

UDK 681.324.04

B. Jerman-Blažič
Institut Jožef Stefan
Jamova 39, Ljubljana

Ključne besede

Referenčni model OSI, nabor grafičnih znakov, nivo predstavitve, implementacija standardov, ISO, CCITT, telematske storitve, odprti sistemi, heterogene računalniške mreže, računalniški sistemi.

Key words

OSI Reference Model, Graphic character sets, presentation level, standard implementation, ISO, CCITT, Telematic services, open systems, heterogeneous computer networks, computer systems.

"Choice of the Coded Character Set for Text Communication in the Presentation Level of RM OSI"

Abstract

The ISO reference model for open system interconnection is a theoretical framework for development of heterogeneous computer networks with implemented telematic services. The sixth level of the model is the presentation level, responsible for data presentation. The paper deals with the problems which arised after the adoption of two international 8bit coded character sets for text communication within the international bodies for standardization. Some guidelines for implementation of the standards in connection with the application field are presented too.

Povzetek

Referenčni model OSI daje teoretično in praktično podlago za izgradnjo heterogenih mrež, v katere se vključujejo najrazličnejši računalniški sistemi in nudijo različne telematske storitve. Sesti nivo modela OSI opredeljuje predstavitev podatkov. V zadnjih dveh letih sta bila za ta nivo izdelana in sprejeta dva mednarodna standarda za nabor grafičnih znakov kodiranih z 8 biti. Problematika, ki je nastala v zvezi z implementacijo teh standardov v okviru RM OSI še ni razrešena. Prispevek obravnava nastale probleme in ponuja nekatere rešitve v odvisnosti od področja implementacije.

0. Uvod

Problem kodiranja nabora znakov sega v začetek razvoja računalništva in njemu sorodnih tehnik. Kodiranje in kodne strukture so bile pred pojavom računalniške tehnike predmet telegrafije. Z razvojem računalniške tehnike je nastala potreba za kodiranje grafičnih in krmilnih znakov. Tako so bile postopoma uvajane 6, 7 in 8-bitne kode v odvisnosti od stopnje uvedene računalniške tehnologije. Razvoj kodnih struktur je danes skupni problem računalništva, telekomunikacij in njihovega otroka - telematike. Pri tem imamo v mislih nove telematske storitve (teletex, videotex, telekonferenco ipd.) in široko področje različnih računalniških mrež (LAN, VAN, PDPT, PBX ipd.) ter distribuirane sisteme za obdelavo podatkov.

Razvoj kodnih struktur je treba opazovati iz dveh zornih kotov: kot kodne strukture, ki se uporabljajo znotraj zaprtih računalniških sistemov in kodne strukture, ki se uporabljajo za prenos in izmenjavo podatkov. Sedanji trendi razvoja distribuiranih sistemov, pri katerih ima ključni pomen prenos in izmenjava podatkov, se gibljejo v smeri izgradnje v skladu z referenčnim modelom OSI mednarodne organizacije za standardizacijo.

Model OSI daje teoretično in praktično podlago za izgradnjo heterogenih mrež, v katere se vključujejo najrazličnejši računalniški sistemi in nudijo različne telematske storitve. Sesti nivo modela OSI opredeljuje predstavitev podat-

kov. Ta nivo predpostavlja, da je prenos podatkov opravljen brez napake. V nadaljevanju prispavka bo obravnavana problematika izbora kodiranja (dodeljevanja odgovarajoče kombinacije bitov grafičnim in krmilnim znakom iz nekega vnaprej dogovorjenega repertoarja) in dekodiranja (predstavitve sprejetih znakov) v 6. nivoju modela OSI.

1. Razvoj 8-bitnih kod za potrebe prenosa in izmenjave podatkov

Izbor in definicija kodnih struktur za potrebe prenosa in izmenjave podatkov je predmet dela SC2 (podkomiteja 2) Mednarodne organizacije za standardizacijo in VIII študijske skupine CCITT-ja. Razvoj prvih standardov za 8-bitne kode sega v konec osemdesetih let. Prva standardizirana 8-bitna kodna tabela za potrebe prenosa podatkov je bila izdelana konec osemdesetih let. Standard je dobil naslov "Nabor grafičnih znakov kodiranih z 8 biti za potrebe izmenjave podatkov ISO 6937". Standard je nastal zaradi potreb CCITT po standardiziranem naboru grafičnih znakov za latinično pisavo evropskih jezikov, ki se je uporabljal pri prenosu in prezentaciji podatkov v okviru teletexa in videotexa. CCITT je objavil predlog standarda v obliki priporočila S.61-1981 oz. T.61-1984 (za potrebe storitve teletexa) in S.101 in T.101 - 1984 (za potrebe videotexa).

