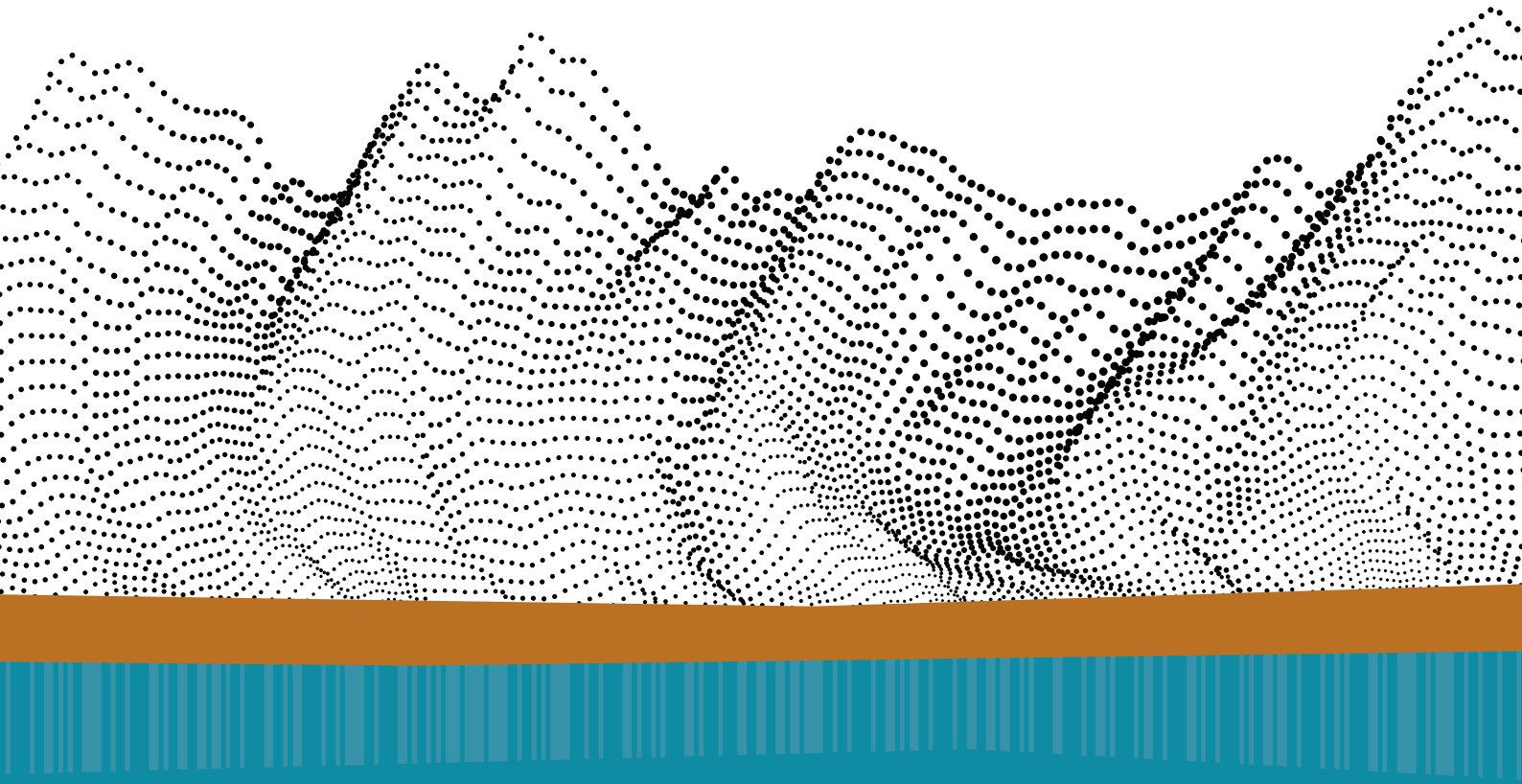
INFORMATIK

Daten verwalten, schützen und auswerten



Grundlagen der Informatik
für Schweizer Maturitätsschulen

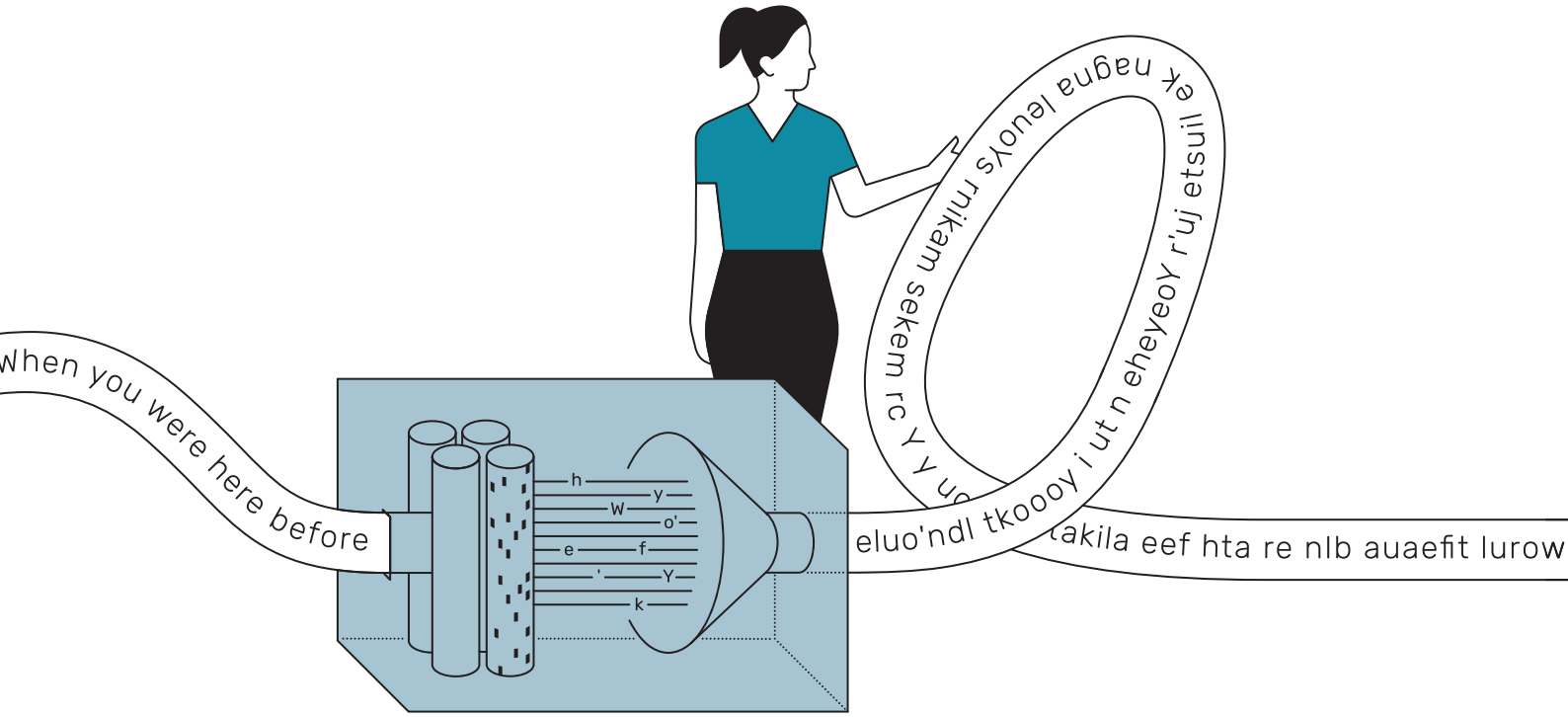
Mit Digital Book, Lösungen
und kostenloser Programmier-
umgebung TigerJython

Michael Barot, Britta Dorn, Ghislain Fourny,
Jens Gallenbacher, Juraj Hromkovič, Regula Lacher

Klett und Balmer Verlag

Klett

Inhaltsverzeichnis



2 Geheimschriften und Datensicherheit

Daten zu schützen gehört zu den Kernaufgaben der Informatik. Was bedeutet das genau? Es bedeutet, gespeicherte Daten vor dem Zugriff von Unbefugten zu schützen oder die in zugänglichen Daten enthaltenen Informationen geheim (unlesbar) zu halten. Darüber hinaus bedeutet es, Daten vor Beschädigung und Manipulation zu schützen. Dieser Schutz gilt sowohl für die auf einem Medium abgespeicherten Daten als auch für die Daten während des Transports von einem Ort zum anderen.

In diesem Kapitel fokussieren wir auf solche Darstellungen (Kodierungen) von Daten, die die Daten für Unbefugte unlesbar machen. Wir sprechen von der **Kryptologie**, der Lehre der Geheimschriften (von griechisch κρυπτός für «geheim, verborgen» und λογία für «Lehre»). Hier erlernen Sie Konzepte, mit denen Sie eigene Geheimschriften entwickeln können oder Geheimschriften von anderen knacken und somit Geheimnisse lüften können.

Geheimschriften der Antike

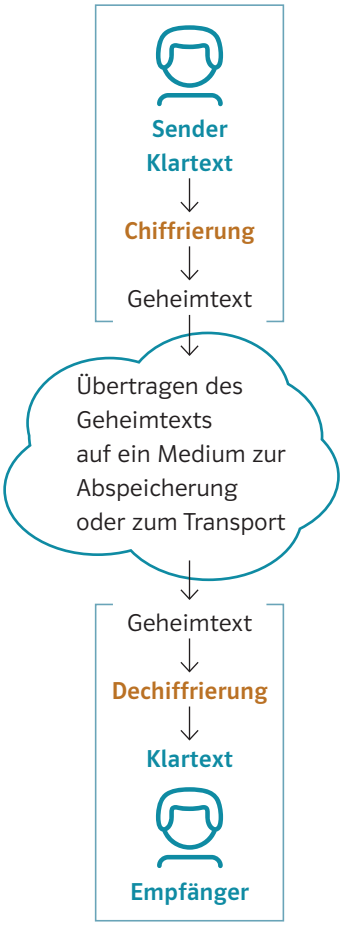
Jede Schrift ist ein System von grafischen Zeichen. Die Menge der verwendeten Zeichen (Symbole) in einem solchen System nennen wir **Alphabet**. Die Zeichen eines Alphabets bezeichnen wir auch als dessen **Buchstaben**. Die Gestalter oder Entwickler einer Schrift wählen oder entwerfen diese Zeichen und bestimmen somit auch das Alphabet. Wörter und Texte als Darstellungen (Kodierungen) von Informationen sind Folgen von Buchstaben des entworfenen Alphabets. Das Wort «kodieren» wird in der Informatik in unterschiedlichen Zusammenhängen verwendet. Kodieren bedeutet, eine Informationsdarstellung in eine andere umzuwandeln, aber auch eine ganz neue (erste) Darstellung für Informationen zu entwickeln. Ganz allgemein kann man unter «kodieren» die Entwicklung einer Schriftsprache verstehen. In diesem Kapitel verbinden wir die Kodierung mit den Umwandlungen von Informationsdarstellungen.

Die Kunst des Lesens und Schreibens wurde sehr lange nur von sehr wenigen Menschen beherrscht, sodass man die ersten Schriften an sich als «Geheimschriften» betrachten kann. Aber je mehr Menschen die Lese- und Schreibkompetenz erwarben, desto grösser wurde der Bedarf an der Geheimhaltung von schriftlich festgehaltenen Informationen, damit nur «Eingeweihte» an diese Inhalte gelangen konnten.

Die ersten Versuche, Geheimschriften zu bauen, sind ungefähr 4000 Jahre alt. Diese ältesten Geheimschriften basierten auf dem geordneten Austausch der Reihenfolge von Buchstaben in Texten.

2.1 Das Geheimnis der folgenden zwei verwendeten (unterschiedlichen) Geheimschriften liegt im Austausch der Positionen der Buchstaben. Versuchen Sie, aus den Geheimtexten die beiden ursprünglichen Texte wiederherzustellen.

a RKPYOTOLIGEEMREOLGCITHEGEHMIINSSE
b HCSFIRILTAHCZFUEBUHAWNERDNUKUZMMOINUEIZNER



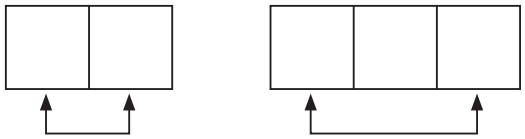
Neue Konzepte und Begriffe

Die Basis jeder Schrift ist ein **Alphabet**. Ein Alphabet ist eine endliche, nicht leere Menge von Zeichen, die wir uns zur Informationsdarstellung aussuchen. Wir sind am häufigsten mit dem lateinischen Alphabet {A,B,…,Z} der Buchstaben, dem Alphabet {0,1,2,…,8,9} der dezimalen Ziffern oder dem binären Alphabet {0,1} der Bits 0 und 1 konfrontiert. Die Folgen von Zeichen eines Alphabets bezeichnet man in der Informatik als **Wörter** oder **Texte**. Die Anzahl der Symbole eines Textes bezeichnet man als die **Länge** des Textes. Einen lesbaren Text in einer natürlichen Sprache bezeichnet man in der Kryptologie als **Klartext**. Die Umwandlung eines Klartextes in einen **Geheimtext** nennen wir **Chiffrierung**. Die Rekonstruktion des Klartextes aus einem Geheimtext nennen wir **Dechiffrierung**. Somit ist Chiffrieren ein Spezialfall von Kodieren, wenn der Zweck der Kodierung der Geheimhaltung der dargestellten Informationen dient.

Eine Geheimschrift soll so gestaltet sein, dass ihre Dechiffrierung eindeutig zum ursprünglichen Klartext führt. Wenn wir eine solche Chiffrierung und die dazugehörige Dechiffrierung haben, dann sprechen wir von einer **Geheimschrift**. Zum **Medium** wird jeder Gegenstand (Papier, Stein, Leder, Computer, …), auf dem eine Informationsdarstellung eingetragen (z. B. durch Schreiben, Malen, Ritzen, Meisseln oder mittels Elektrizität und Magnetismus) und somit abgespeichert wurde.

Bei jeder Form von Informationsaustausch (Kommunikation) sind immer ein **Sender** und ein **Empfänger** beteiligt. Wird bei dieser Kommunikation eine Geheimschrift verwendet, chiffriert der Sender seinen Klartext in einen Geheimtext und der Empfänger dechiffriert den Geheimtext, um den ursprünglichen Klartext zu erhalten.

2.2 Die beiden Vorgehensweisen bei der Chiffrierung und Dechiffrierung in Knobelaufgabe 2.1 kann man grafisch sehr einfach und anschaulich beschreiben. Man schneidet den Klartext in gleich lange Stücke – Länge 2 bei a) und Länge 3 bei b) – und beschreibt die Umwandlungsprozesse aller Stücke uniform mit folgenden Bildern:



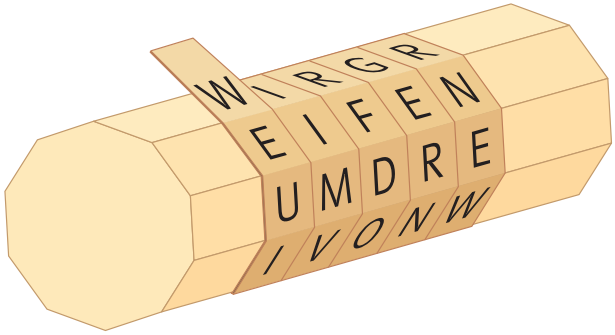
a Entwickeln Sie in Python Programme für die Chiffrierung und die Dechiffrierung der Texte nach den oben gezeichneten Mustern.

b Überlegen Sie sich ein anderes kurzes Muster zur Chiffrierung der Texte auf der Basis des Austauschens der Positionen der Buchstaben. Chiffrieren Sie damit einen Klartext und geben Sie den erzeugten Geheimtext Ihrer Nachbarin / Ihrem Nachbarn zur Dechiffrierung.

