צופן vigenere

צופן [ויז'ינר (vigenere)](https://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher) הינו מורכב יותר מהצפנים שלמדנו עד כה. במשך כ-300 שנים הוא היה נחשב כצופן לא ניתן לפיענוח. הצופן הומצא ע"י הקריפטוגרף האיטלקי ג'ובאני בטיסטה בלאסו. מספר שנים לאחר מכן הצופן הוצג ע"י הדיפלומט הצרפתי חובב הקריפטוגרפיה בלז דה ויז'ינר למלך צרפת. לאחר שבמאה ה-19 נמצאה הדרך לפרוץ את הצופן, הוא יוחס בטעות לויז'ינר ומכאן קיבל את שמו.

הצופן מבוסס על צופן קיסר, אך במקום שיהיה מפתח יחיד שאיתו נצפין את כל ההודעה, המפתח מורכב מרצף של מספר אותיות אשר חוזר על עצמו לאורך ההצפנה. עבור כל אות מתבצעת הצפנת קיסר כאשר המפתח הינו ההיסט של האות במתאימה במפתח. לפני עידן המחשבים היה נהוג להשתמש בטבלה הבאה:

|  | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z |
| b | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a |
| c | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b |
| d | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c |
| e | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d |
| f | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e |
| g | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f |
| h | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g |
| i | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h |
| j | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i |
| k | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j |
| l | l | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k |
| m | m | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l |
| n | n | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m |
| o | o | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n |
| p | p | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o |
| q | q | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p |
| r | r | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q |
| s | s | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r |
| t | t | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s |
| u | u | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t |
| v | v | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u |
| w | w | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v |
| x | x | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w |
| y | y | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x |
| z | z | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x | y |

לדוגמה אילו נרצה להצפין את ההודעה “el el hapoel” בעזרת המפתח “lemon”. נתחיל בהצבת המפתח באופן חוזר מתחת להודעה, כאשר תווים שאינם אותיות אינם מוצפנים. ע"מ להצפין נמצא את העמודה של האות שאנו רוצים להצפין והאות המוצפנת תמצא בשורה של האות המתאימה במפתח.

כאשר נצפין את ‘e’ בעזרת ‘l’ נסתכל בעמודה ‘e’ ובשורה ‘l’ ונראה שהאות בהצפנה היא ‘p’.

כאשר נצפין את ‘l’ בעזרת ‘e’ נסתכל בעמודה ‘l’ ובשורה ‘e’ ונראה שהאות בהצפנה היא ‘p’.

| e | l |  | e | l |  | h | a | p | o | e | l |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| l | e |  | m | o |  | n | l | e | m | o | n |
| p | p |  | q | z |  | u | l | t | a | s | y |

הפיענוח מתקיים בדרך הפוכה. ניגשים לשורה של המפתח, מאתרים את האות מהצופן, האות בהודעה הינה האות בראש העמודה.

לדוגמה עבור אות המפתח ‘l’ נמצא את העמודה של אות הצופן ‘p’ ונראה שהיא שייכת לעמודה של האות ‘e’ ועל כן ‘e’ הינה האות בהודעה המקורית.

| l | e |  | m | o |  | n | l | e | m | o | n |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| p | p |  | q | z |  | u | l | t | a | s | y |
| e | l |  | e | l |  | h | a | p | o | e | l |

1. כתבו קוד ב-python אשר מקבל מהמשתמש את ההודעה להצפנה ואת המפתח. על התוכנית להדפיס את ההודעה המוצפנת.
2. כתבו קוד ב-python אשר מקבל מהמשתמש את ההודעה המוצפנת ואת המפתח. על התוכנית להדפיס את ההודעה המפוענחת.
3. מזגו את שתי התוכניות לתוכנית אחת המקבלת ב-sys.argv את מוד הריצה:
   1. Encrypt – להצפנה.
   2. Decrypt לפיענוח.
4. פענחו את ההודעה הבאה בעזרת המפתח ‘GRIMM’:

Uekq gvfv m fodm ftkim imy r lqmx cqffrv oudr npa igj tahku jk qbvzkatv eta rfwwqj rb tqx, scf yujb ar gct nk nvz sdgelyazymd, mtu btqxv eme tfbtutx btmz jpq iultp zuk pmhk xqhqt kw ftk tpuxj. Fvoq yym smbv pqd g cqffrv paaj fn dqj mmxhkk, etuiy aguzvl tqx jw iqrc btmz jpq iultp zkmmd ikrz mzekpuzm vteq; yf atq cra mxcrge ogctqp 'Rzbfxk Imp-Douqzs-Nfwp.' Atv lmk nvz yazymd egzl fa nvz: 'Oasv, Tufzcm Dqj-Iqputx-Paaj, ymdq oj i buktm ar irsq mtu i nazktq al nqzq; zrsq fnvu fa efcd sxrvpyukpqd, yym ue oct mzj nmmw, gel ftkp euxr uw tqx xwap. Yvb agz smraxv qf skka taz, rvp invv kaa rzq suzvs, igcs zuivtk mtu ygukktk mtu la zuk zgz uwn ftk gift, ui gag srg rmrc izp himmw zym nazktq, mtu btqt pwgd miizpsfbtqx nqxx mvb zazyqzs; gel itke gag mf qzfu ymd dufu, pat'k nadmvb fa yrg, 'Sauu uadtzvs', mtu laz'z gmqb oeba qbvzk ouivqd hvnadk pwg pu zb. 'U ioct fmqv odqgk kmdk,' jiup Rzbfxk Imp-Douqzs-Nfwp fu ymd yukpqd, gel smbv pqd nrvp at zb. Ftk xzmzjdwftki tuhku wgf oe btq cfwp, tgcn m xkrogq liwy fnv duxrroq, mtu rgez ra Xuzktq Dku-Zupoeo-Tauu mzfkimp fnv eaaj, r eaxl dmf tki. Zqp-Xzluzm-Ywap jzl zaz bvai cyif m czkwqj tzqmzlzq tk nie, mtu eme tfb mf gct mrxrqp al yqy. 'Sufl pme, Cqffrv Zqp-Xzluzm-Ywap,' yrqp tk. 'Kpmzq pwg woelxk, cftr.' 'Inzbtqx remk yf mmdrp, Tufzcm Dqj-Iqputx-Paaj?' 'Kw yk miizpsfbtqx'j.' 'Etmz yihq efc saz zv kaai ibdue?' 'Kmwk rvp ioem; kqykmdpgp eme hrsuzm-uik, eu gwad yzkw sxrvpyukpqd oj ba tgmm easvbtutx oaaj, kw ymqv pqd ykzazmvz.' 'Itkim pakj gagx xzmzjdwftki tuhk, Cqffrv Zqp-Xzluzm-Ywap?' 'G xwap wlidfki wr m rvisgk widfnvz az oe btq cfwp; tki pagyv afmtua gzjvz ftk kpdqk cidsk fiw-fxvme, fnv vgf-zimqe gim vgyk jqxun; gag ylzqxe dcef qewi uz,' imbxovl Xuzktq Dku-Zupoeo-Tauu.

העשרה – פריצת צופן vigenere

אחת הבעיות הנפוצות בצופן ויז'ינר הינה המפתח, על מנת לא לשכוח את הצופן, הייתה לרבים נטייה להשתמש במפתח שהוא לא סתם אוסף של אותיות, אלא מילה אמיתית בשפה האנגלית. על כן, קל לבנות תוכנית אשר משתמשת במילון המכיל את כל המילים באנגלית (מילונים אלו קיימים באינטרנט וישנן כ-45,000 מילים בשפה). לאחר שימוש במילה מסוימת כמפתח, ניתן לדרג את מילות ההודעה ולראות אם הן אכן שייכות לשפה האנגלית. במידה וכן המפתח שהיה בשימוש הינו מועמד להיות מפתח הצופן.

אך ישנה גם דרך מתמטית לפצח את הקוד, אשר עובדת גם במקרה והמפתח הוא רצף אקראי של אותיות.

בשנת 1854צ'ארלס באבג' (מתמטיקאי אשר נחשב בעיני רבים כאבי המחשבים המודרניים) הצליח להוכיח שהוא יודע לפצח את צופן ויז'ינר, אך הוא מעולם לא פרסם כיצד עשה זאת. רק לאחר מותו, מתוך עיון בכתביו הסתבר שהוא השתמש באלגוריתם שפורסם רק מאוחר יותר ע"י המתמטיקאי פרדיך קסיסקי.

**מבחן קסיסקי (שיטת קסיסקי)** הינו אלגוריתם שנועד לגלות את אורך המפתח.

**שלב ראשון** נשתמש בעובדה שבקוד ויז'ינר המפתח חוזר על עצמו ועל כן נחפש חזרות בצופן (בעלות 3 אותיות או יותר). בטקסט שכיח שישנם חזרות של מילים ועל כן יש סיכוי שמילים אלו יוצפנו ע"י אותו מפתח ולכן נזהה גם חזרות בקוד המוצפן. לאחר שנזהה את החזרות, נבדוק את המרחק בין הרצפים, מאחר שהנחנו שהמסר בחזרות הוצפן עם אותו מפתח, המרחק בין החזרות הינו כפולה של המפתח.

