

ELEKTRONINIS PAŠTAS

Kodų lentelių, vartojamų elektroniniame pašte, analizė

Gintautas Grigas

Matematikos ir informatikos instituto Programavimo metodologijos skyriaus
vadovas, docentas, daktaras

Institute of Mathematics and Informatics, Department of Programming Methodo-
logy, Associate Professor, Doctor

Akademijos 4, Vilnius

Elektroninis paštas gali būti laikomas pilnaverte bendravimo priemone tik tada, kai juo galima persiųsti tekstą, turintį pagrindinius kalbos ženklus: raides, skaitmenis, skirtukus. Apie lietuviškų rašmenų persiuntimo elektroniniu paštu problemas bei galimus jų sprendimus buvo kalbama 1997 metais vykusiose Pasaulio liuanistų bendrijos (Dagienė, 1998; Grigas, 1998) ir Lietuvos kompiuterininkų sąjungos (Grigas, 1987) konferencijose. Per metus techninės problemos buvo išspręstos: parinkta ir lietuviškiems rašmenims pritaikyta bent po vieną nemokamai platinamą elektroninio pašto programą Lietuvoje plačiau naudojamose operacinėse sistemose DOS, „Windows“, „Mac OS“ ir UNIX. Apie tai buvo rašyta populiarioje spaudoje (Grigas, 1998; Grigas, 1998; Misevičius, 1998). Taigi atsirado reali galimybė kiekvienam elektroninio pašto abonentui sutvarkyti savo programinę įrangą taip, kad jis būtų tikras, jog jo parašytas laiškas pasieks adresatą ir laiško gavėjo programa turės visą informaciją, kaip ekrane parodyti laišką, kad jame būtų matomos tokios lietuviškos raidės, kokias parašė siuntėjas, nepriklausomai nuo to, kokias operacines sistemas naudoja ir kaip koduoja lietuviškas raides laiško siuntėjo ir gavėjo kompiuteriai.

Problema sprendžiama taip: laišką siunčianti programa jį perkoduoja iš kompiuteryje naudojamos kodų lentelės (trumpiau – kodo) į kodą, naudojamą

elektroniniame pašte, o laišką priimanti programa laišką išskodoja iš elektroniniame pašte naudojamo kodo į laiško gavėjo kompiuterio kodą, kuris gali būti kitoks, negu laiško siuntėjo kompiuteryje. Tačiau lietuviškiems laiškam persiūsti elektroniniame pašte naudojamas ne vienas, o trys kodai: ISO/IEC 8859-4 (1 pav.), ISO/IEC 8859-13 (trumpiau ISO 8859-4 ir ISO 8859-13) (2 pav.) ir „Windows1257“ (trumpiau – 1257) (3 pav.). Kuriuo iš šių kodų siunčiamas laiškas, siuntėjo programa informuoja gavėją įrašydama į laiško antraštę vieną iš trijų eilučių:

Content-type: text/plain; charset=iso-8859-4

Content-type: text/plain; charset=iso-8859-13

Content-type: text/plain; charset=x-cp1257

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
80	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
90																
A0		À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î
	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
B0	°	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î
	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
C0	Ā	Ā	Ā	Ā	Ā	Ā	Æ	Ķ	Č	É	Ē	Ē	Ē	Ē	Ī	Ī
	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
D0	Đ	Ņ	Ō	Ķ	Ô	Õ	Ö	×	Ø	Ū	Ú	Û	Ü	Ũ	Ū	ß
	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
E0	ā	á	â	ã	ä	å	æ	ī	č	é	ē	ē	ē	ē	ī	ī
	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
F0	đ	ņ	ō	ķ	ô	õ	ö	÷	ø	ų	ú	û	ü	ũ	ū	
	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

1 pav. Kodų lentelė ISO 8859-4. Pirmoji lentelės dalis (kodai 0–127), sutampanti su ASCII kodu, paveiksle neparodyta

Kelių kodų naudojimas gerai sutvarkytame elektroniniame pašte techninių problemų neturėtų sukelti, kadangi šiuolaikinės elektroninio pašto stotys bei pašto abonentų programos vienodai gerai susidoroja su visais čia paminėtais kodais. Tačiau vis tik yra tam tikrų skirtumų, kuriuos verta paanalizuoti.