Chiffrierung mit Tabellen

Beispiel 2.1

W	I	R	G	R
E	I	F	E	N
U	M	D	R	E
I	V	O	N	W
E	S	T	E	N
A	N	S	C	H
L	E	I	C	H
E	N	Z	U	F
U	S	S	A	N
X	Y	Z	O	P



In Sparta hat man vor 2500 Jahren eine Geheimschrift, genannt SKYTALE, verwendet, die man am anschaulichsten mit einer Tabelle beschreiben kann. Man schreibt einen Klartext zeilenweise in eine Tabelle, auch **Matrix** genannt, von links nach rechts. Den Geheimtext erzeugt man, indem man die Buchstaben spaltenweise von oben nach unten liest und die Spalten von links nach rechts liest. Zum Beispiel betrachten wir die Tabelle (8 × 12) wie folgt:

D	I	E	S	C	H	R	I	F	T	E	N
E	R	M	O	E	G	L	I	C	H	T	E
N	D	A	S	E	R	S	T	E	M	A	L
I	N	F	O	R	M	A	T	I	O	N	E
N	A	U	S	S	E	R	H	A	L	B	D
E	S	M	E	N	S	C	H	L	I	C	H
E	N	G	E	H	I	R	N	S	A	B	Z
U	S	P	E	I	C	H	E	R	N	X	Y

Wenn die Tabelle nicht vollständig gefüllt ist, schreibt man beliebige (am besten zufällig gewählte) Buchstaben in die letzten freien Felder. Wenn ein Text mehr Buchstaben umfasst als die Anzahl Felder der Tabelle, schneidet man den Text in Stücke der Länge, die der Anzahl der Felder der Tabelle entspricht. Dann chiffriert man jedes Stück mit der Tabelle auf die gleiche Art und Weise. Das Geheimnis der Geheimschrift ist die Tabellengrösse (Zeilen mal Spalten) und die Reihenfolge der Positionen, in der man die Tabelle bei der Erzeugung des Geheimtextes durchläuft.
Wenn man diese Tabelle spaltenweise von links nach rechts und innerhalb der Spalten von oben nach unten liest, entsteht der folgende Geheimtext:

Bei der Dechiffrierung muss man den Geheimtext in die Tabelle spaltenweise von oben nach unten eintragen und danach wird der Klartext wie üblich zeilenweise lesbar.
Bei der Verwendung der ursprünglichen SKYTALE war es ein bisschen einfacher. Das Geheimnis war nur eine Zahl i, und zwar der Umfang eines Holzstabes (die Anzahl der Zeilen der Tabelle). Egal, wie lang der Text war, man hat den Text in i Zeilen geschrieben. Somit konnte man die Anzahl der Spalten selbst bestimmen als

$$\frac{\text{Länge des Textes}}{i}$$ nach oben aufgerundet.

D E N I N E E U I R D N A S N S E M A F U M G P S O S O S E E E
C E E R S N H I H G R M E S I C R L S A R C R H I I T T H H N E
F C E I A L S R T H M O L I A N E T A N B C B X N E L E D H Z Y

2.3 Der folgende Geheimtext wurde mit einer Tabelle der Grösse 5 × 10 erzeugt. Die gewählte Reihenfolge der Buchstaben, in der man aus der Tabelle den Geheimtext erzeugt hat, ist unbekannt. Schaffen Sie es trotzdem, den Klartext herzustellen?

A T N I T U H O N F E C I E I R I T T R A W A H H W R M C C
N U R A S E Z O M E G N F N I I E N E D

Wie viele unterschiedliche Möglichkeiten für die Chiffrierung mit der 5 × 10-Tabelle gibt es, wenn:

- **a** nur «zeilenweise» oder «spaltenweise» (ganze Zeile oder ganze Spalte in gleicher Richtung) gelesen werden kann und
- **b** beim zeilenweisen Lesen alle Zeilen einheitlich von links nach rechts oder von rechts nach links gelesen werden und
- **c** beim spaltenweisen Lesen alle Spalten einheitlich von unten nach oben oder von oben nach unten gelesen werden?

2.4 Beschreiben Sie die Dechiffrierung der Geheimschriften aus Beispiel 2.1.

- **a** Wie geht man vor und wo ist der Unterschied zur Chiffrierung?
- **b** Ist es möglich, den Geheimtext zeilenweise in eine Tabelle einzutragen und dann den Klartext spaltenweise zu lesen?
- **c** Gibt es Tabellengrössen, in denen der Algorithmus (Programm) für die Chiffrierung identisch mit dem Algorithmus für die Dechiffrierung sein kann?

2.5 Entwickeln Sie ein Programm, das für eine gegebene Tabellengrösse und einen Geheimtext alle möglichen Lesarten aus Aufgabe 2.3 ausgibt. Erzeugen Sie mit unterschiedlichen «Leseregeln» Geheimschriften und testen Sie Ihr Dechiffrierungsprogramm an diesen Geheimschriften.

Geschichtlicher und gesellschaftlicher Kontext

Zeichen (Buchstaben, Symbole) sind die atomaren Bausteine einer Schriftsprache. Bevor die Schriftsprache entstand, dienten vereinfachte Abbildungen von Tieren, Menschen oder Dingen als Zeichen. Ein solches Zeichen entsprach ursprünglich einem Konzept (einem Wort).
Die ältesten archäologischen Funde sind rund 20 000 Jahre alt (z. B. die Höhlenzeichnungen von Lascaux in Frankreich). Solche abbildenden Zeichen entwickelten sich im Verlauf der kulturellen Entwicklung von konkreten Abbildungen mit ikonischem Charakter zu abstrakten Zeichen mit symbolischem Charakter.

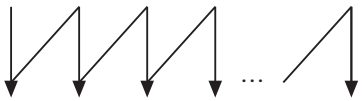
Die ältesten Schriftfunde sind ungefähr 6000 Jahre alt (Irak und Mesopotamien) und es waren **Bilderschriften**. Das bedeutet, dass einzelne Objekte zeichnerisch durch ihre Formen dargestellt wurden. Werden die Zeichen abstrakter und entsprechen inhaltlich immer noch einem Objekt beziehungsweise dem Wort für dieses Objekt, so bezeichnen wir solche Schriften als **Wortschriften**. Der Nachteil dieser Schriftart lag in der ständig wachsenden Anzahl der Zeichen, da man für jedes neue Wort ein neues Zeichen brauchte. Das führte zu der Idee, die grosse Anzahl von Wörtern als Folgen von Zeichen aus einem begrenzten Zeichenrepertoire darzustellen und damit die Anzahl der notwendigen Zeichen zu reduzieren.

Der nächste Schritt in der Entwicklung waren die **Silbenschriften**, in denen ein einzelnes Zeichen einer bestimmten Silbe entspricht (z. B. Ägypten, Mayas). Man vermutet den Ursprung in den einsilbigen Wörtern, bei denen das Zeichen nicht mehr für das Wort, sondern für die Silbe des Wortes (Wortlaut) umgenutzt wurde. Diese phonetische Art der Schriftentwicklung hat sich in vielen Kulturen durchgesetzt und führte letztendlich zur **Buchstabenschrift**, die heute meistverbreitete Schriftart. Hier entspricht jedem Zeichen ein bestimmter Laut. Zuerst entstanden unterschiedliche Mischformen. Die erste reine Buchstabenschrift wurde vor ungefähr 4000 Jahren unter dem Einfluss von ägyptischen Hieroglyphen entwickelt. Die meisten europäischen Alphabete (Zeichensysteme) haben ihren Ursprung im griechischen Alphabet, das die Etrusker nach Italien gebracht hatten und so die Grundlage für die Entwicklung des lateinischen Alphabets legten.

Im Unterschied zu natürlichen Sprachen entwickelt man in der Informatik Schriften, die nichts mit den Phonemen der gesprochenen Sprache oder der visuellen Darstellung von Objekten gemeinsam haben. Das bekannteste Beispiel ist das binäre Alphabet {0,1}, mit dem man beliebige Texte darstellen (kodieren) kann.

Es ist interessant zu beobachten, dass die ältesten Funde von Geheimsprachen ungefähr so alt sind wie die Buchstabenschriften. Diese Geheimschriften basierten auf dem Austausch der Positionen der Buchstaben in den Originaltexten oder -botschaften. Über die Richtung des Schreibens und des Lesens (die Zeichen von links nach rechts und die Zeilen von oben nach unten wie in unserer Kultur) bestand und besteht kein Konsens, weshalb unterschiedliche Kulturen auch unterschiedliche «Schreibrichtungen» entwickelten (z. B. die Zeichen von rechts nach links oder von oben nach unten). Innerhalb eines Schriftsystems ist die Schreibrichtung aber festgelegt und die natürlichste Idee für die «Verschleierung» von Texten ist die Änderung der Schreibrichtung.

2.6 Überlegen Sie sich zu Beispiel 2.1, welche andere natürliche Reihenfolge der Positionen der Buchstaben (als in Aufgabe 2.3) in der Tabelle bei der Chiffrierung verwendet werden kann. Wie viele Möglichkeiten sehen Sie? Beschreiben Sie alle. Die Chiffrierung aus Beispiel 2.1 kann man z. B. wie folgt visuell darstellen:



Kodierung von Buchstaben

Beispiel 2.2 Vor mehr als 2500 Jahren begann sich eine neue Methode für das Chiffrieren durchzusetzen. Man ersetzte unabhängig von deren Position im Klartext einzelne Buchstaben des Alphabets durch festgelegte Kodierungen. Am häufigsten «kodierte» (chiffrierte) man die Buchstaben, indem man sie durch andere Symbole ersetzte, manchmal auch durch Symbolfolgen. Die älteste bekannte Geheimschrift dieser Art stammt vom griechischen Geschichtsschreiber Polybios (ungefähr 2200 Jahre alt). Für das Chiffrieren und Dechiffrieren nutzte er die folgenden Tabellen (links angewandt auf das griechische Alphabet mit 24 Buchstaben und rechts auf das auf 24 Buchstaben reduzierte lateinische Alphabet).

	1	2	3	4	5
1	Α	Β	Γ	Δ	Ε
2	Ζ	Η	Θ	Ι	Κ
3	Λ	Μ	Ν	Ξ	Ο
4	Π	Ρ	Σ	Τ	Υ
5	Φ	Χ	Ψ	Ω	

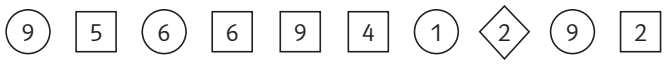
	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I / J	K
3	L	M	N	O	P
4	Q	R	S	T	U / V
5	W	X	Y	Z	

In seiner Tabelle wurde jeder Buchstabe des Alphabets durch eine Folge von zwei Ziffern kodiert. Die erste Ziffer entsprach der Nummerierung der Zeile und die zweite Ziffer der Nummerierung der Spalte. Somit wurde beim lateinischen Alphabet A durch 11 chiffriert, M durch 32, Y durch 53 usw.

Diese Idee führte im Raum des alten Palästina zur Entwicklung ganz neuer Zeichen für die Chiffrierung. Eine Verallgemeinerung dieser Idee zeigt die folgende Tabelle. Für jeden Buchstaben der Schrift wird ein neues Zeichen zusammengesetzt, das durch die Position des Buchstabens in der Tabelle bestimmt wird.

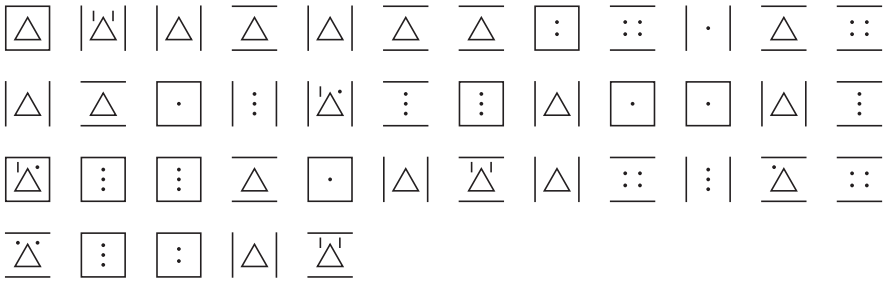
	1	2	3	4	5	6	7	8	9
○	A	B	C	D	E	F	G	H	I
□	J	K	L	M	N	O	P	Q	R
◇	S	T	U	V	W	X	Y	Z	

Das Zeichen für die entsprechende Zeile (Kreis, Quadrat, Raute) wird mit dem Zeichen der entsprechenden Spalte zu einem neuen Zeichen kombiniert. Somit wird der Klartext INFORMATIK wie folgt chiffriert.



2.7 Entwickeln Sie ein Programm, das Klartexte mit der Geheimschrift von Polybios chiffriert. Wenn die Klartexte mit dem lateinischen Alphabet geschrieben worden sind, welche Probleme entstehen bei der Dechiffrierung? Wie würden Sie damit bei der Entwicklung eines Programms zur Dechiffrierung umgehen?

2.8 Der folgende Geheimtext wurde mit einer 3 × 9-Tabelle wie in Beispiel 2.2 chiffriert. Die Bezeichnungen für Zeilen und Spalten sind aber unbekannt. Das einzig Bekannte ist, dass im ursprünglichen Klartext die am häufigsten vorkommenden Buchstaben E und N sind. Versuchen Sie, den Geheimtext zu dechiffrieren und die Chiffrierungstabelle komplett herzustellen.



Neue Konzepte und Begriffe

Die Beschreibung einer **Geheimschrift** besteht aus 4 Teilen:

- **Klartextalphabet**, in dem die Klartexte geschrieben sind.
- **Geheimtextalphabet**, in dem die Geheimtexte geschrieben sind.
- **Chiffrierungsalgorithmus** zur Umwandlung beliebiger Klartexte in Geheimtexte.
- **Dechiffrierungsalgorithmus** zur Umwandlung von Geheimtexten in die ursprünglichen Klartexte.

Das Klartextalphabet und das Geheimtextalphabet dürfen gleich sein, wie es z. B. bei der Verschiebung der Positionen der Buchstaben der Fall war. Der Chiffrierungsalgorithmus und der Dechiffrierungsalgorithmus sind manchmal auch identisch. Das Chiffrieren soll so erfolgen, dass man mittels Dechiffrieren den originalen Klartext eindeutig rekonstruieren kann.

Die Geheimschrift von Polybios kann man wie folgt beschreiben:

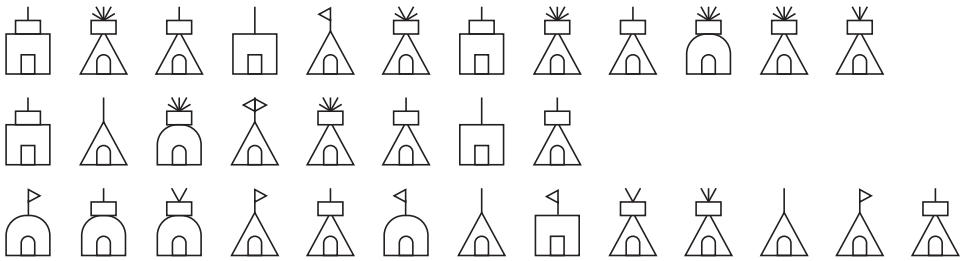
- **Klartextalphabet:** griechisches (oder lateinisches) Alphabet
- **Geheimtextalphabet:** {1,2,3,4,5}
- **Chiffrierungsalgorithmus:** Ordnen Sie jedem Buchstaben des Klartextalphabets eine Folge von zwei Ziffern aus dem Geheimtextalphabet zu. Die erste Ziffer ist die Zeilennummer der Position der Buchstaben in der Tabelle. Die zweite Ziffer ist die Spaltennummer der Position der Buchstaben in der Tabelle.
- **Dechiffrierungsalgorithmus:** Schneiden Sie den Geheimtext in Stücke der Länge 2, jedes Stück besteht also aus zwei Ziffern. Ersetzen Sie jedes Stück ab, für a, b ∈ {1,2,3,4,5}, mit den Buchstaben auf der Kreuzung der Zeile a und der Spalte b.

Wir haben hier zwei Chiffrierungsmethoden kennen gelernt. Die erste basiert auf **Transpositionen**, d. h. der Änderung der Positionen der Buchstaben im Klartext. Die zweite basiert auf **Substitutionen**, d. h. dem Ersetzen von Buchstaben durch ihre Kodierungen (einzelne Zeichen oder Symbolfolgen).

2.9 Jemand hat die Chiffrierungstabelle von Polybios geändert. Er hat das Geheimtextalphabet von {1,2,3,4,5} zu {A,B,X,Y,Z} geändert und die Buchstaben in einer anderen Reihenfolge in die Tabelle eingetragen. Dechiffrieren Sie den Geheimtext und stellen Sie die Chiffrierungstabelle zusammen. Wir wissen zusätzlich, dass E und N in dieser Reihenfolge die häufigsten Buchstaben des Klartextes sind.

A Z Y B Z A A Z A A Z A X B A X Z A X X X Y Y B Z A
Z A Y B X X Z A X X B B Z Y Y Y Z A X X B Y Z A X A X B X X Z A B Y

2.10 Der folgende Geheimtext basiert auf der Erzeugung von neuen Zeichen mit einer 3 × 9-Tabelle. Dechiffrieren Sie den folgenden Geheimtext, wenn bekannt ist, dass E am häufigsten, I am zweithäufigsten und die Buchstaben W und A genau dreimal vorkommen.



