CO558 Cryptography Assignment

Tess of the d'Urbervilles

Educational value

Or: other than a chance to score a really good coursework mark, what else does this get you?

This piece of coursework has two objectives:

- it gets you working in a practical sense with encryption, decryption, plaintext, ciphertext, and keys. This gives you a practical sense of what these are all about, which should be useful for the entirety of the crypto section of the module. It also lets you experience the impact of having certain quantities of ciphertext available.
- it makes sure you know some of the basic tricks in cryptology: substitution, permutation, possibly frequency analysis, applied to a few basic ciphers that you need to know about anyway.

All of this is essential in a course on cryptography, but much nicer to examine through coursework where you have time and computing resources available, rather than in an exam.

Introduction

This assessment comprises seven exercises, unimaginatively called Exercise 1 to Exercise 7. For each exercise you will be given an extract from Thomas Hardy's novel *Tess of the d'Urbervilles* (first published in 1891), encrypted with one of the ciphers we discussed earlier in the course. Your job is to decrypt the extract, recovering the plaintext. In each exercise, every student will be given a different extract to decipher, and the encryption key will vary from student to student; however, all the extracts within a particular exercise will be encoded with the same sort of cipher, as described below. For example, in Exercise 1 everyone will have an extract encoded with a Caesar cipher.

You are free to work with colleagues in devising methods for decrypting particular sorts of cipher, and in putting together programs to help you do so; however, **when it comes to decrypting your own individual pieces of ciphertext, you must work individually, without help from anyone else**.

Brute-force methods are not ruled out; however, you may find the assessment more rewarding if you introduce an analytical component into your decryption tactics.

Preliminaries

On `raptor`, in the directory `\\raptor.kent.ac.uk\exports\courses\co558\assignment` (or equivalent `/courses/co558/assignment`) you will find the following files:

`tess.txt`

This is the ASCII text of *Tess of the d'Urbervilles* from the Project Gutenberg website. Provided mainly for interest, and for the licensing information at the end.

`tess27.txt`

This is `tess.txt` reduced to a 27-character alphabet in the following way:

1. Prefatory and ending material from Project Gutenberg that isn't part of the novel itself has been removed;
2. Apostrophes have been removed;
3. Lower-case letters have been converted to upper case;
4. Each sequence of one or more non-letter characters (including for example digits, punctuation, hyphens and whitespace) has been replaced by a single '|' (vertical bar) character.

`tess26.txt`

A further reduction to a 26-character alphabet, obtained by omitting the vertical bars from `tess27.txt`.

For most exercises, the plaintext is a randomly chosen string of 840 characters taken from `tess26.txt`; the exceptions are Exercise 2, where the extract is only 30 characters long, and Exercise 7, where the extract comes from `tess27.txt`. **You therefore cannot assume that the extract will start at the beginning of a word in the original novel, nor that it ends at the end of a word.**

The Exercises

In the `assignment` sub-directory, you will find further subdirectories named with your login. Within each of these directories you should find seven text files in the form `cexercise1.txt` etc. (If you don't, contact c.perez@kent.ac.uk as soon as possible.) These files contain the ciphertexts that you have to decrypt for the exercises in question, terminated by a newline which does not form part of the ciphertext.

Exercise 1 (2 marks)

The plaintext comes from `tess26.txt` and is encoded with a Caesar cipher.

Exercise 2 (3 marks)

The plaintext comes from `tess26.txt` and is encoded with a Vigenere cipher using the 21-letter key `TESSOFTHEDURBERVILLES`.

Exercise 3 (4 marks)

The plaintext comes from `tess26.txt` and is encoded with a Vigenere cipher. The key is an arbitrary sequence of six letters (i.e. not necessarily forming an English word).

Exercise 4 (5 marks)

The plaintext comes from `tess26.txt` and is encoded with a Vigenere cipher. The key is an arbitrary sequence of between 4 and 6 letters.

Exercise 5 (5 marks)

The plaintext comes from `tess26.txt` and is encoded with a transposition cipher, as follows: the plaintext is written row-wise across a certain number of columns,

between 4 and 6. (You must figure out how many columns were used.) The ciphertext is formed by reading out successive columns from left to right.

