

Von Ugarit bis zum Unicode

Vortrag bei der Mitgliederversammlung des Normenausschusses Technische Grundlagen (NATG),
 Fachbereich A - Einheiten und Formelgrößen (AEF) - des Deutschen Instituts für Normung (DIN)
 Freudenstadt im Schwarzwald, 24. März 2004



Bild 1: Ein "Brief" aus Ugarit
 (etwa 1400 v.Chr.)
 Damaskus, Nationalmuseum




1 Einleitung: Blick zurück nach Ugarit

Ich erlaube mir, einen ziemlich weiten Bogen zu spannen, vom Altertum bis zur Gegenwart (und Zukunft). Und weil von Alphabeten (im weiteren Sinn) die Rede sein soll, fange ich mit einem der ältesten Alphabete an, mit dem Keilschrift-Alphabet von *Ugarit*. Warum gerade Ugarit ... das werden Sie am Schluss des Vortrages verstehen.

Bild 1 zeigt einen "Brief" aus Ugarit, mit dem Siegel des Urhebers in der Mitte. Ugarit war ein Stadtstaat im Gebiet des heutigen Syrien (nahe der Hafenstadt Latakia), der seine Blütezeit von etwa 1400 bis 1200 v.Chr. hatte. Dass es Ugarit gegeben haben muss, wusste man schon lange aus altägyptischen und hethitischen Texten, aber entdeckt wurde es erst 1928 und ausgegraben von 1929 an. Die Ugariter schufen um 1400 v.Chr. für ihre semitische Sprache eine alphabetische Keilschrift mit 31 Zeichen, die je einen Konsonanten bedeuteten. Das war ein großer Fortschritt gegenüber der mesopotamischen (zuerst sumerischen) Keilschrift mit etwa 600 Zeichen, die teils Wörter, teils Silben darstellten. Es

scheint, dass die Ugariter sich an das phönizische Alphabet anlehnten. Statt der phönizischen graphischen Zeichen, die letzten Endes von den ägyptischen Hieroglyphen abgeleitet sind, nahmen sie aber Keilschriftzeichen. Manche Gelehrte meinen übrigens, das ugaritische Alphabet sei das älteste. Jedenfalls überraschen die Namen der ersten Buchstaben des Alphabets (Bild 2) nicht:

Bild 2

Name	alpa	beta	gamla
Bedeutung	a*	b	g
Keilschrift			
		*Knacklaut	

Auf das Ugaritische komme ich am Schluss im Zusammenhang mit dem Unicode zurück.

2 Weitere historische Betrachtungen

Einerlei, ob nun das phönizische oder das ugaritische Alphabet zuerst da war: Die Griechen übernahmen um 900 v.Chr. das phönizische in umgestalteter Form, und zu unserem Glück kam ein Grieche auf den Gedanken, auch die Vokale zu schreiben. Offenbar teilte er nicht die Meinung, die *Thomas Mann* in seiner Erzählung *Das Gesetz* dem Moses unterstellt: Nämlich dass die "tönenden Leerlaute" (also die Vokale) sich von selbst verstünden.

Die Römer übernahmen, vielleicht über die Etrusker, die griechische Schrift, veränderten sie ein wenig, und schließlich hatten wir das weitere Glück, dass wir die Schrift der Römer übernehmen konnten; nur ein paar kleine Ergänzungen waren nötig.

Durch die ganzen Jahrtausende bis heute ist ein Satz von Schriftzeichen eine Konvention: In welcher Reihenfolge und zu welchem Muster muss man mit keilförmigen Griffeln Zeichen in den Ton

drücken? -- und später, in welcher graphischen Gestalt muss man Buchstaben schreiben oder drucken?
 Durch das Computerwesen kam etwas Neues dazu.

