Lisp in Summer Projects Submission

Submission Date	2013-10-02 11:23:22
Full Name	Denis Papathanasiou
Country	USA
Project Name	tweet-secret
Type of software	command-line/terminal app
General category	web tool
LISP dialect	Clojure
GitHub URL	https://github.com/dpapathanasiou/tweet-secret
Did you start this project?	Yes, all the code is written by me
Project Description	I want to describe my project in this form.
Purpose	This is a text steganography application optimized for use on Twitter, written in Clojure. The idea is to create innocuous-looking tweets which really have a secret, or hidden meaning.
Function	The tweet-secret application is essentially a book cipher, using a corpus text known only among the sender and intended recipients. Hidden messages are converted into tweets by using sentences selected from the corpus text.
Motivation	Twitter accounts are either completely private or public; if you have a public account and want to tweet something privately to a select list of people, there is no simple way to do it. Twitter does have a direct message feature, but it has several drawbacks: you can only send messages to people one at a time, you're limited to people who follow you

explicitly, and worst of all, it may be subject to 3rd party snooping, governmental or otherwise.

Audience

I got the idea for this from both the recent disclosures about NSA spying on social media in this country, and the stories of political activists in countries such as Egypt, Iran, China, etc. going to jail (or worse) for what they posted on Twitter.

Methodology

The tweet-secret application is essentially a book cipher, using plain text files available on the web or commonly-shared among a group of people as the "book" or corpus text. The corpus, which consists of one or more texts, is known only among the sender and intended recipients, and should be changed frequently to prevent eavesdroppers from picking up the pattern or method.

The corpus text is parsed into a list of sentences, filtered by character length to exclude anything longer than 140 characters in size. This list of sentences is known as the set of eligible tweets.

Messages are encoded by looking up each word token in a common dictionary, then using that pointer reference (sequence in one or more dictionary lists) to a sentence in the set of eligible tweets whose preceding cumulative string length matches the pointer. That sentence becomes the tweet to be broadcast, to correspond to the word token.

Since the exact pointer position may not match the start or end of a corpus sentence exactly, an unobtrusive marker is used at that point in the text (by default a unicode middle dot character, which can be changed in the config.properties file), which is important for decoding.

Tweets are decoded by finding their position in the set of eligible tweets, counting the cumulative string size up to that point, and adding the amount of the offset marker, if present. The resulting number is the dictionary pointer, which is used to lookup the corresponding word.

Conclusion

While the current release is functional, in that it does what it is supposed to, the use of an offset pointer which inserts a foreign character (the default is a middle dot character) in the selected corpus sentences takes away from the steganographic aspect somewhat.

Ideally, I would like the algorithm to evolve such that that marker character is no longer required.

I also have a list of future enhancements I would to add at some point:

* Come up with a better strategy for handling message words which are not defined in the default dictionary-files texts

* Pack multiple short tweets together into a single broadcast tweet, space-permitting, so that it's not always a 1:1 correspondence between words in the message to tweets (not only harder to break, but also more efficient use of bandwidth)

* Create a graphical user interface in Swing, Standard

Widget Toolkit, or Seesaw as an alternative to the command line interface

* Use the Twitter API to post tweets automatically, if an application has been defined, and the relevant application OAuth settings (Consumer key, Consumer secret, etc.) have been defined in config.properties

Build Instructions

Get and install Leiningen if you do not already have it.

Next, clone this repo to your computer, go to the folder where it exists, and run the following commands from a terminal:

```
$ lein deps
$ lein uberjar
```

If the build succeeds, you should now have a jar file created in the repo's target folder named tweet-secret-1.0-standalone.jar.

Test Instructions

There are two test cases, containing four assertions, included which test the encoding and decoding of an English-language message, using a static corpus from gutenberg.org.

The dictionary for testing is re-bound to a static, universally available dictionary text found online, since the linux words file as defined by default in config.properties can vary from distro to distro and computer to computer.

