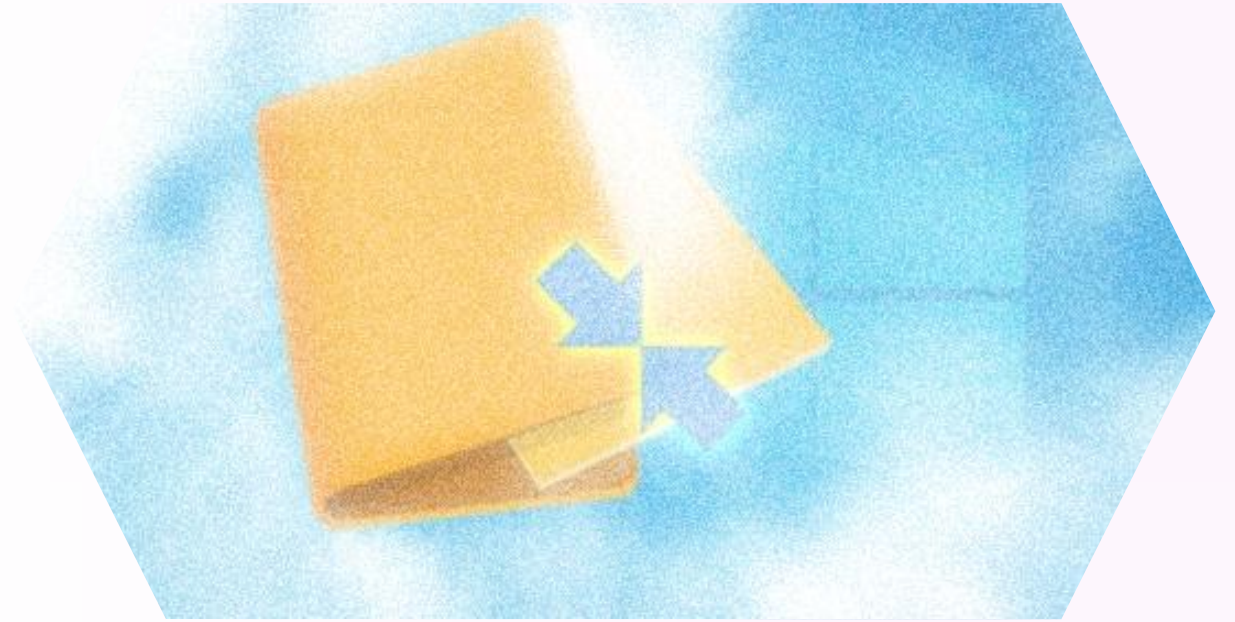


Įvairaus tipo duomenų glaudinimas



Šiandien pamokoje:

- *Susipažinsime su glaudinimo sąvoka.*
- *Aptarsime glaudinimo programas.*
- *Aptarsime duomenų glaudinimą, kai prarandame dalį duomenų detalumo (taikoma grafikos, vaizdo ir garso duomenyse).*
- *Diskutuosime apie didelių duomenų kiekių perdavimą.*
- *Mokysimės glaudinti skaitinius, tekstinius, grafinius, vaizdinius, garsinius duomenis, naudojant įvairius glaudinimo būdus.*

Ivadas

- **Kompiuterio atminties talpa didėja kasdien** – per paskutinius 25-ius metus standartinių kompiuterių ji padidėjo net milijoną kartų, bet matyt tai dar ne pabaiga.
- Kompiuteriuose gali būti laikomos knygos ar net ištisa biblioteka, muzikos įrašai ir filmai, jei tik yra laisvos vietos.
- **Didelės apimties dokumentai** internete yra labai **didelė problema**, nes ilgai užtrunka juos atsisiųsti.
- Dabar stengiamasi sumažinti kompiuterius – net telefonai ir rankiniai laikrodžiai gali laikyti labai daug informacijos.
- Vienas šios problemos sprendimų štai toks: užuot pirkę kompiuterius, turinčius daugiau atminties, ar įsigiję spartesnį interneto ryšį, **galime duomenis taip suspausti, kad jie užimtų mažiau vietos.**
- Šis duomenų kodavimo ir iškodavimo procesas dažniausiai atliekamas kompiuteriu automatiškai. Mes to proceso beveik nepastebime.