Kodna tabela ISO 6937, ki je v osnovi identična s T.61 in T.101, vsebuje dva nabora grafičnih znakov. Na levi strani kodne tabele se nahaja

mednarodna referenčna verzija nabora grafičnih znakov, bolj znana po svetu kot ISO 646, oziroma 7-bitni ASCII, na desni strani kodne tabele so diakritični znaki, s katerimi je možno generiranje akcentiranih latiničnih znakov. Diakritični znaki so znaki brez pomika naprej. Vsaka akcentirana črka latinice se generira kot dvozlgovni znak, ki na terminalni napravi zavzame eno mesto. Kodno tabelo ISO 6937 smo ponazorili na sliki 1. Tehnika, definirana v T.61 in T.101, omogoča predstavitev repertoarja grafičnih znakov, ki je po številu veliko večji od nabora znakov iz tabele 1. ISO 6937 omogoča predstavitev 350 grafičnih znakov. S tem standardom je CCITTa rešil problem prenosa grafičnih znakov za potrebe zahodnoevropskih in latinskoameriških držav. Veliki proizvajalci računalniške opreme, zlasti IBM, so po objavi standarda ISO 6937 sklenili, da potrebujejo enozlogovno 8-bitno ASCII kodo. Na spodbudo IBM-a je ISO TC97/SC2 razvil ISO 8859, v katerem je vsak znak kodiran z enim zlogom. Osnovna kodna tabela tega osemdelnega standarda je prikazana na sliki 2. Kodna tabela, prikazana na sliki 2, skupaj z definiranim naborom krmilnih znakov C0 in C1, je bila sprejeta tudi v ANSI kot 8-bitni ASCII (1.1985) in je znana v svetu pod imenom Latinica 1. Enako kodno tabelo z dodatnimi nabori krmilnih znakov je sprejela tudi ECMA kot standard ECMA-94. Tabela na sliki 2 nam kaže prvi del ISO 8859, ki je namenjen govornemu področju 44 dežel (za angleščino, španščino, nemščino, francoščino, portugalsščino, flamsščino in italijanščino).

00000000		00000001		00000010		00000011		00000100		00000101		00000110		00000111		00001000		00001001		00001010		00001011		00001100		00001101		00001110		00001111	
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
00000000	00000001	00000010	00000011	00000100	00000101	00000110	00000111	00001000	00001001	00001010	00001011	00001100	00001101	00001110	00001111	00010000	00010001	00010010	00010011	00010100	00010101	00010110	00010111	00011000	00011001	00011010	00011011	00011100	00011101	00011110	00011111
00000000	00000001	00000010	00000011	00000100	00000101	00000110	00000111	00001000	00001001	00001010	00001011	00001100	00001101	00001110	00001111	00010000	00010001	00010010	00010011	00010100	00010101	00010110	00010111	00011000	00011001	00011010	00011011	00011100	00011101	00011110	00011111

Sl. 1. ISO 6937/2

Ostali deli ISO 8859 z izjemo 8859/2 so v fazi DIS (draft international standard) ali DP (draft proposal). 8859/2 definira latinico št.2, ki je namenjena prenosu teksta, napisanega v enem od naslednjih jezikov: albanskem, češkoslovaškem, nemškem, madžarskem, poljskem, romunskem in slovenskem (oziroma srbohrvaškem). Enako kodno tabelo, kot je definirana v ISO 8859/2, je sprejela tudi Komisija za kode in difirne sisteme pri Zveznem zavodu za standardizacijo kot JUS 1.81.013. 8859/3 pokriva govorno področje Francije, Nemčije, Italije, Malte, Južne Afrike in Turčije. 8859/4 pa je namenjena Danski, Finski, Nemčiji, Grenlandiji, Norveški, Švedski in sovjetskim republikam Estoniji, Latviji in Litvaniji. Tem kodnim tabelam je skupno to, da ima v primerih, ko se ista črka pojavlja v dveh različnih kodnih tabelah, zmeraj isto mesto, kar pomeni tudi isto kodno kombinacijo (tako se črka ž pojavlja v 8859/2 in 8859/4 na istem mestu in ima kodno kombinacijo 10/14).