3 Die ersten Computer-Zeichensätze

3.1 ASCII

Als die Computer aufkamen, entstand der Bedarf an einer einheitlichen Zuordnung von Binär-Codewörtern zu den Schriftzeichen des Alphabets und den Satzzeichen usw. Da das Computerwesen sich zuerst in den USA entwickelte, dachte man zunächst nur an die Schriftzeichen, die dort gebraucht wurden und die auf den Schreibmaschinen-Tastaturen der USA zu finden waren. Man musste damals aber noch Speicherkapazität sparen und sah deshalb nur 7 bit je Zeichen vor, so dass $2^7 = 128$ Zeichen möglich waren. So ergab sich der *American Standard Code for Information Interchange*, kurz ASCII, der in der folgenden Tabelle 1 dargestellt ist.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	

Tabelle 1: ASCII

ASCII wurde 1968 in der US-Norm ANSI X3.4 -- letzte Fassung von 1977 -- veröffentlicht (**ANSI** = American National Standards Institute). ASCII ist, wie sich aus den folgenden Informationen ergeben wird, immer noch der Kern aller neueren genormten Alphabete. Man kann damit Sprachen schreiben, die das lateinische Alphabet benutzen und keine Sonderzeichen erfordern, also neben Englisch z.B. Afrikaans, Hawaiianisch, Latein und Suaheli.

In der obigen Tabelle sind die Plätze hexadezimal (eigentlich, nach DIN 1333 "Zahlenangaben", *sedezimal*) nummeriert, wie es meistens gemacht wird: Links von der Tabelle die erste "Hex"-Ziffer, über der Tabelle die zweite. Die Umrechnung von hexadezimal auf dezimal oder binär (eigentlich dual) oder in der entgegengesetzten Richtung ist leicht zu machen z.B. mit dem Rechner im Zubehör zu MS Windows oder dem Rechner KCalc bei Linux/KDE (dort *Basis* entsprechend wählen). Auf den Plätzen 00 bis 1F (dez. 0 bis 31), die hier nicht dargestellt sind, liegen Steuerzeichen (C0 controls) wie CR (carriage return, Zeilenwechsel); auf Platz 20 (dez. 32) ist die normale (nicht geschützte) Leerstelle. 7F ist wieder ein Steuerzeichen (delete).

3.2 ISO 646 mit nationalen Varianten

Aus ASCII entstand 1972 die Internationale Norm ISO 646 (**ISO** = International Organization for Standardization), jetzt ISO/IEC 646:1991. Nun sollten auch andere Sprachen geschrieben werden; deshalb wurden 15 nationale Varianten geschaffen, bei denen man ASCII-Zeichen, die als entbehrlich angesehen wurden, durch nationale Sonderzeichen ersetzte. So wurden für die "deutsche" Variante, die als DIN 66003 auch deutsche Norm wurde, die in Tabelle 1 grau unterlegten Zeichen ersetzt:

Hexadezimal-Nr.	40	5B	5C	5D	7B	7C	7D	7E
ASCII-Zeichen	@	[\]	{		}	~
Zeichen in ISO 646-DE	§	Ä	Ö	Ü	ä	ö	ü	ß

In Deutschland brauchte man eben, jedenfalls in der Bürokratie, das Zeichen §, während das @ ganz ungebräuchlich war ...

Diese nationalen Varianten sind für den internationalen Verkehr ungeeignet und somit veraltet.

3.3 Nicht-genormte Erweiterungen von ASCII

Um die Beschränkung auf 128 Zeichen zu überwinden, ging man von 7 bit auf 8 bit über. Zunächst gab es Lösungen einzelner Hersteller; zum Beispiel etablierte Microsoft die "Codepage 850", die für die "neuen" Plätze 80 bis FD (128-253) Zeichen enthält, mit denen nun auch u.a. deutsch, französisch und spanisch geschrieben werden konnte. Außerdem gab es graphische Elemente zum "Zeichnen" von Rahmen. Die Schriftart "Terminal", die vom weitverbreiteten MS-DOS-Editor (nicht zu verwechseln mit dem Windows-Editor) benutzt wird, beruht auf dieser "Codeseite". Die Schriftart "Terminal" wurde außerdem von den "Terminal-Programmen" benutzt, die "Mailboxen" (wie man sie damals nannte - in den USA "Bulletin Board Systems", BBSs) unterstützten. Diese waren populär, solange es noch kein Internet gab oder dieses noch keine große Verbreitung gefunden hatte.