Duomenų glaudinimas

- **Glaudinimas** – bylos duomenų apimties sumažinimas. Kitaip dar vadinamas **archyvavimu** arba **pakavimu**.
- Glaudinimas galimas dėl to, kad duomenų apimtis dažnai yra didesnė nei jų teikiama informacija.

Duomenų glaudinimas

- Glaudinimo programos įvairiapusiškai ***analizuoja*** duomenis, ***atranda pasikartojančių*** kodų grupes, jų pasikartojimo dėsningumus ir jais pasinaudodamos duomenis ***suglaudžia***.
- Duomenų **pertvarkymas taip, kad jie užimtų mažiau vietos atmintyje** ir jiems persiųsti reikėtų siauresnės dažnių juostos arba persiuntimas būtų spartesnis.
- Kai juos reikia panaudoti, ta pati programa juos išpakuoja – pakeičia ankstesniu pavidalu.
- Glaudinami pakuojami, archyvuojami ir elektroniniu paštu persiunčiami duomenys.

Duomenų glaudinimas

- Parenkant metodus ir juos realizuojančias programas, reikia rasti optimalų laiko ir sutaupytos atminties santykį.
- **Ne visi duomenys vienodai gerai susispaudžia.**
 - *labai gerai* susispaudžia **.txt, .doc, .bmp** tipo bylos.
 - *nepavyks* suspausti **.mp3, .jpg** tipo bylų (šis formatas yra jau glaudintų duomenų formatas).
- Tačiau jas vis vien tikslinga archyvuoti.

Archyvų tipai

- Glaudinimo programų darbo rezultatas – byla ar jų rinkinys, vadinamas *archyvu*.
- Archyvai paprastai turi plėtinį, atitinkantį programą, kuria jie buvo sukurti – **.zip**, **.rar**, **.arj**, **.ice** ir t. t.
- Todėl **glaudinimo programos dažniausiai vadinamos archyvavimo programomis** arba tiesiog archyvatoriais.
- Archyve saugoma ir papildoma informacija: bylų vardai, modifikavimo datos, kontrolinės sumos (pvz., bendras vienetukų skaičius bylos duomenyse).
- Tai padeda patikrinti, ar bylos iš archyvo atkurtos teisingai.

