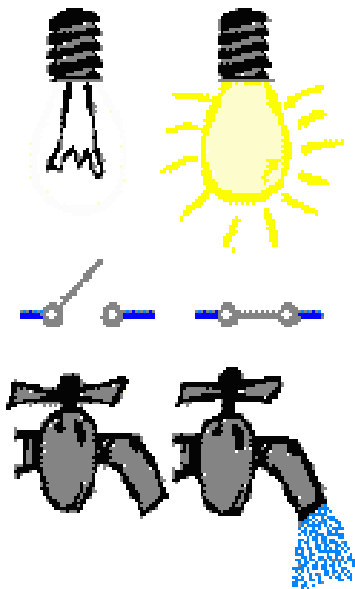


# Informacijos kodavimas.

Perduodami pranešimai koduojami. Kodavimas - tai vienos abėcėlės ženklų keitimas kitos abėcėlės ženklais. Kodavimas reikalingas tam, kad pranešimas būtų perduodamas kuo tiksliau, kad jis būtų kuo mažiau iškraipomas, kad jį suprastų gavėjas ir kad būtų galima persiųsti pasirinktu mainų kanalu. Koduojant dvejetainiais simboliais (tai aktualu kompiuteriams), iš n dvejetainių simbolių galima sudaryti 2<sup>n</sup> skirtingų kombinacijų, t.y. galima užkoduoti abėcėlę, turinčią ne daugiau kaip 2<sup>n</sup> ženklų.

0 1



Keletas dviejų būsenų pavyzdžių.

Kompiuteriu apdorojama informacija turi būti išreikšta jam patogiu pavidalu. Konstruojant kompiuterį paaiškėjo, kad geriausiai jį sudaryti iš tokių elementų, kurie turėtų dvi būsenas. Pavyzdžiui: laidu teka elektros srovė - neteka; įmagnetintas plotelis - neįmagnetintas plotelis, aukšta įtampa - žema įtampa. Šias būsenas paprasčiausiai žymėti vienetu ir nuliu (paprasčiau nė nesugalvosi).

Taigi galime sakyti, kad kompiuteris turi abėcėlę, kuri susideda iš dviejų ženklų: 0 (nulio) ir 1 (vieneto).

Dviejų ženklų abėcėlė vadinama **dvejetainė**.

Kai kalbame apie konkrečią abėcėlę, jos ženklus galime vadinti simboliais. Pavyzdžiui, sakome, kad kompiuterio dvejetainę abėcėlę sudaro du simboliai: 0 ir 1. Simboliais labiausiai įprasta vadinti kompiuterio abėcėlės ženklus (rodomus ekrane ar spausdinamus).

Dviejų ženklų abėcėlė vartojama ne tik kompiuteryje; ją galima aptikti daugelyje situacijų, pavyzdžiui:

- duodant sutikimą (linktelint galvą), neduodant sutikimo (papurtant galvą);
- nustatant lytį – vyras (♂), moteris (♀);
- ženklų pora – įjungta (⊕), išjungta (⊖);
- algebriniai ženklai – pliusas (+), minusas (-);
- atsiskaitymo forma, pavyzdžiui, bilietas pažymėtas (taip), nepažymėtas (ne).

Vienos abėcėlės ženklus galima išreikšti (užrašyti) kitos abėcėlės ženklais. Pavyzdžiui, lyties požymių abėcėlę (vyras, moteris) ♂, ♀

galima pažymėti atitinkamais simboliais (kompiuterio abėcėlė) 1, 0.



Vienos abėcėlės ženklų keitimas kitos abėcėlės ženklais vadinamas **kodavimu**. Taisyklės, nustatančios, kaip koduoti ženklus, vadinamos **kodu**.

Pamąstykime, kam reikalingas kodavimas, ar ne paprasčiau būtų apsieiti be jo.

**Pirma.** Visai be kodavimo neišsiversime, nes būtina garsinę kalbą paversti rašytine - į tai galime žiūrėti kaip į tam tikrą kodavimą.

