Vigenère cipher cryptanalysis

FIRST ATTEMPT

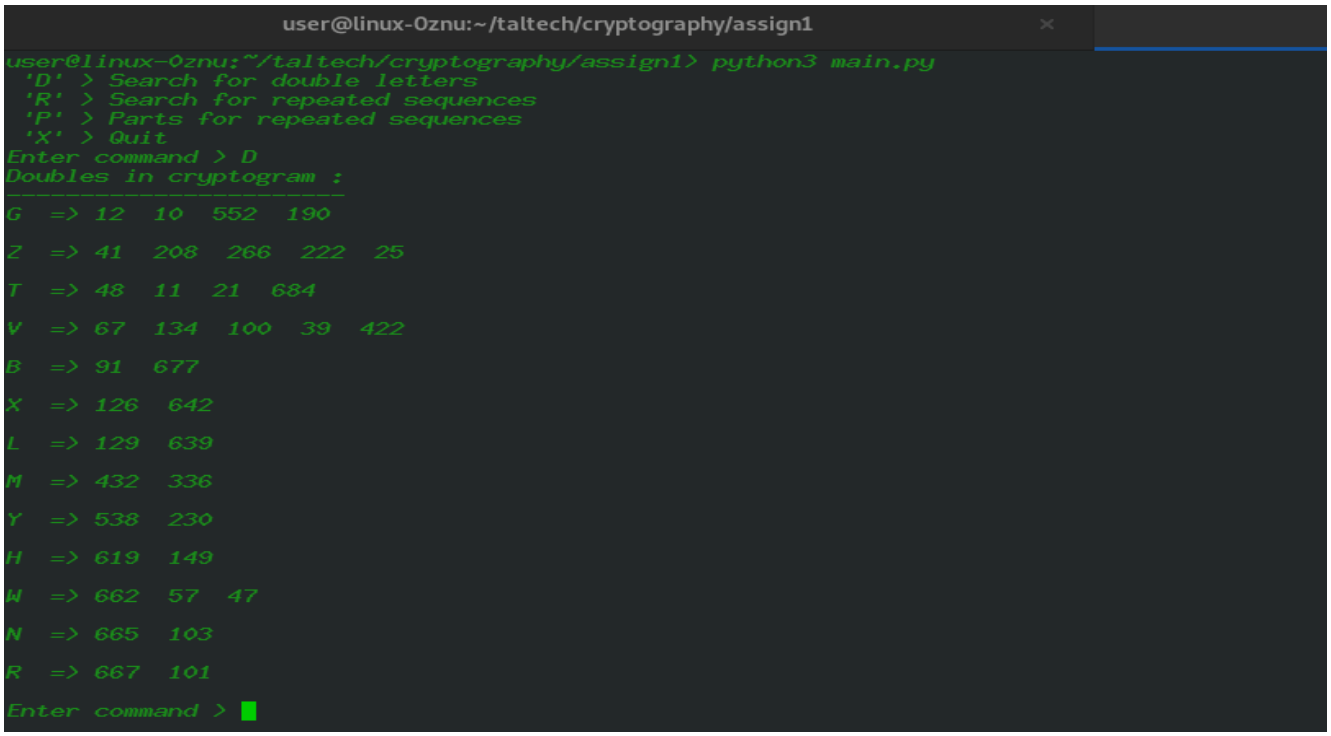
Student code: 194633IV

The cryptogram to break is:

