

Häufigkeitsanalyse

Knacken monoalphabetischer Verschlüsselungen

Monoalphabetische Verschlüsselungsverfahren ordnen jedem Klartextzeichen genau ein Geheimtextzeichen zu. Beispiele für eine solche Zuordnung hast du bereits kennengelernt (s. Tabelle 1).

Klartext- zeichen	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
Geheim- textzeichen	G	M	K	Q	W	T	L	C	S	H	E	Y	N	D	Z	O	A	X	I	V	P	F	R	B	U	J

Tabelle 1: Beispiel für eine monoalphabetische Verschlüsselung

Wenn man die Zuordnungstabelle nicht kennt, ist es auf den ersten Blick ziemlich schwierig den verschlüsselten Text zu lesen. Denn es würde viel zu lange dauern, die 403 Trilliarden möglichen Zuordnungen alle auszuprobieren.

Bei einem längeren verschlüsselten Text können wir aber ausnutzen, dass in jeder Sprache manche Buchstaben besonders häufig vorkommen. In der deutschen Sprache ist der häufigste Buchstabe das *e* und der zweithäufigste das *n*. Da das *e* bei einer monoalphabetischen Substitution immer durch das gleiche Zeichen ersetzt wurde, wird dieses Geheimtextzeichen im Geheimtext ebenfalls das häufigste Zeichen sein. Die Häufigkeit der Klartextzeichen überträgt sich also auf die Geheimtextzeichen.

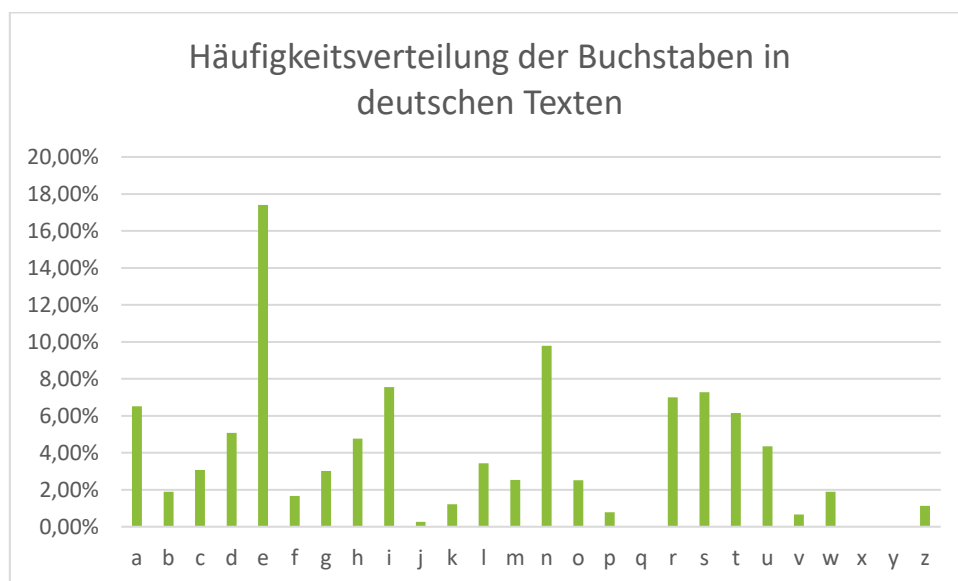


Abbildung 1: Häufigkeitsverteilung der Buchstaben in deutschen Texten¹

Aufgabe 1: Angenommen ein längerer deutscher Text wurde mit der Zuordnung in Tabelle 1 verschlüsselt. Stelle eine Vermutung auf, ...

- ... welches Zeichen in dem Geheimtext am häufigsten vorkommt.
- ... welche fünf bis sechs Zeichen sehr häufig vorkommen.
- ... welche Zeichen kaum oder gar nicht vorkommen.

¹ Quelle: A. Beutelspacher (2009). *Kryptologie*. 9. Aufl. Wiesbaden: Vieweg+Teubner.

Indem wir das häufigste und das zweithäufigste Geheimtextzeichen finden, können wir also diese Zeichen schon mal durch das *e* bzw. das *n* ersetzen. Damit haben wir schon etwa ein Viertel des Textes entschlüsselt! Wir erhalten einen Lückentext, in dem wir die restlichen Zeichen durch intelligentes Raten zuordnen können. Dazu können wir uns zum einen die Häufigkeiten der übrigen Zeichen anschauen, um herauszufinden, ob es sich um einen eher häufigen Buchstaben wie *a*, *i*, *r*, *s* oder *t* handelt oder um einen sehr seltenen wie z. B. *q*, *x* oder *y*. Zum anderen finden wir Wörter, in denen die Lücken nur mit bestimmten Buchstaben gefüllt werden können, damit sie Sinn ergeben. In dem Wort *e?n* wird das Fragezeichen höchstwahrscheinlich für das *i* stehen. In dem Wort *?er* kann das Fragezeichen nur für die Buchstaben *d*, *h*, *p* oder *w* stehen.

