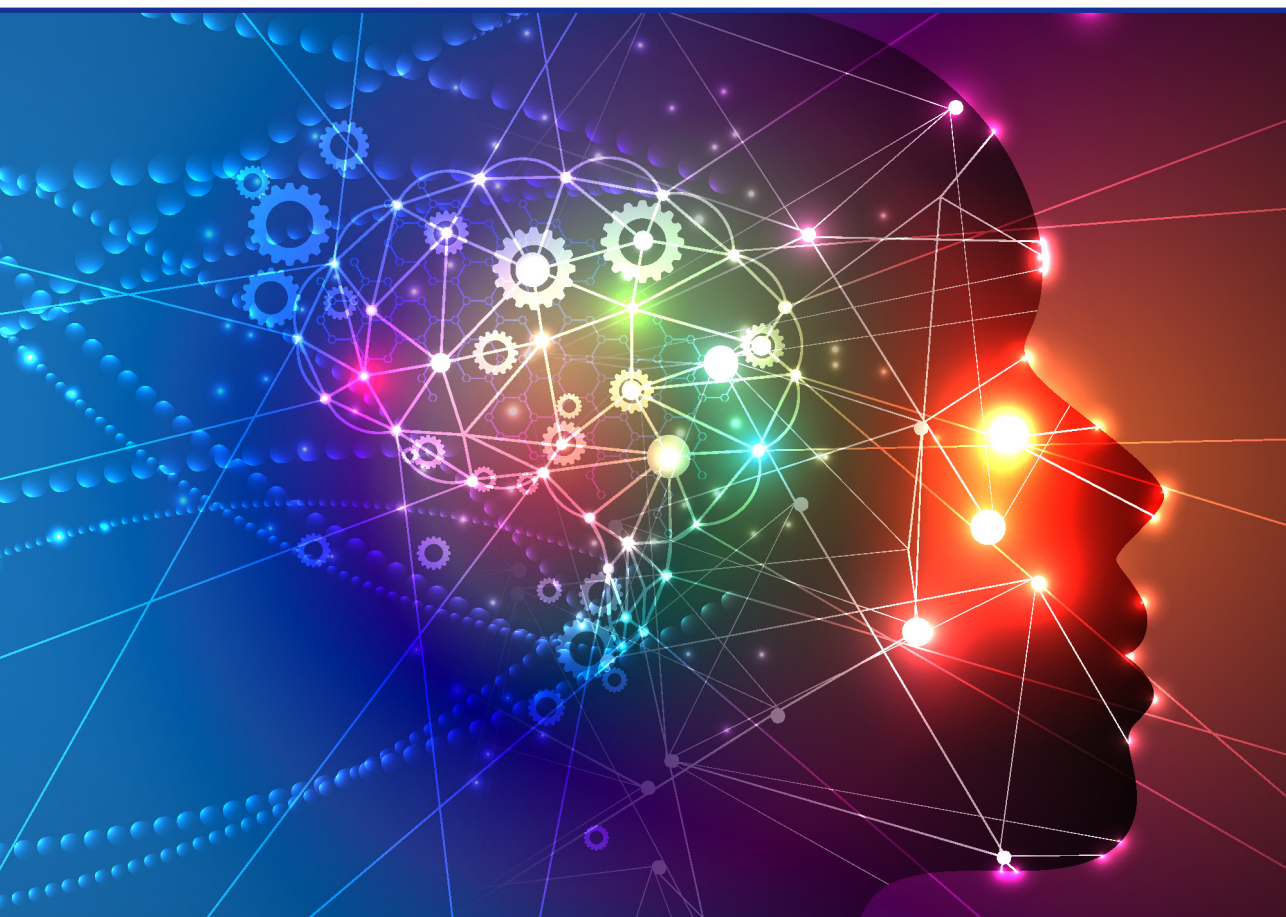


Linda Alksne

## VIDEO LEKCIJAS KĀ KODA ANALĪZE KOMBINĒTAJĀ APMĀCĪBU MODELĪ

Promocijas darba kopsavilkums



Liepājas Universitāte  
Dabas un inženierzinātņu fakultāte



## VIDEO LEKCIJAS KĀ KODA ANALĪZE KOMBINĒTAJĀ APMĀCĪBU MODELĪ

Linda Alksne

Promocijas darba kopsavilkums  
inženierzinātņu doktora grāda iegūšanai  
elektrotehnikas, elektronikas, informācijas un komunikāciju tehnoloģijas nozare,  
e-studiju tehnoloģiju un pārvaldības apakšnozare

Promocijas darba zinātniskais vadītājs

*Dr. habil. phys.* Andris Ozols

Liepāja  
Rīga  
RTU Izdevniecība  
2023

Alksne L. Video lekcijas kā koda analīze kombinētajā apmācību modelī. Promocijas darba kopsavilkums. – Rīga: RTU Izdevniecība, 2023. – 58 lpp.

Iespiests saskaņā ar promocijas padomes “RTU P-21” 2022. gada 6. decembra lēmumu, protokols Nr. 5.



<https://doi.org/10.7250/9789934228797>

ISBN 978-9934-22-879-7 (pdf)

Promocijas darbs izstrādāts Liepājas Universitātes Dabas un inženierzinātņu fakultātē no 2014. līdz 2022. gadam.

Darbā ir ievads, trīs daļas, 13 nodaļas, 52 apakšnodaļas, rezultāti, nobeigums un secinājumi, izmantotās literatūras un avotu saraksts, kā arī pielikums. Promocijas darbā iekļauti 57 attēli un 19 tabulu.

Darba forma: promocijas darbs elektrotehnikas, elektronikas, informācijas un komunikāciju tehnoloģijas nozarē, e-studiju tehnoloģiju un pārvaldības apakšnozarē.

Darba zinātniskais vadītājs

*Dr. habil. phys.* Andris Ozols, Rīgas Tehniskā universitāte

Darba recenzenti:

- *Dr. paed.* Sarma Cakula, Vidzemes Augstskola, Latvija
- *Dr. sc. ing.* Jurgis Poriņš, Rīgas Tehniskā universitāte
- *Ph. D. Carlos Turro Ribalta*, Valensijas Politehniskā universitāte, Spānija

Promocijas darba aizstāvēšana notiks Rīgas Tehniskajā universitātē 2023. gada 20. martā Elektrotehnikas, elektronikas, informācijas un komunikāciju tehnoloģijas nozares promocijas padomē “RTU P-21” (elektrotehnikas, elektronikas, informācijas un komunikāciju tehnoloģijas nozare).

Rīgas Tehniskās universitātes promocijas padomes priekšsēdētājs *Dr. phys.* Atis Kapenieks,  
padomes sekretāre Mārīte Treijere.

### APSTIPRINĀJUMS

Apstiprinu, ka esmu izstrādājusi šo promocijas darbu, kas iesniegts izskatīšanai Rīgas Tehniskajā universitātē zinātnes doktora (*Ph. D.*) grāda iegūšanai. Promocijas darbs zinātniskā grāda iegūšanai nav iesniegts nevienā citā universitātē.

Linda Alksne ..... (paraksts)

Datums: .....

## Anotācija

E-studijās kā viens no pamata informācijas nodošanas veidiem tiek izmantota video lekcija. Šajā darbā video lekcija ir apskatīta kā kods, ar kura palīdzību studentiem tiek nodota informācija. 1948. gadā K. Šenons rakstā “Sakaru matemātiskā teorija” (“*A Mathematical Theory of Communication*”) izstrādāja informācijas teoriju, lai apskatītu būtiskākos sakaru sistēmu aspektus. Galvenā nozīme K. Šenona informācijas teorijā ir entropijai – vidējai informācijai uz vienu ziņojumu no ziņojumu avota. Šenona entropija ir skaitlis, kas ir atkarīgs tikai no informācijas avota statistiskās dabas. Entropija ir ziņojumu avota informācijas mērs.

Līdz šim entropija video lekciju uztveramības kvantitatīvai novērtēšanai nav lietota. Tāpēc darba pētījuma mērķis ir veikt video lekciju analīzi ar entropijas palīdzību, lai noskaidrotu sakarību starp video lekcijas veidu un entropiju, kā arī to, kā vadlīnijas, kas zinātniskajā literatūrā pierādītas kā noteicošās studentu uztveres spējas palielināšanai, ietekmē entropiju. Tas dod iespēju kvantitatīvi novērtēt video lekciju uztveramību.

Darba gaitā ir apkopotas un definētas vadlīnijas viegli uztveramas video lekcijas filmēšanai, pētīti studentu video lekciju skatīšanās parametri un izvēlētas 11 dažāda veida video lekcijas, kam ar programmas *Matlab* palīdzību noteikta vidējā video un audio entropija. Iegūtie rezultāti ir analizēti, pētot vadlīniju ietekmi uz entropiju, kā arī entropijas atkarību no izvēlēta video lekcijas filmēšanas veida. Konstatēta tendence, ka video lekcijas, kas pēc vadlīnijām būtu jāuztver labāk, ir ar zemāku entropiju.

## Saturs

Anotācija.....	4
1. Informācijas teorija.....	13
1.1. Informācija un tās definīcija sakaru teorijā.....	13
1.2. K. Šenona definīcija diskrētu ziņojumu ansamblim.....	15
1.3. Ziņojuma ansambļa galvenie raksturlielumi .....	16
1.3.1. Entropija.....	16
1.3.2. Savstarpējā informācija.....	17
1.3.3. Redundance.....	18
2. Valodu entropija.....	21
3. Latviešu valodas kā koda salīdzinājums dažādos komunikāciju kanālos.....	23
4. Informācija no studentu video lekciju skatīšanās darbību pierakstiem .....	34
5. Vadlīnijas video lekcijas radīšanai .....	41
6. Video lekcijas maksimālā sintaktiskā informācija .....	42
6.1. Kopējā maksimālā lekcijas informācija un lekcijas kā komunikācijas kanāla kapacitāte.....	44
7. Video lekciju raksturojums un analīze .....	47
7.1. Video lekciju raksturojums .....	47
7.2. Video lekcijas kā koda analīze.....	47
Nobeigums un secinājumi.....	56
Izmantotā literatūra.....	57

## **Darbā izmantotie saīsinājumi**

DW – *data warehouse* – datu noliktava

*ETL* – *extract, transform, load* – izvilkšana, transformēšana un ielāde, viens no biznesa inteliģences procesiem

*H* – entropija (biti)

kHz – frekvences mērvienība, kiloherci

LiepU – Liepājas Universitāte

*LZW* – *Lempel-Ziv-Welch* datu saspiešanas algoritms

NSTI – Nacionālais standartu un tehnoloģijas institūts, ASV

RTU – Rīgas Tehniskā universitāte

*RLE* – *Run-length encoding* – darbības garuma kodējums

*Covid-19* – 2019. gada koronavīrusa slimība (angļu val. *coronavirus disease*), elpceļu infekcijas slimība, ko izraisa *SARS-CoV-2* vīruss

## Ievads

Video lekcija ir video ieraksts, kurā tiek piedāvāts izglītojošs materiāls par tēmu, kas ir jāapgūst. Video ieraksts ir elektroniska tehnoloģija vizuālas informācijas ierakstīšanai, kas tiek pasniegts video signāla vai video datu digitālās straumes veidā fiziskā datu nesējā, lai saglabātu šo informāciju un iespēju to pēc tam atskaņot un parādīt izvades ierīcē (monitorā, ekrānā vai displejā). Studiju procesu organizējot kombinētajā studiju modulī, lai video lekcijas publicētu, nepieciešama arī tāda tehnoloģijas kā e-studiju vide vai platforma.

Neatkarīgi no tā, vai jums ir pieci vai 95 gadi, internets mūsdienās piedāvā daudz. Īpaši, ja meklētā tēma ir izglītība, tad resursi šķiet bezgalīgi. Tādas e-studiju platformas kā *Coursera*, *Khan Academy*, *Open Culture Online Courses*, *Udemy*, kā arī lielāko universitāšu – Hārvardas, Jēlas, Bērklijas un citu pasaules ranga augšgalā esošo universitāšu – tiešsaistes studiju platformas, kas ietver tūkstošiem mācību priekšmetu, kas balstīti galvenokārt video lekciju publicēšanā, parāda šīs tēmas aktualitāti.

Arī Rīgas Tehniskajā universitātē jau 2012. gadā moderni tika aprīkotas auditorijas video lekciju ierakstīšanai un video materiālu izvietošanai pielāgotajā *Moodle* e-studiju platformā. Šim piemēram sekoja arī Latvijas Universitāte, Rīgas Stradiņa universitāte, Latvijā e-studiju platformas lieto vispārizglītojošās skolas, izmantojot savu vai *eduspace.lv* piedāvātu e-studiju platformas risinājumu.