GFOLNVSTHAIZGGOTOFRHUAYKGGRSVYELWUOJZTFWYZZMWFMZTTGLXVKMGAJVIT
TWCRZVVATXVXCHRVMHWFMZTTBVIFSBBYULZHTSNFALSGTKNXCLNVXLWVKFLMV
WYKXXSLLZXLHLNVELSWZNUISGVCKZCATXOGHGZYKACMYVZBAWVRYLSKGWZXF
SCYOESLNVEGCLOTKMVJKVVXFKUEYVCEOEMHILUWZASZULYXHZKGNRGAIZYMGS
EJZASATZZBODSVGLIJKDKGHOGJTMOUILXTHWZYKUWGRFMBGLYRELHZKPNTJWXV
VKCVATKWHZKDGMVWSRZBQAGEYTMKOWKQOUZCEHBWVVXLCFKEZXFKZYKACM
YVZASFOKCBZDHVKFDLERMTWFCYKGVWTIEDWKYZTZSJRVLMVSXMGKRSTUCXBLZ
FCTGZOEMMCFZFYXFNKZTMVWTZDHBSJDOGWKZIGMWGTYKPOKGJQXRTEFTXVAYE
KPQGRCKTUMKJGUCMZKNXDGRKOVODOELBUZZZTZWFGTGWSEOROGKSYYOGULUE
CXFWLRSHIKLFXICDOKOVODOEZKWYAVOMGGAIPHPKUDKHBWGJQXRTAKCXFWVZQ
XFKIFSIOJKUZHHZKSGVYKZRHUWFMRTWRAXKEICDOKOVGSZLTBJWXJOMWWYNNR
RGEFAISGVCKYWYNKRBYWZYGMYAYJOGUWXZYLOAJKUAONKIKLDGTUKWWFNZYE
COMIGOSDRPBHWUKZZLPWIRALSLNVYMOCKJGKSKUCUP

FIRST ATTEMPT

While observing the cryptogram at first, I noticed many double letters repeated. So the first intuition I thought of was to write a functionality in the program, which search for redundant double letters (feature ‘D’).



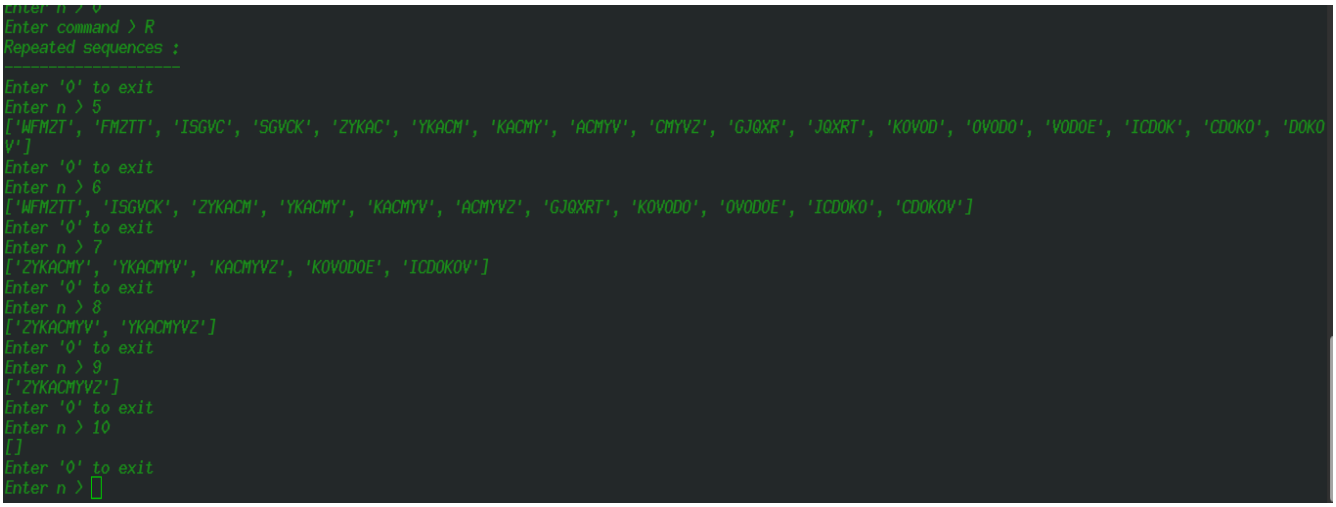
For example, here I can notice that the string ‘GG’ appears three times in the cryptogram, and the program gives me the spacing between each instance. The point from printing these values for each double letter, is to try to conclude something or look for some interesting values. Unfortunately, nothing interesting appeared here. This is not surprising as the procedure came out of an intuition, nothing scientific.

KASISKI APPROACH

Then, I thought of proceeding according to the Kasiski Analysis. So the first step was to find some repeated sequences in the cryptogram. The purpose behind is obviously to find the length of the key. Except that some repetitions might be just coincidences, and this can already be a misleading repeated sequence.