Neben Microsoft haben auch andere Hersteller ihre Erweiterungen von ASCII eingeführt, so dass es wieder Probleme der Kompatibilität gab.

3.4 Die Norm ISO 8859 (jetzt ISO/IEC 8859)

Allgemeines

Als das Computerwesen sich international immer weiter ausbreitete, wurde klar, dass internationale Normen für die Alphabete aller möglichen Sprachen notwendig sind. Sie wurden im Rahmen der Norm ISO 8859 erstellt, die ebenfalls auf einer ANSI-Norm beruht. Man legte eine ganze Anzahl von Codetabellen mit 8 bit je Zeichen fest. In allen diesen Tabellen sind auf den Plätzen 00 bis 1F (dezimal 0-31) die Steuerzeichen gemäß ASCII und auf 20 bis 7E (32-126) die schreibbaren Zeichen nach ASCII; somit kann man mit allen Alphabeten der Familie "ISO 8859" auch Englisch usw. schreiben (vgl. oben bei 1). Die Plätze 80 bis 9F (128-159) wurden in der Norm frei gelassen (und prompt von Microsoft in der "Codepage 1252" für zusätzliche Zeichen benutzt, wodurch neue Probleme der Kompatibilität entstanden). Neu ist in der Norm also die Verwendung der Plätze A0 bis FF (dezimal 160 bis 255).

Die Norm ISO 8859 wird jetzt vom "Joint Technical Committee 1" (JTC1) von ISO und IEC bearbeitet und heißt dementsprechend ISO/IEC 8859 (aber man findet noch oft die alte Bezeichnung).

"Latein-Alphabete" 1 bis 10

- *Latin Alphabet No. 1* (bei Microsoft "Westeuropa" oder "Western"): ISO 8859-1.

Diese Codetabelle ist für die west- und nordeuropäischen Sprachen sowie Albanisch vorgesehen. Einige Zeichen, wie die französischen Ligaturen (aus O und E bzw. o und e) "Œ" und "œ" und der Euro € fehlen dennoch. Die Anwendung der Norm ISO 8859-1 ist derzeit (2004) auf PCs und im Internet wohl noch am meisten verbreitet.

Tabelle 3

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		ı	ø	£	¤	¥	ı	§	"	©	ª	«	¬		®	-
B	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Einige Zeichen sind eine Erläuterung wert:

- Bei A0 ist "no-break space", d.h. Leerstelle, an der ein Zeilenwechsel vermieden werden soll;
- Bei AD ist "soft hyphen", d.h. der "geschützte Bindestrich", ebenfalls ohne Zeichenwechsel;
- Bei B7 ist der mal-Punkt ".", bei D7 das mal-Kreuz × (für Vektorprodukt u.a.).

Die weiteren "**Latein-Alphabete**" (Nr. 2 bis 10) ermöglichen es, auch Deutsch zu schreiben, weil die Umlaute und das "ß" an den selben Plätzen sind wie bei Nr. 1. Mit Nr. 3, 5, 8 und 9 kann man auch Französisch schreiben.

– *Latin Alphabet No. 2* (bei Microsoft "Osteuropa" oder auch "Mitteleuropäisch"): ISO 8859-2.

Für slawische Sprachen, die lateinische Schriftzeichen benutzen, einschließlich des Sorbischen (der Sprache der Sorben in der Lausitz), sowie für Ungarisch und Rumänisch.

– *Latin Alphabet No. 3*: ISO 8859-3.

Für Maltesisch und Esperanto, ursprünglich auch für Türkisch (vgl. jedoch Latin Alphabet No. 5).

– *Latin Alphabet No. 4* (bei Microsoft "Baltisch"): ISO 8859-4.