Leva stran kodne tabele je povsod enaka t.j. ISO 646 ali 7-bitni ASCII. 8859/5 definira kodno tabelo za nabor znakov, ki se uporabljajo v tehničnih in matematičnih znanostih (razni specialni znaki za pisanje matematičnih in tehničnih izrazov oz. formul). 8859/6 definira kodno tabelo za nabor znakov, ki se uporabljajo

00000000		00000001		00000010		00000011		00000100		00000101		00000110		00000111		00001000		00001001		00001010		00001011		00001100		00001101		00001110		00001111	
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
00000000	00000001	00000010	00000011	00000100	00000101	00000110	00000111	00001000	00001001	00001010	00001011	00001100	00001101	00001110	00001111	00010000	00010001	00010010	00010011	00010100	00010101	00010110	00010111	00011000	00011001	00011010	00011011	00011100	00011101	00011110	00011111
00000000	00000001	00000010	00000011	00000100	00000101	00000110	00000111	00001000	00001001	00001010	00001011	00001100	00001101	00001110	00001111	00010000	00010001	00010010	00010011	00010100	00010101	00010110	00010111	00011000	00011001	00011010	00011011	00011100	00011101	00011110	00011111

Sl. 2. ISO 8859/1

v založništvu za formatiranje tekstov, 8859/7 definira kode za grško abecedo in 8859/8 definira cirilico za ruski, srbski, makedonski, bolgarski, ukrajinski in beloruski jezik. Znaki iz kodnih tabel 8859/5/6/7/8 ne nastopajo v naboru znakov v ISO 6937/1/2 in zato so v pripravi dodatki k ISO 6937, ki naj bi vsebovali manjkajoče znake.

2. Izbor kodne strukture

Po objavi ISO 6937/1/2 in ISO 8859/1/2 je v okviru uradnih mednarodnih teles za standardizacijo, kot je CCITT, ISO in zlasti ANSI prišlo do velikega razburjenja. Implementatorji iz krogov proizvajalcev računalniške in telematske opreme so se znašli pred uganko. Katero 8-bitno kodo implementirati v nivoju 6, za posamezno

00000000		00000001		00000010		00000011		00000100		00000101		00000110		00000111		00001000		00001001		00001010		00001011		00001100		00001101		00001110		00001111	
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
00000000	00000001	00000010	00000011	00000100	00000101	00000110	00000111	00001000	00001001	00001010	00001011	00001100	00001101	00001110	00001111	00010000	00010001	00010010	00010011	00010100	00010101	00010110	00010111	00011000	00011001	00011010	00011011	00011100	00011101	00011110	00011111
00000000	00000001	00000010	00000011	00000100	00000101	00000110	00000111	00001000	00001001	00001010	00001011	00001100	00001101	00001110	00001111	00010000	00010001	00010010	00010011	00010100	00010101	00010110	00010111	00011000	00011001	00011010	00011011	00011100	00011101	00011110	00011111

Sl. 2. ISO 8859/2

aplikacijo znano je, da učinkovita komunikacija in izmenjava podatkov v odprtih sistemih povezovanja ni mogoča brez enotnih in standardiziranih kod. Odstavki, ki sledijo obravnavajo problem

izbora kodne strukture in stališča bodočih uporabnikov oziroma proizvajalcev. Odločitve, ki bodo sprejete v svetu po tem vprašanju, bodo vplivale tudi na razvoj telematskih storitev in izdelkov v Jugoslaviji (kot so na primer terminali za teletex, videotex, večjezične tipkovnice terminalov za vnos podatkov ter različne aplikacije v okviru JUPAKa).