Juridinis statusas. Kodai ISO/IEC 8859-4 ir ISO/IEC 8859-13 yra Tarptautinės standartų organizacijos (ISO) standartai. Pirmasis buvo priimtas 1988, antrasis – 1998 metais ir šiuo metu dar tik rengiamas publikuoti, tačiau tarptautinėje visuomenėje jau žinomas.

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
80																
90																
A0		”	¢	£	€	„	¡	§	Ø	©	®	«	¬	-	®	Æ
B0	°	±	²	³	“	µ	¶	·	ø	¹	²	»	¼	½	¾	æ
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ĝ	Ĥ	İ	Ł
D0	Š	Ń	Ň	Ó	Ô	Õ	Ö	×	Ů	Ł	Ś	Ū	Ů	Ž	Ž	ß
E0	ą	ı	ā	ć	ä	å	ę	ē	č	é	ž	è	ğ	ķ	ī	ł
F0	š	ń	ň	ó	ô	õ	ö	÷	ų	ł	ś	ū	ü	ž	ž	'

2 pav. Kodų lentelė ISO 8859-13. Pirmoji lentelės dalis (kodai 0–127), sutampanti su ASCII kodu, paveiksle neparodyta

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
80			,		„	...	†	‡		‰		<				
90		‘	’	“	”	•	—			™		>				
A0		¢	£	¤		¦	§	Ø	©	®	«	¬	-	®	Æ	
B0	°	±	²	³		µ	¶	·	ø	¹	²	»	¼	½	¾	æ
C0	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ĝ	Ĥ	İ	Ł
D0	Š	Ń	Ň	Ó	Ô	Õ	Ö	×	Ů	Ł	Ś	Ū	Ů	Ž	Ž	ß
E0	ą	ı	ā	ć	ä	å	ę	ē	č	é	ž	è	ğ	ķ	ī	ł
F0	š	ń	ň	ó	ô	õ	ö	÷	ų	ł	ś	ū	ü	ž	ž	

3 pav. Kodų lentelė 1257

Kodų lentelė ISO/IEC 8859-4 buvo priimta 1989 metais kaip Lietuvos standartas LST 1093-89 (LST, 1989) ir galiojo iki 1993 metų, kol dar nebuvo priimtas naujas standartas LST 1283:1993 (LST, 1993) su kodų lentele „Baltic Rim“, sutampančia su ISO 8899-13.

Kodų lentelę 1257 firma „Microsoft“ naudoja „Windows“ terpėje, todėl ją reikėtų laikyti šios firmos standartu.

Elektroniniame pašte aukščiausią statusą turi tarptautinio ISO 8859–x standarto kodų lentelės. Laikomasi nuostatos, kad tarptautinius standartus turi pripažinti visos pašto stotys bei programos.

Ženklių aibės ir jų kodai. Visose minėtose trijose kodų lentelėse yra visos specifinės lietuviškos raidės (Aą, Čč, Eę, Èè, Įį, Šš, Ūų, Ūū, Žž).

Visų lietuviškų raidžių kodai lentelėse ISO 8859–13 ir 1257 sutampa.

Visos trys kodų lentelės turi iš ASCII kodo perimtus skyrybos ženklus, reikalingus lietuviškai rašyti: tašką (.), kablelį (,), kabliataškį (;), brūkšnelį (-), klausuką (?), šauktuką (!), skliaustus (()).

Kodų lentelė ISO 8859–13 turi dar tris papildomus skyrybos ženklus, reikalingus lietuviškai rašyti: atidarantią kabutę („ (kodas 165)), uždarančią kabutę (“ (kodas 180)), apostrofą (’ (kodas 255)).

Kodų lentelė 1257 turi penkis papildomus skyrybos ženklus, reikalingus lietuviškai rašyti: brūkšnį (– (kodas 150) arba — (kodas 151)), atidarantią kabutę („ (kodas 132)), uždarančią kabutę (“ (kodas 147)), apostrofą (’ (kodas 146)). Vienne dokumente paprastai vartojamas tik vienos rūšies pasirinkto ilgio brūkšnys. Todėl iš tikrųjų yra keturi papildomi ženklai.

Trijų papildomų skyrybos ženklų, esančių abiejose lentelėse, kodai skiriasi. Tai galima paaiškinti skirtingais šių kodų lentelių projektavimo principais.