To run the tests, use this command (you will need an internet connection to have them run successfully, since both the corpus and dictionary texts used in the texts are defined as remote URLs):

```
$ lein test
```

If successful, you should see this:

```
lein test tweet-secret.core-test

Ran 2 tests containing 4 assertions.
0 failures, 0 errors.
```

Execution Instructions

The current version works via the command line. Use the --help switch when invoking the standalone jar file to get the list of options:

```
$ java -jar target/tweet-secret-1.0-standalone.jar --help

tweet-secret: Text steganography optimized for Twitter

Switches Default Desc
-------- ------- ----
-c, --corpus REQUIRED: at least one url or full path filename
of the secret corpus text(s) known only by you and your
friends
-d, --decode Decode this tweet into plaintext (if none
present, text after all the option switches will be encoded)
-h, --no-help, --help false Show the command line usage
help
```

For example, suppose we want to encode the message "Tonight we take Paris by storm" as a series of innocuous-looking tweets.

Let's use The History Of The Conquest Of Mexico (http://textfiles.com/etext/NONFICTION/mexico) by William Hickling Prescott on textfiles.com as the randomly selected corpus text.

Open a terminal, go to the target folder containing the repo, and type this command:

```
$ java -jar target/tweet-secret-1.0-standalone.jar --corpus
http://textfiles.com/etext/NONFICTION/mexico
"Tonight we take Paris by storm"
```

On Mac OSX, you should also include -Dfile.encoding=utf-8 as a command line argument to the java interpreter so that the tweet strings are output correctly:

```
$ java -Dfile.encoding=utf-8 -jar target/tweet-secret-1.0-
standalone.jar --corpus
http://textfiles.com/etext/NONFICTION/mexico
"Tonight we take Paris by storm"
```

This results in the following six tweets, one for each word of the original message:

```
On the following morning, the gener·al requested permission
to return the emperor's visit, by waiting on him in his palace.
A pitched battle follow·ed.
But the pride of Iztapalapan, on which its lord had freely
l·avished his care and his revenues, was its celebrated
gardens.
This form of governm·ent, so different from that of the
surrounding nations, subsisted till the arrival of the
Spaniards.
The Mexica·ns furnish no exception to this remark.
He ·felt his empire melting away like a morning mist.
```

Followers who know the corpus text can decode these tweets with this command:

```
$ java -jar target/tweet-secret-1.0-standalone.jar --corpus
http://textfiles.com/etext/NONFICTION/mexico
--decode "On the following morning, the gener·al requested
permission to return the emperor's visit, by waiting on him in
his palace."
--decode "A pitched battle follow·ed."
--decode "But the pride of Iztapalapan, on which its lord had
freely l·avished his care and his revenues, was its
celebrated gardens."
--decode "This form of governm·ent, so different from that of
the surrounding nations, subsisted till the arrival of the
Spaniards."
--decode "The Mexica·ns furnish no exception to this
remark."
--decode "He ·felt his empire melting away like a morning
mist."
```

Which results in this list of words, corresponding to the original message:

```
tonight
```

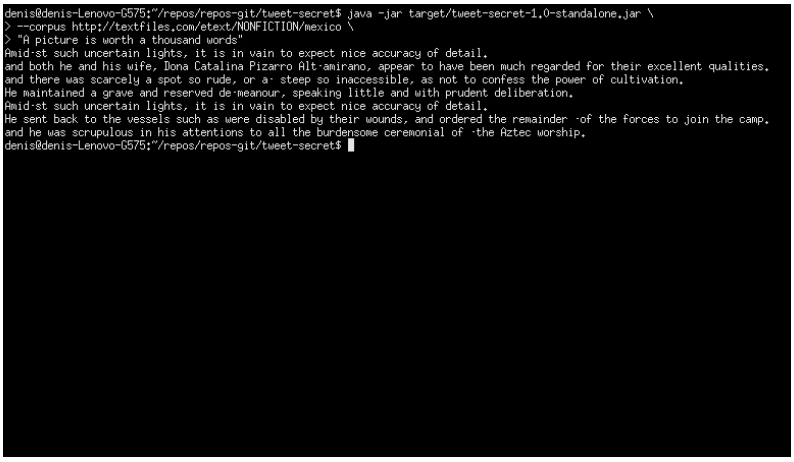
we
take
Paris
by
storm

Describe any bugs or caveats

One issue with encoding is that each word of the hidden message must exist at least one of the dictionary files defined in config.properties.