2.1 Nekatera znana dejstva

ISO 6937 je aplikacijsko orientiran standard, ki poleg grafičnih znakov za pisanje teksta tega definira tudi krmilne funkcije. ISO 6937 je izdelan za potrebe nivoja predstavitev pri prenosu teksta za aplikacije kot so Teletex in Videotex. Za te storitve so v uporabi posebej razviti terminali, ki omogočajo sprejem in predstavitev sprejetih podatkov. Sprejeti podatki se prikazujejo na zaslonu ali pa na tiskalniku. Posebna izbira in shranjevanje teh podatkov zaradi nadaljnje obdelave ni predvidena. V postopku izdelave je četrti del standarda, ki definira krmilne funkcije za preformatiranje teksta (tistega, ki je že formatiran in tistega, ki ni).

Večina proizvajalcev in uporabnikov je pričakovala, da bo ISO 6837 standard z najbolj splošno uporabo. Ta standard je zadovoljeval potrebe

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0000	00	sp	0	ā	P	'	p	A	ā	ā	ā	ā	ā
0001	01	!	1	A	q	a	q	H	h	ā	ā	ā	ā
0010	02	"	2	B	R	b	r	^	^	ā	ā	ā	ā
0011	03	#	3	C	S	c	s	£	£	ā	ā	ā	ā
0100	04	\$	4	O	T	d	t	z	z	ā	ā	ā	ā
0101	05	%	5	E	U	e	u	μ	μ	ā	ā	ā	ā
0110	06	&	6	F	V	f	v	ħ	ħ	ā	ā	ā	ā
0111	07	'	7	G	W	g	w	š	š	ā	ā	ā	ā
1000	08	(8	H	X	h	x	-	-	ā	ā	ā	ā
1001	09)	9	I	Y	i	y	ı	ı	ā	ā	ā	ā
1010	10	*	*	J	Z	j	z	š	š	ā	ā	ā	ā
1011	11	+	+	K	L	k	l	ğ	ğ	ā	ā	ā	ā
1100	12	,	,	L	L	l	l	ı	ı	ā	ā	ā	ā
1101	13	-	-	M	J	m	j	ı	ı	ā	ā	ā	ā
1110	14	.	.	>	N	n	n	ı	ı	ā	ā	ā	ā
1111	15	/	/	?	O	-	-	ı	ı	ā	ā	ā	ā

Sl. 2. ISO 8859/3

zahodnega dela sveta, ki je istočasno tudi najbolj razviti del sveta. V tem delu sveta se nahaja 90 najmočnejših proizvajalcev računalniške in komunikacijske opreme ter največje število uporabnikov sodobnih telematskih storitev. Kmalu po sprejemu ISO 6937 se je izkazalo, da rešitev, ki jo določa standard, ni sprejemljiva za sisteme za obdelavo podatkov, ker se po tem standardu nekateri od grafičnih znakov prenašajo kot dvoznakovne informacije, nekateri pa kot enoznakovne. Takšen način prenosa je nesprejemljiv pri prenosu podatkov za nadaljnjo obdelavo, kot je na primer prenos datotek (file transfer). Razprava v zvezi z ISO 6937 se je začela po opozorilu ISO TC97/SC2/WG4, da je tehnika uporabe grafičnih znakov z različno dolžino kode (eno in dvoznakovno) za večino prevajalnikov programskih jezikov nesprejemljiva. Večina obstoječe programske opreme je narejena po principu en znak en zlog.

Po sprejetju teh pripomb se je CCITT javno ogradil od splošne uporabnosti ISO 6937 in izjavil, da je bil ISO 6937 razvit in namenjen le za prenos in ne za obdelavo tekstovnih podatkov, torej le v sklopu storitev Teletexa in

Videotexa. V teh telematskih storitev je tekst zapisan v strogo formatirani obliki (strani), se v taki obliki prenaša in ni namenjen nadaljni obdelavi. Torej se tekst prenaša v končni obliki po omrežju od računalnika (host) do terminala ali od terminala do terminala. Uporabnik teksta ne vidi ali pa ga izpiše. Sodobni sistemi za avtomatizacijo pisarniškega poslovanja pa zahtevajo uporabo telematskih storitev, ki omogočajo tudi obdelavo in predelavo sprejete oziroma odposlane teksta, kot je preformatiranje, dodajanje in arhiviranje teksta. Mogoče je ravno ta omejenost Videotexa ter izjemno zastarele metode preiskovanja informacij (drevesne strukture, ki so prepočasne), botrovala slabi uveljavitvi te usluge na tleh ZDA. Primerjava obeh standardov nas pripelje do sledečih ugotovitev:

2.2 Prednosti in pomanjkljivosti ISO 6937 in ISO 8859

Gledano iz zornega kota razvitih držav zadostuje za latinično pisavo le nabor grafičnih znakov iz ISO 6937/1/2. To pomeni, da v 8-bitnem okolju, če imamo opravka s čistim tekstom (isto velja tudi za slovensko latinico), ni potrebno preklapljanje iz ene kodne tabele v drugo s pomočjo tehnike za razširitev 8-bitnega okolja (za referenco glej ISO 2022 ali ISO 4873 ali I.B1.010 - tehnike razširitve v 7 in 8-bitnem okolju). 8. bit v kodni tabeli nam pokaže, kateri del kodne tabele je v uporabi, levi (8. bit je ničla) ali desni (8. bit je enica). ISO 8859 za prikaz vseh latiničnih pisav sveta potrebuje 4 kodne tabele. V primerih, ko potrebujemo črke vseh 4 jezikovnih področij, je potrebno preklapljanje (switching). Ta trditev drži le do neke meje, ker nekatere od kodnih tabel ISO 8859/1-4 vsebujejo znake za več kot eno področje oziroma so redundancne. Ta prednost ISO 6937 se takoj spremeni v pomanjkljivost, če bi zahtevali znake iz abeced, ki niso latinične. Tak primer je Jugoslavija, kjer bi potrebovali istočasno latinično in cirilno pisavo. Podobno velja za Grčijo, kjer uporabljajo dve pisavi, latinično in grško. Namen ISO standardov je, da zadosti potrebam večjemu številu dežel in ne le deželam, kjer se uporablja izključno latinična pisava. Torej dolgoročno gledano ta prednost ISO 6937 ne drži.

Druga prednost ISO 6937 je, da je to standard, ki se že nekaj časa uporablja (sprejet je bil l. 1980), za razliko od ISO 8859/3-8, ki je še v fazi DIS. Ta prednost se bo zgubila v kratkem, ko bo ISO 8859/3-8 postal veljaven standard.

Izbor kodnih tabel v ISO 8859 je narejen po regijah, v katerih se nahajajo sorodne države. Leva kodna tabela je ASCII, kar ne velja za ISO 6937. Večina dežel sveta je svoje nacionalne kodne tabele razvila na podlagi ASCII kodne tabele (ISO 646) z nacionalnimi razširitvami, specifičiranimi v ISO 646, mednarodna verzija. Tako je ISO 8859 kompatibilen z veliko večino nacionalnih standardov.

ISO 8859 predstavlja lažji prehod iz obstoječih 7bitnih kod v 8bitno kodo. ISO 8859 predstavlja tudi lažji prehod iz 8-bitnega okolja v 16 bitno okolje (dvoznakovno kodiranje), vsak znak v 16-bitni kodni tabeli bo potreboval za predstavitev le eno 8-bitno kodno kombinacijo. ISO SC2/WG6 trenutno dela na razvoju standardizirane 16-bitne kodne tabele. Zaradi široke uporabe ASCII kodne tabele v veliki večini dežel lahko rečemo, da je ISO 8859 kompatibilen z obstoječo programsko opremo in obstoječimi podatkovnimi bazami.

ISO 8859 je bil že v fazi nastajanja sprejet v večini dežel kot nacionalni standard za nabor znakov, kodiranih z 8bitno kodo, tak primer je

tudi Jugoslavija. ISO 6937 ni bil sprejet kot nacionalni standard nikjer, z izjemo ZR Nemčije, ki je kljub temu sprejela še ISO 8859. ISO 8859 omogoča lažje povezovanje (interworking) izdelkov z ASCII in EBCDIC kodo. Izmenjava znakov iz ISO 8859 v okolju z ISO 6937 je možna z uporabo IGS-a (Identity Graphic Sub-repertoire).