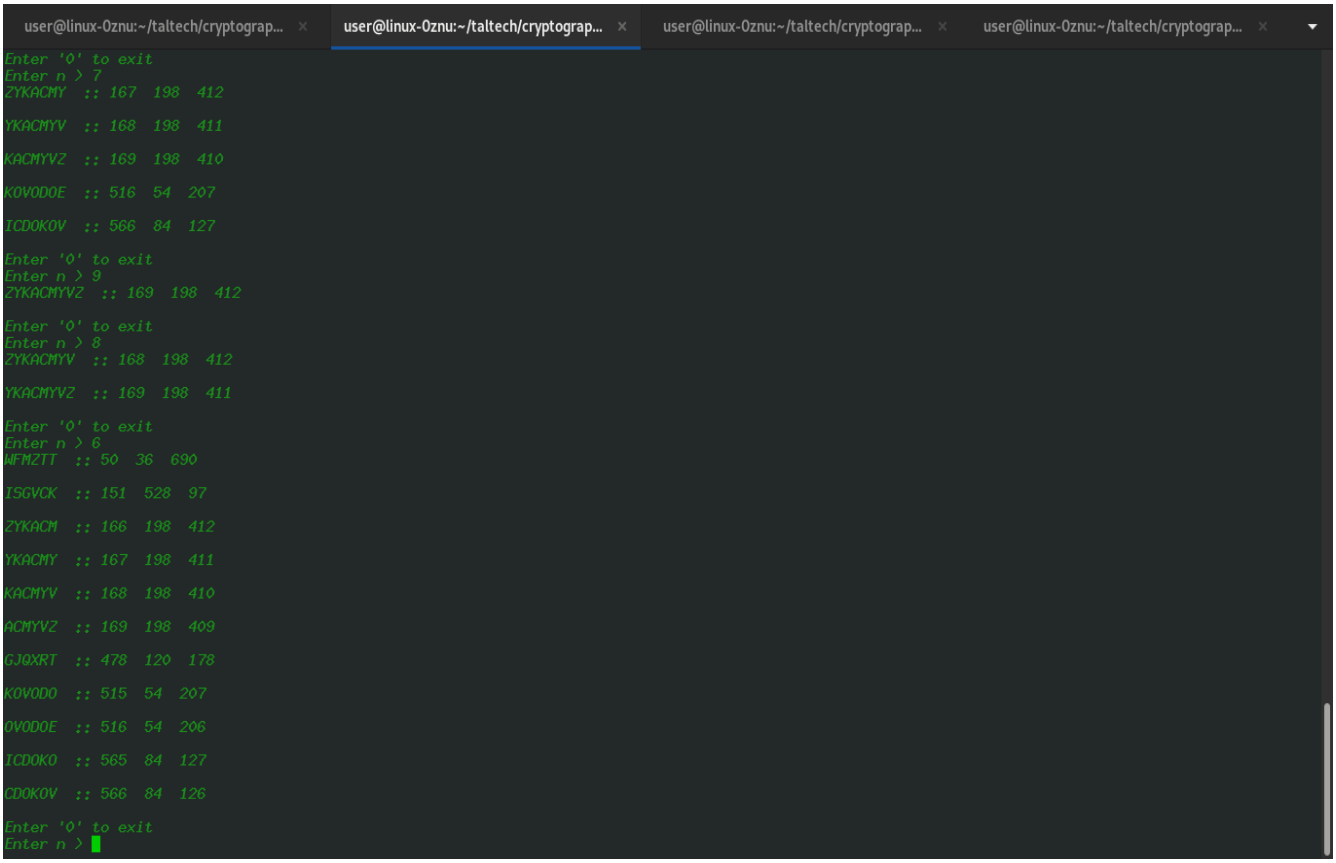
For this step, I developed two more functionalities. One gives the repeated sequences based on a given sequence length (feature ‘R’), which already allows noticing some details. The second also given a sequence length, gives in addition to the repeated sequences of this length, the spacings for each sequence (feature ‘P’).