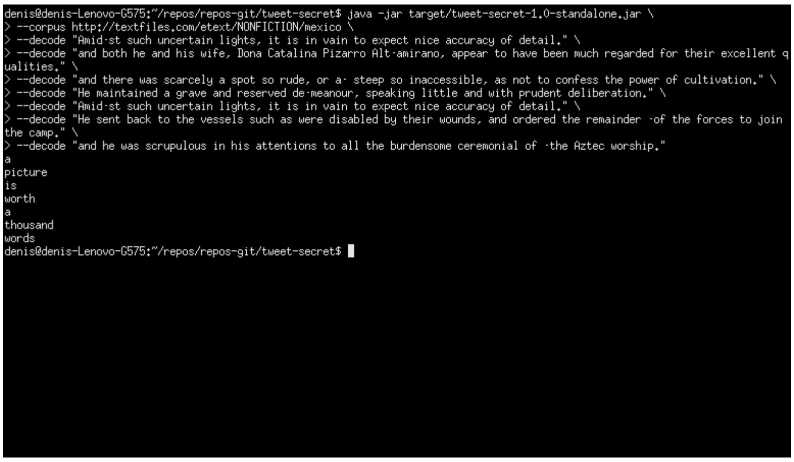
If that's not the case, the application issues a warning, but ideally, the application should be more robust.

Screen shots



```
denis@denis-Lenovo-G575:~/repos/repos-git/tweet-secret$ java -jar target/tweet-secret-1.0-standalone.jar \
> --corpus http://textfiles.com/etext/NONFICTION/mexico \
> "A picture is worth a thousand words"
Amid·st such uncertain lights, it is in vain to expect nice accuracy of detail.
and both he and his wife, Dona Catalina Pizarro Alt·amirano, appear to have been much regarded for their excellent qualities.
and there was scarcely a spot so rude, or a· steep so inaccessible, as not to confess the power of cultivation.
He maintained a grave and reserved de·meanour, speaking little and with prudent deliberation.
Amid·st such uncertain lights, it is in vain to expect nice accuracy of detail.
He sent back to the vessels such as were disabled by their wounds, and ordered the remainder ·of the forces to join the camp.
and he was scrupulous in his attentions to all the burdensome ceremonial of ·the Aztec worship.
denis@denis-Lenovo-G575:~/repos/repos-git/tweet-secret$ █
```

Screenshot from 2013-10-02 11_19_07.png



```
denis@denis-Lenovo-G575:~/repos/repos-git/tweet-secret$ java -jar target/tweet-secret-1.0-standalone.jar \
> --corpus http://textfiles.com/etext/NONFICTION/mexico \
> --decode "Amid·st such uncertain lights, it is in vain to expect nice accuracy of detail." \
> --decode "and both he and his wife, Dona Catalina Pizarro Alt·amirano, appear to have been much regarded for their excellent q
ualities." \
> --decode "and there was scarcely a spot so rude, or a· steep so inaccessible, as not to confess the power of cultivation." \
> --decode "He maintained a grave and reserved de·meanour, speaking little and with prudent deliberation." \
> --decode "Amid·st such uncertain lights, it is in vain to expect nice accuracy of detail." \
> --decode "He sent back to the vessels such as were disabled by their wounds, and ordered the remainder ·of the forces to join
the camp." \
> --decode "and he was scrupulous in his attentions to all the burdensome ceremonial of ·the Aztec worship."
a
picture
is
worth
a
thousand
words
denis@denis-Lenovo-G575:~/repos/repos-git/tweet-secret$ █
```

Screenshot from 2013-10-02 11_22_53.png

Official

I have read rules and have abided by them.
I am 18 years of age or older.
I am not living in Brazil, Quebec, Saudi Arabia, Cuba, Iran, Myanmar (Burma), North Korea, Sudan, or Syria.