**Antra.** Kalbėdami apie informacijos mainus, informacijos perdavimą, nurodėme, jog pranešimai turi būti išreikšiami taip, kad suprastų abu dalyviai: tiek siuntėjas, tiek gavėjas. Jeigu, pavyzdžiui, siunčiame pranešimą aklam, tai turime jo tekstą užkoduoti akliesiems skirtu Brailio raštu.

**Trečia.** Kodavimas reikalingas norint kuo tiksliau perduoti pranešimą.

**Ketvirta.** Kompiuteriai sukonstruoti taip, kad visa informacija juose turi būti išreikšta dviem būdais, t.y. atvaizduota 0 ir 1.

**1 pavyzdys.** Tarkime, mums reikia pasiųsti žinutę norvegui, nemokančiam lietuvių kalbos. Darome įprastai: savo pranešimą rašome kuria nors abiemis suprantama kalba, pavyzdžiui, vokiečių. Vadinasi, atliekame kodavimą: siuntėjas iš lietuvių kalbos perkoduoja į vokiečių kalbą, o gavėjas iš pastarosios perkoduoja į norvegų. Suprantama, šis kodavimas labai sudėtingas: tai ne vienų ženklų perrašymas kitais ženklais, - čia kodavimas virsta žodžių, fazių, sakinių sandaros lygmeniu. Tai vertimas iš vienos kalbos į kitą. Tačiau vis tiek tai yra kodavimas.

**2 pavyzdys.** Turbūt esate žaidę "sugedusį telefoną": vienas vaikas pašnibžda antram į ausį kokį nors žodį, antrasis jį persako trečiajam ir t.t. Kol žodis pasiekia vaikų eilės galą, dažniausiai pasikeičia tiek, kad jo ne atpažinti neįmanoma. Taip atsitinka dėl kelių priežasčių: neišgirstame kai kurių garsų (aplink kas nors triukšmauja), neteisingai suprantame kai kuriuos artimus garsus (pavyzdžiui, "blynas" ir "plynas"), pasitaiko neįprastas, mums nežinomas žodis (todėl jį nesąmoningai bandome keisti mums žinomam) ir t.t.

Jeigu žodžių raides užkoduotume skaičiais, pavyzdžiui, kad ir nusakančiais raidės vietą abėcėlėje, tai iškraipymas būtų gerokai mažesnis. Pagalvokite patys ir pasakykite, kodėl.

Jei galvosime apie pranešimą perduodančius įtaisus, tai vėlgi turime prisitaikyti prie jų galimybių, vadinasi, vėl reikia koduoti. Pavyzdžiui, Morzės abėcėlė atsirado tuomet, kai buvo sukonstruotas telegrafo aparatas, kuris galėjo perduoti tik dviejų rūšių signalus: trumpus ir ilgus.

## Dvejetainiai kodai

Taigi kompiuteriai sukonstruoti iš elementų, kurie gali turėti tik dvi būsenas. Taigi norėdami kurį nors pranešimą įrašyti į kompiuterį, turime jį užkoduoti. (Tik nemanysite, kad žmogus kompiuteriui savo pranešimą turi perrašyti nulių ir vienetų kalba ... Taip niekas nedaro. Kadangi šis veiksmas gana paprastas, griežtai nusakomas taisyklėmis, tai galima jį automatizuoti - kodavimą atlieka pats kompiuteris.)

Koduoti labai paprasta, kai abi abėcėlės turi vienodą skaičių ženklų - tereikia sudaryti abiejų abėcėlių ženklų poras. Taigi vienu dvejetainės abėcėlės ženklu galima koduoti tik, pavyzdžiui, dvi raides turinčią abėcėlę. Norint dvejetaine abėcėle koduoti abėcėlę, turinčią daugiau negu du ženklus, kiekvieną pastarosios ženklą tenka žymėti keliais dvejetainiais simboliais. Iš dviejų dvejetainių simbolių galima sudaryti keturias skirtingas kombinacijas:

00 10 01 11

Taigi dviem dvejetainiais simboliais galima koduoti abėcėlę, turinčią ne daugiau kaip keturis ženklus. Pavyzdžiui, šitaip galima užkoduoti keturių aritmetinių operacijų ženklus:

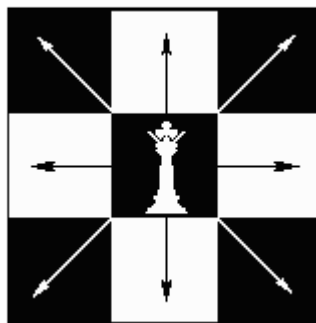
- 00      × 10      + 01      : 11

Iš trijų dvejetainių simbolių galima sudaryti 8 kombinacijas:

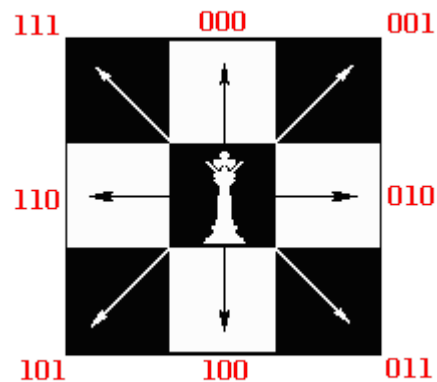
000 010 100 110 001 011 101 111

**3 pavyzdys.** Pabandykime dvejetainiais ženklais užkoduoti šachmatų karaliaus ėjimo kryptis (a pav.).

Kiek dvejetainių ženklų tam prireiks? Pirmiausia suskaičiuojame visus variantus - karaliaus ėjimo kryptis. Jų yra 8. Jei koduotume po du dvejetainius simbolius, tuomet žinotume, kad galėsime užkoduoti tik keturias kryptis. Vadinasi, teks koduoti po tris dvejetainius simbolius - kaip tik galėsime užkoduoti 8 kryptis:  $2^3 = 8$  (b pav.).



a)



b)

**Karaliaus galimų ėjimų kodavimo pavyzdys.**

Vertėtų pridurti, kad trimis dvejetainiais simboliais galėsime užkoduoti ir savaitės dienas: jų septynios, viena kodo kombinacija liks nepanaudota.

Bendru atveju iš  $n$  dvejetainių ženklų galima sudaryti  $2^n$  skirtingų kombinacijų ir jais koduoti abėcėlę, turinčią ne daugiau kaip  $2^n$  ženklų. Pateikiame keletą  $n$  ir  $2^n$  reikšmių:

<b>n</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b><math>2^n</math></b>	<b>2</b>	<b>4</b>	<b>8</b>	<b>16</b>	<b>32</b>	<b>64</b>	<b>128</b>	<b>256</b>	<b>512</b>	<b>1024</b>	<b>2048</b>	<b>4096</b>

Mes rašome lietuvių abėcėlės raidėmis. Taip ją pateikiame ir kompiuteriui. O kaip šią abėcėlę išreikšti dvejetainiais kodais?

Kompiuterio klaviatūroje yra daugiau nei 100 simbolių. Tai didžiosios ir mažosios raidės, skaitmenys, skyrybos ženklai, specialūs simboliai. Šešiais dvejetainiais ženklais galima koduoti 64 ženklų abėcėlę, o septyniais - 128 (žr. aukščiau pateiktas reikšmes). Pirmuosiuose kompiuteriuose buvo pasirinktas septynženklis kodavimas, nes vartotojams pakako 128 ženklų.

Šiuolaikiniuose kompiuteriuose vartojama daugiau kaip 128 simboliai. Todėl vienam simboliui koduoti dažniausiai skiriami 8 dvejetainiai ženklai. Jais galima užkoduoti abėcėlę, turinčią 256 simbolius. Kol kas daugumai kompiuterių vartotojų šitiek simbolių pakanka. Taigi viena raidė ar bet kuris kitas ženklas koduojamas aštuoniais dvejetainiais simboliais, pavyzdžiui,  $A = \{01000001\}$ ,  $a = \{01100001\}$ ,  $B = \{01000010\}$ ,  $b = \{01100010\}$ .

