

Standardi EVS 8:2000 rakendusjuhend

**User's Guide to EVS 8:2000 Requirements on information
technology in Estonian language and cultural environment.**

(seisuga 23. detsember 2000. a.)

Võtmesõnad: keele- ja kultuurikonventsioonid; tarkvara lokaliseerimine; eesti-ladina tähestik;
latinisatsioon; kooditabelid; Eesti lokaat.

<http://www.eki.ee/itstandard/>

Sisukord

SISSEJUHATUS	4
<i>Eesti kohta käivad määrangud standardis EVS 8:2000</i>	4
TARKVARA ARENGUSUUNAD	5
Unicode	5
UTF-8	5
Internatsionaliseerimine ja lokaliseerimine	6
Lokaat	7
Operatsioonisüsteemide kasvav tähtsus	7
ESET1	9
Peatükis kasutatud terminid ja mõistekäsitlused	9
Tähtede liigid. Diakriitikutabel	10
ESET1 tarvitusjuhiseid	11
Laiema tähevalikuga kaasnevaid eksimisvõimalusi	13
Jutumärkide tarvitus	14
SORTIMINE	15
KLAVIATUUR	21
Teksti sisestamisel ja andmevahetusel tekkivaid tüüpvigu	23
TELEFONINUMBRITE KASUTUS	25
Peatükis kasutatud terminid ja mõistekäsitlused	25
Numbri valimine geograafiliselt numbrilt	25
Telefoninumbrite õigekiri	26
OPERATSIOONISÜSTEEMIDE KOHANDAMINE EESTI KEELE KASUTUSEKS	28
Eesti keele tugi Linuxis (Mandrake 7.1 näitel)	28
Sortimine, kuupäeva, numbrite ja rahasummade formaat	28
Klaviatuuripaigutus	29
Ajavöönd	29
Kokkuvõtteks	29
Eesti keele tugi Windowsis	29
Kokkuvõtteks	30
HTML, ELEKTRONPOST JA BRAUSERI SEADED	31
Internet Explorer, Outlook Express ja Netscape Communicator	32
KOODITABELID	34
TARKVARA VASTAVUSE NÕUDED STANDARDILE EVS 8:2000	35
Operatsioonisüsteemid	35
Rakendustarkvara	35
Riistvara	36
LISA 1. MAAILMA RIIGID JA NENDE AMETIKEELED	37
LISA 2. KEELED JA NENDES ESINEVAD LISATÄHED	42
Keeled, mille tähestik (või latinisatsioon) ei sisalda ladina alustähestikust erinevaid tähti	42
Keeled, mille puhul kasutatakse eesti transkriptsiooni	42
Keeled, mille tähestik või latinisatsioon sisaldab ladina alustähestikust erinevaid tähti	42
LISA 3. VABARIIGI VALITSUSE MÄÄRUS VÕORKEELSETE ISIKUNIMEDE KIRJUTUSREEGLITE KOHTA ...	49
Võorkeelsete isikunimede ümberkirjutusreeglite kehtestamine	49
16. novembril 1998 valminud Haridusministeeriumi eesti-ladina töörühma aruanne	52
LISA 4. EESTIS KASUTATAV LADINA TÄHTEDE VALIK ESET1	54
Märkused	54
Muud isiku- ja kohanimes esinevad sümbolid	74
Koostamispõhimõtted	75
LISA 5. VALIK ENAMLEVINUD KOODITABELEID	76
ISO 8859-15	77
ISO 8859-1	78
Windows CP1252 (Western)	79
DOS CP 850	80
ISO 8859-13	81
Windows CP 1257 (Baltic)	82
DOS CP 775	83
CENTEURO (Macintoshi Kesk- ja Ida Euroopa kooditabel)	84
LISA 6. TERMINISÕNASTIK	85

Sissejuhatus

Standard EVS8:2000 "Infotehnoloogia reeglid eesti keele ja kultuuri keskkonnas" ilmus ametlikult jaanuaris 2000. Vaatamata suhtelisele mahukusele (70 lehekülge) ja üldorientatsioonile jäi sealt siiski välja rida konventsioone, kirjeldusi, tähelepanekuid ja näiteid, mis oma olemuselt ei sobinud standardi rängesse raami. Kui EVS 8:2000 oli oma põhisuunitluselt abimaterjal neile, kes vaatlesid Eestit väljastpoolt, on rakendusjuhendis oma koha leidnud põhiliselt materjal, mis abistab Eesti arvutikasutajat laiema kultuurikonteksti nägemisel infotehnoloogias. Käesolev rakendusjuhend täidab seega eeskätt täpsustavat rolli, lisades alusstandardile kommentaare, põhjendusi ühe või teise tehtud valiku osas, standardile EVS 8:2000 vastavuse tingimused ja rakendussoovitusi. Vastavalt rakendusjuhendi eesmärgile on püütud, kus võimalik, vältida liigset formaliseeritust probleemide kirjeldamisel ja puudub ingliskeelne paralleeltekst.

EVS 8:2000 veebiversiooni, käesoleva rakendusjuhendi ja viidad täiendavale materjalile leiate aadressil <http://www.eki.ee/itstandard/>

Eesti kohta käivad määrangud standardis EVS 8:2000

- Arvude murdosa eraldaja on **koma**
- Rahaühikute eraldaja on **punkt**
- Suveaega ei kasutata. Eesti ajavööndi tähistus on **EET** (vahe UTCga 2 tundi)
- ISO 3166 kahetäheline kood (alpha-2) **EE** on üldiselt kasutuses Eesti maakoodina
- ISO 3166 kolmetäheline kood (alpha-3) **EST**
- ISO 3166 koodnumber **233**
- CEPT-MAILCODE **EE**
- UN Genève 1949:68 liiklusvahendi kood **EST**
- ITU E.164 rahvusvaheline sissehelistamiskood **372**
- telefonisides: rahvusvaheline prefiks **00**, riigisisene prefiks **0**
- ITU X.121 numbriline maakood **248**
- EAN prefiks **474**
- ISO 2108 ISBN raamatukood **9985**
- Keele kood vastavalt standardile ISO 639-1 on **et** (mitte ee, nagu sageli ekslikult kasutatakse!)
- Keele kood vastavalt standardile ISO 639-2/T (terminoloogia) on **est**
- Keele kood vastavalt standardile ISO 639-2/B (raamatukogundus) on **est**
- Krooni tähistus ISO 4217 järgi on **EEK**. Krooni lühend on **kr**
- Tulevikku silmas pidades: rahaühik euro tähistus on EUR (kasutatakse analoogiliselt EEKga), märk € või lühendamata väljakirjutatud sõna 'eurot'. Seega EUR 1 = 1 € = 1 euro ja EUR 43.21 = 43.21 € = 43.21 eurot.

Tarkvara arengusuunad

Unicode

Unicode on rahvusvaheline standard, mille eesmärgiks on eraldada kõigi maailma keelte kõikidele märkidele üheselt määratud koodid ning tagada andmevahetus. Unicode on juba laialdaselt kasutusel Internetis (sellel tuginevad HTML ja XML), operatsioonisüsteemides (Windows NT, Solaris), programmeerimisel (Java) jm. Unicode'i kolmas versioon valmis 2000 aasta alguses, sellesse on kantud umbes 50000 märki (neist üle poole on hieroglüüfid). Rahvusvahelise Standardiorganisatsiooni vastav standard kannab numbrit **ISO/IEC 10646** ja on täpses vastavuses Unicode'iga; käesolev rakendusjuhend tarvitab neid sünonüümidena.

Kuigi hetkel on standardi põhitähelepanu suunatud kahe baidi abil kirjutatavate märkide kodeerimisele (neid on kokku u 65000, seda koodipiirkonda tähistatakse lühendiga BMP e *Basic Multilingual Page*), on standard juba ette näinud edasise laienemise ja kuni miljoni märgi kodeerimise. Unicode kirjeldab märke abstraktsel tasemel – märgid on nt ladina suur R (U+0052), ladina väike r (U+0072) või bengali number 5 (U+09EB), konkreetne märgi kuju tekstis ehk glüüf (R, R, **R**, R) sõltub rakendusprogrammist ja seda Unicode ei sätesta. Eri koodiga on ladina suur A ja kirillitsa suur А, sisulise vahe toob kaasa juba nende väliskujult identsete märkide esinemine eri kirjades. Sisulisest erinevusest tulenevalt võivad erineda nende väiketähelised kujud, nad käituvad erinevalt sortimisel jne. Erinev tähendus viib lahku ka nt miinuse ja sidekriipsu, ladina s-i sediiiga ja alakomaga, apostroofi, ühekordse lõpetava jutumärgi ja akuudi. Erilist tähelepanu nõuavad kombineeruvad diakriitilised märgid, mis sarnanevad oma funktsioonilt sammuta klahvidele klaviatuuril. Kombineeruv märk ei saa kunagi esineda tekstis iseseisvalt vaid alati liitmärgi koosseisus. Liitmärke, nende hulgas diakriitilisi märke sisaldavaid tähti, on võimalik kirja panna mitmel viisil. Selline on nt 'ü' (U+00FC), mis on taandata kahele liitmärgile: 'u' ja kombineeruvad täpid (U+0308). Kombineerimine võimaldab väga mitme diakriitilise märgiga tähtede kirjeldamist nt keeleteaduslikes tekstides. Vastavalt on defineeritud kolm Unicode'i rakendustaset (*implementation level*):

1. kombineeruvaid tähti ei kasutata (EVS 8-s defineeritud ESET1 on sellel rakendustasemel);
2. kombineerumine on lubatud kindlalt piiritletud osas;
3. kombineeruvaid märke võib piiranguteta kasutada.

UTF-8

Unicode lubab mitmeid esitusviise, millest tuntumad on UCS-2 e harilik tekst, kus iga märk pannakse kirja kahe järjestikuse baidiga ja UTF-8, mis kasutab muutuva pikkusega baididajasid.

UTF-8 iseloomulikud omadused ja kodeerimisalgoritm:

- Unicode'i märgid U+0000 kuni U+007F (ASCII e hariliku kooditabeli esimene pool) jäävad selle kodeeringu puhul muutmata kujule. See tähendab, et tekstid, mis sisaldavad ainult tähemärke a-z ja lihtsamaid kirjavahemärke ei muutu selle kodeeringuga üldse.
- Kõik muud Unicode'i märgid kodeeritakse kahe või rohkema järjestikuse baidi abil. Kõigi sellisesse rühma kuuluvate baitide kood on suurem kui 7F, mis tähendab, et ASCII tähemärgid ei ole kunagi kasutusel muutuva pikkusega kodeerimise puhul vaid tähistavad alati vaid iseennast.
- Esimene bait muutuva pikkusega rühmas on alati vahemikus C0 kuni FD ning sisaldab infot selle kohta, kui mitu talle järgnevat baiti rühma kuulub. Kõik alustavale järgnevad baidid on vahemikus 80 kuni BF. See võimaldab tekstis lihtsate vahenditega tuvastada, kus algavad ja lõpevad kodeeritud märgid.
- Sel viisil on võimalik kirja panna kõiki Unicode'i miljoni märki.
- Üks märk UTF-8 kodeeringus võib olla maksimaalselt 6 baiti pikk, kui piirduakse kuueteistbitiste märkidega, on see aga vaid kuni kolm baiti pikk.
- Baite FE ja FF ei esine kunagi.

Järgnev tabel illustreerib Unicode'i märkide kodeerimist UTF-8 baidijärjenditega.

U-00000000 - U-0000007F:	0xxxxxxx
U-00000080 - U-000007FF:	110xxxxx 10xxxxxx
U-00000800 - U-0000FFFF:	1110xxxx 10xxxxxx 10xxxxxx
U-00010000 - U-001FFFFF:	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
U-00200000 - U-03FFFFFF:	111110xx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx
U-04000000 - U-7FFFFFFF:	1111110x 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx

x-iga tähistatud positsioonid täidetakse kodeeritava märgi bittidega tagant ette, võimalikud tühjaksjäänud kohad järjendi esimeses baidis täidetakse 0-ga.

Näide:

Unicode'i autoriõiguse märk U+00A9 kodeeritakse kahe baidiga

```
110xxxxx 10xxxxxx
    10    101001 (A9 kahendkood on 10101001)
11000010 10101001 (ehk C2 A9)
```

Euromärgi U+20AC kodeerimiseks on vaja kolm baiti

```
1110xxxx 10xxxxxx 10xxxxxx
    0010    000010    101100 (20AC=0010000010101100)
11100010 10000010 10101100 (ehk E2 82 AC)
```

Kultuuriliselt sobib UTF-8 hästi nende keeltega, kus suurem osa tekstist koosneb ladina alus-tähestiku tähtedest ja diakriitiliste märkidega tähtede osakaal on väike, mitte eriti hästi nende mitte-ladina kirja kasutavate keeltega, mille puhul on võimalik kasutada ökonoomsemaid kodeerimisviise. Esimesse hulka kuuluvad peaaegu kõik ladina kirja kasutavad keeled, v.a ehk vietnami keel, teise hulka nt vene ja hiina keel. Kui ladina kirja puhul pikeneb tekst UTF-8 kodeeringus vaid paar protsenti, siis vene ja hiina keele puhul kasvab teksti maht kahekordseks.

Internatsionaliseerimine ja lokaliseerimine

Tarkvara arendamisel pööratakse üha rohkem tähelepanu **internatsionaliseerimisele** ja **lokaliseerimisele**. Internatsionaliseerimise (kasutatakse ka lühendust **i18n**, mis on tuletatud ingliskeelsest terminist *internationalization* = i, veel 18 tähemärki ja n) eesmärk on tarkvara tööalgoritmide ja kasutajaliidese suhteline eraldamine, mis võimaldaks kasutajaliidese elemente – eelkõige teksti menüüdes, nuppudel, teadetes jm – muuta ilma programmi põhiosa ümber tegemata. Järgmine samm on lokaliseerimine (ka **i10n** terminist *localization*) – internatsionaliseeritud programmi kohandamine konkreetse keele ja kultuuri konventsioonidele. EVS 8:2000 abistab eeskätt lokaliseerimisprotsessi juures, loetledes rea keele- ja kultuuri-elemente, mida võtta arvesse tarkvara kohaldamisel Eesti turule. EVS 8 ei ole mõeldud juhendama internatsionaliseerimist, mis on sisuliselt terve kompleks meetmeid, kuid eeldab, et vähemalt olulisemad neist on programmide loomisel tarvitusele võetud. Tarkvara juures eeldatakse, et:

- programmi teated ja kiirmenüüdes kasutatavad klahvikombinatsioonid asuvad eraldi ressursifailides;
- dialoogikastide ja menüüriba mõõtmed muutuvad dünaamiliselt vastavalt teksti pikkusele;
- ei piirduta ühe sisseehitatud kooditabeliga, sobiv lahendus oleks Unicode'i kasutamine;
- märkide jaotus tähtedeks ja erimärkideks, suur- ja väiketähtede asendused, numbrite ja kuupäeva vorming ning sortimine arvestavad lokaati;
- ortograafiamoodulid võimaldavad lisada eestikeelse õigekirjakontrolli.

Täiuslikult lokaliseeritud tarkvara arvestab kultuurikontekstiga märksa laiemalt: näidetes ei esine mitte John ja Mary vaid Jüri ja Mari; ikoonid ei kubise tänapühade või *halloweeni* butafooriast, mis on Eestis suhteliselt tagasihoidlikumad tähtpäevad; miilid teisendatakse

kilomeetriteks; välditakse võõrkeelsel sõnademängul, idioomidel ja reaaliatel põhinevaid lahendusi (ameerikapärane kirjakeel või nõobikujuline nupp) jne. Reeglina tuleks tarkvara kohaldamine Eesti turule läbi viia Eestis.

Lokaat

Lokaadiks nimetatakse lokaliseerimiseks vajalike andmete sobivalt korrastatud ja kindlate reeglite järgi kirja pandud hulka. Peaaegu tervikliku andmefailina on lokaat kasutusel UNIXil põhinevais operatsioonisüsteemides, MS Windows kasutab lokaliseerimisel põhimõtteliselt samu andmeid, kuid need on laialipillutatult väga mitmes kohas ning paljusid neist ei saa tavakasutaja muuta. Eriti põhjalik andmestik on esitatud EVS 8:2000s, kus lokaadi kirjeldamiseks kasutati ISO rahvusvahelist standardit ISO/IEC 14652. Lokaadi andmestikku kuuluvad tüüpiliselt:

Määratlus	Windows	FreeBSD
Kasutatav kooditabel	Sisse ehitatud MS CP 1257 e Balti kooditabel.	ISO/IEC 8859-15
Kooditabelis leiduvate märkide jaotus numbriteks, tähtedeks ja juhtkoodideks	sisse ehitatud	lokaadifailis
Suur- ja väiketähtede omavaheline vastavus	sisse ehitatud	lokaadifailis
Sortimisreeglid	sisse ehitatud ja seotud Balti kooditabeli kasutamisega	lokaadifailis
Klaviatuuri kirjeldus	sisse ehitatud	muudetav
Ajavöönd	Control panel, muudatusi saab teha registrifailis või spetsiaalprogrammidega	/etc/timezone/, muudatusi saab teha failis /usr/share/zoneinfo/Europe ja programmiga zic

Operatsioonisüsteemide kasvav tähtsus

Riist- ja tarkvara on jõudnud sellise keerukuse astmeni, kus programmeerijal on juba võimatu tunda detailselt riistvara tööpõhimõtteid ja käia sammu uuendustega. Kaasaegne tarkvara koosneb suhteliselt eraldiseisvatest moodulitest, kus iga moodul täidab täpselt piiritletud ülesandeid ning moodulite omavaheliseks suhtluseks kasutatakse standardseid liideseid. Suurem osa sedalaadi teenustest on kättesaadav operatsioonisüsteemi (Windows, FreeBSD, BeOS vm) vahendusel ja programmeerija on vabastatud kohustusest juhtida riistvara tööd. Kuid operatsioonisüsteemide roll ei piirdu pelgalt riistvara haldamisega. Üha enam tarkvaralisi abivahendeid, kasutajaliidese elemente ja üldteenuseid leiab koha operatsioonisüsteemi* poolt osutatavate teenuste hulgas ning suur osa neist seondub otseselt lokaliseerimisega. Nt vajutus klahvile 'Ö' ja tähemärgi 'ö' ilmumine ekraanile käib läbi pika ahela:

- Klaviatuuridraiver registreerib vajutuse konkreetsele klahvile
- Operatsioonisüsteemi edasine tegevus sõltub suuresti lokaliseerimisasetustest. Kontrollida tuleb klaviatuuriseadeid, klaviatuuri registreid, sammuta klahve jpm. Lõpptulemusena saab operatsioonisüsteem aru, et kasutaja soovis saada tähte 'ö' (nt vene klaviatuuril oleks seda tõlgendatud 'ж'-na)
- Operatsioonisüsteem saadab aktiivsele aknale teate "vajutati klaviatuurisõrmist ö"
- Aktiivne programm otsustab, kuidas klahvivajutusele reageerida. Kui hetkel asub kursor tekstikastis, tuleb tähemärk sisestada teksti kursori asukohale. Vajutus funktsiooniklahvile kutsunuks tõenäoliselt esile muu tegevuse, näiteks abitekstide esitamise või programmi töö lõpetamise

* Siinjuures tõlgendame operatsioonisüsteemi juba veidi laiemalt tarkvarakeskkonna tähenduses. Vaatamata sellele, et graafiline kasutajaliides või elektronposti saatmiseks sobiv programm pole rangelt võetuna operatsioonisüsteemi osad, peetakse selliseid teenuseid sedavõrd enesestmõistetavaks, et kõiki tarkvarakomponente koos vaadelda omaette tervikliku keskkonnana.

- Operatsioonisüsteem leiab vajaliku fondi, sellest õige glüüfi ja teisendab selle sobiva suurusega mustvalgeks pildiks
- Pilt kantakse aktiivsele aknale nõutud kohta ning kuvatakse monitoril

Rakenduse roll kogu ahelas piirdus otsustusega, mida ta soovib ette võtta konkreetse klahvi-vajutusega, kogu ülejäänud töö tegi ära operatsioonisüsteem. Lokaliseerimise kontekstis tähendab selline tööjaotus, et operatsioonisüsteemi arendamisele tuleb pöörata erilist tähelepanu. Tavaprogramm võib lokaliseerimisel piirduda menüü ja paari teate eestindamisega, jättes operatsioonisüsteemi vastutama märksa olulisemate valdkondade eest. Oletagem, et meil oli tegemist tekstitöötlusprogrammiga ja kasutaja oli sisestamas '...vajab müügi-assistenti. tö...'. Programm võib ja peakski näitlikult kasutama reegleid nagu:

- Kui *lause_alguses_kasutatakse_suurtähte* (eesti keel) ja *on_täht* (eesti keel, ö) ja *on_täht* (eesti keel, t) ja eelnes *tühik* ja sellele eelnes *lauselõpumärk* (eesti keel), siis *teisenda_suurtäheks* (eesti keel, t).
- Kui ei ole *lubatud_sõna* (eesti keel, müügiassistenti) siis *anna_kasutajale_märku*.
- Kui ei *mahu_reale* (müügiassistenti) siis *poolita* (eesti keel, müügiassistenti).

Kursiivis toodud funktsioonid pöörduvad otse operatsioonisüsteemi poole ja programm ei pea iseseisvalt otsustama, kas keel, milles teksti kirjutatakse, kasutab suurtähti (enamus maailma keeltest ei kasuta!), kas 'ö' ja 't' on eesti keeles tähed, millised võivad olla lauselõpumärgid (vrd hispaania ¡ ja ¿ hüüd- ja küsilause lõpus), milline suurtäht vastab t-le (vrd i ja Í paar türgi keeles) ning kas tähejärjendid on eestikeelsed sõnad ja kuidas neid poolitada.

ESET1

Peatükis kasutatud terminid ja mõistekäsitlused

Ladina tähestik on järjestatud tähejada A B C D E F G H I J K L M N O P Q R S T U V W X Y Z (**ladina alustähestik**, alustähed) koos eri keeltes kasutatavate täheteisenditega. Iga suurtähe juurde kuulub ka vastav väiketäht.

Täheteisendid on moodustatud alustähtedest **diakriitikute** ehk diakriitiliste märkide, s.o alustähe alla, peale, sisse, kõrvale pandud või alustähest läbi tõmmatud lisamärgikeste abil, osalt ka alustähe kuju muudatuste teel. Graafiliselt kokkukujundatud tähti, nt Æ, Œ, ß nimetatakse **ligatuurideks**.

Keeletähestik ladina alustähestiku baasil on ühe kirjakeele õigekirja, sõnastike, teatmeteoste, dokumentide järjestuse jne jaoks kasutatav valik ladina tähestikust järjestustähestikuna. Keeletähestiku tähtedel on oma järjestuskohad, muud tähed (täheteisendid) järjestatakse keeletähestiku tähtedega kokku.

Eesti tähestik on keeletähestik eesti keele ja Eesti riigi asjaajamine tarbeks: A B C D E F G H I J K L M N O P Q R S Š Z Ž T U V W Õ Ä Ö Ü X Y. Konkreetsetes järjestustes võidakse vajadust mööda kasutada V ja W kokkujärjestust. Eesti tähestikku nimetatakse *eesti-ladina tähestikuks*, kui tahetakse toonitada, et ta kasutab ladina kirja, s.t et tegemist pole kirillitsa vm kirjasüsteemiga.

Eesti tähestikus on kolme liiki tähti: 1) tähti C, Q, W, X, Y koos kõigi eesti tähestikus eraldi järjestuskohta mitteomavate ladina tähtedega nimetatakse võõrtähtedeks ja neid kasutatakse ainult võõrnimede ning võõrkeelsete tsitaatide kirjutamisel; 2) tähti F, Š, Z, Ž nimetatakse võõrsõnatähtedeks, sest nad on võõrsõnade tunnuseid; 3) ülejäänud on omatähed*.

Kiri e kirjasüsteem on ühisest alustähestikust lähtuvate KEELETÄHESTIKE kogusumma. Ühist kirjasüsteemi kasutatavatele keeltele on tavaliselt omane kasutada teatud juhtudel, nt nimekirjutuses ka teiste samakirjaliste keelte tähestike tähti. **Ladina kiri** on kogum, mis koosneb ladina alustähestikust ja sellel baseeruvatest keeletähestikest.

Latinisatsioon on mitteladinatähelise keele sõnade või nimede ülekandmine kindlate reeglite alusel ladina kirja, kus lähtekeele kirjamärkidele on pandud vastama tulekkeele tähestiku tähed. Eristatakse täpset **transliteratsiooni**, kus tähtede vastavus on üksühene ja nime lähtekirjapilt eksistusteta taastatav, ning vähem täpset **transkriptsiooni**, kus kasutatakse üksnes tulekkeele keeletähestiku tähti ja püütakse nendega edasi anda lähtekeele hääldust. Nt vene *Горький* on translitereerituna *Gor'kij*, vene *Горьку* aga *Gorki*; mõlemad nimed eesti transkriptsioonis on ühtviisi *Gorki*. **Omaladina** on vastavas riigis eneses kehtestatud latinisatsioon.

Tekstikodeerimisstandardid on välja töötatud rahvusvahelise andmevahetuse hõlbustamiseks peamiselt Rahvusvahelise Standardiorganisatsiooni (ISO) poolt. Seni väga levinud 8bitise jada standardid, koondnimetusega **ISO/IEC 8859**, võimaldavad korraka fikseerida ja üheselt edasi anda umbes 230 eri märki, mis aga pole tänapäeval piisav. **ISO/IEC 10646-2** ja sellega üksüheses vastavuses olev **Unicode**'i konsortsiumi standard on mõeldud 16bitise andmevahetuse jaoks ning see võimaldab anda unikaalse koodi 65000 eri märgile. Standard käsitleb kõigi tähestike ja kirjasüsteemide märke ning sümboleid võrdväärselt ning lubab neid kirja panna mis tahes koosluses. Praegu on kood antud u 50000 märgile, sealhulgas on kõik maailma tähtsamate keelte tähestikud, k.a hieroglüüfid. Tulevikus on ISO 10646/Unicode kõige arvestatavam tekstikodeerimisstandard, millele orienteerub enamik tarkvararakendusi.

* Tulenevalt sellest liigitusest on eesti tähestikuks vahel nimetatud ka sellist valikut, kus puuduvad võõrtähed (eesti õigekirjatähestik), või seda, kust puuduvad ka võõrsõnatähed F Š Z Ž (eesti omasõnatähestik).

Tähtede liigid. Diakriitikute tabel

Ladina tähestikus on:

alustähestiku tähed: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z;

diakriitikutega tähed: Á À Â Ã Ä Å Æ Ç Ć Ĉ Ċ Đ Ę È Ê Ë Ē Ę jne;

ligatuurid e liittähed: IJ Œ;

muud täheteisendid: Æ Þ Œ Š Ž.

Tähtedele lisatavad **diakriitikud** on allpool tabelis.

Diakriitik	Nimetus*	Kood**	Näiteid
´	akuut (<i>acute accent</i>)	0301	á é í ó ú ý
`	graavis (<i>grave accent</i>)	0300	à è ì ò ù
˘	kaar (<i>breve</i>)	0306	ă ă ǫ ŭ
^	tsirkumfleks ("katus", <i>circumflex</i>)	0302	â ê î ô û w
ˇ	haak ("linnuk", <i>caron</i>)	030C	š č ě š ž
◌̊	ülasõõr (<i>ring above</i>)	030A	å ŭ
◌̈	täpid e treema (<i>diaeresis</i>)	0308	ä ë ï ö ü ÿ
ˆ	topeltakuut (<i>double acute accent</i>)	030B	ő ú
˜	tilde (<i>tilde</i>)	0303	ã ñ õ
◌̇	ülapunkt (<i>dot above</i>)	0307	č é ž
◌̄	ülakriips e makron (<i>macron</i>)	0304	ā ē ī ō ū
◌̆	keskpunkt (<i>middle dot</i>)	-	ł
/	kaldkriips (<i>slash</i>)	0338	ø
		0337	ł
˘	sedii (<i>cedilla</i>)	0327	ç ț ș
◌̣	pöörsedii (<i>ogonek</i>)	0328	ą ę ĭ
◌̈	alapunkt (<i>dot below</i>)	0323	ạ ì ı ỵ
◌̅	alakriips (<i>line below</i>)	0331	h n r z
◌̆	konks (<i>hook</i>)	0309	á é í ó ŭ
◌̇	sarv (<i>horn</i>)	031B	σ υ
◌̈	alatäpid (<i>diaeresis below</i>)	0324	× Û
◌̇	alakoma (<i>comma below</i>)	0326	ț

* jutumärkides on levinud, ent ebatäpsed argivasted. Kaldkirjas on ingliskeelsed vasted.

** märgitud on kood ISO10646/Unicode'i kombineerivate diakriitikute tabelis. Seda koodi võib vaja minna juhul, kui vastava tähe lihtkood standardis puudub. Niisugusel juhul kodeeritakse täht kahe koodi abil: alustähe ja kombineeruva diakriitiku koodiga. Tabelis on üksnes valimik diakriitikutest, mis võivad nimedes esineda.