Pavyzdys, kaip gali būti glaudinamas tekstas

vasara pavasaris



vasara

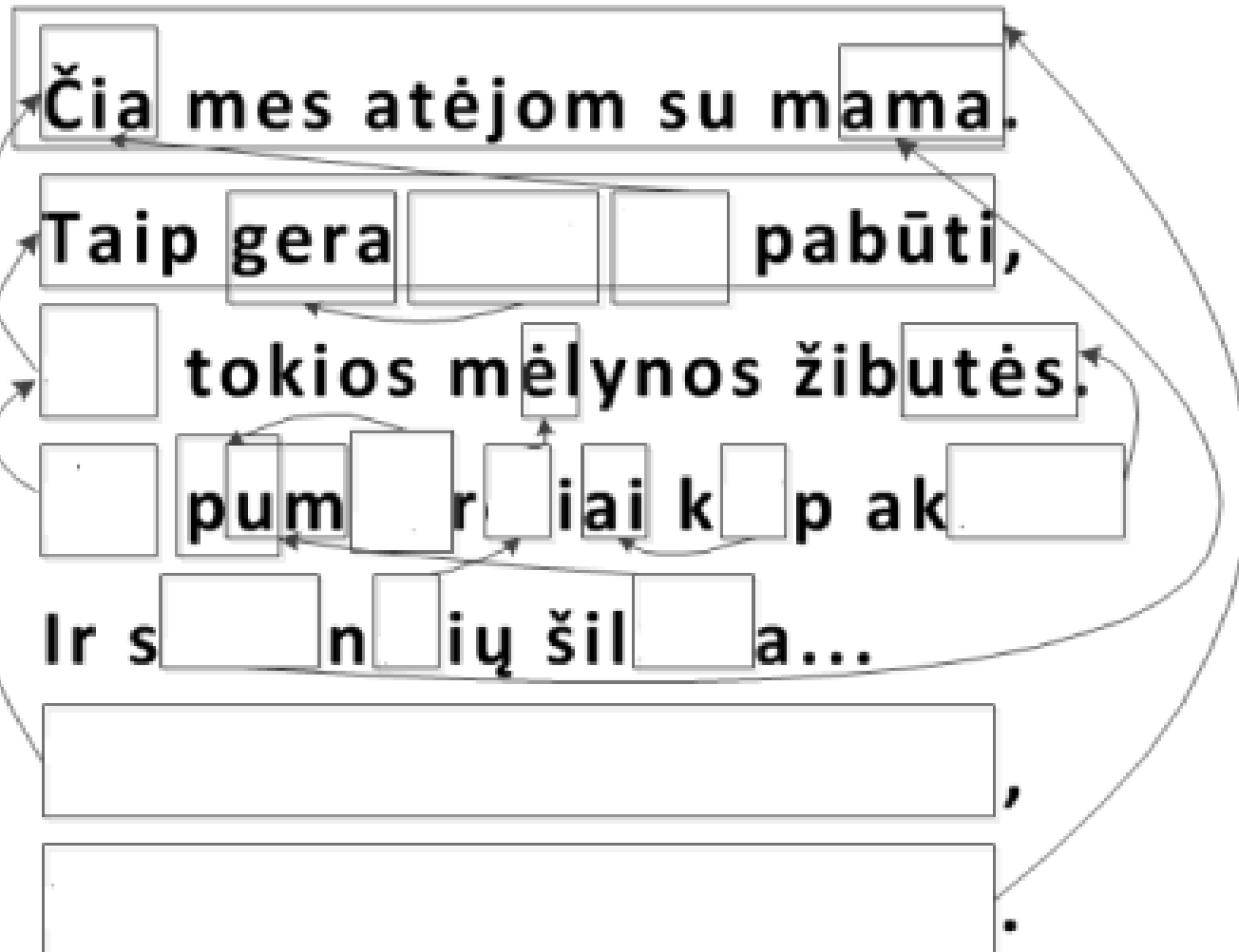
pa



is

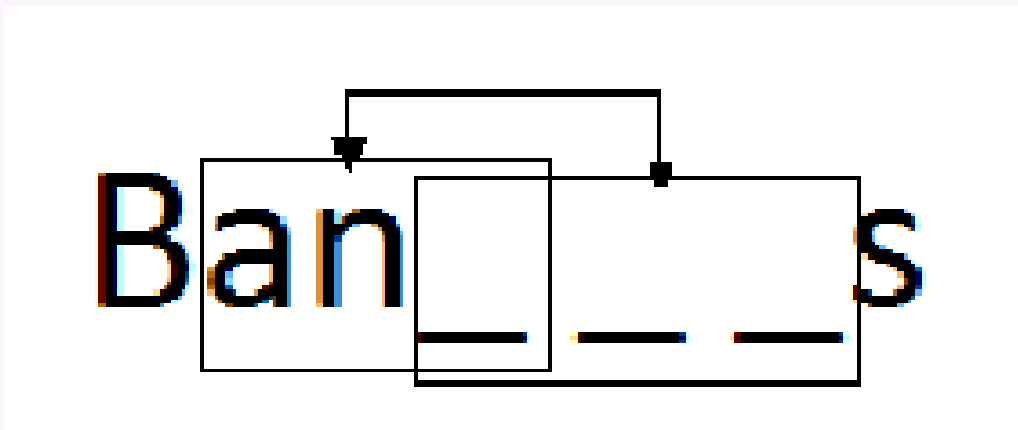
Pasikartojantys žodžiai,
žodžių dalys arba frazės,
pakeičiami kvadratėliais.

*Pakartokime **Kaip atrodytų nesuglaudintas eilėraštis?***



Čia mes atėjom su mama.
Taip gera gera čia pabūti:
Čia tokios mėlynos žibutės.
Čia pumpurėliai kaip akutės
Ir samanėlių šiluma...
Taip gera gera čia pabūti,
Čia mes atėjom su mama.

Kaip išspręstumėte šį galvosūkį?



Bananas

- Kartais trūkstamos teksto dalys rodo pačios į save.
- Šiuo atveju žodis gali būti iškoduotas teisingai, jeigu kopijuojame raides iš kairės į dešinę.
- Tada kiekviena raidė gali būti kopijuojama prieš tai, kai įrašoma.
- Toks kopijavimo būdas labai naudingas kompiuteriams, kai yra daug pasikartojimų.