Visose ISO 8859–x kodų lentelėse yra neišnaudoti 128–160 kodai: čia pateiktuose kodų lentelių paveiksluose (žr. 1 ir 2 pav.) dvi eilutės (80 ir 90) yra tuščios. ASCII lentelėje pirmieji 32 kodai yra skirti valdymui. Tarptautinė standartų organizacija analogiškai pasielgė ir su pirmaisiais antrosios lentelės pusės kodais – juos rezervavo valdymo reikalam. Kita priežastis, susijusi su elektroniniu paštu – šios eilutės paliktos tuščios, kad išliktų suderinamumas su ankstesne telekomunikacijų įranga, pritaikyta tik 7 bitų kodams. Kai į ją patenka 8 bitų kodas, aštuntasis bitas dažniausiai lieka nepastebėtas ir priimami tik pirmieji 7 bitai. Kodas sumažėja dešimtainiu 128 arba šešioliktainiu F0. Jeigu originalaus simbolio kodas buvo iš intervalo 161–255, tai jis virsta kitu teksto simboliu. Duomenys sugadinami, bet įrangai tokia klaida nekenkia – jai nesvarbu, kokie duomenys persiunčiami. Tačiau kai teksto simbolis virsta valdančiuoju simboliu, tai jis suprantamas kaip komanda. Ta komanda yra neteisinga ir gali sutrikdyti įrangos darbą.

Visose „Windows“ kodų lentelėse 125x minėtas nelegalias pozicijas panaudojo rašto ženklams koduoti. Todėl į jos kodų lenteles telpa daugiau

ženklų. Šitaip atsirado galimybė turėti daugiau specialių simbolių. Ženkilai, sudėti į šias dvi eilutes, beveik visose kodų lentelėse 125x yra tie patys ir tose pačiose vietose. Tačiau dalis šių ženklų jau buvo kitose (legaliose) kai kurių ISO 8859 lentelių pozicijose. Dviejose lentelės vietose tas pats simbolis negali būti. Todėl tokie simboliai buvo perkelti į 128–160 pozicijas (ten, kur ir kitose „Windows“ lentelėse yra tie patys simboliai).

Laiškus rašyti patogiau, kai galima vartoti didesnę ženklų aibę. Todėl laiškų rašymo požiūriu geriausia yra kodų lentelė 1257, nes jos ženklų aibė turtinčiausia: turi lietuviškas kabutes, apostrofą, brūkšnį. Nuo jos mažai atsilieka ISO 8859–13 lentelė, turinti kabutes ir apostrofą.

Kaip minėjome, telekomunikacijose nerekomenduojami kodai 127–160. Todėl patikimiau siunčiamuose laiškuose šių kodų nenaudoti, t. y. elektriniame pašte naudoti ISO 8859–4 arba ISO 8859–13 kodų lenteles.

Rašto ženklų kodavimas šiuolaikinėms ir tvarkingai veikiančioms pašto programoms nedaro įtakos, kadangi jos vis tiek perkoduoja laiškus, o į kokį kodą perkoduoti – nesvarbu. Tačiau reikėtų atsižvelgti ir į išimtinius atvejus, laikytinus sutrikimais elektroninio pašto trakte. Pavyzdžiui, gali pasitaikyti, kad laiškas patenka adresatui, kurio pašto programa neatlieka visų jai skirtų funkcijų ir negali perkoduoti laiško (neteisingai sureguliuota, netenkinanti MIME protokolo reikalavimų, neturinti lietuviškų kodavimo lentelių ir pan.). Tokiu atveju laiško gavėjas matys neperkoduotą laiško tekstą. Jeigu gautame laiške lietuviškų raidžių kodai sutaps su kodais gavėjo kompiuteryje, tai laiškas bus matomas teisingas. Atsižvelgiant į tai, kad šiuo metu daugiausia paplitusi operacinė sistema „Windows“, tai laiškus, koduotus 1257 lentele, matys ir tie šios sistemos vartotojai, kurių pašto programos nesugeba jų perkoduoti. Jie taip pat teisingai matys ir visus ženklus, išskyrus kabutes ir apostrofą, laiškuose, koduotuose ISO 8859–13 kodu.

Situacija kitų kalbų kodų lentelėse. Kitose valstybėse, vartojančiose lotyniškąją abėcėlę, yra analogiška situacija. Atitikmenys tarp ISO 8859–x ir 125x kodų pateikti 1 lentelėje. Situacija Vakarų Europos ir Vidurio Europos valstybėse nuo lietuviškosios skiriasi tuo, kad jų tarptautiniai standartai ISO 8859–1 ir ISO 8859–2 buvo sudaryti daugiau kaip prieš dešimtmetį (1987 m.), todėl tapo gerai žinomi tarptautinei visuomenei, pagal juos buvo sudarytos kodų lentelės 1250 ir 1252.

