Computers and Information Sciences
Informatics

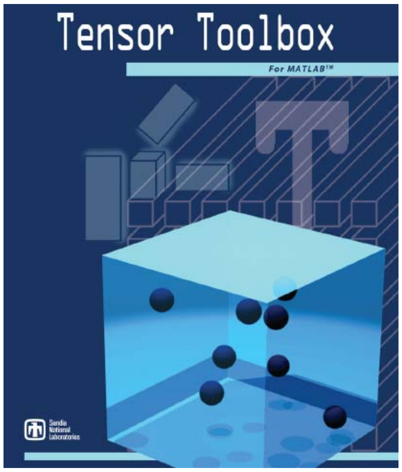

Figure 1: Tensor Toolbox software package

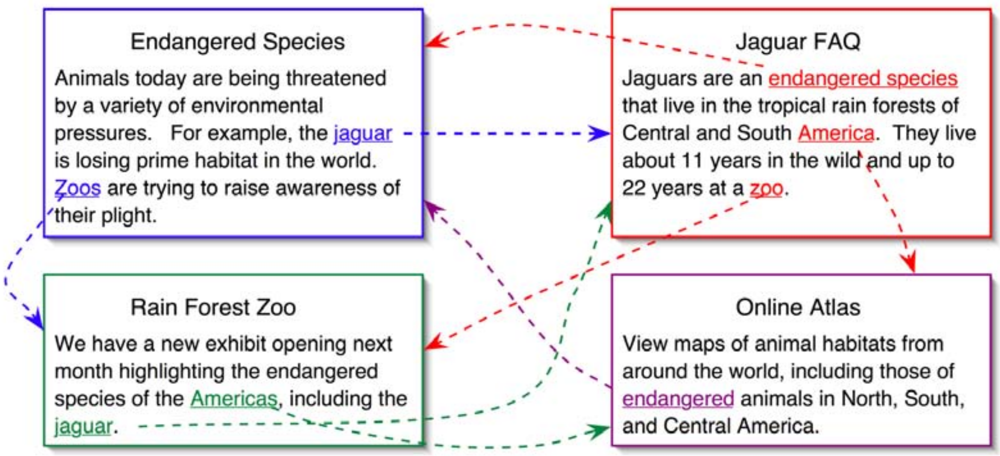

Figure 2: Four web pages with connecting links

Tensor Toolbox for MATLAB™

Sandia's Toolbox allows scientists to solve large, sparse tensor calculations on very large data sets.

Life is multi-dimensional, so it might be surprising to realize that nearly all calculations and simulations are done using two-dimensional mathematical arrays, better known as matrices. Even large-scale, three-dimensional engineering simulations, four-dimensional physics calculations, and multi-dimensional data analysis methods have been structured and optimized to work as two-dimensional matrix calculations. These computational approximations result in slower or less accurate calculations.

Sandia scientists and colleagues are at the forefront of new research in algorithms and software for applying multi-dimensional arrays, called tensors, to solve multi-dimensional problems that arise in data analysis, signals processing, image recognition and analysis, and other fields.

A major roadblock to the use of these multi-dimensional techniques was the absence of any software for large, sparse tensor calculations. Sparse tensors have a majority of entries that are zero. Only the non-zero entries are usually stored. Sandia scientists developed the Tensor Toolbox for MATLAB™ (Figure 1) to address this need. The free software integrates with MATLAB™, the matrix-based high-level language and interactive environment that enables users to perform computationally intensive tasks faster than with traditional programming languages. The Tensor Toolbox makes working with tensors in MATLAB™ as easy as working with matrices. The user need not worry about the low-level details to do complex, high-level operations, and the tool can handle very large problems such as sparse tensors the size of 10,000x10,000x10,000 with a half-million nonzero entries.

Sandia's Tensor Toolbox has enabled new and more accurate analyses in multiple application domains, particularly those involving large amounts of data ("data mining"). For example, several web pages can have connecting links (Figure 2). The links in those web pages can be analyzed in a graph form with labeled edges (Figure 3), and thence stored as a sparse tensor (Figure 4). This higher-order web link analysis allows for better automatic grouping and labeling of web pages through the TOPHITS algorithm (also developed at Sandia).

Another application is in bibliometric analysis using multiple linkages (authors, documents, terms), including understanding author-keyword trends. An example is

For more information:

Technical Contact:
Tamara G. Kolda, Ph.D.
925-294-4769
tgkolda@sandia.gov

Science Matters Contact:
Alan Burns, Ph.D
505-844-9642
aburns@sandia.gov

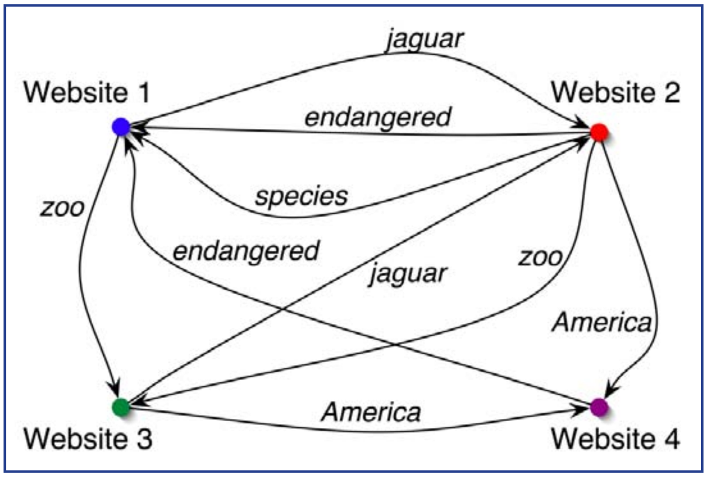

Figure 3: The linked web pages in Figure 2 are converted to a graph with labeled edges (stored as a tensor) and the link text is used in the analysis.

