

Real-time Disaster Event Extraction from Unstructured Text Sources

¹Joint Centre for Disaster Research, Massey University

²Institute of Natural and Mathematical Sciences, Massey University

Abstract

We present a system for real-time event extraction to support emergency response. Automatically extracting events from the unstructured text can address the challenge of information scarcity currently faced by emergency responders. The task is to identify the main event from online text sources such as online news and tweets by answering 5W1H questions (who did, what, when, where, why and how).

Introduction

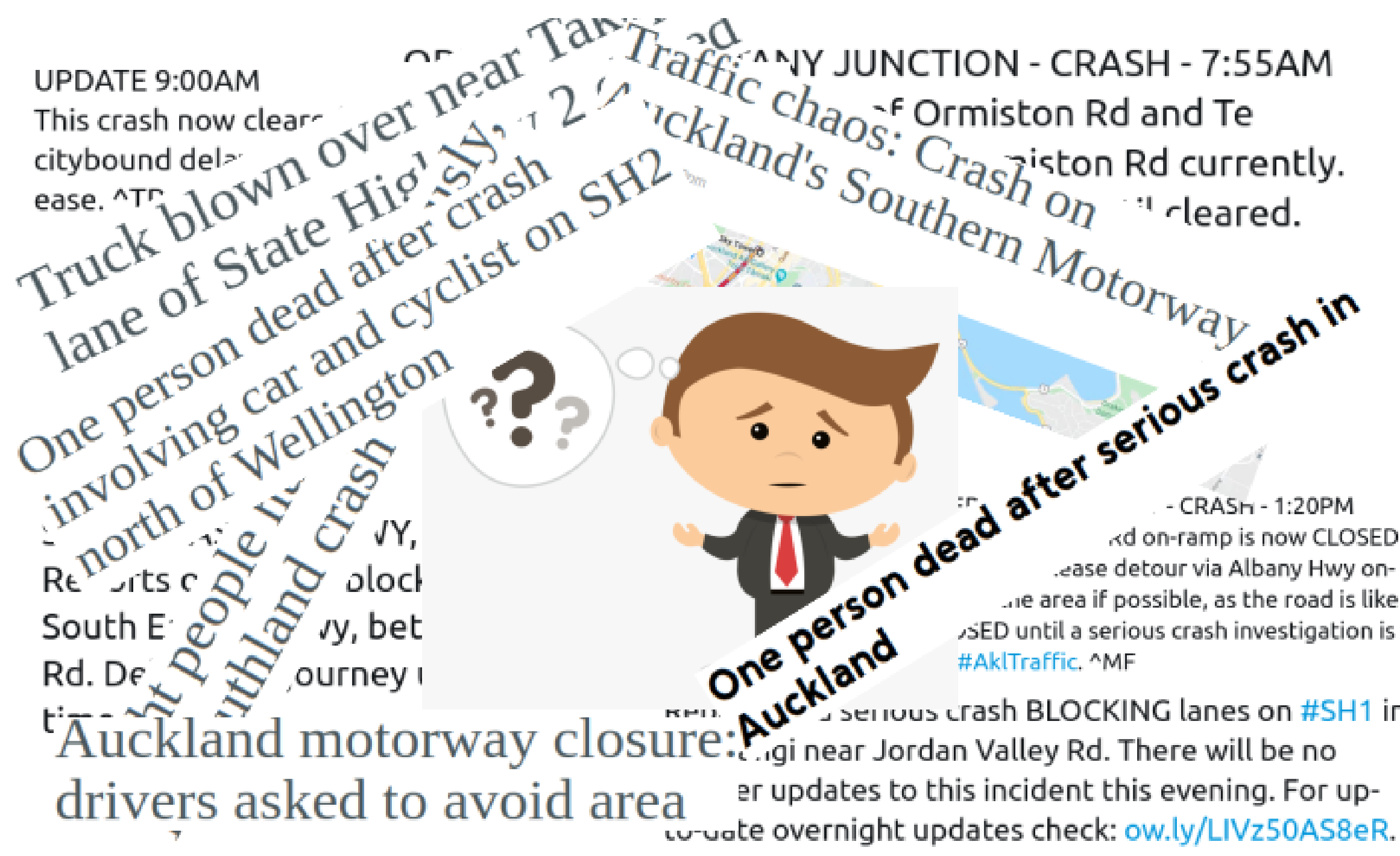


Figure 1: The problem with unstructured text

Multiple events happen in the environment around us, and they are reported in different media sources with different styles. It is beyond the capacity of the human brain to identify events in real-time from multiple sources. Converting unstructured text to a structured format can help emergency responders to better understand the context and make informed decisions.