I start with the feature R. The more I increase the sequence size, the less repeated sequences I get, which is quite logic. But observing those sequences can be very useful to detect some eventual misleading sequences. It’s also interesting to observe some sequences share a common part, and when increasing the sequence size, we find them again, but regrouped.



For example, for a sequence length of 5, we have ‘KOVOD’, ‘OVODO’ and ‘VODOE’. While for a sequence length of 7, we find ‘KOVODOE’. This feature also shows that repeated sequences have a length range between 3 and 9.

Then using feature P, we can observe the spacing between each part for each sequence. When specifying the length each time, we now already know that there is no need to go further than 9.



We display the spacing' values for the different sequence length value. We get something like a table. Except for some 3-letter sequences, all the sequences, whatever their length, give us three spacing values. The third column is obviously the less reliable. That's due to the Vigenère system itself, because the last key repetition is usually truncated in order to fit the length of our plain text, and so we can't really trust the behavior of the cryptogram at his end. Then, by looking at the first column, I can see that some values are just subsequent integers. This can be explained with sequences which share some common parts, which we already noticed previously.

Therefore, I will focus on the second column for the moment. It's an intuitive hypothesis I'm doing based on previous observations. This column seems the most promising track to follow, so I try to eliminate some "bad" candidates. By looking at the factorization into dividers' list of values present in this column, we have 3 permanent dividers which are 2, 3 and 6.

6=2x3 and 6 seems the most likely to be the length of the key. Keys of length 2 or 3 are rarely used, and we usually prefer the longest key. So once again, I'm here making a hypothesis and basing myself on intuition and observation, to select the most promising track to follow. In case 6 fails as key length, then I try 2 and 3.