Häufigkeitsanalyse mithilfe des Rechners

Aufgabe 2: Durch das oben beschriebene Verfahren stehen die Chancen eine monoalphabetische Verschlüsselung zu knacken schon viel besser.

Versuche mithilfe der Schritte 1 bis 5 den Geheimtext in der Datei *geheimtext1* zu knacken. Du lernst dabei, wie du ein Textverarbeitungsprogramm einsetzen kannst, um dir die Arbeit noch ein wenig zu erleichtern.

1. Schritt

Lege dir auf einem Blatt Papier oder am Rechner eine Tabelle mit drei Zeilen an. In der ersten Zeile stehen alle Geheimtextzeichen. In der zweiten Zeile wird in Schritt 2 für jedes Zeichen eingetragen, wie oft es in dem Geheimtext vorkommt. In der dritten Zeile werden später die passenden Klartextzeichen eingetragen.

Geheimtextzeichen	A	B	C	...	Z
Anzahl					
Klartextzeichen					

2. Schritt

Öffne die Datei *geheimtext1* in einem Textverarbeitungsprogramm. Dort hast du die Möglichkeit, Zeichen zu ersetzen. Wenn du die Option *alle ersetzen* wählst, wird gezählt, wie viele Ersetzungen vorgenommen wurden (s. Abbildung 2). Indem du ein Zeichen durch sich selbst ersetzt, also z. B. das A durch A, findest du heraus wie viele As in dem Text enthalten sind, ohne den Text zu verändern. Führe die Ersetzung durch sich selbst für jedes Geheimtextzeichen aus und trage die jeweilige Anzahl in deiner Tabelle in der Zeile *Anzahl* ein.

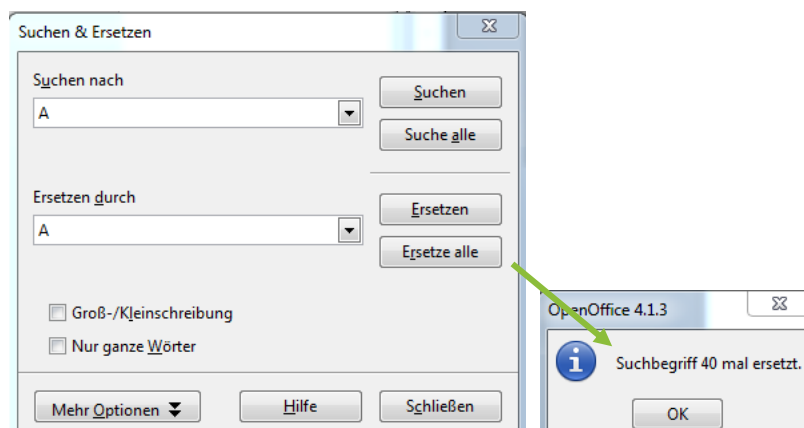


Abbildung 2: Ersetzen eines Zeichens durch sich selbst

3. Schritt

Vergleiche die Anzahl der Geheimtextzeichen miteinander und trage in deiner Tabelle die Klartextzeichen e und n in der passenden Spalte ein.

4. Schritt

Ersetze in dem Geheimtext das häufigste und das zweithäufigste Geheimtextzeichen durch e bzw. n . Bei den **Geheimtext**zeichen handelt es sich um **Groß**buchstaben. Verwende für die **Klartext**zeichen **Klein**buchstaben, damit du das Geheimtextzeichen E nicht mit dem Klartextzeichen e verwechselst.

Setze deshalb auch den Haken bei Groß-/Kleinschreibung.

