8.4 Polyalphabetische Verschlüsselung: Vigenère

Die Antwort auf die Häufigkeitsanalyse liegt nahe: Man benutzt nicht ein Geheimalphabet, sondern mehrere im Wechsel. Das bekannteste dieser Verfahren veröffentlichte der französische Diplomat Blaise de Vigenère 1586.

Unter den Klartext wird ein **Schlüsselwort** geschrieben und so oft wiederholt, bis der Text zu Ende ist. Jeder Buchstabe des Schlüsselworts gibt an, um wie viele Stellen der darüberstehende Klartextbuchstabe verschoben wird (A = 0, B = 1, …).

```
Klartext:   K  R  Y  P  T  O  G  R  A  P  H  I  E
Schlüssel:  H  A  L  L  O  H  A  L  L  O  H  A  L
Geheimtext: R  R  J  A  H  V  G  C  L  D  O  I  P
```

Von Hand liest man die Geheimtextbuchstaben im **Vigenère-Quadrat** ab: Die Zeile gibt der Schlüsselbuchstabe vor, die Spalte der Klartextbuchstabe; im Kreuzungspunkt steht der Geheimtextbuchstabe. Für das erste Zeichen des Beispiels sucht man also die Zeile H und darin die Spalte K und findet R.

```
    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
```

Jede Zeile des Quadrats ist nichts anderes als ein um eine Stelle weiter verschobenes Alphabet – also eine Cäsar-Chiffre. Das Vigenère-Verfahren besteht demnach darin, im Takt des Schlüsselworts zwischen 26 Cäsar-Chiffren hin- und herzuspringen. Genau diese Beobachtung wird uns beim Angriff in Abschnitt 8.4.1 wieder begegnen.

Rechnerisch ist es eine Addition mit Rest, nur dass sich der Summand von Buchstabe zu Buchstabe ändert:

$$c_i = (m_i + s_{i \bmod l}) \bmod 26$$

Dabei ist l die Länge des Schlüsselworts.

Der Gewinn ist deutlich: Gleiche Klartextbuchstaben werden je nach Position verschieden verschlüsselt, die Häufigkeitsverteilung des Geheimtextes wird flach. Drei Jahrhunderte lang galt das Verfahren als unknackbar – „le chiffre indéchiffrable“.

8.4.1 Der Kasiski-Test

1863 zeigte der preußische Major Friedrich Wilhelm Kasiski, wo die Schwäche liegt: Der Schlüssel **wiederholt sich**. Trifft dieselbe Klartextstelle zufällig auf dieselbe Stelle des Schlüsselworts, entsteht dieselbe Geheimtextstelle. Der Abstand solcher Wiederholungen ist deshalb ein Vielfaches der Schlüssellänge.

Daraus ergibt sich der Angriff in drei Schritten:

1. Im Geheimtext wiederholte Buchstabenfolgen (Länge 4 oder 5) suchen und ihre Abstände berechnen.
2. Den größten gemeinsamen Teiler der Abstände bilden – das ist die vermutete Schlüssellänge l. Ein einzelner Abstand genügt nicht, denn er ist nur ein Vielfaches von l.
3. Den Text in l Spalten zerlegen. Jede Spalte ist mit *einem* festen Alphabet verschlüsselt, also eine gewöhnliche Cäsar-Chiffre – und gegen die hilft die Häufigkeitsanalyse aus 8.3.4.

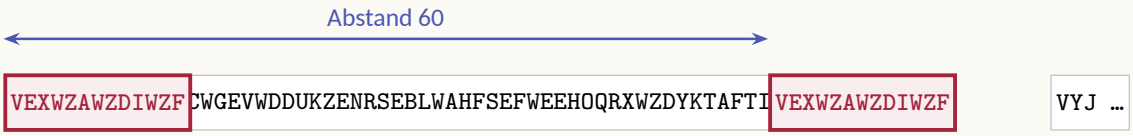
Beispiel: Ein Geheimtext, vier Abstände

Der folgende Text wurde mit einem Schlüsselwort verschlüsselt, dessen Länge wir nicht kennen:

```
  0  VEXWZAWZDIWZFDCWGEVWDDUKZENRSEBLWAHFSEFWEEHOQRXWZDYKTAFTII
 58  LVVEXWZAWZDIWZFVYJECBDGEMKQLNOQRXAQNUUTRCUTTFAQSNKUEBLZULT
116  GCBKFAVWZSUDMTQWDDYFECBDGEMKQLBSFSCWTTXAQNUUTRCUTTMGQIHXMC
174  BAETXWDGLMZDAWPAHCQUHVEOMUTWYJUSNKQIHWGMMWFZOFSIHVQRJJMXCK
```

Vier Folgen kommen doppelt vor:

Folge	Positionen	Abstand
VEXWZAWZDIWZF	0 und 60	60
ECBDGEMKQL	76 und 136	60
XAQNUUTRCUTT	90 und 154	64
OQRX	43 und 87	44



Dieselbe Folge an zwei Stellen. Der Abstand muss ein *Vielfaches* der Schlüssellänge sein — welches, verrät ein einzelner Abstand nicht.

$$\text{ggT}(60,\ 64,\ 60,\ 44) = \mathbf{4}$$

erst mehrere Abstände legen die Schlüssellänge fest

Zwei Fundstellen derselben Folge und ihr Abstand.

Der größte gemeinsame Teiler von 60, 64, 60 und 44 ist **4**. Das Schlüsselwort hat also vier Buchstaben — es war `MAUS`. Mit nur einem Abstand wäre man nicht weitergekommen: 60 allein lässt 1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30 und 60 zu.