ESET1 tarvitusjuhiseid

Valikusse kuuluvate tähtede kasutamisevõimalusi:

1. Üldjuhul on sobivaim moodus küllalt laia tähevalikuga tekstide sisestamiseks Unicode'i tekstieditori kasutamine. Yudit (Linux jt) ja Ultraedit (Windows) on selleks ülesandeks igati sobivad. Windowsi programmid kalduvad rohkem kasutama operatsioonisüsteemi võimalusi, Yudit on iseseisvam ja laiendatav.

2. Windows 95 ja 98: Kasutatakse operatsioonisüsteemi standardfonte (*Times New Roman, Courier New, Arial*), kus mõne tähe jaoks tuleb kasutada nende erikomplekte laienditega *Baltic, CE* (=Central European) või *Tur* (=Turkic), nt *Times New Roman Baltic, Times New Roman CE* või *Times New Roman Tur*. Niisugune moodus sobib kõigi kaasaegsete Windowsi tekstitöötlus- ja küljendusprogrammide *Word, Excel, Corel Draw* jts puhul, mis lubavad tekstis fonte vahetada. Füüsiliselt neid fonte arvutis pole, kataloogist `/windows/fonts/` leiate ainult fondifailid `times.ttf, timesi.ttf` jt. Nõutud fondi koostab operatsioonisüsteem vastavalt sellele, millised keeleteod arvutil installeeritud on. (Installeerimiseks avage arvutil *Control Panel*, valige ikoon *Add/Remove Programs*, selle alt *Windows Setup* ja järgnevalt *Multilanguage Support*, edasi märkige vastavalt vajadusele komponendid *Baltic Language Support, Central European Language Support* ja *Turkish Language Support* ning installeerige need. Installeerimiseks on vaja omada Windowsi installeerimise tarkvara, mis enamasti on CD-l.)

Ladina alustähestikule lisanduvate tähtede koodid on järgmised:

Täht	Lühinimetus	Kood (hoida all klahvi AltGr ja valida numbrisõrmistikul ja font (laiend))*
Á / á	A akuudiga	0193 / 0225
À / à	A graavisega	0192 / 0224
Ā / ā	A kaarega	0195 CE / 0227 CE
Â / â	A tsirkumfleksiga	0194 W / 0226 W
Å / å	A ülasõõriga	0197 / 0229
Ă / ă	A täppidega	0196 / 0228
Ȧ / ȧ	A tildega	0195 W / 0227 W
Ą / ą	A pöördsemiiga	0192 B / 0224 B
Ȧ / ȧ	A ülakriipsuga	0194 B / 0226 B
Æ / æ	AE	0198 W (0175 B) / 0230 W (0191 B)
Ć / ć	C akuudiga	0195 B (0198 CE) / 0227 B (0230 CE)
Č / č	C haagiga	0200 B / 0232 B
Ç / ç	C semiiga	0199 W / 0231 W
Ð / ð	ETH	0208 W / 0240 W
Ǻ / ǻ	D haagiga	0207 CE / 0239 CE
Đ / đ	D läbiva kriipsuga	0208 CE / 0240 CE
É / é	E akuudiga	0201 / 0233
È / è	E graavisega	0200 W / 0232 W
Ê / ê	E tsirkumfleksiga	0202 W / 0234 W
Ë / ë	E haagiga	0204 CE / 0236 CE
Ě / ě	E täppidega	0203 W / 0235 W
Ě / ě	E ülapiiktiga	0203 B / 0235 B
Ę / ę	E pöördsemiiga	0198 B (0202 CE) / 0230 B (0234 CE)
Ē / ē	E ülakriipsuga	0199 B / 0231 B
Ĝ / ĝ	G kaarega	0208 Tur / 0240 Tur

* algul suurtähe kood, kaldkriipsu järel väiketähe kood. Laiendid: *B* = Baltic, *CE* = Central European, *Tur* = Turkic, *W* = Western (*Latin 1* valik, fondimenüüdes harilikult ilma laiendita). Kui laiend puudub, siis saab tähe kätte nii *Western*'i kui ka *Baltic*'u fondist, mis mõlemad on Eestis enim kasutuses.

Ġ / ġ	G sediiiga	0204 B / 0236 B
Ī	suur I ülakupunktiga	0221 Tur
ı	väike punktita I	0253 Tur
Í / í	I akuudiga	0205 W / 0237 W
Ì / ì	I graavisega	0204 W / 0236 W
Î / î	I tsirkumfleksiga	0206 W / 0238 W
Ï / ï	I täppidega	0207 W / 0239 W
Ĭ / ĭ	I pöördseidiiga	0193 B / 0225 B
Ī / ī	I ülakriipsuga	0206 B / 0238 B
Ķ / ķ	K sediiiga	0205 B / 0237 B
Ļ / ļ	L akuudiga	0197 CE / 0229 CE
Ļ / Ļ	L haagiga	0188 CE / 0190 CE
Ļ / Ļ	L läbiva kriipsuga	0217 B (0163 CE) / 0249 B (0179 CE)
Ļ / ļ	L sediiiga	0207 B / 0239 B
Ņ / ņ	N akuudiga	0209 B / 0241 B
Ņ / ņ	N haagiga	0210 CE / 0242 CE
Ņ / ņ	N tildega	0209 W / 0241 W
Ņ / ņ	N sediiiga	0210 B / 0242 B
Ō / ó	O akuudiga	0211 / 0243
Ò / ò	O graavisega	0210 W / 0242 W
Ô / ô	O tsirkumfleksiga	0212 W / 0244 W
Ö / ö	O täppidega	0214 / 0246
Ö / ö	O topeltakuudiga	0213 CE / 0245 CE
Ö / ö	O tildega	0213 / 0245
Ø / ø	O läbiva kaldkriipsuga	0216 W (0168 B) / 0248 W (0184 B)
Ö / ö	O ülakriipsuga	0212 B / 0244 B
Œ / œ	ligatuur OE	0140 W / 0156 W
Ŕ / ř	R akuudiga	0192 CE / 0224 CE
Ř / ř	R haagiga	0216 CE / 0248 CE
Ř / ř	R sediiiga	0170 B / 0186 B
Š / š	S akuudiga	0218 B (0140 CE) / 0250 B (0156 CE)
Š / š	S haagiga	0138 W (0208 B) / 0154 W (0240 B)
Ş / ş	S sediiiga	0170 CE (0222 Tur) / 0186 CE (0254 Tur)
ß	väike SZ	0223
Ť / ť	T haagiga	0141 CE / 0157 CE
Ť / ť	T sediiiga	0222 CE / 0254 CE
Ů / ú	U haagiga	0218 W / 0250 W
Ù / ù	U graavisega	0217 W / 0249 W
Û / û	U tsirkumfleksiga	0219 W / 0251 W
Ů / ů	U ülasõoriga	0217 CE / 0249 CE
Û / ü	U täppidega	0220 / 0252
Ů / ů	U topeltakuudiga	0219 CE / 0251 CE
Ů / ů	U pöördseidiiga	0216 B / 0248 B
Ů / ů	U ülakriipsuga	0219 B / 0251 B
Ý / ý	Y akuudiga	0221 W / 0253 W
Ÿ / ÿ	Y täppidega	0159 W / 0255 W
Ž / ž	Z akuudiga	0202 B (0143 CE) / 0234 B (0159 CE)
Ž / ž	Z haagiga	0142 W (0222 B) / 0158 W (0254 B)
Ž / ž	Z ülakupunktiga	0221 B (0175 CE) / 0253 B (0191 CE)
Þ / þ	THORN	0222 W / 0254 W

3. Windows NT ja 2000: Fondilaidendite kasutamine kui üleminekuetapp ei sobi enam Unicode'i-põhisele operatsioonisüsteemile. Uuem Windows eelistab kasutaja-keelsele klaviatuurile mitte mahtunud üksikuid märke sisestada programmiga Character Map, MS Wordis aga menüüst Insert ja Symbol. Päril Unicode'i-põhist sisestusviisi Windows vaikimisi ei toeta, st isegi kui te teate vajaliku märgi koodi kuueteistkümnend- või kümnendsüsteemis, pole võimalik seda numbrisõrmistikult koodina sisestada. Teatud programmides on siiski võimalik

sisestada Unicode'i märke numbrisõrmistikult kümnendsüsteemis või tekstis kuuteistkümnendarvuna, mille järel tuleb vajutada ALT+x.

4. Teksti sisestus ESET1 rakendustes: Lineaarsesse, fonte mittesisaldavasse teksti (nt andmebaasifailides) saab nimetatud valimikku kuuluvaid tähti lülitada üksnes Unicode'i koodide abil. Lisaks laialtlevinud kodeerimis- ja edastusstandarditele nagu nt UTF-8, defineerib EVS 8:2000 järgneva kodeerimisviisi: andmevahetusfailides pannakse eritähed kirja Unicode'i koodide abil neljakohaliste kuuteistkümnendarvudena, pannes nende ümber noolsulud: <0102>, <00C2> jne. Erandina võib eesti tähestikku kuuluvaid tähti š, ž, õ, ä, ö, ü sisestada kaheksabitiste koodidena vastavalt aluskooditabelile ISO 8859-15. Kasutatav tarkvara peab võimaldama vajaduse korral Unicode'i koodidega antud tähti printida ja nimedes päringuid sooritada. Märkigem, et see viis on väga kitsalt seotud ESET1 kasutusfääriga – isiku- ja kohanimede kodeerimine (kus mitte kunagi ei esine eraldajana kasutatud märke < ja >) ning selle laiema kasutamine on sel põhjusel võimatu.

Laiema tähevalikuga kaasnevaid eksimisvõimalusi

- **Tühik ja sisetühik:** koodiga <0020> on tühik, <00A0> sisetühik, mille funktsioon on mitte lubada selle tühiku kohal rida vahetada. Sisetühiku tüüpiline kasutus on numbrite grupeerimisel (nt 22 500) või initsiaalide ja perekonnanime vahel (A. Obukakk).
- **Hüüumärk, küsimärk:** hispaania keeles kasutatakse pöördhüüumarki (¡ <00A1>) ja pöördküsimarki (¿ <00BF>).
- **Jutumärgid:** Unicode'i järgi tähistab märk " tolli, jutumärkidena kasutatakse muid märke: „, “, ”, «, » jt. Ungari keeles kasutatav topeltakuut (ő) ja minutimärk sarnanevad tollimärgile ainult kuju poolest.
- **Apostroof:** ei ole klaviatuuril sammuta märk. Sammuta diakriitikud on graavis (`) ja akuut (´).
- **Koma:** võib minna segamini ühekordse alustava jutumärgiga (, <201A>).
- **Sidekriips ja miinus:** tihti kasutatakse segiläbi tähenduses. Terve grupp märke on sidekriips-miinus (- <0045>), poolituskriips (- <2010>), enn-kriips (- <2013>), Windowsis AltGr+0150) ja emm-kriips (— <2014>). Unicode pakub veel ülikorrektseks vormistuseks miinusele oma koodi (<2212>).
- **Valuutamärk:** ♂ <00A4> asendati paljudes kooditabelites euro märgiga, valuutamärgi kasutusest oleks parem üldse hoiduda.
- **Paragrahvimärk (§) ja USA paragrahvimärk (§):** USAs kasutatakse juriidilistes tekstides märki ¶, Euroopas aga §.
- **I täpiga ja ilma:** Unicode lähtus algselt Euroopa keelte tavast kirjutada suur I ilma täpita, väike i aga täpiga ja need kaks moodustavad tähepaari. Aserbaidžaaani ja türgi keeles moodustavad suur- ja väiketähe paarid hoopis Í <0130> ja i ning I ja ı <0131>. Suur I on kasutusel rooma number ühena.
- **ß:** Saksa esszett on ainult väiketäht. Suurtähena kirjutatakse see harilikult kujul SS. Kreeka tähe beeta kood on aga <03B2>.
- **ETH ja kriipsuga d:** ESET1 tähevalikus on kaks tähte, mis suurtähena on sarnased: islandi ETH (Ð <00D0>) ja suur D kriipsuga (Ɔ <0110>), mida kasutatakse horvaadi, sloveeni ja saami keeles. Sama kujuga on veel ka Aafrika suur D (Ɔ <0189>), mis ei kuulu Eesti tähevalikusse. Erinevad nende tähtede väiketähekujud: väike ETH (ð <00F0>) ja väike D kriipsuga (đ <0111>). Suurtähtede kodeerimisel tuleb seega olla eriti tähelepanelik ja lähtuda keelest, milles vastav täht on kirja pandud.
- **O läbiva kaldkriipsuga:** Ehkki Ø meenutab kujult tühja hulga märki (kood <2205>), on selle taani, norra ja fääri keele tähe kasutamine matemaatilise sümbolina kohatu.
- **Haagiga tähtede väiketähelised variandid:** Kesk-Euroopa keelte ortograafias kasutatakse kõrgetel väiketähtedel haagi asemel tihti ülakoma. Vrd Đ aga đ', Ě aga ě'. L haagiga on ka suurtähena Ě (ja väiketähena ě).
- **Väike g sediiiga:** läti g sediiiga (täpsemalt alakomaga) kirjutatakse kujul ģ.

- **Haagi ja kaare eristamine:** väikeses või ebaselgelt trükitud kirjas on tihti raske eristada tähe kohal olevat haaki (ˇ) ja kaart (˘). Kaar esineb märksa harvemini, arvestatavalt vaid A peal (Ăă) rumeenia keeles ja G peal (Ğğ) türki ja aserbaidžaani keeles. Diakriitik haak esineb aga paljudes keeltes (sh eesti š ja ž), harilikult konsonantide peal (ČĎǦŇŘŠŽ).
- **Alakoma ja sedii eristamine:** ehkki nende kahe diakriitiku samastamine ei tekita olulisi vigu ja isegi paljud tähenimetused Unicode'is tarvitavad alakoma asemel sediid, on korrektsuse huvides siiski oluline neid diakriitikuid eristada. Alakoma kasutavad rumeenia, läti ja liivi keel, sediid prantsuse ja türki keel. Üldjuhul on C alati sediiiga (Ç), Ğ, Ķ, Ļ, Ņ ja Ŗ alakomaga. Ş-i ja Ţ puhul tuleb lähtuda kasutuskeelest. Rumeenia keele korrektne kasutus eeldab alakoma, kuid sel juhul ei võimalda ESET1 märgivalik enam rumeenia keelt kodeerida.

Jutumärkide tarvis

Lisaks EVS 8:2000s eelistusjärjekorras toodud jutumärgipaaridele (<,>...<> (e 99...66) <<>...<>> ja <">...<">) soosib praegune keeletarvis paari <,>...<"> (e 99...99), mille võiks eelistuselt paigutada teisele kohale. Vaatamata sellele, et ametlikes ortograafiajuhendis pole jutumärkide tarvist reglementeeritud, on seal kasutusel nimelt see paar. Igal juhul tuleks aga vältida praegu üldkasutatavat ameerikapärast jutumärkide tarvist (<">...<"> e 66...99), avav jutumärk peab asuma all, mitte üleval.

Soovitus MS Wordi kasutamisel: Korrektsete jutumärkide kasutamiseks on mitu võimalust:

1. Sisestada neid klaviatuurilt koodidena („ kood on ALT+0132, ” kood on ALT+0148, “ kood on ALT+0147).
2. Teha kasutuseks ümber kiirklahvikombinatsioonid (avav jutumärk CTRL+' ja ", lõpetav jutumärk CTRL+' ja ").

Sortimine

Sortimine on toiming, mille üldine eesmärk on teha kasutajale informatsiooni leidmine hõlpsamaks, esitades selle struktureeritud viisil, nt temaatiliselt või geograafiliselt lahterdatult, ulatusliku registriga jne. **Tähestikuline järjestamine** on levinumaid sortimise liike ja on kombineeritud sageli mõne eespool mainituga.

Nagu standardi eelmine versioon, annab ka EVS 8:2000 juhiseid eestikeelsete stringide tähestikulisse järjekorda seadmise kohta, kuid sortimisalgoritmi esitus standardi uustöötluses EVS 8:2000 erineb standardi esimeses versioonis toodust suuresti. Standardi versioonis EVS 8:1993 esitatud sortimisalgoritmi oli põhimõtteliselt võimalik rakendada ainult piiratud tähevaliku ulatuses ning see ei sobi Unicode'i kasutuselevõtuks. Standardi uus versioon tugineb sortimisel rahvusvahelisele standardile ISO/IEC 14651 'International String Ordering' ja sortimisreeglid esitatakse lokaadi osana.

Igas kirjakeeles on tavaliselt kindlaks kujunenud standardne tähtede järjestus, mida nimetataksegi selle keele **tähestikuks**. Keeletähestikus on üldjuhul igal tähel oma järjestuskoht. Eri keeltes võib ühe ja sama tähe koht olla erinev, näiteks eesti tähestikus järjestatakse *ä* erinevalt kui saksa tähestikus. Kuigi üldjuhul kasutavad ladinatähelised keeled ühesugust alustähestiku järjestust, on ka siin erandeid: eesti tähestikus paigutub Z ja ühes sellega Ž S-i ja Š järele, mitte tähestiku lõppu nagu enamiku teiste keelte tähestikes.

Eesti tähestikjärjestus on selline:

A B C D E F G H I J K L M N O P Q R S Š Ž T U V W Õ Ä Ö Ü X Y

Igal eesti tähestiku tähel on oma järjestuskoht, mõõndusena võib nt eesti perekonnanimesid sisaldavates loendites järjestada V ja W-ga nimed kokku.

Võõrtähed, millel eesti tähestikus omaette järjestuskohta ei ole, järjestatakse vastavalt üldisele tavale, millest lähtus ka Emakeele Seltsi keeleteoimkonna soovitus 24.09.1998, järgmiselt.

1. Diakriitikuga täheteisendid järjestatakse kokku oma diakriitikuta alustähega, nt nii Á, À, Â, Å kui ka ã järjestatakse kokku A-ga.
2. Ligatuurid, s.o kokkukujundatud tähed, järjestatakse nii nagu oleks tegemist kahe eraldi tähega, nt Œ järjestatakse nii nagu OE. Saksa ß (*eszett*) võrdsustatakse järjestamisel ss-ga.
3. Üksikud tähemodifikatsioonid järjestatakse kokku oma analoogtähega:
 - a) aserbaidžaani Ə — nii nagu E (konkreetses rakenduses võib ka järjestada kokku Ä-ga, mis käibib tema asendajana);
 - b) islandi Ð/ð — nagu D;
 - c) islandi Þ/þ — vastavalt üldisele kokkuleppele järjestatakse kõige lõppu*;
 - d) saami Ń — nii nagu N;
 - e) saami Ț, Ț — nii nagu Z.

Esimeses lähenduses võib tunduda, et eelpooltoodud reeglid on sortimiseks piisavad, kuid lähemal vaatlusel kerkib terve rida küsimusi. Tähestikuliselt järjestatakse ju mitte üksiktähti, vaid **stringe** e märgijärjendeid. Võrreldes omavahel kaht sarnastest märkidest koosnevat stringi, tuleb vastused leida sellistele probleemidele nagu:

- kui üks string on kirjutatud suurtähtedega, teine aga väiketähtedega (ABC ja abc), siis kumb neist peab olema eespool;

* levinud on ka selline islandi tähtede eth ja thorn järjestamine, kus nad arvestatakse tähetühendiks: Ð = DH, Þ = TH. Need ühendid on olnud käibel vastavate tähtede asendajatena.

- kui võrrelda tuleb kahte erinevate diakriitikutega kuid sama alustähega tähte (Éva ja Èva), siis kumb neist on eespool;
- mis on olulisem, kas suur- ja väiketähe erinevus või diakriitiku tarvitus (Éva ja eva);
- kui stringides esineb muid märke peale tähtede, siis milline on nende järjestus (Tallinn-Tartu ja Tallinn—Tartu).

Mõned keeletähestikud arvestavad järjestamisel eraldi ka täheühendeid: kõmri (uelsi) keeles satuvad *ng*-algulised sõnad *g*-alguliste sõnade järele, sest ühendil *ng* on omaette järjestuskoht. Samas võib mõni keeletähestik sisaldada tähti, millele eraldi järjestuskohta pole ette nähtud, nt saksa keeles järjestatakse *ä, ö, ü* kokku vastavate alustähtedega *a, o, u*. Keeletähestikku kuulumatud tähed (täheteisendid) järjestatakse enamasti kokku vastavate alustähtedega. Ladina kirja kasutatavate maade sortimistavad ei pruugi aga üldse rakendada teistes kirjasüsteemides. Nii nt sorditakse korea keelt häälduse alusel, Hiina ja Jaapani hieroglüüfikirja põhiliselt hieroglüüfide koostisesse kuuluvate joonte arvu järgi, edasine jagamine toimub hieroglüüfi komponentide kaupa jne. Laiema märgivaliku kasutamisel tuleb omavahel võrrelda juba ka erinevates kirjasüsteemides stringe. Eestis on see probleem esinenud kirillitsa ja ladina kirja segajärjestusel nt teadustööde viidatud materjalide või telefoniraamatu tegevusvaldkondade loendite sortimisel.

EVS 8:1993 võttis aluseks IBM-i käsiraamatu "National Language Support. Reference Manual, volume 2", kus sortimine põhines astete e kaalude süsteemil (vt standard EVS 8:1993 „Infotehnoloogia reeglid eesti kultuuri ja keele keskkonnas“ lk. 31–38). EVS 8:2000 kasutab küll sortimisreeglite kirjapanekuks teistsugust kuju, kuid sisuline lähenemine on mõlemal juhul sama. Sortimine on alati mitmeastmeline protsess ja mida rohkem astmeid arvesse võtta, seda täpsem on tulemus. Esimesel astmel järjestatakse tavaliselt ainult põhitähti, järgneval astmel võetakse arvesse tähtedele lisatud diakriitikud, siis erinevused väike ja suurtähtede vahel. Ladina kirja puhul on kolmeastmeline protsess enamasti piisav, kuid vajaduse korral võib seda veelgi täpsustada.

EVS 8:2000 mainib eesti sortimisreegleid kolmes kohas: kirjeldava teksti kujul peatüki 6 alajaotises 1. Sortimine, Lisa B alajaotis LC_COLLATE esitab reeglitiku standardile ISO/IEC 14651 sobival kujul, lisa C aga ainult aluskooditabeli piires. Kirjelduses on esile toodud ainult need reeglid, mille poolest eesti tähestiku alusel sortimine erineb ISO/IEC 14651 vaikimisi sortimisjärjekorrast (š tuleb sortimisel peale *s-i, z* ja *ž* koos kõigi teiste *z* modifikatsioonidega nende järel, *õ ä ö ü* tähestiku lõpuosas *w* järel). ISO/IEC 14651 eelistabki konkreetse keele puhul **delta** kirjeldust, mis sisaldab ainult **modifitseerimisreegleid** vaikimisi paikapandud sortimistabeli kohaldamiseks konkreetse keele või ülesande vajadustele. Ühelt poolt võimaldab selline esitusviis kogu Eesti andmestiku kirja panna ülimalt lakooniliselt ja rakendaja ei pea tegelema talle võõraste keelte tähestike õppimisega või otsustustega, kas üks või teine tähemodifikatsioon on juba uus esimese astme täht või endiselt taandatav alustähele, teiselt poolt aga jäävad paljud üksikasjad varju. Näiteks ei saa Eesti delta põhjal teada, kas string kirillitsas on enne või pärast kreekakeelset stringi.

Kirjeldavas tekstis on siiski mainitud kõige olulisemat lisareeglit, millele tähelepanu pöörata. See on: “muu võrdsuse korral paigutatakse väiketähed enne suurtähte”. Et see on nimetatud standardis vaikimisi realiseeritud, pole seda formaalses kirjelduses vaja esitada.

Suurtähtede sortimiseks enne väiketähti:

Formaalsesse kirjeldusse tuleb lisada modifitseerimisreegel:

```
reorder after <CIRCLECAP>
<MIN>
<WIDE>
<COMPAT>
<FONT>
<CIRCLE>
reorder-end
```


Muud üldreeglid, mis rahvusvahelises standardis vaikumisi rakenduvad on järgmised:

Üldjuhul on järjestus:

1. lisamärgid
2. numbrid
3. tähemärgid

Tuntumate (ESET1-s esinevate) diakriitiliste märkide omavaheline järjestus on:

- | | |
|----------------------|------|
| 1. akuut | á |
| 2. graavis | à |
| 3. kaar | ã |
| 4. katus | â |
| 5. haak | ă, ĉ |
| 6. ring | å |
| 7. täpid | ä |
| 8. topeltakuut | õ |
| 9. tilde | ã |
| 10. ülakupunkt | è |
| 11. sedii | ç |
| 12. pöörsedii | ą |
| 13. ülakriips | ā |
| 14. läbiv kriips | đ, ł |
| 15. läbiv kaldkriips | ø |

See tähendab, et muu võrdsuse korral (tavapärasel sortimisalgoritmil esimese astme võrdsuse korral, kus võrreldakse alustähti) sorditakse akuudiga a enne graavisega a-d või haagiga c enne c-d sediiiga. Põhjalikuma diakriitiliste märkide loendi leiate viidatud rahvusvahelise standardi juurest.

Kirjasüsteemide omavaheline järjekord on vaikumisi:

1. ladina
2. kreeka
3. kirillitsa
4. gruusia
5. armeenia
6. heebrea
7. araabia
8. devanaagari
9. bengali
10. gurmukhi
11. gudžarati
12. oria
13. tamili
14. telugu
15. kannada
16. malajalami
17. tai
18. lao
19. tiibeti
20. hanguli
21. hiragana
22. katakana
23. hieroglüüfkiri

Ligatuure käsitletakse kahe tähe ühendina ja vastavalt ka sorditakse, nt Oasis < Œrsted (=Œrsted) < Ogilvy, Æ-d loetakse aga eritäheks ja sorditakse Alfred < Ælfred (≠ AElfred) < Berta.

Oluline on ka vahet teha sõnakaupa järjestamisel ja tähtede kaupa järjestamisel. Sõnakaupa järjestamine võtab arvesse sõnadevahelise tühiku, tähtede kaupa järjestamisel tühikuid ei arvestata. Eesti Entsüklopeedia ja ilmunud sõnastikud kasutavad tähtede kaupa järjestamist. Näide EE-st:

Sõnakaupa järjestamine	Tähtede kaupa järjestamine (EE-s kasutusel)
Adam	Adam
Adam Bremenist	adamantaan
adamantaan	Adamaua mägismaa
Adamaua mägismaa	Adam Bremenist
adamiidid	adamiidid

ISO/IEC 14651 juhhib täiendavalt tähelepanu võimalikele asendustele enne ja pärast sortimist (vajalik võib olla Rooma numbrite teisendamine kümnendsüsteemi, sõnade ümberpaigutamine I.V. Kalašnikov → Kalašnikov, I.V. või muu taoline eeltöötlus), arvude sortimisele (kuidas tagada, et 9 ei paigutuks 10 järele kuigi 10 algab numbriga 1, mis on sortimisjärjekorras 9 ees) ja muudele erijuhtudele.

Rakendus, mis tugineb EVS 8:2000le peab laiema tähevaliku korral soovitatavalt rakendama standardit ISO/IEC 14651, kuid see pole nõutav, kui rakendus suudab tagada muude vahenditega samasuguse tulemuse, kui standardis näidatud sortimisalgoritmi rakendades. Tähelepanu tuleks aga pöörata sellele, et sortimisalgoritmi tuleb edaspidi paratamatult laiendada vastavalt kasutusel oleva tähevaliku laienemisele, mistõttu see ei tohi olla liiga kitsalt piiritletud tähevaliku või reeglistikuga.

Isegi minimaalse tähevaliku korral nagu ESET1 või ISO/IEC 8859-15 pole võimalik sortimist korraldada lihtsa loendi abil, sest sortimine toimub paratamatult vähemalt kahes etapis, kus esimesel etapil kontrollitakse märkide võrdsust, teisel etapil suur-väiketähtede erinevusi.

