

ՅՈՒՆԵՍԵՔՆԻ ԵՐԵՎԱՆԻ ԲՆԱԿԱՐԱՆԱԿԱՆ ԳՐԱԴԱՐԱՆ

ՔԱՆՈՆ

ՄԻՆԻՍՏԵՐՍՏՅԱՆ ԿԵՆՏՐԱԼԻ

Otvoreno računarstvo

Norme

- Uvod u norme
- Primjeri norma
- Norme kodnih stranica
- W3C norme

Mario Žagar



nôrma ž 〈G mn -ā/-ī〉

pravilo, propis, obrazac, kriterij prema kojem se određuje kako bi nešto trebalo biti

lat.: drvodjelski kvadar, pravilo, predložak

- Upravljanje kakvoćom – općenito o normama, uloga norma u poslovnom procesu
- Otvoreno računarstvo – norme u računarstvu i njihova primjena

Otvoreni sustavi, norme, normacija



- Skup definiranih norma
 - Olakšavaju suradnju
 - Olakšavaju povezivanje
 - Omogućuju konkurentnost
 - Omogućuju neovisnost o proizvođaču
- Govorimo o stvarnim sustavima i stvarnom životu, ne o apstraktnim pojmovima:
 - SAD - ANSI/IEEE
 - Europa - ISO
 - Hrvatska - ISO (Zavod za norme)
- Neki više vole udruge financirane od proizvođača
- Postojanje okvira norma potiskuje inovacije



Ključni pojmovi:

- donositelji norma moraju biti neovisni od proizvođača,
- otvoreni sustavi znače funkcionalne produkte, ne apstraktne pojmove,
- suradnja između različitih sklopovskih platforma,
- prenosivost aplikacija na razne platforme,
- lakoća prelaska sa starih sustava na nove,
- zaštita investicija u sadašnje sustave,
- umrežavanje i komunikacija bez obzira na fizičku lokaciju.

Postanak norme



- Skupina
 - ima
- Potreba
 - skupina ili netko drugi daje
- Prijedlog rješenja
 - te se dolazi do
- Dogovor (konsenzus) o prihvatanju
- Izrada i objavljivanje norme
- Osvremljenjivanje norme



Postanak norme – primjer AES

- Skupina
 - Američka vlada (NIST)
- Potreba
 - Zaštita podataka kript algoritmom nove generacije
- Prijedlog rješenja
 - Debata o postupku izbora algoritma – siječanj-rujan 1997.
 - Kandidati: CAST-256, CRYPTON, DEAL, DFC, E2, FROG, HPC, LOKI97, MAGENTA, MARS, RC6, Rijndael, SAFER+, Serpent, Twofish – kolovoz 1998.

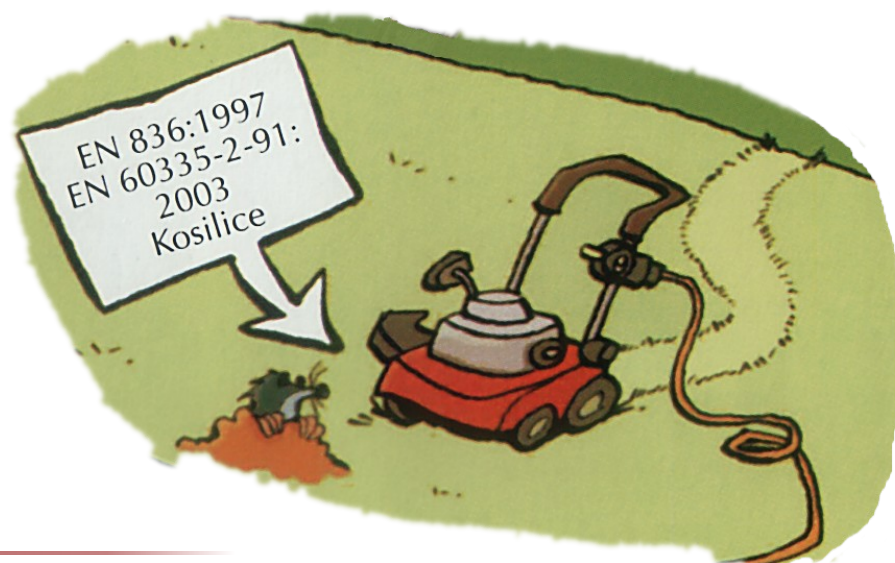
Postanak norme - AES

- Dogovor (konsenzus) o prihvatanju
 - Diskusija o kandidatima (2 skupa) – kolovoz 1998., ožujak 1999.
 - Finalisti – kolovoz 1999.
 - Rijndael, Serpent, Twofish, RC6, MARS
 - Analiza finalista (3. skup) – travanj 2000.
 - AES = Rijndael algoritam – listopad 2000.
- Izrada i objavljivanje norme
 - Odobrena FIPS (Federal Information Processing Standard) norma – veljača-studen 2001.