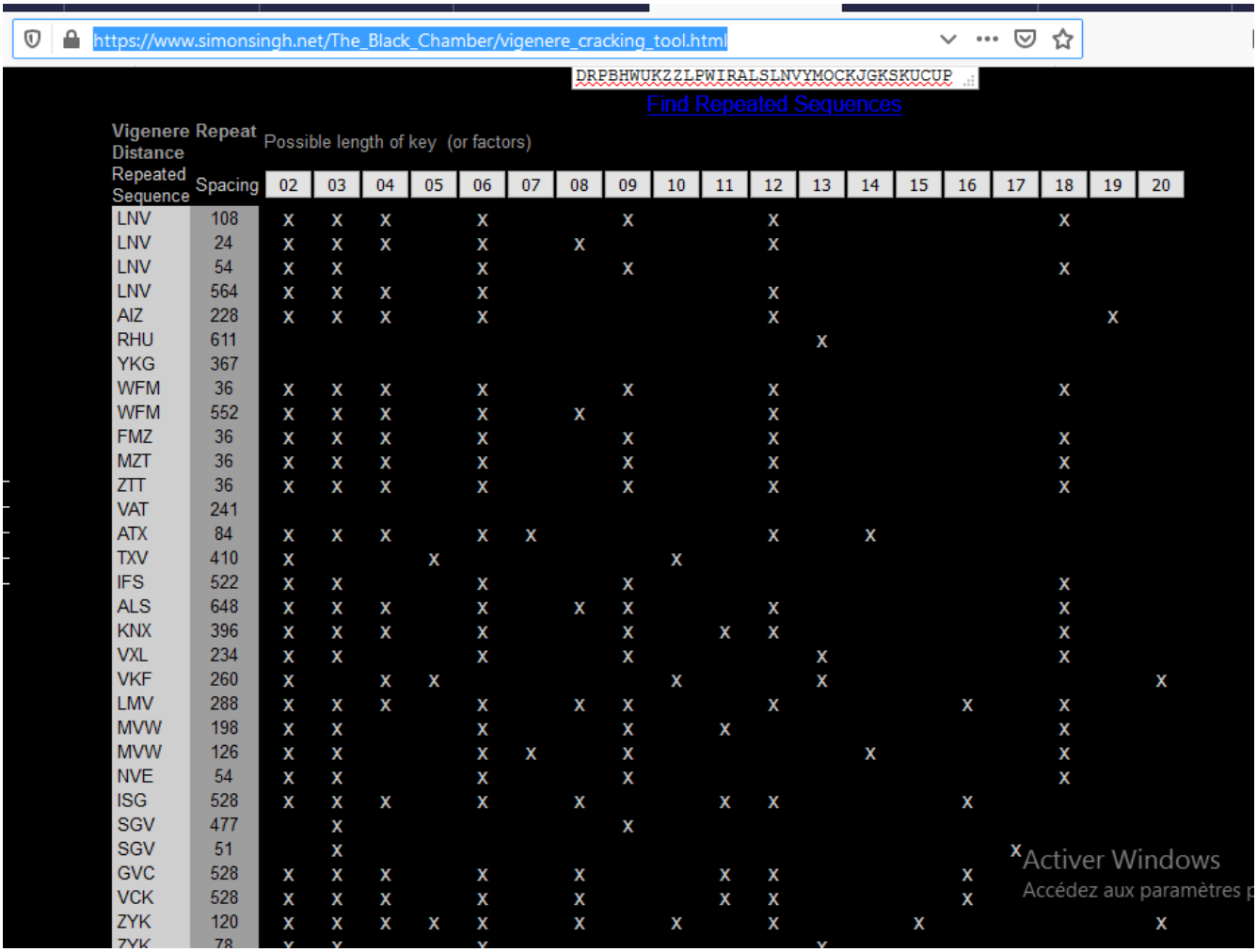
The next step of the Kasiski approach is to to make 6 groups of letters, where each group represent letters encrypted with the same key-letter. For this, I finally added a fourth functionality to the program, which after grouping the letters, it should make a frequency analysis, and therefore conclude about each key-letter (feature 'A').

```
user@linux-0znu:~/taltech/cryptography/assign1> python3 main.py
 'D' > Search for double letters
 'R' > Search for repeated sequences
 'P' > Parts for repeated sequences
 'A' > Cipher analyse
 'X' > Quit
Enter command > A
Analyse cipher :
---------------
Enter key length > 6
Letter group  1  =>  ['G', 'S', 'G', 'R', 'G', 'E', 'Z', 'Z', 'T', 'K', 'I', 'Z', 'V', 'M', 'T', 'S', 'Z', 'A', 'N', 'X', 'L', 'X', 'X', 'E', 'U', 'K'
, 'O', 'K', 'Z', 'Y', 'Z', 'O', 'E', 'K', 'V', 'Y', 'M', 'Z', 'Y', 'N', 'Y', 'Z', 'Z', 'G', 'K', 'T', 'X', 'K', 'M', 'E', 'N', 'V', 'K', 'G', 'Z', 'Y'
, 'K', 'E', 'X', 'Z', 'K', 'Z', 'C', 'K', 'M', 'K', 'E', 'T', 'L', 'G', 'C', 'C', 'M', 'Y', 'T', 'D', 'O', 'G', 'K', 'Q', 'T', 'K', 'K', 'G', 'N', 'O'
, 'L', 'T', 'G', 'O', 'O', 'C', 'S', 'X', 'O', 'Z', 'O', 'P', 'K', 'Q', 'C', 'O', 'S', 'Z', 'G', 'H', 'T', 'E', 'O', 'T', 'O', 'N', 'A', 'K', 'R', 'G'
, 'O', 'Y', 'U', 'K', 'K', 'Y', 'G', 'B', 'Z', 'A', 'Y', 'G', 'U']
Letter group  2  =>  ['F', 'T', 'G', 'H', 'G', 'L', 'T', 'M', 'T', 'M', 'T', 'V', 'X', 'H', 'T', 'B', 'H', 'L', 'X', 'L', 'M', 'X', 'L', 'L', 'I', 'Z'
, 'G', 'A', 'B', 'L', 'X', 'E', 'G', 'M', 'X', 'V', 'H', 'A', 'X', 'R', 'M', 'A', 'B', 'L', 'G', 'M', 'T', 'U', 'B', 'L', 'T', 'K', 'W', 'M', 'B', 'T'
