BioCaster in 2021: Automatic Disease outbreaks detection from global news media

Ontology-based Text mining:

Ontology is an explicit specification of conceptualization and a formal way to define the semantics of knowledge and data. The formal structure of ontology makes it a nature way to encode domain knowledge for the data mining use.

For example :

An example of ontology is when a physicist establishes different categories to divide existing things into in order to better understand those things and how they fit together in the broader world

Source: google

Problem Statement:

How to extract or understand the huge volume of unstructured data about the events related to outbreak of disease from the open source news media?

Outbreak of disease news cause serious concerns and the information is vital for people of any demographic.

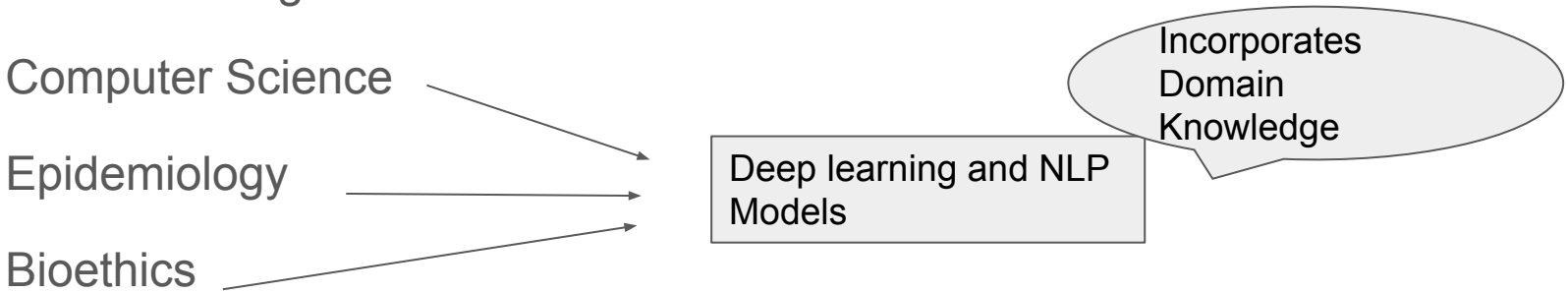
Information pertaining to particular region or race can be misleading and discriminating and thus lead to false interpretation.

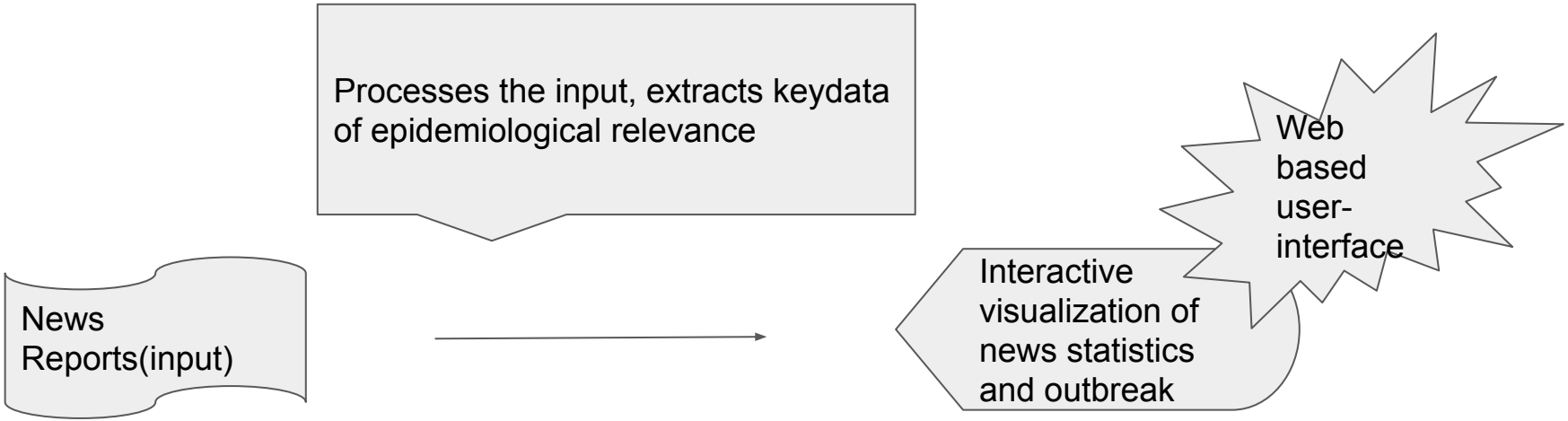
Certain organizations provide little detail and attention on data equality which leads to biases in the data.