Real-time event extraction processing chain

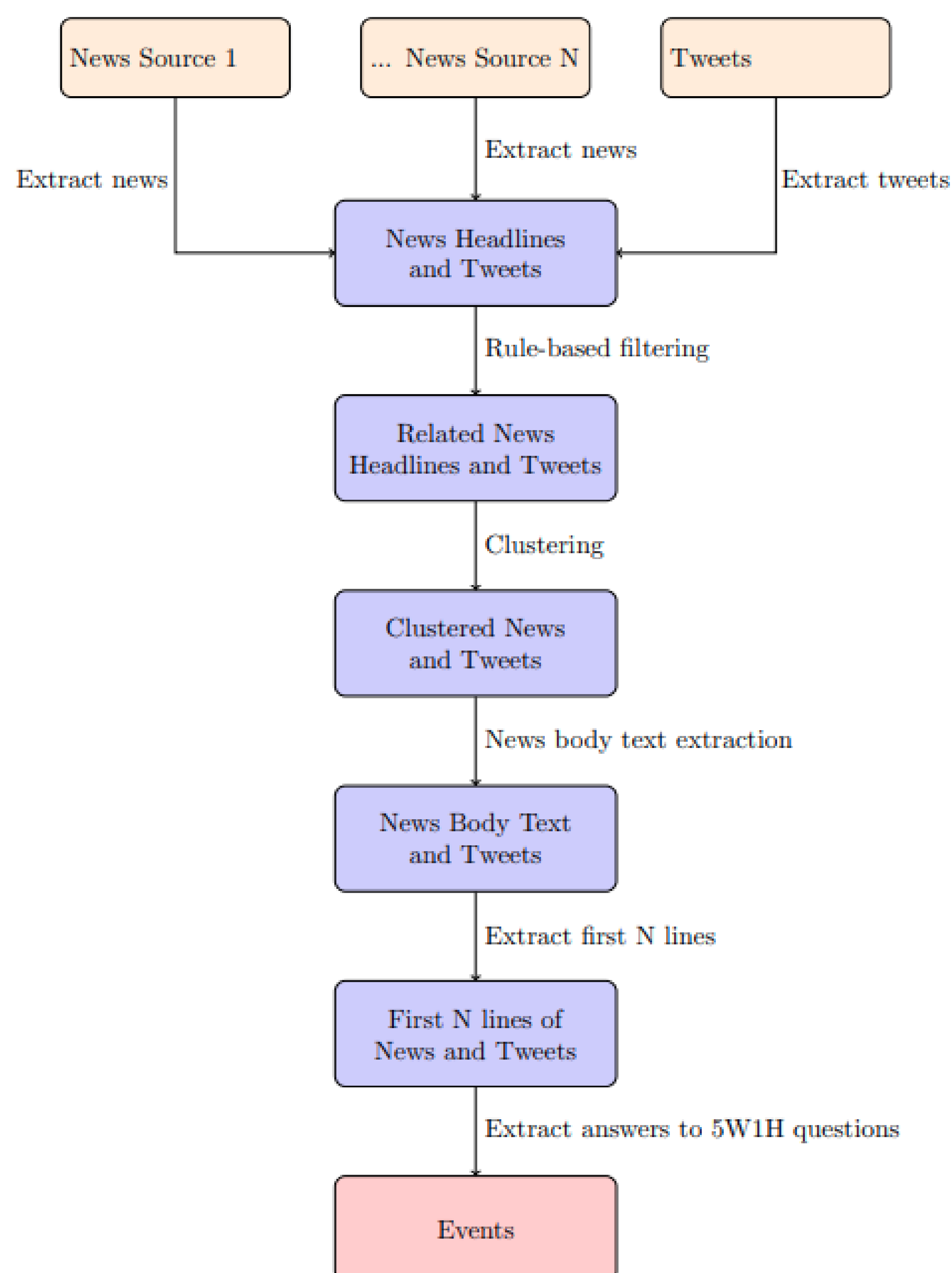


Figure 2: The architecture of the real-time event extraction system

The Dataset

The Dataset is collected from New Zealand Historic Weather Events Catalogue available at National Institute of Water and Atmospheric Research (NIWA) website.