Näiteks ladina alustähestiku tähe i kohta annab ISO/IEC 14651 sortimistabel järgmise info:

```
<U0069> <S0069>;<BASE>;<MIN>;<U0069> % väiketäht I
<UFF49> <S0069>;<BASE>;<WIDE>;<UFF49> % täislauses väiketäht I
<U2170> <S0069>;<BASE>;<COMPAT>;<U2170> % väike rooma number üks
<U24D8> <S0069>;<BASE>;<CIRCLE>;<U24D8> % väiketäht I ringis
<U0049> <S0069>;<BASE>;<CAP>;<U0049> % suurtäht I
<UFF29> <S0069>;<BASE>;<WIDECAP>;<UFF29> % täislauses suurtäht I
<U2160> <S0069>;<BASE>;<COMPATCAP>;<U2160> % rooma number üks
% [...]
<U00ED> <S0069>;<BASE><AIGUT>;<MIN><MIN>;<U00ED> % väike I akuudiga
<U00CD> <S0069>;<BASE><AIGUT>;<CAP><MIN>;<U00CD> % suur I akuudiga
<U00EC> <S0069>;<BASE><GRAVE>;<MIN><MIN>;<U00EC> % väike I graavisega
<U00CC> <S0069>;<BASE><GRAVE>;<CAP><MIN>;<U00CC> % suur I graavisega
<U012D> <S0069>;<BASE><BREVE>;<MIN><MIN>;<U012D> % väike I kaarega
<U012C> <S0069>;<BASE><BREVE>;<CAP><MIN>;<U012C> % suur I kaarega
<U00EE> <S0069>;<BASE><CIRCF>;<MIN><MIN>;<U00EE> % väike I tsirkumfleksiga
<U00CE> <S0069>;<BASE><CIRCF>;<CAP><MIN>;<U00CE> % suur I tsirkumfleksiga
<U01D0> <S0069>;<BASE><CARON>;<MIN><MIN>;<U01D0> % väike I haagiga
<U01CF> <S0069>;<BASE><CARON>;<CAP><MIN>;<U01CF> % suur I haagiga
<U00EF> <S0069>;<BASE><TREMA>;<MIN><MIN>;<U00EF> % väike I täppidega
<U00CF> <S0069>;<BASE><TREMA>;<CAP><MIN>;<U00CF> % suur I täppidega
<U1E2F> <S0069>;<BASE><TREMA><AIGUT>;<MIN><MIN><MIN>;<U1E2F> % väike I täppide ja akuudiga
<U1E2E> <S0069>;<BASE><TREMA><AIGUT>;<CAP><MIN><MIN>;<U1E2E> % suur I täppide ja akuudiga
<U0129> <S0069>;<BASE><TILDE>;<MIN><MIN>;<U0129> % väike I tildega
<U0128> <S0069>;<BASE><TILDE>;<CAP><MIN>;<U0128> % suur I tildega
<U0130> <S0069>;<BASE><POINT>;<MIN><MIN>;<U0130> % suur I ülakupunktiga
<U012F> <S0069>;<BASE><OGONK>;<MIN><MIN>;<U012F> % väike I pöördseediiga
<U012E> <S0069>;<BASE><OGONK>;<CAP><MIN>;<U012E> % suur I pöördseediiga
<U012B> <S0069>;<BASE><MACRO>;<MIN><MIN>;<U012B> % väike I ülakriipsuga
<U012A> <S0069>;<BASE><MACRO>;<CAP><MIN>;<U012A> % suur I ülakriipsuga
<U1EC9> <S0069>;<BASE><CROOK>;<MIN><MIN>;<U1EC9> % väike I ülakonksuga
<U1EC8> <S0069>;<BASE><CROOK>;<CAP><MIN>;<U1EC8> % suur I ülakonksuga
<U0209> <S0069>;<BASE><2GRAV>;<MIN><MIN>;<U0209> % väike I topeltgraavisega
<U0208> <S0069>;<BASE><2GRAV>;<CAP><MIN>;<U0208> % suur I topeltgraavisega
<U020B> <S0069>;<BASE><BREVR>;<MIN><MIN>;<U020B> % väike I pööratud ülakaarega
<U020A> <S0069>;<BASE><BREVR>;<CAP><MIN>;<U020A> % suur I pööratud ülakaarega
<U1ECB> <S0069>;<BASE><POINS>;<MIN><MIN>;<U1ECB> % väike i alapunktiga
```

```

<U1ECA> <S0069>; "<BASE><POINS>"; "<CAP><MIN>"; <U1ECA> % suur I alapunktiga
<U1E2D> <S0069>; "<BASE><TILDS>"; "<MIN><MIN>"; <U1E2D> % väike I alatildega
<U1E2C> <S0069>; "<BASE><TILDS>"; "<CAP><MIN>"; <U1E2C> % suur I alatildega
<U2171> "<S0069><S0069>"; "<BASE><BASE>"; "<COMPAT><COMPAT>"; <U2171> % väike rooma number kaks
<U2161> "<S0069><S0069>"; "<BASE><BASE>"; "<COMPATCAP><COMPATCAP>"; <U2161> % rooma number kaks
% [...]
<U0133> "<S0069><S006A>"; "<BASE><BASE>"; "<COMPAT><COMPAT>"; <U0133> % väike ligatuur IJ
<U0132> "<S0069><S006A>"; "<BASE><BASE>"; "<COMPATCAP><COMPATCAP>"; <U0132> % suur ligatuur IJ
<U33CC> "<S0069><S006E>"; "<BASE><BASE>"; "<SQUARED><SQUARED>"; <U33CC> % ruuttolli märk
<U2173> "<S0069><S0076>"; "<BASE><BASE>"; "<COMPAT><COMPAT>"; <U2173> % väike rooma number neli
% [...]
<U0131> <S0131>; <BASE>; <MIN>; <U0131> % väike punktita I
<U026A> <S026A>; <BASE>; <MIN>; <U026A> % vähendatud suurtäht I
<U0268> <S0268>; <BASE>; <MIN>; <U0268> % väike I kriipsuga
<U0197> <S0268>; <BASE>; <CAP>; <U0197> % suur I kriipsuga
<U0269> <S0269>; <BASE>; <MIN>; <U0269> % ladina väiketäht ioota
<U0196> <S0269>; <BASE>; <CAP>; <U0196> % ladina suurtäht ioota

```

Esimeses veerus on selles tabelis tähe Unicode'i kood, järgnevad neli semikoolonitega eraldatud veergu kirjeldavad nelja võrdlusastet, protsendimärk alustab kommentaari, mis originaalis on selle tähe nimi Unicode'i järgi. Kuigi read on siin esitatud järjestatult, ei mõjuta nende omavaheline järjekord üldjuhul algoritmi töö tulemust, sest Unicode'i kood esimeses veerus võetakse sortimisel arvesse alles sortimise neljandal astmel.

Sortimisalgoritm alustab tööd esimese astme kaalude võrdlemisega, mis nii väikese, suure kui ka täppidega i puhul on sama – <S0069>. Rooma number kaks, ligatuur IJ ja ruuttolli märk sorditakse esimesel tasemel kui kahest tähest koosnevad stringid (vastavalt II, IJ ja I2), väike punktita i aga omab juba uut esimese astme kaalu <S0131>. See, milline on esimese astme kaalude omavaheline järjekord, selgub ridadest:

```

<S0069> % (väike) I
<S0131> % väike punktita I
<S026A> % vähendatud suurtäht I
<S0268> % väike I kriipsuga
<S0269> % ladina väiketäht ioota
<S006A> % ladina suurtäht ioota

```

Nende ridade omavahelise järjestuse muutmine muudab juba algoritmi töö tulemust. Samuti on siin näha, et suur- ja väiketähe eristus toimub järgnevatel astmetel, sest nii suure kui ka väikese i kaal on esimesel astmel sama.

Teisel astmel eristatakse harilikud ja diakriitikutega tähed. Tavalise i (nii suur- kui väiketähe) teise astme sortimiskaal on <BASE> (tähistab alustähte ehk ilma igasuguste diakriitikuteta tähte), akuudiga i-l <BASE><AIGUT> ja graaviseiga i-l <BASE><GRAVE>. Järelikult alustäht i on igal juhul eespool kui diakriitikutega i variandid, kuid viimaste omavahelise järjekorra teadasaamiseks peab algoritm võrdlema abisümboleid <AIGUT> ja <GRAVE> (taas on oluline ridade järjekord):

```

<BASE>
<LOWLINE> % kombineeruv alakriips
<PSILI> % kombineeruv diakriitik koma
<DASIA> % kombineeruv diakriitik pööratud koma
<AIGUT> % kombineeruv akuut
<GRAVE> % kombineeruv graavis
<BREVE> % kombineeruv kaar
<CIRCF> % kombineeruv tsirkumfleks
[...]

```

Kui kahe sõna omavaheline võrdlus pole ikka veel andnud selgust, kumb neist on eespool, võetakse kolmandal astmel arvesse suur- ja väiketähe erinevus. <MIN> tähistab siin väiketähte, <CAP> suurtähte. Neljas tase järjestab märke nende Unicode'i koodi alusel ning peab tagama, et algoritm oskab mistahes kahe stringi puhul üheselt otsustada, milline on nende omavaheline järjestus sortimisel. Võrdsed saavad olla vaid täpselt identsed stringid.

Näiteid:

- Stringid *Ízmail* ja *Izmail* eristuvad esimesel astmel, sest Í ja I kaalud erinevad. Samal põhimõttel eristatakse eesti tähestiku järgi sortides Z ja Ž või A ja Ä, mis muudes tähestikes on harilikult teise astme eristus;

- Stringid *Isak* ja *isák* eristuvad teisel astmel. Oluline on a ja á kaalude omavaheline järjekord, mis määrab, et *Isak* peab olema eespool. Kuid kui teisel astmel arvestatakse väike- ja suurtähe erinevust, oleks eespool hoopis *isák*;
- Stringid *Islandi* ja *islandi* on võrdsed nii esimesel kui teisel astmel, nende järjestus selgub alles kolmandal astmel.

Klaviatuur

Eesti klaviatuuril on kolm **registrit**:

- alumine register
- ülemine register (klahviga “shift” või “püstnool”)
- lisaregister (klahviga “AltGr” või “topeltpüstnool”)

Klahvidele joonistatud märkide asetus vastab registrile, mille abil neid sisestatakse. Klahvi alumises vasakus nurgas on alumise registri klahv, ülemises vasakus ülemise registri klahv ja paremal poolel (ISO klaviatuuristandardi järgi paremal all) asub lisaregistri abil sisestatav märk. Tähemärkide klahvidele pole ülevaatlikkuse huvides eraldi joonistatud väike- ja suurtähti.

Ülemise registri märkide sisestamiseks (nt suurtähed) tuleb vajutada ülemise registri klahvile ja seda all hoides vajutada soovitud klahvile. Samal viisil toimub lisaregistri märkide sisestamine klahviga AltGr.

Klaviatuuri eesmärk on teha tavakasutajale teksti sisestamine arvutisse võimalikult hõlpsaks. Kõige sagedamini tekstides esinevad märgid tehakse kättesaadavaks ühe klahvivajutusega, harvemini esinevaid märke saab sisestada kahe klahvi kombineerimisega. Et kõikmõeldavad tekstides esinevad märgid ei mahu parimagi tahtmise juures klaviatuurile, lubab operatsioonisüsteem harilikult sisestada ülejäänud märke kas spetsiaalprogrammi abil (Windowsis on selleks abivahendiks Character Map) või koodi abil.

Soovitus Windowsi kasutamisel: teades vajamineva märgi koodi (vt lisa Kooditabelid), on võimalik sisestada mistahes kooditabeli märki. Esmalt vajutage ja hoidke all klahvi Alt (klaviatuuril all vasakul). Seejärel tippige numbrisõrmistikul (mitte tähtede kohal asuvaid numbriklahve kasutades!) vajalik kood. Veenduga ka, et klaviatuuri olekulamp Num Lock oleks sisse lülitatud. Nulliga algavad koodid vastavad Windowsi kooditabelile, kõik muud DOSi kooditabelile. Nii nt Põhjamaade keeltes esineva Å sisestamiseks tippige tema kood alt+0197.

Sammuta klahvideks nimetatakse selliseid abiklahve, mis eraldi tähendust ei oma ja mis töötavad kombinatsioonis teiste klahvidega. Eesti klaviatuuril on defineeritud viis sammuta klahvi diakriitiliste märkidega tähtede sisestamiseks. Sammuta klahve ei tule all hoida — vajutage esmalt sammuta klahvile ja seejärel põhitähe klahvile, nt é sisestamiseks vajutage kõigepealt klahvile ´ (üleval paremal; näiliselt ei juutu selle vajutuse peale veel midagi), siis klahvile ‘e’. Mitte kõik klahvikombinatsioonid ei oma tähendust. ~ ja X kombinatsioon ei tekita X-i, mille peal oleks tilde, sest sellist märki valmiskujul kooditabeleis pole.

Vaatamata sellele, et eesti klaviatuuri defineeris üksikasjalikult juba EVS 8:1993, leidub ka siin küsitavusi.

1. Tähtede paigutus eesti klaviatuuril eeldab kooditabeli ISO 8859-1 või sellele sarnase tähevalikuga kooditabeli kasutust, Windows kasutab eesti klaviatuuri koos Balti kooditabeli tähevalikuga.

Probleeme ei tekita eesti tähestiku tähtede sisestamine. Õ, ä, ö ja ü on klaviatuuril alumise registri märgid, š ja ž sisestamiseks on kaks viisi: sammuta klahvi ‘haak’ kombineerimine s või z-ga ja lisaregistri klahv koos klahviga s või z. Mainigem, et viimast võimalust standard EVS 8:1993 ei maini, uues standardi väljaandes on seda nimetatud vaid kui soovituslikku.

Pilt muutub kardinaalselt, kui vaadelda sammuta klahvide tõlgendamist Windowsis. Sammuta klahve on eesti klaviatuuril viis: vasakus ülannurgas tilde ja haak, paremas ülannurgas graavis ja

akuut, ä kõrval lisaregistris katus. Sammuta klahvide kasutamisel vajutatakse esmalt sammuta klahvile (tilde ja graavise saamiseks koos ülemise registri klahviga SHIFT või püstnool), seejärel vajaliku põhitähe klahvile. Sammuta klahvide kasutus Windowsis:

Sammuta klahv	Märk	Tulemus	Kommentaar
Tilde	o	õ	Eelduspärane.
	n	~n	Tildet ja n-i Windows kombineerida ei luba, sest märk ñ puudub Balti kooditabelis.
	tühik	~	eelduspärane, kuid sisuliselt muudab ñ puudumine mõttetuks tilde kasutamise sammuta klahvina, sest ainus tildet kasutatav täht õ on juba eraldi klaviatuurile paigutatud.
Haak (caron)	s	š	Eelduspärane, kuid saadud š kood vastab Balti kooditabelile (alt+0240), Windowsi Western kooditabeli järgi peaks š kood olema (0154). Sama kommentaar kehtib suurtähe kohta.
	z	ž	Eelduspärane, kuid saadud ž kood vastab Balti kooditabelile (alt+0254), Windowsi Western kooditabeli järgi peaks ž kood olema (0158). Sama kommentaar kehtib suurtähe kohta.
	c	č	Eelduspärane, kuid ainult Balti kooditabelit kasutades. Č ei ole eesti tähestiku täht, mistõttu see täht puudub standardis EVS 8:2000 sätestatud aluskooditabelist ISO 8859-15 ja standard ei nõua, et seda tähte peaks saama klaviatuurilt sisestada.
	tühik	ˇ	
Graavis	a	ā	Äärmiselt ebaloogiline. Graavist tõlgendatakse siin kui makronit (pikenduskiipsu, esineb läti keeles), mida eesti keeles ei kasutata. Graavisega a kood Western kooditabelis (0224) ei vasta makroniga a koodile Balti kooditabelis (0226).
	i	ı̇	Graavist ja i-d Windows kombineerida ei luba, sest märk ï puudub Balti kooditabelis.
	tühik	ı̇	Eelduspärane kuid samuti mõttetu. Balti kooditabelis pole ühtegi graavise tähte, seega võiks graavis olla harilik klahv, mitte sammuta diakriitiline märk.
Akuut	a	á	Akuuti ja a-d Windows kombineerida ei luba, sest märk á puudub Balti kooditabelis.
	z	ẏ	Eelduspärane, kuid ainult Balti kooditabelit kasutades. Täht ‘z akuudiga’ esineb ainult poola keeles.
	e	é	Eelduspärane.
	tühik	´	
Katus (tsirkumfleks)	a	ą	Äärmiselt ebaloogiline. Katust tõlgendatakse siin kui pöördsemiid (kasutusel leedu keeles).
	o	ˆo	Katust ja o-d Windows kombineerida ei luba, sest märk ô puudub Balti kooditabelis.
	tühik	ˆ	Eelduspärane kuid samuti mõttetu. Balti kooditabelis pole ühtegi katusega tähte, seega võiks katus olla harilik klahv, mitte sammuta diakriitiline märk.

2. Eesti klaviatuuril on märke, mida ei leidu standardi uustöötluses alustabeliks võetud kooditabelis ISO 8859-15.

Aluskooditabelis ISO 8859-15 puuduvad märgid □ ja ½. Et need märgid on endiselt kasutusel paljudes teistes kooditabeleis, polnud otsest vajadust klaviatuuripaigutust ümber teha.

3. Eesti klaviatuuri on ebamugav kasutada programmeerijal.

Programmeerimiskeeled üldjuhul taunivad muutujate või protseduuride nimedes kaheksabitiste märkide kasutamist, teiselt poolt tingib nende süntaks ehk programmi vormistusreeglid loomuliku keele tekstides harva esinevate sümbolite sagedase kasutamise. Peaaegu kõigis programmeerimiskeeltes kasutatakse sageli sümboleid {, }, [,], \$ ja |. Kõik loetletud märgid asuvad eesti klaviatuuril lisaregistris, mis muudab nende sisestamise tülilaks. Eriti tülilaks on aga sisestada samuti laialdaselt kasutusel olevaid märke ~, ` ja ^, nende sisestamiseks tuleb need esmalt valida lisaregistri abil ja seejärel vajutada tühikule. Osadel klaviatuuridel on jäänud katus ä kõrvale märkimata. Samas klaviatuuril hõlpsalt kasutatavad täpitähtede klahvid jäävad programmeerimisel sageli hoopis kasutamata.

Teksti sisestamisel ja andmevahetusel tekkivaid tüüpvigu

Tüüpvead on esitatud tabelis, kas vaatluse all on tüüpilise eesti ä saatus.

Mis teieni jõudis	Võimalik seletus tekkinud veale
a	Kõige tõenäolisemalt on teksti sisestajal olnud ebamugav ä-d klaviatuurilt sisestada või olid tal ebamäärased kahtlused, et ä kasutamine võiks olla problemaatiline. Sel viisil sisestatud teksti taastamine nõuab olulist käsitööd ja kogu teksti ülevaatamist. Mis veel hullem, eriti isikunimedes on taastamine juba võimatu – kui väljend ‘... silmade sara’ on veel tuvastatav kui ‘silmade sära’ siis J. Sara võib olla nii Sara kui Sära (või hoopis Särä?). Informatsioon on pöördumatult kadunud ja tekst tuleb uuesti sisestada.
ae	Nn telegrammistilis sisestatud tekst (Palju oenne suennipaeevaks!). Sedagi sisestusviisi tuleks tingimata vältida, tekst on raskesti loetav ja ka siin pole originaalkuju sageli võimalik taastada. Teatud juhtudel on võimalik, et tekst on sisestatud korrektselt. Et ae on saksa keeles täiesti mõõndav ‘ä’ lihtsusviis, on selline asendus kasutusel ka programmides, mille põhieesmärk on tekst kasutajale esitada loetaval kujul, peaasi, et sellest oleks võimalik aru saada. Kui programm (nt uudiselugeja tin) otsustab, et 'ä'-d pole võimalik näidata, esitab ta selle ae-na. Vale küll, aga parem kui tühi kast, a või d.
a''	Elektronpostis ja uudisgruppides tihti kasutatust leidev asendusviis. Nagu iga surrogaatkirjaviisiga, kaasneb ka sellega andmete kadumise oht kuid võrreldes kahe eelnevaga on oht siiski väiksem. Kui <i>tingimata</i> on vaja piirduda ASCII kooditabeliga, siis seda kirjepilti tulekski eelistada... tingimusel et seda tehakse järjekindlalt (diakriitikut kujult meenutav märk on alati põhitähe taga, diakriitikutena kasutatavaid märke kasutataksegi ainult selleks otstarbeks ja hälbeid nagu u'' asemel y ei esine).
d	Andmeid sisestati korrektselt (kasutades mõnda enamlevinud ISO või Windowsi kooditabelit, neis kõigis on ä kood sama) kuid andmete edastusel kadus mingil vaheetapil kaheksas bitt, st ä kood muutus 228-128 = 100 ehk d-täheks. Andmeedastushäireid põhjustavad enamasti vananenud postiprogrammid.
Ã¤	Postiprogramm või brauser ei suuda teisendada UTF-8 kodeeringus saabunud teksti.
+AOQ-	Postiprogramm või brauser ei suuda teisendada UTF-7 kodeeringus saabunud teksti.
ì	Postiprogramm või brauser ei suuda teisendada Roman-8 kodeeringus saabunud teksti.

=E4	Postiprogramm või brauser ei suuda teisendada Quoted-Printable kodeeringut. Algandmed on sisestatud Windowsis või mõnd ISO kooditabelit kasutades (teisendades kuueteistkümnendarvu E4 tagasi kümnendarvuks, saame tulemuseks 228).
&auml;	Andmed on pärit HTML tekstist.
&#228;	Andmed on pärit HTML tekstist.
%o	Tõenäoliselt on tekst sisestatud korrektselt, kuid adressaat arvas miskipärast, et see on sisestatud Macintoshil. Teisendades seda Windowsi tarvis ümber, asendatakse 228 (millele Macintoshi kooditabelis vastab promillimärk 'korrektseks' Windowsi promillimärgiks koodiga 137).
„	Te üritate avada DOSis kirjutatud teksti Windowsis seda eelnevalt teisendamata. DOSi kooditabelites 850 ja 775 on ä kood 132, millele Windowsi kooditabeleis vastab jutumärk.
õ	Te üritasite sisestada Windowsis ä-d koodina kuid unustasite koodi alustada nulliga. Korrektne sisestusviis on alt+0228, alt+228 annab tulemuse DOSi kooditabeli järgi, kus sellel kohal asub õ.
Š	Te üritate avada Macintoshis kirjutatud teksti Windowsis seda eelnevalt teisendamata. Macintoshi kooditabelis on ä kood 138, millele Windowsi Western kooditabelis vastab Š.

Telefoninumbrite kasutus

Peatükis kasutatud terminid ja mõistekäsitlused

Rahvusvaheline number koosneb **riigikoodist** (*Country Code, CC*) ja sellele järgnevast **riigisisest numbrist** (*National Significant Number, NSN*). Eesti riigikood on 372. Riigisisene number koosneb **suunanumbrist** (*National Destination Code, NDC*) ja sellele järgnevast **tarbija numbrist** e **abonentnumbrist** (*Subscriber Number, SN*).

Tarbija number on numbrimärkide kombinatsioon, mille järgi eristatakse telefonivõrgu või mobiiltelefonivõrgu tarbijat. Tarbija numbril on viis kuni seitse numbrimärki (geograafiline number) või seitse kuni kaheksa numbrimärki (mittegeograafiline number, sealhulgas mobiiltelefonivõrgu number).

Numeratsiooniala on numbril algusosa järjestikuste numbrite hulk.

Geograafiline number on number, mis on seotud konkreetse geograafilise asukohaga. Eesti geograafilised numbrid asuvad järgmistes numeratsioonialades (vt suunanumber): 32, 33, 35, 38, 39, 43—48, 60—67, 69, 73, 74, 76—79. **Suunanumber** on numbrimärk või numbrimärkide kombinatsioon, mis identifitseerib riigisisest geograafilist piirkonda, Eestis nt 32 (Lääne-Virumaa), 7 (Tartumaa), 79 (Põlvamaa) jne. Tallinna ja Harjumaa geograafilistel numbritel suunanumbril ei ole.

Mittegeograafilised numbrid on mobiiltelefonivõrkude numbrid ja muud numbrid, mida telekommunikatsioonivõrgu operaator või telekommunikatsiooniteenuse osutaja ei ole sidunud konkreetse piirkonnaga. Eestis kasutatavad mittegeograafilised numbrid, välja arvatud teenusenumbrid ja lühinumbrid, asuvad järgmistes numeratsioonialades:

- mobiiltelefonivõrkude numbrid 50—57;
- kaugotsinguteenuseks kasutatavad numbrid 58, 59;
- lühinumbrid 10—19;
- teenusenumbrid 800, 900, 901, 907.

Prefiks on numbrimärk või numbrimärkide kombinatsioon, mis liidetakse numbril valimisel riigikoodi või suunanumbril ette ja mis võimaldab valida teises riigis või riigi siseselt teise piirkonna või mobiiltelefonivõrgu numbril. Prefiks ei kuulu rahvusvahelise numbril koosseisu.

Rahvusvaheline prefiks (00) (*International prefix*) on numbrimärkide kombinatsioon, mis võimaldab valida rahvusvahelist numbril (kaugekõne). **Riigisisene (kaugvalimise) prefiks (0)** (*National [trunk] prefix*) on numbrimärk või numbrimärkide kombinatsioon, mis võimaldab valida teises geograafilises piirkonnas või teises telekommunikatsioonivõrgus asuvat numbril (riigisisene kaugekõne, kõne mobiiltelefonile, kõne teise võrku).

Numbril valimine geograafiliselt numbril

Geograafiline number samas piirkonnas:	tarbija number ilma prefiksil ja suunanumbril
Geograafiline number teises piirkonnas:	0 – suunanumber – tarbija number
Mittegeograafiline number:	0 – tarbija number
Tasuta teenusenumbrid (800):	0 – teenuse number
Eritasuga teenusenumbrid (900):	0 – teenuse number
Andmesideteenuse numbrid (901):	0 – teenuse number
Taksofonide numbrid (907):	0 – taksofoni number

Lühinumber on kuni viiekohaline number, mis tagab telekommunikatsioonivõrgu operaatoril või telekommunikatsiooniteenuse osutajal (**operaatorikoodid**) või muude adressaatide, seal-

hulgas politsei, kiirabi ja päästeteenistuse (teenuste lühinumbrid) väljakutse lihtsustatud valimise. Numeratsioonialas 1 (777 ja numeratsiooniala 9 on esialgu kasutamata) paiknevad järgmised lühinumbrid:

- üleriigiline politsei väljakutse lühinumber 110;
- üleriigiline hädaabiteenistuse väljakutse lühinumber 112;
- numbriinformatsiooni neljakohalised numbrid 1181 kuni 1188;
- muude teenuste kolme- ja neljakohalised numbrid.

Operaatorikoodi kasutades valitakse see kõigepealt, nt valitakse riigisiseseid numbreid järgmiselt: operaatorikood – 0 – suunanumber – tarbijanumber.

Vastavalt teede- ja sideministri määrusega kinnitatud numeratsiooniplaanile jõustub nii välismaale kui ka riigisisese helistamise uus kord 2001. aasta 1. jaanuaril. Välismaale helistades tuleb prefiksi 8 00 asemel valida 00 ning riigisiselt helistades prefiksita 8 2 ja 8 asemel 0, millega prefiks 8 kaob kasutusest.

Aastal 2003 on ette nähtud üleminek ühtsele üle-eestilisele numeratsioonile ilma suunanumbrita.

Telefoninumbrite õigekiri (allikas: <http://www.et.ee/etinfo/0410>).

Seoses üleminekul uuele valimiskorrale on Eesti Telefoni poolt välja töötatud soovituslik juhend telefoninumbrite grupeerimiseks, mis on vastavuses ITU (*International Telecommunication Union*) soovitustega.

1. Abonentnumbri siseselt eristatakse grupid ühe tühikuga.
2. 7-kohalisi abonentnumbreid grupeeritakse alates tagantpoolt 4 ja 3 kaupa:
679 xxxx
3. 6-kohalisi abonentnumbreid grupeeritakse alates tagantpoolt 3 kaupa:
527 xxx
4. 5-kohalisi abonentnumbreid grupeeritakse alates tagantpoolt 3 ja 2 kaupa:
45 xxx
5. Dokumentides, mis on mõeldud **riigisiseseks** kasutuseks, sh eestikeelsetel visiitkaartidel, asetatakse riigisisene prefiks 0 ja sellega **kokku kirjutatav** suunakood sulgudesse.

Riigisisesed dokumendid

(0) 679 xxxx (Tallinn)
(038) 5x xxx (Järvamaa)
(07) xxx xxx (Tartu)
(044) xx xxx (Pärnu)

NB! Abonentnumber on sulgudes olevast riigisisesest prefiksist 0 ja ka sulgudes olevast, prefiksiga kokku kirjutatud suunakoodist eraldatud **kahe tühikuga**.

6. Dokumentides, mis on mõeldud Eestis **riigisiseseks** kasutuseks ja milles on vaja kirjeldada valimiskorda detailselt (st kus ei piisa ainult rahvusvahelise riigikoodi väljatoomisest, vaid on vaja esitada ka rahvusvaheline prefiks 00), kirjutatakse rahvusvaheline prefiks 00 ilma sulgudeta.

00 371 xxxx xxxx (Läti)

Eestist Lätisse helistades valitakse number **alati** prefiksiga, olenemata piirkonnast (maakonnast), kust helistatakse. Sulgudesse asetatakse aga ainult need prefiksivahemikud ja suunakoodid, mida olenevalt helistaja asukohast (kas piirkonna sees või väljaspool piirkonda) kord valitakse, kord mitte.

Rahvusvahelise kaugside prefiks 00 on suunakoodist eraldatud ühe tühikuga, suunakood abonentnumbrist **kahe tühikuga**.

7. Dokumentides, mis on mõeldud **rahvusvaheliseks** kasutuseks, nt ingliskeelsetel visiitkaartidel ja välismaale saadetavates kirjades, riigikoodi sulgudesse ei panda. Rahvusvaheline riigikood algab märgiga “+”

Rahvusvaheline kasutus

+372 xxx xxxx	(Tallinn)
+372 38 xx xxx	(Järvamaa)
+372 7 xxx xxx	(Tartu)
+372 44 xx xxx	(Pärnu)
+358 9 xxx xxxx	(Helsingi)

NB! Suunakood on riigikoodist eraldatud ühe tühikuga, abonentnumber suunakoodist **kahe tühikuga**.

