Justin Aguila
AIT-580

10/14/2023

M8 Lab 1B – Text Analysis with Python and NLTK

In this part of the exercise, we utilized the function “tokenize()” which is an integral part of the Natural Language Processing technique. It divides the large paragraph in our text into smaller pieces. We used this to dissect textual descriptions of biking incidents into individual sentences, allowing us to better understand the nature and extent of these injuries.

As shown in *Figure 1*, this shows a sample of the tokenization output. Overall, we produced 30 distinct sentences detailing a variety of injuries from concussions, dislocated shoulders, broken ribs, deep tissue damage, and many more. Although we ran the tokenize() to populate 35 distinct sentences, we were able to capture a majority of the context and idea from the text. Some injuries are cited in conjunction with protective gear such as helmets, suggesting their importance in minimizing injury severity.

```
Sentence 1: "Concussion, head and neck pain"
"Concussion, dislocated shoulder, broken rib."
Sentence 2: "I have deep tissue damage to my left side, and multiple bruises over my body.
Sentence 3: My lower back also went out."
```

Figure 1. Tokenization Sample (out of 30 outputs)

Furthermore, the sentences also provide qualitative data such as the duration of hospital stays and the need for extensive physical therapy. This indicates the severity and long-term implications of certain injuries. This analysis could offer valuable insights into how we can innovate and improve protective gear for bikers, potentially leading to more effective products that specifically target the most commonly injured body parts.

The use of the “tokenize()” function proves invaluable in breaking down complex and multi-faceted textual datasets into manageable units of meaning. Through this, we were able to identify recurring types of injuries and gain insights into their severity and long-term consequences. This information can be instrumental for policy recommendations and public awareness campaigns aimed at improving biking safety.