Postanak norme - AES



- Osvremenjivanje norme
 - Debata o svojstvima algoritma i primjenjivosti u pojedinim situacijama
 - Stalna kriptanaliza i pronalaženje ograničenja i slabosti algoritma
 - Primjer: evolucija DES u 3DES



Tko predlaže norme?

- Proizvođač
- Udruga
 - rezultat dogovora udruženih proizvođača
 - **udruga/proizvod** (*alliance/product*)
- Interesna udruga
 - vrijeme pojavljivanja krajnjeg proizvoda nepoznato
 - **udruga** (*alliance/marketing*)
- Norme nemaju vrijednost bez pravne podrške i podrške tržišta

Tko *donosi* norme?

- Formalno, neovisno tijelo
 - pravna norma
 - **de iure**
 - slijedi vodeću tehnologiju (dobitak na vremenu)

- Tržište
 - široko prihvaćen proizvod postaje norma
 - **de facto/proizvod** (*product*)
 - široko prihvaćen proizvod licenciran drugima
 - **de facto/licenca** (*licensable*)

Sloboda i rizik

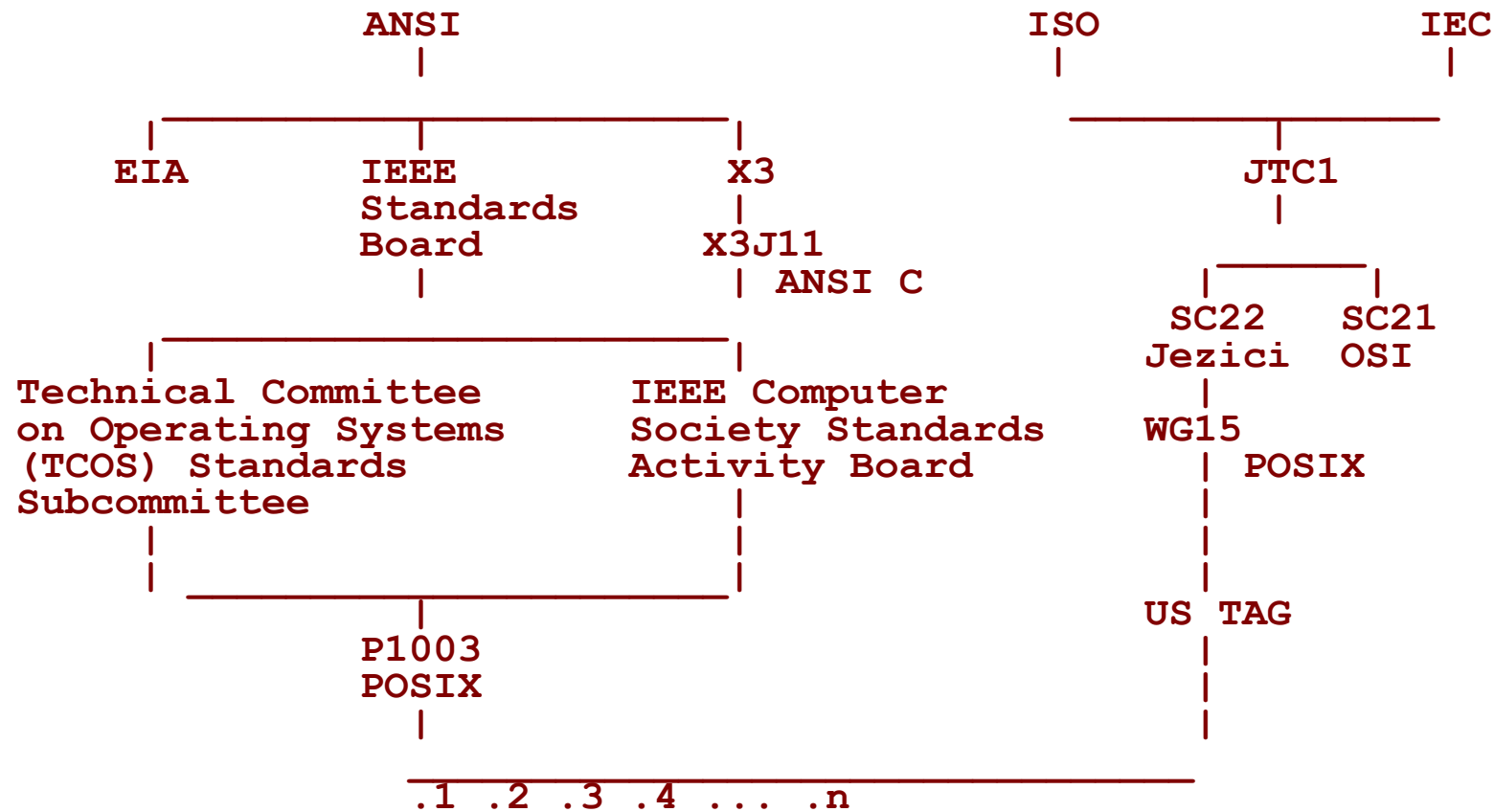
- Odabir norme uvjetuje
 - Mogućnost izbora krajnjeg proizvoda na tržištu
 - Rizik (kratkoročan i dugoročan)