**Promocijas darba nosaukums** – “Video lekcijas kā koda analīze kombinētajā apmācību modelī (*blended learning*)” – liecina, ka tiks analizēti dažādi mācību video ieraksti – gan lekciju ieraksti no auditorijas, gan gatavoti mācību materiāli, kas ir filmēti kā studiju papildmateriāli. Kombinētās mācības padara izglītības vidi elastīgu, jo zināšanas un informācija mācībām ir pieejama jebkurā laikā un jebkurā vietā.

**Darba problēma.** Autore vēlas kvantitatīvi noskaidrot, kāda ir viegli uztverama video lekcija. Entropija ir sakaru teorijas svarīgākais kvantitatīvais mērs, kas plaši tiek izmantots ne tikai sakaru tehnoloģijās, bet arī biznesa inteliģencē, reklāmā, lingvistikā un citur, tāpēc ir loģiski to izmantot arī video lekciju raksturošanai, kas līdz šim, cik autorei zināms, nav izdarīts. Promocijas darbā tiek pārbaudīts, kā iepriekš veiktajos citu autoru pētījumos definētās labu video lekciju kvalitatīvās vadlīnijas ietekmē entropiju un kā entropija korelē ar video lekcijas veidu. Video lekcija tiek analizēta kā (optisks un akustisks) kods. Dažādām video lekcijām atbilst dažādi kodi. Promocijas darbā analizēta vadlīniju ietekme uz šiem kodiem, lai studentiem tie būtu pēc iespējas labāk uztverami.

Jēdziens “Video lekcija” aptver ļoti plašu ekspozīcijas stilu klāstu. Autore savā darbā aplūko tikai 11 dažāda stila materiālus, ko ir iespējams definēt kā video lekcijas un kas aptver galvenos video lekciju veidus. Lai sniegtu kvantitatīvu kritēriju, kas no šiem ir visvieglāk uztverams skatītājam un klausītājam, tiek izmantota entropija.

Tēmu “Video lekcijas kā koda analīze kombinētajā studiju modulī” autore izvēlējās divu iemeslu dēļ. Sākotnēji tika nolemts, ka autore vēlas saistīt savu promocijas darbu ar video lekcijām,



jo *Erasmus* programmā kā Rīgas Tehniskās universitātes toreizējā darbiniece autore apmeklēja Valensijas Politehnisko universitāti (*Universitat Politècnica de València*), Valensijā, Spānijā, kur iepazinās ar to, cik aktīvi šajā universitātē tiek veikti video lekciju ieraksti, kā arī gatavoti video ieraksti studijās par konkrētām tēmām.

Otrkārt, tālmācība, e-studijas, studiju kursi tiešsaistē kļūst arvien populārāki dažādās pazīstamās augstskolās visā pasaulē. Laikā, kad top darbs, pasaule ir *Covid-19* pandēmijas varā. Ne tikai universitātes, bet arī vidusskolas un pamatskolas pāriet uz attālinātajām mācībām. Saturs, kas top, galvenokārt tiek radīts spontāni, nevis pārdomāti, lai nosegtu vielu un sniegtu informāciju, jo neviens šādai situācijai nebija gatavs.

Tomēr tas ir devis arī iespēju cilvēkiem pārbaudīt savas spējas mācīties attālināti, kā arī ir devis vairāk laika, lai pievērstos neklātienas apmācībām.

Lai arī ir veikti dažādi pētījumi, nav konkrētu noteikumu vai vadlīniju, kā jāveido video lekcijas, lai tās būtu iespējami efektīvas (*Chandler, 1991*). Katrs mācībspēks vai katra augstskola var veidot vai ierakstīt lekcijas tā, kā vēlas. Tomēr studentiem ir svarīgi, lai lekcijas būtu pieejamas, viegli saprotamas. Tāpēc autore vēlējas iedziļināties šajā tēmā, jo Latvijā augstskolas vēl ir iesācējas šajā jomā un vadlīnijas vai sīkāka informācija par to, kādai jābūt video lekcijai, lai studentiem tā būtu vieglāk uztverama, palīdzētu gan mācībspēkiem, gan studentiem.

Būtu jāpievērš uzmanība arī tam, ka notiek aktīvo studentu paaudžu maiņa. Ja kādreiz tā sauktai *millenium* paaudzei pietika ar lekcijām sagatavotu *Power Point* prezentāciju, tad šobrīd, lai informācija sasniegtu  $Z^1$  paaudzes studentus, lekcijā jāizmanto dažādi, tajā skaitā arī interaktīvi elementi satura pasniegšanai (*Rubene, 2020*).

Tomēr video materiāliem vajadzētu būt tādiem, kas samazina kognitīvo slodzi un optimizē darba atmiņas izmantošanu (*Chandler, 1991*).

Nākamais jautājums – kā noskaidrot, cik laba un informācijas piesātināta ir video lekcija un kā to izmērīt. Kā izmērīt informāciju, ko sniedz video lekcija.

Precīza un visos gadījumos pieņemama informācijas definīcija nav izveidota. Tomēr ir izveidota t. s. sintaktiskās informācijas kvantitatīvā teorija, kas sekmīgi apraksta informācijas pārraidi un uztveršanu sakaru kanālos. Informācijas materiālais nesējs ir signāls. Signāls ir fizikāls process, kura parametra maiņa atveido pārraidāmo informāciju. Informatīvs signāls kalpo informācijas pārņemšanai. Informatīvo signālu veido informācijas avots, izmainot atbilstošās zīmes atbilstoši noraidāmajai informācijai.

1948. gadā K. Šenons savā rakstā “Sakaru matemātiskā teorija” (“*A Mathematical Theory of Communication*”) pirmo reizi izklāstīja sintaktiskās informācijas teoriju, lai apskatītu būtiskākos sakaru sistēmu aspektus. Šīs teorijas galvenās īpatnības ir, pirmkārt, lielais uzsvars uz varbūtību teoriju un, otrkārt, uz kodēšanas un dekodēšanas lielo nozīmi. No tā laika informācijas teorija ir

---

<sup>1</sup> Z paaudze ir demogrāfiskā grupa, kas seko tūkstošgades paaudzei pirms alfa paaudzes. Pētnieki un populārie plašsaziņas līdzekļi izmanto 20. gadsimta 90. gadu vidu un beigas kā sākuma dzimšanas gadus un 21. gadsimta otrā gadu desmita sākumu kā beigu dzimšanas gadus.

vairāk precizēta, paplašināta un tiek vairāk lietota praktiski tehniskajās sakaru pārraides sistēmās (Shannon, 1948).

Galvenā nozīme Šenona informācijas teorijā ir entropijai – vidējai informācijai uz vienu ziņojumu no ziņojumu avota.

Šenona entropija ir skaitlis, kas ir atkarīgs tikai no informācijas avota statistiskās dabas. Ja informācijas avotam ir vienkāršs modelis, tad entropiju var aprēķināt. Entropija ir ziņojumu avota informācijas mērs. Šajā darbā aprakstītā un analizētā entropija ir Šenona entropija, kas tiek mērīta bitos.

Savā bakalaura un maģistra darbā autore ir veikusi entropijas aprēķinus dažādiem rakstiskiem informācijas avotiem latviešu valodā, kā arī meklējusi dažādas sakarības, kas izskaidrotu entropijas izmaiņas.

Tā autore ir nonākusi līdz promocijas darba tēmai, vēloties izpētīt video lekciju kā kodu, tas ir, kā noteiktā veidā kodētu optisku un akustisku ziņojumu ansambli. No šī koda ir atkarīga video lekcijas kvalitāte. Aprēķināt video lekcijas entropiju un veikt izpēti, vai, ievērojot rekomendācijas par to, kāda ir studentiem visvieglāk uztveramā video lekcija, šādas video lekcijas entropijas aprēķins atšķiras no nejauši izvēlētas video lekcijas, un pierāda to, ka sniegtā sintaktiskā informācija lekcijai, kas ir filmēta pēc rekomendācijām, ir zemāka.

**Pētījuma objekts.** Dažādu e-studijās izmantoto video tehnoloģiju un mācību materiālu veidošanas programmatūru korelācija ar informācijas teorijas entropiju.

**Subjekts.** Video lekcijas kā koda analīze, nosakot entropiju ar aprēķinu un datorsimulācijas palīdzību.

**Pētījuma mērķis** ir ar entropijas palīdzību veikt video lekciju kā informācijas pārraides tehnoloģiju analīzi, lai noskaidrotu viegli uztveramas video lekcijas vadlīniju saikni ar entropiju. Globālais darba mērķis ir uzlabot video lekciju kvalitāti, lai tās tiktu veidotas un tehniski veidotas tā, lai studentiem tās būtu vieglāk uztveramas.

**Pētījuma apakšmērķis** ir izveidot metodi video lekcijas maksimālās entropijas aprēķināšanai. Realizējot šo mērķi, iespējams noteikt video un audio kanālu ietekmi uz entropiju un kopējo videolekcijas kā sakaru kanāla informāciju.

**Hipotēze.** E-studiju video materiālu entropiju nosaka izmantotās video lekcijas veids, izmantotās izstrādes tehnoloģijas un iepriekš definēti video lekciju kvalitātes nosacījumi.

#### **Aizstāvāmās tēzes**

1. Video lekcijas veids nosaka video lekcijas entropiju.
2. Entropija video lekcijai ir aprēķināms lielums. Aprēķināta dabīgas lekcijas nestā maksimālā absolūtā entropija, kas kalpo kā video lekcijas entropijas augšējā robeža, kā arī citi ar to saistītie lekcijas maksimālie informatīvie raksturlielumi – tās nestā informācija, lekcijas kanāla un tās apakškanālu informācijas caurlaides spējas.
3. Jo precīzāk ir ievērotas vadlīnijas lekcijas ierakstīšanā, jo zemāka ir video lekcijas entropija. Pastāv negatīva korelācija starp lekcijas vidējo video entropiju un izpildīto vadlīniju skaitu.

Pētījuma aktualitātes raksturojums un problēmas apzināšana ļauj **formulēt pētījuma jautājumus**.

1. Kāda ir video entropijas atkarība no izvēlētās tehnoloģijas un zinātniskajos pētījumos definēto vadlīniju ievērošanas.
2. Kā aprēķināt informācijas teorijas maksimālo entropiju video lekcijai.
3. Vai pastāv iespēja izmainīt audio un video vidējo entropiju.
4. Kā izmantot pētījuma rezultātus video lekciju kvalitātes uzlabošanai.

Lai rastu atbildes uz pētījuma jautājumiem, promocijas darbā tika definēti vairāki **pētījuma uzdevumi**.

1. Veikt zinātnisko rakstu literatūras analīzi, lai atrastu tos tehnoloģiskos parametrus, kas jāievēro, lai radītu viegli uztveramas video lekcijas noteikumus jeb vadlīnijas.
2. Veikt video lekcijas maksimālās entropijas aprēķinu.
3. Atlasīt dažāda veida video lekcijas, ņemot vērā tehnoloģiju un izmantoto programmatūru, kurām veikt analīzi.
4. Noteikt katras video lekcijas vidējo video entropiju, vidējo spektrālo audio entropiju un vidējo temporālo audio entropiju.
5. Veikt rezultātu analīzi un izstrādāt dokumentāciju atbilstoši vadlīnijām.
6. Sniegt ierosinājumus video lekciju kvalitātes uzlabošanai.

Pētījuma robežas iezīmē pētījuma priekšmets, kas nosaka, ka tiek analizētas video lekcijas, kas filmētas dažādos veidos – gan auditorijā, gan video studijā.

Pētījuma bāzi veido 11 video lekcijas (analizēto lekciju skaits).

### **Pētījuma metodes**

1. Teorētiskās izpētes metodes:

1.1. zinātniskās un metodiskās literatūras analīze – deskriptīvā un salīdzināšanas metode.

2. Empīriskās izpētes metodes:

2.1. datu ieguves metodes:

2.1.1. datu izgūšana no informācijas sistēmas.

2.2. Datu apstrādes un analīzes metodes:

2.2.1. grafiskās analīzes metode kvalitatīvo un kvantitatīvo pētījumu datu klasificēšanai un analīzei izmantotas programmas – datu apstrādes programma *Matlab*;

2.2.2. datu grafiskā attēlošana – *Matlab* un *MS Excel*;

2.2.3. kvalitatīvo datu apstrāde;

2.2.4. kvantitatīvo un kvalitatīvo datu interpretācija.

Promocijas darba izstrāde notika četros posmos.