Für die baltischen Sprachen (Litauisch, Lettisch, - vgl. jedoch No. 7 -, Estnisch, ursprünglich auch für Samisch (Sprache der Samen oder "Lappländer") und die Sprache der grönländischen Inuit ("Eskimo"), vgl. jedoch No. 6.

– *Latin Alphabet No. 5*: ISO 8859-9, für Türkisch.

– *Latin Alphabet No. 6*: ISO 8859-10, für Samisch und Inuit-Sprache (vollständiger als No. 4).

– *Latin Alphabet No. 7*: ISO 8859-13, für baltische Sprachen, insbes. Lettisch (ersetzt No. 4).

– *Latin Alphabet No. 8*: ISO 8859-14, für keltische Sprachen, insbesondere Gälisch und Walisisch.

– *Latin Alphabet No. 9*: ISO 8859-15, für denselben geographischen Bereich wie No. 1, enthält Zeichen wie die schon erwähnten französischen Ligaturen und den Euro, wird sich aber nicht mehr durchsetzen, weil es den Unicode schon gibt.

– *Latin Alphabet No. 10*: ISO 8859-16, für Französisch, Niederländisch und Türkisch besser als No. 1, sonst ähnlich No. 9 ... es wird wohl auch keine Bedeutung gewinnen.

Die übrigen Alphabete in ISO 8859: Sie gelten für Sprachen, die **nicht lateinische** Schriftzeichen benutzen. Da auch sie ASCII enthalten, kann man auch mit ihnen Englisch schreiben.

– *Latin/Kyrillic Alphabet*: ISO 8859-5, für Bulgarisch, Makedonisch, Russisch, Serbisch, Ukrainisch, Weißrussisch.

– *Latin/Arabic Alphabet*: ISO 8859-6, für Arabisch.

Es ist unvollkommen, weil es nicht alle vier Varianten der Schriftzeichen vorsieht (für die Stellung am Anfang eines Worts, innerhalb des Worts, am Ende, oder allein!). Erst der Unicode schafft Besserung.

– *Latin-Greek Alphabet*: ISO 8859-7, für Neugriechisch.

– *Latin-Hebrew Alphabet*: ISO 8859-8, für Hebräisch (Ivrit) und Jiddisch.

– *Latin-Thai Alphabet*: ISO 8859-11, für Thailändisch.



4 Der Unicode (Universal Character Set, UCS)

4.1 Allgemeines

Die Alphabete nach ISO/IEC 8859 sind zum Teil unvollständig; einige Zeichen fehlen im wichtigsten Alphabet, und viele Sprachen sind gar nicht vertreten. In den späten 1980er Jahren begannen zwei zunächst voneinander unabhängige Versuche, alle Mängel zu beseitigen durch die Schaffung eines einzigen universellen Zeichensatzes: ISO begann mit der Arbeit an der internationalen Norm ISO 10646 (diese Bezeichnung soll an den Ursprung, ISO 646 für ASCII, erinnern) und ein Konsortium von (hauptsächlich US-amerikanischen) Herstellern mit dem Projekt *Unicode*. Zum Glück werden die Arbeiten seit 1991 gemeinsam von ISO/IEC JTC1/SC2/WG2 und dem Unicode Consortium durchgeführt. Im Hinblick auf die Zuordnung von Codes zu Schriftzeichen usw. gibt es somit keine Unterschiede zwischen der Norm ISO/IEC 10646 und Unicode. Das Unicode Consortium definiert aber noch zusätzliche Algorithmen für Darstellungsformen mancher Schriften wie Arabisch sowie andere Hilfen für die Texterstellung. Man spricht heute allgemein nur vom Unicode.

Der Unicode wird in *einem* einheitlichen Nummerierungssystem die Schriftzeichen für alle Alphabete, die auf unserem Planeten im Gebrauch sind, erfassen, sowie für viele historische Alphabete, die nur Fachgelehrte benutzen; das meiste von dieser Arbeit ist bereits geleistet.