Kaip kompiuteryje vaizduojamas glaudinimas

Kompiuteryje stačiakampiai ir rodyklės vaizduojami skaičiais. Pavyzdžiui,

Pasas

gali būti užrašytas kaip **Pas(2,2)**. Pirmasis skaičius 2 rodo antrą vietą nuo galo – kopijavimo vietos pradžią

Pas—

Antrasis skaičius 2 reiškia, kad kopijuosime dvi iš eilės einančias raides.

Pasa—

Pasas



Kadangi žodžiui užkoduoti naudojami du skaičiai, tai verta suspausti dviejų ar daugiau raidžių grupes, kitaip nebus sutaupoma vietos. Iš tiesų failo dydis gali padidėti, jeigu dviem skaičiais koduosime vieną raidę.

Pakartokime: pabandykite suglaudinti šias eilutes

Saulė

Vasara pavasaris

Vasara pavasaris

Saulė šypsosi

Vasara pavasaris

Vasara pavasaris

Į laukelį kviečia

Saulė

Vasara pavasaris

Vasara pavasaris

Saulė šypsosi

Vasara pavasaris

Vasara pavasaris

Į laukelį kviečia

Duomenų glaudinimo metodai

- Sugalyvota daug duomenų suspaudimo metodų. Atliekant šią veiklą, buvo naudojamas metodas, kai nuorodomis rodoma į prieš tai buvusius pasikartojančius simbolius tekste.
- Šis metodas vadinamas tiesiog „zip“, jį 1970 m. sugalvojo Izraelio mokslininkai / Phil Katz /. Metodas tinka bet kuriai kalbai, juo lengvai suspaudžiami ir perpus sumažinami duomenys.
- Zip yra glaudinimo be praradimų formatas: *po išskleidimo duomenys yra tokie patys, kokie buvo prieš glaudinimą*. Jis veikia rasdamas perteklinius pradinius duomenis ir efektyviau juos atvaizduodamas. (<https://www.hanshq.net/zip.html>)

Interaktyvi glaudinimo priemonė

<https://studio.code.org/s/text-compression/lessons/1/levels/2>

Teksto pavyzdys:

Apvali, saldi gražuolė...

Šmurkšt, po suolu jau nupuolė.

„Kurgi bėgi? Grįžk į puodą...”

Bet, pakėlus savo kuodą,

Bulvė lekia per laukus:

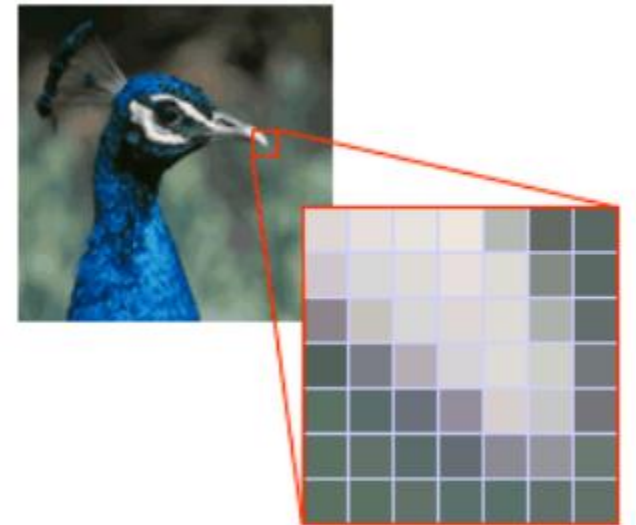
„Nebegrišiu pas vaikus.”

Praktinė užduotis

- Išbandykite interaktyvią glaudinimo priemonę su savo tekstu
- Paieškokite kokie yra dar duomenų glaudinimo metodai, kai neprarandami pradiniai duomenys?

