Grundbegriffe der Informatik

Einheit 3: Alphabete
(und Relationen, Funktionen, Aussagenlogik)

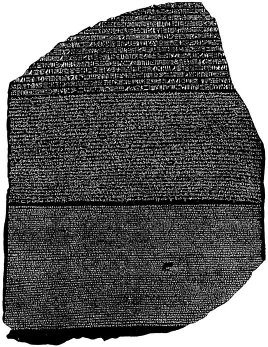
Thomas Worsch

Universität Karlsruhe, Fakultät für Informatik

Oktober 2008

Überblick

J.-F. Champollion
1790 – 1832



Rosetta-Stein
196 v. Chr.

wesentliche Beiträge zur Entzifferung der ägyptischen Hieroglyphen

Bildquellen: http://en.wikipedia.org/wiki/Image:Jean-Francois_Champollion_2.jpg
http://www.gutenberg.org/files/16352/16352-h/images/p1.jpg

- Ein *Alphabet* ist eine endliche nichtleere Menge sogenannter *Zeichen* oder *Symbole*.
- Was ist ein „Zeichen"?
  - Wir tun so, als wüssten wir das:
  - einfach (?) die elementaren Bausteine, aus denen Inschriften zusammengesetzt sind.
- Beispiele:
  - $A = \{|\}$
  - $A = \{\texttt{a}, \texttt{b}, \texttt{c}\}$
  - $A = \{\texttt{0}, \texttt{1}\}$
  - Manchmal erfindet man auch Zeichen: $A = \{\texttt{1}, \texttt{0}, \rotatebox{180}{\texttt{1}}\}$
  - $A = \{\texttt{0}, \texttt{1}, \texttt{2}, \texttt{3}, \texttt{4}, \texttt{5}, \texttt{6}, \texttt{7}, \texttt{8}, \texttt{9}, \texttt{A}, \texttt{B}, \texttt{C}, \texttt{D}, \texttt{E}, \texttt{F}\}$
- Gelegentlich etwas abstrakterer Standpunkt: Jeder der folgenden „Kästen" soll jeweils *ein* Zeichen eines gewissen Alphabetes sein: $\boxed{\texttt{int}}$ $\boxed{\texttt{adams}}$ $\boxed{\texttt{=}}$ $\boxed{\texttt{42}}$ $\boxed{\texttt{;}}$

ASCII-Zeichensatz

- ASCII: *American Standard Code for Information Interchange*.
- umfasst insbesondere eine Liste von 94 „druckbaren“ und einem „unsichtbaren“ Zeichen
- Buchstaben aus nichtenglischen Alphabeten, wie ä, ç, è, ğ, ñ, œ, ß, ů usw., fehlen,
  von Kyrillisch, Japanisch, . . . ganz zu schweigen.

ASCII-Zeichensatz

- Jedes Zeichen hat eine Nummer aus dem Bereich der natürlichen Zahlen zwischen 32 und 126.
- Beachte: das „Leerzeichen“ ␣

		40	(	50	2	60	<	70	F
		41	)	51	3	61	=	71	G
32	␣	42	*	52	4	62	>	72	H
33	!	43	+	53	5	63	?	73	I
34	"	44	,	54	6	64	@	74	J
35	#	45	-	55	7	65	A	75	K
⋮									⋮
84	T	94	^	104	h	114	r	124	|
85	U	95	_	105	i	115	s	125	}
86	V	96	‘	106	j	116	t	126	~
87	W	97	a	107	k	117	u		
88	X	98	b	108	l	118	v		
89	Y	99	c	109	m	119	w		

Unicode

- Unicode: http://www.unicode.org
- Der Unicode-Standard definiert viele Dinge
- Ausgangspunkt: eine *umfassende* Liste von Zeichen
  - rund 100 000 Zeichen
  - siehe http://www.unicode.org/charts/
  - Das ist also ein (großes) Alphabet.
- Weitere Aspekte:
  - Jedes Zeichen hat einen Namen. Z. B. „LATIN SMALL LETTER C WITH CEDILLA“ für ç
  - Sortierreihenfolge von Buchstaben (im Schwedischen: ö *nach* z, im Deutschen: ö *vor* z)
  - Zuordnung von Groß- zu Kleinbuchstaben und umgekehrt
  - und vieles mehr …

Unicode: Code Points

- Es wird ein Alphabet A_U festgelegt, und
- eine Nummerierung dieser Zeichen
  Jedenfalls in einem gewissen Sinne:
  - Jedem Zeichen aus A_U ist eine nichtnegative ganze Zahl zugeordnet, der *Code Point* des Zeichens.
  - Die Liste der benutzten Code Points ist aber nicht „zusammenhängend“.