Uporaba obeh kodnih sistemov ISO 6937 in ISO 8859 v okviru enega računalniškega sistema je komplicirana in neučinkovita. Uporaba ISO 6937 bi morala biti omejena na telematske storitve, konverzija kode iz ISO 6937 v ISO 8859 bi bila potrebna le pri vходу podatkov v računalnik. Taka strategija bi bila zlasti pomembna za sledeče kategorije uporabnikov in proizvajalcev: a) za implementatorje in razvijalce izdelkov kot so: terminali, tiskalniki in programski paketi, b) za nacionalne odbore in organizacije za standardizacijo za področje kodiranja, avtomatizacije pisarniške opreme in komunikacij ter c) za vse ostale odbore ISO in CCITT, ki se ukvarjajo s standardizacijo medijev in komunikacij.

V primeru, da bo to sprejeto in da se zahteva izločitev ISO 6937 in pospešena uporaba ISO

Pri tiskalnikih z marjeticami, kjer je število znakov omejeno, ima prednost ISO 8859. Tiskalniki z dvema marjeticama imajo možnost uporabe vseh črk ISO 8859, ena marjetica (190 znakov) bi pokrivala ASCII in druga obe desni strani kodnih tabel dveh delov ISO 8859. Pri tiskalnikih, ki uporabljajo verige, če je veriga relativno velika (na primer od 350 znakov) ni problem implementirati obeh standardov, seveda s čim manjšim številom dvojnih črk. Pri teh tiskalnikih ni korespondence med kodo in prikazovanjem določenega znaka, medtem ko je v sami napravi poskrbljeno za ustrezno korespondenco. Pri teh napravah bi se pojavil le problem počasnosti zaradi velikega števila znakov v verigi.

Gledano s stališča določenega geo območja je bolj varčno in gospodarno implementirati ISO 8859. Če želimo implementirati vse kodne tabele ISO 8859, potem ta prednost glede varčnosti odpade. Enako velja, če hkrati implementiramo latinično in cirilično pisavo, ker v tem primeru spet potrebujemo dve kodni tabeli in uporabo tehnike razširitve v 8-bitnem okolju. Če uporabljamo tehnologije z izmenljivimi tiskalniki elementi, potem ima prednost ISO 8859. Vsak tiskalni element ustreza določeni kodni tabeli. To pomeni, da obstoječe tiskalnike z marjetico še naprej uporabljamo in po potrebi izmenjujemo marjetico v skladu s kodno tabelo. Kot zaključek temu lahko povemo še, da je cena tiskalnikov in terminalov sorazmerno nepomembna glede na ceno programske opreme.

3.2 Implementacija ISO 6937 in ISO 8859 v programski opremi

Prednost kode, v kateri je vsak znak določen z enim zlogom, je neprimerljiva s kodo z nekonstantno dolžino. Velika večina obstoječe programske opreme zelo lahko obdeluje 8-bitno kodo iz ISO 8859. Problemi se pojavljajo v primerih, ko želimo, da naša programska oprema sprejema dve kodni tabeli, na primer ISO 8859/2 in ISO 8859/3. V teh primerih uporabljamo ali: -poseben krmilni znak ("single shift" funkcijo), ki nam kaže, da so prihajajoči znaki predstavljeni z dvema zlogoma, prvi zlog je koda za SSI in drugi za sam znak, -poseben znak ("locking shift" funkcijo), s katerim spremenimo celo desno stran kodne tabele. Nekateri od zagovornikov ISO 6937 trdijo, da je prednost te kodne tabele v možnosti njene uporabe kot 7 in 8-bitno, v odvisnosti od sprejemne naprave. Pri terminalih se da problem sprejemanja znakov z različno dolgimi zlogi dokaj lahko rešiti, pri programski opremi pa se zatakne, ker vemo, da je težko pisati programe, ki sprejemajo hkrati 7 in 8-bitne zloge. Pri 7-bitnih zlogih večina zahodno evropskih pisav uporablja SSI in LSI funkcije za potrebe generiranja akcentiranih znakov. Pri tem ISO 6937 kaže na določene prednosti, ker ni pojava redundantnega kodiranja (v kodni tabeli ni enakih znakov). To velja le, če se uporablja samo ISO 6937/1/2. Pri uporabi drugih kodnih tabel iz ISO 6937, ki so v pripravi (na primer grške črke iz grške abecede in iz kodne tabele za tehnično uporabo) se ta prednost zgubi.