8. Mobiiltelefoni numbril kuulub riigisisene kaugside prefiks 0 mobiiltelefoni **võrgukoodi koosseisu**. Mobiiltelefoni võrgukoodi **ei märgita sulgudesse**, kuna tegemist on kohustusliku koodiga, mis valitakse ka võrgusisese kõne korral:

050 xx xxx; 051 xx xxx; 052 xx xxx; 053 xxx xxx	(EMT)
056 xx xxx; 056 xxx xxx; 057 xx xxx, 057 xxx xxx	(Radiolinja)
055 xx xxx; 055 xxx xxx	(Q-GSM)

9. 900-seeria teenuste puhul kirjutatakse 0 järgnevast 900-st **lahku**:

0 900 xxxx

Vältimaks olukorda, kus 900-seeria ja 800-seeria numbreid hakatakse nimetama näiteks 09-seeria numbrites, tuleb numbreid vastavat korda silmas pidades ka lugeda. Seega tuleb 900-seeria numbrites koht informatsiooni andes lugeda "null-üheksasada", mitte "null-üheksa-null-null".

10. 901-seeria (andmeside sissehelistamiskeskused) teenuste puhul kirjutatakse 0 järgnevast 901-st **lahku**:

0 901 xxxx

11. 800-seeria teenuste puhul kirjutatakse 0 järgnevast 800-st **lahku**:

0 800 xxxx

12. Neljakohaline numbriinfo number kirjutatakse kokku:

1182

Operatsioonisüsteemide kohandamine eesti keele kasutuseks

Eesti keele tugi Linuxis (Mandrake 7.1 näitel)

Linux (nagu ka kõik teised) Unixipõhised operatsioonisüsteemid kasutab lokaliseerimiseks lokaate. Lokaadis defineeritakse kasutuskeel (mõjutab süsteemi poolt väljastatavate teadete vormistust), eelistatav kooditabel, sortimisreeglid, rahaühikute, numbrite, kuupäeva ja kellaaja vorming. Programmile edastatakse süsteemis või kasutaja poolt valitud lokaat keskkonnamuutujate kaudu kuid programmid võivad valida ja kasutada ka teisi süsteemis defineeritud lokaate ja vahetada neid vastavalt vajadusele (nt sortida tekstifail esmalt serbia keele reeglite põhjal, väljastada ja sortida järgmine tekstifail juba prantsuse keele tähestiku alusel). Mõnevõrra eraldi seisavad süsteemi ajavööndi määramine (andmestik maailma riikide ajavööndite ja nende muutuste kohta tuleb kasutajal endal leida ja kompileerida programmiga `zic`) ning klaviatuuri definitsioon. Linuxi uuemad versioonid kasutavad Eesti puhul kooditabelit ISO 8859-15, mis on EVS 8:2000 järgi aluskooditabel. Lokaadi kasutamine on viidud operatsioonisüsteemi teenuste hulka, mis välistab vajaduse programme eraldi konfigureerida.

Keskkonnamuutujad, mida kasutatakse keeleelistuste määramiseks, on järgmised:

LC_ALL

LC_ALL on ülemuslik kõigi teiste lokaadimuutujate suhtes v.a. LANGUAGE.

LANGUAGE=et

Muutujat LANGUAGE kasutab GNU teek ja see omab tähendust esialgu ainult Linuxis. LANGUAGE mõjutab operatsioonisüsteemi poolt väljastatavate teadete keelt. On võimalik defineerida ka mitu keelt, operatsioonisüsteem kontrollib vajalikke tõlgitud teadete olemasolu eelistuste järjekorras.

LC_CTYPE=et

Kui LC_ALL on defineerimata, määrab LC_CTYPE lokaadis kasutatava tähevaliku.

LC_COLLATE=et

Kui LC_ALL on defineerimata, määrab LC_COLLATE sortimisreeglid.

LC_MONETARY=et

Kui LC_ALL on defineerimata, määrab LC_MONETARY rahasummade kirjutuse.

LC_NUMERIC=et

Kui LC_ALL on defineerimata, määrab LC_NUMERIC arvude kirjutuse.

LC_TIME=et

Kui LC_ALL on defineerimata, määrab LC_TIME kuupäeva ja kellaaja formaadi.

LANG=et

LANG annab väärtuse kõigile määramata jäänud muutujatele (st puudub nii LC_ALL kui ka mõni üksikkategooria LC_CTYPE, LC_COLLATE, LC_MONETARY, LC_NUMERIC, LC_TIME).

Sortimine, kuupäeva, numbrite ja rahasummade formaat

Tehke vajalikud muudatused tekstifailis `/usr/share/i18n/locales/et_EE`

Käivitage programm `localedef` käsuga

```
localedef -c -i et_EE -f ISO-8859-15 et_EE.uus
```

Lokaadi andmestik muudetakse operatsioonisüsteemi jaoks mõnevõrra sobivamale kujule ja kirjutatakse kataloogi `/usr/share/locales/et_EE.uus`.

NB! Korrektnete uue lokaadi nimi on `et_EE.ISO-8859-15` (keele kood `et` pluss maa kood `EE`, punkt ja kooditabeli nimi). Kasutage seda, kui olete veendunud, et kõik töötab ja te teate, mida teete.

Parandage vastavalt faili `/etc/profile`, muutes või lisades rea

```
export LC_ALL=et_EE.uus
```

Mandrake kasutab faili `/etc/sysconfig/i18n`, mis samuti vajab muutmist. Muutused on näha pärast uut sisselogimist. Kontrollida võite ka Perli käsuga

```
perl -V:d_setlocale
```

Kui kõik on korras, vastab Perl `d_setlocale='defined'`, pikem nurin, mis lõpeb hoiatusega `Falling back to the standard locale ("C")` tähendab, et midagi on viltu läinud.

Klaviatuuripaigutus

Klaviatuuripaigutus pannakse paika installeerimisel. Hilisemaks muutmiseks kasutage programmi `/usr/sbin/kbdconfig` ja valige menüüst 'eesti'.

Ajavöönd

Mugav on süsteemi ajavööndi seadmiseks kasutada programmi `linuxconf`, tavakasutajal või vööndiaja ajutiseks muutmiseks aga keskkonnamuutuja `TZ=EET-2` abil. Kui Eesti valitsus siiski otsustab taas kord suveaja kasutusele võtta, tõmmake värske vööndiaegade fail Internetist (nt `ftp://elsie.nci.nih.gov/pub/`, sisuliselt on sealt vaja ainult Euroopa osa faili Tallinn) ja kasutage programmi

```
zic vööndiajafail
```

Hetkel töötavad daemonid automaatselt uue ajaga ei kohandu ja teil tuleb masinale uus algladimine teha.

Kokkuvõtteks

Unixilaadsete operatsioonisüsteemide keeletugi on üsna kaootilises seisus, põhjuseks standardite (v.a POSIX) ja üldiste kokkulepete puudumine. Kõrvu on kasutuses eesti keelde tõlgitud ja laitmatult keele- ja kultuurikeskkonda toetavad programmid (graafiline kasutajaliides KDE) ja sellised, millele igasugune ASCII piiridest väljuv märk raskusi valmistab (postiserver `smtpd`, tekstiotsinguteks loodud `glimpse` jt). Enamasti nõuab operatsioonisüsteemi ja rakendusprogrammide lokaliseerimine suuremal või vähemal määral kasutajapoolset sekkumist. Näiteks Mandrake 7.1 ei paigaldanud installeerimisel eesti keele tuge veatult ja uue lokaadiandmestiku kompileerimine tõi lisaks esile pisivea — märkide mnemoonilistes nimetustes ei tehtud vahet suur- ja väiketähtedel. Operatsioonisüsteemi avatus lokaliseerimiseks annab aga vigadele vaatamata olulise eelistuse täiesti suletud operatsioonisüsteemide ees.

Eesti keele tugi Windowsis

Windows 3.11 ei sisaldanud eesti keele tuge, kuid standardi EVS 8:1993 rakendamiseks loodud tarkvara võimaldas installeerida eesti klaviatuuri ja Western-koodilehekülje alusel loodud süsteemsed fondid, millesse oli lisatud ž-täht. **Windows 95**-ga algas Unicode'i kasutuselevõtt; kaheksabitised kooditabelid küll ei kadunud, kuid üleminekulahendusena oli võimalik kasutada mitut erinevat koodilehekülge ('Western' sisaldas Lääne-Euroopa, Cyr kirillitsa, Tur Türgile sobivaid, CE aga Kesk-Euroopa tähemärke). Eesti liideti koos Läti ja Leeduga ühise Balti kooditabeli alla. Positiivne oli see samm küll lõunanaabrite jaoks, Eestile tähendas see aga uut *de facto* standardit ja juba väljatöötatud tarkvaralahenduste ümber-tegemist. Sobivate fontide nappus ja Lääne-Euroopa tähemärkide sage kasutus raamatutes ja ajalehetekstides tõi kaasa ebakõla tavakasutajate ja kirjastajate vahele — tavakasutajad leppisid vaikumisi pealepandud Balti kooditabeliga, niipea aga, kui sisestatud tekst oli vaja trükki anda, mindi üle Lääne-Euroopa kooditabelile. **Windows NT 4.0** on juba üleni Unicode'il põhinev operatsioonisüsteem, kus erinevate koodilehekülgede kasutus omab tähtsust ainult piiratud keskkondades ja tulevikus on ette näha nende kasutuselt kadumist.

Nagu teisedki operatsioonisüsteemid, kasutab Windows NT lokaate. Praegu tunneb Windows 126 lokaati ja nende arv suureneb pidevalt. Eristatakse **süsteemilokaati** (*system locale*) mis määrab vaikimisi kasutatava kooditabeli ja süsteemifondid ja mida kasutavad Unicode'i mitte tundvad programmid (määratakse installeerimisel ja Windows 95 seda muuta ei luba), **kasutajalokaati** (*user locale*), mis on iga kasutajakonto lahutamatu omadus ja mille muutmine on võimalik ilma uue alglaadimiseta ning **sisestuslokaati** (*input locale*), mis määrab sisestuskeele ja tekstisisestusviisi ning mida tihti võrdsustatakse klaviatuuri-paigutusega. Windows võimaldab kasutada vanema operatsioonisüsteemi DOS programme. Kooditabelite ja klaviatuuri kasutus DOSis pannakse küll paika koos süsteemilokaadiga, kuid seda võib edaspidi muuta. Windows 95 sisaldab selleks eraldi utiliiti `changepcp.exe` (CD-l kataloogis `/other/changepcp/`).

Lokaadi kasutamiseks peab eelnevalt installeerima **keeltegrupi**, kuhu vajalik keel kuulub. Hetkel on Windowsis kasutusel 17 keeltegrupi — araabia, armeenia, balti, gruusia, heebrea, india, jaapani, keskeuroopa, kirillitsa, korea, kreeka, lääneeuroopa, tai, türgi, vietnami ja kaks hiina keeltegrupi, vajalike gruppide lisamine toimub Control Paneli menüüs Regional Options. Osadel lokaatidel puudub vastav koodilehekülge (nt gruusia ja sanskriti), ning neid on võimalik kasutada ainult Unicode'i toe abil. Sisestuslokaadi (klaviatuuripaigutuste) lisamine toimub samuti läbi Control Paneli, valige Keyboard – Input Locales (vajalik keeltegrupp peab eelnevalt installeeritud olema). Kui teil on installeeritud mitu klaviatuuripaigutust, on nende vahetamine võimalik klaviatuuri abil klahvikombinatsiooniga vasakpoolne Alt koos ülemise registri klahviga Shift.

Kokkuvõtteks

Erinevalt Unixist on Windows arenenud suletult ning ainult ühe firma tootena. Microsoft ei ole nõustunud standardima oma kooditabeleid, kirjeldama lokaadi andmestikku ega klaviatuuripaigutuse muutmise võimalusi. Tulemuseks on suhteliselt kompaktne ja kasutusmugavuselt Linuxit ületav tervik, mis aga oma suletuse tõttu põhjustab tihti kokkusobimatust eesti keele ja kultuuri tavadega. Vead, millega kasutajal tuleb leppida, varieeruvad pisi-seikadest nagu 'Sugust' 'augusti' asemel kuu nimetusena kuni globaalseteni nagu eesti keele paigutamine põhimõtteliselt sobimatusse balti keeltegrupi. Windowsi kasutajaliidese tõlkimine pole mõeldav ilma tootjapoolse heakskiiduta ning on isegi sel juhul kulukas.

HTML, elektronpost ja brauseri seaded

Internet on muutunud arvutikasutajale sama tavaliseks meediakanaliks kui seda on ajaleht või televisioon. Olukord, kus info saatja ja vastuvõtja töökeskkond on identne, on pigem erand kui reegel. Isegi ühe ja sama operatsioonisüsteemi kasutajail on käepärast erinev komplekt fonte, programme, erinevalt installeeritud keeltetugi ja oma eelistuste kohaselt paikapandud seaded. Igal sellisel juhul võib ette tulla olukord, kus informatsiooni (rakendusjuhendi kontekstis teksti kodeeringut) tuleb teisendada ja iga teisenduse juures on võimalik, et midagi läheb kaduma. Teine võimalik moonutuste allikas on tarkvaratootjate poolt kasutatavad kinnised failiformaadid. Tekstiteisendused ühest programmist teise ja isegi sama tootja sama programmi uue versiooni kasutusele võtmine toob kaasa vajaduse juba salvestatud faile uuesti avada ja uues vormingus salvestada ja ikka muutub seejuures midagi, reeglina ebasoovitavas suunas. Seetõttu on pikemas perspektiivis mõistlik kasutada ainult standardseid vorminguid.

Kadude vältimiseks on loodud terve rida standardeid info kodeerimiseks (nt Unicode) salvestamiseks (nt XML ja Rich Text Format) kui ka edastamiseks (nt MIME ja UTF-8). Sellest hoolimata esineb ikkagi olukordi, kus info adekvaatne edasitoimetamine pole võimalik, katsuge nt teisendada nimesid 'Jānis Bīlziņš' ja 'à la François Viète' Balti ja Lääne-Euroopa tähevalikut kajastavate kooditabelite vahel. EVS 8:2000 soovib aluskooditabelina kasutada Lääne-Euroopa tähemärke sisaldavaid kooditabeleid (ISO 8859-15 ja Windows CP 1252), mis viiks miinimumini vähemalt sedalaadi kaod tüüpilises Eesti keele ja kultuuri-keskkonnas.

Üks võimalus üheselt teksti autori poolt mõeldud märke edastada on mnemooniline kirjaviis ja kood, mis on valdavalt kasutusel veebilehekülgedel. Märki nime eristamiseks muust tekstist pannakse selle ette &, numbrilise koodi puhul lisaks # ja lõpetatakse semikooloniga. Märki kood on siin Unicode'i kood kümnendsüsteemis. Sellega välditakse kaheksabitiste kooditabelite kasutamist infoedastusel ja teksti saaja (täpsemalt brauser) võib olla kindel, et tekstis leiduv ä on nimelt a umlaudi e täppidega.

"Eesti riigiasutuste WWW-lehekülgede koostamise juhend" soovib lehekülgede autoritel kasutada eestikeelsete lehekülgede puhul kooditabelit ISO 8859-1, mis võib küll olla hetkel parim üldkasutatav lahendus, kuid mis ei lähe kokku standardi EVS 8:2000 soovitustega ning nõuab eraldi kommentaare. Teksti kodeerimisviisi valikul tuleb arvestada nii eesti tähtede kasutusvõimalusi, kui ka brauserite kooditabelite toetust, viimast ka erinevates operatsioonisüsteemides.

Andmeedastusel kasutatav kodeering on enamasti kirjas veebilehekülje päises, nt
<META HTTP-EQUIV="Content-type" CONTENT="text/html; charset=ISO-8859-1">
ja vene keele puhul

<META HTTP-EQUIV="Content-type" CONTENT="text/html; charset=KOI8-R">
Edastuskodeering ISO 8859-1 tähendab, et tekstis esinevad kaheksabitised märgid peavad vastama nimetatud kooditabelile, sellest välja jäävad märgid võib esitada mnemoonilisel kujul. Koodi 228 tõlgendatakse 'ä'-na, koodi 150 kasutamine mõttekriipsu tähenduses on viga, sest ISO 8859-1 ei sisalda vahemikus 128-159 üldse graafilisi märke. Kuid ISO 8859-1 ei sisalda ka tähti š ja ž, neid saab kasutada ainult kujul &#kood; Märksa sobivam oleks eestikeelsete lehekülgede kodeeringuks kasutada ISO 8859-15, ISO 8859-13, Windows CP 1252 jt, kus kõik eesti tähestiku tähed mahuvad ühte kaheksabitisesse kooditabelisse. Kõigi neil on aga omakorda puudujääke. ISO 8859-15 on sedavõrd uus, et esialgu on see kasutatav vaid Netscape 6-ga, ISO 8859-13 ei sobi muu tähevaliku poolest, Windowsi kooditabelid pole standardiseeritud. Kaaluda võib hoopis UTF-8 üldkasutusse võtmist, hoolimata selle kodeeringu suhtelisest "ebasõbralikkusest" kasutajate vastu (erinevalt teistest kodeerimisviisidest on UTF-8 kodeeringus tekst suhteliselt loetamatu, kui brauser seda dekodeerida ei suuda). Brauserite kasutusstatistika näitab vananenud tarkvara osakaalu pidevat vähenemist internetikasutuses (hinnanguliselt u 5% jääb UTF-8 kodeeringuga hätta – juuni 2000). See on juba jõudnud piirini, kus piiratud ja eesti keelele mitte kõige sobivamate kooditabelite kasutusest välja jätmine võiks üles kaaluda üksikute kasutajate ebamugavuse.

Eesti tähestiku tähed HTMLis:

Õ	Õ	Õ*	* Kümnekkoodid leiate märkide all tabelis ISO-8859-1 (vt Lisa 5: Kooditabelid)
õ	õ	õ	
Ä	Ä	Ä	
ä	ä	ä	
Ö	Ö	Ö	
ö	ö	ö	
Ü	Ü	Ü	
ü	ü	ü	
Š	Š	Š	Š sai nime alles hiljuti ning selle kasutamine vane- mate brauserite puhul on problemaatiline. Siin kordab HTMLi standard Windowsi möödalaskmist – š ilma ž-ta on tarbetu, sest eranditult kõigis keeltes, kus esineb š, kasutatakse ka ž-tähte. Et ž-l ametlikult mnemoonilist nime pole, ei ole ka &scaroni; kasutamisel mõtet.
š	š	š	
Ž	-	Ž	Analoogia põhjal š-ga oleks ž nimetus &#zcaron;
ž	-	ž	

Elektronposti puhul kasutatakse MIME vormingut, mis koos tekstiga saadab kirja päises selle teksti kodeerimisel kasutatud kooditabeli nimetuse. Tüüpiline kirja päis sisaldab nt järgmisi ridu:

```
Received: from madli.ut.ee [...]
[...]
MIME-Version: 1.0
Content-Type: TEXT/PLAIN; charset=ISO-8859-15
Content-Transfer-Encoding: 8BIT
Status: RO
```

Rida `Content-Type: TEXT/PLAIN; charset=ISO-8859-15` ütleb, et saadeti harilik tekst (mitte nt HTML või helifail) ja et tekstis leiduvad koodid vastavad kooditabelis ISO 8859-15 toodule. Rida `Content-Transfer-Encoding: 8BIT` annab teada, et tekstiedastusel seda täiendavalt ei teisendatud ja kõik kaheksabitised koodid esindavad iseennast. Hetkel on Eestis veel jõus soovitus kasutada elektronpostis edastuskodeeringuna 'Quoted Printable' vormingut (ä kodeeritakse =E4), käesolev rakendusjuhend soovitab aga alustada üleminekut UTF-8 kodeeringule koos kaheksabitise edastusega. Veenduge siiski, et kirja saajal ei teki selle avamisega raskusi.

Internet Explorer, Outlook Express ja Netscape Communicator

Internet Exploreri olulisemad seaded leiate menüüst Tools > Internet Options > General. Nupu all Fonts võite asendada fondi Times New Roman (Internet Exploreri vaikimisi font ladina kirja kasutatavate keelte näitamiseks) rohkem tähemärke sisaldava Lucida Sans Unicode fondi vastu, kui see teil installeeritud on.

Nupu all Languages lisage kindlasti eesti keel, sest mõned veebisaidid arvestavad kasutaja keelevelikuga ja võivad teile eestikeelse avalehekülje asemel pakkuda hoopis ingliskeelse. Uue keele lisamiseks klikkige nupul Add, kirjutage reale User defined Estonian [et] ja seejärel klikkige OK. Nihutage eesti keel nimekirja algusotsa, klikkides real, mille lisasite ja nupul Move up, kuni eesti keel on nimekirjas esimene.

Outlook Express: avage menüü Tools > Options > Send.

International Settings määrab kirjade kodeeringu, selleks valige soovitatavalt UTF-8 või Western European (Windows). Kodeeringut ISO 8859-15 Outlook Express hetkel kasutada ei võimalda. Veidi allpool olev Mail Sending Format määrab, kas kirja üritatakse kujundada nagu veebilehekülge (HTML) või saadetakse tavatekstina. Uudisegruppidesse kirjutamisel jääge kindlasti tavateksti juurde (Plain Text), kirjade puhul heale tuttavale võib kaaluda ka HTMLi kasutamist, kuid ärge unustage seda pärast tagasi muutmast. Nupp Plain Text Settings

vahetab andmevahetusel kasutatavat kodeeringut. Eesti piires ei tekita kaheksabitine andme-edastus raskusi, seega valige MIME formaat ja Encode Text Using > None.

Netscape (4.7): avage menüü Edit > Preferences.

Appearance > Fonts võimaldab kasutada määrata, milliseid fonte kasutatakse eri kodeeringute puhul. Avage Navigator > Languages ja lisage eesti keel. Selleks klikkige Add, reale Others kirjutage Estonian [et] (Netscape 6 puhul ainult et) ja klikkige OK. Nihutage eesti keel nimekirja algusotsa, klikkides real, mille lisasite ja nupul, millel ülespoole suunatud nool, kuni eesti keel on nimekirjas esimene.

Mail & Newsgroups > Messages (all) saate valida, kas edastusel kasutatakse Quoted Printable vormingut või saadetakse kiri teele kaheksabitisena. Rakendusjuhend soovib valida 'As is' e kaheksabitise variandi. Sama põhimenüü teine alajaotis Formatting määrab, kas teie kirjad (ja postitused uudisgruppidesse) kujundatakse HTMLiga või saadetakse tavalise tekstina. Siin tuleks küll jääda tavateksti kasutuse juurde (Use the plain text editor to compose messages), edasised valikud omavad tähtsust ainult HTMLi kasutamisel.

Kooditabelid

Kooditabelite kasutus ja teksti teisendamine ühest kooditabelist teise põhjustab sageli vigu ja toob kaasa moonutusi esialgses tekstis. Vaatamata sellele, et operatsioonisüsteemid ja tekstitöötlusprogrammid on üle minemas Unicode'i kasutamisele, on esmane kasutuskeskkond endiselt piiratud ühe 256-märgilise kooditabeli ja kindlale keelele orienteeritud klaviatuuriga. Kui aga operatsioonisüsteem võimaldab kasutada Unicode'i, tuleks loomulikult eelistada seda võimalust.

EVS 8:2000 kinnitas Eesti **aluskooditabeliks** ISO 8859-15. See ei tähenda tingimata, et kõik operatsioonisüsteemid oleksid kohustatud selle kasutusse võtma, sest kõigil neil on oma väljakujunenud kooditabelite süsteemid ja ISO 8859-15 otsene kasutus on mõeldav ainult Unixil põhinevates operatsioonisüsteemides. Küll aga eeldab aluskooditabeli selline valik, et ka muudes operatsioonisüsteemides kasutataks eesti keele ja kultuurikonteksti arvestades võimalikult lähedase märgivalikuga kooditabeleid. Märgivalikus peaksid tingimata olema kõik eesti tähestiku tähed (see välistab CE e Kesk-Euroopas kasutatavad kooditabelid ISO 8859-2, Windows CP 1250 jt, kus puudub õ-täht) ning on tungivalt soovitatav, et peale eesti tähestiku tähtede kuuluksid märgivalikusse Eestis avaldatavates trükistes sageli esinevad Lääne-Euroopa keelte tähed ja Euroopa ühisraha euro märk (mis muudab ebasoovitavaks Balti tähevalikuga kooditabelite ISO 8859-13, Windows CP 1257 ja DOS CP 775 kasutamise). Kooditabelite eelistatav kasutus operatsioonisüsteemide kaupa on järgmine:

Operatsioonisüsteem (üldnimetus)	Esmane eelistus	Teisene eelistus	Viimane võimalus
Unix	ISO 8859-15	ISO 8859-13	ISO 8859-1
Windows	CP 1252 (Western)	CP 1257 (Baltic)	-
DOS	CP 1116	CP 775	CP 850
Macintosh	-	CENTEURO	-

Kui esimeses veerus toodud kooditabelit kasutada ei õnnestu, tuleks eelistada teises veerus toodud kooditabelit. Kolmandas veerus esitatud kooditabeli kasutus on võimalik ainult juhul kui ühtki sobivamat valikut pole.

Valiku laiemalt kasutatavaid kooditabeleid leiate eraldi lisast.

Tarkvara vastavuse nõuded standardile EVS 8:2000

Tarkvara vaadeldakse diferentseeritult – omaette nõuded esitatakse operatsioonisüsteemidele ja rakendusprogrammidele ning standardile vastavus on jagatud kolme kategooriasse:

Operatsioonisüsteemid

A

- Operatsioonisüsteem ja selle juurde kuuluvad abifailid, utiliidid, teegid jm vahendid on täielikult tõlgitud eesti keelde;
- Operatsioonisüsteem võimaldab piiranguteta sisestada, töödelda ja salvestada kõiki Unicode'i märke vähemalt rakendustasemel 1 ning toetab kodeeringut UTF-8;
- Operatsioonisüsteemi koosseisus peab olema vähemalt üks font, mille tähevalikusse kuuluksid kõik ESET1 märgid;
- Operatsioonisüsteem võimaldab kasutada eesti paigutusega klaviatuuri ja ühte järgnevaist kooditabelitest ISO 8859-15, Windows CP 1252 (nagu see on esitatud Unicode'i lisamaterjalide hulgas, st. koos euro märgi ja ž-tähega) või DOS CP 1116;
- Operatsioonisüsteem lubab kasutajal muuta põhilisi lokaliseerimisel arvesse tulevaid väärtusi või vorminguid – ajavööndit, kuupäeva ja kellaaja esitusviise, numbrite ja rahasummade esitust, klaviatuuripaigutust, sortimisreegleid, paberi formaati.

B

- Operatsioonisüsteem toetab kodeeringut UTF-8 ja võimaldab sisestada ning kasutada Unicode'i märke vähemalt standardis EVS 8:2000 esitatud tähevaliku ESET1 ulatuses;
- Operatsioonisüsteem võimaldab kasutada eesti paigutusega klaviatuuri ja eesti tähestikku kuuluvate tähtede kasutamine ei põhjusta raskusi;
- Operatsioonisüsteem lubab kasutajal muuta põhilisi lokaliseerimisel arvesse tulevaid väärtusi või formaate – ajavööndit, kuupäeva ja kellaaja esitusviise, numbrite ja rahasummade esitust, klaviatuuripaigutust, sortimisreegleid, paberi formaati.

C

- Operatsioonisüsteem võimaldab kasutada eesti paigutusega klaviatuuri ja eesti tähestikku kuuluvate tähtede kasutamine ei põhjusta raskusi.

Rakendustarkvara

A

- Programm ja selle juurde kuuluvad abifailid, utiliidid, teegid jm vahendid on täielikult tõlgitud eesti keelde;
- Programmiga on kaasas eestikeelne abimaterjal, mis oma mahult ei ole väiksem kui programmi originaalversiooniga kaasnev materjal;
- Programm võimaldab piiranguteta sisestada ja kasutada kõiki Unicode'i märke, vajadusel tuginedes selleks operatsioonisüsteemi poolt pakutavatele standardvahenditele;
- Programm arvestab kõikide süsteemsete lokaliseerimisasetustega ega piira nende kasutust;
- Programm, mille funktsioonide hulka kuulub loomuliku keele töötlus (teksti või kõne), võimaldab kohandada eesti keelele kõiki neid reegleid, mida ta ise oma töös kasutab — st kui programm kasutab õigekirjamoduleid, peab liides võimaldama eesti keele õigekirja-kontrolli lisada, kui programm teisendab klaviatuurilt sisestatud jutumärke, peab saama muuta jutumärkide kuju jne);

B

- Programmiga on kaasas eestikeelne abimaterjal, mis oma mahult ei ole väiksem kui programmi originaalversiooniga kaasnev materjal;
- Programm ei piira neid operatsioonisüsteemi poolt pakutavaid võimalusi, mis kuuluvad standardiga EVS 8:2000 käsitletud valdkondade alla;

C

- Programm on lokaliseeritav;

Riistvara**A**

- Seadmes kasutatavad loomuliku keele teated ja sildid on eestikeelsed;
- Seadmega on kaasas eestikeelne abimaterjal, mis oma mahult ei ole väiksem kui muu(de)s keel(te)s esitatud abimaterjal;
- Seade, mille töö spetsiifika puudutab standardis EVS 8:2000 piiritletud valdkondi, vastab ühtlasi samas standardis esitatud eelistustele (klaviatuuriga varustatud seade kasutab eesti klaviatuuri, printer kasutab paberi formaati A4, GPS mõõdab meetreid, mitte jalgu jne.);

B

- Seadmega on kaasas eestikeelne abimaterjal, mis oma mahult ei ole väiksem kui muu(de)s keel(te)s esitatud abimaterjal;
- Seade ei piira neid operatsioonisüsteemi poolt pakutavaid võimalusi, mis kuuluvad standardiga EVS 8:2000 käsitletud valdkondade alla;

C

- Seade mis võimaldab sisestada, kuvada või väljastada teksti, lubab raskusteta kasutada eesti tähestiku tähti (subtitreerimismasinad võimaldavad kuid näiteks mobiiltelefonides kasutatav GSM kooditabel ei vasta isegi minimaalsetele EVS 8ga sobivuse nõuetele).

