Sept. 17, 2008

Visual Analytics with Jigsaw - the VAST 2007 Contest

John Stasko, Carsten Görg, Zhicheng Liu

Information Interfaces Research Group
School of Interactive Computing
Georgia Institute of Technology

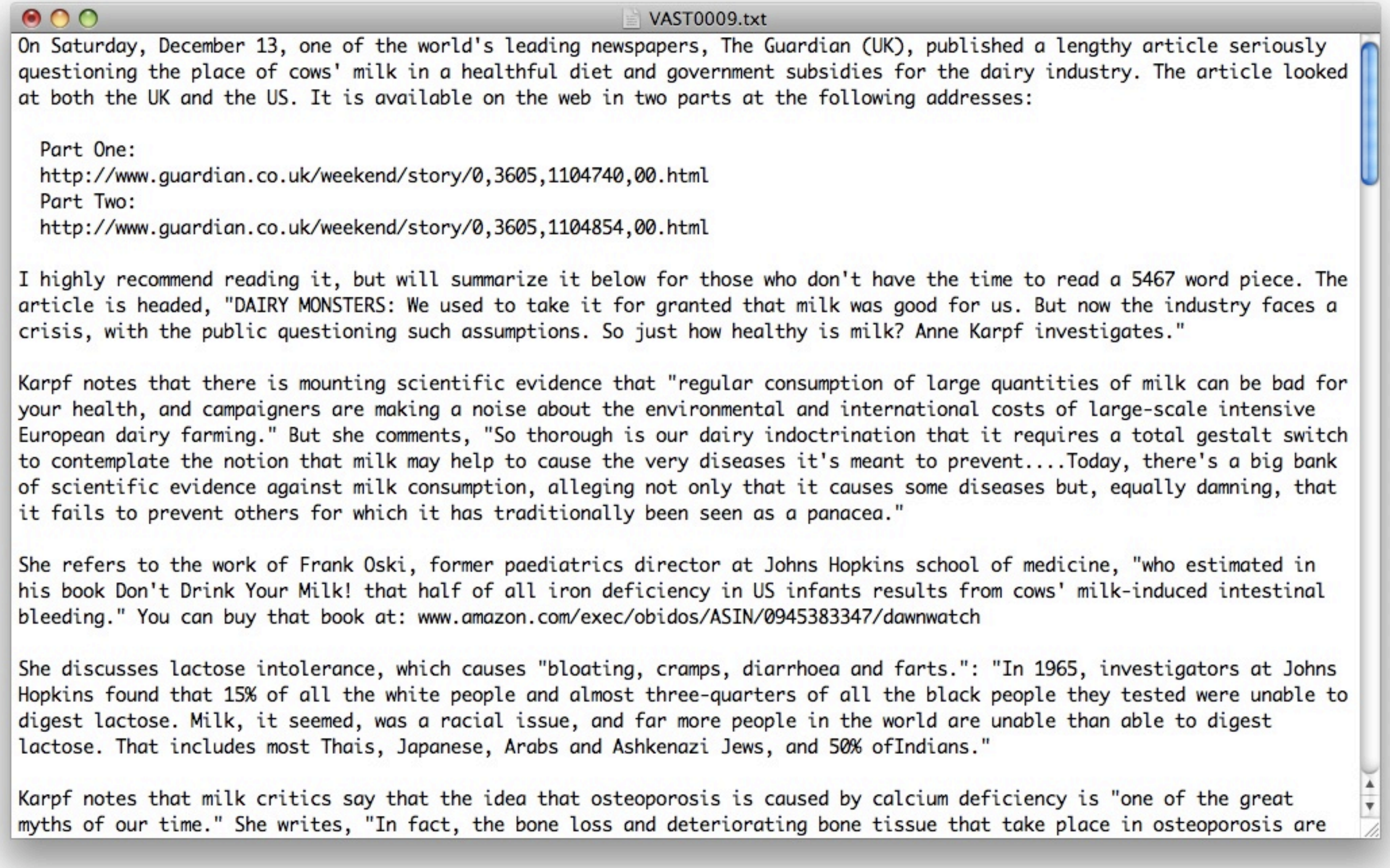
SouthEast RVAC

Georgia Tech College of Computing
School of Interactive Computing

The VAST 2007 Contest

- The dataset
  - news stories, blog entries, some multimedia materials
- The goal
  - investigate a major law enforcement/ counter-terrorism scenario, form hypotheses, and collect supporting evidence.

