**Applying Grits Analytic to WHO MERS Articles**

POC: Dr. Andrew Huff [huff@ecohealthalliance.org](mailto:huff@ecohealthalliance.org)

**Abstract**

This report summarizes the results of running various GRITS Analytics on a collection of 80 WHO articles about MERS/CoV infections that were provided to EHA by Prof. Naren Ramakrishnan.

**Article Classification**

GRITS is capable of classifying news articles by giving them one or more disease labels. All articles in the collection were classified with the labels ‘MERS’ and ‘Coronavirus,’ except for article 9 and article 44. Article 9 was only given the ‘Coronavirus’ label perhaps because it does not mention MERS by name. It does mention a ‘novel coronavirus’ in Qatar. Article 44 was labeled as ‘Swine Flue H1N1’ and ‘Influenza’. It discusses a patient that has a H1N1 infection and a novel coronavirus infection. The ‘Coronavirus’ label was erroneously not applied.

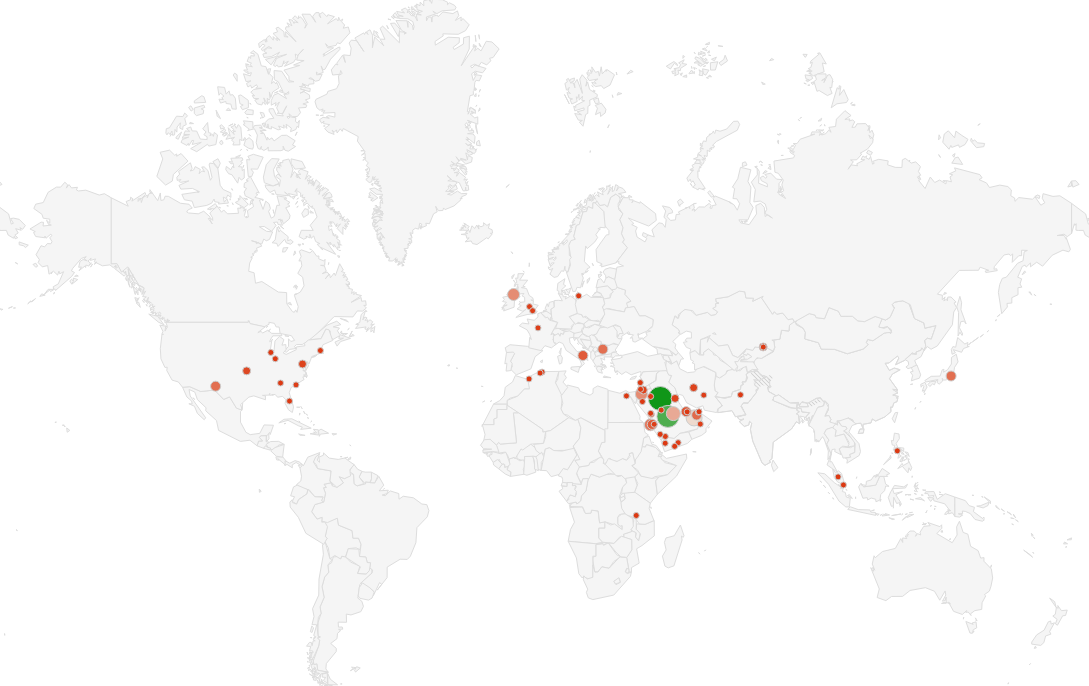
**Symptoms**

GRITS uses a combination of ontologies to identify epidemiological keywords in a variety of categories. The following table contains the top 10 symptom keywords ranked by the number of articles they appeared in. I thought it was unexpected that `diarrhea` appeared more frequently that `cough`.

|  |  |
| --- | --- |
| **Symptom keyword** | **Articles the keyword appears in** |
| infection | 78 |
| respiratory infection | 36 |
| diarrhea | 31 |
| renal failure | 24 |
| acute respiratory distress | 15 |
| shock | 13 |
| respiratory disease | 13 |
| cough | 10 |
| fever | 9 |
| pneumonia | 9 |

**Location Resolution**

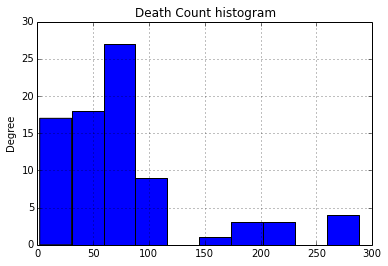
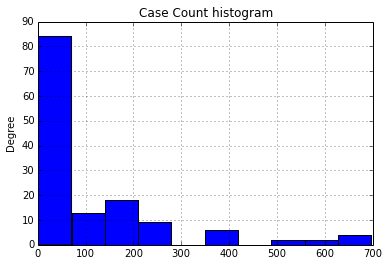
GRITS is capable of identifying location names in articles and resolving them to coordinates. The map below shows all the locations mentioned in the collection of WHO articles. The size of the dots is proportional to the number of times the location is mentioned.