- Pētījuma pirmais posms (2014. gada 1. septembris–2015. gada 1. septembris). Pētāmās tēmas izvēle. Sava esošā pētījuma sasaiste ar e-studiju un video lekciju tēmu.
- Pētījuma otrais posms (2015. gada septembris–2017. gada septembris). Pētāmās problēmas un pretrunu konstatācija, pētāmās problēmas aktualitātes raksturojums un izpēte, pētījuma dizaina izstrāde. Padziļināta literatūras analīze par video lekciju veidošanu, kombinēto

apmācību moduli, tā priekšrocībām un trūkumiem, par entropijas izmantošanu video analizē, kā arī par tehniskiem risinājumiem entropijas mērīšanā. Pētījuma zinātniskās literatūras un avotu atlase un klasifikācija. Video lekcijas uzņemšanas rekomendāciju izstrāde.

- Pētījuma trešais posms (2018. gada septembris–2020. gada decembris). Maksimālās strukturālās informācijas aprēķināšana. Turpmāko darbu plāna izveide. Video lekciju skatīšanās paradumu analīze. Tehnisko risinājumu apzināšana un izvēle.
- Pētījuma ceturtais posms (2020. gada janvāris–2021. gada decembris). Video lekciju atlasīšana un analizēšana. Kopējā pētniecības darba rezultātu analīze un interpretācija.

Promocijas darba priekšizstāvēšana fakultātē notika 2022. gada 27. janvārī. Darba recenzents – profesors Atis Kapenieks.

Promocijas darba struktūru veido anotācija, ievads, trīs daļas, 12 nodaļas, 53 apakšnodaļas, rezultāti, nobeigums un secinājumi, izmantotās literatūras un avotu saraksts, kā arī pielikums. Promocijas darbā iekļauti 57 attēli un 20 tabulas. Kopumā minēti 106 literatūras avoti latviešu un angļu valodā, bet darba gaitā tika izlasīts vismaz piecas reizes lielāks skaits.

#### **Pētījuma zinātniskā novitāte un teorētiskā nozīmība**

1. Pirmo reizi veikts maksimālās informācijas un entropijas aprēķins dabīgai lekcijai kā informācijas pārraides kanālam. Iegūtie rezultāti ir augšējā robeža video lekcijai.

2. Veikti detalizēti pētījumi *Matlab* vidē par dažādu video lekciju veidošanai izmantoto tehnoloģiju un programmatūru ietekmi uz entropiju un studentu uztveršanas spēju.

#### **Pētījuma praktiskā nozīmība**

1. Sniegtas tehniskas vadlīnijas video lekciju ierakstīšanai.

2. Izveidota datortehnoloģijās balstīta metodika, kā veikt e-studiju vajadzībām radīta video sakaru kanāla kā koda analīzi.

3. Izstrādāti iespējamie inovatīvie risinājumi video lekciju izvērtēšanā jauniešu izglītošanai.

**Autores ieguldījums un pētījumu rezultātu aprobācija. Promocijas darbā aprakstītie pētījumi ir izstrādāti Liepājas Universitātes Dabas un inženierzinātņu fakultātē.**

Dati par video lekciju skatīšanos iegūti no Valensijas Politehniskās universitātes, sadarbojoties ar *Carlos Turro Ribalta*, tīkla un mediju pakalpojumu direktora vietnieku. *Matlab* programmas pielāgošana vidējās entropijas noteikšanai tika izstrādāta kopā ar Lauri Cikovski, Rīgas Tehniskās universitātes HPC centra vadītāju un vadošo pētnieku. Darbs izstrādāts Rīgas Tehniskās universitātes emeritētā profesora *Dr. habil. phys. Andra Ozola* vadībā.

Zinātniskās pētniecības darba rezultāti publicēti piecos zinātniskos recenzētos izdevumos.

1. Alksne L., Cikovskis L., Ozols A. (2022). Entropy of video lecture (iesniegts publicēšanai BJMC).
2. L. Alksne, A. Ozols. Maximum Shannon information delivered in a lecture. *Latvian Journal of Physics and Technical Sciences*, 2022, No. 2, pp. 12–22. DOI:10.2478/lpts-2022-0008. SCOPUS.

3. Alksne, L., Jansone, A., & Bērzkalne, Z. (2019). Benefits from analyzing video lecture logs with leading business analytics tools. *Baltic Journal of Modern Computing*, 7 (3), 393–404. doi:10.22364/bjmc.2019.7.3.06.
4. Alksne, L. (2016) How to produce video lectures to engage students and deliver the maximum amount of information. *Society. Integration. Education Proceedings of the international scientific conference*. Volume ii, may 27th–28th, 2016. 503–516.
5. Bajarune, L., & Ozols, A. (2015). Latvian language as a code in different communication channels. Paper presented at the *Vide. Tehnologija. Resursi – Environment, Technology, Resources*, 11–16. doi:10.17770/etr2015vol3.182 Retrieved from [www.scopus.com](http://www.scopus.com).

# 1. Informācijas teorija

## 1.1. Informācija un tās definīcija sakaru teorijā

Promocijas darba teorētiskajā daļā sniegts ieskats par to, kas ir informācija un kāda ir tās definīcija sakaru teorijā. Darbā sniegts skaidrojums par to, kā informācija ir saistīta ar entropiju – galveno parametru, kas tiek noteikts un analizēts darba pētniecības daļā. Informācijas pārraidīšanai ir bijusi milzīga nozīme no pagātnes līdz pat mūsdienām, tāpēc darbā apskatīta arī informācijas kodēšana, kodu izcelsme un vēsture, kā arī datu saspišana. Šajā promocijā darbā uzsvars tiek likts uz Šenona informāciju, kas tiek izmantota sakaru kanālos, jo video lekcijas satura pārraidi klausītājiem un skatītājiem arī var uzskatīt par sakaru kanālu. Šenona informācija ir apskatīta un analizēta promocijas darba gaitā, atšķirot to no informācijas satura analizēšanas.

Ar informācijas apmaiņas sistēmu ir iespējams regulēt attiecības starp atsevišķiem sabiedrības locekļiem un to grupām; tā nodrošina cilvēku kopas pastāvēšanu ārējā vidē. Katrs indivīds sazinoties saņem vai arī sniedz informāciju. Latīņu valodā vārds “*informatio*” nozīmē izskaidrojumu, izklāstu, skaidrojumu. Informācijas apmaiņas sistēma cilvēku sabiedrībā ir tik tālu attīstījusies, ka tajā jau kopš cilvēces pastāvēšanas sākumiem vienlaikus ar dabiskajiem informācijas apmaiņas līdzekļiem plaši izmanto arī mākslīgos. Attīstot un pilnveidojot mākslīgos informācijas pārraides un apmaiņas līdzekļus, cilvēks mūsdienās ir izveidojis bagātīgu tehnisko līdzekļu arsenālu. Cilvēku sabiedrības attīstības līmeni lielā mērā raksturo tās rīcībā esošā tehniskā informācijas apmaiņas sistēma.

Informācija ir viena no matērijas eksistences formām; tā ir neatņemama materiālās pasaules īpašība. Var pat uzskatīt, ka katrā materiālās pasaules stāvoklī vai norisē slēpjas informācija. Informācijas jēdzienu stipri izcēla zinātnes nozare – kibernetika. Vārds “kibernetika” ir cēlies no grieķu valodas. Tulkojumā tas nozīmē – mākslu vadīt. Mūsdienās ar kibernetiku saprot zinātnes nozari par informācijas iegūšanas, glabāšanas, pārraides un pārveidošanas likumībām un to pētīšanu sarežģītās sistēmās. Viduslaikos šo zinātņi praktiski aizmirsā. Savos zinātniskajos darbos kibernetikas jēdzienu no jauna sāka lietot izcilais franču zinātnieks A. Ampērs (1834). Pēc Ampēra nāves (1836) to atkal aizmirsā, līdz 1984. gadā mūsdienu kibernetikas un informācijas teorijas pamatlicējs, ASV zinātnieks Norberts Vīners šo maz pazīstamo vārdu izmantoja savas grāmatas nosaukumā. Līdz ar šīs grāmatas iznākšanu (1984) sāka strauji attīstīties informācijas teorija.

Informācijas jēdzienu parasti saista ar divu objektu – informācijas avota un patērētāja – eksistenci. Precīza un visos gadījumos pieņemama informācijas definīcija nav izveidota.

Pētot tās dabas un sabiedrības parādības, kas saistītas ar informāciju, un pētot informācijas īpašības, jāsaprot ar informācijas jēdzienu visdažādākajos aspektos – var interesēties par informāciju, kas slēpjas atsevišķu vārdu vai to sakopojumu jēgā (semantiskā jeb metriskā informācija); informāciju, kas raksturo tās saņēmēja relatīvo zināšanu pieaugumu (pragmatiskā jeb sintaktiskā informācija); informāciju, kas ietverta dažādos mākslas darbos (idealizētā informācija). Liela un praktiska nozīme ir informācijas tehniskajai jēgai, iedomājoties informāciju kā simboliku,

ko izmanto sazināšanās nolūkiem. Informācija ir definēta arī šādi: informācija ir ziņas, ar kurām operē kā ar objektu, tās noraidot, sadalot, pārveidojot, uzglabājot vai tieši izmantojot.

Informācijas materiālais nesējs ir signāls. Signāls ir fizikāls process, kura parametra maiņa atveido pārraidāmo informāciju. Informatīvs signāls kalpo informācijas pārveidošanai. Informatīvo signālu veido informācijas avots, izmainot atbilstoši noraidāmajai informācijai atbilstošās zīmes. Zīmes var būt secīgi rakstīti burti. Ziņojums ir zīmju kopa, kas attālo pārraidāmo informāciju. Ziņojuma ansamblis ir cipari, alfabēts u. tml. Ziņojuma ansambļa secīgs sakārtojums veido ziņu. Tātad – informatīvo signālu lieto ziņojuma pārveidošanai jeb ziņošanai. Signāla uztvērējs uztver signālu kā ziņojumu, tālāk ziņojums tiek pārveidots informācijā, ko saņem informācijas patērētājs. Informācijas pārveidošanas procesu ziņojumā sauc par kodēšanu, savukārt pretējo procesu – par dekodēšanu.

Komunikācija ir apmaiņšanās ar ziņojumiem – sazināšanās. Tehniskās informācijas pārraidē vispārējās komunikācijas shēmas galvenie elementi ir ziņojuma avots, sakaru kanāls un ziņojuma saņēmējs. Šāda sistēma ir domāta informatīva ziņojuma pārraidei (Grabinskis, 1984).

### **Secinājumi**

1. Dots vispārīgs priekšstats par informācijas jēdzienu.
2. Definēti sintaktiskās informācijas pamatjēdzieni – zīme, ziņojums, signāls, kodēšana un dekodēšana.

## 1.2. K. Šenona definīcija diskrētu ziņojumu ansamblim

1948. gadā K. Šenons savā rakstā “Sakaru matemātiskā teorija” (“*A Mathematical Theory of Communication*”) izstrādāja sintaktiskās informācijas teorijas pamatus, lai matemātiski apskatītu būtiskākos sakaru sistēmu aspektus un gūtu iespēju optimizēt sakaru sistēmas (Shannon, 1948). Šīs teorijas galvenie momenti ir, pirmkārt, lielais uzsvars uz varbūtību teoriju, otrkārt, kodera un dekodera lielā nozīme. No tā laika informācijas teorija ir vairāk precizēta, paplašināta un tiek vairāk praktiski lietota tehniskajās sakaru pārraides sistēmās.

Galveno nozīmi Šenons piešķir tieši informācijas pārraidīšanas veidam un uztveršanai. Šenons pilnībā ignorē to, vai teksts ir nozīmīgs, saprotams, pareizs, nepareizs vai bez nozīmes. Līdzīgi tiek izslēgti arī jautājumi par to, kas informāciju raida un kas ir uztvērējs. Pēc Šenona informācijas teorijas izriet, ka ir mazsvarīgi, vai teksts ir secīgs un nozīmīgs, vai burti ir izvēlēti uz labu laimi. Šeit parādās paradokss – uz labu laimi izvēlēti burti sasniedz maksimumu informācijas saturā, taču, kur teksts ir ar lielāku nozīmi un lingvistisko dažādību, tas atbilst mazākai vērtībai (Garcez, 2011).

Ziņojumu ansamblis jeb ziņojumu avota alfabēts ir ziņojumu kopa, ko ziņojumu avots izmanto informācijas pārraidei. Tie var būt burti, cipari, pikseļi utt.

Teletaips<sup>2</sup> un telegrāfs ir divi visvienkāršākie piemēri informācijas pārraidīšanai diskrētā kanālā. Diskrēts kanāls nozīmē sistēmu, kur tiek pārraidīts galīgs skaits elementāru simbolu  $S_1, \dots, S_n$  no viena punkta uz otru. Katrs simbols  $S_i$  pieņem noteiktu laiku  $t_i$  sekundēs.