Einen Überblick über die Alphabete, die im Unicode erfasst werden, bietet die WWW-Seite **www.unicode.org/charts**; die "charts" sind Tabellen für Teilbereiche des Unicode. Sie können im PDF-Format heruntergeladen werden. Hier können nur einige Beispiele gegeben werden:

- Am leichtesten ist der Übergang zum Unicode für die Benutzer des "Latein-Alphabets Nr. 1" nach ISO/IEC 8859-1. Die Zeichen von 00 bis 7F (dez. 0 bis 127), jetzt "Basic Latin", und A0 bis FF (dez. 160 bis 255), jetzt "Latin 1 Supplement", behalten ihre Nummerierung. Für die Nummern 80 bis 9F (128 bis 159), die in ISO/IEC frei gehalten wurden (siehe oben!), fügt Unicode jedoch neue Steuerzeichen (C1 controls) ein, z.B. 86 "Start of selected area".

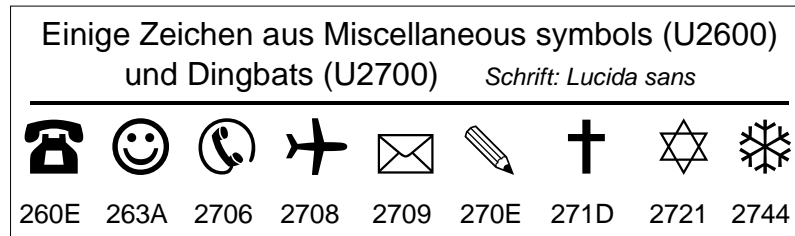
Wir finden somit im Unicode neben vielem anderem:

Bezeichnung	Ab Nr.	Zeichen für
Basic Latin	0000	ASCII (Englisch, Lateinisch u.a.)
Latin-1 Supplement	0080	West- und nordeuropäische Sprachen
Latin Extended-A	0100	Slaw. Sprachen, Ungarisch, baltische Sprachen u.a.
Greek	0370	Neugriechisch, Koptisch u.a.
Cyrillic	0400	Russisch, Ukrainisch, Bulgarisch u.a.
Hebrew	0590	Hebräisch (Ivrit), Jiddisch, Ladino (= Spaniolisch)
Arabic	0600	Arabisch, Persisch, Urdu u.a.
Devanagari	0900	Indische Sprachen: Hindi, Marathi
Currency symbols	20A0	Euro-Symbol u.a.

Tabelle 4: Einige Bereiche des Unicode (ohne die Steuerzeichen)

Außer Zeichen für die eigentlichen Text-Alphabete gibt es auch, ähnlich wie bei den nicht-genormten (herstellereigenen) Fonts wie etwa von Microsoft *Wingdings* und *Webdings*, eine große Anzahl von graphischen Elementen, z.B. *Miscellaneous symbols* (Tabelle U2600) und *Dingbats* (Tabelle U2700). Daraus sind die Kostproben, die in Bild 3 dargestellt sind.

Bild 3



Weiter gibt es sehr viele mathematische Zeichen (soweit ich es bisher feststellen konnte, außer einem Zeichen alle, die in DIN 1302 und im Entwurf zu ISO 80000-2 vorkommen), "Musical symbols" (Tabelle UID100) und vieles mehr. Wer den Rahmenelementen, die ich bei Codepage 850 erwähnt habe, nachtrauern sollte, kann sie in der Tabelle Box drawing (U2500) wiederfinden.

4.2 UTF-8 (UCS transformation format)

Damit in einem zusammenhängenden Text der Beginn des Codes jedes einzelnen Zeichens erkennbar ist, setzt man die Unicode-Werte in einer bestimmten Weise in eine neue Ziffernfolge um. Die Umsetzung wurde ebenfalls vom *Unicode Consortium* festgelegt. Der gegenwärtig gebräuchliche Stand ist das *UCS transformation format* UTF-8 (erfunden 1992). Die Rechenregel für die Umsetzung von Unicode in UTF-8 sieht so aus:

– Die Hex-Zahlen 00 bis 7F für die ASCII-Zeichen bleiben unverändert. Diese Zeichen werden auch weiterhin durch je *ein* Byte (zwei Hex-Ziffern) wiedergegeben. Jeder Texteditor, der auf lateinischer Schrift beruht, stellt sie richtig dar.