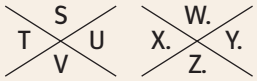
Geschichtlicher und gesellschaftlicher Kontext

Die Entwicklung der Kryptologie als Lehre der Geheimschriften ist stark verbunden mit dem Begriff der Sicherheit, d. h. in der Antike mit der Frage: «Wann ist eine Geheimschrift sicher?» Das Wort «sicher» soll bedeuten, dass kein Uneingeweihter es schafft, die Geheimtexte zu dechiffrieren.

Interessanterweise beruhte im Altertum die Sicherheit fast ausschliesslich auf der Geheimhaltung der Geheimschrift. Es wurden noch keine Methoden entwickelt, mit denen man aus den Geheimtexten selbst herausfinden könnte, wie die Geheimschrift funktioniert. Deswegen war das oberste Prinzip der Sicherheit, die verwendete Geheimschrift geheim zu halten. In der Praxis bedeutete das meistens, dass man die Beschreibung der Geheimschrift nicht schriftlich aufbewahren durfte. Deswegen waren alle Geheimschriften so gebaut, dass man sie leicht auswendig lernen konnte.

A	B	C
D	E	F
G	H	I

J.	K.	L.
M.	N.	O.
P.	Q.	R.



Die Geheimschrift der Freimaurer im Mittelalter ist das wohl bekannteste Beispiel dieser Art. Sie wurde von antiken Geheimschriften aus dem Gebiet des alten Palästina abgeleitet. Die Freimaurer zeichneten ihre Chiffrierungstabellen in Sand oder in Mehl während des Lernprozesses, um das aufgezeichnete Geheimnis nach dem Erlernen leicht verwischen, also unkenntlich machen zu können.

2.11 Arbeiten Sie in Gruppen. Jede Gruppe entwickelt eine neue Geheimschrift, die man leicht auswendig lernen kann, und chiffriert mit dieser Geheimschrift einen Klartext. Die Beschreibung (Tabelle, Zeichnungen, …) der Geheimschrift überlässt man für die Dauer von 1–2 Minuten einer anderen Gruppe zum Auswendiglernen. Nach der Rückgabe der Beschreibung offeriert man den Geheimtext zum Dechiffrieren.

Wenn das Geheimtextalphabet kleiner ist als das Klartextalphabet, bleibt uns nichts anderes übrig, als die Buchstaben des Klartextalphabets durch Folgen von Symbolen des Geheimtextalphabets zu kodieren. Wenn wir als Geheimtextalphabet das heute häufig verwendete binäre Alphabet {0,1} auswählen, müssen wir zur Buchstabenkodierung Folgen von Nullen und Einsen nehmen. Das geht aber nicht beliebig wegen der Eindeutigkeit der Dechiffrierung. Solange die Kodierungen aller Buchstaben (wie bei der standardisierten ASCII-Kodierung) die gleiche Länge haben, ist die Dechiffrierung einfach und eindeutig. Man schneidet den Geheimtext in Stücke genau dieser Länge und interpretiert jedes Stück als einen Buchstaben. Wie schwierig es aber sein kann, wenn die Kodierungen unterschiedliche Längen haben, zeigt die folgende Knobelaufgabe.

2.12 Für die binäre Kodierung von lateinischen Buchstaben hat man folgende binäre Folgen verwendet. Die Buchstaben, die im Klartext nicht vorkommen, sind nicht aufgelistet.

A → 1010	B → 11010	D → 01111	E → 0
G → 1101	H → 101	I → 00	K → 1
L → 01	M → 1001	N → 01	O → 0010
R → 001	S → 0010	T → 1111	U → 110011
W → 1101			

a Finden Sie zwei unterschiedliche, sinnvolle Wörter, die durch den gleichen Geheimtext chiffriert werden.

```
1 1 0 1 0 0 1 1 0 1 0 1 1 0 0 1 1
```

b Finden Sie zwei unterschiedliche, sinnvolle Sätze, deren Chiffrierung zu dem gleichen folgenden Geheimtext führt.

```
0 1 1 1 1   0 0 0 1 1   0 1 0 0 1   0 0 1 1 0   0 1 0 0 1   0 0 0 0 1
0 1 1 1 1   1 1 0 1 0   0 1 1 0 1   0 1 1 0 0   1 1
```

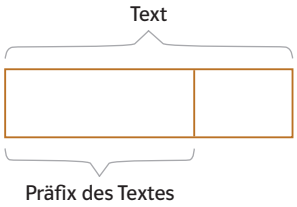
Die Verteilung in Stücke der Länge 5 hat nichts mit der Buchstabenkodierung oder den Leerzeichen zwischen Wörtern im Klartext zu tun. Sie soll nur die Übersichtlichkeit erhöhen, insbesondere, wenn man den Geheimtext übertragen muss.

Neue Konzepte und Begriffe

Eine Chiffrierung kann man als einen Algorithmus oder eine Funktion beschreiben, der oder die jedem Text über ein Klartextalphabet (als Eingabe oder Argument) einen Text über ein Geheimtextalphabet zuordnet. Weil ein Algorithmus detailliert beschreibt, wie man den Klartext in einen Geheimtext umwandelt, garantiert er eine ausführliche Beschreibung der Chiffrierung. Jeden Chiffrierungsalgorithmus kann man als die Beschreibung einer Funktion ansehen, die jedem Argument (Text im Klartextalphabet) einen Funktionswert (Text im Geheimtextalphabet) zuordnet. Die Beschreibung einer Funktion kann, muss aber nicht den Umwandlungsweg darstellen. Sie bestimmt nur das Resultat für jedes mögliche Argument.

Eine Funktion f von A nach B (z. B. A beinhaltet alle Texte über ein Klartextalphabet und B alle Texte über das Geheimtextalphabet) heisst **injektiv**, wenn für zwei beliebige unterschiedliche Argumente x und y aus A die Funktionswerte f(x) und f(y) auch unterschiedlich sind. Die gute Eigenschaft injektiver Funktionen ist, dass man aus den Funktionswerten das Argument eindeutig bestimmen kann. Deswegen streben wir an, dass entworfene Chiffrierungen injektiven Funktionen entsprechen. Die Injektivität der Chiffrierungen garantiert, dass man jedem Geheimtext eindeutig den ursprünglichen Klartext zuordnen kann.

2.13 Ein Text v (Folge von Symbolen) ist ein **Präfix** eines Textes u, wenn für einen Text x gilt: u = vx, d. h., wenn der Text u mit dem Text v anfängt. Somit sind z. B. alle Buchstabenfolgen (Texte) L, LE, LER, LERN, LERNE und LERNEN Präfixe des Textes LERNEN. Den ganzen Text betrachten wir auch als ein Präfix von sich selbst.



Betrachten wir eine Kodierung von Buchstaben durch Folgen von Symbolen (wie in Knobelaufgabe 2.12). Wir sagen, dass diese Kodierung **präfixfrei** ist, wenn keine Symbolfolge ein Präfix einer anderen Folge ist. Die Kodierung in Aufgabe 2.12 ist nicht präfixfrei, weil z. B.
- 01 (Kodierung von L) Präfix von 01111 (Kodierung von D) ist oder
- 0 (Kodierung von E) Präfix von 00 (Kodierung von I) und von 001 (Kodierung von R) ist.

Bestimmen Sie alle Folgen aus Knobelaufgabe 2.12, die Präfixe von anderen Folgen sind.
Wenn man bei der Chiffrierung eine präfixfreie Kodierung von Buchstaben verwendet, entspricht die Chiffrierung garantiert einer injektiven Funktion. Erklären Sie, wie man einen Algorithmus für die Dechiffrierung so gestalten kann, dass er eindeutig aus jedem Geheimtext den ursprünglichen Klartext «ausrechnen» kann.
Wählen Sie eine präfixfreie Kodierung von Buchstaben des lateinischen Alphabets, indem Sie mindestens drei unterschiedliche Kodierungslängen verwenden. Entwickeln Sie Programme zur Chiffrierung und zur Dechiffrierung für Ihre Kodierung und testen Sie beide an unterschiedlichen Texten.

2.14 Schaffen Sie es, eine Kodierung zu entwickeln, die nicht präfixfrei ist und trotzdem einer injektiven Funktion von Texten über das Klartextalphabet auf Texte über das Geheimtextalphabet entspricht? Entwickeln Sie basierend auf dieser Kodierung Programme zur Chiffrierung und Dechiffrierung mit dieser Kodierung.
Versuchen Sie eine ganze Klasse (wie präfixfreie Kodierungen) von Kodierungen zu entwickeln, in der die Kodierungen nicht notwendigerweise präfixfrei sind und trotzdem eine eindeutige Dechiffrierung ermöglichen.

Was Sie gelernt haben

Die Geheimschriften dienen zur Geheimhaltung von Daten (verschriftlichten Informationen). Zwei grundlegende Methoden zur Entwicklung von Geheimschriften basieren auf Transpositionen (dem Austausch der Positionen von Buchstaben) und Substitutionen (der Kodierung von Buchstaben durch andere Symbole oder Symbolfolgen).

Eine Geheimschrift besteht aus den vier Elementen
- Klartextalphabet,
- Geheimtextalphabet,
- Algorithmus zur Chiffrierung und
- Algorithmus zur Dechiffrierung.

Die Chiffrierung sollte einer injektiven Funktion entsprechen, um aus jedem Geheimtext eindeutig den ursprünglichen Klartext bestimmen zu können.

Kryptosysteme

Wenn man zur Geheimhaltung von Daten immer die gleiche Geheimschrift verwendet, wächst das Risiko, dass das Geheimnis irgendwann gelüftet wird. Deswegen hat man sich schon in der Antike überlegt, dass man abwechselnd unterschiedliche Geheimschriften verwendet, z. B. bei einem Briefwechsel.

2.15 Versuchen Sie, diesen Geheimtext zu dechiffrieren. Als Hilfe geben wir preis, dass die Chiffrierung auf dem Ersatz von Buchstaben durch andere Buchstaben des lateinischen Alphabets basiert und dass der Klartext den Namen Polybios beinhaltet. Als Erleichterung gibt es im Geheimtext Leerzeichen zwischen den Chiffrierungen der einzelnen Wörter.

S R O B E L R V Z D U H L Q J H V F K L F K W V V F K U H L E H U

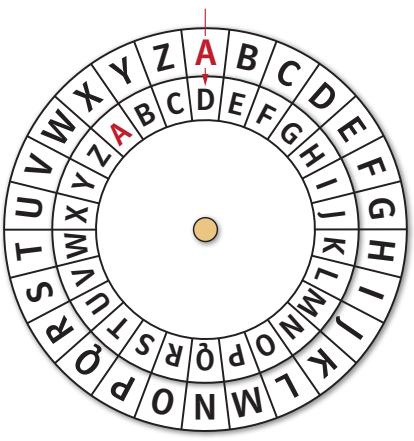
Hinweis: Falls man weitere Hilfe braucht, geben wir preis, dass das Trigramm SCH im Klartext zweimal vorkommt.

Beispiel 2.3

Kryptosystem CAESAR

Das bekannteste System von Geheimschriften aus der Antike ist CAESAR, das tatsächlich der Feldherr Gaius Julius Caesar verwendet hat. Das System besteht aus 25 Geheimschriften, die man mit Zahlen 1, 2, …, 25 bezeichnete. Alle benutzten das lateinische Alphabet für Klartexte sowie für Geheimtexte. Wenn man die Geheimschrift i verwendet hat, hat man jeden Buchstaben des Klartextes bei der Chiffrierung durch den Buchstaben ersetzt, der i Positionen rechts davon im Alphabet steht. Wenn man über Z hinausgegangen ist, setzte man am Anfang des Alphabets fort. Am einfachsten kann man die Chiffrierung und Dechiffrierung mit der CAESAR-Scheibe veranschaulichen. Die Scheibe besteht aus zwei Rädern 26 lateinischen Buchstaben. Weil sie Chiffrierung und Dechiffrierung ziemlich erleichtert, empfehlen wir, sich so eine Scheibe zu basteln. Die Räder kann man gegeneinander drehen. Wenn man, wie im Bild, das innere Rad um drei Positionen gegen den Uhrzeigersinn dreht, zeigt jeweils der Buchstabe des inneren Rades die Chiffrierung des Buchstabens des äusseren Rades für die Geheimschrift 3. In Knobelaufgabe 2.15 war der Klartext mit der Verschiebung 3 von CAESAR chiffriert.

Mathematisch beschreibt man die Funktionsweise von CAESAR mit der modularen Addition. Zuerst ordnet man jedem Buchstaben des

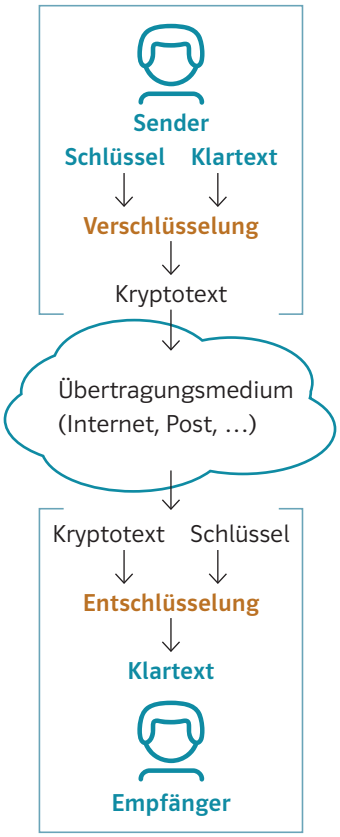


Eine Kopiervorlage für die Scheibe finden Sie bei den Online-Materialien auf http://www.meinklett.ch.

Alphabets seine Ordnung zu, wobei man bei 0 zu zählen beginnt:
Ordnung (A) = 0, Ordnung (B) = 1, Ordnung (C) = 2, …, Ordnung (Z) = 25.
Die Verschiebung des Buchstabens ☐ um i Positionen nach rechts in der Chiffrierung geschieht wie folgt:

- Wenn Ordnung (☐) + i ≤ 25, wird ☐ durch den Buchstaben mit der Ordnung Ordnung (☐) + i ersetzt,
- wenn Ordnung (☐) + i > 25, wird ☐ durch den Buchstaben mit der Ordnung Ordnung (☐) + i − 26 ersetzt.

Die folgende kurze Beschreibung besagt, dass ☐ durch den Buchstaben mit der Ordnung (Ordnung (☐) + i) mod 26 ersetzt wird. Dabei bedeutet allgemein der Ausdruck x mod y den Rest vom Teilen der Zahl x durch y (für zwei positive ganze Zahlen x und y).



Neue Konzepte und Begriffe

Ein **Kryptosystem** ist eine Sammlung von Geheimschriften, wobei jede Geheimschrift einen eindeutigen Namen (Bezeichnung) hat. Die einzelnen Namen der Geheimschriften nennen wir **Schlüssel**. Dass die Geheimschriften eines Kryptosystems etwas Gemeinsames haben, drückt man dadurch aus, dass man Chiffrierung und Dechiffrierung als injektive Funktionen Ver und Ent von zwei Argumenten bezeichnet:

Chiffrierung: Ver (Klartext, Schlüssel) = Geheimtext
Dechiffrierung: Ent (Geheimtext, Schlüssel) = Klartext

Wenn wir Kryptosysteme verwenden, sprechen wir auch über **Verschlüsselung** statt Chiffrierung und **Entschlüsselung** statt Dechiffrierung. Die Darstellung des geheimen Informationsaustausches sieht dann so aus wie in der Abbildung links. Das Wort **Kryptotext** verwendet man auch als Synonym für «Geheimtext».

Es ist zu beobachten, dass Sender und Empfänger den gleichen Schlüssel nutzen und dieser Schlüssel bei der Kommunikation nicht übertragen wird. Dieser geheime Kommunikationsprozess kann also nur dann stattfinden, wenn sich Sender und Empfänger im Voraus auf einen gemeinsamen Schlüssel geeinigt haben.

2.16 Der Geheimtext

X G T Y G P F G U E J N W G U U G N B Y G K

wurde mit CAESAR verschlüsselt, der Schlüssel ist aber unbekannt. Entschlüsseln Sie den Geheimtext, ohne alle Schlüssel auszuprobieren, wenn Sie wissen, dass der häufigste Buchstabe im Klartext E ist.

2.17 Der folgende Geheimtext wurde mit CAESAR verschlüsselt. Bestimmen Sie den Schlüssel, wenn Sie wissen, dass das Trigramm EIT am häufigsten von allen Trigrammen im Klartext vorkommt.