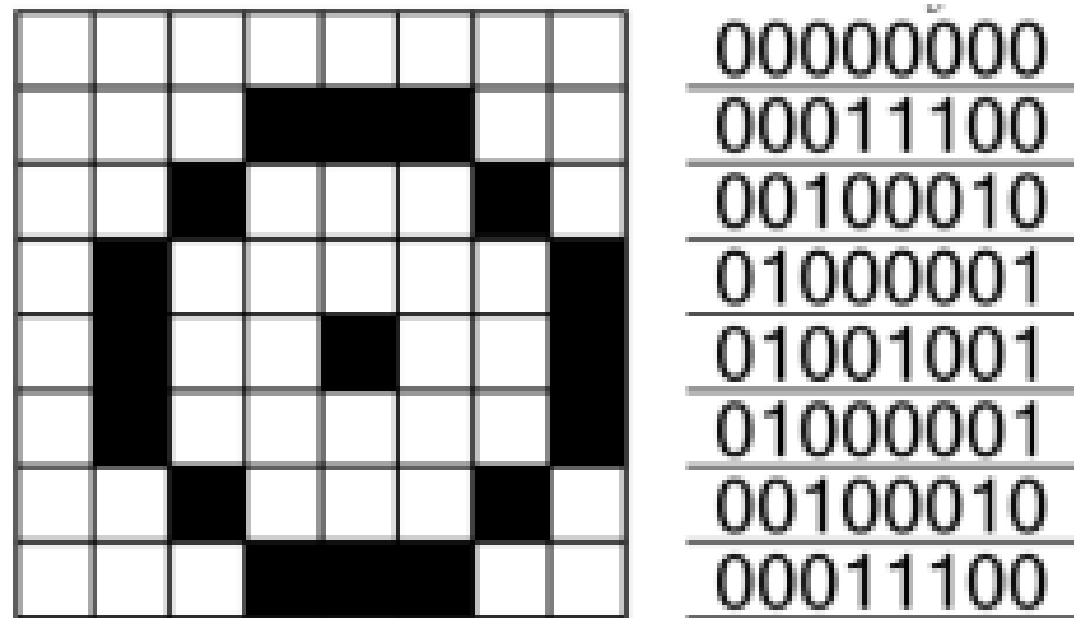
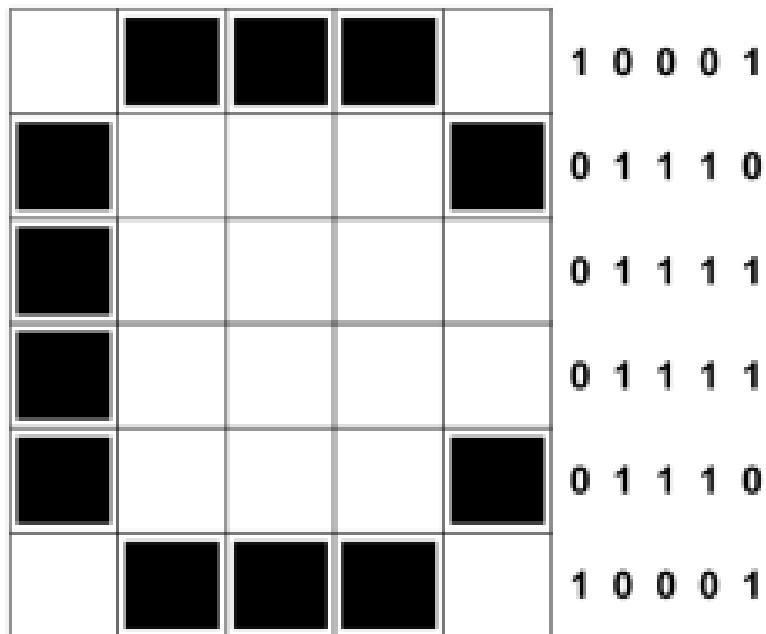
Vaizdų kodavimas ir glaudinimas

- Nuotraukos ir kiti vaizdai ekrane yra pateikiami kaip spalvotų taškų (pikselių) rinkinys.
- Kompiuterių ekranai suskirstyti į tinklelę mažų kvadratėlių (pikselių, taškų). Kiekvienas iš pikselių gali būti skirtingos spalvos, o kadangi pikseliai maži, atskirų taškų ekrane nematome. Jie susilieja ir formuoja rodomą vaizdą.