Riistvara vastavuse nõuetele lisanduvad rakendustarkvarale esitatud nõuded selles ulatuses, millises seadmesse on vastav tarkvara sisse ehitatud.

Riist- või tarkvarahangete juures tuleks taotleda vähemalt B-taseme vastavust standardile EVS 8:2000. Kui hankija saab vähemalt kaks pakkumist A- või B-tasemel, tuleks valik teha nende vahel, jättes C-taseme ja standardile üldse mitte vastavad pakkumised kõrvale. Kui vähemalt kaks pakkumist vastavad A-, B- või C-tasemele, tuleks valik teha nende vahel.

Lisa 1. Maailma riigid ja nende ametikeeled

Riik	Keel*	Kirjasüsteem (latinisatsioon)**
Afganistan	puštu, dari	araabia
Albaania	albaania	
Alžeeria	araabia	araabia (prantsuse trkr)
Ameerika Ühendriigid	inglise, (hawai, hispaania, samoa, tšamorro)	
Andorra	katalaani, prantsuse, hispaania	
Angola	portugali	
Antigua ja Barbuda	inglise	
Araabia Ühendemiraadid	araabia	araabia (inglise trkr)
Argentina	hispaania	
Armeenia	armeenia	armeenia (eesti trkr, VÕK 1976)
Aserbaidžaan	aserbaidžaaani	
Austraalia	inglise	
Austria	saksa, (sloveeni)	
Bahama	inglise	
Bahrein	araabia	araabia (inglise trkr)
Bangladesh	bengali	bengali
Barbados	inglise	
Belau	belau, inglise	
Belgia	hollandi, prantsuse, saksa	
Belize	inglise, hispaania	
Benin	prantsuse	
Bhutan	dzongkha	tiibeti (omaladina)
Boliivia	hispaania, aimara, ketšua	
Bosnia ja Hertsegoviina	bosnia, horvaadi, serbia	kirillitsa (rööpne horvaadi tähestik)
Botswana	inglise, tsvana	
Brasiilia	portugali	
Brunei	inglise, malai	
Bulgaaria	bulgaaria	kirillitsa (omaladina, ÜRO 1977)
Burkina Faso	prantsuse	
Burundi	rundi, prantsuse	
Cabo Verde	portugali	
Colombia	hispaania	
Costa Rica	hispaania	
Côte d'Ivoire	prantsuse	
Djibouti	prantsuse, araabia	araabia (prantsuse trkr)
Dominica	inglise	
Dominikaani Vabariik	hispaania	
Ecuador	hispaania	
Eesti	eesti	

* sulgudes on regionaal- e vähemuskeel(ed)

** viimane veerg on täidetud vaid siis, kui tegemist ei ole ladina kirjaga; kui võimalik, on lisatud latinisatsiooni nimetus või kehtestaja lühend. Lühendid: lat = latinisatsioon, trkr = transkriptsioon, VÕK = vabariiklik õigekeelsuskomisjon, ÜRO = Ühinenud Rahvaste Organisatsioon.

Egiptus	araabia	araabia (inglise trkr)
Ekvatoriaal-Guinea	hispaania	
El Salvador	hispaania	
Eritrea	tigrinja, araabia, inglise	etioopia araabia
Etioopia	amhari	etioopia (omaladina, ÜRO 1967)
Fidži	inglise, fidži	
Filipiinid	tagali, inglise	
Gabon	prantsuse	
Gambia	inglise	
Ghana	inglise	
Grenada	inglise	
Gruusia	gruusia	gruusia (eesti trkr, VÕK 1976)
Guatemala	hispaania	
Guinea	prantsuse	
Guinea-Bissau	portugali	
Guyana	inglise	
Haiti	prantsuse, haitikreooli	
Hiina	hiina	hiina (omaladina, ÜRO 1977)
Hispaania	hispaania, (baski, galeegi, katalaani)	
Holland	hollandi, (friisi, papiamento)	
Honduras	hispaania	
Horvaatia	horvaadi	
Iiri	iiri, inglise	
Iisrael	heebrea, araabia	heebrea (omaladina, ÜRO 1977) araabia
India	hindi, inglise, (assami, bengali, gudžarati, kannada, malajalami, marathi, oria, pandžabi, tamili, telugu, urdu)	devanaagari (inglise trkr) (assami, bengali, gudžarati, kannada, malajalami, devanaagari, oria, gurmukhi, tamili, telugu, arabia)
Indoneesia	indoneesia	
Iraak	araabia	araabia (inglise trkr)
Iraan	pärsia	araabia (omaladina, ÜRO 1967)
Island	islandi	
Itaalia	itaalia, (prantsuse, saksa)	
Jaapan	jaapani	hiina-jaapani (Hepburni trkr; <i>kunrei</i>)
Jamaica	inglise	
Jan Mayen	norra	
Jeemen	araabia	araabia (inglise trkr)
Jordaania	araabia	araabia (inglise trkr)
Jugoslaavia	serbia, (albaania, ungari)	kirillitsa (rööpne horvaadi tähestik)
Kambodža	khmeeri	khmeeri (ÜRO lat 1972)
Kamerun	prantsuse, inglise	
Kanada	inglise, prantsuse	
Kasahstan	kasahhi, vene	kirillitsa (eesti trkr, VÕK 1977) kirillitsa (<i>vt Venemaa</i>)
Katar	araabia	araabia (inglise trkr)
Kenya	suahiili, inglise	

Kesk-Aafrika Vabariik	prantsuse	
Kiribati	inglise, kiribati	
Komoorid	prantsuse, araabia	araabia (prantsuse trkr)
Kongo DV	prantsuse	
Kongo Vabariik	prantsuse	
Kreeka	kreeka	kreeka (omaladina, ÜRO 1987)
Kuuba	hispaania	
Kuveit	araabia	araabia (inglise trkr)
Kõrgõzstan	kirgiisi, vene	kirillitsa (eesti trkr, VÕK 1977) kirillitsa (vt <i>Venemaa</i>)
Küpros	kreeka, türgi	kreeka (omaladina, ÜRO 1987)
Laos	lao	lao (prantsuse trkr)
Leedu	leedu, (poola)	
Lesotho	sotho, inglise	
Libeeria	inglise	
Liechtenstein	saksa	
Liibanon	araabia	araabia (prantsuse trkr)
Liibüa	araabia	araabia (inglise trkr)
Luksemburg	prantsuse, saksa, letseburgi	
Lõuna-Aafrika Vabariik	inglise, afrikaani, ndebele, pedi, sotho, tsvana, svaasi, suulu, tsonga, venda, koosa	
Lõuna-Korea	korea	korea (McCune-Reischaueri trkr)
Läti	läti, (liivi)	
Madagaskar	malagassi, prantsuse	
Makedoonia	makedoonia	kirillitsa (ÜRO lat 1977)
Malaisia	malai	
Malawi	njandža, inglise	
Maldiivid	maldiivi	divehi (omaladina)
Mali	prantsuse	
Malta	malta, inglise	
Maroko	araabia	araabia (prantsuse trkr)
Marshalli Saared	inglise, maršalli	
Mauritaania	araabia, prantsuse, (soninke, <i>toucouleur</i> ? i, volofi)	araabia (prantsuse trkr)
Mauritius	inglise, prantsuse	
Mehhiko	hispaania	
Mikroneesia	inglise, kosrae, nukuoro, pohnpei, tšuugi, ulithi, <i>woleai</i> , japi	
Moldova	moldova (rumeenia)	
Monaco	prantsuse	
Mongoolia	mongoli	kirillitsa, mongoli
Mosambiik	portugali	
Myanmar	birma	birma
Namiibia	inglise	
Nauru	nauru, inglise	
Nepal	nepali	devanaagari (inglise trkr)
Nicaragua	hispaania	

Nigeeria	inglise, hausa, (ibo, joruba)	
Niger	prantsuse, hausa	
Norra	norra, uusnorra, (põhja-, lõuna-, Lule saami)	
Omaan	araabia	araabia (inglise trkr)
Paapua Uus-Guinea	inglise, uusmelaneesia (tok-pisini)	
Pakistan	urdu, inglise	araabia (inglise trkr)
Panama	hispaania	
Paraguay	hispaania, guaranii	
Peruu	hispaania, ketšua, aimara	
Poola	poola	
Portugal	portugali	
Prantsusmaa	prantsuse, (bretooni, oksitaani)	
Põhja-Korea	korea	korea
Rootsi	rootsi, (põhja-, Lule saami)	
Rumeenia	rumeenia	
Rwanda	ruanda, prantsuse, inglise	
Saalomoni Saared	inglise	
Saint Kitts ja Nevis	inglise	
Saint Lucia	inglise	
Saint Vincent	inglise	
Saksamaa	saksa, (alam-, ülemsorbi)	
Sambia	inglise	
Samoa	samoa, inglise	
San Marino	itaalia	
São Tomé ja Príncipe	portugali	
Saudi Araabia	araabia	araabia (inglise trkr)
Seišellid	seišellikreooli, inglise, prantsuse	
Senegal	prantsuse	
Sierra Leone	inglise	
Singapur	inglise, malai, hiina, tamil	hiina (inglise trkr) tamil (inglise trkr)
Slovakkia	slovaki, (ungari)	
Sloveenia	sloveeni, (itaalia, ungari)	
Somaalia	somaali, araabia	araabia
Soome	soome, rootsi, (Inari, põhja-, koltasaami)	
Sri Lanka	singali, tamil, inglise	singali (inglise trkr) tamil (inglise trkr)
Sudaan	araabia, inglise	araabia (inglise trkr)
Suriname	hollandi	

Suurbritannia	inglise, (gaeli, iiri, kõmri)	
Svaasimaa	svaasi, inglise	
Süüria	araabia	araabia (prantsuse v inglise trkr)
Šveits	saksa, prantsuse, itaalia, retoromaani	
Zimbabwe	inglise	
Taani	taani, (fääri, grööni, saksa)	
Tadžikistan	tadžiki	kirillitsa (eesti trkr, VÕK 1977)
Tai	tai	tai (omaladina, ÜRO 1967)
Tansaania	suahiili, inglise	
Togo	prantsuse	
Tonga	tonga, inglise	
Trinidad ja Tobago	inglise	
Tšaad	prantsuse, araabia	(prantsuse trkr)
Tšehhi	tšehhi	
Tšiili	hispaania	
Tuneesia	araabia	araabia (prantsuse trkr)
Tuvalu	inglise	
Türgi	türgi	
Türkmenistan	türkmeeni	kirillitsa* (eesti trkr, VÕK 1977)
Uganda	inglise, suahiili	
Ukraina	ukraina	kirillitsa (eesti trkr, VÕK 1975)
Ungari	ungari	
Uruguay	hispaania	
Usbekistan	usbeki	kirillitsa* (eesti trkr, VÕK 1977)
Uus-Meremaa	inglise, maoori	
Valgevene	valgevene, vene	kirillitsa (eesti trkr, VÕK 1976) kirillitsa (vt <i>Venemaa</i>)
Vanuatu	bislama	
Vatikan	ladina, itaalia	
Venemaa	vene	kirillitsa ¹ eesti trkr, 1998; ² omaladina, ÜRO 1987
Venezuela	hispaania	
Vietnam	vietnami	

* toimub üleminek ladina transkriptsioonile

Lisa 2. Keeled ja nendes esinevad lisatähed

Keeled, mille tähestik (või latinisatsioon) ei sisalda ladina alustähestikust erinevaid tähti

aimara, belau, fidži, grööni, hiina*, inglise, ketšua, kiribati, koosa, kosrae, kreeka*, ladina, malai, nauru, ndebele, nukuoro, papiamento, pedi, pohnpei, ruanda, rundi, seišellikreooli, somaali, soninke, sotho, suahiili, suulu, svaasi, tai*, *toucouleur*'i, tsonga, tšuugi, tuvalu, uusmelaneesia (tok-pisini), *woleai*

Keeled, mille puhul kasutatakse eesti transkriptsiooni

armeenia*, gruusia*, kasahhi*, kirgiisi*, tadžiki*, türkmeeni (turkmeeni)*, ukraina*, usbeki*, valgevene*, vene*

Keeled, mille tähestik või latinisatsioon sisaldab ladina alustähestikust erinevaid tähti

Iga keele puhul on näidatud nendes lisaks ladina alustähestikule kasutatavad tähed. Nool-sulgudes on vastava tähe kood ISO10646/Unicode'is, eesti keeletähestikku kuuluvatel tähtedel seda ei esitata. Mõne tähe kirjanepuuduseks (vt nt bislama *Á*-i) tuleb kasutada kaht koodi, sel juhul on esimene alustähe kood, teine kombineeruva diakriitika kood. Kursiivis on nende keelte nimetused, mille kogutähestik ei mahu Eestis kasutatava ladina tähtede valikusse.

Allpool ei ole järgmiste lisas 1 mainitud keelte latinisatsioonides sisalduvate tähtede tabelleid: amhari*, araabia*, assami*, bengali*, birma*, gudžarati*, hindi*, kannada*, khmeeri*, lao*, malajalami*, marathi*, mongoli*, nepali*, oria*, pandžabi*, puštu*, pärsia*, tamili*, telugu*, urdu*.

AFRIKAANI

È <00C8> è <00E8>,
É <00C9> é <00E9>,
Ê <00CA> ê <00EA>,
Ë <00CB> ë <00EB>,
Î <00CE> î <00EE>,
Ï <00CF> ï <00EF>,
Ô <00D4> ô <00F4>,
Û <00DB> û <00FB>

ALBAANIA

Ç <00C7> ç <00E7>,
Ë <00CB> ë <00EB>

ASERBAIDŽAANI

(Ä ä),
Ə <018F> ə <0259>,
Ç <00C7> ç <00E7>,
Ğ <011E> ğ <011F>,

İ <0130> i <0131>,
Ö ö,
Ş <015E> ş <015F>,
Ü ü

Märkus: Eraldi tähepaarid on İ i ja I i. Ä ja Ə märgivad tegelikult üht ja sama tähte, Ä on teise tähe mõõndav asendus.

BASKI

Ç <00C7> ç <00E7>,
Ñ <00D1> ñ <00F1>,
Û ü

BISLAMA

É <00C9> é <00E9>,
Ï <00CF> ï <00EF>,
À <004D><0304>
Á <006D><0304>,

Â <0050><0304>

Ã <0070><0304>,

Ü ü

BOSNIA

Ć <0106> ć <0107>,
Č <010C> č <010D>,
Đ <0110> đ <0111>,
Š š,
Ž ž

BRETOONI

Â <00C2> â <00E2>,
Ê <00CA> ê <00EA>,
Ñ <00D1> ñ <00F1>,
Ô <00D4> ô <00F4>,
Û <00D9> ù <00F9>,
Ü ü

* tärniga on lisas 2 läbivalt tähistatud mitteladina tähestikust erinevad keeled, mille puhul loend sisaldab nende latinisatsioonidest kasutatavaid lisamärke

BULGAARIA*

Ă <01CD> ă <01CE>,
 Ć <010C> ć <010D>,
 Š š,
 Ž ž

EESTI

Ä ä, Ö ö, Š š,
 Ü ü, Ž ž

FRIISI

Â <00C2> â <00E2>,
 Ê <00CA> ê <00EA>,
 Î <00CE> î <00EE>,
 Ô <00D4> ô <00F4>,
 Ú <00DA> ú <00FA>,
 Û <00DB> û <00FB>

FÄÄRI

Á <00C1> á <00E1>,
 Æ <00C6> æ <00E6>,
 Ð <00D0> ð <00F0>,
 Í <00CD> í <00ED>,
 Ó <00D3> ó <00F3>,
 Ø <00D8> ø <00F8>,
 Ú <00DA> ú <00FA>,
 Ý <00DD> ý <00FD>

GAELI (Šoti gaeli)

À <00C0> à <00E0>,
 Á <00C1> á <00E1>,
 È <00C8> è <00E8>,
 É <00C9> é <00E9>,
 Ì <00CC> ì <00EC>,
 Ò <00D2> ò <00F2>,
 Ó <00D3> ó <00F3>,
 Ù <00D9> ù <00F9>

GALEEGI

Á <00C1> á <00E1>,
 É <00C9> é <00E9>,
 Í <00CD> í <00ED>,
 Ñ <00D1> ñ <00F1>,
 Ó <00D3> ó <00F3>,
 Ú <00DA> ú <00FA>,
 Û ü

GUARANII

Á <00C1> á <00E1>,
 ã <00E3>,
 É <00C9> é <00E9>,
 ě <1EBD>,
 Å <0067><0303>,
 Í <00CD> í <00ED>,
 ã <0129>,
 Ñ <00D1> ñ <00F1>,
 Ó <00D3> ó <00F3>,
 õ,
 Ú <00DA> ú <00FA>,
 ã <0169>,
 ÿ <1EF9>

HAUSA

Ɓ <0181> ɓ <0253>,
 ɗ <018A> d <0257>,
 ƙ <0198> k <0199>

HAVAI

‘ <02BB>,
 Ā <0100> ā <0101>,
 Ē <0112> ē <0113>,
 Ī <012A> ī <012B>,
 Ō <014C> ō <014D>,
 Ū <016A> ū <016B>

HEEBREA*

‘ <02BB>,
 É <00C9> é <00E9>,
 ͜ <0048><0331>
 ͝ <1E96>,
 ͞ <1E94> ͟ <1E95>

HISPAANIA

Á <00C1> á <00E1>,
 É <00C9> é <00E9>,
 Í <00CD> í <00ED>,
 Ñ <00D1> ñ <00F1>,
 Ó <00D3> ó <00F3>,
 Ú <00DA> ú <00FA>,
 Ü ü

HOLLANDI

Ä ä,

Ë <00CB> ë <00EB>,
 Ī <00CF> ī <00EF>,
 Ö ö,
 Ü ü

HORVAADI

Ć <0106> ć <0107>,
 Č <010C> č <010D>,
 Đ <0110> đ <0111>,
 Š š,
 Ž ž

IBO

Ị <1ECA> ì <1ECB>,
 Ọ <1ECC> ọ <1ECD>,
 Ụ <1EE4> ụ <1EE5>

IIRI

Á <00C1> á <00E1>,
 É <00C9> é <00E9>,
 Í <00CD> í <00ED>,
 Ó <00D3> ó <00F3>,
 Ú <00DA> ú <00FA>

INDONEESIA

É <00C9> é <00E9>

ISLANDI

Á <00C1> á <00E1>,
 Æ <00C6> æ <00E6>,
 Ð <00D0> ð <00F0>,
 É <00C9> é <00E9>,
 Í <00CD> í <00ED>,
 Ó <00D3> ó <00F3>,
 Ö ö,
 Þ <00DE> þ <00FE>,
 Ú <00DA> ú <00FA>,
 Ý <00DD> ý <00FD>

ITAALIA

À <00C0> à <00E0>,
 È <00C8> è <00E8>,
 É <00C9> é <00E9>,
 Ì <00CC> ì <00EC>,
 Í <00CD> í <00ED>,

Ò <00D2> ò <00F2>,
Ó <00D3> ó <00F3>,
Ù <00D9> ù <00F9>,
Ú <00DA> ú <00FA>

JAAPANI*

Ō <014C> ō <014D>,
Ū <016A> ū <016B>
Märkus: Rööpselt kasu-
tusel olevas nn riiklikus
latinisatsioonis (*kunrei-*
shiki) on tsirkumfleksiga
tähed Ô ô ja Û û.

JAPI

Ä ä,
Ë <00CB> ë <00EB>,
Ö ö

JORUBA

Ẹ <1EB8> ẹ <1EB9>,
Ọ <1ECC> ọ <1ECD>,
Ş <1E62> ş <1E63>

KATALAANI

À <00C0> à <00E0>,
Ç <00C7> ç <00E7>,
È <00C8> è <00E8>,
É <00C9> é <00E9>,
Í <00CD> í <00ED>,
Ï <00CF> ï <00EF>,
Ľ <013F> ľ <0140>,
Ò <00D2> ò <00F2>,
Ó <00D3> ó <00F3>,
Ú <00DA> ú <00FA>,
Û ü

KOREA*

Ŏ <014E> ǒ <014F>,
Ŭ <016C> ǔ <016D>

KŌMRI (uelsi)

À <00C0> à <00E0>,
Á <00C1> á <00E1>,
Â <00C2> â <00E2>,
Ä ä,
È <00C8> è <00E8>,

É <00C9> é <00E9>,
Ê <00CA> ê <00EA>,
Ë <00CB> ë <00EB>,
Ì <00CC> ì <00EC>,
Í <00CD> í <00ED>,
Î <00CE> î <00EE>,
Ï <00CF> ï <00EF>,
Ò <00D2> ò <00F2>,
Ó <00D3> ó <00F3>,
Ô <00D4> ô <00F4>,
Ö ö,
Ù <00D9> ù <00F9>,
Ú <00DA> ú <00FA>,
Û <00DB> û <00FB>,
Ü ü,
Ẁ <1E80> ẁ <1E81>,
Ẃ <1E82> ẃ <1E83>,
Ẅ <0174> ẅ <0175>,
Ẇ <1E84> ẇ <1E85>,
Ỳ <1EF2> ỳ <1EF3>,
Ỵ <00DD> ỵ <00FD>,
Ỷ <0176> ỷ <0177>,
Ỹ <0178> ỷ <00FF>

LETSEBURGI (luksemburgi)

À ä,
É <00C9> é <00E9>,
Ë <00CB> ë <00EB>,
Æ <004D><0302>
Ç <006D><0302>,
Ě <004E><0302>
É <006E><0302>,
Ö ö,
Ü ü

LEEDU

Ạ <0104> ạ <0105>,
Č <010C> č <010D>,
È <0116> è <0117>,
Ẹ <0118> ẹ <0119>,
Ị <012E> ị <012F>,
Š š,
Ū <016A> ū <016B>,
Ụ <0172> ụ <0173>,

Ž ž

LIIVI

Ā <0100> ā <0101>,
Ä ä,
< <01DE> > <01DF>,
% <1E10> ™ <1E11>,
Ē <0112> ē <0113>,
Ī <012A> ī <012B>,
Ļ <013B> ļ <013C>,
Ņ <0145> ņ <0146>,
Ō ō,
¬ <00D5><0304>
¼ <00F5><0304>,
Ō <014C> ō <014D>,
Ŗ <0156> ŗ <0157>,
Š š,
Ţ <0162> ț <0163>,
Ū <016A> ū <016B>,
Ž ž

Märkus: tähe all olev
diakriitik on täpsemalt
alakoma

LÄTI

Ā <0100> ā <0101>,
Č <010C> č <010D>,
Ē <0112> ē <0113>,
Ģ <0122> ģ <0123>,
Ī <012A> ī <012B>,
Ķ <0136> ķ <0137>,
Ļ <013B> ļ <013C>,
Ņ <0145> ņ <0146>,
Š š,
Ū <016A> ū <016B>,
Ž ž

Märkus: tähe all olev
diakriitik on täpsemalt
alakoma

MAKEDOONIA*

Ć <0106> ć <0107>,
Č <010C> č <010D>,
Đ <0110> đ <0111>,
Š š, Ž ž

MALAGASSI

Á <00C1> á <00E1>,
Đ <004E><0308>
Ñ <006E><0308>,
Ô <00D4> ô <00F4>

MALTA

À <00C0> à <00E0>,
Ċ <010A> ċ <010B>,
È <00C8> è <00E8>,
Ġ <0120> ġ <0121>,
Ħ <0126> ħ <0127>,
Ì <00CC> ì <00EC>,
Î <00CE> î <00EE>,
Ò <00D2> ò <00F2>,
Ù <00D9> ù <00F9>,
Ż <017B> ż <017C>

MAOORI

Ā <0100> ā <0101>,
Ē <0112> ē <0113>,
Ī <012A> ī <012B>,
Ō <014C> ō <014D>,
Ū <016A> ū <016B>

MARŠALLI

Ā <0100> ā <0101>,
Ĺ <013B> ĺ <013C>,
Ĕ <004D><0327>
Ė <006D><0327>,
Ĭ <004E><0304>
Ī <006E><0304>,
Ŋ <0145> ŋ <0146>,
Ō <014C> ō <014D>,
Ī <004F><0327>
Ĭ <006F><0327>,
Ū <016A> ū <016B>

NJANDŽA (tšitševa)

Ŵ <0174> ŵ <0175>

NORRA

Å <00C5> å <00E5>,
Æ <00C6> æ <00E6>,
É <00C9> é <00E9>,
Ò <00D2> ò <00F2>,
Ô <00D4> ô <00F4>,

Ø <00D8> ø <00F8>

PILIPINO vt TAGALOGI

POOLA

Ą <0104> ą <0105>,
Ć <0106> ć <0107>,
Ę <0118> ę <0119>,
Ł <0141> ł <0142>,
Ń <0143> ń <0144>,
Ó <00D3> ó <00F3>,
Ś <015A> ś <015B>,
Ż <0179> ż <017A>,
Ź <017B> ź <017C>

PORTUGALI

À <00C0> à <00E0>,
Á <00C1> á <00E1>,
Â <00C2> â <00E2>,
Ã <00C3> ã <00E3>,
Ç <00C7> ç <00E7>,
É <00C9> é <00E9>,
Ê <00CA> ê <00EA>,
Í <00CD> í <00ED>,
Ó <00D3> ó <00F3>,
Ô <00D4> ô <00F4>,
Õ õ,
Ú <00DA> ú <00FA>,
Û ü

PRANTSUSE

À <00C0> à <00E0>,
Â <00C2> â <00E2>,
Ç <00C7> ç <00E7>,
È <00C8> è <00E8>,
É <00C9> é <00E9>,
Ê <00CA> ê <00EA>,
Ë <00CB> ë <00EB>,
Î <00CE> î <00EE>,
Ï <00CF> ï <00EF>,
Ô <00D4> ô <00F4>,
Œ <0152> œ <0153>,
Ù <00D9> ù <00F9>,
Û <00DB> û <00FB>,

Ü ü,
ÿ <0178> ÿ <00FF>

RETOROMAANI

À <00C0> à <00E0>,
È <00C8> è <00E8>,
É <00C9> é <00E9>,
Ì <00CC> ì <00EC>,
Î <00CE> î <00EE>,
Ò <00D2> ò <00F2>,
Ö ö,
Ù <00D9> ù <00F9>

ROOTSI

Ä ä,
Å <00C5> å <00E5>,
É <00C9> é <00E9>,
Ö ö

RUMEENIA

Â <00C2> â <00E2>,
Ă <0102> ă <0103>,
Î <00CE> î <00EE>,
Ș <015E> ș <015F>,
Ț <0162> ț <0163>
Märkus: tähe all olev diakriitik on alakoma ja korrektset, kuid ESET1st välja jäävad koodid on
Ș <0218> ș <0219>,
Ț <021A> ț <021B>

SAAMI (põhjaasaami)

Á <00C1> á <00E1>,
Č <010C> č <010D>,
Đ <0110> đ <0111>,
Ŋ <014A> ŋ <014B>,
Š š,
Ț <0166> ț <0167>,
Ž ž

SAAMI (Inari saami)

Á <00C1> á <00E1>,
Â <00C2> â <00E2>,
Ä ä,
Č <010C> č <010D>,

Đ <0110> đ <0111>,
Ń <014A> ń <014B>,
Š š,
Ž ž

SAAMI (koltasaami)

Â <00C2> â <00E2>,
Ă ä,
Å <00C5> å <00E5>,
Č <010C> č <010D>,
Đ <0110> đ <0111>,
Ʒ <01B7> ʒ <0292>,
Ž <01EE> ž <01EF>,
Ǧ <01E6> ǧ <01E7>,
Ǧ <01E4> ǧ <01E5>,
Ǧ <01E8> ǧ <01E9>,
Ń <014A> ń <014B>,
Ö ö,
Š š,
Ž ž

SAAMI (Lule saami)

Á <00C1> á <00E1>,
Ă ä,
Å <00C5> å <00E5>,
Ñ <00D1> ñ <00F1>

SAAMI (lõunasaami)

Ă ä,
Å <00C5> å <00E5>,
Ö ö

SAKSA

Ă ä,
Ö ö,
ß <00DF>,
Ü ü

SAMOA

‘ <02BB>

SERBIA

Ć <0106> ć <0107>,
Č <010C> č <010D>,
Đ <0110> đ <0111>,
Š š,

Ž ž

SLOVAKI

Á <00C1> á <00E1>,
Ă ä,
Č <010C> č <010D>,
Ď <010E> ď <010F>,
É <00C9> é <00E9>,
Í <00CD> í <00ED>,
Ĺ <0139> ľ <013A>,
Ľ <013D> ľ <013E>,
Ň <0147> ň <0148>,
Ó <00D3> ó <00F3>,
Ô <00D4> ô <00F4>,
Ř <0154> ř <0155>,
Š š,
Ť <0164> ť <0165>,
Ú <00DA> ú <00FA>,
Ý <00DD> ý <00FD>,
Ž ž

SLOVEENI

Č <010C> č <010D>,
Š š,
Ž ž

SOOME

Ă ä,
Ö ö

SORBI (alamsorbi)

Ć <0106> ć <0107>,
Č <010C> č <010D>,
Ě <011A> ě <011B>,
Ł <0141> ł <0142>,
Ń <0143> ń <0144>,
Ř <0154> ř <0155>,
Ś <015A> ś <015B>,
Š š,
Ž <0179> ż <017A>,
Ž ž

SORBI (ülemsorbi)

Ć <0106> ć <0107>,
Č <010C> č <010D>,
Ě <011A> ě <011B>,
Ł <0141> ł <0142>,
Ń <0143> ń <0144>,
Ř <0154> ř <0155>,
Ś <015A> ś <015B>,
Š š,
Ž <0179> ż <017A>,
Ž ž

Ł <0141> ł <0142>,
Ń <0143> ń <0144>,
Ó <00D3> ó <00F3>,
Ř <0158> ř <0159>,
Š š,
Ž <0179> ż <017A>,
Ž ž

ŠOTI GAELI vt GAELI

TAANI

Å <00C5> å <00E5>,
Æ <00C6> æ <00E6>,
É <00C9> é <00E9>,
Ø <00D8> ø <00F8>

TAGALOGI (pilipino)

À <00C0> à <00E0>,
Á <00C1> á <00E1>,
Â <00C2> â <00E2>,
È <00C8> è <00E8>,
É <00C9> é <00E9>,
Ê <00CA> ê <00EA>,
Ä <0047><0303>
Å <0067><0303>,
Ì <00CC> ì <00EC>,
Í <00CD> í <00ED>,
Î <00CE> î <00EE>,
Ñ <00D1> ñ <00F1>,
Ò <00D2> ò <00F2>,
Ó <00D3> ó <00F3>,
Ù <00D9> ù <00F9>,
Ú <00DA> ú <00FA>,
Û <00DB> û <00FB>

TONGA

‘ <02BB>

TSVANA

Ê <00CA> ê <00EA>,
Ô <00D4> ô <00F4>

TŠAMORRO

Â <00C2> â <00E2>,
Ñ <00D1> ñ <00F1>,
Ü ü

TŠEHHI

Á <00C1> á <00E1>,
 Č <010C> č <010D>,
 Ď <010E> đ <010F>,
 Ě <00C9> é <00E9>,
 ě <011A> ě <011B>,
 Í <00CD> í <00ED>,
 Ň <0147> ň <0148>,
 Ó <00D3> ó <00F3>,
 Ř <0158> ř <0159>,
 Š š,
 Ť <0164> ť <0165>,
 Ú <00DA> ú <00FA>,
 Ů <016E> ů <016F>,
 Ý <00DD> ý <00FD>,
 Ž ž

TŠITŠEVA vt NJANDŽA**TÜRGI**

Â <00C2> â <00E2>,
 Ç <00C7> ç <00E7>,
 Ğ <011E> ğ <011F>,
 Î <00CE> î <00EE>,
 İ <0130>,
 ı <0131>,
 Ö ö,
 Ş <015E> ş <015F>,
 Û <00DB> û <00FB>,
 Ü ü

Märkus: türgi tähestikus moodustavad eraldi tähepaarid İ i ja I ı.

