VisWeek 2010
VIS • INFOVIS • VAST
24 - 29 OCTOBER, 2010 SALT LAKE CITY, UTAH, USA

Data Samples for the VAST Challenge 2010

(Posted March 14)

Data provided in these challenges are synthetic and should not be construed as involving any real people or events.� There may be similarities to real events you may have heard of, but the information you need to puzzle out the challenges are contained in the dataset provided.

The VAST 2010 Challenge features a rather strange tale in the general topic areas of arms dealing and public health.� You will be given three datasets each representing a mini-challenge, which can be integrated to form an overall picture of what is transpiring with a certain set of people of interest.�

Mini-Challenge 1:

The first dataset focuses on some suspected arms deals that have multi-national implications.� You will be provided several different types of text reports from �US Government Intelligence Sources�, newspaper reports, web sites and blogs, and so on.� The following is an example of the data that will be provided in this mini-challenge:

10. US GOVERNMENT TELEPHONE INTERCEPT: Translated from Russian, from an Internet Caf� in Moscow to a pre-paid cell phone in Yemen, 4 February 2009.

    The caller says, �Your brother is dead, but I will be like a brother to you.� The person receiving the call says in Russian, but with an Arabic accent, �I am truly blessed for your friendship.� I will need your support to continue to build the family farm. Shall we plan a family reunion?�� The caller says, �Soon, my brother, very soon.�

Mini-Challenge 2:

The second dataset looks at tracking and characterizing the outbreak of a disease in disparate locations.� Health officials have pooled their data for these locations in the hopes of learning more about the disease and its causes.� We will provide hospital admittance and records of death for various cities, and participants will be asked to generate visualizations of the course of the disease over the time span of the data.� The following is an example of the data that will be provided:

Hospital Admittance:

USER_WARNING	DATE	GENDER	AGE	SYNDROME	ID
SYNTHETIC_DATA	4/15/2005	M	53	LEFT KNEE PAIN	1
SYNTHETIC_DATA	4/15/2005	F	55	HEAD ACHE	2
SYNTHETIC_DATA	4/15/2005	F	57	NAUSEA, VOMITING	3
SYNTHETIC_DATA	4/15/2005	F	40	R ANKLE INJ	4
SYNTHETIC_DATA	4/15/2005	F	55	STOMACH CRAMPS	5
SYNTHETIC_DATA	4/15/2005	F	81	TREMORS	6
SYNTHETIC_DATA	4/15/2005	F	45	LOWER ABDOMINAL PAIN	7
SYNTHETIC_DATA	4/15/2005	M	28	SKIN RASHABCESS	8
SYNTHETIC_DATA	4/15/2005	M	38	LEFT KNEE PAIN	9
SYNTHETIC_DATA	4/15/2005	F	40	LACERATION TO FOREHEAD	10

Patient Death Records:

USER_WARNING	DATE	ID
SYNTHETIC_DATA	4/15/2005	2
SYNTHETIC_DATA	4/15/2005	3
SYNTHETIC_DATA	4/15/2005	23
SYNTHETIC_DATA	4/15/2005	32
SYNTHETIC_DATA	4/15/2005	101

Mini-Challenge 3:

This challenge focuses on medical information collected from some hospital patients that is relevant to the overall scenario.� For this mini-challenge, you will be asked to analyze some genetic data and provide visualizations of sequence variations and their evolution.

The data consists of multiple text files containing genetic sequences collected from viral mutants present in human blood samples.

Each sequence is composed of hundreds to several thousands of bases. Each base is encoded as A, T, C, or G.

A genetic sequence file looks like this:

```
>Sequence24
ATGGATTCCAACACTGTGTCAAGTTTCCAGGACATACTATTGAGGATGTCAAAAATGCAATTGGGGTCCTATGGTTCATACATGTTGGAAAGGGAACTG

>Sequence32
```

ATGAGTAATGAGAATGGGGGACCTCCACTTACTCCAAAACAGAAACGGAAAATGGCGAGAACAGCTAGGTGTATCAGCGGATCCACTGGCATCACTGC