News story #1



On Saturday, December 13, one of the world's leading newspapers, The Guardian (UK), published a lengthy article seriously questioning the place of cows' milk in a healthful diet and government subsidies for the dairy industry. The article looked at both the UK and the US. It is available on the web in two parts at the following addresses:

Part One:
http://www.guardian.co.uk/weekend/story/0,3605,1104740,00.html
Part Two:
http://www.guardian.co.uk/weekend/story/0,3605,1104854,00.html

I highly recommend reading it, but will summarize it below for those who don't have the time to read a 5467 word piece. The article is headed, "DAIRY MONSTERS: We used to take it for granted that milk was good for us. But now the industry faces a crisis, with the public questioning such assumptions. So just how healthy is milk? Anne Karpf investigates."

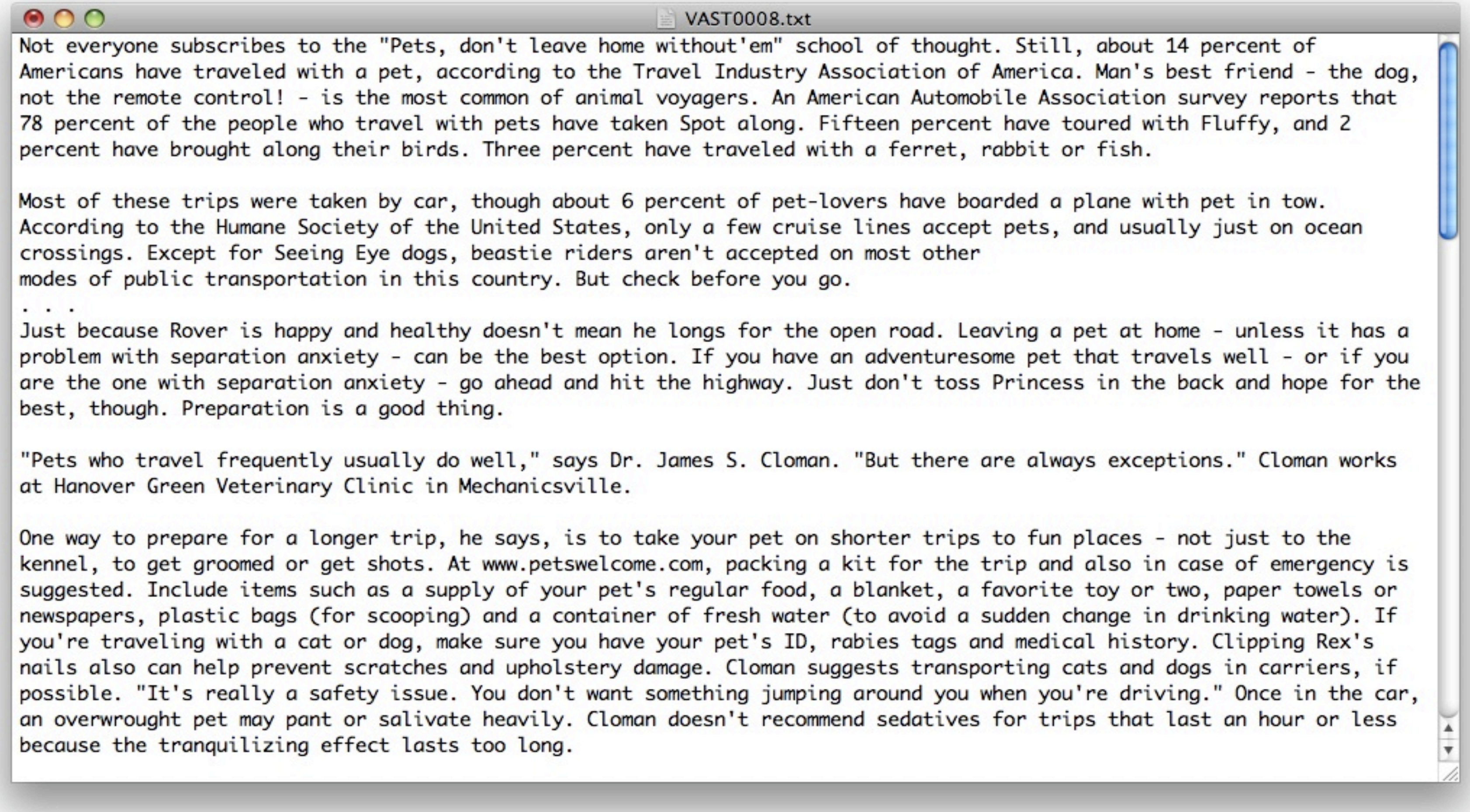
Karpf notes that there is mounting scientific evidence that "regular consumption of large quantities of milk can be bad for your health, and campaigners are making a noise about the environmental and international costs of large-scale intensive European dairy farming." But she comments, "So thorough is our dairy indoctrination that it requires a total gestalt switch to contemplate the notion that milk may help to cause the very diseases it's meant to prevent....Today, there's a big bank of scientific evidence against milk consumption, alleging not only that it causes some diseases but, equally damning, that it fails to prevent others for which it has traditionally been seen as a panacea."