1 lentelė. Kodų lentelės, skirtos kalboms, vartojančioms lotyniškąją abėcėlę

ISO	„Windows“	Regionas	Kalbos (pagal „Windows95PE“)
ISO 8859-1	1252	Vakarų Europa	afrikanų, anglų, baskų, danų, indoneziečių, islandų, ispanų, italų, kataloniečių, norvegų, olandų, portugalų, prancūzų, suomių, švedų, vokiečių
ISO 8859-2	1250	Vidurio Europa	albanų, čekų, kroatų, lenkų, slovaku, slovenų, rumunų, vengrų
ISO 8859-9	1254	Turkija	turkų
ISO 8859-13	1257	Baltijos valstybės	estų, latvių, lietuvių

Pastaba. Į lentelę įtrauktos tik tos ISO 8859-x kodų lentelės, kurios turi atitikmenis antrame stulpelyje išvardytų „Windows“ kodų grupėje.

Šiaurės Europos ir Baltijos šalims 1988 metais buvo sudarytas standartas ISO 8859-4. Į šį kodą buvo stengtasi įtraukti kuo daugiau kalbų abėcėlių ir dėl to mažai vietos liko specialiesiems (skyrybos) ženklams. Dėl to jis nepriėjo nei Šiaurės, nei Baltijos šalyse. Kai firma „Microsoft“ kūrė kodus 125x, Baltijos šalys jau buvo sudariusius jų poreikiams geriau tinkantį standartą „Baltic Rim“ (Misevičius, 1998). Pagal jį buvo sudarytas kodas 1257 ir tarptautinis standartas ISO 8859-13. Tarptautinio standarto priėmimas ir tvirtinimas yra ilga procedūra. Todėl daug kas užsienyje mano, kad Lietuvoje (taip pat ir Latvijoje) dar tebenaudojamas kodas ISO 8859-4, ir į programinę įrangą įdeda jo lentelę. Dėl to Lietuvoje susidarė išskirtinė situacija – elektroniniame pašte naudojamas kodas ISO 8859-4, neturintis atitikmens kodų 125x šeimoje.

Skirtumai tarp kitų valstybių ISO 8859-x ir 125x kodų porų taip pat turi savų ypatumų. Trumpai apžvelgsime juos.

Į Vakarų Europos valstybių kalboms skirtą kodų lentelę 1252, į tas pačias pozicijas yra perkelti visi standarto ISO 8859-1 ženklai. Todėl lentelė 1252 buvo papildyta dviem eilutėmis (4 pav.) su skyrybos ir kitokiais specialiais ženklais, kurių nėra standarte ISO 8859-1.

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
80			,	ƒ	„	…	†	‡	^	‰	Š	<	Œ			
	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
90		‘	’	“	”	•	—	—	~	™	š	>	œ			ÿ
	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159

4 pav. Dvi kodų lentelės 1252 eilutės, kurių nėra ISO 8859-1 standarte

Jei laiško gavėjo programoje neveiks perkodavimas, laiško gavėjas, esantis „Windows“ terpėje, teisingai matys visus ženklus, išsiųstus ISO 8859–1 kodu. Lyginant su lietuviška situacija, tai bus lyg ir privalumas: laiške lietuvių kalba analogišku atveju bus neteisingi kabučių ir apostrofo simboliai, kurių kodai analogiškoje lentelių poroje yra skirtingi. Tačiau tai tik formalus privalumas – kode ISO 8859–1 minėtų trijų ženklų išvis nėra, nors jie ir vartojami kai kuriose Vakarų Europos kalbose (islandų, vokiečių).

Standartas ISO 8859–9 skiriasi nuo standarto ISI 8859–1 tuo, kad jame trys islandiškos raidės pakeistos trimis turkiškomis raidėmis. Visa kita, kas buvo pasakyta apie ISO 8859–1, tinka ir šiam standartui.