Teletaipa gadījumā, kur visi signāli ir ar vienu pārraidīšanas ilgumu un jebkādu kārtību no 32 simboliem, tas nozīmē, ka katrs simbols attēlo 5 bitu informāciju. Ja sistēma pārraida  $n$  simbolus vienā sekundē, tad var teikt, ka kanāla caurlaides spēja ir  $5n$  biti sekundē. Tas nenozīmē, ka teletaipa kanāls vienmēr pārraidīs ar šādu ātrumu – tas ir maksimālais ātrums, tas, vai maksimums tiks sasniegts un kāds būs pārraidīšanas ātrums, ir ļoti atkarīgs no informācijas avota (Shannon, 1948).

Diskrēta kanāla definīcija ir tikai viena piektā daļa no Šenona raksta, vēl tiek dotas definīcijas diskrētam trokšņainam kanālam, nepārtrauktas pārraides kanālam, kā arī dažādas teorēmas un aprēķinu formulas (Carlson, 1986; Ozols, 2005).

### Secinājumi

1. Raksturota K. Šenona pieeja diskrētu ziņojumu ansambļa definīcijai.
2. Uzsvērta Šenona pieejas statistiskā daba.

---

<sup>2</sup> Teletaips – tādu ierīču kā lentes perforatoru un lappušprinteru kopums, ko izmanto telekomunikācijai.



### 1.3. Ziņojuma ansambļa galvenie raksturlielumi

Entropija ir ziņojumu ansambļa vidējā informācija uz vienu ziņojumu. Ziņojumu piemēri ir burti, vārdi, attēli. Viena ziņojuma nesto informāciju, ja dots ziņojumu ansamblis  $a_1, a_2, \dots, a_n$  ar varbūtībām  $p_1, \dots, p_n$ , definē kā

$$I(a_k) = \log_2 \frac{1}{p_k}.$$

Visu varbūtību  $p_k$  summai jābūt vienāgai ar 1. Sakaru kanāla caurlaides spēja ir vienāda ar tajā iespējamo maksimālo informācija pārraides ātrumu.

$$C = R_{\max},$$

kur  $C$  – kanāla caurlaides spēja;  $R$  – pārraides ātrums:

$$R = \frac{\partial I(t)}{\partial t}. \quad (1)$$

#### 1.3.1. Entropija

Entropija ir situācijas nenoteiktības kvantitatīvs mērs, ko plaši lieto termodinamikā un informācijas teorijā. Entropijas jēdzienu izmanto dažādos informācijas optimālās kodēšanas problēmu pētījumos. Entropijas jēdziens informāciju teorijā apraksta, cik daudz nejaušības ir signālā vai nejaušā notikumā.

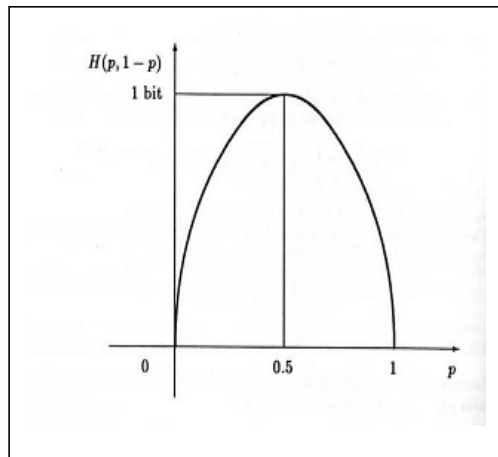
Klods E. Šenons savā rakstā “Sakaru matemātiskā teorija” pierādīja, ka pastāv noteikta robeža informācijas saspiēšanai bez zudumiem. Šī robeža ir entropija, kas tiek apzīmēta ar  $H$ . Entropijas  $H$  vērtība ir atkarīga no informācijas avota. Lai bez zudumiem saspieestu informāciju, saspiešanas koeficientam jābūt tuvu entropijas lielumam.

Entropija ir skaitlis, kas ir atkarīgs tikai no informācijas avota statistiskās dabas. Ja informācijas avotam ir vienkāršs modelis, tad entropiju var aprēķināt.

Izvēloties patvaļīgu ziņojumu ansambli  $X = (X_1, X_2, X_3, X_4, \dots)$ , kas ir teksts, un ziņojumi ir burti, ir iespējams aprēķināt vairākas entropijas kārtas. Katra kārta atšķiras – jo augstāka kārta, jo tiek aprēķināta vairāk blakus stāvošo burtu ietekme uz vienu burtu.

Entropijas  $H$  īpašības:

- 1)  $H(X) \geq 0$ ;
- 2) entropija ir aditīvs lielums diviem neatkarīgiem ziņojumu ansambļiem  $X$  un  $Y$



1.3.1. att. Entropijas atkarība no varbūtības divu ziņojumu ansambļa gadījumā.

$$H(X + Y) = H(X) + H(Y); (2)$$

3) entropija ir ierobežots lielums  $H \leq H(\max) = \log_2 N$ , kur  $N$  ir ziņojumu skaits ansamblī.

Nulltās kārtas entropiju rēķina pēc formulas

$$H(0) = H(\max) = \log_2 N \text{ biti/simbolu.} \quad (3)$$

Ja  $N$  ir burtu skaits alfabētā, tad tiek ņemta vērā arī atstarpe, ja tiek rēķināta entropija rakstītai valodai. Šajā gadījumā tiek pieņemts, ka burti nav atkarīgi viens no otra.  $H(0)$  tiek arī uzskatīta kā  $H(\max)$ , jo nevienas kārtas entropija nevar būt augstāka par nulltās kārtas entropiju.

Pirmās kārtas entropiju aprēķina pēc formulas

$$H(1) = \sum_{i=1}^m p_i \log_2 p_i \text{ biti/simbolu,} \quad (4)$$

kur  $p_i$  – ziņojuma (simbola) varbūtība, arī šajā gadījumā tie ir neatkarīgi viens no otra.

Otrās kārtas entropiju aprēķina pēc formulas

$$H(2) = \sum_{i=1}^m p_i \sum_{j=1}^m p_{ji} \log_2 p_{ji} \text{ biti/simbolu,} \quad (5)$$

kur  $p_{ji}$  – nosacītā varbūtība ziņojumam  $j$ , ja iepriekšējai ir bijis, piemēram, burts, tomēr ziņojumam nav obligāti jābūt burtam. Tas var būt arī, piemēram, pikselis. Lai aprēķinātu otrās kārtas entropiju, ir jāņem vērā tas, kāds simbols atrodas pirms simbola, jeb visas iespējamās divu simbolu kombinācijas.

Trešās, ceturtās un lielākas kārtas entropiju aprēķina pēc tādas pašas formulas, tikai ar katru kārtu tiek ņemti vērā vairāk ziņojumu, kas atrodas pirms attiecīgi definētā ziņojuma.

Entropijas minimums  $H(X) = 0$ , un to sasniedz tie ziņojumi, kuriem sniegtās informācijas varbūtība ir vienāda ar 1. Maksimālā entropija  $n$  – simbolu ziņojumam  $A$  ir  $H(X) = \log_2 m$  biti, un tā tiek sasniegta tad, ja simboliem ir vienāda varbūtība

$$p_i = \frac{1}{m}. \quad (6)$$

Beztrokšņu kanālā vienmērīgas pārraides gadījumā

$$R = \frac{H(X)}{\tau}, \quad (7)$$

kur  $\tau$  – viena ziņojuma pārraides vidējais ilgums. Nevienmērīgas pārraides gadījumā:

$$R = \frac{\partial I(t)}{\partial t}. \quad (\text{Jehonovičs, 1984})$$

### 1.3.2. Savstarpējā informācija

Savstarpējā informācija starp diviem diskrētiem, brīvi izvēlētiem ziņojumu ansambļiem  $X$  un  $Y$  tiek apzīmēta ar  $I(X, Y)$ . Savstarpējā jeb relatīvā informācija ir lielums, kas norāda informācijas

daudzumu, kas tiek dalīts starp trokšņaina kanāla ieeju (ansamblis  $X$ ) un izeju (ansamblis  $Y$ ). Savstarpējā informācija tiek mērīta bitos.

Tātad savstarpējā informācija ir informācijas mērs starp  $X$  un  $Y$ . Ja  $X$  un  $Y$  ir neatkarīgi, tad  $X$  nesatur nekādu informāciju par  $Y$ , un otrādi, tādā gadījumā to kopējā informācija ir nulle. Ja  $X$  un  $Y$  ir identiski, tad visa  $X$  nodotā informācija tiek dalīta ar  $Y$ . Ja  $X$  nesāņem nekādu atbildi no  $Y$ , un otrādi, tad kopējā informācija ir vienāda ar viena paša  $X$  (vai  $Y$ ) nosūtīto informāciju, sauktu par  $X$  entropiju.

Savstarpējā informācija tiek aprēķināta šādi:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X), \quad (8)$$

kur  $H(X|Y)$  un  $H(Y|X)$  ir ansambļu nosacītās entropijas –  $X$  attiecībā pret  $Y$  un  $Y$  attiecībā pret  $X$ . Savstarpējās informācijas maksimālā vērtība ir vienāda ar trokšņaina informācija kanāla caurlaides spēju (biti/simbols).

Savstarpējās informācijas īpašības

1. Savstarpējā informācija ir simetriska

$$I(X;Y) = I(Y;X).$$

2. Savstarpējā informācija vienmēr ir nenegatīva

$$I(X;Y) \geq 0.$$

3. Savstarpējā informācija ir saistīta ar kanāla kopējo ieejas un izejas entropiju šādi

$$I(X;Y) = H(X) + H(Y) - H(X,Y), \quad (9),$$

kur  $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$  ir ansambļu  $X$  un  $Y$  kopējā entropija.

Relatīvo informāciju lieto trokšņainu sakaru kanālu gadījumā informācijas caurlaides spējas noteikšanai, kā arī informācijas pārraides ātruma definēšanai

$$R = \frac{H(X,Y)}{\tau}. \quad (10) \text{ (Carlson, Bruce, 1986)}$$

### 1.3.3. Redundance

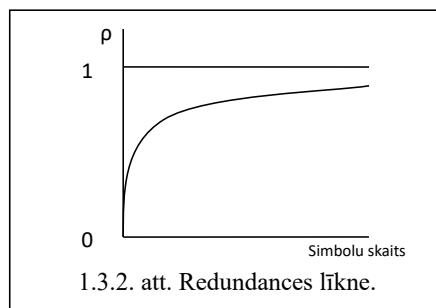
Aprēķinot entropiju, ir iespējams aprēķināt arī ziņojuma avota redundanci (1.3.2. att).

$$\rho = 1 - \frac{H(A)}{H_{\max}(A)} = 1 - \frac{H_n(A)}{H_0(A)} \quad (11)$$

Ar informatīvo redundanci saprot tādu dublējošu vai papilddatu ieslēgšanu sistēmas datu masīvos, kuru izņemšana nemazina šo masīvu adekvātumu

reālajiem objektiem, ko tie apraksta (Mackay, D. J. C, 2006). Redundance informācijas teorijā bitu skaits, kas tiek lietoti, lai nosūtītu ziņu mīnus patiesās informācijas daudzums bitos (Latvian Academy of Sciences, 2015).

Datu saspiešana ir veids, kā izslēgt nevēlamo redundanci, taču, ja ziņojuma pārraide jāveic trokšņainā kanālā ar ierobežotu ietilpību, tad redundance ir vēlama. Tātad redundance, piemēram,



1.3.2. att. Redundances līkne.

mūsu valodā ir tie vārdi, kurus mēs pasakām, bet arī bez šiem vārdiem informācija būtu saprotama. Par to var pārliecināties ikreiz, kad tiek sūtīta īsziņa, cenšoties pateikt pēc iespējas vairāk, izmantojot pēc iespējas mazāk simbolu, lai veidotos tikai viena īsziņa – tiek izlaisti vārdi, burti vai pat rakstīti bez atstarpēm, taču teksts vienlīdz ir saprotams.

Tksts bz ptkšņm. Tekstsbezatstarpēm. Ja pēc šī teksta var rekonstruēt teikumu nozīmi, kas ir rakstīts bez patskaņiem un bez atstarpēm, tas nozīmē, ka šie patskaņi vai atstarpes šai ziņai ir redundanti.

Tomēr redundantie vārdi vai burti ir tie, kas ļauj mums saprast informāciju, ja, piemēram, runājam trokšņainā vietā. “Kur ir sāls paciņa?” Ja, pasakot šo teikumu, garām brauks trokšņaina smagā mašīna, sāls var izklausīties kā “māls” vai “tāls”, bet redundantais vārds “paciņa” tomēr liks saprast jautājumu. Šis ir ļoti primitīvs piemērs, bet vairumā gadījumos, ja teikumā būs izlaists viens vārds, uzminēt trūkstošo vārdu ir vienkārši, ja būs izlaists burts, tad tas, ka to uzminēsim, ir gandrīz viennozīmīgi.