C H D U N K K J N L L D M G D H S A D R S D G S Z T R J K D
H M H F J D H S D M C N B G C H D U N K K J N L L D M G
D H S R D K A R S H R S J D H M D J K D H M H F J D H S

2.18 CAESAR kann man als Kryptosystem von 25 Geheimschriften mit Verschiebungen der Buchstaben um 1, 2, 3, …, 25 Positionen im lateinischen Alphabet betrachten. Kann man die SKYTALE von Sparta auch als ein Kryptosystem betrachten? Was wären dann die Schlüssel?

2.19 Verschlüsselung und Entschlüsselung sind bei CAESAR zwei unterschiedliche Algorithmen. Bei der Verschlüsselung verschiebt man die CAESAR-Scheibe im Uhrzeigersinn, bei der Entschlüsselung verschiebt man die Scheibe gegen den Uhrzeigersinn.
Gibt es Schlüssel bei CAESAR, für die Verschlüsselung und Entschlüsselung identisch sind?

2.20 Wenn man weiss, dass ein Geheimtext durch CAESAR verschlüsselt wurde, kann man es immer schaffen, den Klartext herzustellen. Im schlimmsten Fall betreibt man den Aufwand und probiert alle 25 Schlüssel aus, bis man einen bedeutungsvollen Text erhält.
Cleverius überlegt sich ein neues Kryptosystem, um einem potenziellen Gegner den Aufwand für eine Dechiffrierung dank einer grösseren Anzahl von Schlüsseln zu erhöhen.
Die Schlüssel sind:

(i, k, j) i, j ∈ {1,2,…,25}, k ∈ {1,…,10}

Den Schlüssel wendet man auf einen Klartext wie folgt an:
1. Wenden Sie CAESAR mit Verschiebung i auf den Klartext an.
2. Anschliessend wenden Sie die ursprüngliche SKYTALE auf den in (1) erhaltenen Text mit dem Schlüssel k (Anzahl der Zeilen) an.
3. Zum Schluss wenden Sie CAESAR mit Verschiebung j auf den in (2) erhaltenen Text an.

a Der folgende Text wurde mit dem Schlüssel (11, 3, 18) verschlüsselt. Dechiffrieren Sie den Text:

JHH HJO KXH HQX GGW HOH LDU QVH HVG QGH ZLQ

b Wie viele unterschiedliche Geheimschriften besitzt das Kryptosystem von Cleverius?

2.21 Verallgemeinern Sie das Kryptosystem CAESAR wie folgt:
1. CAESAR ordnet mit Schlüssel i jedem Buchstaben □ den Buchstaben △ zu, wenn Ordnung (△) = (i + Ordnung (□)) mod 26. Was würde passieren, wenn wir die Addition durch die Multiplikation ersetzten?
   Ordnung (△) = (i · Ordnung (□)) mod 26. Wir bezeichnen dieses Kryptosystem als MULTCAESAR. Für welche i ∈ {1,2,…,25} erhalten wir eine Geheimschrift im Sinne einer injektiven Funktion? Wie sieht der Dechiffrierungsalgorithmus aus?
2. Eine Aussage von Louis Pasteur wurde mit MULTCAESAR verschlüsselt. Man weiss, dass E der häufigste Buchstabe im Klartext ist. Dechiffrieren Sie den Geheimtext und bestimmen Sie den Schlüssel.
   WNEDWUKYDDKJMONPPOUTUNOEUNPUNUNADDUMDAHOMR
3. Wir schlagen ein neues Kryptosystem LINCAESAR mit den Schlüsseln (a,b) ∈ {1,2,…,25} vor. Die Zahlen a und b verwendet man zur Verschlüsselung wie folgt:
   Ordnung (△) = (a · Ordnung (□) + b) mod 26

Wie viele unterschiedliche Geheimschriften im Sinne von injektiven Abbildungen beinhaltet LINCAESAR? Für welche Zahlenpaare (a, b) funktionieren die Verschlüsselung und die Entschlüsselung als injektive Funktionen? Wie sieht die Formel aus, mit der man für jeden Buchstaben □ des Geheimtextes den entsprechenden Buchstaben △ des Klartextes bestimmen kann?

Kryptosystem mit vielen Schlüsseln

Beispiel 2.4 Alle bisher vorgestellten Kryptosysteme hatten die in der Antike gewünschte Eigenschaft, dass man sie leicht auswendig lernen konnte. Ihre Schwäche lag in der geringen Anzahl der Schlüssel, sodass man, sofern das Kryptosystem bekannt war, durch Ausprobieren alle leicht rekonstruieren konnte.
Auch wenn man in Klartexten sowie in Geheimtexten nur mit dem lateinischen Alphabet gearbeitet hat, bestand immer die Möglichkeit der Erstellung sehr vieler Geheimschriften. Beispielsweise könnte ein Schlüssel wie folgt aussehen:

Man kann jedem Buchstaben einen beliebigen Buchstaben zuordnen. Um die eindeutige Dechiffrierung zu garantieren, darf man aber in keinem Fall zwei Buchstaben des Klartextes den gleichen Buchstaben im Geheimtext zuordnen. Die Anzahl der möglichen Schlüssel ist so hoch, dass niemand versuchen würde, alle auszuprobieren.

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

2.22 Dechiffrieren Sie den folgenden Geheimtext, der mit dem Schlüssel aus Beispiel 2.4 chiffriert wurde:

Y C X C B C W T C T S C Y L T B V Z R R B W C V C L Q C K C B C T

2.23 Wie gross ist die Anzahl aller möglichen Schlüssel (aller Geheimschriften) des Kryptosystems in Beispiel 2.4?

Neue Konzepte und Begriffe

Betrachten wir eine auf Substitution basierende Geheimschrift mit Klartextalphabet Σ und Geheimtextalphabet Γ. Wir nennen die Geheimschrift **monoalphabetisch**, wenn bei der Chiffrierung gilt: «Jeder Buchstabe aus Σ wird durch ein und denselben Buchstaben aus Γ ersetzt, unabhängig von seiner Position in dem Klartext.»

Ein Kryptosystem ist monoalphabetisch, wenn alle seine Geheimschriften monoalphabetisch sind. Das Kryptosystem CAESAR und seine Verallgemeinerungen in Beispiel 2.4 sind monoalphabetische Kryptosysteme.

Man kann den Begriff «monoalphabetisch» für Geheimschriften verallgemeinern, die Buchstaben durch Buchstabenfolgen kodieren. Man fordert analog, dass die Kodierungen der einzelnen Buchstaben immer gleich sind, egal wo sich der Buchstabe im Klartext befindet.

Eine Geheimschrift heisst **polyalphabetisch**, wenn sie nicht monoalphabetisch ist, d. h., wenn mindestens ein Buchstabe des Klartextalphabets unterschiedlich kodiert wird, abhängig von seiner Position im Klartext.

2.24 Welche der folgenden Geheimschriften (Kryptosysteme) sind monoalphabetisch und welche sind polyalphabetisch? Begründen Sie Ihre Antworten.

a Die Geheimschrift von Polybios aus Beispiel 2.2

b Das Kryptosystem mit den Schlüsseln (i, j), i, j ∈ {1,2,…,25}, wobei bei der Verschlüsselung alle Buchstaben des Klartextes auf geraden Positionen mit Schlüssel i mit CAESAR chiffriert sind und alle an ungeraden Positionen mit Schlüssel j

c Die Geheimschrift aus Knobelaufgabe 2.10

d Die Geheimschrift mit einer Zahl j, in der jeder Buchstabe □ des lateinischen Alphabets durch den Buchstaben △ mit der Ordnung
Ordnung (△) = (Ordnung (□) + Position von □ im Klartext + j) mod 26
ersetzt wird

2.25 Chiffrieren Sie den folgenden Klartext mit zwei Geheimschriften aus **b** und **d** von Aufgabe 2.24 für i = 1 und j = 2.

DAS NEUE WIRD ZUERST VERSPOTTET,
DANN WIDERLEGT UND LETZTENDLICH ALS ETWAS
SELBSTVERSTAENDLICHES AKZEPTIERT

Was Sie gelernt haben

Kryptosysteme sind Sammlungen von Geheimschriften, die mit gleichem Klartextalphabet und gleichem Geheimtextalphabet arbeiten und deren Chiffrierung «ähnlich» ist. Ähnlich bedeutet, dass es einen Chiffrierungsalgorithmus gibt, der für den gegebenen Namen einer Geheimschrift des Kryptosystems und gegebenen Klartext den Klartext mit dieser Geheimschrift chiffriert. Somit kann man sich für den geheimen Nachrichtenaustausch eine immer wechselnde Geheimschrift des Kryptosystems aussuchen und so das Risiko mindern, dass das Kryptosystem durch Verwenden der immer gleichen Geheimschrift gebrochen wird. Die Namen der einzelnen Geheimschriften eines Kryptosystems bezeichnen wir als Schlüssel und sprechen deswegen von Verschlüsselung und Entschlüsselung.
Eine Geheimschrift, die auf dem Prinzip der Substitution (der Ersetzung von Buchstaben durch ihre Kodierungen) basiert, ist monoalphabetisch, wenn jeder Buchstabe durch die gleiche Kodierung (Symbolfolge), unabhängig von seiner Position im Klartext, ersetzt wird. Wenn man für unterschiedliche Vorkommen des gleichen Buchstabens unterschiedliche Kodierungen bei der Chiffrierung anwendet, sprechen wir von polyalphabetischen Geheimschriften.

Stochastische Kryptoanalyse

Die Vorstellung aus der Antike, dass die Geheimhaltung des Kryptosystems die wichtigste Massnahme für die Geheimhaltung von Daten und Nachrichten ist, erwies sich im 7. und 8. Jahrhundert als naiv. Man entwickelte Methoden, mit denen man allgemein beliebige monoalphabetische Kryptosysteme brechen kann, vorausgesetzt, dass die Klartexte sinnvollen Texten in natürlichen Sprachen entsprechen. Und das gilt auch für Kryptosysteme mit einer so hohen Anzahl an Schlüsseln, dass es wegen des entsprechend hohen Aufwandes nicht möglich wäre, den verwendeten Schlüssel durch Ausprobieren zu identifizieren.

Die Kunst, Kryptosysteme zu brechen, bezeichnet man als **Kryptoanalyse**. Wir sprechen dann vom Brechen, wenn man nur aus dem Geheimtext (ohne den Schlüssel) das Geheimnis der verwendeten Geheimschrift bestimmen und somit den Geheimtext dechiffrieren kann. Entsprechend besteht die Kryptologie aus zwei Teilgebieten, nämlich der Kryptoanalyse und der **Kryptografie**, die sich dem Entwurf von sicheren Kryptosystemen widmet.
Die erste Methode der Kryptoanalyse aus dem 7. Jahrhundert beruhte auf der unterschiedlichen Häufigkeit der einzelnen Buchstaben in natürlichen Sprachen. Die folgende Tabelle gibt den durchschnittlichen prozentualen Anteil der Buchstaben in einem deutschsprachigen Text an.

Wir sehen, dass die prozentuale Häufigkeit einzelner Buchstaben in einem vom Umfang und vom Inhalt her repräsentativen deutschsprachigen Text sehr unterschiedlich ist. Im Einzelfall können natürlich die Texte von dieser Statistik auch stark abweichen. Aber in ausreichend langen Texten ist man sehr nah an der typischen Häufigkeitsverteilung der einzelnen Buchstaben.
Der Anhaltspunkt für die Kryptoanalyse ist die Tatsache, dass sich in monoalphabetischen Geheimschriften die Häufigkeit der Buchstaben im Klartext auf die Häufigkeit der Buchstaben im Geheimtext überträgt. Wenn z. B. E 144-mal im Klartext vorkommt, dann kommt die Kodierung von E (das Symbol, das E kodiert) genau 144-mal im Geheimtext vor. Somit kann man schätzen, dass das häufigste Symbol in einem (deutschen) Geheimtext die Chiffrierung von E ist und das zweithäufigste Symbol im gleichen Geheimtext N kodiert.

E	17,74 %
N	10,01 %
I	7,60 %
R	6,98 %
S	6,88 %
A	6,43 %
T	5,94 %
H	5,22 %
D	5,12 %

U	4,27 %
L	3,49 %
C	3,26 %
M	2,75 %
G	2,69 %
O	2,39 %
B	1,85 %
W	1,73 %
F	1,56 %

K	1,40 %
Z	1,10 %
P	0,64 %
V	0,64 %
J	0,23 %
Y	0,04 %
X	0,02 %
Q	0,01 %

2.26 Der folgende deutschsprachige Text wurde mit dem Kryptosystem CAESAR verschlüsselt. Finden Sie den Schlüssel und dechiffrieren Sie den Geheimtext.

ZDIHZ INXCF VIIZD BZIZA ZCGZM GDZWZ I

2.27 Entwickeln Sie ein Programm, das als Eingabe einen Text erhält und als Ausgabe für jeden Buchstaben auflistet, wie viele Male der Buchstabe im Text vorkommt und wie viel Prozent des Textes (die Anzahl des Vorkommens des Buchstabens im Text geteilt durch die Textlänge) der Buchstabe ausmacht.
Fügen Sie als Eingabe für Ihr Programm einen Text mit mehr als 1000 Wörtern ein und vergleichen Sie die Ausgabe Ihres Programms mit der oben in der Tabelle präsentierten Statistik.

Neue Konzepte und Begriffe

Sei T ein Text und □ ein Buchstabe aus dem Alphabet, über das T geschrieben worden ist. Wir bezeichnen die Anzahl des Vorkommens des Buchstabens □ in T als

$$H_{\square}(T)$$

und nennen es die **absolute Häufigkeit** von □ in T. Die **Länge** von T ist die Anzahl aller Buchstaben in T und wird als |T| bezeichnet. Die **relative Häufigkeit** $h_{\square}(T)$ von □ in T ist definiert als $h_{\square}(T) = \frac{H_{\square}(T)}{|T|}$.
Somit bezeichnet $H_{\square}(T)$ den proportionalen Anteil der Buchstaben □ im Text T und $h_{\square}(T) \cdot 100\,\%$ ist der proportionale Anteil von □ in T in Prozenten ausgedrückt.

Beispiel 2.5 **Häufigkeitsanalyse**

Wir betrachten den Geheimtext:

M U M M U X J U Q Y M H Q O U M S U T U Q N G W T
J Q V H U M X Q U Q T M U M S U V Y P P U M

Wir vermuten, dass der Klartext in deutscher Sprache geschrieben worden ist und das verwendete Kryptosystem monoalphabetisch ist. Wir zählen das Vorkommen von Buchstaben im Geheimtext und stellen fest:
- U kommt 11-mal vor,
- M kommt 9-mal vor,
- Q kommt 6-mal vor.