Vidurio Europos kalbų abėcėlės turi daug specifinių raidžių su įvairiais diakritiniais ženklais. Standarte ISO 8850–2 net 20 raidžių (5 pav.) pateko į kodų lentelės eilutes A ir B, kur 125x lentelėse priimta dėstyti specialiuosius ženklus. Dėl to minėtas 20 raidžių 1250 lentelėje teko perkelti į kitas pozicijas – nelegalią zoną (6 pav.). Dėl to laiške, išsiųstame ISO 8859–2 kodu ir skaitomame su pašto programa, kurioje neveikia perkodavimai, dalis raidžių bus pakeista kitais ženklais. Tai žymiai blogesnė situacija, negu su lietuviškų kodų ISO 8859–13 ir 1257 pora.

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
A0		À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î
B0	°	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î
	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191

5 pav. Dvi kodų lentelės ISO 8859–2 eilutės

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
80		‚	ƒ	„	…	†	‡	ˆ	‰	Š	‹	Ś	Ŧ	Ž	Ż	
90	‘	’	“	”	•	—	~	™	š	›	ś	ŧ	ž	ż		
A0	˘	˙	Ł	ł	Ą	ą	Ś	ś	©	Ş	«	¬	–	®	Ž	
B0	°	±	˘	ı	˙	µ	¶	·	¸	ş	»	Ł	ł	Ŧ	Ż	
	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191

6 pav. Keturios kodų lentelės 1250 eilutės

Išvados

Palyginę kodų lenteles, turinčias lietuviškus rašto ženklus, ir įvertinę analogišką situaciją kitose valstybėse vartojamose kodų lentelėse, galime

tvirtinti, kad siunčiamų laiškų kodavimui geresnes perspektyvas turi kodas ISO 8859–13, kuris yra ir Lietuvos, ir tarptautinis standartas. Vienintelis jo trūkumas lyginant su 1257 kodu yra tas, kad jame nėra lietuviškame rašte vartojamo brūkšnio. Tačiau su šiuo trūkumu galima taikstyti, nes brūkšnį elektroniniame laiške galima pakeisti dviem brūkšneliais, ką ir daro tvarkingai rašantys laiškus. Kitų dviejų kodų lentelių privalumai (arba ISO 8859–13 trūkumai) yra laikini, nes reikia tikėtis, kad laikui bėgant kodas ISO 8859–13 taps daugiau žinomas negu ISO 8869–4, o pasenusių arba blogai sureguliuotų pašto programų skaičius mažės.

Literatūra

1. Dagienė V. Šiuolaikinės informacinės problemos švietime: kalbos problema. – *Lituanistika pasaulyje šiandien: darbai ir problemos*. T. 3. Vilnius: Baltos lankos, 1998. P. 55–64.
2. Grigas G. Lietuviškų rašmenų panaudojimo kompiuteriuose ir jų tinkluose problemos. – *Lituanistika pasaulyje šiandien: darbai ir problemos*. T. 3. Vilnius: Baltos lankos, 1998. P. 65–72.
3. Grigas G. Lietuviškų rašmenų problemos kompiuterių pašte. – *Kompiuterininkų dienos'97: Renginio medžiaga*. Vilnius: Žara, 1997. P. 125–128.
4. Grigas G. Tarptautiniai standartai, lietuviškos kodų lentelės ir elektroninis paštas. – *Kompiuterija*, 1998, Nr. 5(9). P. 36–38.
5. Grigas G, Laurinavičius T. UUPC jau gali „susikalbėti“ su kitomis pašto programomis. – *Kompiuterija*, 1998, Nr. 3(7). P. 28–30.
6. LST 1093–89. Informacijos apdorojimo sistema. Daugiakalbis koduotų ženklų rinkinys su lietuviškomis raidėmis. Aštuonių bitų kodai. Vilnius, 1989.
7. LST 1282:1993. Informacijos technologija. 8 bitų lotyniškos abėcėlės koduotų ženklų rinkinys, papildytas lietuviškos abėcėlės raidėmis, skirtas naudoti WINDOWS, UNIX aplinkoje. Vilnius, 1993.
8. Misevičius R. Pašto programa „Pine“ ir lietuviški rašmenys. – *Kompiuterija*, 1998, Nr. 6(10). P. 38.

[teikta 1998 m. rugsėjo mėn.]

Analysis of Code Tables Used in Electronic Mail

Summary

The technical problems of sending Lithuanian texts by electronic mail and their solutions using MIME protocol are shortly discussed. Comparative analysis of three code tables: ISO 8859–13, Windows–1257 and ISO 8859–4 mostly used in Lithuanian electronic mail is presented. The code tables are compared according to their standardization status, character sets and prevalence in software. It is shown that the code table ISO 8859–13 has the best perspective.