



Bei der **Kryptoanalyse** wird versucht, durch Eigenschaften des Geheimtextes den Schlüssel oder den Klartext herauszufinden.

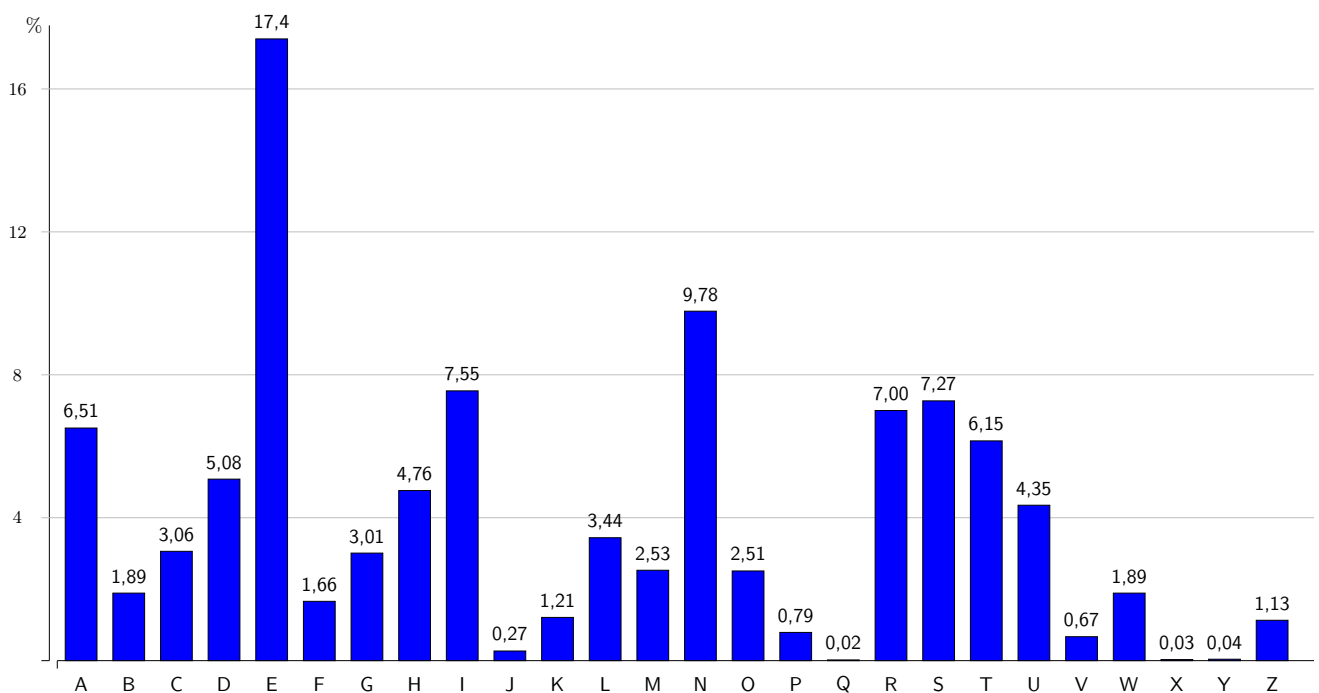
Um einen Geheimtext, der durch eine monoalphabetische Substitution entstanden ist, ohne bekannten Schlüssel zu entschlüsseln, verwendet man die **Häufigkeitsanalyse**.

Verschiedene Buchstaben werden in der deutschen Sprache unterschiedlich häufig verwendet. Dies kann bei einer Substitution dazu führen, dass einige Buchstaben auffällig häufig vorkommen. Das beste Beispiel dafür ist das **E**, da statistisch gesehen jeder sechste Buchstabe in einem deutschen Text ein **E** ist, muss auch in einem verschlüsselten Text ein Buchstabe besonders häufig vorkommen, der höchstwahrscheinlich dem **E** entspricht. Die Häufigkeiten der Buchstaben sind in der folgenden Tabelle und in der Grafik dargestellt. Dabei werden Umlaute ausgeschrieben (z.B. AE statt Ä).

A	B	C	D	E	F	G	H	I
6,51 %	1,89 %	3,06 %	5,08 %	17,4 %	1,66 %	3,01 %	4,76 %	7,55 %

J	K	L	M	N	O	P	Q	R
0,27 %	1,21 %	3,44 %	2,53 %	9,78 %	2,51 %	0,79 %	0,02 %	7,00 %

S	T	U	V	W	X	Y	Z
7,27 %	6,15 %	4,35 %	0,67 %	1,89 %	0,03 %	0,04 %	1,13 %



Neben den Wahrscheinlichkeiten der einzelnen Buchstaben werden noch andere statistische Daten bei der Kryptoanalyse verwendet. In deutschen Texten kommen einige Buchstabenkombinationen häufiger vor. Die Buchstabenkombination zweier aufeinander folgender Buchstaben nennt man *Bigramm*. Insgesamt gibt es 676 ($=26^2$) verschiedene Bigramme. Die zehn häufigsten Bigramme der deutschen Sprache sind in der folgenden Tabelle aufgelistet. Dabei fällt auf, dass die beiden eher seltenen Buchstaben **C** und **H** gemeinsam relativ häufig auftreten.

ER	EN	CH	DE	EI	ND	TE	IN	IE	GE
4,09 %	4,00 %	2,42 %	2,27 %	1,93 %	1,87 %	1,85 %	1,68 %	1,63 %	1,47 %

Bei der Analyse von Geheimtexten fallen Bigramme, die aus Doppelbuchstaben bestehen, besonders auf. Die zehn häufigsten Doppelbuchstaben sind:

SS	NN	LL	EE	MM	TT	RR	DD	FF	AA
0,76 %	0,43 %	0,42 %	0,23 %	0,23 %	0,23 %	0,15 %	0,13 %	0,12 %	0,08 %

Im Deutschen werden auch sehr viele kurze Wörter benutzt. Dies hilft bei der Entschlüsselung ebenfalls weiter. Die zehn häufigsten Wörter sind mit absteigender Häufigkeit:

die der und den am in zu ist dass es