Exercise 6 (5 marks)

The plaintext comes from `tess26.txt` and is encoded with a transposition cipher, as follows: the plaintext is written row-wise across six columns. The ciphertext is formed by reading out successive columns in an arbitrary order (which you must figure out to decipher the message). *Hint:* look for common pairs of letters, such as 'th'.

Exercise 7 (6 marks)

The plaintext comes from `tess27.txt` and is encoded with a general substitution cipher, using a randomly chosen mapping from the 27-character alphabet onto itself. Note that normally (i.e. except by chance) a vertical bar will be mapped onto some other letter of the alphabet.

Not Cricket

The files `tess26.txt` and `tess27.txt` are made available to you to enable you to *check* your answers, not to assist you in finding the answers. For example it would be contrary to the spirit of this assessment to write programs to search these files as part of the decryption process. It is not straightforward to detect this, but if there is evidence, marks will be deducted.

However, it is entirely in order to analyse these files to determine language statistics, e.g. letter frequencies.

Use of Others' Software

I strongly prefer you write your own software (individually or in groups). That said, it is permissible to use software provided by others under very strict conditions, which **all** need to be satisfied:

- The software acts (more or less) like a calculator. In other words, it is OK to use an online program to automate a simple repetitive process (as long as you know how to that process yourself). It is NOT OK to use a program that automates the entire decryption process without any help from you (unless you write that program yourself).
- The software is freely and immediately available, and you indicate how;
- The software is well documented, just as would be required if you had written the software yourself;
- Your description includes a description in your own words of how the software works, in sufficient clarity and detail to **clearly demonstrate your understanding on the decryption approach used** and to allow someone to replicate the approach without using the software.

If these conditions are not sufficiently satisfied, you may even **get 0 marks** despite providing the correct decryption.

Dyslexia

It is possible that dyslexia may make one or two of the exercises particularly difficult. If, having made a serious attempt, you find that you cannot make progress with an exercise because of this, please contact me (c.perez@kent.ac.uk).

How to Submit Your Work

The deadline is on the Moodle page as well as SDS. Deadline on the Moodle takes precedence over any other information). In `/proj/co558c/decryption/` (or `\\raptor.kent.ac.uk\exports\proj\co558c\decryption\`), you will find a folder with your login on it. Into this folder you should place a text file called `exercise1.txt` with your answer to Exercise 1. (To reiterate, it's got to be a *text* file and be called `exercise1.txt`: calling it `Exercise1.txt`, `excercise1.txt`, `exercise1.txt`, `exercise1.doc` or `exercise1.txt.txt` ("hide known extensions" - you're doing a module on Security!?) etc. is a surefire way to get 0 marks for Exercise 1.) Similarly place in this folder files called `exercise2.txt`, ... `exercise7.txt`.

Each of these files must have the following format:

- Line 1 must contain **(only) the first 30 characters** of the plaintext you have obtained by decrypting your ciphertext extract. To get full marks for an exercise, it is necessary that these characters be exactly right, so it's worth checking that they correspond to a sequence of 30 characters from `tess26.txt` or `tess27.txt` as the case may be. Indeed it will normally be necessary to get these characters exactly right to get *any* marks from the exercise.
- Line 2 must be blank.
- Lines 3 **onwards** should contain a description of how you carried out the decryption. This description need not be very long - in some cases a single sentence may do - but should be sufficient to allow a reader to reproduce your approach. If you used a program to help you, you should identify it, and include the source as a separate file (or files) in your submission folder: **this source must be properly set out and commented, and include the names of all the authors**.

  It will not be possible to look at these descriptions or the source files in every case, but spot checks will be made, and if the description/source is missing or inadequate, marks shall be deducted.

For example, this is what `exercise1.txt` might look like:

```
RINTANDFROMTHEBACKOFHERHEADAKI

Using pen and paper, I worked through all possible shifts until
I found one that converted the ciphertext into recognisable
English. I checked that the resulting extract occurred in
tess26.txt. All this took some time, so I then boarded the
Tardis and travelled back to the beginning of the 21st century
to submit my work.
```

(The exercise and description were originally written by Andrew Runnalls. Further modified by Palani Ramaswamy. This version is made by C. Perez-Delgado.)