Deswegen schätzen wir, dass U den Buchstaben E, M den Buchstaben N und Q den Buchstaben I kodiert. Wir probieren es aus und erhalten den folgenden Lückentext:

```
NENNE--EI-N-I-EN-E-EI-----I--EN-IEI-NEN-E----EN
```

Die Annahme von E und N sieht gut aus und für I könnte sie auch zutreffen. Jetzt hat man drei Möglichkeiten fortzufahren. Man rät weitere fehlende Buchstaben im Lückentext, um sinnvolle Wörter zu erhalten, oder sucht nach Doppelbuchstaben (die SS, NN oder LL entsprechen könnten) oder nach den in relativ grosser Häufigkeit vorkommenden Bigrammen und Trigrammen (Tabelle):

ER	3,89 %	EIN	1,14 %
EN	3,74 %	ICH	1,12 %
CH	2,97 %	DER	0,92 %
TE	2,21 %	SCH	0,84 %
ND	2,11 %	UND	0,81 %
DE	2,06 %	DIE	0,74 %

Je länger der Geheimtext ist, desto mehr Anhaltspunkte bekommen wir und desto einfacher ist es für uns, den Geheimtext zu dechiffrieren. Unser Text ist kurz und deswegen stehen wir vor einer grösseren Herausforderung.
Aus dem Lückentext (Vorkommen von - I E) sowie aus der relativen Häufigkeit des Trigramms DIE vermuten wir, dass X den Buchstaben D kodieren könnte und somit das zweite Wort des Klartextes möglicherweise DREI ist (was bedeuten würde: J kodiert R). Wir vermuten ebenfalls, dass I - NEN das Wort IHNEN sein könnte, was bedeuten würde, dass T den Buchstaben H kodiert. Der Doppelbuchstabe PP am Ende des Geheimtextes kann nicht NN sein, also schätzen wir, dass es S oder L ist. Diesen Annahmen zufolge erhalten wir inzwischen den folgenden Lückentext. Bringen Sie die Kryptoanalyse mit weiteren Nachforschungen zu Ende.

```
NENNEDREI-N-I-ENGEHEI-----I--ENDIEIHNEN-E--
S/LS/LEN
```

2.28 Ein deutschsprachiger Klartext wurde monoalphabetisch chiffriert. Entschlüsseln Sie in Teamarbeit den Geheimtext und ermitteln Sie den verwendeten Schlüssel. Um Ihnen die Lösung der Aufgabe zu erleichtern, haben wir die Interpunktions- und Leerzeichen im Geheimtext stehen lassen. In der Realität macht man das bei der Verwendung von Kryptosystemen aber nie.

```
RLJ AJHFJ RJF OJHJLETGHFLKNJM MJMMN EVM
XFPUNDADOLJ (SDE OFLJGHLTGHJM XFPUNDT, SJFQDFOJM,
YMR ADODT, WDFN, AJHFJ, XYMRJ). RLJ VXNLSLNVJNJM
RJF XDEEYMLBLJFJMRJM YMR XFPUNDVMVAPNLXJF NJLAJM
RLJ XFPUNDADOLJ LM BWJL OJQLJNJ JLM. RLJ
XFPUNDOFVUHLJ (VYGH: XFPUNDOFVKLJ, SDE OFLJGHLTGHJM
XFPUNDT, SJFQDFOJM, YMR OFVUHJLM, TGHFJLQJM)
LTN RLJ WLTTJMTGHVKN RJF JMNWLGXAYMO SDM
XFPUNDTPTNJEJM, YMR RLJ XFPUNDVMVAPTJ LTN RLJ
AJHFJ RJF VMVAPTJ SDM OJHJLENJCNJM YMR
XFPUNDTPTNJEJM, RLJ BYE XMVGXJM RJF VMVAPTLJFNJM
XFPUNDPTNJEJ KYJHFJM TDAA.
```

2.29 Gehen Sie auf die Seite https://einfachinformatik.inf.ethz.ch/application/topics und klicken Sie auf «Datenschutz und Geheimschriften». Dort wählen Sie «Entschlüsseln ohne Schlüssel». Hier steht eine Aufgabe, in der die Häufigkeit der Buchstaben im Geheimtext schon gezählt wurde und im Säulendiagramm mit der stochastischen Verteilung der Buchstaben in deutschsprachigen Texten visuell verglichen wird. Hier kann man schätzen, welcher Buchstabe des Geheimtextes welchen Buchstaben des Klartextes kodiert.
Die Schätzungen werden automatisch in den Lückentext eingesetzt und können auch rückgängig gemacht werden. Nutzen Sie das Tool, um den Geheimtext zu dechiffrieren.

2.30 Der folgende Geheimtext ist durch eine monoalphabetische Chiffrierung aus einem englischen Klartext entstanden. Wir verraten noch, dass R durch H chiffriert ist. Versuchen Sie den Geheimtext zu dechiffrieren.

```
EBNCP TAICH JAHBD JAMES IAHCM HPSBD RAWBD
DCFAF RKEGI SKCEF BAWBT AMCLG ABIAH AKCBE
```

Der Satz des Klartextes ist noch nicht beendet – es fehlt eine Erklärung. Formulieren Sie die fehlende Erklärung selbst.

2.31 Arbeiten Sie in Gruppen.

a Alle Gruppen nehmen einen deutschsprachigen Text, in dem die Buchstaben E und N am häufigsten vorkommen. Jeden Buchstaben chiffrieren Sie monoalphabetisch durch eine Folge von fünf Bits und halten die Chiffrierung geheim. Für zwei unterschiedliche Buchstaben verwenden Sie immer zwei unterschiedliche Bitfolgen. Dann tauschen die Gruppen die Geheimtexte untereinander und versuchen den fremden Text zu dechiffrieren.

b Dies ist die gleiche Aufgabe wie in **a**, jedoch wird jeder Buchstabe des Klartextes durch eine Folge von zwei Buchstaben statt fünf Bits kodiert.

Geschichtlicher und gesellschaftlicher Kontext

In der Epoche zwischen 750 und dem 11. Jahrhundert war das islamische Reich im Nahen Osten und Mittelasien ein blühendes Zentrum der Kultur und der Wissenschaft. Das galt auch für die Kryptografie – und nicht nur für die Chiffrierung von Botschaften, sondern für den Datenschutz im Allgemeinen (beispielsweise wurden schon damals Steuerunterlagen verschlüsselt). Die in diesem Lehrmittel vorgestellten Chiffrierungsmethoden der **Transposition** (Änderung der Positionen von Buchstaben) und der **Substitution** (Ersetzen von Buchstaben durch Symbole oder Symbolfolgen) wurden weiterentwickelt.
Über das Studium der Häufigkeiten der Wörter insbesondere in den religiösen Texten kamen die Menschen auch zum Studium der Häufigkeiten der einzelnen Buchstaben. Das war die Geburtsstunde der Kryptoanalyse, weil die monoalphabetischen Chiffrierungen die Häufigkeiten in die Geheimtexte übertragen. Man weiss nicht genau, wer die Methode der Häufigkeitsanalyse erfunden hat und wann sie erfunden wurde. Vielleicht wurde sie auch aus verständlichen Gründen zuerst geheim gehalten. Die älteste bekannte schriftliche Offenbarung um das Jahr 900 kommt vom Gelehrten al-Kindī. Seine Schriften wurden erst 1987 in einem Archiv in Istanbul entdeckt. Interessant ist zu beobachten, dass bereits damals wie auch heute Folgendes zu den derart geschützten Themen gehörte: Militärgeheimnisse, Finanzen und Technologien.

Was Sie gelernt haben

Die monoalphabetischen Chiffrierungen übertragen die Häufigkeiten von Buchstaben vom Klartext auf den Geheimtext. Das öffnet den Zugang zum Knacken der monoalphabetischen Kryptosysteme auch in Fällen, in denen die Anzahl der Schlüssel so gross ist, dass man nicht alle ausprobieren kann. Wenn der Klartext hinreichend lang ist und somit die Buchstabenhäufigkeiten statistisch repräsentativ sind (entsprechend der Verteilung der Buchstabenhäufigkeit in typischen Texten der zugrunde liegenden Sprache), dann kann man davon ausgehen, dass die häufigsten Zeichen des Geheimtextes auch die statistisch am häufigsten vorkommenden Buchstaben kodieren. Wenn die Annahmen stimmen, kann man relativ schnell per Hand, also auch ohne Computer, aus dem Lückentext den Klartext rekonstruieren. Dabei können auch die häufigsten Bigramme, Trigramme und die Doppelbuchstaben helfen.

Stochastik und polyalphabetische Kryptosysteme

Nachdem mit dem Wissen um die Buchstabenhäufigkeiten die Kryptoanalyse alle monoalphabetischen Kryptosysteme ohne allzu grossen Aufwand knacken konnte, entstand der Bedarf, Kryptosysteme zu entwickeln, bei denen die Analyse der Buchstabenhäufigkeit keine Chance hat. Obwohl es in jener Zeit noch keine Wahrscheinlichkeitstheorie gab, ist die Buchstabenhäufigkeitsanalyse sowie die Entwicklung von neuen Kryptosystemen und Angriffen auf dieselben stark mit stochastischen Überlegungen verbunden. In diesem Abschnitt lernen Sie die Grundkonzepte der stochastisch basierten Kryptologie, die bis zum Anfang des 20. Jahrhunderts die Lehre der Geheimschriften geprägt hat.

2.32 Der folgende Geheimtext wurde mit einer speziellen Verallgemeinerung des CAESAR-Kryptosystems mit einem unbekannten Schlüssel (i, j) chiffriert.

```
OGOPF FJGDN FXFTT VFPHG IGJOT EITJH UGOFF
THGTE IKDJU G
```

Alle Buchstaben auf ungeraden Positionen wurden mit CAESAR mit Schlüssel i chiffriert und alle auf geraden Positionen mit Schlüssel j. Dechiffrieren Sie den Geheimtext.

2.33 Man hat sich ein neues Kryptosystem MASKIERT-CAESAR überlegt. Das Kryptosystem hat als Schlüssel ein Paar (Bitfolge, j). Zuerst wird die Bitfolge unter einem Klartext wiederholt hingeschrieben, sodass unter jedem Buchstaben des Klartextes ein Bit steht. Die Bitfolge verwendet man als eine Maske. Wenn eine 0 unter einem Buchstaben steht, wird der Buchstabe einfach in den Geheimtext kopiert. Falls dort eine 1 steht, wird der Buchstabe mit dem Schlüssel j chiffriert.
Den ersten Teil der Chiffrierung nennen wir Maskierung. Zum Beispiel sieht die ganze Chiffrierung für den Schlüssel (10110001, 3) wie folgt aus:

```
G U T E K R Y P T O G R A P H E N M U E S S E N P F I F F I G S E I N
1 0 1 1 0 0 0 1 1 0 1 1 0 0 0 1 1 0 1 1 0 0 0 1 1 0 1 1 0 0 0 1 1 0 1
J U W H K R Y S W O J U A P H H Q M X H S S E Q S F L I F I G V H I Q
```

Der folgende Geheimtext ist mit dem Schlüssel (10110001, j) chiffriert und j ist unbekannt. Dechiffrieren Sie den Geheimtext:

```
Q I S Q N E G Q  T E U Y S C H D  U F F Q N Z U G  Z T I U C K E X
Z E D R O R D Q  D T E O H O E Z  Q N Q G E I D Q  Q N
```

Die Kryptosysteme in den Knobelaufgaben 2.32 und 2.33 sind polyalphabetisch. Die Kodierung der einzelnen Buchstaben hängt von ihrer Position im Klartext ab. Worin besteht der Vorteil? Zum Beispiel wird der häufigste Buchstabe E abhängig von der Position im Klartext durch zwei unterschiedliche Buchstaben chiffriert. Somit wird es im Geheimtext keinen Buchstaben mehr geben, der so häufig vorkommt wie E im Klartext. Also werden die Differenzen zwischen Buchstabenhäufigkeiten im Geheimtext viel kleiner. Das ist die Geburtsidee für stochastische Kryptosysteme, in denen in Geheimtexten alle Buchstaben ungefähr gleich häufig vorkommen.

Jules Verne und Kryptoanalyse

Beispiel 2.6

In einem von Jules Vernes Romanen geht es von Beginn an bis zum Ende um die Dechiffrierung eines Geheimtextes. Der Schlüssel ist eine Dezimalzahl. Ähnlich wie bei der Maskierung mit der Bitfolge schreibt man die Zahl Ziffer für Ziffer unter einen Klartext. Die Chiffrierung der Buchstaben erfolgt mit CAESAR. Jeder Buchstabe wird um so viele Positionen verschoben, wie die Ziffer unter den Buchstaben angibt.
Der Geheimschlüssel in dem Roman war 432513 und der Klartext wurde wie folgt chiffriert:

Der Schlüssel 432513 enthält 5 unterschiedliche Ziffern, nämlich 1, 2, 3, 4 und 5. Somit wird jeder Buchstabe abhängig von seiner Position im Text durch 5 unterschiedliche Buchstaben kodiert. Zum Beispiel wird jeder Buchstabe E im Klartext abhängig von seiner Position (abhängig von der darunter stehenden Ziffer des Schlüssels) durch einen der fünf Buchstaben F, G, H, I oder J im Geheimtext chiffriert. Somit verteilt sich die Häufigkeit von E auf 5 Buchstaben und die Differenzen der Buchstabenhäufigkeiten im Geheimtext schmelzen massgeblich.

```
D E R W I R K L I C H E U R H E B E R D E S D I A M A N T E N R A U B S …
4 3 2 5 1 3 4 3 2 5 1 3 4 3 2 5 1 3 4 3 2 5 1 3 4 3 2 5 1 3 4 3 2 5 1 3 …
H H T B J U O O K H I H Y U J J C H V G G X E L E P C S U H R U C Z C V …
```

2.34 In dem folgenden Geheimtext ist der Name des Romans von Jules Verne chiffriert. Der Schlüssel ist 1830. Wie heisst der Roman?

```
B K K T   I C Q D   F Z W M   F Q O E   O I X F   E M P A   N I C O   O I V
```

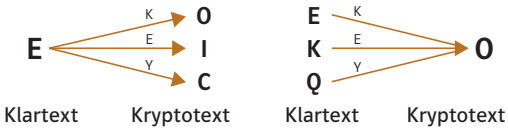
Kryptosystem VIGENÈRE

Beispiel 2.7

Das Kryptosystem aus Beispiel 2.6 hat einen Nachteil. Durch die Verwendung von Ziffern sind Verschiebungen nur um 0, 1, 2, …, 9 Positionen möglich. Wie ermöglicht man die nicht verwendeten Verschiebungen?
Das Kryptosystem VIGENÈRE verwendet statt dezimalen Zahlen (Ziffernfolgen) Wörter oder Texte (Buchstabenfolgen) als Schlüssel. Genau wie in Beispiel 2.6 schreibt man den Schlüssel wiederholend unter den Klartext, bis unter jedem Buchstaben des Klartextes genau ein Buchstabe des Schlüssels steht. Beim Chiffrieren verschiebt man jeden Buchstaben im Klartext um so viele Positionen im Alphabet, wie die Ordnung des darunter stehenden Buchstabens ist.
Wenn der Schlüssel KEY ist, sieht die Verschlüsselung wie folgt aus:

```
D A S I S T D A S T O R I N D A S D E R S C H L U E S S E L P A S S T
K E Y K E Y K E Y K E Y K E Y K E Y K E Y K E Y K E Y K E Y K E Y K E
N E Q S W R N E Q D S P S R B K W B O V Q M L J E I Q C I J Z E Q C X
```

Was bedeutet das für die Häufigkeiten der Buchstaben in Geheimtexten? Der Buchstabe E wird durch drei Buchstaben O, I und C chiffriert (Bild unten links).



Man kann ungefähr eine gleichmässige Verteilung der relativen Häufigkeit 17,7 % des Vorkommens von E auf die drei Buchstaben O, I und C erwarten, nämlich jeweils rund 5,9 %. Deswegen ist die Abbildung rechts noch aussagekräftiger. Wir können die geschätzte relative Häufigkeit des Buchstabens O im Geheimtext wie folgt ausrechnen:

$$h_{\text{Geheimtext}}(O) = \frac{h_{\text{Klartext}}(E)+h_{\text{Klartext}}(K)+h_{\text{Klartext}}(Q)}{3} = \frac{17{,}7\,\% + 1{,}4\,\% + 0{,}01\,\%}{3} = \frac{19{,}2\,\%}{3} = 6{,}4\,\%$$

2.35 Berechnen Sie die erwartete relative Häufigkeit der Buchstaben A, B und C in Geheimtexten, falls die Geheimtexte mit dem Schlüssel KEY im Kryptosystem VIGENÈRE verschlüsselt worden sind.

2.36 Entwickeln Sie ein Programm, das für einen gegebenen Schlüssel von VIGENÈRE für jeden Buchstaben des Geheimtextes die erwartete relative Häufigkeit des Vorkommens des Buchstabens ausrechnet.
Wenden Sie das Programm an, um die erwarteten relativen Häufigkeiten der Buchstaben in Geheimtexten für die folgenden Schlüssel zu berechnen:

a KEY
b SECRETKEY
c AXBMNEURSTCYDHDG

Sie können die Resultate mit Tabellen oder grafisch als Balkendiagramme darstellen. Was beobachten Sie? Welcher der Schlüssel gleicht die relativen Häufigkeiten der Buchstaben in Geheimtexten am besten aus? Wie würden Sie einen Schlüssel wählen, um gleiche relative Häufigkeiten aller Buchstaben in Geheimtexten zu erreichen?

Neue Konzepte und Begriffe

Die Schlüssel des Kryptosystems VIGENÈRE sind beliebige Texte (Buchstabenfolgen), auch **Schlüsselwörter** genannt. Vor der Chiffrierung schneidet man den Klartext in Stücke der Länge des Schlüsselwortes und jedes Stück wird mit der gleichen Chiffrierungsmethode verschlüsselt. Man schreibt das Schlüsselwort so unter das Stück des Klartextes, dass unter jedem Buchstaben des Klartextes genau ein Buchstabe des Schlüsselwortes steht. Dann wird jeder Buchstabe des Klartextes modulo 26 um so viele Positionen verschoben wie die Ordnung des darunter stehenden Buchstabens des Schlüsselwortes.