She refers to the work of Frank Oski, former paediatrics director at Johns Hopkins school of medicine, "who estimated in his book Don't Drink Your Milk! that half of all iron deficiency in US infants results from cows' milk-induced intestinal bleeding." You can buy that book at: www.amazon.com/exec/obidos/ASIN/0945383347/dawnwatch

She discusses lactose intolerance, which causes "bloating, cramps, diarrhoea and farts.": "In 1965, investigators at Johns Hopkins found that 15% of all the white people and almost three-quarters of all the black people they tested were unable to digest lactose. Milk, it seemed, was a racial issue, and far more people in the world are unable than able to digest lactose. That includes most Thais, Japanese, Arabs and Ashkenazi Jews, and 50% ofIndians."

Karpf notes that milk critics say that the idea that osteoporosis is caused by calcium deficiency is "one of the great myths of our time." She writes, "In fact, the bone loss and deteriorating bone tissue that take place in osteoporosis are

News story #2



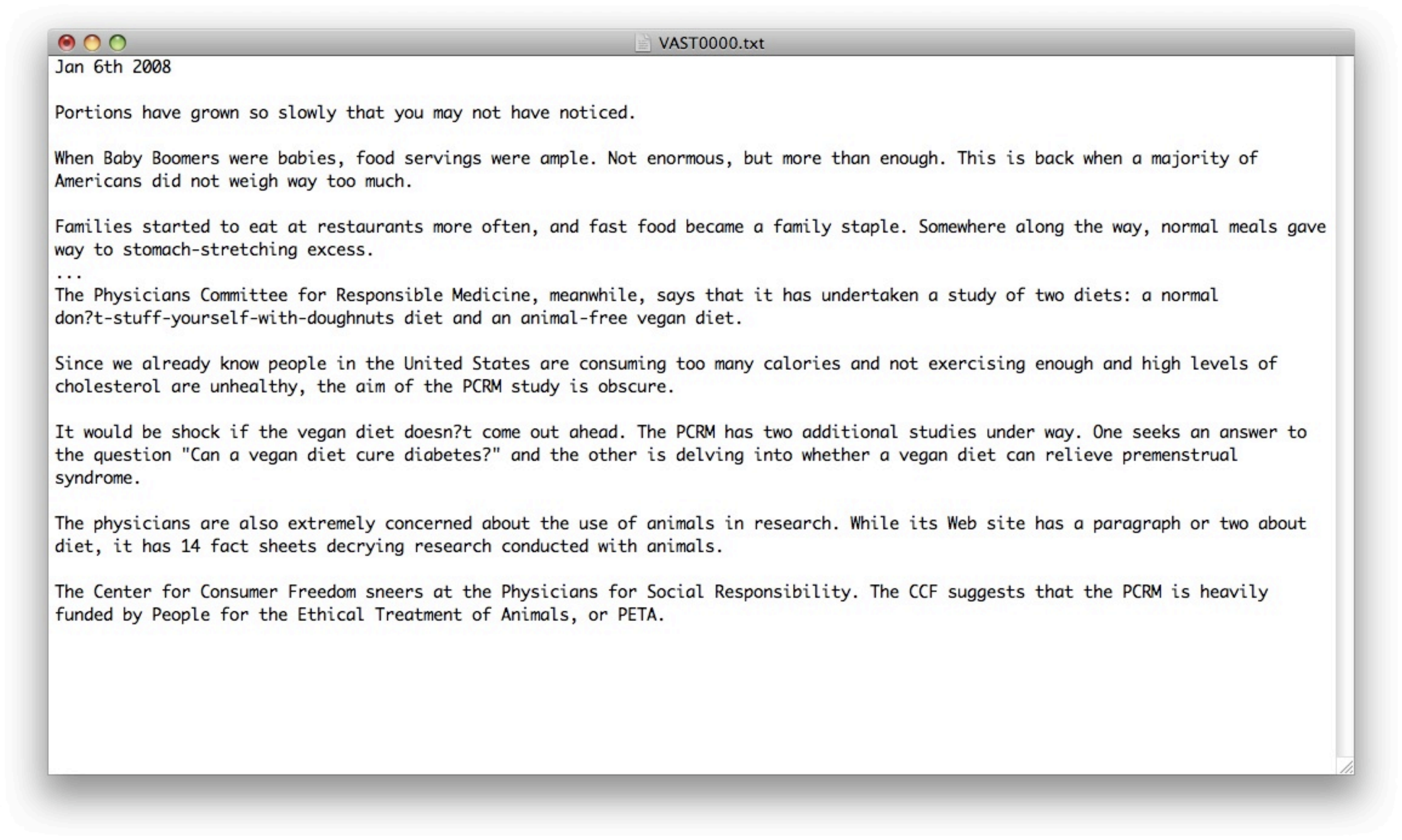
Not everyone subscribes to the "Pets, don't leave home without'em" school of thought. Still, about 14 percent of Americans have traveled with a pet, according to the Travel Industry Association of America. Man's best friend - the dog, not the remote control! - is the most common of animal voyagers. An American Automobile Association survey reports that 78 percent of the people who travel with pets have taken Spot along. Fifteen percent have toured with Fluffy, and 2 percent have brought along their birds. Three percent have traveled with a ferret, rabbit or fish.

Most of these trips were taken by car, though about 6 percent of pet-lovers have boarded a plane with pet in tow. According to the Humane Society of the United States, only a few cruise lines accept pets, and usually just on ocean crossings. Except for Seeing Eye dogs, beastie riders aren't accepted on most other
modes of public transportation in this country. But check before you go.
. . .
Just because Rover is happy and healthy doesn't mean he longs for the open road. Leaving a pet at home - unless it has a problem with separation anxiety - can be the best option. If you have an adventuresome pet that travels well - or if you are the one with separation anxiety - go ahead and hit the highway. Just don't toss Princess in the back and hope for the best, though. Preparation is a good thing.

"Pets who travel frequently usually do well," says Dr. James S. Cloman. "But there are always exceptions." Cloman works at Hanover Green Veterinary Clinic in Mechanicsville.