The Dataset Details

Description

The duration	20.5 years (2000-01-01 to 2020-07-01)
#of weather events before preprocessing	215
#of weather events after preprocessing	195
#of news headlines in the dataset	2824

Table 1: Details of the dataset

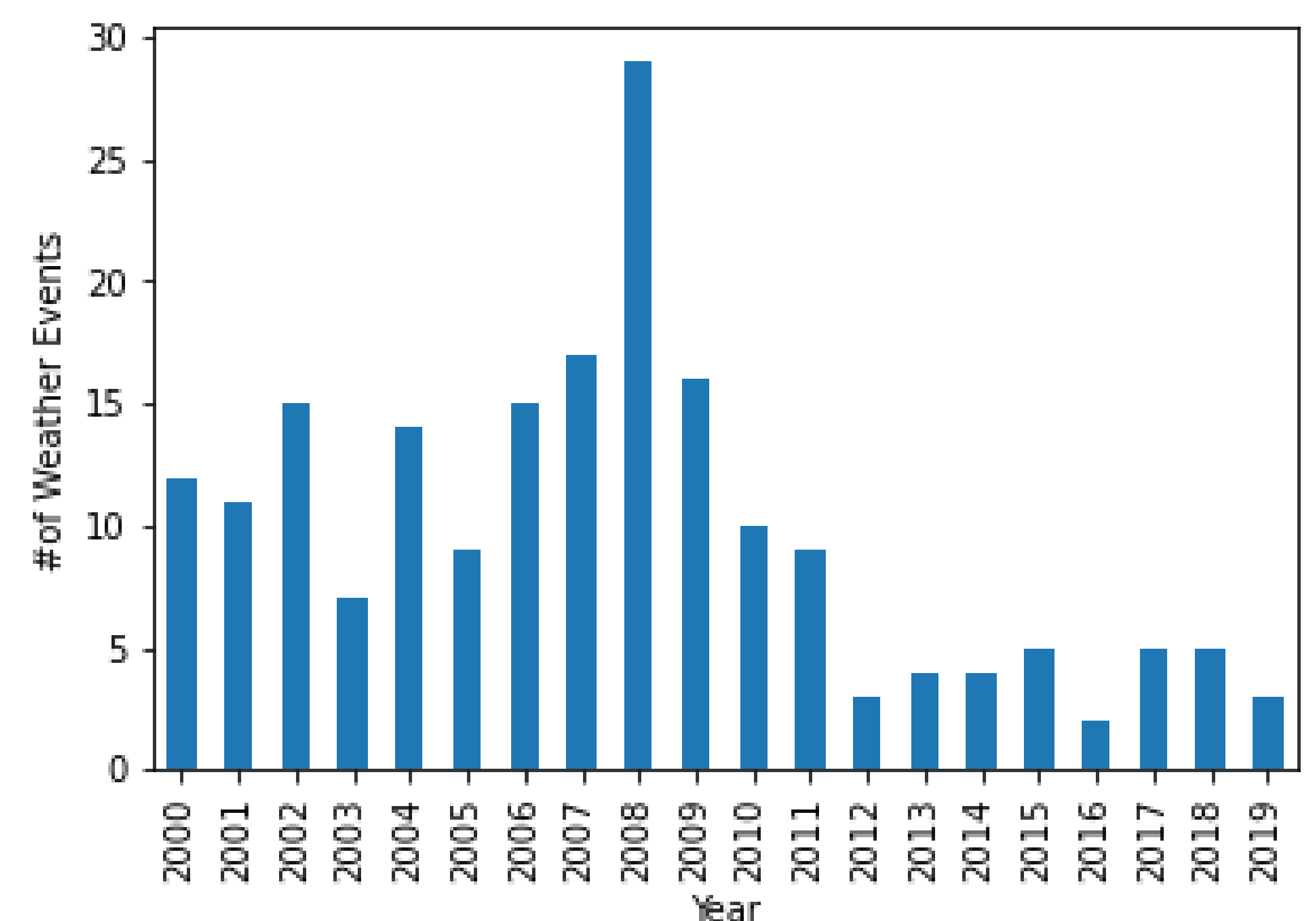


Figure 3: The distribution of weather events by year

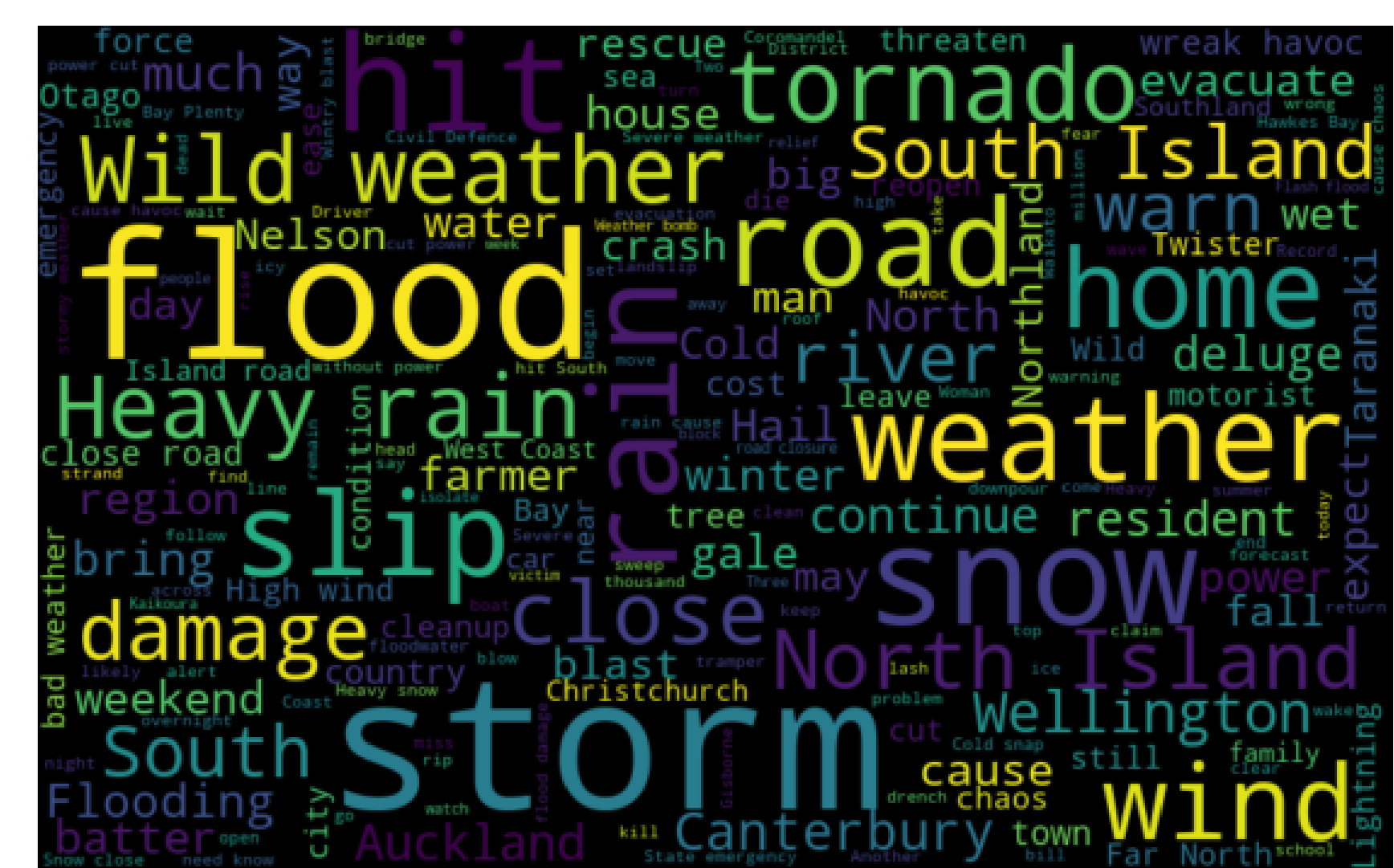


Figure 4: The word-cloud based on the frequency of words in the dataset

Initial Results

Motorists face delays as main road to West Coast remains shut

Figure 5: The news headline

```
Who? : Motorists
What? : face delays
When? : until at least Monday
Where? : West Coast
Why? : main road to West Coast remains shut
How? : No matching found!
```