Tip norme	Mogućnost izbora	Rizik
pravna (de iure)	velika	mali
proizvod	mala	mali-velik
licenca	srednja	mali-srednji
proizvod/udruga	srednja	srednji
udruga	mala	vrlo velik

Formalna tijela

- ISO – International Standards Organization
- ITU-T – International Telecommunication Union
Telecommunication Standardization Sector
- W3C – World Wide Web Consortium
- IETF – Internet Engineering Task Force
- IEEE – Institute of Electrical and Electronic Engineers
- IEC – International Electrotechnical Commission
- ANSI – American National Standards Institute

Primjer: Odnosi u društvima za norme



Što je *normirano*? *Otvoreno*?

- Sklopovlje
 - Procesori (x86, ARM)
 - Memoriije (DDR_x, flash)
 - Komunikacijsko sklopovlje (RS-232, Ethernet, ATM)
- Programski jezici
 - ISO C, C99, ISO/IEC C++ 98, Java, Fortran 95, ...
- Drugi jezici
 - Talijanski, španjolski, njemački, norveški, ...
 - HTML 4.0, CSS 2, CSS 3, ...

Što je *normirano*? *Otvoreno*?



- Mrežni protokoli
 - Ethernet, TCP/IP, HTTP, FTP, IMAP, ...
- Usluge
 - WWW, CORBA, ...
- Alati
 - GDB, Eclipse, Java, .Net, ...
- Oblici zapisa
 - JPEG, PNG, PDF, ODF, ...
- Algoritmi...



Izvor: HZN, brošura 20 godina novog pristupa

Kako do otvorenosti

- Norme neovisne o proizvođaču donose
 - Prenosivost
 - Povezivost
 - Suradnju
 - Zamjenjivost (dobavljača, starih sustava s novim)
 - Zaštitu investicija
 - Smanjenje troškova
- Inovativnost?
 - Potisnuta (u drugim smjerovima), ali ne i ugušena

Primjer: Postupak razvoja IEEE norma

- Dobrovoljno sudjelovanje tehničkih stručnjaka
- Njihovo poznavanju industrije i profesionalni sud,
- Odluke se donose konsenzusom
 - dakle uvjet je teži od natpolovične većine.
- Kod glasovanja potrebno **75%** glasova **ZA**
 - obično 85-90%.
- Technical Committee on Operating Systems (TCOS)
 - odobrava projekt i formiranje radne grupe.
- IEEE Standards Board autorizira početak projekta.
 - tehnički stručnjaci definiraju radnu inačicu norme,
 - to obično traje dvije do tri godine
 - jednotjedni sastanci svaka tri mjeseca na različitim lokacijama, stimulira se interes šire svjetske zajednice.
 - tokom rada na normi, ključan je visoki stupanj usklađenosti i konsenzusa.

Primjer: Postupak razvoja IEEE norma

- Druga grupa stručnjaka određuje se u grupu za glasovanje
- Obično su to svi IEEE članovi te stručnjaci koji iz bilo kojeg razloga nisu mogli sudjelovati u prethodnoj fazi pripreme dokumenta
- Kada se postigne visoki stupanj suglasja (>75%), dokument se upućuje IEEE Standards Board-u
 - za vrijeme procesa IEEE glasovanja uključen je i ISO JTC1 WG15
- Na taj način norma postaje međunarodno priznata

Što je novi pristup?