4. Aplikacije in uporaba ISO 6937 in ISO 8859

Problemi, ki lahko nastanejo pri uporabi ISO 6937 pri obdelavi podatkov, so mogoče najbolj vidni pri sortiranju. Kako lahko primerjamo dva znaka, od katerih je eden brez akcenta in je zapisan s 8-biti in drug akcentiran znak, ki je zapisan s 16 biti če v programski opremi za sortiranje to ni posebej specificirano, lahko nastanejo težave. Pogosta tehnika pri sortiranju je transformacija vsakega znaka v numerično vrednost in uporaba te vrednosti za primerjavo.

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0000	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
0001	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	
0010	02	03	04	05	06	07	08	09	10	11	12	13	14	15		
0011	03	04	05	06	07	08	09	10	11	12	13	14	15			
0100	04	05	06	07	08	09	10	11	12	13	14	15				
0101	05	06	07	08	09	10	11	12	13	14	15					
0110	06	07	08	09	10	11	12	13	14	15						
0111	07	08	09	10	11	12	13	14	15							
1000	08	09	10	11	12	13	14	15								
1001	09	10	11	12	13	14	15									
1010	10	11	12	13	14	15										
1011	11	12	13	14	15											
1100	12	13	14	15												
1101	13	14	15													
1110	14	15														
1111	15															

Sl. 2. ISO 8859/8

8859, bi lahko nastali sledeči problemi: -obstoječi terminali z implementirano ISO 6937 kodno tabelo bi postali neuporabni (na primer Xerox terminali). -11 evropskih proizvajalcev je obljubilo, da bodo podpirali ISO 6937 in bi nova odločitev povzročila zmedo -CEN/CENELEC, ki je sprejel obveznost, da podpre ISO 6937 bo postavljen v nerodno situacijo.

3. Stališča glede implementacije ISO 6937 in ISO 8859

3.1 Implementacija na terminalih in tiskalnikih

Če je v terminalu implementiran ISO 6937, potem ni problem implementirati tudi ISO 8859, ker je množica znakov enaka. Spremeniti je treba le del strojne opreme (firmware), ki namesto generiranja znaka kot kombinacije diakritičnega znaka in navadnega znaka generira le en znak.

V primeru, ko imamo naprave z visoko resolucijo, kjer za vsak znak potrebujemo veliko število bitov, se akcentirani znaki generirajo kot kombinacija zgornega in spodnjega akcenta ter same črke. Pri napravah nizke resolucije se akcentirani znak generira naenkrat brez kombinacij.

Večina CPU-jev ima (tako kot IBM 370), vgrajen ukaz prav za to operacijo. Kaj se bo zgodilo, če namesto na 8-bitni zlog naleti ukaz na 16 bitov. Tudi ostali programi za konverzijo med ASCII in EBCDIC kodo so narejeni na podoben način - translacija le 8-bitnih zlogov.

Drug problem, ki lahko nastane pri uporabi ISO 6937, je urejanje teksta. Večina zaslonko orientiranih urejevalnikov je narejenih na podlagi predpostavke, da mesto, ki ga zavzame črka na zaslonu, odgovarja enemu znaku, zakodiranemu z 8-biti. Tehnika, s katero pozicioniramo kursor, sloni na predpostavki, da je vsak znak zakodiran z enako dolgim zlogom. Pozicioniranje kursorja ter ažuriranje zaslona v primeru implementacije ISO 6937 bo zahtevalo spremembe v opremi zaradi prilagajanja na spremenljivo dolžino kode. Podobni problemi lahko nastajajo pri specifikacijah dolžin polj. Koliko znakov bomo spravili v polju z določeno dolžino, bo odvisno od tega, kakšni znaki bodo nastopali. Na primer: ukaz "enter last name, up to 12 character" nam ne zagotavlja, da bomo dobili pravilen zapis v podatkovni bazi, ker je to, koliko znakov bomo pobrali, odvisno od tega, kateri so ti znaki.