*Saksaņā ar kdāas aņņļu uvinersiātetes pjuētīmu, nav sravīgi kādā sbebē ir srākatoti bturi vādrā, veigīni savgīri lai primias un pdējēais butrs btūu sāavs vtieās. Pēārije var būt plnīgi saukjati un jūs jjooprām vēiearst lsāt bez pomblrēm. Tas ntoeik tēpāc, ka pabestiā mēs nāealsm kratu btruu atšseiķvi, bet gan vrādu kā vneiu velesu (<http://klab.lv/>).*

Ja 50 % valodas ir redundanta, tad būtu iespējams ietaupīt 50 % ieguldījumu, kas nepieciešami šīs valodas ziņu elektroniskai nosūtīšanai. Kaut kas līdzīgs notiek tad, kad fails tiek saspiešts, bet ja kaut kur pārraidīšanā gadās kāds troksnis un tiek iznīcināts kāds no simboliem, ko ietver saspieštais fails, tad ir neiespējami atjaunot oriģinālu failu.

Ziņu digitālajā kodēšanā redundancei ir svarīga nozīme, izmantojot kodēšanu ar vieninieku pāra skaitu. Burtam A binārajā sistēmā atbilst 01000001. Tātad, lai pārraidītu burtu A, nepieciešami 8 biti, ir jāpārraida šīs astoņas zīmes. Bet, ja līnija ir ar traucējumiem un ieviešas kļūda, un tiek saņemts 010000?1, un nevar noteikt, kāds burts tas ir. Tas var būt A, ja trūkstošā zīme ir 0, bet, ja šī zīme ir bijusi 1, tad tas ir bijis burts C. Protams, ka normālā kontekstā nav problēmu to saprast, taču, kas notiek, ja redundance jau ir izmantota un fails ir saspiešts. Tādā gadījumā, pievienojot paritātes bitu, kas būtu vēl viens redundants bits, būtu iespējams problēmu atrisināt. Ja skaitļu summa ir pāra skaitlis, tad tiek pievienota 0, ja nepāra, tad – 1. Tātad, saņemot 010000?10, pievienotā nulle pasaka priekšā, ka vajadzēja saņemt A (01000001); ja tiktu saņemts 010000?11, tad pievienotais 1 pasaka, ka saņemtais burts ir C (01000011). Ja ir gaidāms patiesi trokšņains kanāls, ir iespējams vienoties par paritātes bitu sūtīšanu pēc katriem 4 bitiem. Tas var šķīst nevajadzīgi – sūtīt liekus bitus, kas patiesībā nav vajadzīgi. Taču, ja teksts no 10 000 zīmēm būtu saspiešts līdz 8000, izslēdzot redundantos simbolus, katrai nosūtītajai zīmei jāpievieno arī viens paritātes bits – tie būtu 8000 paritātes biti. 8000 biti ir ekvivalenti 1000 zīmēm, tas nozīmē, ka tas ir izdevīgāk (*Underwood J. M., 2006*).

Šī metode ir koriģējošais kods.

Redundance ir cieši saistīta ar saspiežamības koeficientu, ko aprēķina šādi:

$$r = \frac{H(A)}{H_{\max}(A)} = \frac{H_n(A)}{H_0(A)} = 1 - \rho. \quad (12)$$

Autore secina, ka, lai arī informācijas teorija nav tik sena, tā ir ļāvusi strauji attīstīties informācijas pārraides tehnoloģijām. K. E. Šenona veiktajos aprēķinos un atklājumos tiek balstīta mūsdienu informācijas apmaiņa un video kodēšana.

### **Secinājumi**

1. Definēti Šenona informācijas teorijas galvenie jēdzieni diskrētā kanālā – entropija, savstarpējā informācija, redundance, saspiestamība, pārraides ātrums un to saistība.
2. Uzsvērta tās lielā nozīme informācijas pārraides tehnoloģijās.

## 2. Valodu entropija

Dabiskā valoda ir viens no galvenajiem komunikācijas līdzekļiem. Kā zīmju sistēma tā ir instruments informācijas novadišanai un izplatīšanai. Ar savas zīmju sistēmas, simbolu un to kombinācijas, savienošanas un salikuma likumu palīdzību valoda ir unikāls komunikācijas kods, un saskaņā ar to tiek lietoti arī citi neverbālie kodi.

Valoda kopš pirmsākumiem ir arī informācijas uzkrāšanas līdzeklis un veido struktūru un “navigācijas sistēmu” glabātajai informācijai. Tieši tādēļ valoda veido konceptuālus un pat uztveres likumus. Valoda kā komunikācijas līdzeklis ir atvērts kods ar īpašiem attīstības likumiem.

Ar informāciju pārblīvētā telpā, kurā modernās tehnoloģijas atļauj ciešu mijiedarbību starp cilvēkiem, kuru atrašanās vietas ir lielā attālumā viena no otras un kurā informācija plūst un tiek apmainīta tādos ātrumos kā mūsdienās, jāmaina arī valodas pētīšanas un dabiskās valodas normu noteikšanas pieeja, lai valoda attīstītos nevis kā abstrakts līdzeklis abstraktai komunikācijai, bet kā bagāts, radošs rīks, ar kura palīdzību var glabāt un nodot tālāk informāciju par jauno un ātri mainīgo realitāti. Tāpēc ir nepieciešamas ne tikai valodas lietojuma, gramatikas, sintakses un stila pētījumi, bet ir svarīgi pievērst uzmanību arī valodas statistiskajām īpašībām (*Papadimitriou, 2010*).

Jau rakstot bakalaura darbu par valodas entropijas tēmu, autorei diemžēl nācās secināt, ka par mūsu valodas entropiju un statistiskajām īpašībām nav nekādu pētījumu, nevienu aprēķinu, turpretī angļu un krievu valodā runājošie savu valodu ir izpētījuši ļoti smalki – veikuši aprēķinus dažāda žanra literatūrai – gan dzejai, gan romāniem, kā arī telegrammām. Entropijas aprēķini veikti ļoti augstās pakāpēs, krievu valodā – pat līdz astotajai. Bakalaura darbā autore aprēķināja entropiju preses sakaru kanālam, savukārt maģistra darbā dažādiem avotiem, lai tos varētu salīdzināt un meklēt sakarības starp žanru, kādā darbs rakstīts, un entropijas skaitli.

Valodu entropija

$H(N)$	Angļu valoda	Krievu valoda
$H(0)$	4,75	5
$H(1)$	4,07	4,05
$H(2)$	3,36	3,52
$H(3)$	2,77	
$H(8)$		2

Šenons šādus skaitļus ieguvis, analizējot septiņus literāros darbus. Pirms aprēķinu sagatavošanas lielie burti tika pārveidoti uz mazajiem, tika izņemti visi skaitļi un speciālās rakstzīmes, kopā bija 5 086 936 simboli (*Kulkarni, 2002*). Valodu entropija tiks aplūkota arī promocijas darba 5. nodaļā.

Autore secina, ka entropijas lietojums ir ļoti plašs. To iespējams izmantot dažādās nozarēs un dažādiem informācijas nesējiem.

### Secinājumi

1. Detalizēti izskatīta Šenona entropijas lietošana biznesa inteliģences procesos. Izcelta tās svarīgā nozīme.
2. Analizēta entropijas kā informācijas drošības mēra loma.
3. Aplūkota valodu entropijas nozīme lingvistikā. Konstatēts, ka latviešu valodas entropija līdz autores darbiem nav aprēķināta un analizēta. Tas tiks darīts tālāk promocijas darba izklāstā.

### 3. Latviešu valodas kā koda salīdzinājums dažādos komunikāciju kanālos

Videolekcija ietver skaņas kanālu, kur lektors runā kādā valodā, tāpēc promocijas darbā valoda tiks aplūkota kā kods. Konkrēti – latviešu valoda. Jebkura valoda ir kods, ar ko cilvēki kodē savstarpējās saziņas informāciju. Šajā gadījumā tiks aplūkota rakstītā valoda.

Valodā par ziņojumu var uzskatīt:

- 1) burtu;
- 2) vārdu;
- 3) teikumu.

Latviešu valodā skaņa atbilst burtam, tāpēc burtu var uzskatīt par koda simbolu, kas atbilst ziņojumam – skaņai.

Promocijas darba autore izvēlējās burtu kā kodu simbolu, kas raksturo kodu, jo burts ir valodas pamats. Valoda ir kā kods, ar ko nodot informāciju, un burti ir kodu kombinācijas, ar ko kodē informāciju.

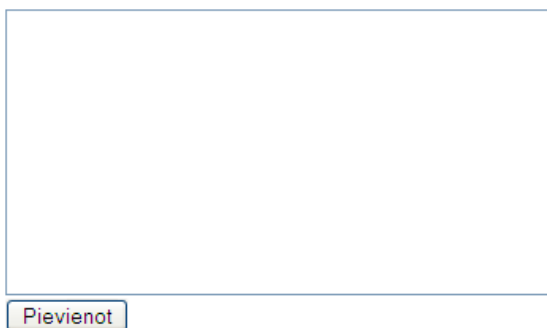
Savā pētījumā autore izvēlējās salīdzināt entropiju, ņemot latviešu valodas tekstus no preses (trīs raksti), literatūras avotiem un latviešu valodā rakstītiem likumiem.

Lai veiktu aprēķinus, tika izveidota programma. Tā veidota interneta mājaslapas veidā, izmantotā programmēšanas valoda ir *php*, izmantota ir arī *html* valoda. Visi aprēķini veikti, izmantojot *php* skriptu un *MySQL*<sup>3</sup> datubāzes vadības sistēmu, kas izvietoti uz *Apache web* servera (*web.hc.lv*, 2008).

Programmas skripts – 1. pielikumā.

#### Programmas darbošanās princips

1. Analizējamais teksts – tiek ierakstīts vai iekopēts teksts pievienošanas formā (3.1. att).



3.1. att. Programmas logs.

Pēc teksta ievadīšanas jāspiež poga “Pievienot”.

---

<sup>3</sup> Daudzpavedienu daudzlietotāju *SQL* datubāzu vadības sistēma (*DBVS*), kas ir uzstādīta uz vairāk nekā sešiem miljoniem datoru visā pasaulē.



2. Programma aizstāj visus lielos burtus ar mazajiem, noņemtas visas pieturzīmes, atstarpes aizvietotas ar simbolu „\_”.

3. Teksts tiek apstrādāts, izdalot trīs burtu kombinācijas, divu burtu kombinācijas un burtus pa vienam. Programma, analizējot tekstu, datubāzē ieraksta katru jauno kombināciju, bet, ja kombinācija tiek atrasta atkārtoti, ierakstīto kombināciju skaits datubāzē palielinās par vienu.

4. Kad teksta analīze pabeigta, iegūto kombināciju skaits tiek ielikts formulās, lai aprēķinātu teksta entropijas pirmo, otro un trešo kārtu, tas ir, tiek izvadīti rezultāti  $H(1)$ ,  $H(2)$  un  $H(3)$ .

Maksimālo latviešu valodas entropiju autore aprēķināja pēc šādas formulas:

$$H(0) = \log_2 m \text{ biti/simbolu,} \quad (13)$$

kur  $m$  – burtu skaits alfabētā un atstarpe, latviešu valodā  $m = 33 + 1$ .

$$H(\max) = H(0) = 5,0875 \text{ biti/simbolu}$$

Pirmās kārtas entropijas formula ir šāda:

$$H(1) = - \sum_{i=1}^m p_i \log_2 p_i \text{ biti/simbolu,} \quad (14)$$

$$p_i = \lim_{N \rightarrow \infty} \frac{N_i}{N} \quad (15)$$

kur  $p_i$  – simbola parādīšanās varbūtība, tas ir, atsevišķā simbola skaits pret visu teksta simbolu skaitu, ja šo simbolu skaits ir liels.