One way to prepare for a longer trip, he says, is to take your pet on shorter trips to fun places - not just to the kennel, to get groomed or get shots. At www.petswelcome.com, packing a kit for the trip and also in case of emergency is suggested. Include items such as a supply of your pet's regular food, a blanket, a favorite toy or two, paper towels or newspapers, plastic bags (for scooping) and a container of fresh water (to avoid a sudden change in drinking water). If you're traveling with a cat or dog, make sure you have your pet's ID, rabies tags and medical history. Clipping Rex's nails also can help prevent scratches and upholstery damage. Cloman suggests transporting cats and dogs in carriers, if possible. "It's really a safety issue. You don't want something jumping around you when you're driving." Once in the car, an overwrought pet may pant or salivate heavily. Cloman doesn't recommend sedatives for trips that last an hour or less because the tranquilizing effect lasts too long.

News story #3



Jan 6th 2008

Portions have grown so slowly that you may not have noticed.

When Baby Boomers were babies, food servings were ample. Not enormous, but more than enough. This is back when a majority of Americans did not weigh way too much.

Families started to eat at restaurants more often, and fast food became a family staple. Somewhere along the way, normal meals gave way to stomach-stretching excess.

...

The Physicians Committee for Responsible Medicine, meanwhile, says that it has undertaken a study of two diets: a normal don?t-stuff-yourself-with-doughnuts diet and an animal-free vegan diet.

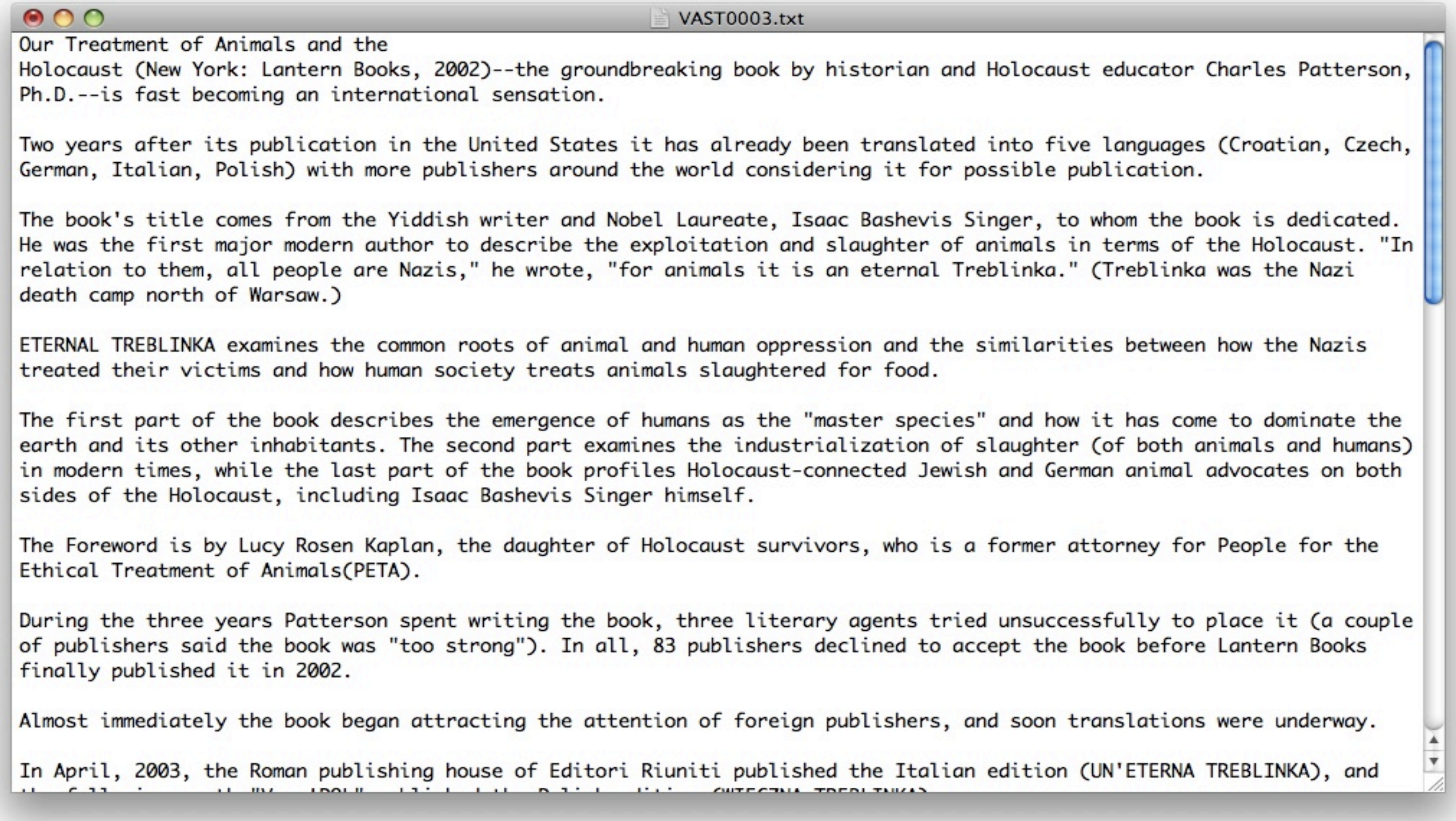
Since we already know people in the United States are consuming too many calories and not exercising enough and high levels of cholesterol are unhealthy, the aim of the PCRM study is obscure.