- Jedenfalls liegt eine Beziehung zwischen Unicode-Zeichen und nichtnegativen ganzen Zahlen vor.
- Man spricht von einer Relation.
- Womit wir bei ein bisschen Mathematik wären ...

- Wie kann man die Beziehung zwischen Unicode-Zeichen in A_U und nichtnegativen ganzen Zahlen beschreiben?
- Z. B. durch die Angabe aller Paare (a, n),
  für die $a \in A_U$ ist und n der zu a gehörenden Code Point
- Für die Menge U aller dieser Paare gilt: $U \subseteq A_U \times \mathbb{N}_0$.

- *kartesisches Produkt $A \times B$* der Mengen A und B:
  Das ist die Menge *aller* Paare (a, b) mit $a \in A$ und $b \in B$:

$$A \times B = \{(a, b) \mid a \in A \text{ und } b \in B\}$$

- Eine Teilmenge $R \subseteq A \times B$ heißt auch eine *Relation* oder genauer *binäre Relation* von A in B.
- Beispiele:

$$(\mathtt{A}, 65) \in U \qquad \text{aber} \qquad (\mathtt{B}, 4711) \notin U$$

Die Unicoderelation $U \subseteq A_U \times \mathbb{N}_0$ hat zusätzliche „schöne“ Eigenschaften:

- Für jedes Zeichen $a \in A_U$ existiert (mindestens) ein $n \in \mathbb{N}_0$ mit $(a, n) \in U$.
  $R \subseteq A \times B$ heißt *linkstotal*, wenn für jedes $a \in A$ ein $b \in B$ existiert mit $(a, b) \in R$.
- Für kein Zeichen $a \in A_U$ gibt es mehrere $n \in \mathbb{N}_0$ mit der Eigenschaft $(a, n) \in U$.
  $R \subseteq A \times B$ heißt *rechtseindeutig*, wenn es für kein $a \in A$ zwei $b_1 \in B$ und $b_2 \in B$ mit $b_1 \neq b_2$ gibt, so dass sowohl $(a, b_1) \in R$ als auch $(a, b_2) \in R$ ist.
- Relationen, die linkstotal und rechtseindeutig sind, heißen *Abbildungen* oder *Funktionen*. Schreibweise $R : A \to B$.
- Gelegentlich (bei uns erst mal lange nicht) betrachtet man *partielle Funktionen*: man verzichtet auf die Linkstotalität und fordert nur Rechtseindeutigkeit.

Spezielle Abbildungen

- Bei Unicode gibt es keine zwei verschiedene Zeichen a_1 und a_2, denen der gleiche Code Point zugeordnet ist.
  $R \subseteq A \times B$ heißt *linkseindeutig*, wenn für alle $(a_1, b_1) \in R$ und alle $(a_2, b_2) \in R$ gilt:

  $$\text{wenn } a_1 \neq a_2, \text{ dann } b_1 \neq b_2 \; .$$

- Eine Abbildung, die linkseindeutig ist, heißt *injektiv*.
- $R \subseteq A \times B$ heißt *rechtstotal*, wenn für jedes $b \in B$ ein $a \in A$ existiert, für das $(a, b) \in R$ ist.
- Eine Abbildung, die rechtstotal ist, heißt *surjektiv*.
- Eine Abbildung, die sowohl injektiv als auch surjektiv ist, heißt *bijektiv*.

- *„Die Abbildung $U : A_U \to \mathbb{N}_0$ ist injektiv.“*
  Das ist eine Aussage. Sie ist *wahr*.
- *„Die Abbildung $U : A_U \to \mathbb{N}_0$ ist surjektiv.“*
  Das ist auch eine Aussage. Sie ist *falsch*.
- Aussagen sind Sätze, die „objektiv“ wahr oder falsch sind. Dazu braucht man eine Interpretation der Zeichen, aus denen die zu Grunde liegende Nachricht zusammengesetzt ist.
- Wir bauen ganz massiv darauf, dass es keine Missverständnisse durch unterschiedliche Interpretationsmöglichkeiten gibt.
  Was ist $\mathbb{N}$? Enthält es die 0?
- Manche umgangssprachlichen Sätze sind nicht wahr oder falsch sondern sinnlos:

  > *„Ein Barbier ist ein Mann, der alle Männer rasiert,*
  > *die sich nicht selbst rasieren.“*

  Rasiert sich ein Barbier selbst ...?