, 'Q', 'H', 'L', 'X', 'A', 'A', 'B', 'F', 'T', 'G', 'D', 'Z', 'M', 'K', 'X', 'T', 'M', 'X', 'M', 'H', 'G', 'M', 'P', 'X', 'X', 'P', 'T', 'U', 'X', 'V'
, 'B', 'Z', 'W', 'G', 'G', 'X', 'H', 'I', 'V', 'K', 'M', 'H', 'H', 'X', 'X', 'X', 'I', 'H', 'V', 'U', 'W', 'I', 'V', 'B', 'M', 'R', 'I', 'Y', 'B', 'M'
, 'G', 'L', 'A', 'L', 'W', 'E', 'O', 'H', 'L', 'L', 'M', 'K', 'P']
Letter group  3  =>  ['O', 'H', 'O', 'U', 'R', 'W', 'F', 'W', 'G', 'G', 'T', 'V', 'C', 'W', 'B', 'B', 'T', 'S', 'C', 'W', 'V', 'S', 'H', 'S', 'S', 'C'
, 'H', 'C', 'A', 'S', 'F', 'S', 'C', 'V', 'F', 'C', 'I', 'S', 'H', 'G', 'G', 'S', 'O', 'I', 'H', 'O', 'H', 'W', 'G', 'H', 'J', 'C', 'H', 'V', 'O', 'M'
, 'O', 'B', 'C', 'F', 'C', 'S', 'Z', 'D', 'W', 'V', 'W', 'S', 'V', 'R', 'B', 'G', 'C', 'F', 'V', 'B', 'W', 'W', 'O', 'R', 'V', 'Q', 'U', 'C', 'D', 'O'
, 'U', 'W', 'S', 'K', 'U', 'F', 'I', 'C', 'O', 'W', 'G', 'P', 'B', 'R', 'F', 'F', 'O', 'H', 'Y', 'W', 'R', 'C', 'G', 'J', 'W', 'R', 'S', 'W', 'Y', 'Y'
, 'U', 'O', 'O', 'D', 'W', 'C', 'S', 'W', 'P', 'S', 'O', 'S']
Letter group  4  =>  ['L', 'A', 'T', 'A', 'S', 'U', 'W', 'F', 'L', 'A', 'W', 'A', 'H', 'F', 'V', 'Y', 'S', 'G', 'L', 'V', 'W', 'L', 'L', 'W', 'G', 'A'
, 'G', 'M', 'W', 'K', 'S', 'L', 'L', 'J', 'K', 'E', 'L', 'Z', 'Z', 'A', 'S', 'A', 'D', 'J', 'O', 'U', 'W', 'G', 'L', 'Z', 'W', 'V', 'Z', 'W', 'A', 'K'
, 'U', 'W', 'F', 'K', 'M', 'F', 'D', 'L', 'F', 'W', 'K', 'J', 'S', 'S', 'L', 'Z', 'F', 'N', 'W', 'S', 'K', 'G', 'K', 'T', 'A', 'G', 'M', 'M', 'G', 'D'