Some live streaming news are high-volume, multilingual and biased and so the organizations have to implement state of the art technology in NLP(Natural Language Processing) techniques in order to process such information and provide predictive models with more accurate results (less biased) from the outbreak.

Solution:

BioCaster uses latest AI technologies for real-time disease outbreak understanding and detection.




Processes the input, extracts keydata of epidemiological relevance
Web based user-interface
News Reports(input)
Interactive visualization of news statistics and outbreak

Goal :

Fully automated real-time media monitoring based on streaming news data

Steps:

Input : Data Sources such as Google news and RSS news feeds(CNN, New York Times, USA today etc)

Translate : Translates various documents from different languages to English

Identifies bio-medicine related news documents through neural models

Understand : Converts these documents to event semantics by using rule based methods, (goal of finding regularities in data that can be expressed in the form of an IF-THEN rule)

Named Entity recognition(NER) and Entity normalization techniques.

NER

The named entity recognition (NER) is one of the most popular data preprocessing task. It involves the identification of key information in the text and classification into a set of predefined categories. An entity is basically the thing that is consistently talked about or refer to in the text.

Source: geeksforgeeks

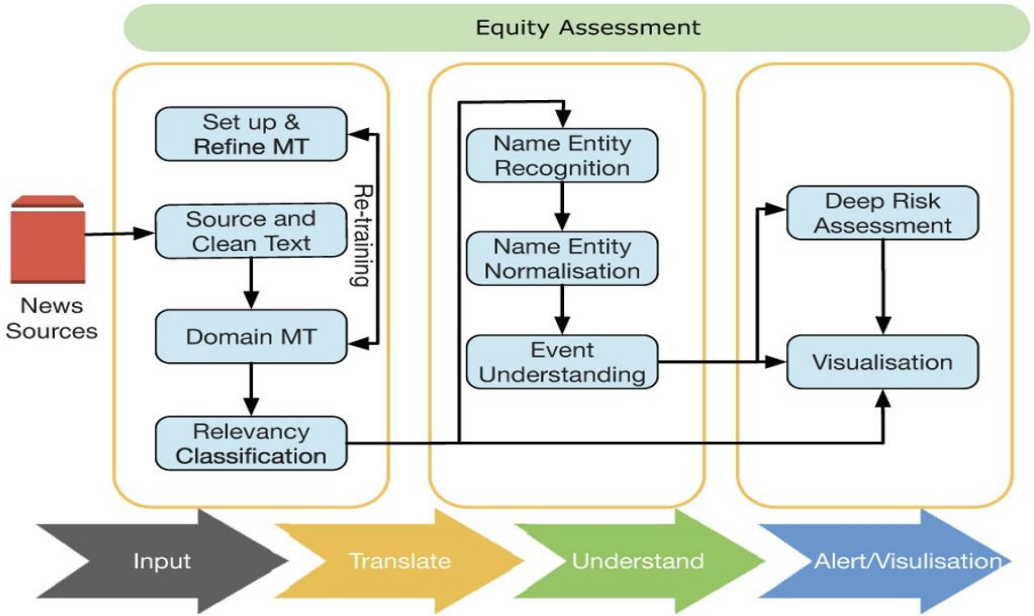
Entity Normalization Technique

Entity normalization (also called entity disambiguation, entity grounding, or entity linking) is an important subtask of information extraction that addresses this issue by linking entity mentions in text to categories or concepts of a reference vocabulary.