Für den Bau eines Kryptosystems ist der Begriff der Sicherheit von zentraler Bedeutung. Dieser hat sich seit der Antike, wo es nur um die Geheimhaltung der Geheimschriften ging, wesentlich verändert. August Kerkhoffs formulierte im 19. Jahrhundert die folgende Definition der Sicherheit, die heute als **Kerckhoffs' Prinzip der Sicherheit** bekannt ist:

«Ein Kryptosystem ist **sicher**, wenn das Kryptosystem öffentlich bekannt ist und trotzdem ohne die Kenntnis des verwendeten Schlüssels es nicht möglich ist, aus einem Geheimtext den ursprünglichen Klartext abzuleiten.»

Das Kryptosystem VIGENÈRE galt drei Jahrhunderte lang aus der Sicht des Kerckhoffs' Prinzips als sicher.

2.37 Der folgende Geheimtext entstand durch die Verschlüsselung mit dem Kryptosystem VIGENÈRE. Wir wissen, dass der Klartext auf Deutsch ist und dass E der häufigste Buchstabe (20 Vorkommnisse) im Klartext ist. Schaffen Sie es, den Klartext zu rekonstruieren und das Schlüsselwort zu bestimmen, wenn bekannt ist, dass die Schlüssellänge 3 ist?

```
ETF ICS KFH NMD ROD NPG NFR EMA SUZ
EOC IHD SIZ NED LOT NEC ACD IBT SFH
GFM EOE EIK ESM DBR VFQ SUZ EOC NJR
ZVU ESS IFE EO
```

Beispiel 2.8

Angriff auf VIGENÈRE

In Knobelaufgabe 2.37 greift man den Geheimtext an, wenn man die Schlüssellänge kennt und wenn man die Tatsache weiss, dass der ursprüngliche Klartext deutschsprachig ist. Dazu gibt es eine allgemeine Strategie. Betrachten wir den folgenden Geheimtext, der wie in Knobelaufgabe 2.37 aus einem deutschsprachigen Klartext durch die Chiffrierung mit dem Kryptosystem VIGENÈRE mit einem Schlüssel der Länge 3 entstanden ist.

```
IBU IXJ LML JHY WVO IKP RXY VXP GAA
MAY ISP IEL RBL NXK ILY ILB PMH XBZ
XYB IKZ MXU YKL MGZ GAY MMA DNT AXP
XXU XYL VGA IGG MXS EFO SKP DHU X
```

Den Geheimtext haben wir absichtlich wegen der Schlüssellänge 3 als eine Folge von Trigrammen (Buchstabenfolgen der Länge 3) geschrieben. Wir wissen, dass in allen 36 Teilen die ersten Buchstaben mit der gleichen Verschiebung von CAESAR verschlüsselt sind. Unter diesen 36 Buchstaben des Geheimtextes kommt am häufigsten der Buchstabe I vor (9-mal). Deswegen vermuten wir, dass I den Buchstaben E kodiert, was eine Verschiebung von 4 bedeuten würde. Somit nehmen wir E mit Ordnung (E) = 4 als den ersten Buchstaben des Schlüsselwortes an.
Die zweite Gruppe von Buchstaben bilden die 35 Buchstaben des Geheimtextes, die auf der zweiten Position der Trigramme stehen. Hier kommt der Buchstabe X am häufigsten (8-mal) vor. Deswegen nehmen wir an, dass X den Buchstaben E kodiert. Die Verschiebung wäre dann 19 und somit wäre T der zweite Buchstabe des Schlüsselwortes.
Bei der dritten Gruppe der 35 letzten Buchstaben der Trigramme ist die Statistik nicht so einfach. Die Buchstaben L, P und Y sind die häufigsten und kommen jeweils 5-mal vor. Als beste Strategie ergibt sich somit, die Vermutungen für die ersten zwei Gruppen einzusetzen und zu versuchen, aus dem Lückentext die richtige Wahl aus L, P und Y für die Kodierung von E zu treffen.

```
EI-EE-HT-FO-SC-ER-NE-RE-CH-IH-EZ-EL-
NI-JE-ES-ES-LT-TI-TF-ER-IE-UR-IN-CH-
IT-ZU-WE-TE-TF-RN-EN-IE-AM-OR-ZO-T
```

Von den drei Möglichkeiten L (Verschiebung 7), P (Verschiebung 15) und Y (Verschiebung 24) führt nur die Verschiebung um 7 zu folgendem sinnvollen Text: «Eine echte Forscherin erreicht ihre Ziele nie. Jedes Resultat ist fuer sie nur ein Schritt zum weit entfernten Ziel am Horizont.» Somit ist das Schlüsselwort ETH.
Man kann nicht ausschliessen, dass der häufigste Buchstabe in einer der Gruppen gerade nicht E kodiert. Das muss man durch Ausprobieren der entsprechenden Verschiebungen der Buchstaben erkennen und die Vermutung zurücknehmen. Dann könnte man versuchen, den zweithäufigsten Buchstaben als Kodierung von E anzunehmen. Alternativ kann man auch vermuten, der häufigste Buchstabe der Gruppe kodiere den Buchstaben N. Wiederholtes Probieren ist oft unvermeidbar und ein regelmässiger Bestandteil der Kryptoanalyse. Besonders bei kurzen Texten oder langen Schlüsselwörtern wird die Grösse der einzelnen Gruppen so klein, dass man sich nicht vollständig auf die statistischen Merkmale verlassen kann.

2.38 Entwickeln Sie ein Programm in Python, das für einen Geheimtext (Folge von Buchstaben) und eine Zahl m die Buchstaben in m Gruppen unterteilt, wobei
- Gruppe 1: Buchstaben auf Positionen 1, m+1, 2m+1, 3m+1, …
- Gruppe 2: Buchstaben auf Positionen 2, m+2, 2m+2, 3m+2, …
- …
- Gruppe m: Buchstaben auf Positionen m, 2m, 3m, …,

und das für jede Gruppe die stochastische Verteilung der Häufigkeiten der Buchstaben berechnet.
Testen Sie Ihr Programm mit den Geheimtexten aus Knobelaufgabe 2.37 und Beispiel 2.8.

2.39 Arbeiten Sie in Gruppen.
Jede Gruppe wählt einen deutschsprachigen Klartext und ein Schlüsselwort. Die Länge des Textes geteilt durch die Länge des Schlüsselwortes sollte mindestens 40 ergeben.
Verschlüsseln Sie den Klartext mit VIGENÈRE und dem gewählten Schlüsselwort.
Danach tauschen die Gruppen die Geheimtexte und die Länge der Schlüsselwörter.
Die Gruppen nutzen die in Aufgabe 2.38 entwickelten Programme, um die Geheimtexte zu knacken.

Geschichtlicher und gesellschaftlicher Kontext

Der Vorschlag, die Verschiebung im CAESAR-Kryptosystem immer nach der Chiffrierung von ein paar Wörtern zu wechseln, kam schon im Jahr 1466 vom italienischen Gelehrten Leon Battista Alberti (1404–1472).

Die VIGENÈRE-Tabelle zeigt für jeden Buchstaben des Klartextes, durch welchen Buchstaben dieser bei einem gegebenen Buchstaben des Schlüssels verschlüsselt wird. So wird z. B. der Buchstabe E, der auf den Buchstaben K des Schlüssels trifft, durch den Buchstaben O verschlüsselt.

Klartextbuchstabe / Schlüsselbuchstabe

	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

Der Ursprung des Kryptosystems VIGENÈRE liegt in der «Tabula recta» (Quadratische Tafel) in der Abbildung, die der deutsche Benediktiner Johannes Trithemius (1462–1516) im Jahr 1508 veröffentlicht hat. Jede Zeile der Tafel entspricht einer CAESAR-Verschiebung und seine Idee war es, während der Chiffrierung die verwendeten Zeilen zu wechseln. Zum Beispiel sollte man nach der Chiffrierung eines Buchstabens die nächste Zeile für den folgenden Buchstaben nehmen.

Das war immer noch eine feste Geheimschrift und kein Kryptosystem, aber sie war polyalphabetisch und glich die Häufigkeiten der Buchstaben im Geheimtext aus. Im Jahr 1553 entwickelte der Italiener Giovan Battista Bellaso anhand der Tafel das vorgestellte Kryptosystem VIGENÈRE. Das Schlüsselwort bestimmte die Reihenfolge der Verwendung der einzelnen Zeilen der Tafel.

Der Franzose Blaise de Vigenère (1523–1596) hat die vorhandenen Ideen in einem Werk zusammengefasst und weitere Verallgemeinerungen vorgeschlagen. Somit erhielt das vorgestellte Kryptosystem den Namen VIGENÈRE und wurde insbesondere in diplomatischen Kreisen das Mass der geheimen Kommunikation. VIGENÈRE galt über 300 Jahre als sicher.

Erst im Jahr 1854 erschuf der berühmte Entwickler des ersten programmierbaren Computers, Charles Babbage (1791–1871), eine Methode zum Knacken von VIGENÈRE. Er hat sie aber nie publiziert. Der Deutsche Friedrich Wilhelm Kasiski (1805–1881) hat diese Methode (heute Kasiski-Test genannt) zum Brechen von VIGENÈRE 1863 veröffentlicht.

Charles Babbage

Blaise Vigenère

Mit den technischen Möglichkeiten des digitalen Zeitalters wurden immer grössere Datenmengen gesammelt, verwaltet, analysiert und nach Bedarf schnell für unterschiedliche Zwecke auf beliebige Entfernungen zur Verfügung gestellt. Dies führte dazu, dass die Geheimhaltung von Daten nicht mehr wie ursprünglich nur auf Finanzen, Militär, Diplomatie und Wirtschaft gerichtet war. Heute sprechen wir von **sensiblen Daten**, die auf Personen bezogen sind. Dazu zählen Daten bezüglich der Gesundheit, des Privatlebens, der politischen Meinung, der Weltanschauung sowie genetische und biometrische Daten. Oft, wie z. B. in sozialen Medien oder beim Publizieren eigener Webseiten, können wir selbst entscheiden, welche Informationen wir über uns veröffentlichen. Dies ist aber nicht immer so. Unsere Gesundheitsdaten beispielsweise sind in Datenbanken des Krankenhauses, bei der Versicherung oder bei unserem Hausarzt gespeichert und wir haben Anspruch auf die Geheimhaltung dieser Daten. Hier sind wir aber abhängig davon, wie gut die Datenbankverwalter ihre Datenbanken schützen.

Welche Möglichkeiten stehen uns zur Verfügung, um unsere Daten zu schützen? Den Zugang zu unserem Computer und den dort gespeicherten Daten schützen wir mit Passwörtern, die so gewählt werden müssen, dass man sie nicht leicht erraten kann. Oft sind aber unsere Daten irgendwo im Netz bei einem Dienstleister aufbewahrt und wir haben keine Ahnung, wo und wie sie abgespeichert sind und wie sie geschützt sind. Falls die Geheimhaltung dieser Daten für uns wichtig ist, können wir sie mit ausgewählten Geheimschriften chiffrieren. Dann sind sie doppelt geschützt, nämlich mit dem Zugangspasswort und mit einer Chiffrierung.

Beispiel 2.9 **Bestimmung der Schlüssellänge mit dem Kasiski-Test**

Wir wissen schon, wie VIGENÈRE mit bekannter Schlüssellänge zu brechen ist. Wie bestimmen wir aber die Schlüssellänge? Die Idee ist folgende: Stellen wir uns vor, in einem Klartext kommt das Trigramm EIN 120-mal vor. Wenn das Schlüsselwort BALMER ist, kann EIN höchstens auf sechs (die Länge des Schlüsselwortes) unterschiedliche Arten chiffriert werden. Welche der sechs Chiffrierungen angewandt wird, hängt von dem Teil des Schlüsselwortes unter EIN ab.

```
  E I N       E I N       E I N       E I N       E I N       E I N
R B A L M E  B A L M E R  A L M E R B  L M E R B A  M E R B A L  E R B A L M
  F I Y       E T Z       P U R       Q M E       I Z R       V J N
```

Somit muss mindestens eine dieser sechs Chiffrierungen von EIN oft (mindestens 20-mal) im Geheimtext vorkommen. Auf diese Weise findet man z. B. sehr viele Trigramme FIY im Geheimtext.
Wenn man annimmt (was stochastisch vertretbar ist), dass das Trigramm FIY die Chiffrierung eines gleichen Trigramms des Klartextes mit dem gleichen Teil (Teilwort) des Schlüssels (BAL in unserem Fall) ist, müssen alle Entfernungen zwischen den ersten Buchstaben F von FIY im Geheimtext Mehrfache der Schlüssellängen sein. Warum ist das so? Das ist so, weil unter EIN immer der gleiche Teil BAL des Schlüssels stehen muss und das Schlüsselwort BALMER wiederholend unter dem Klartext platziert ist.

```
E I N E F R A G E W I E S E I N E S E I N O D E R N I C H T S E I N
B A L M E R B A L M E R B A L M E R B A L M E R B A L M E R B A L M
F I Y               E T Z   F I Y               E T Z

H A T I N D E R W E L T D E R Q U A N T E N K E I N E E I N D E U T I G E
E R B A L M E R B A L M E R B A L M E R B A L M E R B A L M E R B A L M E
                                                E T Z

A N T W O R T
R B A L M E R
```

In dem Beispiel sehen wir, dass das Trigramm FIY zweimal (Positionen 1 und 19) im Geheimtext und ETZ dreimal (Positionen 14, 32 und 62) vorkommt. Wegen FIY muss $19 - 1 = 18 = 2 \cdot 3^2$ ein Vielfaches der Schlüssellänge sein. Wegen ETZ müssen folgende Differenzen ein Vielfaches der Schlüssellänge sein:
$62 - 14 = 48 = 2^4 \cdot 3$
$62 - 32 = 30 = 2 \cdot 3 \cdot 5$
$32 - 14 = 18 = 2 \cdot 3^2$
Weil alle drei erhaltenen Zahlen 18, 30 und 48 Vielfache der Schlüssellänge sein müssen, muss der grösste gemeinsame Teiler GGT (18, 30, 48) = 6 ein Vielfaches des Schlüsselwortes sein.

Somit kann die Schlüssellänge nur 1, 2, 3 oder 6 sein. Bei hinreichend langen Klartexten erhält man so viele Anhaltspunkte, dass meistens der GGT der tatsächlichen Schlüssellänge entspricht. Man könnte bei der Bestimmung der Schlüssellänge auch mit Bigrammen arbeiten. Hier ist aber die Wahrscheinlichkeit höher, dass gleiche Bigramme im Geheimtext nicht durch die Chiffrierung der gleichen Bigramme des Klartextes mit dem gleichen Schlüsselteil, sondern zufällig entstehen.

Neue Konzepte und Begriffe

Das Kryptosystem VIGENÈRE kann man so angreifen, dass man zuerst die Länge des Schlüssels bestimmt. Das tut man mit dem Kasiski-Test, indem man gleiche Trigramme im Geheimtext sucht. Die Abstände dieser Trigramme sind Mehrfache der Schlüssellänge. Wenn man die Schlüssellänge kennt, teilt man den Geheimtext in entsprechend viele Gruppen und wendet in einzelnen Gruppen die Buchstabenhäufigkeitsanalyse an.

2.40 Der folgende Geheimtext ist eine VIGENÈRE-Chiffrierung eines deutschsprachigen Klartextes. Der Geheimtext beinhaltet wiederholt vorkommende Trigramme wie IOO. Bestimmen Sie mit dem Kasiski-Test die Schlüssellänge und dechiffrieren Sie anschliessend den Geheimtext.