UELSI vt KÖMRI**ULITHI:**

Ä <0226> ä <0227>,
 Ę <0116> ę <0117>,
 £ <022E> ³ <022F>

UNGARI

Á <00C1> á <00E1>,
 Ě <00C9> é <00E9>,
 Í <00CD> í <00ED>,
 Ó <00D3> ó <00F3>,
 Ö ö,
 Ő <0150> ő <0151>,
 Ú <00DA> ú <00FA>,
 Ü ü,
 Ű <0170> ű <0171>

Ó <00D3> ó <00F3>,
 Ö ö,
 Ő <0150> ő <0151>,
 Ú <00DA> ú <00FA>,
 Ü ü,
 Ű <0170> ű <0171>

USBEKI**Ò Ó****Ô Õ**

Märkus: valik pärineb ametlikust kavandatastavast ladina tähestikust

VEENE*

" <0034>,
 Č <010C> č <010D>,
 Ě <00C8> è <00E8>,
 Ě <00CB> ë <00EB>,
 Š š,
 Ž ž

Märkus: käesolev valik on vene omaladina (vene nimele translitereerimise tabelist)

VIETNAMEI

À <00C0> à <00E0>,
 Á <00C1> á <00E1>,
 Â <00C2> â <00E2>,
 Æ <1EA6> æ <1EA7>,
 Ā <1EA4> ā <1EA5>,
 Ă <1EAA> ă <1EAB>,
 Ȃ <1EA8> ȃ <1EA9>,
 Ȧ <1EAC> ȧ <1EAD>,
 Ȧ <00C3> ă <00E3>,
 Ȧ <0102> ă <0103>,
 Ȧ <1EB0> ă <1EB1>,
 Ȧ <1EAE> ă <1EAF>,
 Ȧ <1EB4> ă <1EB5>,
 Ȧ <1EB2> ă <1EB3>,
 Ȧ <1EB6> ă <1EB7>,
 Ȧ <1EA2> ă <1EA3>,
 Ȧ <1EA0> ă <1EA1>,
 Đ <0110> đ <0111>,
 È <00C8> è <00E8>,
 É <00C9> é <00E9>,
 Ê <00CA> ê <00EA>,
 Ë <1EC0> è <1EC1>,
 Ë <1EBE> é <1EBF>,
 Ë <1EC4> ẽ <1EC5>,
 Ê <1EC2> ẽ <1EC3>,
 Ê <1EC6> ê <1EC7>,
 Ë <1EBC> ẽ <1EBD>,
 Ê <1EBA> ẽ <1EBB>,
 Ɛ <1EB8> ɛ <1EB9>,
 Ì <00CC> ì <00EC>,
 Í <00CD> í <00ED>,
 Ï <0128> ï <0129>,
 Ì <1EC8> ì <1EC9>,
 Ị <1ECA> ị <1ECB>,
 Ò <00D2> ò <00F2>,
 Ó <00D3> ó <00F3>,
 Ô <00D4> ô <00F4>,
 Ò <1ED2> ò <1ED3>,
 Ỏ <1ED0> ỏ <1ED1>,
 Ỗ <1ED6> ỗ <1ED7>,
 Ỗ <1ED4> ỗ <1ED5>,
 Ộ <1ED8> ộ <1ED9>,
 Ỗ ỗ,
 Ỏ <1ECE> ỏ <1ECF>,
 Ơ <01A0> ơ <01A1>,
 Ỗ <1EDC> ờ <1EDD>,
 Ớ <1EDA> ớ <1EDB>,
 Ỗ <1EE0> ỡ <1EE1>,
 Ỗ <1EDE> ỡ <1EDF>,
 Ợ <1EE2> ợ <1EE3>,
 Ộ <1ECC> ộ <1ECD>,
 Û <00D9> ù <00F9>,
 Ú <00DA> ú <00FA>,
 Û <0168> ù <0169>,
 Ủ <1EE6> ủ <1EE7>,
 Ʋ <01AF> ư <01B0>,
 Û <1EEA> ừ <1EEB>,
 Ứ <1EE8> ứ <1EE9>,
 Ỡ <1EEE> ỡ <1EEF>,
 Ủ <1EEC> ử <1EED>,
 Ự <1EF0> ự <1EF1>,
 Ụ <1EE4> ụ <1EE5>,
 Ỡ <1EF2> ỳ <1EF3>,
 Ý <00DD> ý <00FD>,
 Ỡ <1EF8> ỹ <1EF9>

Ÿ <1EF6> ý <1EF7>,
Ỳ <1EF4> ỳ <1EF5>
Märkus: Vietnami
tähestikus on toonide
märgimiseks kasutusel 5
diakriitikut: ´ (akuut),
` (graavis), ~ (tilde),
ˆ (konks) ja . (alapunkt).
Need võivad kombinee-

ruda ka muid hääldus-
nähtusi märgivate dia-
kriitikutega. Väljaspool
ametlikku dokumentat-
siooni on mööndav tooni-
diakriitikute ärajätmine.

VOLOFI

Â <00C2> â <00E2>,

Ă <0102> ă <0103>,
È <00C8> è <00E8>,
É <00C9> é <00E9>,
Ö <0048><0324>
× <0068><0324>,
Ø <0054><0324>
Û <0074><0324>

¹ ий = ii	(<i>üldjuhul:</i>)	Новороссийск = Novorossiisk, Вий = Vii
² ий = i	(<i>kahe- ja enamasilbilise sõna lõpul:</i>)	Горький = Gorki, Долгорукий = Dolgoruki
¹ й = i	(<i>üldjuhul:</i>)	Майоров = Maiorov, Толстой = Tolstoi
² й = j	(<i>sõna algul vokaali ees:</i>)	Йодное = Jodnoje
к = k		Катя = Katja, Жуковский = Žukovski
л = l		Липецк = Lipetsk, Смоленск = Smolensk
м = m		Мусоргский = Mussorgski, Кострома = Kostroma
н = n		Новгород = Novgorod, Минин = Minin
о = o		Омск = Omsk, Ломоносов = Lomonossov
п = p		Павлов = Pavlov, Степан = Stepan
р = r		Руслан = Ruslan, Игорь = Igor
¹ с = s	(<i>üldjuhul:</i>)	Серов = Serov, Курск = Kursk
² с = ss	(<i>vokaalide vahel ja sõna lõpul vokaali järel:</i>)	Писарев = Pissarev, Денис = Deniss Märkus: <i>Liitsõnalise nime järelkomponendi algul oleva с-i võib asendada ühekordse s-ga (Новосибирск = Novosibirsk)</i>
т = t		Титов = Titov, Иркутск = Irkutsk
у = u		Ушаков = Ušakov, Лунин = Lunin
ф = f		Фадеев = Fadejev, Кафтанов = Kaftanov
¹ х = h	(<i>üldjuhul:</i>)	Хабаровск = Habarovsk, Архипов = Arhipov
² х = hh	(<i>vokaalide vahel ja sõna lõpul vokaali järel:</i>)	Чехов = Tšehhov, Долгих = Dolgihh Märkus: <i>Liitsõnalise nime järelkomponendi algul oleva х võib asendada ühekordse h-ga (Самоходов = Samohodov)</i>
ц = ts		Цыганов = Tsõganov, Кудрявцев = Kudrjajtsev
ч = tš		Чичиков = Tšitšikov, Гатчина = Gattšina
ш = š		Шишкин = Šiškin, Пушкин = Puškin
щ = štš		Щедрин = Štšedrin, Верещагин = Vereštšagin
ь	(<i>jäab märkimata:</i>)	Подьячев = Podjatšev
ы = õ		Крылов = Krõlov, Мощный = Moštšnõi
¹ ь	(<i>üldjuhul jääb märkimata:</i>)	Юрьевец = Jurjevets, Тотьма = Totma
² ь = j	(<i>и ees:</i>)	Ильич = Iljitš, Ильинский = Iljinski
э = e		Электрогорск = Elektrogorsk
ю = ju		Южно-Сахалинск = Južno-Sahhalinsk, Тютчев = Tjuttšev
я = ja		Ярославль = Jaroslavl, Брянск = Brjansk, Евгения = Jevgenija, Лидия = Lidija Märkus: <i>väljaspool dokumente ja teatmeteoseid võib eesnimede lõpul и järel я asendada а-ga (Евгения = Jevgenia, Лидия = Lidia)</i>

Erinevused võrreldes 1976. a "Õigekeelsussõnaraamatus" avaldatud tabeliga:

1. с = ss vokaalide vahel ja sõna lõpul vokaali järel (*varem:* .. sõna lõpul rõhulise silbi vokaali järel)
2. Liitsõnalise nime järelkomponendi algul oleva с-i võib asendada ühekordse s-ga (*varem:* asendatakse ühekordse s-ga)
3. Liitsõnalise nime järelkomponendi algul oleva х võib asendada ühekordse h-ga (*varem:* asendatakse ühekordse h-ga)
4. я = ja (kõigil juhtudel; *varem:* eesnimede lõpul ия = ia; nüüd lubatav üksnes väljaspool dokumente ja teatmeteoseid)

Vene nimede translitereerimine Eesti dokumentides

tähed	näited
a = a	Андрианов = Andrianov, Калинин = Kalinin
б = b	Бородино = Borodino, Тобольск = Tobol'sk
в = v	Волга = Volga, Вавилов = Vavilov
г = g	Григорий = Grigorij, Вологда = Vologda
д = d	Дон = Don, Волгоград = Volgograd
е = e	Петропавловск = Petropavlovsk, Сергей = Sergej, Егоров = Egorov, Алексеев = Alekseev, Васильев = Vasil'ev, Подъездов = Pod"ezdov
ё = ё	Орёл = Orël, Ёлкин = Ėlkin, Жёлтый = Žëltyj, Пугачёв = Pugačëv
	Märkus: <i>Täht e-ga märgitud ё transkribeeritakse vastavalt algkujule kas e või ё-ga</i>
ж = ž	Жуков = Žukov, Нижний Новгород = Nižnij Novgorod
з = z	Зернов = Zernov, Кутузов = Kutuzov
и = i	Исаев = Isaev, Филин = Filin, Иосиф = Iosif
й = j	Майоров = Majorov, Толстой = Tolstoj, Йошкар-Ола = Joškar-Ola
к = k	Катя = Katja, Жуковский = Žukovskij
л = l	Липецк = Lipeck, Смоленск = Smolensk
м = m	Мусоргский = Musorgskij, Кострома = Kostroma
н = n	Новгород = Novgorod, Минин = Minin
о = o	Омск = Omsk, Ломоносов = Lomonosov
п = p	Павлов = Pavlov, Степан = Stepan
р = r	Руслан = Ruslan, Игорь = Igor'
с = s	Серов = Serov, Курск = Kursk, Писарев = Pisarev, Денис = Denis, Новосибирск = Novosibirsk
т = t	Титов = Titov, Иркутск = Irkutsk
у = u	Ушаков = Ušakov, Лунин = Lunin
ф = f	Фадеев = Fadeev, Кафтанов = Kaftanov
х = h	Хабаровск = Habarovsk, Архипов = Arhipov, Чехов = Čehov, Долгих = Dolgih, Самоходов = Samohodov
ц = c	Цыганов = Cyganov, Кудрявцев = Kudrjavcev
ч = č	Чичиков = Čičikov, Гатчина = Gatčina
ш = š	Шишкин = Šiškin, Пушкин = Puškin
щ = šč	Щедрин = Ščedrin, Верещагин = Vereščagin
ъ = "	Подьячев = Pod"jačev
ы = y	Крылов = Krylov, Мощный = Moščnyj
ь = '	Юрьевец = Jur'evec, Тотьма = Tot'ma, Ильич = П'и'ч, Ильинский = П'inskij
э = è	Электрогорск = Èlektrogorsk
ю = ju	Южно-Сахалинск = Južno-Sahalinsk, Тютчев = Tjutčev
я = ja	Ярославль = Jaroslavl', Брянск = Brjansk, Мария = Marija, Евгения = Evgenija, Лидия = Lidija

Haridusminister Mait KLAASSEN

16. novembril 1998 valminud Haridusministeeriumi eesti-ladina töörühma aruanne

Vabariigi valitsuse määruse (vt eespool) täitmiseks loodud ekspertkomisjonis osalesid haridusministeeriumi, Eesti keele instituudi, Kodakondsus- ja migratsiooniameti, Riigikantselei ja AS Andmevara esindajad. Komisjon kaalus võimalusi määruse täitmiseks ning tegi ettepanekuid määruse ümbersõnastamiseks:

[...]

1. Muuta Vabariigi Valitsuse 25. märtsi 1998. a määruse nr 66 „Võõrkeelsete isikunimede ümberkirjutusreeglite kehtestamine“ (RT I 1998, 31/32, 427) nimetust ja esitada see järgmisel kujul: „Võõrkeelsete isikunimede kirjutusreeglite kehtestamine“.

2. Muuta Vabariigi Valitsuse 25. märtsi 1998. a määruse nr 66 punkt 1 ja sõnastada see järgmiselt:

„1. Ladinatähelised võõrnimed kirjutatakse lähtedokumendi nimetava-käändelise nimekuju alusel muutmata kujul „Eestis kasutatava ladina tähtede valiku“ [standard EVS 8 kasutab lühendatud nimetust ESET1] abil, rakendades nimede tähestikulisel sortimisel „Eestis kasutatavat ladina tähtede järjestust“. Selle punkti täitmise kindlustamiseks:

1) [...]

2) Riigikantseleil korraldada Eesti standardi EVS 8:1993 „Infotehnoloogia reeglid eesti kultuuri ja keele keskkonnas“ käesoleva määrusega kooskõlla viimine;

3) [...]

4) Haridusministeeriumil koostada ja avaldada käesoleva punkti rakendamise kindlustamiseks teatmaterjal maailma riikide ja nende ametikeelte loeteluga ning keelte ja nende tähestike valikuga.“

3. Täiendada Vabariigi Valitsuse 25. märtsi 1998. a määrust nr 66 järgmise punktiga:

„5. Ladinatähelised võõrnimed, mis sisaldavad haruldasi, Eestis kasutatavasse ladina tähtede valikusse mittekuuluvaid tähti, kantakse Eesti dokumentidesse ja andmekogudesse nimeteadusliku usaldusasutuse ekspertiisi alusel. Isikuandmetele lisatakse sel juhul märkus isikunime täpse kuju kohta.“

Töörühma üldpõhimõtted ja järeldused olid veel:

- Vaatluse alla tuleb võtta need ladinatähelised keeled, mis on dokumendikeeled, s.o sellised, milles koostatakse ja antakse välja nt isikutõendusdokumente (passe, isikutunnistusi), tõendeid, lepingutekste ja muid isiku- või kohanimesid sisaldavaid akte. Dokumendikeeltena talitlevad kõigi riikide riigi- (ameti-)keeled, ent peale selle ka nt regionaaltasandi keeled.
- Mitteladinatähelised keeled tulevad vaatluse alla sedavõrd, kui võrd nende keelte nimed on dokumentides fikseeritud ladina tähtedega.
- Mitteladinatäheliste keelte nimede kandmiseks Eesti registritesse tuleb kõne alla kaks võimalust:
 - a) nimi võetakse vastava dokumendi ladinatähelisest osast muutmata kujul;
 - b) kui võimalus *a* puudub, kirjutatakse nimi ümber ladina tähestikku Eestis kehtivate või rahvusvaheliselt tunnustatud reeglite alusel.Juhul *a* ei tekita nimekirjutus enamasti probleeme, sest sageli kasutatakse sellistes dokumentides lihtsustatud transkriptsiooni. Omaladinaga maade, nt Bulgaaria puhul tuleb ette ka diakriitikutega tähti.

- ESET1 kasutuselevõtt ei tähenda sugugi, et neil üksikjuhtudel, kus tähevalikust ei piisa ladinatähelise isikunime muutmata kujul kirjanekuks, kuulub see automaatselt lihtsustamisele. Andmeregistrisse tuleb kanda ka nime originaalkuju.

Aruandes soovitatud standardi EVS 8 ajakohastamine on juba toimunud, käesolev rakendusjuhend püüab jõudumööda täita teist töörühma esitatud ettepanekut põhjaliku informatiivse materjali koostamise osas. Suur osa siin esitatud materjalist (nt lisad 1, 2 ja 4) on algselt üle võetud töörühma aruandest.

Lisa 4. Eestis kasutatav ladina tähtede valik ESET1

Märkused

1. Kõik tähemärgid on esitatud **suur- ja väiketähe paaridena**. Erandlik on ainult väiketähena kasutatav ß. Ka suurt punktiga i-d ning väikest punktita i-d ei saa rangelt võttes lugeda paariks. Suur- ja väiketähtedel on ühine keelte loend.
2. **Tähtede järjekord** on rakendusi arvestades esitatud rahvusvahelisel kujul, kus kõik diakriitikuga tähed paigutatakse põhitähe järel. Seega eesti õ, ä, ö, ja ü asuvad vastavalt o, a, o ja u järel. Eestikeelsed registrid järjestatakse loomulikult eesti tähestiku alusel. Omaette koht on rahvusvahelises järjestuses islandi tähel thorn (Ð), mis paigutatakse kokkuleppeliselt ladina alustähestiku lõppu.
3. **Ladina alustähestikku kuuluvad tähed** ja lisasümbolid on olemas kõigis kooditabeleis ja klaviatuuridel. Samuti on eelduseks, et need on kasutusel kõigis keeltes. Nende puhul on toodud ainult nimetus ja Unicode'i kood ning nad on esitatud vaid loendi täielikkuse huvides.
4. **Tähtede juures on esitatud** nende nimetus eesti keeles, Unicode'i kood, levinumad kooditabelid, kus see tähemärk leidub koos koodinumbriga selles tabelis ja keeled, kus see tähemärk kasutusel on. **Kaldkriips eraldab** suur- ja väiketähe nimetust, Unicode'i koodi ja koodi kooditabeleis, nt. <00DA> / <00FA> – suurtähe Unicode'i number on kuueteistkümnendarv DA, väiketähel FA.
5. **Eestikeelses nimetuses** võib lihtsustatult ära jätta täiendi 'ladina'. See on vajalik sama tähekuju eristamiseks muus kirjas, nt ladina väike a – kirillitsa väike a.
6. **Kooditabelite loendis** on ISO ja Windowsi kooditabelid paigutatud eraldi lõikudesse v.a juhul, kui tähemärk esineb ainult kahes kooditabelis. ISO 8859 kooditabeleis on kood esitatud kuueteistkümnendarvuna, Windowsi kooditabeleis aga kümnendarvuna, mille ette on lisatud number 0, nt järgnev rida Windows 1257 'Baltic' (0218/0250); annab Windowsi kasutajale juhise tähemärgi s sisestamiseks klaviatuurilt: vasakpoolset Alt-klahvi all hoides valige numbrisõrmistikul kood 0218 (suurtäht) või 0250 (väiketäht). 'Baltic' tähendab, et kasutada tuleb fonti Times New Roman Baltic vm kooditabelile Windows CP 1257 vastavat fonti. Kõiki kooditabeleid saab kasutada Windowsi süsteemsete fontidega Arial, Courier ja Times New Roman.
7. Tärn keele nimetuse järel **keelte loendis** märgib, et see keel ei kasuta ladina kirja ja konkreetne täht on vajalik selle keele latiniseerimiseks. Nende keelte puhul, kus kasutusel on ladina tähtede miinimumist välja jäänud tähemärke, on sulgudes lisatud märgend (puudulik).

A a

nimetus: ladina suur A / ladina väike A
Unicode'i number: <0041> / <0061>

Á á

nimetus: ladina suur A akuudiga / ladina väike A akuudiga
Unicode'i number: <00C1> / <00E1>
leidub kooditabeleis: ISO 8859-1 (C1/E1); 8859-2 (C1/E1); 8859-3 (C1/E1); 8859-4 (C1/E1); 8859-9 (C1/E1); 8859-10 (C1/E1); 8859-14 (C1/E1); 8859-15 (C1/E1); Windows 1250 'CE' (0193/0225); 1252 'Western' (0193/0225); 1254 'Turkic' (0193/0225);
kasutuskeeled: fääri; galeegi; hispaania; hollandi; iiri; islandi; portugali; rootsi; Šoti gaeli; slovaki; taani; tšehhi; ungari;

À à

nimetus: ladina suur A graavisega / ladina väike A graavisega
Unicode'i number: <00C0> / <00E0>
leidub kooditabeleis: ISO 8859-1 (C0/E0); 8859-3 (C0/E0); 8859-9 (C0/E0); 8859-14 (C0/E0); 8859-15 (C0/E0);
Windows 1252 'Western' (0192/0224); 1254 'Turkic' (0192/0224);
kasutuskeeled: itaalia; norra; portugali; prantsuse; retoromaani; Šoti gaeli; saksa;

Ă ă

nimetus: ladina suur A kaarega / ladina väike A kaarega
Unicode'i number: <0102> / <0103>
leidub kooditabeleis: ISO 8859-2 (C3/E3);
Windows 1250 'CE' (0195/0227);
kasutuskeeled: rumeenia;

Â â

nimetus: ladina suur A tsirkumfleksiga / ladina väike A tsirkumfleksiga
Unicode'i number: <00C2> / <00E2>
leidub kooditabeleis: ISO 8859-1 (C2/E2); 8859-2 (C2/E2); 8859-3 (C2/E2); 8859-4 (C2/E2); 8859-9 (C2/E2); 8859-10 (C2/E2); 8859-14 (C2/E2); 8859-15 (C2/E2);
Windows 1250 'CE' (0194/0226); 1252 'Western' (0194/0226); 1254 'Turkic' (0194/0226);
kasutuskeeled: bretooni; friisi; hollandi; letseburgi (luksemburgi); portugali; prantsuse; rumeenia; tšamorro; türgi;

Å å

nimetus: ladina suur A ülasingiga / ladina väike A ülasingiga
Unicode'i number: <00C5> / <00E5>
leidub kooditabeleis: ISO 8859-1 (C5/E5); 8859-4 (C5/E5); 8859-9 (C5/E5); 8859-10 (C5/E5); 8859-13 (C5/E5); 8859-14 (C5/E5); 8859-15 (C5/E5);
Windows 1252 'Western' (0197/0229); 1254 'Turkic' (0197/0229); 1257 'Baltic' (0197/0229);
kasutuskeeled: norra; rootsi; soome; taani;

Ä ä

nimetus: ladina suur A täppidega / ladina väike A täppidega
Unicode'i number: <00C4> / <00E4>
leidub kooditabeleis: ISO 8859-1 (C4/E4); 8859-2 (C4/E4); 8859-3 (C4/E4); 8859-4 (C4/E4); 8859-9 (C4/E4); 8859-10 (C4/E4); 8859-13 (C4/E4); 8859-14 (C4/E4);

8859-15 (C4/E4);
Windows 1250 'CE' (0196/0228); 1252 'Western' (0196/0228); 1254
'Turkic' (0196/0228); 1257 'Baltic' (0196/0228);
kasutuskeeled: aserbaidžaani (puudulik); eesti; friisi; hollandi; japi; letseburgi
(luksemburgi); rootsi; saksa; slovaki; soome;

Ã ã

nimetus: ladina suur A tildega / ladina väike A tildega
Unicode'i number: <00C3> / <00E3>
leidub kooditabeleis: ISO 8859-1 (C3/E3); 8859-4 (C3/E3); 8859-9 (C3/E3); 8859-10 (C3/E3);
8859-14 (C3/E3); 8859-15 (C3/E3);
Windows 1252 'Western' (0195/0227); 1254 'Turkic' (0195/0227);
kasutuskeeled: portugali;

Ą ą

nimetus: ladina suur A pöörsediiga / ladina väike A pöörsediiga
Unicode'i number: <0104> / <0105>
leidub kooditabeleis: ISO 8859-2 (A1/B1); 8859-4 (A1/B1); 8859-10 (A1/B1); 8859-13
(C0/E0);
Windows 1250 'CE' (0165/0185); 1257 'Baltic' (0192/0224);
kasutuskeeled: leedu; poola;

Ā ā

nimetus: ladina suur A ülakriipsuga / ladina väike A ülakriipsuga
Unicode'i number: <0100> / <0101>
leidub kooditabeleis: ISO 8859-4 (C0/E0); 8859-10 (C0/E0); 8859-13 (C2/E2);
Windows 1257 'Baltic' (0194/0226);
kasutuskeeled: hawaii; läti; maoori;

Æ æ

nimetus: ladina suur AE / ladina väike AE
Unicode'i number: <00C6> / <00E6>
leidub kooditabeleis: ISO 8859-1 (C6/E6); 8859-4 (C6/E6); 8859-9 (C6/E6); 8859-10 (C6/E6);
8859-13 (AF/BF); 8859-14 (C6/E6); 8859-15 (C6/E6);
Windows 1252 'Western' (0198/0230); 1254 'Turkic' (0198/0230); 1257
'Baltic' (0175/0191);
kasutuskeeled: fääri; inglise; islandi; norra; prantsuse; taani;

B b

nimetus: ladina suur B / ladina väike B
Unicode'i number: <0042> / <0062>

C c

nimetus: ladina suur C / ladina väike C
Unicode'i number: <0043> / <0063>

Ć ć

nimetus: ladina suur C akuudiga / ladina väike C akuudiga
Unicode'i number: <0106> / <0107>
leidub kooditabeleis: ISO 8859-2 (C6/E6); 8859-13 (C3/E3);
Windows 1250 'CE' (0198/0230); 1257 'Baltic' (0195/0227);
kasutuskeeled: alamsorbi; bosnia; horvaadi; makedoonia*; poola; serbia* (puudulik);
ülemsorbi;