It would be shock if the vegan diet doesn?t come out ahead. The PCRM has two additional studies under way. One seeks an answer to the question "Can a vegan diet cure diabetes?" and the other is delving into whether a vegan diet can relieve premenstrual syndrome.

The physicians are also extremely concerned about the use of animals in research. While its Web site has a paragraph or two about diet, it has 14 fact sheets decrying research conducted with animals.

The Center for Consumer Freedom sneers at the Physicians for Social Responsibility. The CCF suggests that the PCRM is heavily funded by People for the Ethical Treatment of Animals, or PETA.

News story #4



Our Treatment of Animals and the
Holocaust (New York: Lantern Books, 2002)--the groundbreaking book by historian and Holocaust educator Charles Patterson, Ph.D.--is fast becoming an international sensation.

Two years after its publication in the United States it has already been translated into five languages (Croatian, Czech, German, Italian, Polish) with more publishers around the world considering it for possible publication.

The book's title comes from the Yiddish writer and Nobel Laureate, Isaac Bashevis Singer, to whom the book is dedicated. He was the first major modern author to describe the exploitation and slaughter of animals in terms of the Holocaust. "In relation to them, all people are Nazis," he wrote, "for animals it is an eternal Treblinka." (Treblinka was the Nazi death camp north of Warsaw.)

ETERNAL TREBLINKA examines the common roots of animal and human oppression and the similarities between how the Nazis treated their victims and how human society treats animals slaughtered for food.

The first part of the book describes the emergence of humans as the "master species" and how it has come to dominate the earth and its other inhabitants. The second part examines the industrialization of slaughter (of both animals and humans) in modern times, while the last part of the book profiles Holocaust-connected Jewish and German animal advocates on both sides of the Holocaust, including Isaac Bashevis Singer himself.

