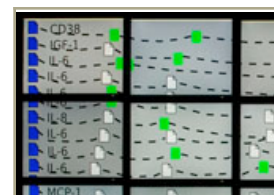
Ramakrishnan works with life scientists to create software for data



Deept Kumar, Naren Ramakrishnan, and Joseph Gresock demonstrate their StoryGrapher interface that helps organize multiple runs of Storytelling and visualize how storylines merge. Photo by John McCormick. View the [complete photo](#) at a larger size.



Kumar developed a ranking mechanism that determines which "leads" to follow in constructing a story and implemented it on System X, the 1100 node Apple Xserve supercomputer at Virginia Tech. Photo by John McCormick.



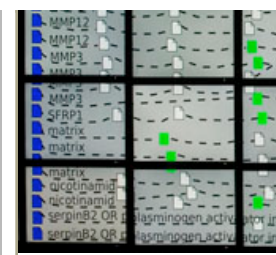
mining and information analysis. The aim is to help researchers make connections in the complex, burgeoning world of scientific discovery. “Every day, there are new research results reported in the literature and there are discoveries to be made by exploring connections,” he says.

“Our minds cannot correlate all available datasets efficiently and with any high degree of confidence without the aid of bioinformatics tools,” says Richard Helm, associate professor of biochemistry. “Attempting to find significant connections within the ocean of online information is daunting. For instance, a biologist desiring to stay abreast of literature on the budding yeast *Saccharomyces cerevisiae* must monitor yeast-specific journals (e.g., *Yeast*), general microbiology journals (e.g., *Eukaryotic Cell*), “methods” journals (e.g., *Nature Methods*), and even clinical and disease-oriented journals (e.g., *Cancer Research*). There would be published experiments in each of these venues that look at particular subsets of a biological process. The Storytelling algorithm helps connect publications by recognizing overlaps in content across them and drawing them together into a storyline. Evaluation of these stories can provide hypotheses that can be tested, potentially resulting in new insights into the role of a particular molecular event in the process you are interested in.”

Helm and colleague Malcolm Potts are studying the processes and strategies used by organisms to enter into and exit from a state of reduced metabolic activity, such as dormancy or suspended animation. A good example of an organism in “deep sleep” is the commercial dry active yeast that can be purchased in your local supermarket. Upon the addition of warm water and a bit of sugar, the microbes spring back to life and begin normal metabolic processes. “Uncovering the processes that allow organisms to enter and exit from metabolic arrest can potentially lead to the development of robust mammalian cell-based biosensors for detection of pathogens, as well as provide the ability to store cells, cell components, and vaccines without reliance upon low temperature storage,” says Helm.

“Human cells also exhibit states of metabolic arrest, termed quiescence and senescence. While both of these states have similar properties, cells entering into quiescence can exit the arrested state and enter into normal cell function. Senescent cells, however, are permanently arrested and cannot proliferate,” Helm says. “It is presently thought that the senescence program is a tumor suppressor mechanism. Exploration of the differences between quiescence and senescence can provide insights into aging processes as well as cancer mechanisms.”

To understand the role played by the Storytelling algorithm, consider the stresses imposed on yeast while undergoing desiccation and rehydration. The changes in cell volume as a result of the stress will cause variations in concentrations of molecules within the cells, and the cells will eventually become starved as nutrients will not be able to be taken up. Crowding of



Stories are mined by the Storytelling algorithm from input documents to an output document. View the [complete photo](#) and a more detailed caption.



Helm and laboratory specialist Tata Wiles review datasets that contain information on “master regulator” proteins that control gene expression. Photo by Josh Armstrong. View the [complete photo](#) and a more detailed caption.

When you ask a supercomputer to tell a story, you might not expect a creative outcome. But a group of Virginia Tech researchers are using System X, the university’s

molecules can lead to disruption of the cellular organization present in active cells. Such a reorganization needs to be coordinated so that exit from a dried state will return the cell to a fully functional state. Evaporation leads to temperature stresses during the drying process, and rehydration can affect cell membranes as they become rehydrated. Membrane reorganization processes is one key aspect of yeast survival and is the reason why one is directed to add warm water, rather than cold, to commercial active dry yeast.