**Case Counts**

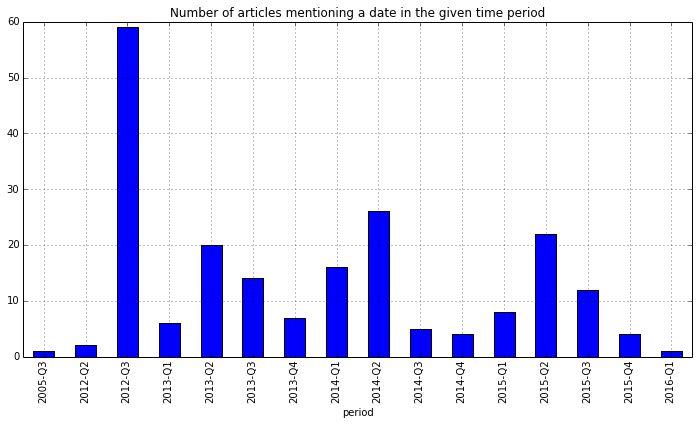
|  |  |
| --- | --- |
| GRITS can extract case and death counts by searching for patterns like “{{number}} new cases” in the bodies of articles. The histogram to the right shows the distribution of the number of counts extracted from the collection of WHO articles. Most of the WHO articles contain multiple case or death counts even though they focus on individual cases. | download.png |

The histograms below shows the distribution of the magnitude of the case and death counts.[[1]](#footnote-1)



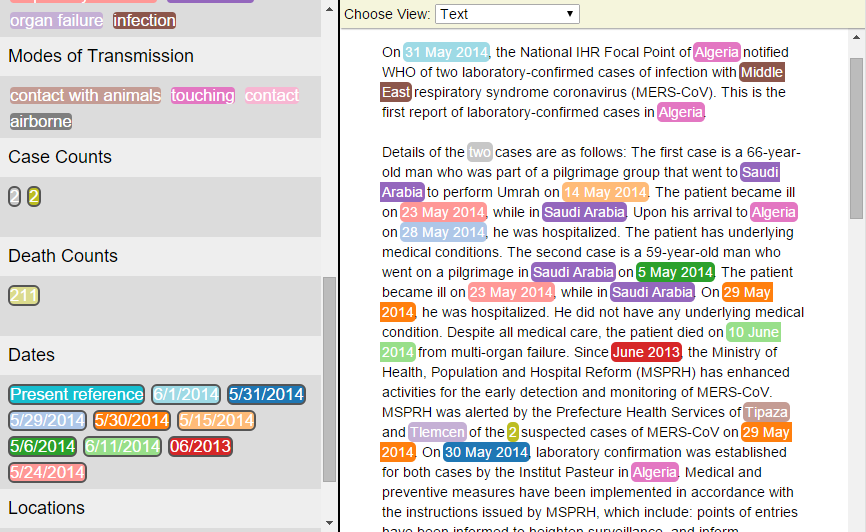
**Time information**

GRITS extracts dates mentioned in articles by using the Stanford SUTime library. The plot below shows the distribution of the dates extracted from the WHO article collection. For example, the plot shows that nearly 60 articles mention dates in the 3rd quarter of 2012 when MERS was first identified.



**Dashboard**

The GRITS Dashboard provides a way to view the features GRITS identifies in the context of the original article. In the screen capture below article 77 is annotated with the dates, case counts and death counts that GRITS could identify.



**Line Lists**

Since IARPA is particularly interested in line lists, we will address how we think GRITS could potentially be used to generate this type of information. Typically a line list will have rows that correspond to single patients.[[2]](#footnote-2) However, news articles frequently present outbreak information at a population level. This population level data could be presented in a line-list-like table by associating features like dates, locations and counts with each other to form rows. Feature association can be done by using the text-offsets returned by the GRITS API to combine features that appear next to each other. In future work, feature association could be performed with greater precision by using NLP techniques like dependency parsing. Furthermore, GRITS has an experimental patient information extractor (PIE) capable of identifying ages and associating them with other features. This information is not included in the sample analytics provided by EHA, as the PIE is not included in the production version of GRITS and it is not in active development. However, significant interest in it may spur us to continue developing it.

1. The death count histogram does not include a few outliers which were years inadvertently picked up as counts. [↑](#footnote-ref-1)
2. Example: http://foodborne.unl.edu/public/role/epidemiologist/lineLists.html [↑](#footnote-ref-2)