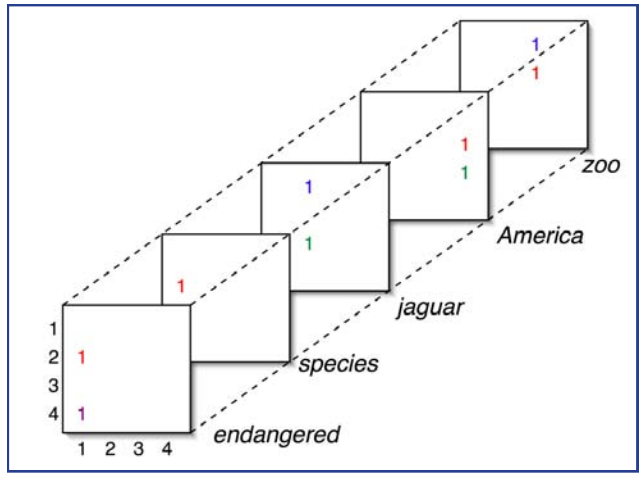

Figure 4: The resulting tensor representation of Figure 3 is extremely sparse. This tensor is decomposed using the Tensor Toolbox.

shown in Figure 5, where connections between cited authors and their publications are plotted in graph form. Similarly, national security applications of the Tensor Toolbox include temporal analysis of email exchanges, including automated discovery of conversation topics and sender/recipient roles over time.

Outside of Sandia, over one thousand registered users of the Tensor Toolbox have reported diverse applications including chatroom data analysis, continuum mechanics, online monitoring of network data, acoustic signal research, chemometrics, finite element computations, studies of bird migration, statistical computations, biochemical analysis, image classification, air traffic control studies, astronomy, models of tumor growth, character animation, computer vision, brain imaging, multidimensional economics, general relativity research, modeling optical systems, physics, multilayer absorption for photovoltaics, signal processing, computational differential geometry, neuro-fuzzy networks, and video analysis.

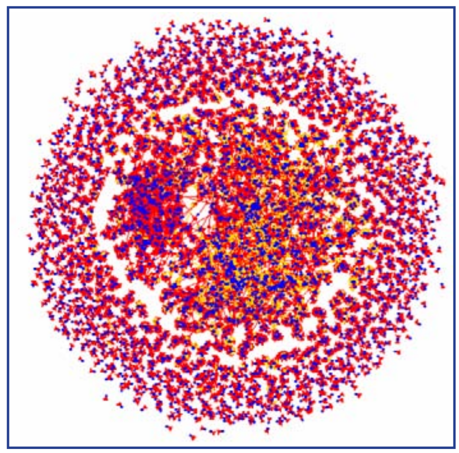

Figure 5: A graphical representation of 5000 papers (in blue), their authors (in red), and the citation and authorship connections between them.

Tensor Toolbox Web Site
http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/

References:

Brett W. Bader, Richard A. Harshman, and Tamara G. Kolda. **Temporal analysis of semantic graphs using ASALSAN**. In *ICDM 2007: Proceedings of the 7th IEEE International Conference on Data Mining*, pages 33–42, October 2007.

Brett W. Bader and Tamara G. Kolda. **Efficient MATLAB computations with sparse and factored tensors**. *SIAM Journal on Scientific Computing*, July 2007.

Brett W. Bader and Tamara G. Kolda. **Algorithm 862: MATLAB tensor classes for fast algorithm prototyping**. *ACM Transactions on Mathematical Software*, 32(4):635–653, December 2006.

Peter A. Chew, Brett W. Bader, Tamara G. Kolda, and Ahmed Abdelali. **Cross-language information retrieval using PARAFAC2.** In *KDD '07: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 143–152. ACM Press, 2007.

Daniel M. Dunlavy, Tamara G. Kolda, and W. Philip Kegelmeyer. **Multilinear algebra for analyzing data with multiple linkages.** Technical Report SAND2006-2079, Sandia National Laboratories, April 2006.

Tamara G. Kolda and Brett W. Bader. **Tensor decompositions and applications.** Technical Report SAND2007-6702, Sandia National Laboratories, November 2007.

Tamara Kolda and Brett Bader. **The TOPHITS model for higher-order web link analysis.** In *Workshop on Link Analysis, Counterterrorism and Security*, 2006.

Tamara G. Kolda, Brett W. Bader, and Joseph P. Kenny. **Higher-order web link analysis using multilinear algebra.** In *ICDM 2005: Proceedings of the 5th IEEE International Conference on Data Mining*, pages 242–249, November 2005.

Sandia National Laboratories

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND 2008-0486P

National Nuclear Security Administration

01/2008