Merke

Die Vigenère-Chiffre ist nur so stark wie die Periode ihres Schlüssels. Ein sich wiederholender Schlüssel zerlegt den Text in mehrere einfache Chiffren.

1922 ergänzte William F. Friedman den **Friedman-Test**: Er schätzt die Schlüssellänge über den *Koinzidenzindex*, also die Wahrscheinlichkeit, dass zwei zufällig gezogene Buchstaben eines Textes gleich sind. Für deutschen Klartext liegt dieser Wert bei etwa 0,076, für zufällige Buchstaben bei $1/26 \approx 0{,}038$. Je länger der Schlüssel, desto näher rückt der Geheimtext an den Zufallswert.

Arbeitsauftrag 4

12. Verschlüsseln Sie `SICHERHEIT` mit dem Schlüsselwort `MAUS`.
13. Entschlüsseln Sie `DLGK TKQT` mit dem Schlüsselwort `TAG`.
14. Erklären Sie, warum ein Schlüsselwort mit lauter gleichen Buchstaben (z. B. `AAAA`) die Vigenère-Chiffre auf ein bekanntes Verfahren zurückführt. Auf welches?
15. Schreiben Sie ein Programm zur Ver- und Entschlüsselung nach Vigenère.

8.4.2 Werkstatt: ein Vigenère-Chiffrat knacken

Der folgende Text ist deutsch, mit einem unbekannten Schlüsselwort verschlüsselt und 509 Buchstaben lang. Umlaute stehen als AE/OE/UE, Satzzeichen und Leerzeichen sind entfallen. Die Zahl am Zeilenanfang gibt die Position des ersten Buchstabens der Zeile an – damit lassen sich Abstände ablesen, ohne nachzuzählen.

```
  1  NIWKZ SNPRW DGXDS GWVWB MFEJA URTEE KAWVB TEVMF LCMHK BSLMX
 51  MABWS VOMSV VOZON STHRI LELCF NDORO QKMPD SGLAO PKEMN AQSOI
101  LMFNM CKDIJ DMHTW AARZO SUPOK MMHWT AOODG WVSEL KRAVV KACSA
151  KZNIC SUPDE MCSWT OYZDW AMVOZ RODBO OZNIW TSOVQ EVMKC KRLMM
201  KCMVS CMFXB NEJHW BTOGL LWXBO XLQFC XKLLM FEVNK FIUUB TEVMW
251  SVJED VWCXK LLMES BNEJO WGWRN LMFRI OUXQY UMSTK IFKTI SWLWB
301  XBEMA KSAMH WWXPQ JIWZX BQODJ QURSK SAACS DORGM XPMXT DQURB
351  ODAMK OVKNY ZAPNL EJMAD ASMBI ZBIMH LHWRV RUFLW BBNRW QMXLC
401  EUPRS OORKC URBOI EKZSN PRSBF KKRWA MVOZR ODBWX JECZA LKJON
451  XWDQM XUFLT OADIE ULOIE SVMJO VKBKB SOVNE FLAOT KEFOW NMCSU
501  PDEMC SWTK
```

Dass die einfache Häufigkeitsanalyse hier nicht weiterhilft, zeigt schon ein Blick auf die Verteilung: Der häufigste Buchstabe des Chiffrats ist M mit 7,7 %, gefolgt von O mit 7,3 % und gleich drei Buchstaben mit 6,1 %. Im Deutschen läge das E dagegen bei 17,4 %. Die Verteilung ist flach – aber, wie Kasiski gezeigt hat, nicht strukturlos.

16 Suchen Sie im Chiffrat Buchstabenfolgen der Länge 5, die zweimal vorkommen, und notieren Sie mindestens vier Funde mit ihren Positionen und dem Abstand. Bestimmen Sie zu jedem Abstand *alle* Teiler und markieren Sie die Zahl, die in jeder Zeile vorkommt. Begründen Sie, warum ein einzelner Fund nicht genügt: Welche Schlüssellängen ließe Ihr kleinster Abstand noch offen?

Legen Sie dafür im Heft eine Tabelle mit den Spalten *Fundstelle*, *Position 1*, *Position 2*, *Abstand* und *alle Teiler des Abstands* an.

17 Zerlegen Sie das Chiffrat in m Spalten und bestimmen Sie für jede Spalte den häufigsten Buchstaben. Nehmen Sie an, dass dies das verschlüsselte E ist, und berechnen Sie daraus den Schlüsselbuchstaben. Arbeiten Sie arbeitsteilig – je Gruppe eine Spalte.

Tabellenspalten: *Spalte*, *Anzahl Buchstaben*, *häufigster Buchstabe*, *Verschiebung*, *Schlüsselbuchstabe*.

18 Notieren Sie das Schlüsselwort. Fällt Ihnen an seinen Buchstaben etwas auf, das Ihre Rechnung bestätigt? Entschlüsseln Sie anschließend die ersten beiden Zeilen von Hand und setzen Sie Wortgrenzen. Der Klartext verrät, wem Sie dieses Verfahren verdanken.

19 Automatisieren Sie den Angriff im Notebook *vigenere-info12.ipynb*: Häufigkeiten zählen, Wiederholungen und Abstände finden, ggT bilden, Spalten auswerten, entschlüsseln.

20 *Für Schnelle:* Lassen Sie Ihr Programm alle Schlüssellängen von 1 bis 12 durchprobieren und jedes Ergebnis über die Abweichung von der deutschen Buchstabenverteilung bewerten (χ^2). Der Angriff läuft dann ganz ohne Vorwissen: Bei diesem Text schneidet die richtige Länge rund zwanzigmal besser ab als die nächstbeste. Beachten Sie dabei eine Falle — auch jedes *Vielfache* der richtigen Länge liefert einen guten Wert. Warum?