Häufig setzt man aus einfachen Aussagen ($\mathcal{A}$ und $\mathcal{B}$) kompliziertere auf eine der folgenden Arten zusammen:

logsiche Negation: „Nicht $\mathcal{A}$."
kurz $\neg\mathcal{A}$.

logisches Und: „$\mathcal{A}$ und $\mathcal{B}$."
kurz $\mathcal{A} \wedge \mathcal{B}$.

logisches Oder: „$\mathcal{A}$ oder $\mathcal{B}$."
kurz $\mathcal{A} \vee \mathcal{B}$.

logische Implikation: „Wenn $\mathcal{A}$, dann $\mathcal{B}$."
kurz $\mathcal{A} \Rightarrow \mathcal{B}$.

- Ob eine so zusammengesetzte Aussage wahr oder falsch ist, hängt dabei *nicht* vom konkreten Inhalt der Aussagen ab!
- Wesentlich ist nur, welche Wahrheitswerte die Aussagen $\mathcal{A}$ und $\mathcal{B}$ haben.

- *aussagenlogische Formel*
  - nach obigen Regeln zusammengesetzt und
  - statt elementarer Aussagen einfach *Aussagevariablen*,
  - die als Werte „wahr“ und „falsch“ annehmen können.

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$
falsch	falsch	wahr	falsch	falsch	wahr
falsch	wahr	wahr	falsch	wahr	wahr
wahr	falsch	falsch	falsch	wahr	falsch
wahr	wahr	falsch	wahr	wahr	wahr

- Das „Oder“ ist „inklusiv“ (und nicht „exklusiv“): Wenn A und B beide wahr sind, dann auch $A \vee B$.
- Man kann für komplizierte Aussagen anhand der obigen Tabellen „ausrechnen“, wenn sie wahr sind und wann falsch.
  - einfaches Rechnen bzw. scharfes Hinsehen: die Aussagen $\neg(A \vee B)$ und $(\neg A) \wedge (\neg B)$ immer gleichzeitig wahr bzw. falsch sind.
  - Gleiches gilt für $\neg\neg A$ und A.
  - Solche Aussagen nennt man *äquivalent*.
- Klammereinsparungsregeln:
  - $\neg$ bindet stärker als $\wedge$, $\vee$ und $\Rightarrow$.

Anmerkung zur Implikation

A	B	$A \Rightarrow B$
falsch	falsch	wahr
falsch	wahr	wahr
wahr	falsch	falsch
wahr	wahr	wahr

- Die Implikation $A \Rightarrow B$ ist auf jeden Fall wahr, wenn A falsch ist, unabhängig vom Wahrheitsgehalt von B.
- Die Aussage „*Wenn* $0 = 1$ *ist, dann ist* $42 = -42$*.*“ ist *wahr*.
- Das wird so noch etwas klarer:
  - Was ist denn das „Gegenteil“ von $A \Rightarrow B$?
  - Doch wohl $A \wedge \neg B$.
  - Also ist $A \Rightarrow B$ äquivalent zu $\neg(A \wedge \neg B)$
  - Das ist äquivalent zu $(\neg A) \vee (\neg\neg B)$ und
  - das ist äquivalent zu $(\neg A) \vee B$.

- Eine nützliche Notation aus der *Prädikatenlogik*:

  Allquantor $\forall$ *Existenzquantor* $\exists$

- In der puren Form hat eine quantifizierte Aussage eine der Formen

$$\forall x\, A(x) \qquad \text{oder} \qquad \exists x\, A(x)$$

- Dabei soll $A(x)$ eine Aussage sein, die von einer Variablen x abhängt (oder jedenfalls abhängen kann). A kann weitere Quantoren enthalten.
- $\forall x\, A(x)$ ist zu lesen als: „Für alle x gilt: $A(x)$".
- $\exists x\, A(x)$ ist zu lesen als: „Es gibt ein x mit: $A(x)$".
- Zum Beispiel:

$$\forall x\, (x \in \mathbb{N}_0 \Rightarrow \exists y\, (y \in \mathbb{N}_0 \wedge y = x + 1))$$

- noch mal dieses Beispiel:

$$\forall x \, (x \in \mathbb{N}_0 \Rightarrow \exists y \, (y \in \mathbb{N}_0 \wedge y = x + 1))$$

- Das hat man oft: Eine Aussage gilt nicht für alle x, sondern nur für alle x aus einer gewissen (Teil-)Menge M.
- Abkürzung: Statt

$$\forall x \, (x \in M \Rightarrow B(x))$$

schreibt man einfach

$$\forall x \in M : \; B(x)$$

- Doppelpunkt sinnvoll, wenn Lesbarkeit dadurch verbessert
- Obiges Beispiel wird zu:

$$\forall x \in \mathbb{N}_0 : \exists y \in \mathbb{N}_0 : y = x + 1$$