Lai aprēķinātu otrās kārtas entropiju  $H(2)$ , formulā jāievieto  $p_i$  – simbola  $i$  parādīšanās varbūtība,  $p_{ij}$  – simbola parādīšanās varbūtība, ja pirms simbola  $i$  ir cits noteikts simbols  $j$ , tas ir, kombināciju skaits no diviem burtiem  $i$  un  $j$  pret simbola  $i$  skaitu, kura parādīšanās varbūtība tiek aprēķināta.

$$H(2) = - \sum_{i=1}^m p_i \sum_{j=1}^m p_{ji} \log_2 p_{ji} \text{ biti/simbolu;} \quad (16)$$

$$p_{ji} = p(a | a_i) = \lim_{N_j \rightarrow \infty} \frac{N_{ij}}{N_j}. \quad (17)$$

Trešās kārtas entropija  $H(3)$  tiek aprēķināta pēc šādas formulas:

$$H(3) = - \sum_{i=1}^m p_i \sum_{j=1}^m p_{ji} \sum_{k=1}^m p_{k|j,i} \log_2 p_{k|j,i} \text{ biti/simbolu;} \quad (18)$$

$$p_{k|j,i} = p(a_k | a_j \cdot a_i) = \lim_{N_{ji} \rightarrow \infty} \frac{N_{k|ji}}{N_{ji}}. \quad (19)$$

Šeit  $p_i$  un  $p_{ij}$  ir tie paši lielumi, kas  $H(1)$  un  $H(2)$  gadījumā,  $p_{k|j,i}$  ir simbola  $k$  nosacītā parādīšanās varbūtība, ja pirms simbola  $k$  ir simboli  $i$  un  $j$  varbūtība, tas ir, kombinācija par trīs

simboliem pret tādas kombinācijas pa diviem simboliem skaitu, kas ir pirms simbola, kura varbūtība aprēķināta pie  $p_i$ .

Programma ir izstrādāta, lai to izmantotu neierobežotu reižu skaitu, tāpēc autore izvēlējās tekstus no katras sākotnēji izvēlētās kategorijas, teksta fragmentiem ir aptuveni vienāds simbolu skaits.

### **Iegūtie rezultāti**

- Izglītības likums (5069 simboli; [www.likumi.lv](http://www.likumi.lv))

$H(1) = 4,349815972854$ ;

$H(2) = 2,923080865874$ ;

$H(3) = 1,5184876867063$ .

- Darba likums (5055 simboli; [www.likumi.lv](http://www.likumi.lv))

$H(1) = 4,2634495552793$ ;

$H(2) = 3,0348646651092$ ;

$H(3) = 1,7572303727057$ .

- Meža likums (5075 simboli; [www.likumi.lv](http://www.likumi.lv))

$H(1) = 4,280832823845$ ;

$H(2) = 3,1072458132097$ ;

$H(3) = 1,6503462658157$ .

- Rainis "Tālas noskaņas zilā vakarā" (5042; Rainis, 1903)

$H(1) = 4,1655032631897$ ;

$H(2) = 3,08465847253$ ;

$H(3) = 1,9725815401199$ .

- A. Pumpurs "Lāčplēsis" (5052; Pumpurs, 1888)

$H(1) = 4,1902519411154$ ;

$H(2) = 3,0167031089157$ ;

$H(3) = 1,8641893821756$ .

- R. Blaumanis "Salna pavasarī" (5057; Blaumanis, 1958)

$H(1) = 4,2996599130587$ ;

$H(2) = 3,3166620129879$ ;

$H(3) = 2,1391020983167$ .

- Brāļi Kaudzītes "Mērnīeku laiki" (5014; Kaudzītes, 1879)

$H(1) = 4,3218684830753$ ;

$H(2) = 3,3380220823268$ ;

$H(3) = 2,2467898718413$ .

- Bakalaura darbā analizētais teksts no avīzes "5 Minūtes"

$H(1) = 4,4171518380824$ ;

$H(2) = 3,3706035633218$ ;

$H(3) = 1,9541757696007$ .

- "Diena" (Diena, 2008)

$H(1) = 4,3788652103022$ ;

$H(2) = 3,3894964595281$ ;

$H(3) = 2,2223302874547$ .

- LETA (LETA, 2008)

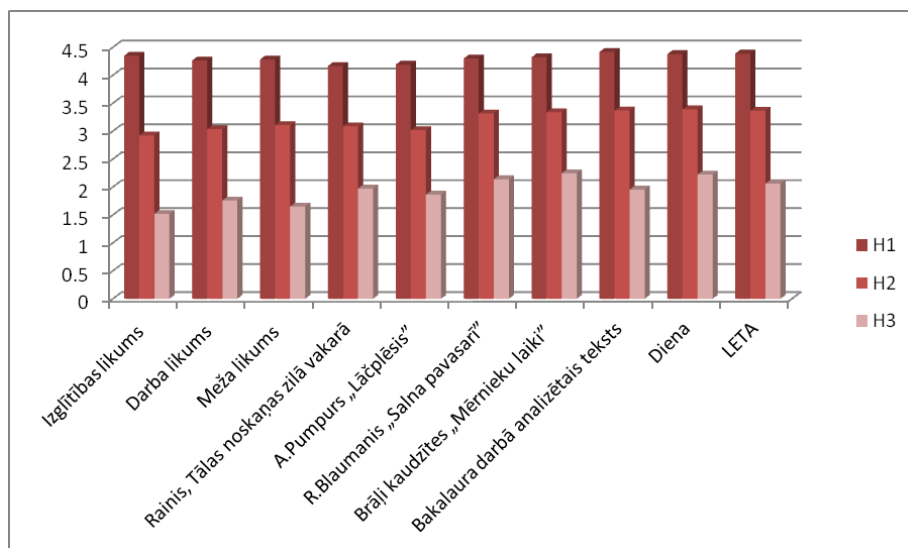
$H(1) = 4,3883573930156$ ;

$H(2) = 3,3656853313569$ ;

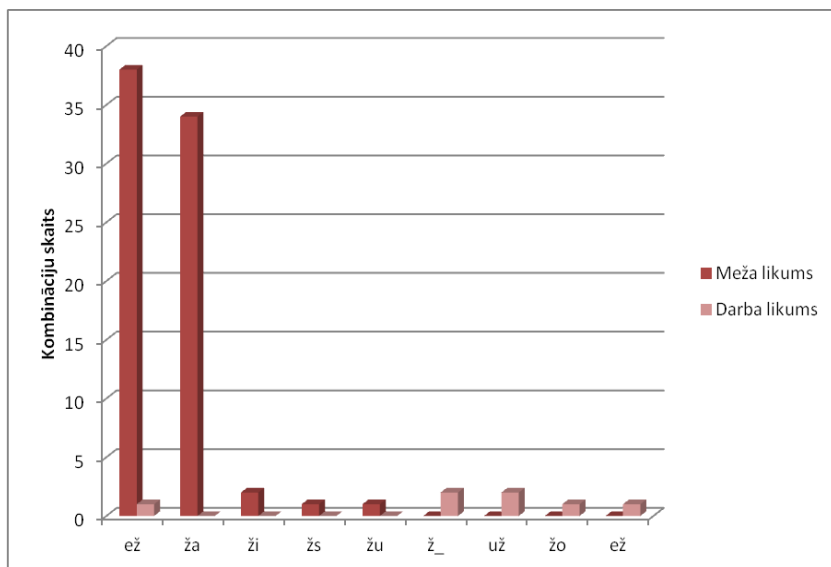
$H(3) = 2,0605372669881$ .

3.2. attēlā redzami grafiski aprēķinu rezultāti visām trīs entropijas kārtām, tas ir,  $H(1)$ ,  $H(2)$  un  $H(3)$ . Redzams, ka vienāda veida tekstiem ir līdzīga entropija.

Visaugstākā entropija ir tekstiem no preses avotiem un prozas darbiem, tas skaidrojams ar bagātīgu vārdu krājuma lietošanu. Tālāk seko likumu teksti, šajā gadījumā ļoti augsta entropija ir "Meža likumam", kas ir ļoti bieži likumā izmantotā burta "ž" dēļ – tas lietots 37 reizes, tādējādi radot kombinācijas ar "ž" burtu, kas netiek bieži izmantotas, ja teksta saturs ir cits, piemēram, Darba likumā, burts "ž" parādās tikai trīs reizes. 3.3. attēlā redzams salīdzinājums, kombinācijām ar burtu "ž" aprēķinot otrās kārtas entropiju Darba likumā un Meža likumā.

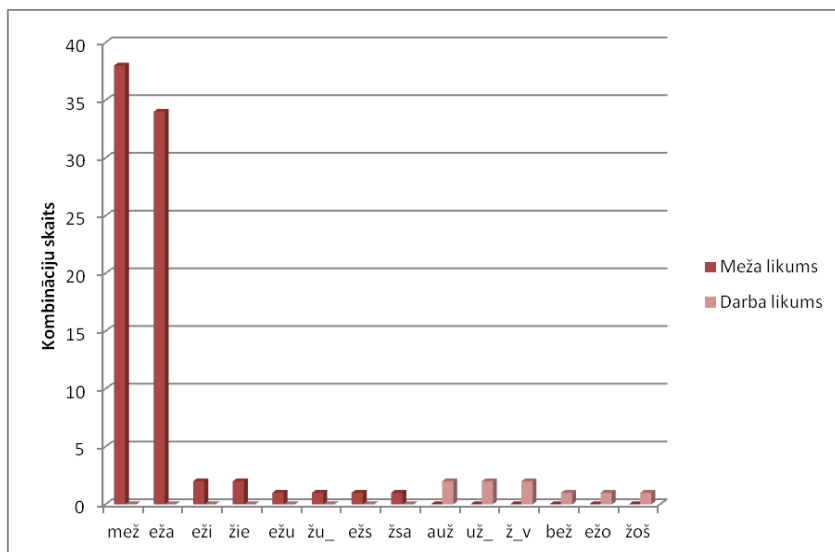


3.2. att. Aprēķinātā entropija.



3.3. att. Burtu “z” kombinācijas  $H(2)$ .

3.4. attēlā redzams salīdzinājums, kombinācijām ar burtu “z” aprēķinot trešās kārtas entropiju šiem diviem likumiem.



3.4. att. Burtu “z” kombinācijas  $H(3)$ .

3.3. attēlā, kas parāda  $H(2)$  kombinācijas, visbiežāk redzamās kombinācijas “ež” un “ža” veido saskaitāmos attiecīgi 0,0208197308569 un 0,00108960894717 Meža likumā, bet “ž\_” un “už”, kas ir kombinācijas, kas biežāk atkārtojas Darba likumā, veido saskaitāmos 0,000233844693472 un 0,00263838210761. Lai gan šeit kombinācijām “ža” un “už” saskaitāmais Meža likumā ir mazāks nekā Darba likumam, tomēr tie nav jāsalīdzina, jo kombinācijās ir svarīgi, kurā pozīcijā, pirmajā vai otrajā, atrodas burts “ž”.

Dzejas un eposa fragmenti sniedza viszemāko entropijas  $H(1)$  vērtību. Pētot iegūtās kombinācijas, autore salīdzināja Raiņa dzeju un R. Blaumaņa noveli “Salna pavasarī”. Dzejā kombināciju skaits burtiem gan pa divi, gan pa trīs bija lielāks, savukārt prozas darbā kombinācijas atkārtojās daudz biežāk un dažādāk, kas gala rezultātā radīja vislielāko entropiju.

Vidējā iegūtā entropija pa žanriem redzama 3.6. attēlā. Pirmās kārtas entropija parāda, ka vislielākā entropija ir preses kanālam, tad seko likumi, pēc tam – proza, pēdējā – dzeja. Savukārt otrās kārtas entropija vismazākā ir likumiem, vislielākā – preseī un prozai. Otrās kārtas entropija parāda viena burta atkarību no pirms tam stāvoša burta. Jo lielāka entropija, jo lielāka ir burtu kombinācijas pa divi parādīšanās varbūtība. Trešās kārtas entropija aprēķina varbūtību summu, ja viena burta parādīšanās ir atkarīga no pirms tam stāvošiem diviem burtiem. Vismazākā trešās kārtas entropija ir likumiem, tad seko dzeja, savukārt visaugstākā entropija ir prozai. Jo mazāka entropija, jo lielāka saspiežamība jeb redundance. Jo mazāka entropija, jo vieglāk ir paredzēt, kāds burts sekos cits citam. Jo vairāk burtu tiek “atklāts”, jo vairāk samazinās entropija, jo burtu secība ir vieglāk paredzama.

Ir izveidota spēle, balstoties Šenona entropijas definīcijā, kurā ir jāmin burti, lai atminētu vārdus. Pēc tā, cik ātri tiek atminēta burtu secība, tiek noteikta entropija. 3.5. attēlā redzams autores mēģinājums izspēlēt spēli. Rezultāti būtu labāki, ja spēle būtu latviešu valodā, jo pēdējo vārdu bija grūti atminēt nezināšanas dēļ. Uzdevums ir veiksmīgi atminēt nākamo burtu, zinot tikai iepriekš atminēto. K. E. Šenons burtu minējumu skaita secību uzskata par teikuma šifrējumu un eksperimenta entropiju vērtē kā entropiju angļu valodas burtiem (Kozłowski, 2008; Mika 2008).