Vaizdų kodavimas

Paprasčiausią nespaltvotą vaizdą
galime užkoduoti dvejetainine sistema

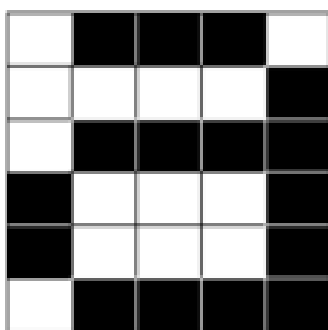


Vaizdų glaudimas

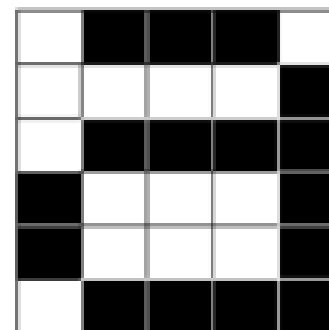
- Ar galima aukščiau užkoduotus vaizdus suglaudinti?

0	1	1	1	0
0	0	0	0	1
0	1	1	1	1
1	0	0	0	1
1	0	0	0	1
0	1	1	1	1

Atkreipkite dėmesį, kad pirmoje eilutėje šalia yra 3 juodi pikseliai, o antroje – 4 balti pikseliai. Taigi užuot užrašę kiekvieną bitą, galime skaičiais apibūdinti raštą.



1, 3, 1



1, 3, 1

4, 1

1, 4

0, 1, 3, 1

0, 1, 3, 1

1, 4

Duomenų glaudinimas

1 paveikslas

2, 7

2, 7

2, 1, 1, 3, 1, 1, 1, 1

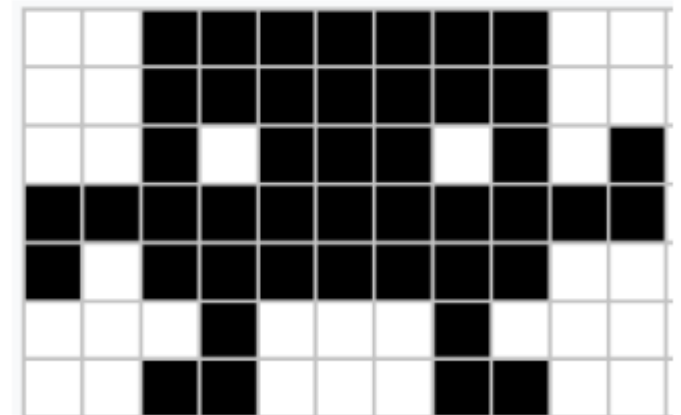
0, 11

0, 1, 1, 7

3, 1, 3, 1

2, 2, 3, 2

1 paveikslas



Duomenų glaudinimas

2 paveikslas

5, 6

4, 1, 6, 1

3, 1, 7, 2

2, 2, 9, 1

1, 1, 12, 1

0, 1, 4, 1, 4, 1, 4, 1

0, 1, 14, 1

0, 1, 14, 1

0, 1, 3, 1, 6, 1, 3, 1

1, 1, 3, 1, 4, 1, 3, 1

2, 1, 3, 4, 4, 1

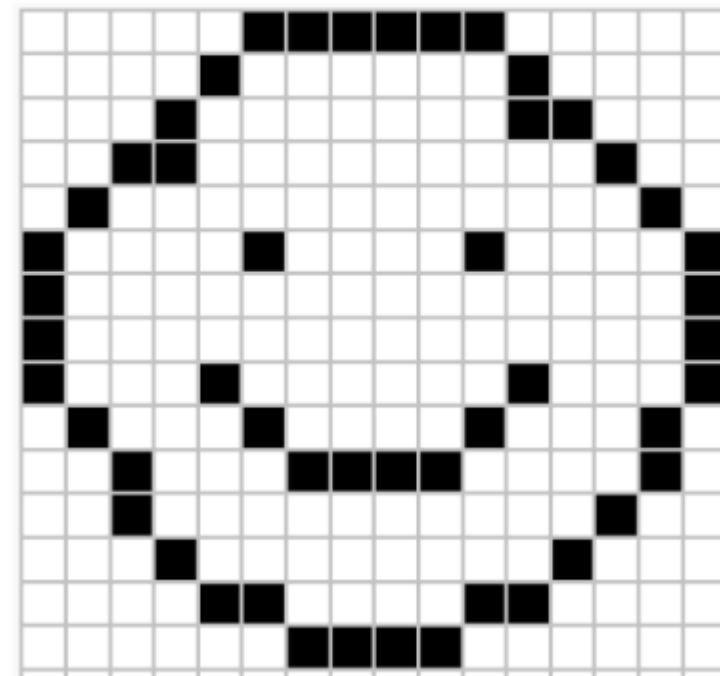
2, 1, 10, 1

3, 1, 8, 1

4, 2, 4, 2

6, 4

2 paveikslas



Praktinė užduotis

Koks paveikslas suglaudintas?