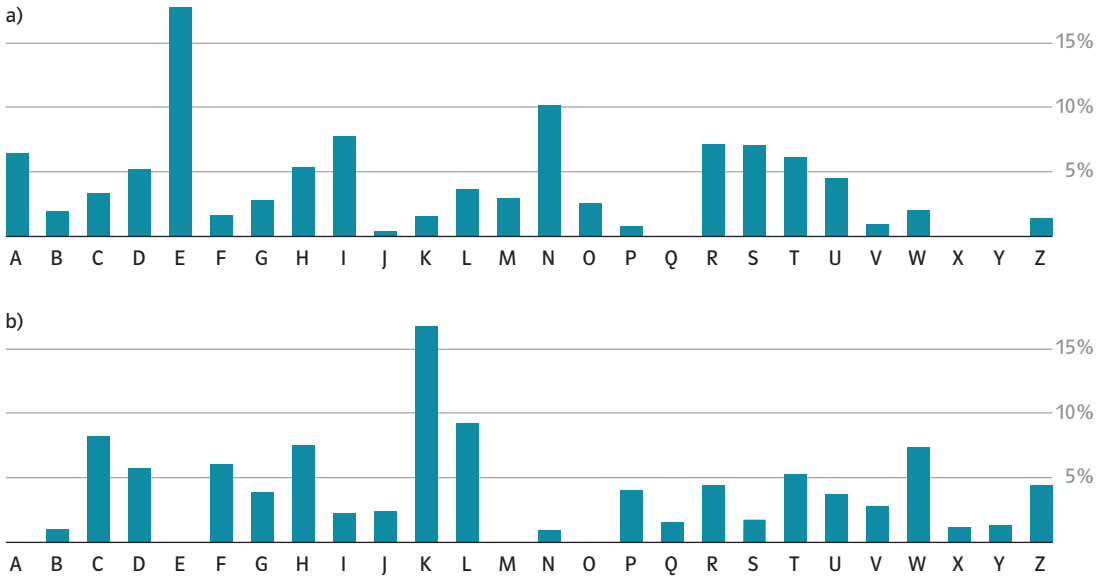
```
U E U I O O C E U D I W I O O W R R C L W V U Q U R R C L W V N D T H X E T H E R R
C F K R T W V S F Y O Q R N J J T G R S V U A V I O O C E Q E I H R U I Y O H I T Q L N
J Z N J V S E V R J R U I L N G U E U I O O C E U D I W I O O W H R V R W V A X W Z X I
O O C E Q I O O W A W D E W V A X W U Q U W R C L W V D L V N D V C K J T H E T D
X E Y F M U F L O V R Q Z C K K S P V H U Y O H I E Q
```

2.41 Entwickeln Sie in Gruppenarbeit ein Programm, das für einen gegebenen Klartext und ein Schlüsselwort den Geheimtext mit VIGENÈRE verschlüsselt als Ausgabe berechnet. Jede Gruppe wendet das Programm an, um einen längeren Klartext mit selbst gewählten Schlüsseln zu verschlüsseln. Dann tauschen die Gruppen die Geheimtexte aus. Die Aufgabe ist dann, die Länge des Schlüsselwortes zu bestimmen. Dazu kann man auch hilfreiche Programme entwickeln. Ein Programm sollte für ein gegebenes Trigramm (oder allgemein ein Wort) alle Vorkommnisse des Trigramms in einem gegebenen Text finden und die Positionen der ersten Buchstaben der Trigramme ausgeben. Ein zweites Programm kann für gegebene Zahlen den grössten gemeinsamen Teiler ausrechnen.

Beispiel 2.10

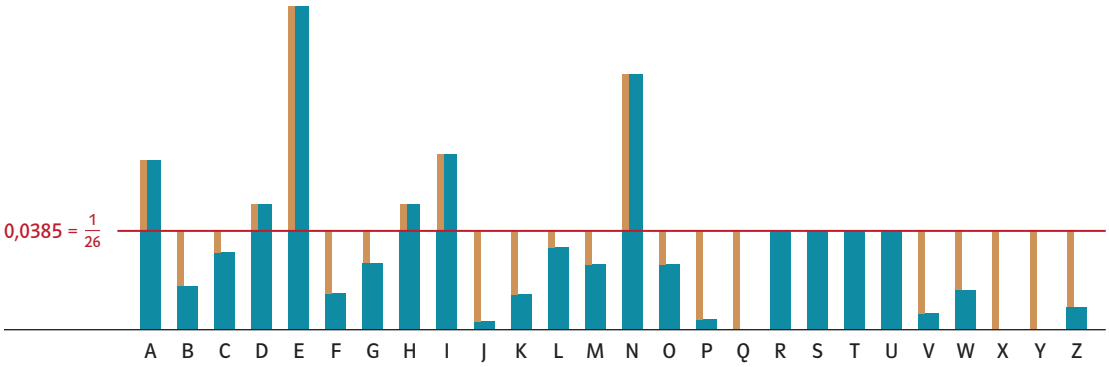
Bestimmung der Schlüssellänge mit Friedman'scher Charakteristik

Die Verteilung der relativen Häufigkeiten der einzelnen Buchstaben in statistisch repräsentativen deutschsprachigen Texten kann man mit folgendem Histogramm gut visualisieren:



Das erste Histogramm (a) zeigt die statistische relative Häufigkeit der Buchstaben in einem typischen deutschsprachigen Text und das zweite Histogramm (b) zeigt die relative Häufigkeit der Buchstaben in einem monoalphabetisch verschlüsselten Kryptotext. Man kann erkennen, dass die Balken ungefähr gleich gross sind, jedoch auf unterschiedlichen Positionen. Wir sagen, dass die Balken «permutiert» sind. Wenn also eine Buchstabenverteilung wie in Histogramm (b) aussieht, nehmen wir sofort an, dass der Klartext monoalphabetisch verschlüsselt wurde. Der Buchstabe K kodiert E und der Buchstabe L kodiert wahrscheinlich N. Wenn das Histogramm eines Geheimtextes ganz anders aussieht (wenn z. B. keine grossen Unterschiede in Buchstabenhäufigkeiten bestehen), schliessen wir auf eine polyalphabetische Chiffrierung.

Eine Vorgehensweise in der Mathematik ist, komplexere Objekte wie z. B. Häufigkeitsverteilungen (Histogramme) durch eine Zahl zu beschreiben, wir sagen auch «zu charakterisieren». Es ist klar, dass dabei viel Information über die Häufigkeitsverteilung verloren geht, also kann man aus der Zahl die Verteilung nicht rekonstruieren. Die Zielsetzung ist aber, dass die Zahl die wesentliche Information trägt, die wir für unsere Zwecke brauchen.



$$|h_A(T) - \tfrac{1}{26}| + |h_B(T) - \tfrac{1}{26}| + |h_C(T) - \tfrac{1}{26}| + \ldots + |h_Y(T) - \tfrac{1}{26}| + |h_Z(T) - \tfrac{1}{26}|$$

Im Histogramm der Häufigkeitsverteilung der Buchstaben in deutschsprachigen Texten (blau) ist eine rote Linie $y = \frac{1}{26} = 0{,}0385$ gezeichnet. Wenn die Buchstaben gleich häufig vorkämen, würden alle Balken genau die Höhe $\frac{1}{26}$ haben und wir hätten die sogenannte Gleichverteilung. Wir wollen mit einer Zahl ausdrücken, wie weit sich die betrachtete Häufigkeitsverteilung von einer Gleichverteilung unterscheidet.

Eine Möglichkeit wäre, die absoluten Differenzen der Häufigkeiten der Buchstaben von $\frac{1}{26}$ (die orangen Balken in der Abbildung) zu addieren. Aus Gründen, die man eventuell im Stochastikunterricht erfahren kann, ziehen die Mathematiker es vor, die Quadrate der Differenzen zu nehmen.

Neue Konzepte und Begriffe

Die Friedman'sche Charakteristik eines Textes T definiert man als:

$$FC(T) = \left(h_A(T) - \tfrac{1}{26}\right)^2 + \left(h_B(T) - \tfrac{1}{26}\right)^2 + \ldots + \left(h_Z(T) - \tfrac{1}{26}\right)^2 = \sum_{\Delta \in \text{Alphabet}} \left(h_\Delta(T) - \tfrac{1}{26}\right)^2$$

Die Summenbezeichnung Σ ⬚ ist eine kurze Beschreibung einer Summe und entspricht einer Schleife beim Programmieren. Man summiert die Quadrate der Differenzen $h_\Delta(T) - \frac{1}{26}$ über alle 26 Buchstaben Δ im Alphabet.

Für einen Geheimtext T, in dem alle Buchstaben ungefähr gleich häufig (in $\frac{1}{26}$ Fällen) vorkommen, ist FC(T) nah an 0. Für typische Texte in deutscher Sprache ist FC(T) ungefähr 0,0385…

2.42 Berechnen Sie die Friedman'sche Charakteristik von Klartexten und Geheimtexten in Beispiel 2.5, Aufgabe 2.28, Beispiel 2.6 und 2.8 und Aufgabe 2.40 und vergleichen Sie sie paarweise. Anstatt dies von Hand zu machen, können Sie zuerst ein Programm entwickeln, das für einen gegebenen Text T (Folge von Buchstaben) die FC(T) berechnet.

2.43 Sei T_2 eine monoalphabetische Chiffrierung eines Klartextes T_1. Wie ist die Beziehung zwischen $FC(T_1)$ und $FC(T_2)$? Könnte die berechnete Information $FC(T_2)$ eines Geheimtextes T_2 helfen, um eine Angriffstrategie zur Dechiffrierung von T_2 zu wählen?

2.44 Bestimmen Sie die Friedman'sche Charakteristik von typischen Texten in englischer, französischer und italienischer Sprache. Kann diese Charakteristik hilfreich bei der Bestimmung der Sprache des Klartextes sein?

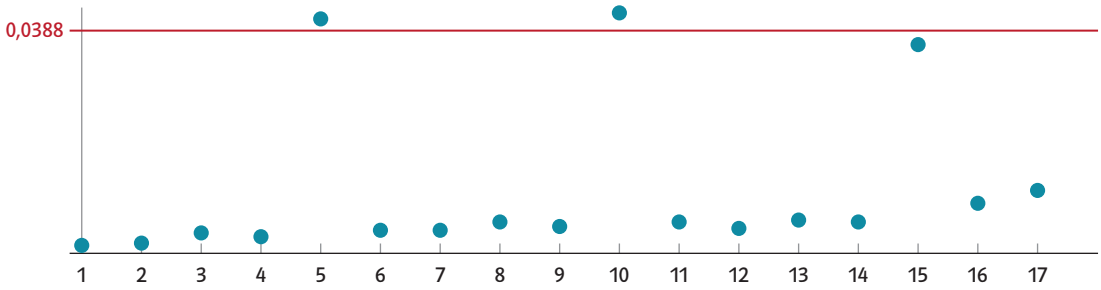
Fortsetzung

Beispiel 2.10 Die Friedman'sche Charakteristik kann man verwenden, um die Bestimmung der Länge des Schlüsselwortes von VIGENÈRE zu automatisieren.

Wenn der Geheimtext T den Wert FC(T) nah an 0,0388 hat, schätzt man, dass T monoalphabetisch chiffriert wurde, und man greift T mit der Buchstabenhäufigkeitsanalyse an. Wenn FC(T) nah an 0 ist oder mindestens viel kleiner als 0,0388, denkt man an eine polyalphabetische Chiffrierung.

In diesem Fall probieren wir für i = 2, 3, 4, … das folgende Verfahren. Für ein festes i schneiden wir den Geheimtext T in Textstücke der Länge i und verteilen Buchstaben aus T in i Gruppen T_1, T_2, …, T_i. Die Gruppe T_j(j = 1, …, i) beinhaltet alle j-ten Buchstaben der Textstücke der Länge i (wie in Beispiel 8 für i = 3). Dann berechnet man $FC(T_j)$ für alle j. Falls alle nah an 0,0388 liegen, kommt ein starker Verdacht, dass i ein Vielfaches der Schlüssellänge ist. Wenn die Werte $FC(T_j)$ näher an 0 liegen, ist i kein guter Kandidat für die Schlüssellänge. Statt alle Werte $FC(T_1)$, $FC(T_2)$, …, $FC(T_j)$ zu studieren, betrachtet man oft nur ihren Durchschnittswert dFC(T).

Man kann mit i = 2 anfangen und so fortfahren, bis man einen guten Kandidaten (dFC(T) nah an 0,0388) für die Schlüssellänge findet. Die Teilung des Geheimtextes T in Gruppen T_1, T_2, …, T_i sowie die Berechnung von dFC(T) kann man programmieren und somit automatisieren. Man kann die Resultate auch visualisieren. Die Abbildung zeigt die Analyse eines Kryptotextes, der mit einem Schlüssel der Länge 5 verschlüsselt wurde. Alle Vielfachen von 5 zeigen, dass dFC(T) wesentlich grösser ist als für Zahlen, die durch 5 nicht teilbar sind. Somit kann man aus der Grafik sofort die Schlüssellänge 5 schätzen.



2.45 Arbeiten Sie in Gruppen. Entwickeln Sie zuerst ein Programm zur Berechnung von FC(T) für eine Buchstabenfolge T. Nutzen Sie dieses Programm als Modul, um ein Programm zu entwickeln, das für eine Buchstabenfolge T und eine Zahl i die Werte $FC(T_j)$ für jedes j = 1, …, i wie in Beispiel 2.10 beschrieben berechnet. Wenden Sie dieses Programm als Modul zur Berechnung des Durchschnittswertes dFC(T) an. Letztendlich schreiben Sie ein Programm, das für die ersten 25 Zahlen i = 1, 2, …, 25 die Werte dFC(T) bestimmt und visualisiert. Wenden Sie das entwickelte Programm auf die Geheimtexte aus Aufgabe 2.40, Knobelaufgabe 2.37 und Beispiel 2.8 an und visualisieren Sie die Resultate. Danach nehmen Sie einen Text in deutscher Sprache von mindestens 3000 Zeichen und verschlüsseln Sie ihn mit VIGENÈRE mit einem Schlüssel der Länge von höchstens 10. Wenden Sie Ihr Programm auf den chiffrierten Geheimtext an. Warum unterscheidet sich die entstandene Grafik wesentlich von den vorherigen drei? Diskutieren Sie die möglichen Gründe dafür.

Geschichtlicher und gesellschaftlicher Kontext

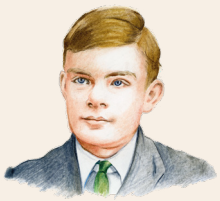
Mit der erfolgreichen Kryptoanalyse von VIGENÈRE hat man wieder die Frage eröffnet, was Sicherheit bei Kryptosystemen bedeutet und ob überhaupt garantiert sichere Kryptosysteme gebaut werden können. Bis zu diesem Zeitpunkt war es eher ein intellektuelles Spiel, das sich wiederholend abspielte. Man entwickelte ein Kryptosystem, das so lange als sicher galt, bis ein anderer Wissenschaftler eine Idee entdeckt hatte, um dieses zu brechen.

Eines der bekanntesten Kryptosysteme nach VIGENÈRE war ENIGMA, das Kryptosystem der deutschen Wehrmacht während des Zweiten Weltkrieges. Das Grundkonzept stammt vom deutschen Ingenieur Scherbius (1878–1929), der solche raffinierten Chiffrierungsgeräte gebaut hat. Die Idee war, die Geheimschrift nach der Chiffrierung jedes einzelnen Buchstabens «automatisch» zu ändern, aber auf eine solch komplizierte Art und Weise, dass es sehr schwierig war, darauf zu kommen. Trotzdem gelang Alan Turing, einem der Gründer der Informatik, die heldenhafte Tat, ENIGMA zu brechen und somit Tausenden von Menschen das Leben zu retten.

Nach all diesem Hin und Her haben die Mathematiker als sicheres Kryptosystem ein solches definiert, für das keine Dechiffrierungsmethode existiert (also kann sie auch kein Genie entdecken), wenn man den geheimen Schlüssel nicht kennt. Leider haben die Mathematiker bewiesen, dass ein solches Kryptosystem in der Terminologie von VIGENÈRE einen Schlüssel braucht, der mindestens so lang ist wie der Klartext. Zusätzlich durfte man den Schlüssel nur ein einziges Mal verwenden. Weil man aber den Schlüssel vorab geheim ohne Kryptosystem verabreden musste, ist dieses Kryptosystem nicht wirklich praktisch.

Den Weg aus dieser Sackgasse haben die Fortschritte in der Entwicklung des Konzeptes der Berechnungskomplexität (Rechenaufwand) in der Informatik geöffnet. Man änderte «leicht» die Definition eines sicheren Kryptosystems: Statt der Nichtexistenz einer Dechiffrierungsmethode forderte man «nur» die Nichtexistenz einer effizienten Dechiffrierungsmethode. Wenn der beste Dechiffrierungsalgorithmus einen Aufwand von mehreren Millionen Jahren Computerarbeit forderte, fühlte man sich auf der sicheren Seite. Ronald L. Rivest (* 1947), Ali Shamir (* 1952) und Leonard Max Adleman (* 1945) entwickelten 1976 das berühmte RSA-Kryptosystem, was die Geburtsstunde der Kryptologie mit öffentlichen Schlüsseln war und die Basis des heutigen E-Commerce (E-Banking, Online-Shopping) ist. Das Besondere dabei sind zwei unterschiedliche Schlüssel – einer für die Chiffrierung und der andere für die Dechiffrierung. Den Schlüssel für die Chiffrierung darf man veröffentlichen und trotzdem schafft es niemand, davon den Dechiffrierungsschlüssel abzuleiten. Der Dechiffrierungsschlüssel ist das Geheimnis des Empfängers, der dieses Geheimnis mit niemandem teilen darf. Mehr über diese faszinierende Entwicklung der Kryptologie präsentieren wir im Lehrmittel für das Ergänzungsfach.

