

IEEE VAST Challenge 2011

Discussion blog

[Home](#)
[Download Dataset](#)
[Task Description](#)
[Criteria for Judging](#)
[Guidelines](#)
[How to Submit?](#)
[Answer Forms](#)
[Results](#)
[Discussion Blog](#)
[History of Changes](#)

Questions about the challenge

If you have any questions regarding the challenge and its rules please enter your questions in the comment form at the bottom of the page.

Hey Kevin,

The answer forms are now available on the Answer Forms page.

Posted by [uml](#) on 06/17 at 10:59 AM

When will the answer forms be available? I see the "answer forms" page is ready, but there are no links to the forms.

Posted by [Kevin](#) on 06/16 at 01:16 PM

Aram,

If possible, please use the scale indicated in the Readme.

Posted by [uml](#) on 05/27 at 11:27 AM

Another distances question:

Instead of computing great-arcs, I've converted the coordinates into a cartesian model (assuming a sphere with 6371 km radius, the average earth radius) and calculated the euclidean distance between the points. This corrects errors slightly, but it's still off by about 10%. Is there any more light you can shed on this?

Posted by [Aram](#) on 05/27 at 11:01 AM

Aram,

These are straight distances and do not account for curvature of the earth.

Posted by [uml](#) on 05/26 at 06:06 PM

I tried to post before and apologize for any duplicates.

In the MC1 Readme, the area for Vastopolis is bounded by 42.1609 N, 93.1923 W and 42.3017 N, 93.5673 W, with the distance between those points given as 17 miles East-West and 8.647 miles North-South. I have tried to verify these distances based on a spherical model and the WGS84 (more accurate Earth model), and I get distances that about 10-12% greater than the ones given. What was the method used to derive these distances?

Posted by [Aram](#) on 05/26 at 05:23 PM

Casey,

It is possible there are errors in the dataset. Handle them as you see fit and focus on key pieces of information.

Posted by [uml](#) on 05/19 at 09:20 AM

A large number of the text files for MC #3 seem to have problems. For example, when the original news articles were

scraped of place names, some of them weren't replaced, leaving a blank.
Example: "The event occurred at ."

In other cases, the replacement names and/or genders for people aren't consistent in the same article. Also, while the names of certain events were changed to be fictional in some places, their actual real-life names were not changed in others.

I know a certain amount of data is going to be messy as part of the challenge, but these seem like awfully big errors for what are supposed to be news articles.

Can I assume these are intentional, or is there a larger issue with the data files?

Posted by [Casey Canfield](#) on 05/18 at 11:10 AM

Ahmed,

We will be providing instructions on how to submit to the competition shortly.

Posted by [UML](#) on 05/16 at 02:12 PM

How and when should we register in the competition?

Posted by [Ahmed Bahey](#) on 05/14 at 03:42 PM

Hi Erik,

As mentioned in a prior posting, we suspect that the collected data contains errors. How to handle them is up to you.

Posted by [UML](#) on 04/21 at 02:40 PM

Hi,

In the MC1 Microblogs.csv file, there seems to be a dozen of entries with ill-formatted timestamps, e.g.:

5592,5/16/2011 Fire-,42.22243 93.3712,The skill to win is important but the will to win is vital. - Joe Paterno

(instead of 'Fire-' there should be a point of time)

I suppose it is OK to remove those lines? They don't seem to contain very valuable information (although Joe Paterno has a point



Posted by [Erik](#) on 04/13 at 08:27 AM

Hi Ranieri,

As mentioned in an earlier posting, the longitudes are specified as West, so they can be represented as negative numbers. The N and W letters are stripped in the table.

Posted by [UML](#) on 04/12 at 02:28 PM

Hi, at the "MC1_Dataset_README" file, the picture of Vastopolis was shown in between Lat,Long coordinates at North (N) and West (W). By convention, we could hide the letters, by using "+" (plus sign) for N and E, and "-" (minus sign) for S and W. So, if we apply this convention to the Vastopolis picture, the coordinates must be 42.3017, -93.5673, for example. But this is inconsistent with the column Location of the table "Microblogs", which shows only positive values for the longitude, and so makes us understand that the messages were posted at N and E. Can you explain me, please?

Posted by [Ranieri](#) on 04/12 at 12:59 PM

Gennady,

We suspect that the collected data contains errors likely generated from data cleaning or data transfer, or GPS and time accuracy transcriptions, and that these need to be dealt with appropriately.

Posted by [UML](#) on 04/12 at 02:58 AM

Concerning Mini Challenge 1.

Just noticed, that sometimes there are pairs of records with same ID and time, but different locations.

Just one example (out of 100,000 such records):

39,5/1/2011 22:42,42.22243 93.33845,Die tuinman met zijn aardappels en vlecht moet dat echt in Zomergasten? Een gemiddelde aflevering van de Pfaffs is spannender.

39,5/1/2011 22:42,42.18262 93.29851,the ball just hit 6 inches foul. and Tim McCarver was begging the ump basically to call it fair. IS WRONG WITH YOU

How to treat such records?

Obviously, one person can't be at two locations simultaneously

Posted by [Gennady Andrienko](#) on 04/11 at 05:42 AM

Asad,

Teams may use any existing commercial product or research prototype and may combine tools.

Posted by [UML](#) on 04/11 at 02:00 AM

Hi,

Can we run it through some already available tool in the market such as SAS

Posted by [Asad Ahamad](#) on 04/11 at 01:13 AM

Hi Greg,

As shown in the README file, the longitudes are specified as West, so they can be represented as negative numbers.

Posted by [uml](#) on 04/03 at 12:02 AM

I suppose all the longitudes in the Microblogs.csv should be considered to be West of Greenwich and thus actually negative?

Posted by [Greg](#) on 04/02 at 11:14 PM

Mini-Challenge 1, 3, and Grand Challenge task descriptions are available and Mini-Challenge 1 data description is available. Please see the Task Description page for more information.

Posted by [uml](#) on 03/23 at 10:08 AM

Melancon, thank you for your interest in the challenge. Sorry for the delay. All datasets will be ready by the end of the month. The submission deadline is the end of June, like in previous years. Please refer to the Task Description page for mini-challenge details and sample datasets. I'll be posting additional information for mini-challenge 1 and 3 shortly.

Posted by [uml](#) on 03/18 at 01:33 PM

I'm getting worried about the data still not being available. I understand this must be a heck of a job but we need time 😊 to prepare good submissions Will the submission dates be adjusted accordingly? Plus, I have students waiting for data to run their term project ...

Posted by [Melançon](#) on 03/18 at 12:55 PM

The full datasets are not available yet. Under Task Description at the top of the page you can download the sample dataset.

Posted by [uml](#) on 02/28 at 09:01 AM

Are the data sets available yet? the download page seems to be down

Posted by [Steve](#) on 02/27 at 02:26 PM

Page 1 of 1 pages

Comment Form

Commenting is not available in this weblog entry.

Powered by [ExpressionEngine](#)