4, 1, 5

3, 3, 4

2, 1, 3, 1, 3

1, 2, 1, 1, 1, 2, 2

0, 1, 7, 1, 1

0, 1, 1, 2, 1, 2, 1, 1, 1

0, 1, 1, 2, 1, 2, 1, 1, 1

0, 1, 4, 2, 1, 1, 1

0, 9, 1

4, 6

Duomenų glaudinimas

Naudodamiesi interaktyvia piešinių kodavimo priemone

<https://csfieldguide.org.nz/en/interactives/run-length-encoding/>

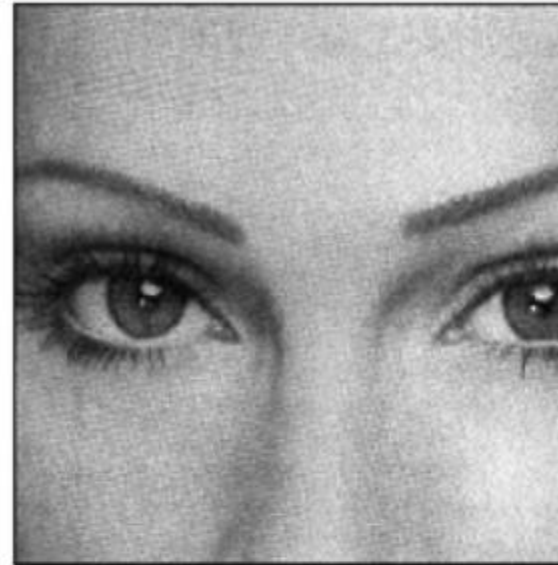
parenkite daugiau piešinių ir jų kodų rinkinių, iš kurių mokiniai pagal pateiktą kodą galėtų nupiešti piešinį arba pagal pateiktą piešinį parašyti jo kodą.

Glaudvinimas prarandant duomenis

- Glaudinimas su nuostoliais - tai skaitmeninių duomenų saugojimo ir perdavimo srityje naudojamas duomenų glaudinimo metodas, kuriuo sumažinamas failo dydis, pašalinant dalį duomenų.
- Skirtingai nuo glaudinimo be nuostolių, kai išskleidus išsaugomi visi pradiniai duomenys, glaudinant su nuostoliais aukojama dalis duomenų, kad būtų pasiektas didesnis glaudinimo koeficientas.
- Tai reiškia, kad išskleidžiant failą, kuriam buvo atliktas nuostolingasis glaudinimas, negaunama tiksli originalių duomenų kopija; vietoj to gaunama apytikslė kopija, kurios kokybė gali šiek tiek nukentėti.

JPEG formatas

- Vaizdo glaudinimas: JPEG (angl. Joint Photographic Experts Group), sumažina vaizdų failų dydį, atsisakydami kai kurių mažiau pastebimų detalių ir spalvų, todėl gaunamas mažesnis failas, tačiau *sumažėja vaizdo kokybė*.
- <https://sipdtdevelopers.wordpress.com/2014/01/23/digital-image-compression/>



(a) Original Image



(b) Image with 10:1 compression



(c) Image with 45:1 compression

Praktinė užduotis

- Suglaudinkite savo darbų aplanką, esantį kompiuterio darbalaukyje.
- Patikrinkite, kaip pasikeitė aplanko dydis?

Namų darbas: praktinė užduotis

- Peržiūrėkite filmuotą medžiagą : <http://tests.lt/glaudinimas/> - įsidėmėkite archyvų plėtinius.
- **Ar pasiruošę tikrai „kietai“ suglaudinti?** Popieriaus lape Jums pateikta istorija buvo apdorota kompiuterio programa, kuri šioje istorijoje aptiko mažiausiai 1000 galimų išbraukti raidžių. Kiek galite rasti jų? Prisiminkite, gali būti išbraukiamos tik grupės po dvi ar daugiau pasikartojančių raidžių. Sėkmės!