Novi pristup predstavlja novo oruđe za tehničko usklađivanje. Europsko je vijeće u svibnju 1985. godine odlučilo provesti jasno razgraničenje odgovornosti između Europske unije, EFTA-e i europskih organizacija za normizaciju: otada CEN, CENELEC i ETSI pomažu ostvarenju slobodnoga kretanja roba i usluga na jedinstvenome europskom tržištu.

Direktive Europske unije poznate kao "direktive novoga pristupa" utvrđuju "bitne zahtjeve" (npr. zahtjeve koji se odnose na zaštitu zdravlja, sigurnost i zaštitu okoliša) koje proizvodi moraju zadovoljiti prije stavljanja na europsko tržište. CEN, CENELEC i ETSI omogućuju da se ti zahtjevi zadovolje s pomoću usklađenih norma.

Kako novi pristup pomaže industriji?

Europske su norme dragovoljne. Potreba za europskim normama može se utvrditi u mnogim izvorima. Industrija je glavni predlagatelj europskih norma. Doista, većina se europskih norma izrađuje kako bi se zadovoljile trenutačne trgovinske potrebe europske industrije, a ne kao potpora europskim direktivama.

U slučajevima kad postoji koja direktiva, proizvođači mogu odabrati bilo kakvo tehničko rješenje koje zadovoljava bitne zahtjeve direktive. Ako primjenjuju usklađene europske norme, uživaju korist od "pretpostavke o sukladnosti" s bitnim zahtjevima utvrđenim u toj direktivi.

Međutim, ako odaberu svoju vlastitu metodu, moraju osigurati složenu tehničku dokumentaciju koja katkad mora uključivati izvještaje priznatih ispitnih kuća koje se nazivaju "prijavljenim tijelima" da su u skladu s odgovarajućom direktivom.

Primjena jedne europske norme daje proizvođačima mogućnost pristupa u 28 država te da dopru do 460 milijuna korisnika.

Što se događalo proteklih 20 godina?

U 20 godina novoga pristupa Europsko povjerenstvo, EFTA i europske organizacije za normizaciju uključene su u partnerski odnos za koji ne postoji primjer u drugim dijelovima svijeta. To partnerstvo osigurava jedan od ključnih elemenata nužnih za dovršenje zajedničkoga tržišta. Europske norme pomažu uvođenju novih proizvoda, pomažu uvođenju novina te potiču konkurentnost.





Primjeri norma

Ponašanje



- Noževi su u upotrebi od kamenog doba
- Za lakše rezanje u upotrebi su dva noža
- U srednjem vijeku se umjesto jednog noža pojavljuje vilica (u Europu je vjerojatno došla s Bliskog Istoka)

Skupina

- Norme se razlikuju od skupine do skupine
- Ne postoji ispravna metoda, samo prihvaćena
- Za prihvaćanje norme moraju postojati:
 - Potreba
 - Dogovor

- **Portable Open System Interface**
 - Potreba za usklađivanjem sučelja operacijskih sustava
 - Programskih
 - Korisničkih
 - Mrežnih
 - Odnosi se na grupu standarda IEEE 1003.*n* i dijelove ISO/IEC 9945
 - IEEE 1003.1-1998 \equiv POSIX.1

IEEE 1003.1-2004

fork()

```
#include <unistd.h>
pid_t fork(void);
```

The `fork()` function **shall** create a new process.

The new process (child process) **shall** be an exact copy of the calling process (parent process) except as detailed below:

- The child process **shall** have a unique process ID.
- The child process ID also **shall not** match any active process group ID.
- ...

Internetske norme

RFC – Internet Request For Comments

STD – Internet Standard

FYI – For Your Information

BCP – Best Current Practices

- Opisuju norme Interneta
 - Protokoli
 - Jezici
 - Algoritmi

MUST

This word, or the terms "**REQUIRED**" or "**SHALL**", mean that the definition is an absolute requirement of the specification.