Figure 6: The answers extracted by the algorithm for 5W1H questions

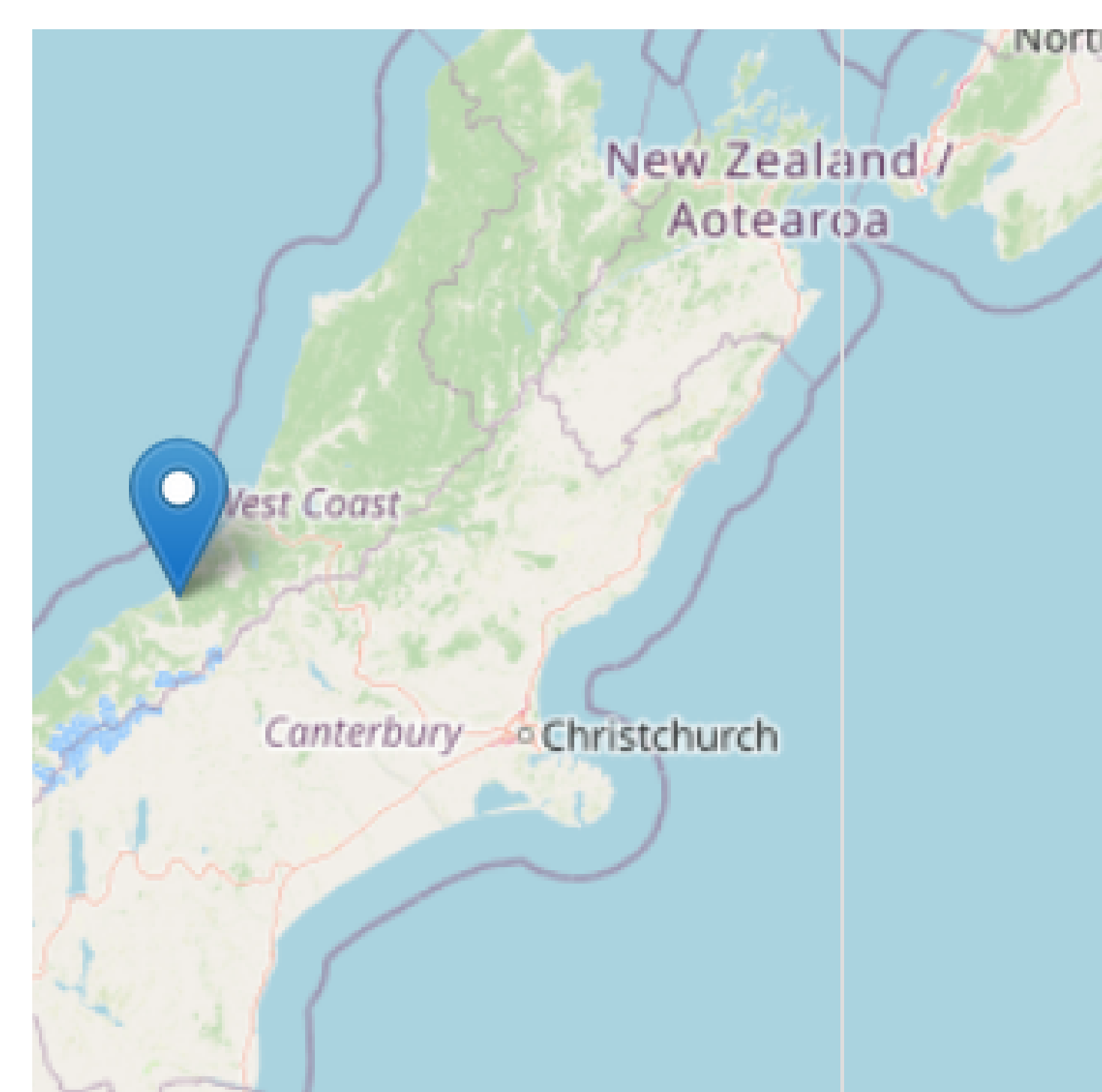


Figure 7: Geo-locating of the event