– Für den Bereich 80 bis 7FF gilt:

Den Hex-Zahlen entsprechen Binärwerte, die mit 11 bit dargestellt werden:

hhhhhiiiiii

Sie werden auf zwei Bytes in der Form

110hhhhh und **10**iiiiii

verteilt, wobei die fett dargestellten Bits fest vorgegeben sind. So wird z.B. der deutsche Umlaut ä im Unicode durch E4 repräsentiert, und die 11-stellige Binärzahl dazu ist 00011100100. Daraus werden nach der obigen Regel die Binärzahlen **110**00011 und **10**100100, denen wiederum die Hex-Zahlen C3 und A4 entsprechen. Das sind *zwei* Bytes. Ein Editor, der auf ISO 8859-1 beruht, stellt sie durch Ä und æ dar (vgl. Tabelle 3). Man sieht: Die Zeichen "oberhalb" von ASCII werden durch zwei Bytes wiedergegeben. Schon deutsche oder französische Texte werden dadurch in einem konventionellen Texteditor schwer lesbar. Texte in griechischen, kyrillischen, hebräischen und arabischen Schriftzeichen haben durchweg zwei UTF-Bytes je Schriftzeichen, also doppelt so viel wie bei Zeichensätzen nach ISO 8859; diese Sprachen sind allerdings auch bisher in einem auf ASCII oder ISO 8859-1 basierenden Texteditor nicht lesbar.

– In ähnlicher Weise geht es bei noch größeren Unicode-Werten weiter. Für den Bereich von 800 bis FFFF, in den z.B. indische Schriften wie Devanagari sowie das Euro-Zeichen und die meisten mathematischen Zeichen fallen, ergeben sich drei UTF-Bytes je Zeichen; darüber noch mehr.

– Für alle Zeichen von Nr. 80 an lässt das jeweils erste UTF-Byte in binärer Darstellung erkennen, wieviele Bytes für das gegebene Zeichen nachfolgen: Die Anzahl der führenden 1-Bits ist nämlich gleich der Anzahl der Bytes (im obigen Beispiel: zwei).

4.3 Der Unicode in der praktischen Anwendung

Durch RFC 2070 und 2247 (vergleichen Sie meinen Vortragstext vom 7. März 2001 über "*Internet - Gremien, Normen*"*) hat die *Internet Society* den Unicode in das Internet eingeführt und seinen Gebrauch in neuen Internet-Protokollen empfohlen. Die Windows-Betriebssysteme NT, 2000 und XP

unterstützen Unicode, ebenso das Betriebssystem *Linux* und die aktuellen Versionen der WWW-Browser *MS Internet Explorer*, *Netscape Navigator*, *Mozilla*, *Opera* und *Konqueror*, sowie die zugehörigen E-Mail-Programme. Der MS Internet Explorer tut das sogar, wenn das Betriebssystem (z.B. Windows 95) noch keine Unicode-Unterstützung hat.

Neuere Ausgaben von Schriften (fonts) wie *Arial*, *Helvetica*, *Times New Roman* und *Adobe Times* enthalten Zeichen gemäß Unicode für slawische Sprachen, Griechisch und Kyrillisch, und einige Schriften wie *Lucida Sans Unicode* haben Tausende von Unicode-Zeichen ... die Schrift *Arial Unicode MS* hat sogar etwa 40 000 Zeichen und eine Dateigröße von 23 MiB.

In *Microsoft Office* und *Open Office* kann man einzelne Unicode-Zeichen per Einfügen/Symbol bzw. Einfügen/Sonderzeichen aus Tabellen dieser Zeichen auswählen und in einen Text einfügen. Zum Schreiben regelrechter Texte in einem fremden Alphabet stellt Microsoft "visual keyboards" zur Verfügung, die auf dem Bildschirm erscheinen.