MUST NOT

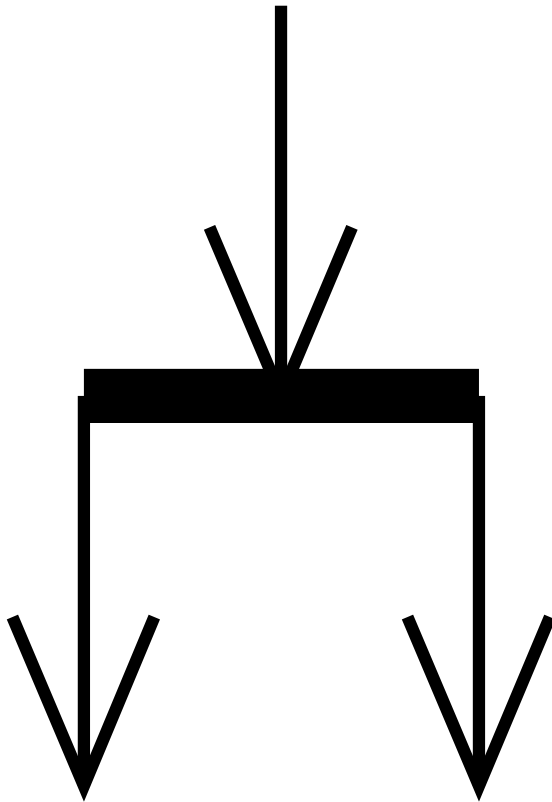
This phrase, or the phrase "**SHALL NOT**", mean that the definition is an absolute prohibition of the specification.

I još jedna...



UML2 – diagram aktivnosti

- Račvanje (fork) označava aktivnosti koje se mogu odvijati paralelno



UML2 - veze



ovisnost (*dependancy*)



proširenje (*extension*)



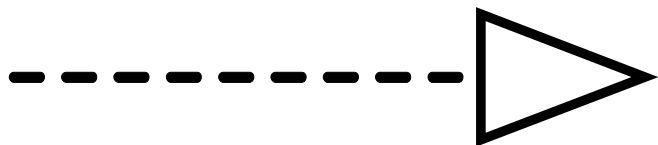
sadržanje (*aggregation*)



sastavljanje (*composition*)



poopćenje (*generalisation*)



realizacija (*realisation*)

Pisanje datuma i vremena

ISO 8601:2004 - *Data elements and interchange formats; Information interchange; Representation of dates and times*

- Zapis datuma, vremena i vremenskih intervala
 - yyyy-mm-dd
 - hh:mm:ss
 - yyyy-mm-ddThh:mm:ss

Matoš se nije rodio 1873-06-12

ISO 8601:2004

- Međunarodna – ISO norma
- Jezično neovisan
- Čitljiv strojno i ljudima
- Jednostavno se sortira i uspoređuje
 - veće jedinice prije manjih
- Dozvoljeno je izostaviti pojedine komponente
 - nije potrebna tolika preciznost
 - potrebno je uštediti na zapisu
 - pravila za interpretaciju

Kako pisati datum i vrijeme?

- Ne postoji hrvatska norma (HRN)
- Postoje upute u *Pravopisnom priručniku* (Jojić Ljiljana, Zagreb : Novi Liber, 2003., pp. 105-107):
“Točka se piše na kraju rednoga broja.”

Matoš se rodio 13. 6. 1873.

Kako pisati vrijeme?

“Točka se piše iza brojeva koji znače količinu vremena u satima, minutama i sekundama.”

“Ako se količina vremena mjeri desetinkom sekunde, između znamenke koja znači sekundu i desetinku sekunde stavlja se zarez.”

“Ako se količina vremena mjeri manjim jedinicama decimalnog sustava, stotinkama i tisućinkama sekunde, opet se stavlja zarez, a desno od njega bilježi potreban broj jedinica.”

Kako pisati vrijeme?

“Na elektroničkim spravama, satovima i sl. pojavljuje se i dvotočka između broja koji znači sat i broja koji znači minutu, odnosno broja koji znači minutu i broja koji znači sekundu. Odatle je moguće da se vrijeme za posebne potrebe, npr. u sportu, obilježava i tako da se između broja koji znači sat i broja koji znači minutu, odnosno između broja koji znači minutu i broja koji znači sekundu umjesto točke stavlja dvotočka. Tada se između sekunde i dijela sekunde umjesto zareza može stavljati točka.”

Kako pisati vrijeme?

- Prema pravopisu:

hh.mm.ss,ddd

- Za posebne potrebe:

hh:mm:ss.ddd



Norme W3C

- World Wide Web Consortium
- međunarodno neovisno tijelo
 - interoperabilnost Web
 - kompatibilnost Web tehnologija, korištenje bilo kakvog hardvera i softvera za rad na Webu, primjenom normi



W3C mission: "To lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web"

Kako do norma?



