TwitLoccan

This Python module makes use of current popular libraries to estimate whether a given profile on social media is from a Canadian. This module will only use information that has been put in the public domain by the individual. The data will be purely for aggregate trend analysis and will not target any particular individuals.

Overview

Statements about broad trends require the ability to sort large volumes of data into various categories. Many web-based interfaces may allow for partial completion of some fields. In other cases, the fields may be entered as informal text strings. Popular modules already exist in this area. The purpose of this module is to focus on Canada. Thus the efforts to ensure that that Kingston, Ontario is identified as in Canada rather than Jamaica is worth the effort.

The Canadian specific characteristics are a product of manual tuning of the existing libraries. Using data from such sources as Twitter, there will be attempts to improve on the initial estimates from multiple sources. Three libraries will be used in various ways. It is expected that there will be a tradeoff between speed and accuracy. The more libraries that are used the greater the accuracy but also the slower the speed.

The module will also contain considerable volumes of manual edits. For example, a reference to a city may be informal and ambiguous. However, an additional quote cheering on a particular sports team may be used to narrow the choice.

Usage Example

In the following examples will demonstrate the usage of the module.

In [8]:
```
import location
import geograpy

testlocation = location.C3(useGEOGRPY = True)
isCan = testlocation.isCan("Winnipeg")
print(isCan)
```

```
True
```

Future Directions

The above example is preliminary. Future directions will include more informal langauge but more formal accuracy tests.