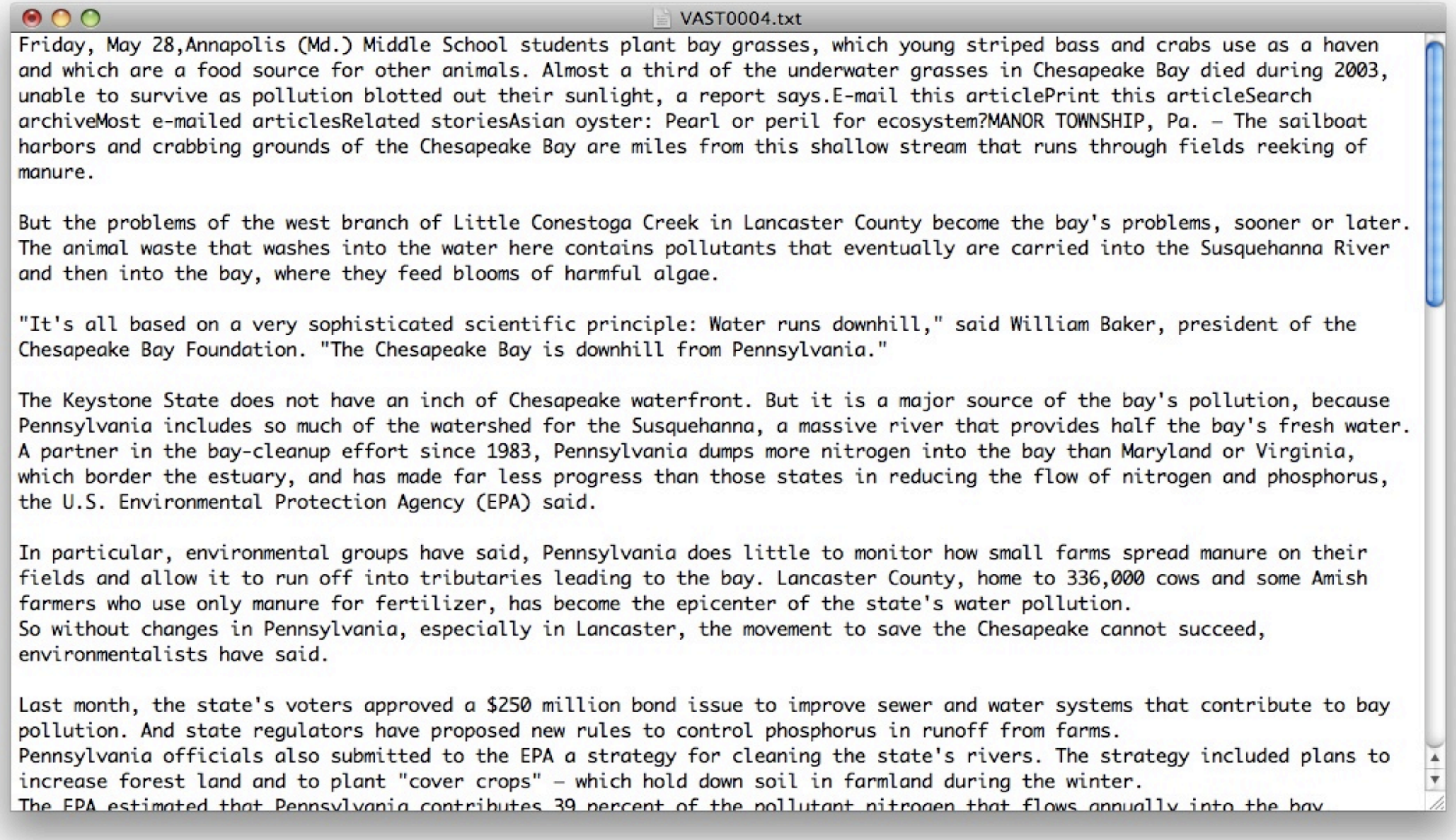
The Foreword is by Lucy Rosen Kaplan, the daughter of Holocaust survivors, who is a former attorney for People for the Ethical Treatment of Animals(PETA).

During the three years Patterson spent writing the book, three literary agents tried unsuccessfully to place it (a couple of publishers said the book was "too strong"). In all, 83 publishers declined to accept the book before Lantern Books finally published it in 2002.

Almost immediately the book began attracting the attention of foreign publishers, and soon translations were underway.

In April, 2003, the Roman publishing house of Editori Riuniti published the Italian edition (UN'ETERNA TREBLINKA), and

News story #5



Friday, May 28,Annapolis (Md.) Middle School students plant bay grasses, which young striped bass and crabs use as a haven and which are a food source for other animals. Almost a third of the underwater grasses in Chesapeake Bay died during 2003, unable to survive as pollution blotted out their sunlight, a report says.E-mail this articlePrint this articleSearch archiveMost e-mailed articlesRelated storiesAsian oyster: Pearl or peril for ecosystem?MANOR TOWNSHIP, Pa. – The sailboat harbors and crabbing grounds of the Chesapeake Bay are miles from this shallow stream that runs through fields reeking of manure.