, 'Z', 'F', 'E', 'S', 'L', 'W', 'K', 'D', 'D', 'Y', 'G', 'K', 'W', 'T', 'W', 'K', 'J', 'Z', 'K', 'F', 'A', 'D', 'S', 'W', 'W', 'G', 'G', 'Y', 'W', 'A'
, 'W', 'A', 'N', 'G', 'F', 'O', 'D', 'U', 'W', 'L', 'C', 'K']
Letter group  5  =>  ['N', 'I', 'O', 'Y', 'V', 'O', 'Y', 'M', 'X', 'J', 'C', 'T', 'R', 'M', 'I', 'U', 'N', 'T', 'N', 'K', 'Y', 'L', 'N', 'Z', 'V', 'T'
, 'Z', 'Y', 'V', 'G', 'C', 'N', 'O', 'K', 'U', 'O', 'U', 'U', 'K', 'I', 'E', 'T', 'S', 'K', 'G', 'I', 'Z', 'R', 'Y', 'K', 'X', 'A', 'K', 'S', 'G', 'O'
, 'Z', 'V', 'K', 'Z', 'Y', 'O', 'H', 'E', 'C', 'T', 'Y', 'R', 'X', 'T', 'Z', 'O', 'Z', 'K', 'T', 'J', 'Z', 'T', 'G', 'E', 'Y', 'R', 'K', 'Z', 'R', 'O'
, 'Z', 'G', 'O', 'Y', 'U', 'L', 'L', 'O', 'O', 'A', 'A', 'U', 'G', 'A', 'V', 'I', 'K', 'K', 'Z', 'M', 'X', 'O', 'Z', 'X', 'Y', 'E', 'V', 'N', 'Z', 'Y'
, 'X', 'J', 'K', 'T', 'N', 'M', 'R', 'K', 'I', 'N', 'K', 'U']
Letter group  6  =>  ['V', 'Z', 'F', 'K', 'Y', 'J', 'Z', 'Z', 'V', 'V', 'R', 'X', 'V', 'Z', 'F', 'L', 'F', 'K', 'V', 'F', 'K', 'Z', 'V', 'N', 'C', 'X'
, 'Y', 'V', 'R', 'W', 'Y', 'V', 'T', 'V', 'E', 'E', 'W', 'L', 'G', 'Z', 'J', 'Z', 'V', 'D', 'J', 'L', 'Y', 'F', 'R', 'P', 'V', 'T', 'D', 'R', 'E', 'W'
, 'C', 'V', 'E', 'Y', 'V', 'K', 'V', 'R', 'Y', 'I', 'Z', 'V', 'M', 'U', 'F', 'E', 'F', 'Z', 'Z', 'D', 'I', 'Y', 'J', 'F', 'E', 'C', 'J', 'K', 'K', 'E'
, 'Z', 'T', 'R', 'Y', 'E', 'R', 'F', 'K', 'E', 'V', 'I', 'D', 'J', 'K', 'Z', 'F', 'U', 'S', 'R', 'R', 'K', 'K', 'L', 'J', 'N', 'F', 'C', 'K', 'Y', 'J'
, 'Z', 'K', 'I', 'U', 'Z', 'I', 'P', 'Z', 'R', 'V', 'J', 'C']
Enter command >
```