Wenn z. B. das P der häufigste Geheimtextbuchstabe wäre, würdest du ihn durch das kleine e ersetzen:

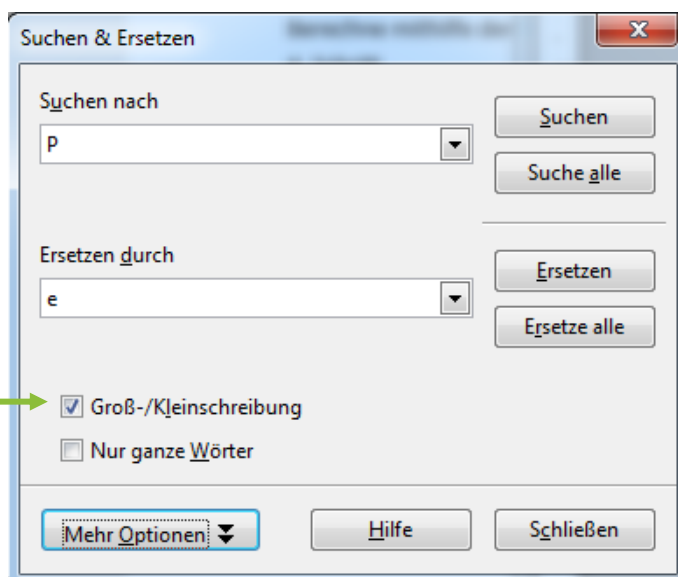


Abbildung 3: Berücksichtigung von Groß- und Kleinschreibung beim Ersetzen

5. Schritt

Ordne nun durch intelligentes Raten die übrigen Klartextzeichen zu. Trage jedes Klartextzeichen, das du einem Geheimtextzeichen zugeordnet hast, in deine Tabelle ein und ersetze es im Text.

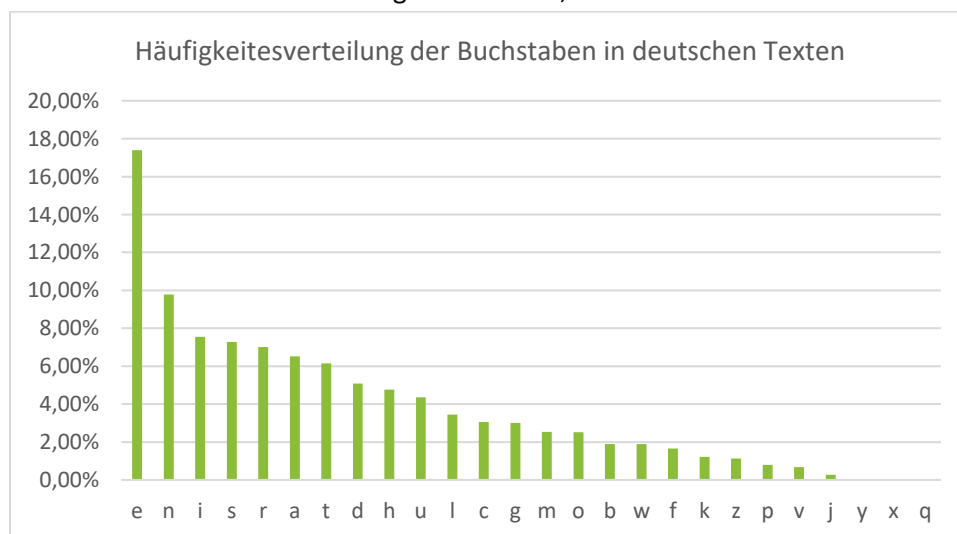


Abbildung 4: Häufigkeitsverteilung sortiert nach Häufigkeiten

Aufgabe 3 (für Schnelle):

- a) In der Datei *geheimtext2* ist ebenfalls ein monoalphabetisch verschlüsselter Text enthalten. Diesmal wurden allerdings andere Zeichen als Buchstaben für die Geheimtextzeichen verwendet. Führe für diesen Text ebenfalls eine Häufigkeitsanalyse durch und versuche ihn zu knacken. Die Satzzeichen Punkt, Komma, Ausrufezeichen und Fragezeichen haben ihre Bedeutung behalten und codieren keine Buchstaben.
- b) In der Datei *geheimtext3* ist ein vergleichsweise kurzer Geheimtext enthalten. Versuche auch diesen Text zu knacken.

Aufgabe 4: Erstelle eine Liste von Voraussetzungen, die ein verschlüsselter Text erfüllen muss, damit er mithilfe einer Häufigkeitsanalyse geknackt werden kann.

Aufgabe 5: Diskutiert, wie ein Ersetzungsverfahren sicherer gemacht werden kann, so dass ein Geheimtext nicht mehr so leicht mithilfe einer Häufigkeitsanalyse geknackt werden kann.

Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz](https://creativecommons.org/licenses/by-nc-sa/4.0/). Sie erlaubt Bearbeitungen und Weiterverteilung des Werks unter Nennung meines Namens und unter gleichen Bedingungen, jedoch keinerlei kommerzielle Nutzung.

