Practical

1 Background

This assignment consists of practical experiments with historical ciphers. The task is to analyze four different ciphertexts, identify the cipher used, and then break it.

The original plaintexts are all written in English from one or more novels written 70-80 years ago. Each plaintext is different and each is encrypted with a different classical cipher algorithm. The four algorithms used, in random order in your set, are:

- Random simple substitution
- Caesar cipher
- Vigenère cipher
- 2×2 Hill cipher

For all of the ciphertexts, the alphabet used is an unusual one consisting of 29 characters, namely the uppercase letters A-Z and the three punctuation symbols, ',' (comma), '.' (full stop or period) and '-' (dash). The frequency distribution of the individual symbols together with the most common digrams and trigrams are shown in the Appendix below. Before encryption, all plaintext characters were changed to upper case and all characters outside this alphabet were deleted.

It is expected that you make use of software and/or artificial intelligence tools to help you. All the tasks can be achieved with publicly available tools. It is allowed and encouraged to write your own scripts or programs to simplify repetitive elements, but this is not essential. Some recommended software tools are the following:

- CrypTool: `https://legacy.cryptool.org/en/cto`
- CyberChef: `https://gchq.github.io/CyberChef`
- JCrypTool: `https://www.cryptool.org/en/jct`
- Cryptii: `https://cryptii.com`
- Dcode: `https://www.dcode.fr/en`

Note that these web tools may try to serve you large amounts of advertising, so you may consider using them with an ad blocker or in "private" modes.

There are 10 points available in total for this assignment, with partial marks as shown below. Your answers should be submitted by midnight of **January 30, 2026**.

Your answers must be delivered through Blackboard as a single text document (feel free to use LaTeXor any text editor). This is an individual assignment. It is acceptable to discuss the general approach with other students, but your answers should be your own.

You need to obtain your individual folder of four ciphertext files. The folders and a list of usernames tied to folder numbers are available as a zip file on the course website at: `https://ttm4135.iik.ntnu.no/latest/#practical-10`. Remember to mark the submission with your username and ciphertext number to speed up grading.

2 Assignment

Your text document should provide solutions to the five following problems.

2.1 Statistical analysis (2 points)

Use a frequency analysis tool to find the distribution of individual characters (1-grams), digrams (pairs of letters or 2-grams) and trigrams (3-grams) for each of your ciphertexts. You will end up with three tables for each ciphertext, similar to those in the Appendix. Include the tables for ciphertext 0 and explain which tool(s) you used.

2.2 Caesar analysis (2 points)

Compare your tables with the Appendix and identify the Caesar cipher. It should be sufficient to study the individual character frequencies, identify the most frequent, and check whether this shift is plausible for all characters. Explain your observations.

Provide the number (0, 1, 2 or 3) of your Caesar ciphertext, the key (shift value), and the (cleaned up) plaintext in your document. Explain which tool(s) you used.

2.3 Substitution analysis (2 points)

Use the frequency analysis for characters, digrams and trigrams to identify the random simple substitution cipher in your set. Find which ciphertext 3-gram corresponds to the letters THE. This is not necessarily the most common 3-gram, so check it against the 1-gram and 2-gram distributions as well to make sure that your guess is correct.

Provide the number (0, 1, 2 or 3) of your random substitution ciphertext and the assumed encryption of THE. Explain how you reached this conclusion. There is no need to find a complete key or the full plaintext (unless you want to).

2.4 Vigenère analysis (2 points)

Compare your tables with those in the Appendix and identify the Vigenère cipher in your set. The index of coincidence tool in Dcode can be useful: `https://www.dcode.fr/index-coincidence`. Given that the period is 5, find the plaintext. You should split your ciphertext into five streams and then apply the same technique that you used to decrypt the Caesar cipher to decrypt each string.

Provide the number (0, 1, 2 or 3) of your Vigenère ciphertext and the key as five letters. Explain how you reached this conclusion and what tools you used.

2.5 Hill cipher analysis (2 points)

Find the key for the remaining Hill cipher. This can be challenging. The following process is recommended:

1. Using your digram frequency table, determine candidate digrams which map to "TH" and "HE". These are probably two of your most common digrams.

2. Once you have identified likely candidates for two digrams, test them by solving the linear system to find the encryption key.

3. For a candidate key, try to decrypt parts of the ciphertext to get English text.

You may need to make a few attempts if the statistics of your individual ciphertexts do not work out nicely. We give points for good answers even if you were not able to find the correct key.

Provide the number (0, 1, 2 or 3) of your Hill ciphertext and the key; write your key as four characters row by row. Explain your guesses and what tools you used.

A Character, digram and trigram distributions

These tables show the distribution of single characters, digrams, and trigrams across the entire source file used to generate the plaintexts.

Char	%
E	11.93
T	8.8
A	7.9
O	7.52
I	6.82
H	6.42
N	6.23
S	5.99
R	5.54
D	4.36
L	4.00
U	2.83
M	2.64
W	2.37
Y	2.24
F	2.11
C	2.1
G	1.89
,	1.7
.	1.5
P	1.47
B	1.37
V	0.92
K	0.75
-	0.21
X	0.14
Q	0.11
J	0.09
Z	0.04

Digram	%
TH	2.81
HE	2.53
IN	1.64
ER	1.62
AN	1.5
HA	1.34
RE	1.28
OU	1.26
AT	1.08
EN	1.04
TO	1.02
IS	1.02
ES	1.02
ED	1.01
ON	1.00
IT	1.00

Trigram	%
THE	1.58
YOU	0.77
AND	0.74
ING	0.71
THA	0.52
HAT	0.52
,AN	0.41
THI	0.36
HER	0.35
HIS	0.32
VER	0.32