Unfortunately, the feature code doesn't work completely. Actually, the letters are well grouped, but then I had a bug in the frequency analysis code which I couldn't fix.

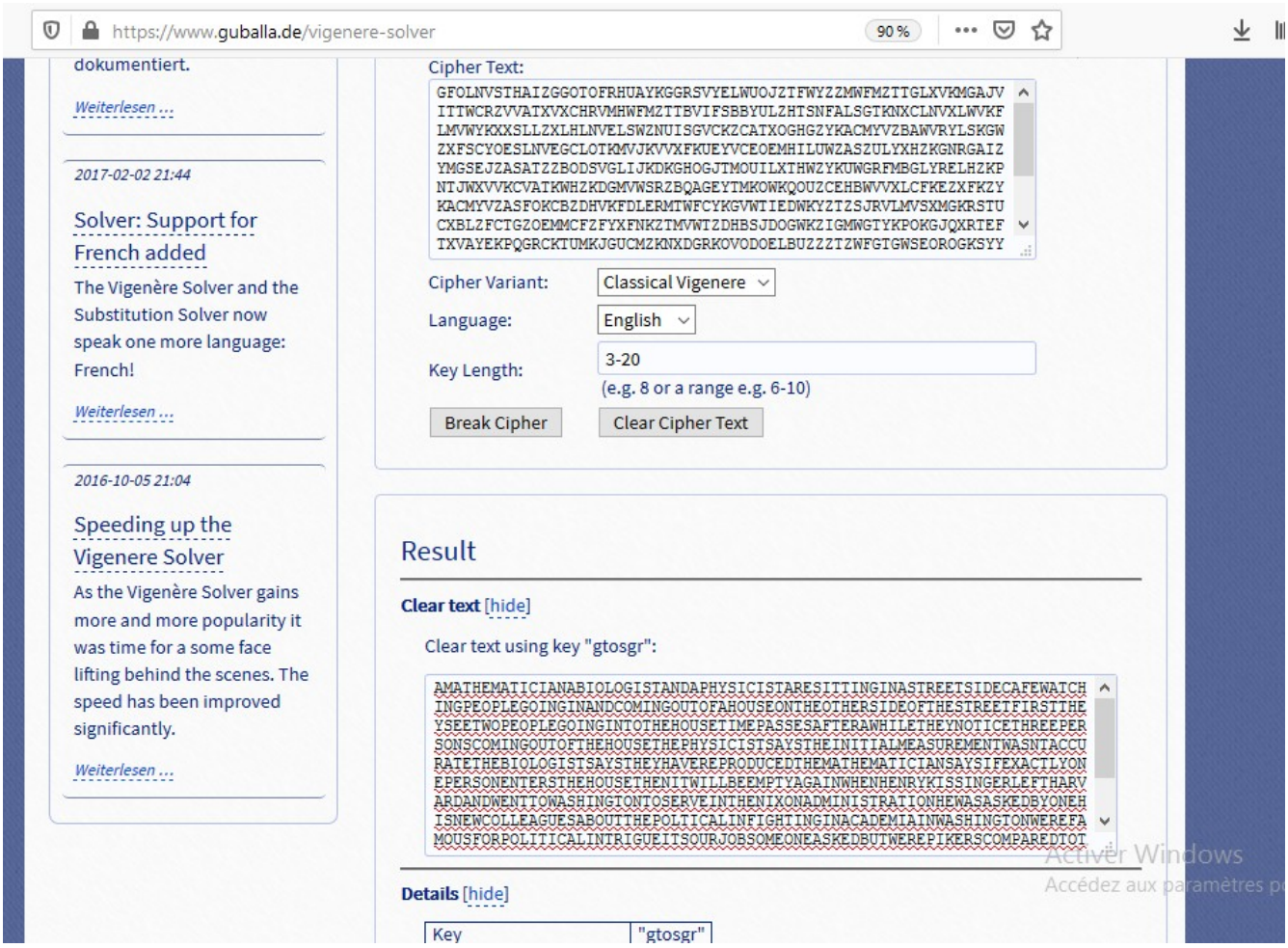
VERIFICATION AND CONCLUSION

As my analysis code doesn't work totally, I used an online automated tool for this analysis (available on https://www.simonsingh.net/The_Black_Chamber/vigenere_cracking_tool.html).



The outcome of the dividers is quite similar to the one I found with my own developed tool. And we also have 2, 3 and 6 as the most probable key lengths.

Then, I used a German automated tool available on the Internet (on https://www.guballa.de/vigenere-solver) to break the cipher. I indicated a key length between 3 and 20, just in case to be sure.



The tool broke the code, and the key found is ‘GTOSGR’, which has a length of 6 characters. This confirms the result found previously.

To conclude, using this key we obtain the following English text (after adding spaces and punctuation):

A MATHEMATICIAN, A BIOLOGIST AND A PHYSICIST ARE SITTING IN A STREET SIDE CAFE WATCHING PEOPLE GOING IN AND COMING OUT OF A HOUSE ON THE OTHER SIDE OF THE STREET. FIRST THEY SEE TWO PEOPLE GOING IN TO THE HOUSE. TIME PASSES AFTER A WHILE, THEY NOTICE THREE PERSONS COMING OUT OF THE HOUSE. THE PHYSICIST SAYS THE INITIAL MEASUREMENT WASN’T ACCURATE. THE BIOLOGIST SAYS THEY HAVE REPRODUCED. THE MATHEMATICIAN SAYS IF EXACTLY ONE PERSON ENTERS THE HOUSE THEN IT WILL BE EMPTY AGAIN.
WHEN HENRY KISSINGER LEFT HARVARD AND WENT TO WASHINGTON TO SERVE IN THE NIXON ADMINISTRATION, HE WAS ASKED BY ONE HIS NEW COLLEAGUES ABOUT THE POLTICAL IN FIGHTING IN ACADEMIA IN WASHINGTON WERE FAMOUS FOR POLITICAL INTRIGUE. IT SOURJOB SOMEONE ASKED BUT WERE PIKERS COMPARED TO THE BACKSTABBING AND DIRTY POLITICS AT UNIVERSITIES. WHY DO YOU PEOPLE FIGHT LIKE THAT ? KISSINGER IS SAID TO HAVE RESPONDED IN HIS LOW GRAVELLY VOICE: IT’S BECAUSE THE STAKES ARE SO LOW.