Clearly the processes of desiccation and rehydration involve many variables acting at once. But laboratory-based investigations require minimizing variables in order to make conclusive statements about cell behavior.

“How can one develop a coherent understanding of the process when the published literature is based upon single variable stresses, such as temperature shock, starvation, or chemical stress? It was our need to look at a wide variety of stresses and apply them to yeast desiccation and rehydration that gave birth to the Storytelling algorithm,” says Helm.

The team used an early version of Storytelling to explore article abstracts from the U.S. National Library of Medicine PubMed database. Abstracts are short summaries of articles that do not have agreed-upon code or nomenclature. Sentences and paragraphs present thoughts from different people using different phrases and jargon, and who are not even necessarily thinking about the same problems. The input to Storytelling took the form of abstracts of 140,000 publications about yeast, which were modeled as sets of terms from the abstracts. The researchers then asked the Storytelling program to discover the relationship between a paper that describes gene expression changes during desiccation and rehydration and other papers in the abstract database. It was discovered that processes involving manipulation of sulfur compounds were central to survival. This new finding was published in the journal *Applied and Environmental Microbiology*.

This initial phase of Storytelling, driven by computer science Ph.D. student **Deept Kumar**, was based on modeling content similarity between abstracts. Kumar developed a ranking mechanism that determines which “leads” to follow in constructing a story and implemented it on System X, the 1100-node Apple Xserve supercomputer at Virginia Tech. Each “node” in the supercomputer is tasked with creating a potential set of connections, and the nodes exchange information among themselves to link papers and formulate the story. “Kumar’s implementation processes hundreds of thousands of papers and can work with up to 200 nodes simultaneously,” says Ramakrishnan.

The project’s newest member, computer science master’s degree student **Joseph Gresock**, generalized the algorithm to support the task of modeling cross-talk in the signaling networks between cells.

**supercomputer,
to test a new
search
program that
can make
connections
that tell the
stories of life.**



Ramakrishnan



Helm

Helm likens cross-talk to an individual making a decision after garnering input from several different people, all of whom had different things to say. “In cells, cross-talk is all too prevalent and can influence how they receive and respond to signals. A cell can be subjected to lots of stimuli, and it will respond to a given stimulus in a certain way. Given a different stimulus, it might respond in a completely different way or in a way that shares some overlap with some other stimulus. Crosstalk can happen when the pathways that carry signals from one stimulus meet another pathway that usually carries signals from another stimulus. Biologists would like to be able to follow, or model, these complex interactions,” says Helm.

In investigating molecular mechanisms underlying organism aging, Helm and colleagues conducted a laboratory experiment that demonstrated that adding the vitamin B compound nicotinamide to the growth medium of primary human fibroblasts extended the lifespan of the connective tissue cell.

To understand what other stimulus might cause similar or related results, Helm used Storytelling to see if there is any cross-talk going on. “The goal is to use the software to generate a hypothesis that we can test experimentally to understand this lifespan extension at the molecular level,” says Helm. “We are asking Storytelling to elucidate relationships between cellular inputs, such as nicotinamide, and outputs, or cell fate decisions, such as lifespan extension.”

“We model cross-talk by creating lots and lots of stories and seeing if these stories cross each other,” says Ramakrishnan. In this approach, repeated runs of Storytelling are organized, from abstracts discussing a given set of “input” molecules to abstracts discussing output molecules (such as poly-ADP ribose). The numerous stories discovered in this manner are then summarized to identify “novellas,” which are repeated, frequently re-used templates of connections. Gresock developed the StoryGrapher, an intuitive visual interface to organize such multiple runs of Storytelling and interactively navigate the results. This work provides further insight into how signaling cascades interact. Helm’s group is exploring several connections as a result.

“The holy grail of applying computing to biology is to understand a particular organism or process at a higher level than we are used to considering,” says Helm. “Storytelling is an important new tool in this endeavor.”

Virginia Tech Office of the Vice President for Research, 301 Burruss Hall (0244), Blacksburg, Virginia 24061

Phone: 540/231-6077 Fax: 540/231-4384 <http://www.research.vt.edu>

August 2007