Č č

nimetus: ladina suur C haagiga / ladina väike C haagiga
Unicode'i number: <010C> / <010D>
leidub kooditabeleis: ISO 8859-2 (C8/E8); 8859-4 (C8/E8); 8859-10 (C8/E8); 8859-13 (C8/E8);
Windows 1250 'CE' (0200/0232); 1257 'Baltic' (0200/0232);
kasutuskeeled: alamsorbi; bosnia; bulgaaria* (puudulik); horvaadi; leedu; läti;
makedoonia*; serbia* (puudulik); slovaki; sloveeni; tšehhi; vene*
(puudulik); ülemsorbi;

Ç ç

nimetus: ladina suur C sediiga / ladina väike C sediiga
Unicode'i number: <00C7> / <00E7>
leidub kooditabeleis: ISO 8859-1 (C7/E7); 8859-2 (C7/E7); 8859-3 (C7/E7); 8859-9 (C7/E7);
8859-14 (C7/E7); 8859-15 (C7/E7);
Windows 1250 'CE' (0199/0231); 1252 'Western' (0199/0231); 1254
'Turkic' (0199/0231);
kasutuskeeled: albaania; aserbaidžaaani (puudulik); baski; hispaania; inglise; portugali;
prantsuse; türgi;

D d

nimetus: ladina suur D / ladina väike D
Unicode'i number: <0044> / <0064>

Ð ð

nimetus: ladina suur ETH / ladina väike ETH
Unicode'i number: <00D0> / <00F0>
leidub kooditabeleis: ISO 8859-1 (D0/F0); 8859-10 (D0/F0); 8859-15 (D0/F0);
Windows 1252 'Western' (0208/0240);
kasutuskeeled: fääri; islandi;

Ď ě

nimetus: ladina suur D haagiga / ladina väike D haagiga
Unicode'i number: <010E> / <010F>
leidub kooditabeleis: ISO 8859-2 (CF/EF); Windows 1250 'CE' (0207/0239);
kasutuskeeled: slovaki; tšehhi;

Đ đ

nimetus: ladina suur D läbiva kriipsuga / ladina väike D läbiva kriipsuga
Unicode'i number: <0110> / <0111>
leidub kooditabeleis: ISO 8859-2 (D0/F0); 8859-4 (D0/F0); 8859-10 (A9/B9);
Windows 1250 'CE' (0208/0240);
kasutuskeeled: bosnia; horvaadi; makedoonia*; serbia* (puudulik);

E e

nimetus: ladina suur E / ladina väike E
Unicode'i number: <0045> / <0065>

É é

nimetus: ladina suur E akuudiga / ladina väike E akuudiga
Unicode'i number: <00C9> / <00E9>
leidub kooditabeleis: ISO 8859-1 (C9/E9); 8859-2 (C9/E9); 8859-3 (C9/E9); 8859-4 (C9/E9);
8859-9 (C9/E9); 8859-10 (C9/E9); 8859-13 (C9/E9); 8859-14 (C9/E9);
8859-15 (C9/E9);

kasutuskeeled: Windows 1250 'CE' (0201/0233); 1252 'Western' (0201/0233); 1254 'Turkic' (0201/0233); 1257 'Baltic' (0201/0233); afrikaani; friisi; galeegi; heebrea* (puudulik); hispaania; hollandi; iiri; indoneesia; islandi; itaalia; letseburgi (luksemburgi); norra; portugali; prantsuse; retoromaani; rootsi; Šoti gaeli; saksa; slovaki; taani; tšehhi; ungari;

È è

nimetus: ladina suur E graavisega / ladina väike E graavisega
Unicode'i number: <00C8> / <00E8>
leidub kooditabeleis: ISO 8859-1 (C8/E8); 8859-3 (C8/E8); 8859-9 (C8/E8); 8859-14 (C8/E8); 8859-15 (C8/E8);
kasutuskeeled: Windows 1252 'Western' (0200/0232); 1254 'Turkic' (0200/0232); afrikaani; hollandi; itaalia; letseburgi (luksemburgi); norra; portugali; prantsuse; retoromaani; Šoti gaeli; vene* (puudulik);

Ê ê

nimetus: ladina suur E tsirkumfleksiga / ladina väike E tsirkumfleksiga
Unicode'i number: <00CA> / <00EA>
leidub kooditabeleis: ISO 8859-1 (CA/EA); 8859-3 (CA/EA); 8859-9 (CA/EA); 8859-14 (CA/EA); 8859-15 (CA/EA);
kasutuskeeled: Windows 1252 'Western' (0202/0234); 1254 'Turkic' (0202/0234); afrikaani; bretooni; friisi; hollandi; letseburgi (luksemburgi); norra; portugali; prantsuse; tsvana;

Ě ě

nimetus: ladina suur E haagiga / ladina väike E haagiga
Unicode'i number: <011A> / <011B>
leidub kooditabeleis: ISO 8859-2 (CC/EC); Windows 1250 'CE' (0204/0236);
kasutuskeeled: alamsorbi; tšehhi; ülemsorbi;

Ë ë

nimetus: ladina suur E täppidega / ladina väike E täppidega
Unicode'i number: <00CB> / <00EB>
leidub kooditabeleis: ISO 8859-1 (CB/EB); 8859-2 (CB/EB); 8859-3 (CB/EB); 8859-4 (CB/EB); 8859-9 (CB/EB); 8859-10 (CB/EB); 8859-14 (CB/EB); 8859-15 (CB/EB);
kasutuskeeled: Windows 1250 'CE' (0203/0235); 1252 'Western' (0203/0235); 1254 'Turkic' (0203/0235); afrikaani; albaania; friisi; hollandi; japi; letseburgi (luksemburgi); prantsuse; rootsi; vene* (puudulik);

Ē ē

nimetus: ladina suur E ülapanktiga / ladina väike E ülapanktiga
Unicode'i number: <0116> / <0117>
leidub kooditabeleis: ISO 8859-4 (CC/EC); 8859-10 (CC/EC); 8859-13 (CB/EB);
Windows 1257 'Baltic' (0203/0235);
kasutuskeeled: leedu;

Ē ē

nimetus: ladina suur E pöördseidiiga / ladina väike E pöördseidiiga
Unicode'i number: <0118> / <0119>
leidub kooditabeleis: ISO 8859-2 (CA/EA); 8859-4 (CA/EA); 8859-10 (CA/EA); 8859-13
(C6/E6);
Windows 1250 'CE' (0202/0234); 1257 'Baltic' (0198/0230);
kasutuskeeled: leedu; poola;

Ē ē

nimetus: ladina suur E ülakriipsuga / ladina väike E ülakriipsuga
Unicode'i number: <0112> / <0113>
leidub kooditabeleis: ISO 8859-4 (AA/BA); 8859-10 (A2/B2); 8859-13 (C7/E7);
Windows 1257 'Baltic' (0199/0231);
kasutuskeeled: havai; läti; maoori;

F f

nimetus: ladina suur F / ladina väike F
Unicode'i number: <0046> / <0066>

G g

nimetus: ladina suur G / ladina väike G
Unicode'i number: <0047> / <0067>

Ģ ģ

nimetus: ladina suur G kaarega / ladina väike G kaarega
Unicode'i number: <011E> / <011F>
leidub kooditabeleis: ISO 8859-3 (AB/BB); 8859-9 (D0/F0);
Windows 1254 'Turkic' (0208/0240);
kasutuskeeled: aserbaidžaani (puudulik); türgi;

Ģ ģ

nimetus: ladina suur G sediiga / ladina väike G sediiga
Unicode'i number: <0122> / <0123>
leidub kooditabeleis: ISO 8859-4 (AB/BB); 8859-10 (A3/B3); 8859-13 (CC/EC);
Windows 1257 'Baltic' (0204/0236);
kasutuskeeled: läti;

H h

nimetus: ladina suur H / ladina väike H
Unicode'i number: <0048> / <0068>

I i

nimetus: ladina suur I / ladina väike I
Unicode'i number: <0049> / <0069>

İ ı

nimetus: ladina suur I ülapunktiga / ladina väike punktita I
Unicode'i number: <0130> / <0131>
leidub kooditabeleis: ISO 8859-3 (A9/B9); 8859-9 (DD/FD);
Windows 1254 'Turkic' (0221/0253);
kasutuskeeled: aserbaidžaani (puudulik); türgi;

Í í

nimetus: ladina suur I akuudiga / ladina väike I akuudiga
Unicode'i number: <00CD> / <00ED>
leidub kooditabeleis: ISO 8859-1 (CD/ED); 8859-2 (CD/ED); 8859-3 (CD/ED); 8859-4
(CD/ED); 8859-9 (CD/ED); 8859-10 (CD/ED); 8859-14 (CD/ED); 8859-
15 (CD/ED);

Windows 1250 'CE' (0205/0237); 1252 'Western' (0205/0237); 1254 'Turkic' (0205/0237);
kasutuskeeled: fääri; galeegi; hispaania; hollandi; iiri; islandi; itaalia; portugali; slovaki; taani; tšehhi; ungari;

Ì ì

nimetus: ladina suur I graavisega / ladina väike I graavisega
Unicode'i number: <00CC> / <00EC>
leidub kooditabeleis: ISO 8859-1 (CC/EC); 8859-3 (CC/EC); 8859-9 (CC/EC); 8859-14 (CC/EC); 8859-15 (CC/EC);
Windows 1252 'Western' (0204/0236); 1254 'Turkic' (0204/0236);
kasutuskeeled: itaalia; retoromaani; Šoti gaeli;

Î î

nimetus: ladina suur I tsirkumfleksiga / ladina väike I tsirkumfleksiga
Unicode'i number: <00CE> / <00EE>
leidub kooditabeleis: ISO 8859-1 (CE/EE); 8859-2 (CE/EE); 8859-3 (CE/EE); 8859-4 (CE/EE); 8859-9 (CE/EE); 8859-10 (CE/EE); 8859-14 (CE/EE); 8859-15 (CE/EE);
Windows 1250 'CE' (0206/0238); 1252 'Western' (0206/0238); 1254 'Turkic' (0206/0238);
kasutuskeeled: afrikaani; friisi; letseburgi (luksemburgi); prantsuse; retoromaani; rumeenia; türgi;

İ i

nimetus: ladina suur I täppidega / ladina väike I täppidega
Unicode'i number: <00CF> / <00EF>
leidub kooditabeleis: ISO 8859-1 (CF/EF); 8859-3 (CF/EF); 8859-9 (CF/EF); 8859-10 (CF/EF); 8859-14 (CF/EF); 8859-15 (CF/EF);
Windows 1252 'Western' (0207/0239); 1254 'Turkic' (0207/0239);
kasutuskeeled: afrikaani; friisi; hollandi; inglise; itaalia; prantsuse;

Ī ī

nimetus: ladina suur I pöördsemiiga / ladina väike I pöördsemiiga
Unicode'i number: <012E> / <012F>
leidub kooditabeleis: ISO 8859-4 (C7/E7); 8859-10 (C7/E7); 8859-13 (C1/E1);
Windows 1257 'Baltic' (0193/0225);
kasutuskeeled: leedu;

I i

nimetus: ladina suur I ülakriipsuga / ladina väike I ülakriipsuga
Unicode'i number: <012A> / <012B>
leidub kooditabeleis: ISO 8859-4 (CF/EF); 8859-10 (A4/B4); 8859-13 (CE/EE);
Windows 1257 'Baltic' (0206/0238);
kasutuskeeled: havai; läti; maoori;

J j

nimetus: ladina suur J / ladina väike J
Unicode'i number: <004A> / <006A>

K k

nimetus: ladina suur K / ladina väike K
Unicode'i number: <004B> / <006B>

Ķ ķ

nimetus: ladina suur K sediiga / ladina väike K sediiga
Unicode'i number: <0136> / <0137>
leidub kooditabeleis: ISO 8859-4 (D3/F3); 8859-10 (A6/B6); 8859-13 (CD/ED);
Windows 1257 'Baltic' (0205/0237);
kasutuskeeled: läti;

L l

nimetus: ladina suur L / ladina väike L
Unicode'i number: <004C> / <006C>

Ľ ľ

nimetus: ladina suur L akuudiga / ladina väike L akuudiga
Unicode'i number: <0139> / <013A>
leidub kooditabeleis: ISO 8859-2 (C5/E5); Windows 1250 'CE' (0197/0229);
kasutuskeeled: slovaki;

L l

nimetus: ladina suur L haagiga / ladina väike L haagiga
Unicode'i number: <013D> / <013E>
leidub kooditabeleis: ISO 8859-2 (A5/B5); Windows 1250 'CE' (0188/0190);
kasutuskeeled: slovaki;

Ł ł

nimetus: ladina suur L läbiva kriipsuga / ladina väike L läbiva kriipsuga
Unicode'i number: <0141> / <0142>
leidub kooditabeleis: ISO 8859-2 (A3/B3); 8859-13 (D9/F9);
Windows 1250 'CE' (0163/0179); Windows 1257 'Baltic' (0217/0249);
kasutuskeeled: alamsorbi; poola; ülemsorbi;

Ļ ļ

nimetus: ladina suur L sediiga / ladina väike L sediiga
Unicode'i number: <013B> / <013C>
leidub kooditabeleis: ISO 8859-4 (A6/B6); 8859-10 (A8/B8); 8859-13 (CF/EF);
Windows 1257 'Baltic' (0207/0239);
kasutuskeeled: läti;

M m

nimetus: ladina suur M / ladina väike M
Unicode'i number: <004D> / <006D>

N n

nimetus: ladina suur N / ladina väike N
Unicode'i number: <004E> / <006E>

Ń ń

nimetus: ladina suur N akuudiga / ladina väike N akuudiga
Unicode'i number: <0143> / <0144>
leidub kooditabeleis: ISO 8859-2 (D1/F1); 8859-13 (D1/F1);
Windows 1250 'CE' (0209/0241); 1257 'Baltic' (0209/0241);
kasutuskeeled: alamsorbi; poola; ülemsorbi;

Ň ň

nimetus: ladina suur N haagiga / ladina väike N haagiga
Unicode'i number: <0147> / <0148>
leidub kooditabeleis: ISO 8859-2 (D2/F2); Windows 1250 'CE' (0210/0242);
kasutuskeeled: slovaki; tšehhi;

Ñ ñ

nimetus: ladina suur N tildega / ladina väike N tildega
Unicode'i number: <00D1> / <00F1>
leidub kooditabeleis: ISO 8859-1 (D1/F1); 8859-3 (D1/F1); 8859-9 (D1/F1); 8859-14 (D1/F1);
8859-15 (D1/F1);
Windows 1252 'Western' (0209/0241); 1254 'Turkic' (0209/0241);
kasutuskeeled: baski; bretooni; galeegi; hispaania; tšamorro;

Ņ ņ

nimetus: ladina suur N sediiga / ladina väike N sediiga
Unicode'i number: <0145> / <0146>
leidub kooditabeleis: ISO 8859-4 (D1/F1); 8859-10 (D1/F1); 8859-13 (D2/F2);
Windows 1257 'Baltic' (0210/0242);
kasutuskeeled: läti;

O o

nimetus: ladina suur O / ladina väike O
Unicode'i number: <004F> / <006F>

Ó ó

nimetus: ladina suur O akuudiga / ladina väike O akuudiga
Unicode'i number: <00D3> / <00F3>
leidub kooditabeleis: ISO 8859-1 (D3/F3); 8859-2 (D3/F3); 8859-3 (D3/F3); 8859-9 (D3/F3);
8859-10 (D3/F3); 8859-13 (D3/F3); 8859-14 (D3/F3); 8859-15 (D3/F3);
Windows 1250 'CE' (0211/0243); 1252 'Western' (0211/0243); 1254
'Turkic' (0211/0243); 1257 'Baltic' (0211/0243);
kasutuskeeled: alamsorbi; fääri; galeegi; hispaania; hollandi; iiri; islandi; itaalia; norra;
poola; portugali; Šoti gaeli; slovaki; taani; tšehhi; ungari;

Ò ò

nimetus: ladina suur O graavisega / ladina väike O graavisega
Unicode'i number: <00D2> / <00F2>
leidub kooditabeleis: ISO 8859-1 (D2/F2); 8859-3 (D2/F2); 8859-9 (D2/F2); 8859-14 (D2/F2); 8859-15 (D2/F2);
Windows 1252 'Western' (0210/0242); 1254 'Turkic' (0210/0242);
kasutuskeeled: itaalia; norra; portugali; retoromaani; Šoti gaeli;

Ô ô

nimetus: ladina suur O tsirkumfleksiga / ladina väike O tsirkumfleksiga
Unicode'i number: <00D4> / <00F4>
leidub kooditabeleis: ISO 8859-1 (D4/F4); 8859-2 (D4/F4); 8859-3 (D4/F4); 8859-4 (D4/F4); 8859-9 (D4/F4); 8859-10 (D4/F4); 8859-14 (D4/F4); 8859-15 (D4/F4);
Windows 1250 'CE' (0212/0244); 1252 'Western' (0212/0244); 1254 'Turkic' (0212/0244);
kasutuskeeled: afrikaani; bretooni; friisi; hollandi; inglise; jaapani*; letseburgi (luksemburgi); norra; portugali; prantsuse; slovaki; tsvana;

Ö ö

nimetus: ladina suur O täppidega / ladina väike O täppidega
Unicode'i number: <00D6> / <00F6>
leidub kooditabeleis: ISO 8859-1 (D6/F6); 8859-2 (D6/F6); 8859-3 (D6/F6); 8859-4 (D6/F6); 8859-9 (D6/F6); 8859-10 (D6/F6); 8859-13 (D6/F6); 8859-14 (D6/F6); 8859-15 (D6/F6);
Windows 1250 'CE' (0214/0246); 1252 'Western' (0214/0246); 1254 'Turkic' (0214/0246); 1257 'Baltic' (0214/0246);
kasutuskeeled: aserbaidžaani (puudulik); eesti; friisi; hollandi; islandi; japi; letseburgi (luksemburgi); rootsi; saksa; soome; türgi; ungari;

Ő ő

nimetus: ladina suur O topeltakuudiga / ladina väike O topeltakuudiga
Unicode'i number: <0150> / <0151>
leidub kooditabeleis: ISO 8859-2 (D5/F5); Windows 1250 'CE' (0213/0245);
kasutuskeeled: ungari;

Ō ō

nimetus: ladina suur O tildega / ladina väike O tildega
Unicode'i number: <00D5> / <00F5>
leidub kooditabeleis: ISO 8859-1 (D5/F5); 8859-4 (D5/F5); 8859-9 (D5/F5); 8859-10 (D5/F5); 8859-13 (D5/F5); 8859-14 (D5/F5); 8859-15 (D5/F5);
Windows 1252 'Western' (0213/0245); 1254 'Turkic' (0213/0245); 1257 'Baltic' (0213/0245);
kasutuskeeled: eesti; portugali;

Ø ø

nimetus: ladina suur O läbiva kaldkriipsuga / ladina väike O läbiva kaldkriipsuga
Unicode'i number: <00D8> / <00F8>
leidub kooditabeleis: ISO 8859-1 (D8/F8); 8859-4 (D8/F8); 8859-9 (D8/F8); 8859-10 (D8/F8); 8859-13 (A8/B8); 8859-14 (D8/F8); 8859-15 (D8/F8);
Windows 1252 'Western' (0216/0248); 1254 'Turkic' (0216/0248); 1257 'Baltic' (0168/0184);
kasutuskeeled: fääri; norra; taani;

Ō ō

nimetus: ladina suur O ülakriipsuga / ladina väike O ülakriipsuga
Unicode'i number: <014C> / <014D>
leidub kooditabeleis: ISO 8859-4 (D2/F2); 8859-10 (D2/F2); 8859-13 (D4/F4);
Windows 1257 'Baltic' (0212/0244);
kasutuskeeled: havai; jaapani*; läti; maoori;

Œ œ

nimetus: ladina suur ligatuur OE / ladina väike ligatuur OE
Unicode'i number: <0152> / <0153>
leidub kooditabeleis: ISO 8859-15 (BC/BD);
Windows 1252 'Western' (0140/0156); 1254 'Turkic' (0140/0156); CP1256 (0140/0156);
kasutuskeeled: prantsuse;

P p

nimetus: ladina suur P / ladina väike P
Unicode'i number: <0050> / <0070>

Q q

nimetus: ladina suur Q / ladina väike Q
Unicode'i number: <0051> / <0071>

R r

nimetus: ladina suur R / ladina väike R
Unicode'i number: <0052> / <0072>

Ř ř

nimetus: ladina suur R akuudiga / ladina väike R akuudiga
Unicode'i number: <0154> / <0155>
leidub kooditabeleis: ISO 8859-2 (C0/E0); Windows 1250 'CE' (0192/0224);
kasutuskeeled: slovaki; ülemsorbi;

Ř ř

nimetus: ladina suur R haagiga / ladina väike R haagiga
Unicode'i number: <0158> / <0159>
leidub kooditabeleis: ISO 8859-2 (D8/F8); Windows 1250 'CE' (0216/0248);
kasutuskeeled: alamsorbi; tšehhi;

Ŕ ŕ

nimetus: ladina suur R sediiga / ladina väike R sediiga
Unicode'i number: <0156> / <0157>
leidub kooditabeleis: ISO 8859-4 (A3/B3); 8859-13 (AA/BA);
Windows 1257 'Baltic' (0170/0186);
kasutuskeeled: läti;

S s

nimetus: ladina suur S / ladina väike S
Unicode'i number: <0053> / <0073>

Š š

nimetus: ladina suur S akuudiga / ladina väike S akuudiga
Unicode'i number: <015A> / <015B>
leidub kooditabeleis: ISO 8859-2 (A6/B6); 8859-13 (DA/FA);
Windows 1250 'CE' (0140/0156); 1257 'Baltic' (0218/0250);
kasutuskeeled: poola; ülemsorbi;

Š š

nimetus: ladina suur S haagiga / ladina väike S haagiga
Unicode'i number: <0160> / <0161>
leidub kooditabeleis: ISO 8859-2 (A9/B9); 8859-4 (A9/B9); 8859-10 (AA/BA); 8859-13 (D0/F0); 8859-15 (A6/A8);
Windows 1250 'CE' (0138/0154); 1252 'Western' (0138/0154); 1254 'Turkic' (0138/0154); 1257 'Baltic' (0208/0240);
kasutuskeeled: alamsorbi; bosnia; bulgaaria* (puudulik); eesti; horvaadi; leedu; läti; makedoonia*; serbia* (puudulik); slovaki; sloveeni; soome; tšehhi; vene* (puudulik); ülemsorbi;

Ș ș

nimetus: ladina suur S sediiga / ladina väike S sediiga
Unicode'i number: <015E> / <015F>
leidub kooditabeleis: ISO 8859-2 (AA/BA); 8859-3 (AA/BA); 8859-9 (DE/FE);
Windows 1250 'CE' (0170/0186); 1254 'Turkic' (0222/0254);
kasutuskeeled: aserbaidžaani (puudulik); rumeenia; türgi;

ß

nimetus: ladina väike SZ
Unicode'i number: <00DF>
leidub kooditabeleis: ISO 8859-1 (DF); 8859-2 (DF); 8859-3 (DF); 8859-4 (DF); 8859-9 (DF);
8859-10 (DF); 8859-13 (DF); 8859-14 (DF); 8859-15 (DF);
Windows 1250 'CE' (0223); 1252 'Western' (0223); 1254 'Turkic' (0223);
1257 'Baltic' (0223);
kasutuskeeled: letseburgi (luksemburgi); saksa;

T t

nimetus: ladina suur T / ladina väike T
Unicode'i number: <0054> / <0074>

Ť t'

nimetus: ladina suur T haagiga / ladina väike T haagiga
Unicode'i number: <0164> / <0165>
leidub kooditabeleis: ISO 8859-2 (AB/BB); Windows 1250 'CE' (0141/0157);
kasutuskeeled: slovaki; tšehhi;

Ț ț

nimetus: ladina suur T sediiiga / ladina väike T sediiiga
Unicode'i number: <0162> / <0163>
leidub kooditabeleis: ISO 8859-2 (DE/FE); Windows 1250 'CE' (0222/0254);
kasutuskeeled: rumeenia;

Ū ū

nimetus: ladina suur U / ladina väike U
Unicode'i number: <0055> / <0075>

Ú ú

nimetus: ladina suur U akuudiga / ladina väike U akuudiga
Unicode'i number: <00DA> / <00FA>
leidub kooditabeleis: ISO 8859-1 (DA/FA); 8859-2 (DA/FA); 8859-3 (DA/FA); 8859-4 (DA/FA); 8859-9 (DA/FA); 8859-10 (DA/FA); 8859-14 (DA/FA); 8859-15 (DA/FA);
Windows 1250 'CE' (0218/0250); 1252 'Western' (0218/0250); 1254 'Turkic' (0218/0250);
kasutuskeeled: friisi; fääri; galeegi; hispaania; hollandi; iiri; islandi; itaalia; portugali; slovaki; taani; tšehhi; ungari;

Ù ù

nimetus: ladina suur U graavisega / ladina väike U graavisega
Unicode'i number: <00D9> / <00F9>
leidub kooditabeleis: ISO 8859-1 (D9/F9); 8859-3 (D9/F9); 8859-9 (D9/F9); 8859-14 (D9/F9); 8859-15 (D9/F9);
Windows 1252 'Western' (0217/0249); 1254 'Turkic' (0217/0249);
kasutuskeeled: bretooni; itaalia; prantsuse; retoromaani; Šoti gaeli;

Ū ū

nimetus: ladina suur U tsirkumfleksiga / ladina väike U tsirkumfleksiga
Unicode'i number: <00DB> / <00FB>
leidub kooditabeleis: ISO 8859-1 (DB/FB); 8859-3 (DB/FB); 8859-4 (DB/FB); 8859-9 (DB/FB); 8859-10 (DB/FB); 8859-14 (DB/FB); 8859-15 (DB/FB); Windows 1252 'Western' (0219/0251); 1254 'Turkic' (0219/0251);
kasutuskeeled: afrikaani; friisi; hollandi; jaapani*; letseburgi (luksemburgi); prantsuse; türgi;

Ŭ ŭ

nimetus: ladina suur U ülasõõriga / ladina väike U ülasõõriga
Unicode'i number: <016E> / <016F>
leidub kooditabeleis: ISO 8859-2 (D9/F9); Windows 1250 'CE' (0217/0249);
kasutuskeeled: tšehhi;

Û ü

nimetus: ladina suur U täppidega / ladina väike U täppidega
Unicode'i number: <00DC> / <00FC>
leidub kooditabeleis: ISO 8859-1 (DC/FC); 8859-2 (DC/FC); 8859-3 (DC/FC); 8859-4 (DC/FC); 8859-9 (DC/FC); 8859-10 (DC/FC); 8859-13 (DC/FC); 8859-14 (DC/FC); 8859-15 (DC/FC); Windows 1250 'CE' (0220/0252); 1252 'Western' (0220/0252); 1254 'Turkic' (0220/0252); 1257 'Baltic' (0220/0252);
kasutuskeeled: aserbaidžaani (puudulik); baski; bretooni; eesti; friisi; galeegi; hispaania; hollandi; letseburgi (luksemburgi); portugali; prantsuse; rootsi; saksa; tšamorro; türgi; ungari;

Ú ú

nimetus: ladina suur U topeltakuudiga / ladina väike U topeltakuudiga
Unicode'i number: <0170> / <0171>
leidub kooditabeleis: ISO 8859-2 (DB/FB); Windows 1250 'CE' (0219/0251);
kasutuskeeled: ungari;

Ū ū

nimetus: ladina suur U pöörsediiga / ladina väike U pöörsediiga
Unicode'i number: <0172> / <0173>
leidub kooditabeleis: ISO 8859-4 (D9/F9); 8859-10 (D9/F9); 8859-13 (D8/F8); Windows 1257 'Baltic' (0216/0248);
kasutuskeeled: leedu;

Ū ū

nimetus: ladina suur U ülakriipsuga / ladina väike U ülakriipsuga
Unicode'i number: <016A> / <016B>
leidub kooditabeleis: ISO 8859-4 (DE/FE); 8859-10 (AE/BE); 8859-13 (DB/FB);
Windows 1257 'Baltic' (0219/0251);
kasutuskeeled: havai; jaapani*; leedu; läti; maoori;

V v

nimetus: ladina suur V / ladina väike V
Unicode'i number: <0056> / <0076>

W w

nimetus: ladina suur W / ladina väike W
Unicode'i number: <0057> / <0077>

X x

nimetus: ladina suur X / ladina väike X
Unicode'i number: <0058> / <0078>

Y y

nimetus: ladina suur Y / ladina väike Y
Unicode'i number: <0059> / <0079>

Ý ý

nimetus: ladina suur Y akuudiga / ladina väike Y akuudiga
Unicode'i number: <00DD> / <00FD>
leidub kooditabeleis: ISO 8859-1 (DD/FD); 8859-2 (DD/FD); 8859-10 (DD/FD); 8859-14
(DD/FD); 8859-15 (DD/FD);
Windows 1250 'CE' (0221/0253); 1252 'Western' (0221/0253);
kasutuskeeled: fääri; islandi; slovaki; taani; tšehhi;

ÿ ÿ

nimetus: ladina suur Y täppidega / ladina väike Y täppidega
Unicode'i number: <0178> / <00FF>
leidub kooditabeleis: ISO 8859-14 (AF/FF); 8859-15 (BE/FF);
Windows 1252 'Western' (0159/0255); 1254 'Turkic' (0159/0255);
kasutuskeeled: prantsuse;

Z z

nimetus: ladina suur Z / ladina väike Z
Unicode'i number: <005A> / <007A>

Ž ž

nimetus: ladina suur Z akuudiga / ladina väike Z akuudiga
Unicode'i number: <0179> / <017A>
leidub kooditabeleis: ISO 8859-2 (AC/BC); 8859-13 (CA/EA);
Windows 1250 'CE' (0143/0159); 1257 'Baltic' (0202/0234);
kasutuskeeled: alamsorbi; poola; ülemsorbi;

Ž ž

nimetus: ladina suur Z haagiga / ladina väike Z haagiga
Unicode'i number: <017D> / <017E>
leidub kooditabeleis: ISO 8859-2 (AE/BE); 8859-4 (AE/BE); 8859-10 (AC/BC); 8859-13
(DE/FE); 8859-15 (B4/B8);
Windows 1250 'CE' (0142/0158); 1252 'Western' (0142/0158); 1257
'Baltic' (0222/0254);
kasutuskeeled: alamsorbi; bosnia; bulgaaria* (puudulik); eesti; horvaadi; leedu; läti;
makedoonia*; serbia* (puudulik); slovaki; sloveeni; soome; tšehhi; vene*
(puudulik); ülemsorbi;

Ž ž

nimetus: ladina suur Z ülakupunktiga / ladina väike Z ülakupunktiga
Unicode'i number: <017B> / <017C>
leidub kooditabeleis: ISO 8859-2 (AF/BF); 8859-3 (AF/BF); 8859-13 (DD/FD);
Windows 1250 'CE' (0175/0191); 1257 'Baltic' (0221/0253);
kasutuskeeled: poola;

Þ þ

nimetus:	ladina suur THORN / ladina väike THORN
Unicode'i number:	<00DE> / <00FE>
leidub kooditabeleis:	ISO 8859-1 (DE/FE); 8859-10 (DE/FE); 8859-15 (DE/FE); Windows 1252 'Western' (0222/0254);
kasutuskeeled:	islandi;

Muud isiku- ja kohanimeses esinevad sümbolid

!'

nimetus:	apostroof e. ülakoma
Unicode'i number:	<0027>
kommentaar:	Keeled, kus ülakoma võib sõnades esineda on: afrikaani, eesti, hollandi, iiri, inglise, itaalia, katalaani, norra, prantsuse, rumeenia, saksa, slovaki, sloveeni, taani. Eesti näide: ole'nd mõistlik; inglise: let's rock'n'roll. Ülakomaga tähistatakse ka pehmendumärki kirillitsa latiniseerimisel vene, ukraina, bulgaaria jt. keeltes.