Im E-Mail-Programm *MS Outlook Express* kann man (wenn man das Format *Rich Text / HTML* und den Zeichensatz *UTF-8* gewählt hat) einzelne Unicode-Zeichen im Quelltext eingeben, in der Form

&#xHEX;

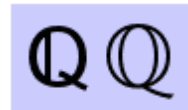
[mit dem Semikolon am Ende!] ... dabei steht HEX für die Hex-Nummer, also z.B. 3B1 für das griechische α . (Auch mit der entsprechenden dezimalen Nummer, hier 945, geht es, dann ohne das "x"). Das Programm rechnet die Eingabe in UTF-8 um.

Andere Textverarbeitungsprogramme und E-Mail-Programme dürften ähnliche Hilfsmittel zur Verfügung stellen.

4.4 Die graphische Gestaltung der Zeichen, die im Unicode erfasst sind

Die Tabellen des Unicode schreiben keine bestimmte graphische Ausprägung eines Zeichens vor. In den Erläuterungen zu diesen Tabellen (code charts) heißt es: "The shapes of the reference glyphs** used in the code charts are not prescriptive ... Considerable variation is to be expected in actual fonts".

Als Beispiele kann man anführen, dass der Code für den Buchstaben **A** (hex. 41, vgl. Tabelle 1) auch für **A** oder **Œ** gilt, oder dass das Zeichen für die Menge der rationalen Zahlen (the set of rational numbers, Nr. 211A) verschiedene Ausprägungen haben kann:



Die erste Fassung, mit dem charakteristischen senkrechten Strich, stammt aus *Adobe Times*, die zweite, m.E. nicht sehr geglückte, aus *Lucida sans*.

5 Noch einmal Ugarit

Jetzt komme ich auf das Ugaritische zurück, um den Bogen zu schließen. Dieses Alphabet ist nämlich das letzte, das in das System des Unicode aufgenommen worden ist ... mit der Tabelle U10380.

Mein letztes Bild zeigt eine Alphabet-Lehrtafel aus dem alten Ugarit und dazu aus der Tabelle U10380 einen Ausschnitt mit den ersten fünf Zeichen des ugaritischen Alphabets ... die Reihenfolge der Zeichen ist dieselbe!



10380 alpa	10381 beta	10382 gamla	10383 kha	10384 delta

Es scheint, dass es auch damals schon verschiedene Spielarten der Gestaltung der "Glyphen" gegeben hat.

Das Studium ugaritischer Texte ist zu einer Wissenschaft geworden, die einen erstaunlichen Umfang angenommen hat. Viele Forscher, ganze Verlage und Periodika widmen sich ausschließlich dem Ugaritischen. Man hat festgestellt, dass alttestamentarische Texte, wie Psalmen, vielfach auf kanaanäische Vorbilder zurückgehen, und diese sind auf ugaritischen Tontafeln überliefert. Und es gibt im Alten Testament Textpassagen, die im Lauf der Zeit beim Abschreiben verfälscht wurden und nun keinen Sinn mehr erkennen lassen. Durch den Vergleich mit den entsprechenden kanaanäischen Texten werden diese Passagen verständlich. Deshalb ist die Bibelwissenschaft an der ugaritischen Sprache und Schrift interessiert. So verwundert es auch nicht, dass ein Bedürfnis besteht, Ugaritisch auf dem PC zu schreiben, und zwar im Rahmen des Unicode. (Nicht-genormte Zeichensätze hat es dafür allerdings schon bisher gegeben).

*<http://home.arcor.de/lutz.schweizer/vortrag.pdf>

**Definition in Encarta World EnglishDictionary (<http://dictionary.msn.com>): "the symbol or symbols that form a single character in a font". Nach deutschen Lexika: Griechisch "glyphe" = Gravur, Schnitt.

Zum Teil aktualisiert: 10. Juli 2004

Konstruktive Kritik zu diesem Text ist willkommen!

Links zu Themen des Vortrags: <http://home.arcor.de/lutz.schweizer/NetLinks2.html>