ניקח לדוגמה את הטקסט:

crypto is short for cryptography

אנו רואים שהמילה crypto חוזרת על עצמה.

נצמצם את הרווחים ונצפין עם המפתח ‘abcd’:

| c | r | y | p | t | o | i | s | s | h | o | r | t | f | O | r | c | r | y | p | t | o | g | r | a | p | h | y |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a | b | c | d | a | b | c | d | a | b | c | d | a | b | C | d | a | b | c | d | a | b | c | d | a | b | c | d |
| c | s | a | s | t | p | k | v | s | i | q | u | t | g | q | u | c | s | a | s | t | p | i | u | a | q | j | b |

אנו רואים שהרצף ‘csastp’ חוזר על עצמו והמרחק בין התחלות הקצף הינו 16 תווים. כפי שציינו, המרחק הינו כפולה של אורך המפתח, על כן אורך המפתח יכול להיות 1,2,4,8,16. מפתחות באורך 1 או 2 אינם סבירים ולכן נשארנו עם אורך מפתח של 4,8 או 16. ככל שהטקסט ארוך יותר, יש סיכוי רב יותר למצוא חזרות נוספות, למשל בצופן הבא:

**VHV**SSP**QUCE**MRVBVBBB**VHV**SURQGIBDUGRNICJ**QUCE**RVUAXSSR

ישנן שתי חזרות:

* **QUCE** – חוזר במרווח של 30 כך שגודל המפתח האפשרי הוא 3,5,6,10,15,30.
* **VHV** - חוזר במרווח של 18 כך שגודל המפתח האפשרי הוא 3,6,9,18.

במקרה כזה נבנה טבלה שסופרת את מספר החזרות של כל גודל מפתח ונמצא את האורכים שחוזרים הכי הרבה פעמים והם יהיו המועמדים לגודל המפתח. בדוגמה הנ"ל 3 ו-6.

ישנה שיטה נוספת למציאת האורך המפתח היא ע"י שכפול הצופן וכתיבתו שורה מתחת לשורה. בכל פעם נזיז בצורה ציקלית את הצופן בשורה התחתונה תו אחד ימינה. לאחר כל הזזה נספור את המספר התווים הזהים שנמצאים אחד מעל השני. ההזזה שמכילה מספר רב ביותר של תווים זהים, היא מכפלה של אורך המפתח, כך שהמפתח הוא אחד המחלקים של מספר זה.

**שלב שני** –נבנה רשימה באורך המפתח, איבר עבור כל אות במפתח. כל איבר ברשימה יכיל את האותיות הצופן שהוצפנו ע"י אותה אות במפתח. לדוגמה, בצופן שלנו, עבור מפתח באורך 4 נבנה את הרשימה הבאה:

[  
‘ctstcta’,  
‘spigspq’,  
‘akqqaij’,  
‘svuusub’

]

במידה ואכן זיהינו את אורך המפתח הנכון (4) אזי:

* כל אותיות האיבר הראשון ברשימה הוצפנו עם אותה אות.
* כל אותיות האיבר השני ברשימה הוצפנו עם אותה אות.
* כל אותיות האיבר השלישי ברשימה הוצפנו עם אותה אות.
* כל אותיות האיבר הרביעי ברשימה הוצפנו עם אותה אות.

**שלב שלישי** – נשתמש בעובדה שעבור כל שפה ידועה לנו שכיחות ההופעה של כל אות. עבור כל איבר ברשימה, ננסה לפענח אותו עם כל האותיות האפשריות (26 אפשרויות באנגלית) ונמצא את המחרוזת עם שכיחויות הכי קרובות לשפה האמיתית. Chart, bar chart

Description automatically generated with medium confidence

למפתח שבעזרתו נקבל את הפיענוח שהכי קרוב לטבלת השכיחויות, יש את הסיכוי המירבי להיות מפתח ההצפנה.

יתכן ושלב זה ימצא מספר תוצאות קרובות, נסנן אותן בשלב הבא.

כיצד נמצא את הפיענוח ששכיחות האותיות בו הכי קרוב לשכיחות בשפה? קיימות מספר שיטות ואנו נציג 2 מהן.