Diemžēl interneta vietnē <https://mathweb.ucsd.edu/~crypto/java/ENTROPY/> palicis tikai apraksts par spēli, bet spēles principam tiek izmantots entropijas kalkulators.

SINCE THE LESSONS ARE FREE IF K  
 6 5 2 20 1 2 15 9 1 1 3 3 13 1 1 1 3 1 2 1 1 2 23 16 3 1 1 4 25 1 4  
 NITTING DOESNT APPEAL TO YOU TH  
 25 1 19 1 1 1 1 1 14 11 10 4 2 2 1 7 22 2 1 1 22 2 3 6 1 3 1 1 1 3 5  
 EN YOU MIGHT WANT TO LEARN TO W  
 1 3 2 1 1 1 1 11 1 1 1 1 1 15 2 1 1 1 1 1 1 22 4 1 1 1 1 4 15 27 21  
 ATERSKI  
 3 14 16 2 20 26 1

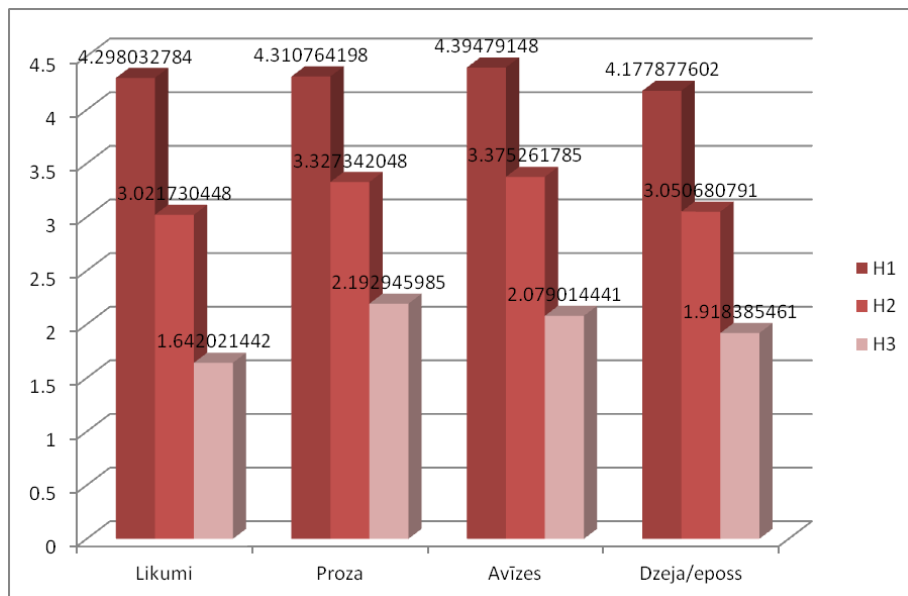
The entropy for this experiment is 3.0859928

Letters

New Quote

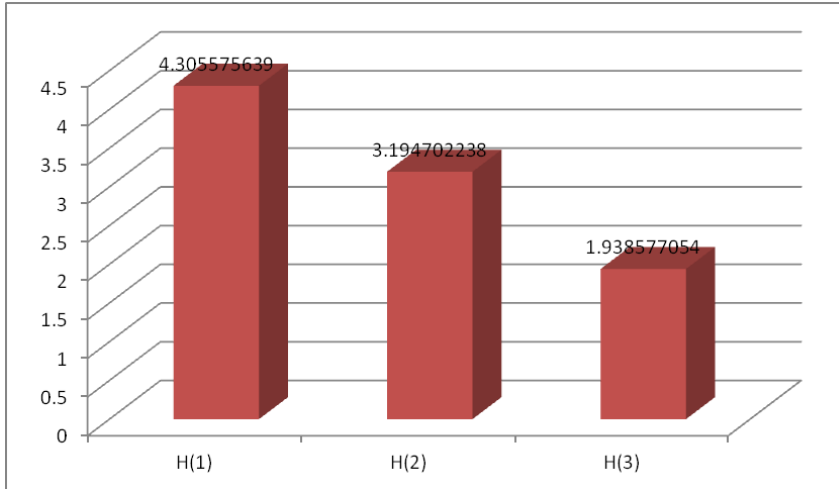
Audio:  On  Off

3.5. att. K. E. Šenona eksperiments entropijas noteikšanai



3.6. att. Vidējā entropija pa žanriem.

Vidējie entropijas rezultāti visiem izmantotajiem tekstiem, kam tika aprēķināta entropija, redzami 3.7. attēlā.



3.7. att. Vidējā entropija pa kārtām.

3.1. tabula

Entropijas salīdzinājums

$H(N)$	Latviešu valoda	Angļu valoda	Krievu valoda
$H(0)$	5,0875	4,75	5
$H(1)$	4,3056	4,07	4,05
$H(2)$	3,1947	3,36	3,52
$H(3)$	1,9386	2,77	

3.1. tabulā redzama latviešu valodas entropija, ko aprēķināja autore, K. E. Šenona iegūtais rezultāts angļu valodai un mācībspēka dotā informācija par krievu valodas entropiju. Latviešu valodas entropija ir lielāka, jo latviešu valodā ir vairāk burtu, bet tā straujāk krīt, jo kombinācijas atkārtojas retāk, līdz ar to saspiežamība ir lielāka.

Ņemot vērā iegūtajās entropijas, ir iespējams aprēķināt arī redundanci (3.2. tab.) un saspiežamības koeficientu. Jo precīzāk šie parametri ir noteikti, jo augstākas kārtas entropija tiek lietota aprēķinos. Redundance:

$$\text{pie } H(1) \rho = 1 - \frac{H(A)}{H_{\max}(A)} = 1 - \frac{H_n(A)}{H_0(A)} = 1 - \frac{4,3056}{5,0875} = 0,1537 = 15,4\%$$

$$\text{pie } H(2) \rho = 1 - \frac{H(A)}{H_{\max}(A)} = 1 - \frac{H_n(A)}{H_0(A)} = 1 - \frac{3,1947}{5,0875} = 0,3720 = 37,2\%$$

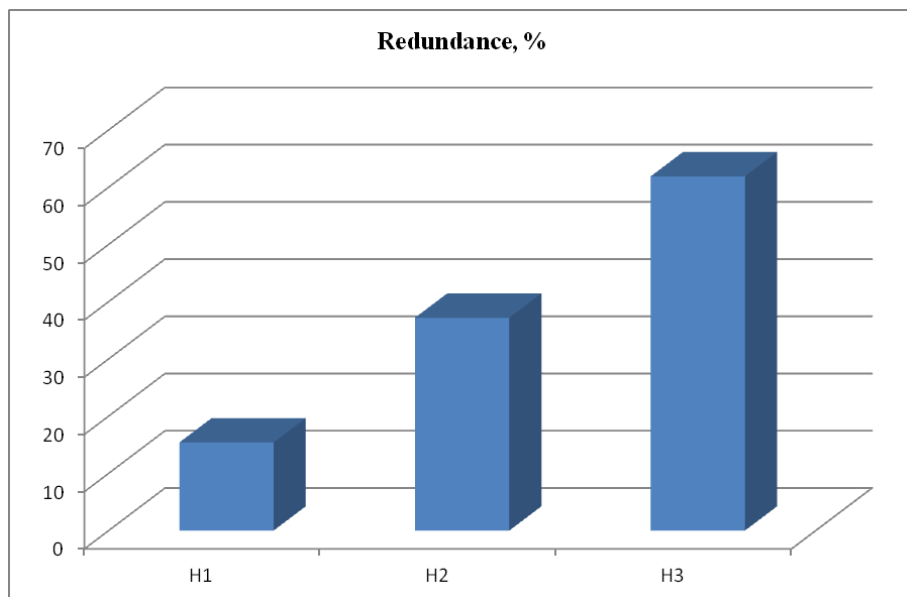
$$\text{pie } H(3) \rho = 1 - \frac{H(A)}{H_{\max}(A)} = 1 - \frac{H_n(A)}{H_0(A)} = 1 - \frac{1,9386}{5,0875} = 0,6189 = 61,9\%$$

3.2. tabula

Redundances salīdzinājums

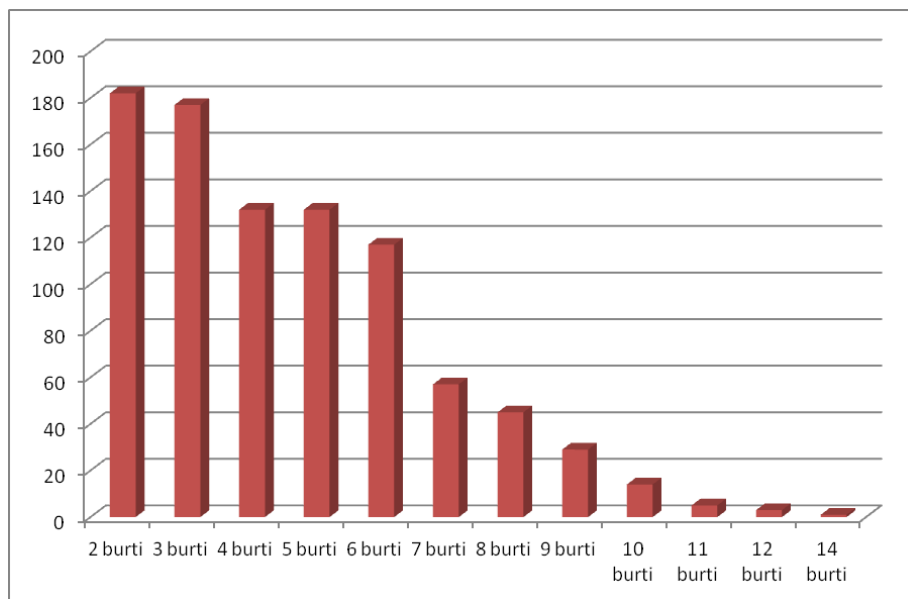
$\rho_{H(N)}$	Latviešu valoda	Angļu valoda	Krievu valoda
$\rho_{H(1)}$	15,4 %	14,3 %	19 %
$\rho_{H(2)}$	37,2 %	29,2 %	29,6 %
$\rho_{H(3)}$	61,9 %	41,7 %	

Salīdzinot latviešu valodas redundanci ar angļu un krievu valodu, redzams, ka latviešu valodai redundance ir vislielākā. Tas ir skaidrojams gan ar garumzīmēm un mīkstinājuma zīmēm, gan arī jau iepriekš pieminētajiem patskaņiem, bez kuriem latviešu valoda arī ir saprotama. Lielāku entropiju var sasniegt, ja ir augstāka dažādība, tas ir, kombinācijas atkārtojas dažādi, arī tas palielina redundanci. Ja visbiežāk lietotā kombinācija tiek izlaista, to varētu uztvert tāpat. Piemēram, vārdu izglītība Izglītības likumā aizstāt ar *izglīt*, tas ļautu atbrīvoties no vairākām kombinācijām un būtu saprotams visos locījumos.

3.8. att. Redundance pie  $H(1)$ ,  $H(2)$  un  $H(3)$ .



Lai noteiktu, kāda būtu visprecīzākā redundance, pie kāda  $H(n)$  tā jāaprēķina, bija jānosaka vārda vidējais garums. Autore izvēlējās tekstu no R. Blaumaņa noveles “Salna pavasarī” un manuāli aprēķināja vārda vidējo garumu. Tas ir 4,52, noapaļojot būtu jāaprēķina  $H(5)$ , lai aprēķinātu visprecīzāko redundanci latviešu valodai. Vārdu sadalījums, ņemot vērā to garumu, redzams 3.9. attēlā (Bajarune, 2015).



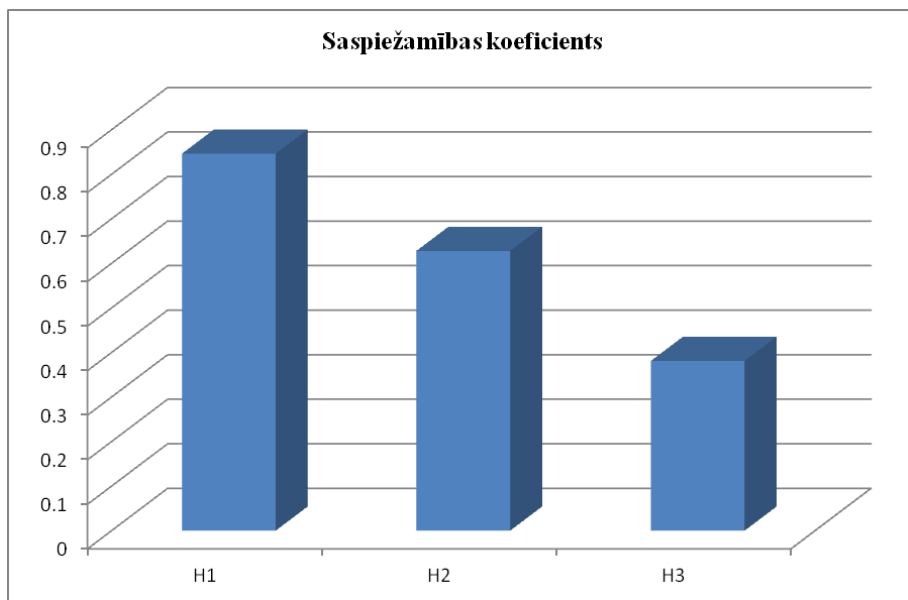
3.9. att. Vārdu sadalījums, ņemot vērā to garumu.