!'

nimetus:	jutumärgid
Unicode'i number:	<0022>
kommentaar:	Jutumärgidega tähistatakse kirillitsa (vene) kõvendumärki selle translitereerimisel.

Koostamispõhimõtted

Ladina tähtede valik koosneb 199 tähemärgist ja kahest lisamärgist ning katab suurema osa Euroopa ametikeeltest. Kõik valikusse kuuluvad tähemärgid leiduvad tähevalikus Windows Glyph List 4 (WGL4) ja vähemalt ühes ISO 8859-seeria kooditabelis ning seega tehnilised raskused (sisestamine klaviatuurilt, fondid ja printimine) ESETi koheseks käibelevõtuks on minimaalsed. Valikule lähim Euroopa standard EN 1923 LL8 sisaldab veel järgnevaid tähemärke:

nimetus eesti keeles	Unicode'i koodid	kasutuskeel
Suur ja väike C tsirkumfleksiga	<0108> / <0109>	esperanto
Suur ja väike C ülapunktiga	<010A> / <010B>	malta
Suur ja väike G tsirkumfleksiga	<011C> / <011D>	esperanto
Suur ja väike G ülapunktiga	<0120> / <0121>	malta
Suur ja väike H tsirkumfleksiga	<0124> / <0125>	esperanto
Suur ja väike H läbiva kriipsuga	<0126> / <0127>	malta
Suur ja väike I tildega	<0128> / <0129>	grööni (vana ortograafia)
Suur ja väike ligatuur IJ	<0132> / <0133>	hollandi (soovitav kirjutada eraldi I ja J)
Suur ja väike J tsirkumfleksiga	<0134> / <0135>	esperanto
Väike KRA	<0138>	grööni (vana ortograafia)
Suur ja väike L keskpunktiga	<013F> / <0140>	katalaani
Väike N eelneva apostroofiga	<0149>	afrikaani (soovitav kirjutada eraldi ' ja n)
Suur ja väike ENG	<014A> / <014B>	saami
Suur ja väike S tsirkumfleksiga	<015C> / <015D>	esperanto
Suur ja väike T läbiva kriipsuga	<0166> / <0167>	saami
Suur ja väike U tildega	<0168> / <0169>	grööni (vana ortograafia)
Suur ja väike U kaarega	<016C> / <016D>	esperanto
Suur ja väike W tsirkumfleksiga	<0174> / <0175>	kõmri
Suur ja väike Y tsirkumfleksiga	<0176> / <0177>	kõmri

Ka EN 1923 LL8 ei kata kõiki Euroopa keeli, kus on kasutusel ladina kiri, välja jäävad nt liivi, saami ja kõmri keel. Kõiki ladina kirjal põhinevaid tähemärke on Unicode'is registreeritud kokku üle tuhande, kuid suuremat osa neist aktiivselt ei kasutata. Esimese piiranguga jäid ESETist välja vähetuntud transkriptsioonisüsteemides, vananenud ortograafiates ja muuks otstarbeks (nt hääldusmärkidena) kasutatavad ladina tähed. Valik ei kata ka neid keeli, mis pole dokumendi-keeled (kõmri), muid Euroopa kirju (kreeka, vene, gruusia, armeenia, heebrea) ja väljaspool Euroopat kasutatavaid ladina tähemärke (vietnami keel ja paljud Aafrika keeled).

Lisa 5. Valik enamlevinud kooditabeleid

Kooditabelite esitusviis on ühtlustatud – kõigis järgnevais kooditabelites

- Kõigi siintoodud kooditabelite vasak pool (koodivahemik 0–127) kattub **ASCII** kodeeritud märgistikuga;
- vahemikus 0–31 ja 127 asuvad juhtkoodid (märgivad nt reavahetust või tabulatsiooni) ning see vahemik — kaks esimest veergu — on joonistelt üleni välja jäetud;
- lisaks neile on ISO kooditabeleis vahemik 128–159 — tabeli teise poole kaks esimest veergu — reserveeritud juhtkoodidele ning selles vahemikus graafilisi märke ei ole;
- iga märgi all leiate tema kümnendkoodi, mis Windowsi kooditabelite puhul võimaldab seda märki sisestada numbrisõrmistikult. Tabeli ülemise rea ja kõige vasakpoolsema veeru abil on mugav leida vajaliku märgi koodi kuueteistkümnendarvuna, selle leidmiseks ühendage omavahel number või täht ülemisest reast ja vasakpoolsest veerust. Näiteks Ä kood tabelis Windows CP 1252 on kümnendarvuna 196 (leiate Ä alt), kuueteistkümnendarvuna aga C4 (Ä kohal ülemises reas on C, kõige vasakpoolsemas veerus 4). Et tegemist on Windowsi kooditabeliga, saab järelkult Ä-d sisestada tippides Alt Gr klahvi all hoides numbrisõrmistikul 0196.
- Windowsi kooditabelites vahemikus 128–159 leidub veel üksikuid täitmata kohti, ka need on märgitud rastriga. Tühiku kood on kõigis kooditabelites 32, teises tühjas kuid ilma rastrita ruudus (nt 160 kõigis ISO kooditabeleis) asub sisetühik;
- kooditabel ISO 8859-1 kattub Unicode'i algusega, st selles kooditabelis esitatud koodid on ühtlasi ka vastavate märkide koodid Unicode'is. Nt saksa ß kood on kuueteistkümnendarvuna DF, mis Unicode'i-pärases kirjaviisis oleks U+00DF, standardis EVS 8:2000 pakutud kodeerimisviisil aga <00DF>;
- IT terminoloogias tarvitatakse vahel ka nimetusi C0 — juhtkoodid vahemikus 00–1F, C1 — juhtkoodid vahemikus 80–9F, GL — vahemik 20–7F, GR — vahemik A0–FF, G0 — graafilised märgid GL-is, G1 — graafilised märgid GR-is.

	2	3	4	5	6	7	A	B	C	D	E	F
0		0	@	P	`	p		°	À	Đ	à	đ
	32	48	64	80	96	112	160	176	192	208	224	240
1	!	1	A	Q	a	q	ı	±	Á	Ñ	á	ñ
	33	49	65	81	97	113	161	177	193	209	225	241
2	"	2	B	R	b	r	ç	²	Â	Ò	â	ò
	34	50	66	82	98	114	162	178	194	210	226	242
3	#	3	C	S	c	s	£	³	Ã	Ó	ã	ó
	35	51	67	83	99	115	163	179	195	211	227	243
4	\$	4	D	T	d	t	€	Ž	Ä	Ô	ä	ô
	36	52	68	84	100	116	164	180	196	212	228	244
5	%	5	E	U	e	u	¥	μ	Å	Õ	å	õ
	37	53	69	85	101	117	165	181	197	213	229	245
6	&	6	F	V	f	v	Š	¶	Æ	Ö	æ	ö
	38	54	70	86	102	118	166	182	198	214	230	246
7	'	7	G	W	g	w	§	•	Ç	×	ç	÷
	39	55	71	87	103	119	167	183	199	215	231	247
8	(8	H	X	h	x	š	ž	È	Ø	è	ø
	40	56	72	88	104	120	168	184	200	216	232	248
9)	9	I	Y	i	y	©	¹	É	Ù	é	ù
	41	57	73	89	105	121	169	185	201	217	233	249
A	*	:	J	Z	j	z	ª	º	Ê	Ú	ê	ú
	42	58	74	90	106	122	170	186	202	218	234	250
B	+	;	K	[k	{	«	»	Ë	Û	ë	û
	43	59	75	91	107	123	171	187	203	219	235	251
C	,	<	L	\	l	 	¬	œ	Ì	Ü	ì	ü
	44	60	76	92	108	124	172	188	204	220	236	252
D	-	=	M]	m	}	-	œ	Í	Ý	í	ý
	45	61	77	93	109	125	173	189	205	221	237	253
E	.	>	N	^	n	~	®	ÿ	Î	ß	î	ß
	46	62	78	94	110	126	174	190	206	222	238	254
F	/	?	O	_	o		—	¿	Ï	ß	ï	ÿ
	47	63	79	95	111	127	175	191	207	223	239	255

	2	3	4	5	6	7	A	B	C	D	E	F
0		0	@	P	`	p		°	À	Đ	à	đ
	32	48	64	80	96	112	160	176	192	208	224	240
1	!	1	A	Q	a	q	ı	±	Á	Ñ	á	ñ
	33	49	65	81	97	113	161	177	193	209	225	241
2	"	2	B	R	b	r	ç	²	Â	Ò	â	ò
	34	50	66	82	98	114	162	178	194	210	226	242
3	#	3	C	S	c	s	£	³	Ã	Ó	ã	ó
	35	51	67	83	99	115	163	179	195	211	227	243
4	\$	4	D	T	d	t	¤	´	Ä	Ô	ä	ô
	36	52	68	84	100	116	164	180	196	212	228	244
5	%	5	E	U	e	u	¥	µ	Å	Õ	å	õ
	37	53	69	85	101	117	165	181	197	213	229	245
6	&	6	F	V	f	v	¦	¶	Æ	Ö	æ	ö
	38	54	70	86	102	118	166	182	198	214	230	246
7	'	7	G	W	g	w	§	·	Ç	×	ç	÷
	39	55	71	87	103	119	167	183	199	215	231	247
8	(8	H	X	h	x	¨	,	È	Ø	è	ø
	40	56	72	88	104	120	168	184	200	216	232	248
9)	9	I	Y	i	y	©	¹	É	Ù	é	ù
	41	57	73	89	105	121	169	185	201	217	233	249
A	*	:	J	Z	j	z	ª	º	Ê	Ú	ê	ú
	42	58	74	90	106	122	170	186	202	218	234	250
B	+	;	K	[k	}	«	»	Ë	Û	ë	û
	43	59	75	91	107	123	171	187	203	219	235	251
C	,	<	L	\	l		¬	¼	Ì	Ü	ì	ü
	44	60	76	92	108	124	172	188	204	220	236	252
D	-	=	M]	m	}	-	½	Í	Ý	í	ý
	45	61	77	93	109	125	173	189	205	221	237	253
E	.	>	N	^	n	~	®	¾	Î	Þ	î	þ
	46	62	78	94	110	126	174	190	206	222	238	254
F	/	?	O	_	o		¯	¿	Ï	ß	ï	ÿ
	47	63	79	95	111	127	175	191	207	223	239	255

Windows CP1252 (Western)

0		0	@	P	`	p	€			°	À	Đ	à	đ
	32	48	64	80	96	112	128	144	160	176	192	208	224	240
1	!	1	A	Q	a	q		'	ı	±	Á	Ñ	á	ñ
	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	"	2	B	R	b	r	,	'	ç	²	Â	Ò	â	ò
	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	#	3	C	S	c	s	f	“	£	³	Ã	Ó	ã	ó
	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	\$	4	D	T	d	t	„	”	¤	'	Ä	Ô	ä	ô
	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	%	5	E	U	e	u	...	•	¥	µ	Å	Õ	å	õ
	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	&	6	F	V	f	v	†	—		¶	Æ	Ö	æ	ö
	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	'	7	G	W	g	w	‡	—	§	•	Ç	×	ç	÷
	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	(8	H	X	h	x	^	~	¨	,	È	Ø	è	ø
	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9)	9	I	Y	i	y	‰	™	©	¹	É	Ù	é	ù
	41	57	73	89	105	121	137	153	169	185	201	217	233	249
A	*	:	J	Z	j	z	Š	š	ª	º	Ê	Ú	ê	ú
	42	58	74	90	106	122	138	154	170	186	202	218	234	250
B	+	;	K	[k	{	<	>	«	»	Ë	Û	ë	û
	43	59	75	91	107	123	139	155	171	187	203	219	235	251
C	,	<	L	\	l		Œ	œ	¬	¼	Ì	Ü	ì	ü
	44	60	76	92	108	124	140	156	172	188	204	220	236	252
D	-	=	M]	m	}			-	½	Í	Ý	í	ý
	45	61	77	93	109	125	141	157	173	189	205	221	237	253
E	.	>	N	^	n	~	Ž	ž	®	¾	Î	Þ	î	þ
	46	62	78	94	110	126	142	158	174	190	206	222	238	254
F	/	?	O	_	o			ÿ	¯	¿	Ï	Ɔ	ï	ÿ
	47	63	79	95	111	127	143	159	175	191	207	223	239	255

DOS CP 850

0		0	@	P	`	p	Ç	É	á		L	ð	Ó	-
	32	48	64	80	96	112	128	144	160	176	192	208	224	240
1	!	1	A	Q	a	q	ü	æ	í		⊥	Ð	ß	±
	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	"	2	B	R	b	r	é	Æ	ó		⊥	Ê	Ô	⎯
	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	#	3	C	S	c	s	â	ô	ú		⊥	Ë	Ò	¾
	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	\$	4	D	T	d	t	ä	ö	ñ	⊥	—	È	õ	¶
	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	%	5	E	U	e	u	à	ò	Ñ	Á	⊥	ı	Õ	§
	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	&	6	F	V	f	v	å	û	ª	Â	ã	Í	μ	÷
	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	'	7	G	W	g	w	ç	ù	º	À	Ã	Î	þ	,
	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	(8	H	X	h	x	ê	ÿ	¿	©	ℒ	ï	þ	°
	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9)	9	I	Y	i	y	ë	Ö	®	⊥	⊥	⊥	Ú	¨
	41	57	73	89	105	121	137	153	169	185	201	217	233	249
A	*	:	J	Z	j	z	è	Ü	¬		⊥	⊥	Û	·
	42	58	74	90	106	122	138	154	170	186	202	218	234	250
B	+	;	K	[k	{	ï	ø	½	⊥	⊥		Ù	1
	43	59	75	91	107	123	139	155	171	187	203	219	235	251
C	,	<	L	\	l		î	£	¼	⊥	⊥		ý	3
	44	60	76	92	108	124	140	156	172	188	204	220	236	252
D	-	=	M]	m	}	ì	Ø	¡	¢	≡		Ý	2
	45	61	77	93	109	125	141	157	173	189	205	221	237	253
E	.	>	N	^	n	~	Ä	×	«	¥	⊥	ì	—	
	46	62	78	94	110	126	142	158	174	190	206	222	238	254
F	/	?	O	—	o		Å	f	»	⊥	¤		'	
	47	63	79	95	111	127	143	159	175	191	207	223	239	255

0	2	3	4	5	6	7	A	B	C	D	E	F
		0	@	P	`	p						
	32	48	64	80	96	112	160	176	192	208	224	240
1	!	1	A	Q	a	q	”	±	Į	Ń	į	ń
	33	49	65	81	97	113	161	177	193	209	225	241
2	"	2	B	R	b	r	¢	²	Ā	Ń	ā	ņ
	34	50	66	82	98	114	162	178	194	210	226	242
3	#	3	C	S	c	s	£	³	Ć	Ó	ć	ó
	35	51	67	83	99	115	163	179	195	211	227	243
4	\$	4	D	T	d	t	¤	“	Ä	Ö	ä	ö
	36	52	68	84	100	116	164	180	196	212	228	244
5	%	5	E	U	e	u	„	μ	Å	Õ	å	õ
	37	53	69	85	101	117	165	181	197	213	229	245
6	&	6	F	V	f	v	¡	¶	Ę	Ö	ę	ö
	38	54	70	86	102	118	166	182	198	214	230	246
7	'	7	G	W	g	w	§	•	Ē	×	ē	÷
	39	55	71	87	103	119	167	183	199	215	231	247
8	(8	H	X	h	x	Ø	ø	Č	U	č	u
	40	56	72	88	104	120	168	184	200	216	232	248
9)	9	I	Y	i	y	©	¹	É	Ł	é	ł
	41	57	73	89	105	121	169	185	201	217	233	249
A	*	:	J	Z	j	z	Ŕ	ŗ	Ż	Ś	ż	ś
	42	58	74	90	106	122	170	186	202	218	234	250
B	+	;	K	[k	{	«	»	È	Ū	è	ū
	43	59	75	91	107	123	171	187	203	219	235	251
C	,	<	L	\	l		¬	¼	Ç	Ü	ç	ü
	44	60	76	92	108	124	172	188	204	220	236	252
D	-	=	M]	m	}	-	½	Ķ	Ž	ķ	ž
	45	61	77	93	109	125	173	189	205	221	237	253
E	.	>	N	^	n	~	®	¾	Ī	Ž	ī	ž
	46	62	78	94	110	126	174	190	206	222	238	254
F	/	?	O	_	o		Æ	æ	Ł	ß	ł	'
	47	63	79	95	111	127	175	191	207	223	239	255

Windows CP 1257 (Baltic)

0		0	@	P	`	p	€			°	Ą	Š	ą	š
	32	48	64	80	96	112	128	144	160	176	192	208	224	240
1	!	1	A	Q	a	q		‘		±	Į	Ń	į	ń
	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	"	2	B	R	b	r	,	’	¢	²	Ā	Ņ	ā	ņ
	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	#	3	C	S	c	s		“	£	³	Ć	Ó	ć	ó
	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	\$	4	D	T	d	t	„	”	¤	´	Ä	Ö	ä	ö
	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	%	5	E	U	e	u	...	•		μ	Å	Õ	å	õ
	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	&	6	F	V	f	v	†	—		¶	Ę	Ö	ę	ö
	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	'	7	G	W	g	w	‡	—	§	•	Ē	×	ē	÷
	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	(8	H	X	h	x			Ø	ø	Č	Ų	č	ų
	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9)	9	I	Y	i	y	‰	™	©	¹	É	Ł	é	ł
	41	57	73	89	105	121	137	153	169	185	201	217	233	249
A	*	:	J	Z	j	z			Ŗ	ŗ	Ż	Ś	ż	ś
	42	58	74	90	106	122	138	154	170	186	202	218	234	250
B	+	;	K	[k	{	<	>	«	»	È	Ū	è	ū
	43	59	75	91	107	123	139	155	171	187	203	219	235	251
C	,	<	L	\	l				¬	¼	Ç	Ü	ç	ü
	44	60	76	92	108	124	140	156	172	188	204	220	236	252
D	-	=	M]	m	}	¨	-	-	½	Ķ	Ž	ķ	ž
	45	61	77	93	109	125	141	157	173	189	205	221	237	253
E	.	>	N	^	n	~	ˇ		®	¾	Ī	Ž	ī	ž
	46	62	78	94	110	126	142	158	174	190	206	222	238	254
F	/	?	O	_	o				Æ	æ	Ł	ß	ł	•
	47	63	79	95	111	127	143	159	175	191	207	223	239	255

DOS CP 775

0		0	@	P	`	p	Ć	É	Ā	▒	L	ą	Ó	-
	32	48	64	80	96	112	128	144	160	176	192	208	224	240
1	!	1	A	Q	a	q	ü	æ	Ī	▒	⊥	č	ß	±
	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	"	2	B	R	b	r	é	Æ	ó	▒	⊥	ę	Ō	“
	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	#	3	C	S	c	s	ā	ō	Ž		⊥	è	Ń	¾
	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	\$	4	D	T	d	t	ä	ö	ž	⊥	—	ı	õ	¶
	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	%	5	E	U	e	u	ğ	Ç	ž	Ą	†	š	Õ	§
	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	&	6	F	V	f	v	å	ç	”	Č	Ÿ	ı	μ	÷
	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	'	7	G	W	g	w	ć	Ś		Ę	Ū	ū	ń	„
	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	(8	H	X	h	x	ł	ś	©	È	Ł	ž	Ų	◦
	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9)	9	I	Y	i	y	ē	Ö	®	⊥	⊥	⊥	ķ	•
	41	57	73	89	105	121	137	153	169	185	201	217	233	249
A	*	:	J	Z	j	z	Ŕ	Ü	¬		⊥	⊥	Ł	•
	42	58	74	90	106	122	138	154	170	186	202	218	234	250
B	+	;	K	[k	{	ŗ	ø	½	⊥	⊥	■	ı	1
	43	59	75	91	107	123	139	155	171	187	203	219	235	251
C	,	<	L	\	l		ī	£	¼	⊥	⊥	■	ņ	3
	44	60	76	92	108	124	140	156	172	188	204	220	236	252
D	-	=	M]	m	}	Ž	Ø	Ł	ı	=	■	Ē	2
	45	61	77	93	109	125	141	157	173	189	205	221	237	253
E	.	>	N	^	n	~	Ä	×	«	Š	⊥	■	Ņ	■
	46	62	78	94	110	126	142	158	174	190	206	222	238	254
F	/	?	O	_	o		Å	œ	»	⊥	Ž	■	,	
	47	63	79	95	111	127	143	159	175	191	207	223	239	255

CENTEURO (Macintoshi Kesk- ja Ida Euroopa kooditabel)

	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0		0	@	P	`	p	Ä	ž	†	ı	ŋ	—	ř	ū
	32	48	64	80	96	112	128	144	160	176	192	208	224	240
1	!	1	A	Q	a	q	Ā	Ď	◦	Ī	Ń	—	Š	Ů
	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	"	2	B	R	b	r	ā	í	Ł	≤	¬	“	,	Ú
	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	#	3	C	S	c	s	É	d'	£	≥	√	”	„	ů
	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	\$	4	D	T	d	t	Ą	Ě	§	ī	ń	‘	š	Ů
	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	%	5	E	U	e	u	Ö	ē	•	Ɔ	Ň	’	Š	ů
	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	&	6	F	V	f	v	Ü	È	¶	∂	Δ	÷	ś	Ů
	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	'	7	G	W	g	w	á	ó	ß	Σ	«	◇	Á	ų
	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	(8	H	X	h	x	ą	è	®	ł	»	ō	ř	Ý
	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9)	9	I	Y	i	y	č	ô	©	Ł	...	Ř	ť	ý
	41	57	73	89	105	121	137	153	169	185	201	217	233	249
A	*	:	J	Z	j	z	ä	ö	™	ı		ř	Í	ķ
	42	58	74	90	106	122	138	154	170	186	202	218	234	250
B	+	;	K	[k	{	č	õ	ę	Ł	ň	Ř	Ž	Ż
	43	59	75	91	107	123	139	155	171	187	203	219	235	251
C	,	<	L	\	l		Ć	ú	¨	ł	Ŏ	<	ž	Ł
	44	60	76	92	108	124	140	156	172	188	204	220	236	252
D	-	=	M]	m	}	ć	ě	≠	Ł	Ŏ	>	Ů	ź
	45	61	77	93	109	125	141	157	173	189	205	221	237	253
E	.	>	N	^	n	~	é	ě	ğ	ı	ő	ř	Ó	Ç
	46	62	78	94	110	126	142	158	174	190	206	222	238	254
F	/	?	O	_	o		Ž	ü	ı	Ń	Ŏ	Ř	Ô	ˇ
	47	63	79	95	111	127	143	159	175	191	207	223	239	255

Lisa 6. Terminisõnastik

bitt	bit	(kahendnumber), number 0 või 1 kahendsüsteemis kasutatuna
bait	byte	teatavast arvust bittidest koosnev jada, mida käsitletakse tervikuna. Harilikult on baidis 8 bitti (sellist baiti nimetatakse ka oktetiks ja selle väärtus on vahemikus 0..255). Tavaliselt esitatakse baidiga ühte märki või märgi mingit osa
märk	character	andmete esituseks, korralduseks või juhtimiseks kasutatava elemendihulga liige. Märgid jaotatakse tavaliselt graafilisteks märkideks ja juhtmärkideks
loomulik keel	natural language	keel, mille reeglid põhinevad käibival kasutusel, olemata spetsiaalselt ette kirjutatud
märgistik	character set	terviklik märkide kogum kindlaks otstarbeks, nt eesti tähestiku tähed
kood	code	reegel, mis seab märgile vastavusse mõne teise hulga elemendi e koodiväärtuse (hrl täisarvu – numberkood), nt S vastab numbrile 83. Ka koodiväärtust (83) nimetatakse tihti koodiks.
kodeeritud märgistik, koodimärgistik	coded character set	märke ja neid esindavaid koode üheselt määratlev reeglistik
kodeering	encoding	teksti esitusviis, mille puhul koodiväärtused on kindlal ja tagasi teisendataval viisil konverditud, nt promilli Unicode'I koodiväärtuse esitamine kujul ‰
kooditabel	code page	koodimärgistiku esitus tabeli kujul
kirjamärk, graafiline märk	graphic character	märk, mis (erinevalt juhtmärgist) omab visuaalset kuju, teda võib kirjutada käsitsi, trükkida või kuvada ekraanile
tärk	alphanumeric character	täht, number või muu graafiline märk.
glüüf	glyph	märgi visuaalne kuju
täht	letter	kirjamärk, mis esineb loomuliku keele tähestikus
diakriitik	diacritical mark	lisamärk tähe koostises (nt täpid, akuut vm)
ligatuur	ligature	liitmärk, kahe märgi (rakendusjuhendis tähe) ühend (nt Œ)
tähestik, alfabeet	alphabet	kokkuleppelises järjestuses märgistik (loomuliku keele tähestik koosneb tähtedest)
kiri, kirjasüsteem	script	kirjamärkide hulk, mis on kasutusel ühe või mitme keele kirjutamiseks (nt ladina, kirillitsa, kreeka, devanaagari)

numbrimärk, number	digit	märk, mis esitab naturaalarvu. Kümnennumbrid on numbrid '0'..'9', kuueteistkümnennumbrid '0'..'9', 'A', 'B', 'C', 'D', 'E' ja 'F', kahendnumbrid (bitid) '0' ja '1'
tühimärk	blank (character)	märk, mis esitab tühja positsiooni stringis (tühik on selle eriliik)
tühik	space (character)	märk, mis kutsub esile sisestuskoha, prindi või kuvamise edasiliikumise ühe koha võrra
sisetühik	no-break space	erimärk stringis, mille kohal tekstiprotsessor ei poolita stringi (nt numbrites 32 000, initsiaalide ja nime vahel K. Kõmpus)
erimärk	special character	kirjamärk, mis pole täht, number ega tühimärk (nt koma, pluss, dollarimärk vm)
juhtmärk	control character	märk, mis toimib tegevuse (reavahetuse, sõnumi lõpu jne) käsuna ning millel ei ole iseseisvat graafilist kuju
string	string	tervikuna käsitletav märkide järjend
eraldaja	delimiter	üks või mitu märki, mida kasutatakse stringi alguse või lõpu märkimiseks
poolituskriips, ujukriips	soft hyphen, discretionary hyphen	erimärk sõnas, mis märgib võimalikku poolituskohta. Seda ei kuvata kui sõna asub rea sees ja kuvatakse sidekriipsuna kui sellel kohal toimub poolitus
sidekriips	hyphen	kriips, mida kasutatakse vastavalt õigekirjareeglitele ja mis kuvatakse alati
font	font	sarnase kujundusega glüüfide kogu