But the problems of the west branch of Little Conestoga Creek in Lancaster County become the bay's problems, sooner or later. The animal waste that washes into the water here contains pollutants that eventually are carried into the Susquehanna River and then into the bay, where they feed blooms of harmful algae.

"It's all based on a very sophisticated scientific principle: Water runs downhill," said William Baker, president of the Chesapeake Bay Foundation. "The Chesapeake Bay is downhill from Pennsylvania."

The Keystone State does not have an inch of Chesapeake waterfront. But it is a major source of the bay's pollution, because Pennsylvania includes so much of the watershed for the Susquehanna, a massive river that provides half the bay's fresh water. A partner in the bay-cleanup effort since 1983, Pennsylvania dumps more nitrogen into the bay than Maryland or Virginia, which border the estuary, and has made far less progress than those states in reducing the flow of nitrogen and phosphorus, the U.S. Environmental Protection Agency (EPA) said.

In particular, environmental groups have said, Pennsylvania does little to monitor how small farms spread manure on their fields and allow it to run off into tributaries leading to the bay. Lancaster County, home to 336,000 cows and some Amish farmers who use only manure for fertilizer, has become the epicenter of the state's water pollution.
So without changes in Pennsylvania, especially in Lancaster, the movement to save the Chesapeake cannot succeed, environmentalists have said.

Last month, the state's voters approved a $250 million bond issue to improve sewer and water systems that contribute to bay pollution. And state regulators have proposed new rules to control phosphorus in runoff from farms.
Pennsylvania officials also submitted to the EPA a strategy for cleaning the state's rivers. The strategy included plans to increase forest land and to plant "cover crops" – which hold down soil in farmland during the winter.
The EPA estimated that Pennsylvania contributes 39 percent of the pollutant nitrogen that flows annually into the bay

Difficulty

- How about 1500 of them?
- Many possible threads, lots of details

Jigsaw

- Visualizes documents, entities and connections between them
- Human-centered exploration integrated with (a little) computational analysis

Investigative Focus

- Challenge: Jigsaw does not have capabilities for finding themes or concepts in a document collection
- Instead, Jigsaw acts as a visual index that can reveal connections between entities or reports once we have a viable lead

Contest Strategy

- Four people on the team, each skimmed through 350+ reports to get familiarized with the themes
- Jotted down notes about potential people, organizations or events to study further.

Initial Exploration using Jigsaw

- Generate XML file to feed data into Jigsaw
  - contest data set pre-processed with entities identified
- Looking for connection between entities across document collection
  - no definitive lead after 6 hours, but some interesting connections showing up

Problems with Automated Entity Extraction

- The same entity identified in some reports is not identified in others
- Entity not identified at all

Update the Entity Information

- Scan all documents to identify missed entities
- Remove false positives
- Filter entities appearing in only one report

Demo

More Information

- http://www-static.cc.gatech.edu/gvu/ii/jigsaw/vast07-contest/
  - actual contest entry with our solution, description of the analytical process and video

Lessons Learnt

- Reading reports is crucial
- Needs lots of screen estate / pixels

System Improvement

- Need stable entity processing capability
- Support user-driven entity update
- New visualizations from different perspectives
  - timeline view not very useful, a calendar representation might be better
  - report cluster
  - shoebox
  - new features to the document view

Current State

- Adding entity extracting engines
  - GATE vs. Lingpipe
- Re-designing the system architecture
  - entity hierarchy, attributes, aliases
  - memory / scaling issues
- Interface design choices

Future Plans

- Additional computational analysis
- Collaborative (synchronous) version
- Represent reliability and uncertainty
- Input of structured data and web data
- Representing the investigation history

Acknowledgments

- Earlier work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC

- Also supported by NSF IIS-0414667

This material is supported in part by:
National Science Foundation
Grant #0414667