- Spor proces nastanka
- W3C često preuzima i standardizira *de facto* standarde (npr. XMLHttpRequest)
- Etape:
 - **Working Draft** (...in development, Last Call..)
 - **Candidate Recommendation**
 - Proposed Edited Recommendation
 - Proposed Recommendation
 - **Recommendation**



Primjeri norma

- **HTTP**
- **HTML**
- **CSS**
- **DOM**
- **PNG, SVG**
- **SMIL**
- **MathML**
- **SOAP/XMLP**
- **Xforms**
- **XHTML**
- **XML**
- **XML Schema**
- **Xpath**
- **Xpointer**
- **XSL, XSLT**





Norme kodnih stranica

Prikaz znakova u računalu

- Računala razumiju samo 0 i 1
- Znakovi (grafemi) – skupovi bitova
- Koji skup bitova označava koje slovo?
- Kako će sva računala i svi programi znati ispravno protumačiti bitove?
 - kodne stranice (*code page*)
- Dvije vrste znakova:
 - ispisivi
 - kontrolni (neispisivi)

ASCII

- American Standard Code for Information Interchange (1963)
- ANSI standard
- **7-bitni** zapis $\rightarrow 2^7 = \mathbf{128}$ različnih znakov
- Najviši, 8. bit: paritet ili "0"
- 95 ispisivih znakov
- 33 kontrolna znaka

Tablica ASCII



Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex	Char	Dec	Oct	Hex
(nul)	0	0000	0x00	(sp)	32	0040	0x20	@	64	0100	0x40	`	96	0140	0x60
(soh)	1	0001	0x01	!	33	0041	0x21	A	65	0101	0x41	a	97	0141	0x61
(stx)	2	0002	0x02	"	34	0042	0x22	B	66	0102	0x42	b	98	0142	0x62
(etx)	3	0003	0x03	#	35	0043	0x23	C	67	0103	0x43	c	99	0143	0x63
(eot)	4	0004	0x04	\$	36	0044	0x24	D	68	0104	0x44	d	100	0144	0x64
(enq)	5	0005	0x05	%	37	0045	0x25	E	69	0105	0x45	e	101	0145	0x65
(ack)	6	0006	0x06	&	38	0046	0x26	F	70	0106	0x46	f	102	0146	0x66
(bel)	7	0007	0x07	'	39	0047	0x27	G	71	0107	0x47	g	103	0147	0x67
(bs)	8	0010	0x08	(40	0050	0x28	H	72	0110	0x48	h	104	0150	0x68
(ht)	9	0011	0x09)	41	0051	0x29	I	73	0111	0x49	i	105	0151	0x69
(nl)	10	0012	0x0a	*	42	0052	0x2a	J	74	0112	0x4a	j	106	0152	0x6a
(vt)	11	0013	0x0b	+	43	0053	0x2b	K	75	0113	0x4b	k	107	0153	0x6b
(np)	12	0014	0x0c	,	44	0054	0x2c	L	76	0114	0x4c	l	108	0154	0x6c
(cr)	13	0015	0x0d	-	45	0055	0x2d	M	77	0115	0x4d	m	109	0155	0x6d
(so)	14	0016	0x0e	.	46	0056	0x2e	N	78	0116	0x4e	n	110	0156	0x6e
(si)	15	0017	0x0f	/	47	0057	0x2f	O	79	0117	0x4f	o	111	0157	0x6f
(dle)	16	0020	0x10	0	48	0060	0x30	P	80	0120	0x50	p	112	0160	0x70
(dc1)	17	0021	0x11	1	49	0061	0x31	Q	81	0121	0x51	q	113	0161	0x71
(dc2)	18	0022	0x12	2	50	0062	0x32	R	82	0122	0x52	r	114	0162	0x72
(dc3)	19	0023	0x13	3	51	0063	0x33	S	83	0123	0x53	s	115	0163	0x73
(dc4)	20	0024	0x14	4	52	0064	0x34	T	84	0124	0x54	t	116	0164	0x74
(nak)	21	0025	0x15	5	53	0065	0x35	U	85	0125	0x55	u	117	0165	0x75
(syn)	22	0026	0x16	6	54	0066	0x36	V	86	0126	0x56	v	118	0166	0x76
(etb)	23	0027	0x17	7	55	0067	0x37	W	87	0127	0x57	w	119	0167	0x77
(can)	24	0030	0x18	8	56	0070	0x38	X	88	0130	0x58	x	120	0170	0x78
(em)	25	0031	0x19	9	57	0071	0x39	Y	89	0131	0x59	y	121	0171	0x79
(sub)	26	0032	0x1a	:	58	0072	0x3a	Z	90	0132	0x5a	z	122	0172	0x7a
(esc)	27	0033	0x1b	;	59	0073	0x3b	[91	0133	0x5b	{	123	0173	0x7b
(fs)	28	0034	0x1c	<	60	0074	0x3c	\	92	0134	0x5c		124	0174	0x7c
(gs)	29	0035	0x1d	=	61	0075	0x3d]	93	0135	0x5d	}	125	0175	0x7d
(rs)	30	0036	0x1e	>	62	0076	0x3e	^	94	0136	0x5e	~	126	0176	0x7e
(us)	31	0037	0x1f	?	63	0077	0x3f	_	95	0137	0x5f	(del)	127	0177	0x7f