1. נחשב את שכיחות האותיות בהודעה. נשתמש בסטיית התקן ע"מ לחשב את המרחק שבין שכיחות האותיות בהודעה לשכיחות בשפה. הפיענוח הטוב ביותר ייתן את סטיית התקן הקטנה ביותר.
2. מהסתכלות על טבלת השכיחויות אנו יכולים לראות שהאותיות E,T,A,O,I,N הן האותיות השכיחות ביותר באנגלית ואילו האותיות Z,J,Q,X,K,V הן בעלות השכיחות הנמוכה ביותר בשפה. ננסה לתת לכל פיענוח ציון אשר יסמן את קרבתו של הפיענוח לשפה האנגלית. נעשה זאת בשלבים:
   1. נעבור על הפיענוח ונספור את מספר המופעים של כל אות. נשמור את התוצאה במילון.
   2. נסדר את הפיענוח לפי סדר השכיחויות.
   3. נעבור כל 6 האותיות השכיחות ביותר בפיענוח, על כל אות מתוך E,T,A,O,I,N נוסיף נקודה לציון של הפיענוח. נעבור על 6 האותיות הכי פחות שכיחות בפיענוח, על כל אות מתוך Z,J,Q,X,K,V נוסיף נקודה לציון של הפיענוח.
   4. הפיענוח שיקבל את הציון הגבוהה ביותר הוא בעל הסיכויים הטובים ביותר להיות הפיענוח הנכון שפוענח בעזרת אות הצופן הנכונה.

לדוגמה אם פיענוח עבור אות צופן מסוימת ייצר לנו את המחרוזת

AAEKDVLSRMHFBZEHXLMOILIAEJJBCJJJINAIEPPXEZSGOFMMYMLHRJBJMSLSPIPXNGEYZAEFYAYYOFXCLXMAEXNABOAAJNCNRBHLHAGZIFHXXXRYLNISJSSXIRSHXSZRSSOPPPSRRJIRACOARACCPOHVBANCCFYSRRKKBG

נסדר את הפיענוח לפי מספר ההופעות של כל אות (A מופיעה הכי הרבה פעמים, לאחר מכן S ו-D מופיעה הכי פחות פעמים בטקסט). סידור המחרוזת לפי שכיחויות ייתן את המחרוזת הבאה:

| **A** | S | R | X | J | **I** | P | C | H | L | E | Y | N | O | B | M | F | **Z** | G | **K** | **V** | D |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E | T | A | I | O | N |  |  |  |  |  |  |  |  |  |  | V | K | X | Q | J | Z |

הציון של פיענוח זה יהיה 5 (האותיות A ו-I מופיעות ב-6 האותיות בעלות מספר המופעמים הגדול ביותר בטקסט ואילו האותיות Z,K ו-V מופיעות ב-6 האותיות בעלות מספר המופעים הנמוך ביותר בטקסט).

**שלב רביעי** – יתכן ונגיע לשלב זה עם מפתח אחד, אך מרב הסיכויים שנגיע עם מספר אפשרויות. כפי שראינו בצופן קיסר, ניתן לעבור לזיהוי אנושי אך ניתן גם להשתמש בנתונים נוספים על השפה.

בשפה האנגלית קיימים כ-45,000 מילים, נוכל למצוא רשימה המכילה את כל המילים הללו בכתובת <https://inventwithpython.com/dictionary.txt>.

נוריד את המילון ונשמור אותו כקובץ. בתוכנית הפריצה שלנו נטען את המילון, מומלץ לשמור את המילים כמפתחות של מילון (dictionary) וזאת מהסבה שיותר מהיר למצוא מפתח במילון מאשר חיפוש ברשימה (בשל פונקציית ה-hash שמילון מפעיל על המפתחות).

נוציא מההודעה את כל התווים שהם אינם אותיות או רווחים.

נחלק את המילים מההודעה לרשימה של מילים.

עבור כל מילה נבדוק אם היא נמצאת במילון של המילים בשפה האנגלית או לא.

בסופו של דבר ניתן להודעה ציון לפי אחוז המילים שבה המופיעות במילון השפה האנגלית.

ההודעה עם הציון הגבוהה ביותר היא הפיענוח הנכון של הצופן.

אז עכשיו שיש לכם את כל הכלים להצפין ולפרוץ קוד ויז'ינר, כתבו תוכנית המצפינה הודעה ולאחר מכן מנסה לפרוץ את ההצפנה ולגלות את מפתח ההצפנה. זכרו, ככל שההודעה ארוכה יותר, הסיכויים לפרוץ אותה גדולים יותר.