Public-Key-Kryptosysteme kann man auch zum Schutz eigener Daten verwenden. Falls man solchermassen chiffrierte Daten stehlen würde, wäre es unmöglich, die Daten zu dechiffrieren, etwa mit dem Ziel, davon zu profitieren oder vielleicht sogar erpresserisch tätig zu werden. Weil viele Datensammlungen unchiffriert aufbewahrt werden, passiert es immer öfter, dass Hacker in die Datenbanken einbrechen und versuchen, durch Erpressung oder Verwendung der gewonnenen Informationen Gewinne zu erzielen.


Alan Turing


Ronald L. Rivest


Ali Shamir


Leonard Max Adleman

Was Sie gelernt haben

Die grosse Schwäche der monoalphabetischen Chiffrierungen liegt in der Tatsache, dass man sie mit der Häufigkeitsanalyse mit relativ geringem Aufwand brechen kann. Deswegen entwickelte man polyalphabetische Kryptosysteme, die anstreben, die Häufigkeit der Buchstaben in Geheimtexten gleichmässig verteilt zu haben. Eines der bekanntesten polyalphabetischen Kryptosysteme ist VIGENÈRE. Der Schlüssel ist hier ein Text (Wort), den man Buchstabe für Buchstabe unter den Klartext bei der Chiffrierung schreibt. Jeder Buchstabe des Klartextes wird um genau so viele Positionen im Alphabet verschoben, wie die Ordnung des darunter stehenden Buchstabens ist. Schlüssel, die Permutationen der Buchstaben des Alphabets sind, führen zu einer Gleichverteilung der Buchstabenhäufigkeiten in den Geheimtexten.

Das Kryptosystem VIGENÈRE kann so angegriffen werden, dass zuerst die Schlüssellänge bestimmt wird. Der Kasiski-Test sucht dazu nach gleichen Trigrammen im Geheimtext, die mit hoher Wahrscheinlichkeit gleiche Trigramme des Klartextes chiffrieren. Die Entfernungen zwischen diesen Trigrammen sind mathematisch gesehen Vielfache der Schlüssellänge. Deswegen reicht es, den GGT aller dieser Entfernungen zu berechnen. Dieser GGT oder einer seiner Teiler ist die gesuchte Schlüssellänge.

Eine andere Möglichkeit, die Schlüssellänge von VIGENÈRE zu bestimmen, ist stochastischer Natur. Mit Friedman'scher Charakteristik kann man die Buchstabenverteilung im Text einfach durch eine Zahl charakterisieren. Diese Charakterisierung durch eine Zahl ist so gut, dass man aus dieser Zahl ablesen kann, ob der Text monoalphabetisch oder polyalphabetisch chiffriert wurde. Dadurch kann man mittels Probieren die Schlüssellänge bestimmen.

Mit bekannter Schlüssellänge m teilt man die Buchstaben des Geheimtextes in m entsprechende Gruppen, die jeweils durch die gleiche CAESAR-Verschiebung verschlüsselt worden sind. In jeder dieser Gruppen kann man die Häufigkeitsanalyse verwenden, um die jeweilige Verschiebung zu bestimmen.

Zusammenfassung

Das Bedürfnis, Daten vor Unbefugten zu schützen, ist mindestens so alt wie die Schriften selbst. Die Kryptologie, die Lehre der Geheimschriften, ist mindestens 4000 Jahre alt. Sie besteht aus den Teilbereichen Kryptografie (die Wissenschaft des Entwurfes von Kryptosystemen) und Kryptoanalyse (die Entwicklung von Methoden zum Brechen von Kryptosystemen).
Die ältesten Geheimschriften aus der Antike basierten auf Transpositionen und Substitutionen. Die Datensicherheit basierte damals auf der Geheimhaltung der ganzen Geheimschrift, weshalb die Geheimschrift nicht schriftlich aufbewahrt werden durfte und somit auswendig gelernt werden musste.
Kryptosysteme sind Sammlungen von Geheimschriften, die uns ermöglichen, abwechselnd unterschiedliche Geheimschriften zu verwenden. Die Namen (Spezifikationen) der einzelnen Geheimschriften nennen wir die Schlüssel des Kryptosystems. Nach dem Sicherheitsprinzip von Kerckhoffs darf das Kryptosystem bekannt sein und die Sicherheit nur auf der Geheimhaltung des Schlüssels basieren.
Eine Geheimschrift heisst monoalphabetisch, wenn jeder Buchstabe des Klartextes unabhängig von seiner Position im Text durch die gleiche Folge von Symbolen ersetzt wird. Solche Kryptosysteme kann man mit relativ kleinem Aufwand brechen, wenn man weiss, in welcher natürlichen Sprache der Klartext geschrieben worden ist. Das kommt daher, dass sich die Häufigkeit der Buchstaben im Klartext auf die Häufigkeit ihrer Chiffrierungen im Geheimtext überträgt und dass die Häufigkeiten der einzelnen Buchstaben sehr unterschiedlich sein dürfen.
Bei polyalphabetischen Chiffrierungen kann ein Buchstabe abhängig von seiner Position im Text durch unterschiedliche Symbole oder Symbolfolgen ersetzt werden. Das bekannteste polyalphabetische Kryptosystem des Mittelalters war VIGENÈRE, das über 300 Jahre als sicher galt. Bei diesem System teilt man den Klartext in Stücke der Länge des Schlüsselwortes ein, und alle Stücke werden gleich chiffriert. Das Schlüsselwort schreibt man Buchstabe für Buchstabe unter jedes Stück des Klartextes. Jeder Buchstabe des Klartextes wird im Alphabet um so viele Positionen verschoben, wie die Ordnung des darunter stehenden Buchstabens des Schlüsselwortes ist. VIGENÈRE kann man brechen, indem man zuerst die Schlüssellänge über Distanzen zwischen gleichen Trigrammen des Geheimtextes oder durch stochastische Überlegungen schätzt. Wenn man die Schlüssellänge kennt, kann man mit der Buchstabenhäufigkeitsanalyse den Geheimtext dechiffrieren.

Testen Sie sich selbst

Konzepte und Begriffe

1 Welche zwei Methoden hat man zum Entwurf von Geheimschriften in der Antike entwickelt?
2 Was war das Grundkonzept der Sicherheit in der Antike?
3 Welche Komponente beinhaltet eine vollständige Beschreibung einer Geheimschrift?
4 Listen Sie Beispiele von Geheimschriften auf, bei denen der Chiffrierungsalgorithmus und der Dechiffrierungsalgorithmus identisch sind.
5 Warum entsprechen Chiffrierungen injektiven Funktionen von Texten über das Klartextalphabet nach Texten über das Geheimtextalphabet?
6 Was ist der Unterschied zwischen einer Geheimschrift und einem Kryptosystem?
7 Was bezeichnet das Fachwort «Schlüssel» bei einem Kryptosystem?
8 Wann ist eine Geheimschrift monoalphabetisch? Listen Sie einige Beispiele auf.
9 Wie definiert man die absolute und die relative Häufigkeit von Buchstaben in einem Text? Welche Merkmale des Klartextes werden bei einer monoalphabetischen Chiffrierung in den Geheimtext übertragen?
10 Wie und unter welchen Voraussetzungen funktioniert die Kryptoanalyse von monoalphabetischen Kryptosystemen? Welche Rolle können dabei die häufigsten Bigramme und Trigramme spielen?
11 Was fordert das Kerckhoffs' Prinzip der Sicherheit? Erfüllen alle Kryptosysteme mit riesiger Schlüsselzahl (die man nicht alle ausprobieren kann) dieses Sicherheitsprinzip?

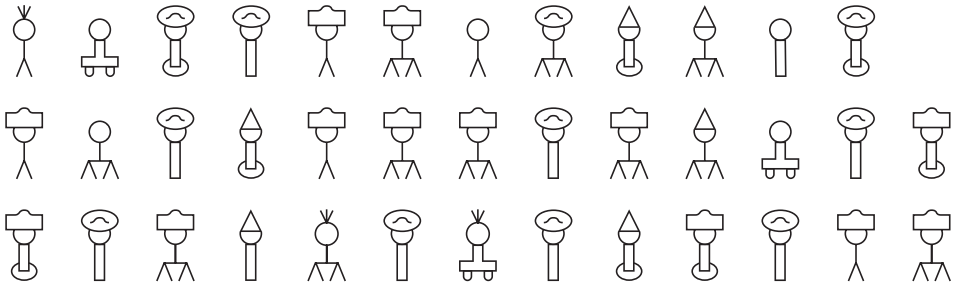
12 Wann ist eine Geheimschrift polyalphabetisch? Aus welchem Grund führte man polyalphabetische Chiffrierungen ein? Listen Sie einige Beispiele auf.
13 Wie funktioniert das Kryptosystem VIGENÈRE? Wie muss man die Schlüsselwörter wählen, um eine Gleichverteilung der Buchstabenhäufigkeiten im Geheimtext zu erreichen?
14 Wie kann man mit der Buchstabenhäufigkeitsanalyse VIGENÈRE knacken, wenn die Schlüssellänge bekannt ist?
15 Wie kann man die Schlüssellänge einer VIGENÈRE-Chiffrierung bestimmen?

Aufgaben

2.1 Der folgende Geheimtext wurde mit einem der vorgestellten antiken Kryptosysteme chiffriert. Der Geheimschlüssel ist 5. Dechiffrieren Sie den Geheimtext, indem Sie zuerst schätzen, um welches Kryptosystem es sich handeln kann.

```
GVMES EOEMS HNRAE EDDDN IAAES MTSRC HEHDH
ANAAA LWUTF TAPET URTND NITWE GMHIS
```

2.2 Mit der Strategie aus Beispiel 2.2 entwickelte man ein neues Alphabet mit 25 neuen Zeichen. Um die 26 Buchstaben des lateinischen Alphabets auf die neuen 25 Zeichen durch eine Substitution abzubilden, hat man die Buchstaben V und W zu einem Symbol zusammengefasst.
Dechiffrieren Sie den Geheimtext und rekonstruieren Sie die Chiffrierungstabelle.



2.3 Berechnen Sie die Friedman'sche Charakteristik des folgenden Geheimtextes:

```
ALHDC RNIXQ HEDHX GLFJH LQHLO HRCXB ERQDC
HORLO HLOHP QHUQL RQALH DOYDE IAHRW NSJNP
PHORA LHRHR YHLBE HORL PQHUQ
```

Schätzen Sie danach, ob der Geheimtext monoalphabetisch oder polyalphabetisch chiffriert wurde.
Dechiffrieren Sie den Geheimtext unter der Annahme, dass die Buchstaben E, I und S in dieser Reihenfolge die häufigsten Buchstaben des Klartextes waren. Bestimmen Sie möglichst vollständig den Schlüssel.

2.4 Der folgende Geheimtext wurde polyalphabetisch basierend auf einer Idee von Leon Battista Alberti chiffriert. Die Leerzeichen wurden für die Vereinfachung der Dechiffrierung bewahrt.

```
LJQJMWYJW QJTS HGZZOYZG GRHRXZO SLNAL
KPL OZCVLTIOM NCMZ MRN NWCFRLTUDWP NOC
UBIZDYCICDOWC GTRPYPCP
```

Die Chiffrierung startet mit einem CAESAR-Schlüssel i (d.h. in der i-ten Zeile der «Tabula recta»). Jeweils nach der Chiffrierung von zwei Wörtern erhöht man den Schlüssel um 1 (geht man zur nächsten Zeile der Tabelle).
Dechiffrieren Sie den Geheimtext, auch wenn der Startschlüssel i unbekannt ist.

2.5 Chiffrieren Sie die Zusammenfassung dieses Kapitels mit der folgenden Geheimschrift, deren Chiffrierung Sie als ein Programm implementieren. Der i-te Buchstabe des Geheimtextes wird um (i−1) mod 26 Positionen verschoben. Anders gesagt: Den ersten Buchstaben chiffrieren Sie mit der ersten Zeile der «Tabula recta». Nach der Chiffrierung jedes Buchstabens nehmen Sie die nachfolgende Zeile der Tabelle für die Chiffrierung des nächsten Buchstabens. Bei der 26-sten Zeile springen Sie zurück zur ersten Zeile.
Berechnen Sie die Friedman'sche Charakteristik des Klartextes und des Geheimtextes und vergleichen Sie die beiden Ergebnisse. Erklären Sie die Gründe für den Unterschied.

2.6 Nutzen Sie den Kasiski-Test, um die Schlüssellänge der VIGENÈRE-Verschlüsselung zu bestimmen.

```
NEQSWRNEQDSPSRBOQBOVQMLJEIQCIJSWR
```

2.7 Der folgende Geheimtext wurde mit VIGENÈRE verschlüsselt. Bestimmen Sie den Schlüssel und dechiffrieren Sie den Geheimtext.

```
MBVOM KAMIG LXVFM BVWMB VHLXZ LMBVF
IVDMP BMKSX QGCGB XZLKA QXL
```

IMPRESSUM

INFORMATIK
Daten verwalten, schützen und auswerten
Grundlagen der Informatik für Schweizer Maturitätsschulen

Autorinnen und Autoren: Dr. Michael Barot, Prof. Dr. Britta Dorn, Dr. Ghislain Fourny, Prof. Dr. Jens Gallenbacher, Prof. Dr. Juraj Hromkovič (Leitung), Regula Lacher
Autorinnen und Autoren der Programmierumgebungen TigerJython und WebTJ: Joël Schneider, Renato Menta, Silvia La, Tobias Kohn, Nicole Trachsler
Entwicklung der E-Tutorials: Lukas Fässler (Leitung), Markus Dahinden, Oliver Probst
Entwicklung interaktiver Lernumgebungen: Tobias Aeschbacher, Juraj Hromkovič, Regula Lacher, Johan Stettler, Marc Widmer
Projektleitung und Redaktion: Edgar Brütsch, Sabina Schleuniger
Sprachliche Bearbeitung: Elke Buelow
Umschlag und Gestaltungskonzept: Hansen Typografische Gestaltung, Luzern; Sager Visuelle Gestaltung, Luzern
Satz: Typogräfin® Petra Wenger
Umschlagbild: iStock.com/octomesecam
Illustrationen: Cornelia Kandler, S. 27 und S. 32, Aurel Märki, S. 4, Vaidotas Kinčius, S. 6
Grafische Illustrationen: Typogräfin® Petra Wenger
Korrektorat: Stefan Zach, z.a.ch gmbh
Rechte und Bildredaktion: Simone Zöckler

1. Auflage, 2022

ISBN 978-3-264-84597-6

klett.ch/lehrwerke
www.klett.ch
info@klett.ch

Das Lehrwerk «Informatik für Schweizer Maturitätsschulen» unterstützt einen kompetenzorientierten Informatikunterricht nach dem neuen Rahmenlehrplan. Die Schülerinnen und Schüler werden angeleitet, selbstständig und selbstentdeckend die Informatik zu erkunden. Sie bauen entsprechendes Wissen und Denken anhand von anschaulichen Beispielen, Projekten, Knobelaufgaben und Übungen gezielt und schrittweise auf, festigen ihre Kompetenzen und überprüfen diese mittels Selbsttests.

Das Lehrwerk gliedert sich in drei Bände zu den Grundlagen der Informatik sowie in einen weiteren Band für Informatik im Ergänzungsfach.

Im Band «Daten verwalten, schützen und auswerten» lernen die Schülerinnen und Schüler …

- Informationen digital so darzustellen, dass die Daten als Informationsdarstellungen effizient bearbeitet werden können;
- Geheimschriften zu entwickeln, um Daten durch Chiffrierungen geheim zu halten, sowie einige Geheimschriften zu analysieren, um Geheimnisse zu lüften;
- Daten zu komprimieren und so robust darzustellen, dass Übertragungsfehler bestimmt und automatisch korrigiert werden können;
- Daten zu verwalten und so zu organisieren, dass gesuchte Informationen schnell gefunden und anschaulich dargestellt werden können;
- aus Daten neue Zusammenhänge und Gesetzmässigkeiten zu entdecken und Maschinen so zu programmieren, dass diese nach einem selbstständigen Training eine Expertise für gewisse Tätigkeiten erwerben.

ISBN 978-3-264-84597-6
9 783264 845976