Source: bmcbioinformatics.biomedcentral.com

Key Features

- ★ multilingual translation,
- ★ deep neural techniques,
- ★ time series user interface and
- ★ equity assessment.


Equity Assessment
Set up & Refine MT
Source and Clean Text
Domain MT
Relevancy Classification
Re-training
News Sources
Name Entity Recognition
Name Entity Normalisation
Event Understanding
Deep Risk Assessment
Visualisation
Input
Translate
Understand
Alert/Visulisation

Conclusion

Thus the detection of real time disease outbreak is achieved by processing large amount of news data using deep neural techniques.

http://biocaster.org/

Limitations

Limitation to cope with negations(for example reports that have prominent decreases of cases)

Lacks in evaluating the aspects of the events

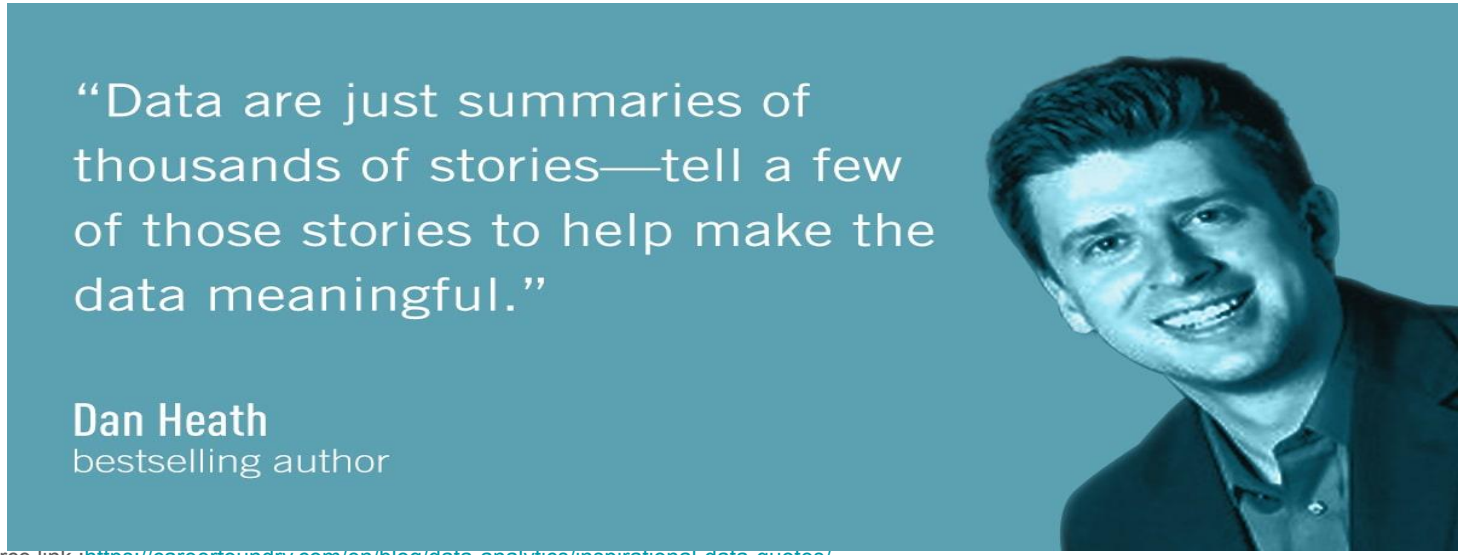
Input is the traditional media, does not cover non traditional media such as memes etc as of now.

Reference link

Download the paper here:
https://academic.oup.com/bioinformatics/article-pdf/38/18/4446/45878238/btac497.pdf

https://www.nejm.org/doi/full/10.1056/nejmp2012910

Thank you!



Source link :https://careerfoundry.com/en/blog/data-analytics/inspirational-data-quotes/

Devi Priya,Data Science Graduate, SJSU