Kad būtų lengviau keistis informacija, priimami tarptautiniai bei nacionaliniai kodų standartai, kuriuose apibrėžiami raidžių ir kitų simbolių kodai. Jie paprastai surašomi į kodavimo lenteles.

Kai simbolis koduojamas 8 dvejetainiais ženklais, tai iš viso galima užkoduoti  $2^8 = 256$  simbolius. Jų pakanka kelių valstybių raidėms koduoti. Todėl viena kodų lentelė sudaroma kelioms ar net keliolikai valstybių.

## Kompiuterių kodų lentelės

Microsoft firma **Windows** operacinei sistemai sugrupavo valstybių kalbas ir sudarė keletą kodų lentelių. Vakarų Europai skirta vadinamoji **Windows-1252** kodų lentelė ([5 lentelė](#))

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	0	16	32	48	@	P	`	p	€		°	À	Ð	à	ð	
1	1	17	33	49	A	Q	a	q		´	ı	±	Á	Ñ	á	ñ
2	2	18	34	50	B	R	b	r	,	’	ƒ	²	Â	Ò	â	ò
3	3	19	35	51	C	S	c	s	f	“	£	³	Ã	Ó	ã	ó
4	4	20	36	52	D	T	d	t	”	”	¤	´	Ä	Ô	ä	ô
5	5	21	37	53	E	U	e	u	…	•	¥	µ	Å	Õ	å	õ
6	6	22	38	54	F	V	f	v	†	–	ı	¶	Æ	Ö	æ	ö
7	7	23	39	55	G	W	g	w	‡	—	§	·	Ç	×	ç	÷
8	8	24	40	56	H	X	h	x	ˆ	˘	˙	,	È	Ø	è	ø
9	9	25	41	57	I	Y	i	y	‰	™	©	˚	É	Ù	é	ù
A	10	26	42	58	J	Z	j	z	Š	š	ª	º	Ê	Ú	ê	ú
B	11	27	43	59	K	[	k	{	<	>	«	»	Ë	Û	ë	û
C	12	28	44	60	L	\	l		€	œ	–	¼	Ì	Ü	ì	ü
D	13	29	45	61	M	]	m	}			-	½	Í	Ý	í	ý
E	14	30	46	62	N	^	n	~			®	¾	Î	Þ	î	þ
F	15	31	47	63	O	_	o			ÿ	–	¿	Ï	ß	ï	ÿ

Joje yra bene daugiausiai kalbų: airių, albanų, anglų, danų, fareriečių, islandų, ispanų, italų, kataloniečių, norvegų, olandų, portugalų, prancūzų, suomių, švedų, vokiečių kalbų raides.

Lietuvai skirta **Windows-1257** kodų lentelė  
(6 lentelė).

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	0	16	32	48	@	P	`	p	€			°	À	Š	ą	š
1	1	17	33	49	À	Q	a	q		'		±	Į	Ń	ı	ń
2	2	18	34	50	B	R	b	r	,	'	ƒ	²	Ā	Ÿ	ā	ņ
3	3	19	35	51	C	S	c	s		“	£	³	Č	Ó	é	ó
4	4	20	36	52	D	T	d	t	”	”	¤	´	Ā	Ō	ā	ū
5	5	21	37	53	E	U	e	u	…	•		μ	Å	Ö	ä	ü
6	6	22	38	54	F	V	f	v	†	–	!	¶	Ę	Ö	ę	ü
7	7	23	39	55	G	W	g	w	‡	—	§	·	Ē	×	ē	÷
8	8	24	40	56	H	X	h	x			Ø	ø	Č	Ū	č	ų
9	9	25	41	57	I	Y	i	y	‰	™	©	¹	É	Ĺ	é	ĺ
A	10	26	42	58	J	Z	j	z			℞	ℓ	Ž	Š	ž	ś
B	11	27	43	59	K	[	k	{	<	>	«	»	Ě	Ů	ě	ů
C	12	28	44	60	L	\	l				–	¼	Ģ	Ū	ģ	ū
D	13	29	45	61	M	]	m	}	~	~	-	½	Ķ	Ž	ķ	ž
E	14	30	46	62	N	^	n	~	˘	˙	®	¾	Ī	Ž	ī	ž
F	15	31	47	63	O	_	o		,	,	Æ	æ	Ł	ß	ł	·