Kontrolni znakovi

- Primjeri:
 - prelazak u novi red, povratak na početak reda, tabulator, zvono, *backspace*, *escape*
- Problem: **višeznačnost** na različitim platformama :-(
- Primjer:
 - prelazak u "novi red"
 - prisjetite se - programski jezik C: `"\r\n"`

Prelazak u "novi red"

- **CR** – *Carriage Return* – pomicanje na početak reda
- **LF** – *Line Feed* – spuštanje za jedan redak

CR Commodore, Mac OS (do v.9)

LF Unix i slični sustavi

CR+LF MS-DOS, Windows

- Uporaba CR+LF za Internet protokole, ponekad se tolerira samo LF

Ispisivi znakovi

- Znamenke, slova, znakovi
- 7 znakova za akcente (mogu se kombinirati sa slovima, ovisno o programskoj podršci)
- Brojevi: 0011 + BCD vrijednost
 - prisjetite se: Binary Coded Decimal
- Slova:
 - abecedni poredak
 - razlika između velikog i malog u jednom bitu
 - jednostavno sortiranje, pretvorbe

ASCII problemi

- Premalo znakova (127)
- Potrebni dodatni znakovi za latinicu
- Potrebna dodatna pisma
 - ćirilica, glagoljica, grčko pismo...
- Rješenje: uvođenje različitih proširenja za različite jezike

IBM PC

- Proširenje ASCII-ja
- IBM-ov zapis
- 8-bitni zapis
- Korišćenje u MS-DOS-u
- verzije po regijama/pismima
- **CP 850** – Latin I - Western European
- **CP 852** – Latin II - Eastern European

MS Windows ANSI

- Temeljen na ANSI prijedlozima, nikad standardiziran!
- **8-bitni** zapis
- **1252** – West European Latin
- **1250** – East European Latin
- Pitanje za programere: **podržati ili ne?**
 - Ne -> nije moguće raditi s ovakvim dokumentima
 - Ne -> pogreške pri radu programa
 - Da -> postaje *de-facto* standard

ISO/IEC 8859



- Alias ISO 8859
- Nadogradnja ASCII-ja (kompatibilnost!)
- **8-bitni** zapis -> **256** znakova
- Podijeljen u numerirane dijelove:
 - ISO 8859-1 ... ISO 8859-16
 - svaki dio prilagođen određenom pismu ili regiji

Neke ISO 8859 kodne stranice



Oznaka	Naziv	Opis
ISO 8859-1	Latin-1 Western European	većina zapadnoeuropskih zemalja
ISO 8859-2	Latin-2 Eastern European	srednjeistočna Europa koja koristi latinicu (Hrvatska!)
ISO 8859-5	Latin/Cyrillic	slavenski jezici koji koriste ćirilicu
ISO 8859-15	Latin-9	nadogradnja ISO 8859-1 (dodani znakovi €, Œ, Ÿ ...) potpuni francuski, finski, estonski
ISO 8859-16	Latin-10 South-Eastern European	Srednjeistočna Europa (Hrvatska!) uz finski, njemački, francuski... € znak

Zajednički problemi



- Kako pisati jedan dokument koji u sebi sadrži više jezika / pisama?
- Kako pisati strana imena u e-mail poruci raspodijeljenom timu?
- Kako podržati azijske jezike, koji sadrže i po nekoliko tisuća različitih znakova?