5. Sklepne misli

Iz vsega povedanega lahko izločimo sledeče:

ISO 6937 je aplikacijsko orientiran standard namenjen za implementacijo v storitvah Videotex in Teletex ter za naprave, ki to aplikacijo omogočajo. Po definiciji ISO-ja in CCITT-ja je tekst informacija, namenjena ljudem, ki to informacijo sprejemajo, jo obdelajo in jo naprej posredujejo, zato sta 8-bitni ASCII in ISO 8859 standarda namenjena predvsem za prenos in obdelavo teksta, ISO 6937 pa le komunikacijam s tekstovnimi podatki. To pomeni, da je treba ISO 6937 uporabljati le pri čistem prenosu in v primerih, ko je tekst shranjen na magnetnih medijih, le za prikazovanje. Prvotna ideja SC2 WG4 pri razvoju ISO 6937 je bila izdelati splošno uporaben standard za prenos in izmenjavo podatkov. Opozorila strokovne javnosti, da je ta standard neprimeren za splošno uporabo zaradi problemov pri obdelavi, so povzročila, da je SC2 WG4 dobil nalogo, da še enkrat obdelata ISO 6937 in omeji natančneje njegovo uporabo.

Eden od glavnih motivov za razvoj ISO 8859 je bilo zavračanje nekaterih proizvajalcev, da implementirajo ISO 6937. Kljub temu, da bo širok sprejem ISOa 8859 obsodil terminale z implementiranimi ISO 6937, večina proizvajalcev trdi, da obstoječe terminalne naprave v največji možni meri podpirajo 8-bitne kode s podobnimi lastnostmi kot ISO 8859. Pričakovati je, da bo to število še večje po sprejemu ISO 8859/1-8 v okviru ISO SC2 in ANSI B8R X3.134.2. Čeže, da je večina ameriških proizvajalcev pripravljena podpreti ASCII terminale, v katerih bo vključena tudi najnovejša ASCII 8-bitna koda. V primeru,

da bo ISO 6937 še naprej podpiran od nekaterih proizvajalcev kot prevladajoč standard za slepi prenos, ni treba pričakovati, da bodo ameriški proizvajalci in njihovi uporabniki prilagajali svoje aplikacije temu standardu.

Dosedanja razprava kaže, da bo najverjetneje v kratkem (na plenarnem zasedanju ISO SC2 marec 1987) sprejeto priporočilo ANSI X3 in ISO SC2, s katerim bo ISO 8859 postal prevladujoč standard za prenos in obdelavo podatkov, vključno z uporabo pri izmenjavi tekstovnih podatkov v okolju avtomatiziranih pisarn. Ugotovitev, da mora uporaba računalnikov sloneti na enakih standardih ne glede na to, kje se računalnik nahaja, ali v pisarni ali nekje drugod, je bila eden od pomembnejših dejavnikov, ki so pripeljali do združitve med ISO TC95 (office systems) in ISO TC97 (information processing). V tem kontekstu je uporaba različnega nabora znakov v avtomatizirani pisarni in v poslovno proizvodnem sistemu podjetja skrajno nezaželena.

Obstaja verjetnost, da ne bo mogoče zaradi obstoječih telematskih mrež popolnoma izpodrinuti ISO 6937. ISO 6937 je koda, ki je narejena za specifično uporabo (Videotex in Teletex) in taka mora tudi ostati. Terminali in tiskalniki z določenimi dodatki lahko interpretirajo oba standarda, zato je treba pričakovati, da bo njihova uporaba in prodaja ostala neprizadeta. Te naprave lahko uporabljamo pri telematskih storitvah, lahko pa kot periferijo računalnikov s splošno uporabo. Konverzija med ISO 6937 in ISO 8859 bo relativno enostavna, ker vsebuje ISO 6937 podmnožico znakov ISO 8859. Translacijske podatke, ki so namenjeni nadaljnji obdelavi iz formatiranih podatkov lahko zmeraj opravimo s programom za preformatiranje strani.

Iz vsega navedenega sledi, da je ISO 8859 boljše tehnično rešitev in da je bolj splošno uporabna. Tako je pričakovati, da bo ISO najverjetneje priporočil kot splošno uporaben standard ISO 8859. Priporočati le eno možno rešitev pomeni, da bodo nekateri pri tem prizadeti in da bodo drugi pridobili na račun prvih. Naša naloga je, da podpremo boljše tehnično rešitev v interesu boljše standardizacije in reševanja naših potreb (uporaba tudi cirilice).

6. Reference

1. Delovni dokumenti ISO TC97/SC2, WG4 in WG2
2. Delovni dokumenti ANSI X3L2