Joje taip pat yra anglų, danų, estų, latvių, lenkų, norvegų, suomių, švedų ir vokiečių raidės. Nors Didžioji Britanija, Norvegija, Suomija, Švedija, Vokietija naudoja Vakarų Europos kodų lentelę, Tačiau šių valstybių raidžių kodai sutampa su mūsų mūšiškos **Windows-1257** kodais.

Lentelių langeliuose surašytos koduojamos raidės, skaitmenys, skyrybos ženklai ir kiti simboliai. Smulkiu šriftu apačioje užrašyti intervalo [0 ; 255] skaičiai yra tų simbolių dešimtainiai kodai. Pavyzdžiui, raidės "N" kodas abiejose lentelėse yra 78.

Bet kurį lentelės simbolių galime gauti nuspaudę klaviatūros klavišus: ALT + 0 + dešimtainis kodas.

Kiekviena kodų lentelė padalinta į dvi dalis. Pirmoji dalis (stulpeliai nuo 0 iki 7) yra bendra visoms valstybėms. Ši lentelės dalis dar vadinama **ASCII kodu** (anglų k. *American Standard Code for Information Interchange* - standartinis JAV informacijos mainų kodas).

Kodų lentelių stulpeliai ir eilutės žymimi šešioliktainiais skaičiais (apie juos sužinosite kitame skyriuje). Simbolio šešioliktainis kodas gaunamas prie stulpelio numerio prirašius eilutės numerį, pavyzdžiui, raidės "N" šešioliktainis kodas 4E.

Kiekvienos dalies pirmieji 32 kodai (0 - 31 ir 128 - 159) yra skirti valdymo simboliams. Jie valdo teksto simbolių skaitymą bei rašymą. Pavyzdžiui, valdymo simbolis, kurio kodas yra 10, nurodo kompiuteriui, kad reikia spausdinti iš naujos eilutės. Valdymo simboliai tekstuose nerašomi. Todėl jie ir nepavaizduoti pateiktose kodų lentelėse.

## Unikodas

Kai vartojami 8 dvejetainiai ženklai, vienoje kodų lentelėje galima sutalpinti tik kelių kalbų rašmenis. Todėl vien Europos valstybių kalboms reikia keliolikos kodų lentelių. O kur Azijos kalbų hieroglifai ... Vis dėlto gerai būtų visų pasaulio kalbų ženklus koduoti viena ir ta pačia lentele - visiems būtų patogiau. Tam reikia vienam simboliui skirti ne 8, bet daugiau bitų - 16 ar net daugiau.

Šiuo tikslu 1988 m. buvo pradėtas kurti Unikodas. Vienam simboliui čia skiriama 16 bitų. Taigi vienoje kodų lentelėje gali tilpti  $2^{16} = 65536$  simboliai. Unikodo projektavimas tęsiamas: pagrindinis darbas - užpildyti laisvas vietas naujais simboliais.

Su Unikodu susiduriame jau dabar: daugelyje operacinės sistemos **Windows98** programų simboliai jau koduojami Unikodu.

Unikodu galima užrašyti labai daug simbolių, kol kas visiems pakanka vietos. Tačiau ateityje gal neužteks ir šios didžiulės kodų lentelės, todėl lygiagrečiai su Unikodu kuriama lentelė, kurios vienam simboliui skiriami 32 bitai, t.y. 4 baitai. Šios lentelės pirmųjų  $2^{16}$  simbolių kodai sutampa su Unikodu. Todėl galima sakyti, kad 32 bitų kodavimo lentelė yra Unikodo tęsinys.