Unicode

- Unicode **nije** kodna stranica!
- **ujedinjavanje regionalnih standarda** u jedan
- Svaki znak – jedna numerička vrijednost (*code point*)
- Oznaka U+numerička vrijednost
- Potencijalno **~1 milijun** znakova
- Trenutno zauzeto: manje od 10% prostora :-)



Unicode - svojstva

- Mapiranje prvih 256 znakova na **numeričke oznake** identično s **ISO 8859-1**
- Znakovi podijeljeni u "ravnine"
- Gotovo svi često korišteni znakovi su u prvih 64K numeričkih oznaka
 - Basic Multilingual Plane - BMP
- Višestruko pojavljivanje nekih znakova
 - lakša konverzija
- Ocrtava standard ISO/IEC 10646

Unicode - kodiranja

- Kako zapisati Unicode numeričke vrijednosti svakog znaka (*code point*)?
- 3 vrste:
 - UTF-8
 - UTF-16
 - UTF-32

Unicode: UTF-8

- **varijabilna** dužina (**1, 2, 3 ili 4** bajta)
- prvih 127: zapis identičan ASCII-ju - 1 B
- ostali često korišteni znakovi (HR) - 2 B
- ostali znakovi iz BMP - 3 B
- znakovi iz ostalih ravnina - 4 B
- **najviše raširen**
- korišćenje:
 - XML, e-mail, Web stranice
 - Unix / Linux

Unicode: UTF-16

- **varijabilna** dužina (**2 ili 4** bajta)
- gotovo svi često korišteni znakovi - 2 B
- ostatak - 4 B
- problem: kojim se redom šalju bajtovi?
 - **little / big endian** -> UTF-16LE, UTF-16BE
 - UTF-16 (BOM – Byte Order Mark - na početku)
- korišćenje
 - interna reprezentacija znakova
 - Windows NT/2000/XP/CE
 - Java i .NET programska okruženja

Unicode: UTF-32

- **fiksna** dužina (**4** bajta)
- trenutno je vrlo rijetko pojavljivanje znakova za koje su doista potrebna 4 bajta
- rijetko korišten

Usporedba?

- veličina datoteke
- varijabilne granice znakova
- korišćenje postojećih API-ja
- jednostavnost konverzije
- kompatibilnost s ASCII

Primjer: Prikaz znakova

- Kolika je dužina ovog dokumenta u oktetima?

Ovisi ;) <ž/>

ASCII: ne ide!

ISO-8859-1 ne ide!

ISO-8859-2: 4 okteta 3C BE 2F 3E

UTF-8 (BOM): 8 okteta EF BB BF 3C C5 BE 2F 3E

UTF-16: 10 okteta FF FE 3C 00 7E 01 2F 00 3E 00

- Koji je kod za znak ž u Unicode? 0x017E (U-017E)

Prepoznavanje kodne stranice

- oznaka na početku datoteke
- ručni odabir u programu
- Web, e-mail: oznaka u zaglavlju

A gdje smo mi? Preporuke!



Prikazano u \ Napisano u	CP852	windows-1250	iso-8859-2	utf-8
CP852	čćšđž ČĆŠĐŽ	ž†çĐ§ ¬ŽćŃ!	☞☞çĐ§ Ž☞ćŃŚ	☞☞☞Ų ☞☞☞A
windows-1250	ŘŠÜ× ĽĂŎďĂ	čćšđž ČĆŠĐŽ	čć☞đ☞ ČĆ☞Đ☞	☞☞☞ ☞DŸ
iso-8859-2	ŘŠ= ž ĽĂęď«	čćąđł' ČĆ©Đ®	čćšđž ČĆŠĐŽ	☞☞☞ ☞ΣЮ
utf-8	–Ž–ç– –ı–Ł–ž–ı–ć–á–É–ž	ĀĀĀŁĹĀĹĹ ĀĀĀŁĹ ĀĀĹ	ĀĀĀŁĹĀĀĀŁĹ ĀĀĀŁĹ ĀĀĹ	čćšđž ČĆŠĐŽ

- Windows kodne stranice nisu standard :(
- Mnogi još uvijek koriste ISO 8859-2
 - u redu, ali postoje mnogi problemi s interoperabilnošću
- UTF-8
 - najbolji dugoročni izbor

Ima li toga još?



- Ima :-)
- UTF-9 i UTF-18
 - **April Fool's** Day RFC dokument
 - tehnički izvediv :-)
- postoji još mnogo (nama manje važnih) kodnih stranica





Pitanja?