Saspiežamības koeficients:

$$\text{pie } H(1) \quad r = \frac{H(A)}{H_{\max}(A)} = \frac{H_n(A)}{H_0(A)} = 1 - \rho = 0,8463;$$

$$\text{pie } H(2) \quad r = \frac{H(A)}{H_{\max}(A)} = \frac{H_n(A)}{H_0(A)} = 1 - \rho = 0,628;$$

$$\text{pie } H(3) \quad r = \frac{H(A)}{H_{\max}(A)} = \frac{H_n(A)}{H_0(A)} = 1 - \rho = 0,3811.$$



3.10. att. Saspiežamības koeficients pie  $H(1)$ ,  $H(2)$  un  $H(3)$ .

### Secinājumi

1. Pirmo reizi ir kvantitatīvi analizēta latviešu valoda kā kods dažādos tekstos. Veikts dažādu tekstu entropiju salīdzinājums.
2. Latviešu valodas entropijas ir salīdzinātas ar angļu un krievu valodas entropijām. Parādīts, ka latviešu valodas tekstu redundance ir lielāka nekā angļu un krievu valodām, jo latviešu valodā ir garumzīmes un mīkstinājuma zīmes.

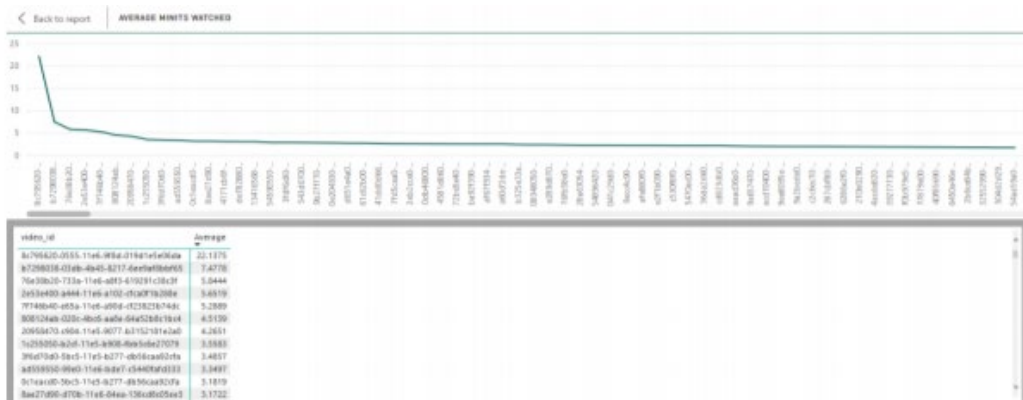
## 4. Informācija no studentu video lekciju skatīšanās darbību pierakstiem

Autore savos pētījumos ir analizējusi divu *UPV (Universitat Politècnica de València)* video lekciju ierakstu sistēmas žurnālu datus.

Pirmais logs ir 2016./2017. gada, otrais – 2018./2019. gada, abi – pirmā semestra (septembris–janvāris). Žurnālos ir divu veidu reģistri – “Darbības” un “Pēdas”. “Darbības” ir darbības ar video atskaņotāju, tāpēc tās parādās pēc tam, kad lietotājs kaut ko dara. Tās lielākoties ir saprotamas. “Darbības” un “Pēdas” ir reģistru nosaukumi, tādēļ tie jāizceļ.

“Pēdas” ir periodiska informācija, ko atskaņotājs nosūta kopā ar video daļu, kas tiek skatīta. “Pēdas” netiek ierosinātas pēc lietotāja izraisītiem notikumiem, jo lietotājs var aizvērt interneta pārlūku vai logu ar video, tad žurnāls zaudētu visu pēdējo informāciju, tāpēc aiz tā glabājas loģika sadalīt datus laika intervālos. Tie pāriet no punkta uz punktu sekundēs. *Inpoint* un *outpoint* ir sekunžu skaits videoklipā, kurā notikusi darbība. Notikumi (piemēram, *PLAY*) abi būs vienādi. “*Course\_id*” ir kurss (*UPV* sāk semestri februārī), “*Mediapackage*” ir videoklipa *ID*. Kurša *ID* ir sērijas *id*. Tātad multivides paketes (video) pieder kursam (sērijas *id*).

Pirmais, ko autore ir analizējusi, ir video skatīšanās sesija, kas ir viens gadījums, kad students skatās konkrētu videoklipu. Ir noteikta vidējā video skatīšanās sesija, kas ir vidējais veltītais laiks video lekcijai. Šis ir iesaistīšanās galvenais raksturlielums, laika ilgums, ko students tērē, skatoties videoklipu (t. i., video skatīšanās sesijas ilgums). Iesaistīšanās (*engagement*) laiks ir standarta metrika, ko izmanto bezmaksas video nodrošinātāji piemēram, *YouTube*, un uzņēmumu pakalpojumu sniedzēji, piemēram, *Wistia*. Tomēr tās ierobežojums vai neprecizitāte ir tāda, ka tas nevar uztvert, vai students aktīvi pievērš uzmanību video vai vienkārši atskaņo to fonā, vienlaikus veicot citus uzdevumus (*Guo*, 2014).



4.1. att. Video skatīšanās laiki. Uz  $y$  ass attēloti vidējie skatīšanās laiki minūtēs, uz  $x$  ass – video lekciju  $id$ . Tabulā zem grafika attēlots video lekciju  $id$  un vidējais skatīšanās laiks minūtēs, kas attiecas uz grafika sākumu.

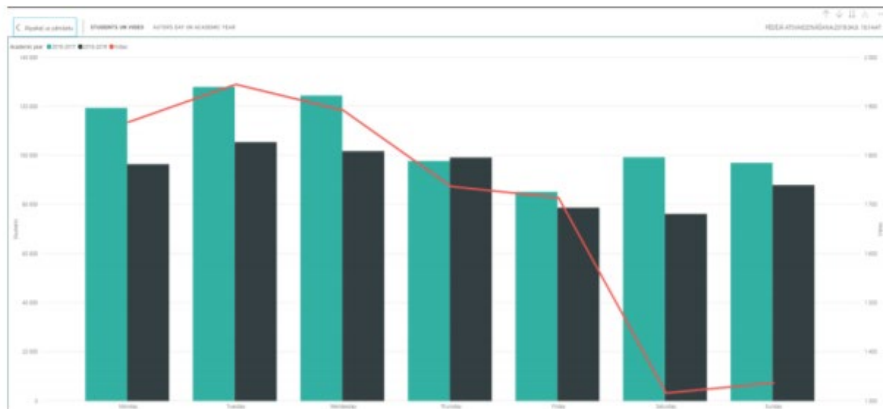
Kā redzams 4.1. attēlā, kur attēlota funkcija un vidējie video skatīšanās laiki, kas izgūti ar *Power Bi*, autore ir ieguvusi ļoti noderīgu informāciju, ko jau minējusi video lekcijas filmēšanas vadlīnijās, t. i., lekciju labāk sadalīt 5–10 minūšu garos, tēmai atbilstošos video, jo tas ir vidējais ilgums, cik aptuveni students skatās lekciju bez pārtraukuma. Lai meklētu dziļākas sakarības, par video vajadzētu iegūt vairāk informācijas.

4.2. attēls liecina, ka autore varēja atrast un attēlot arī informāciju par laiku, kad studenti skatās video lekcijas – vai darba nedēļas laikā, vai vairāk nedēļas nogalēs. Vai šis laiks ir vienāds visu semestri, vai studenti ir aktīvāki, kad eksāmenu laikā semestra beigās.

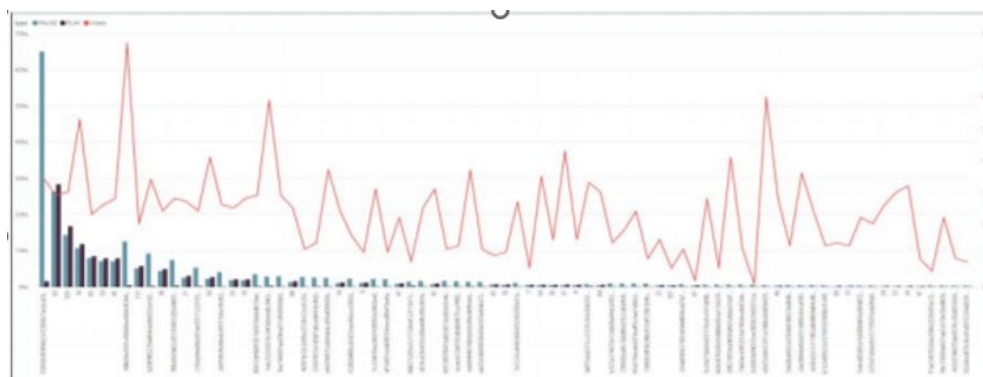
Bija interesanti uzzināt informāciju arī par video lekcijas pārtraukšanu – pastāv zināma saikne starp kursa  $ID$  un pauzes biežumu, iespējams, piezīmju izveidei (4.3. att.).

Datu analizēšanas rīku, piemēram, *Power Bi* vai *Tableau*, izmantošana ir ātrs veids, kā iegūt informāciju universitātes darbiniekiem par populārākajiem video un priekšmetiem. Varētu būt interesanti salīdzināt šo informāciju ar studentu skaitu katrā klasē. Arī videoklipiem, kas ir skatīti visbiežāk, būtu interesanti redzēt tā saturu. Vai tie ir ļoti interesanti, vai tie ir ļoti grūti, vai šajā priekšmetā ir ļoti daudz studenti.

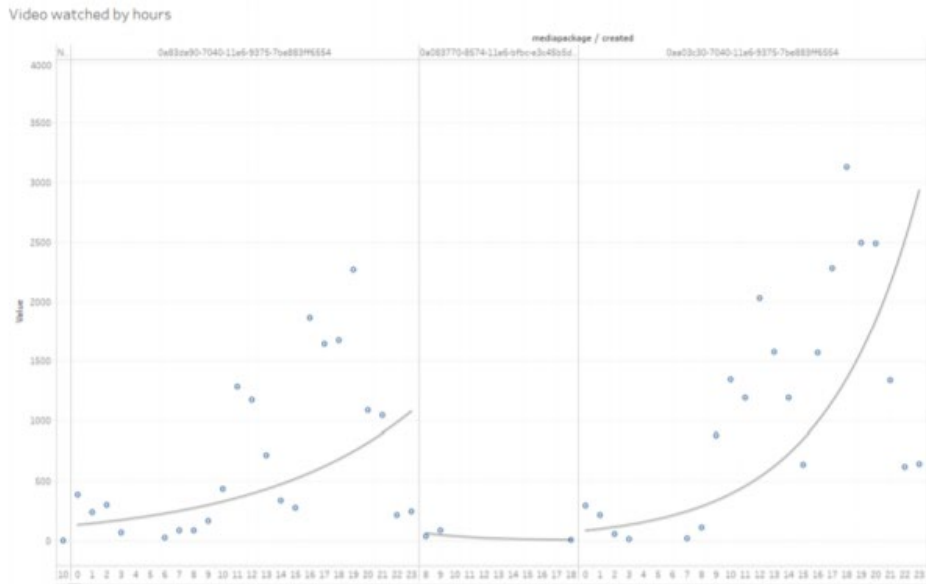
Eksponenciālās funkcijas līkne (4.4. att.) tiek izmantota, lai redzētu tendenci, kurā diennakts stundā studenti skatās video. Ir video, kas tiek skatīti visas dienas garumā – studiju un universitātes laikā, kā arī mācoties mājās, un ir tādi, kas ir skatīti tikai dažas reizes (varbūt vienu reizi) lekcijas laikā. Universitātes darbinieki var novērtēt ne tikai to, cik pieprasīti vai populāri ir videoklipi, bet arī to, cik noslogoti un izmantoti ir serveri dienas garumā dažādos laikos. Tātad – ieguvēji ir ne tikai mācītbspēki un administrācija, bet arī infrastruktūras turētāji, gan nodrošinot pakalpojumu, gan plānojot attīstību.



4.2. att. Video lekciju skatīšanās pa nedēļas dienām.

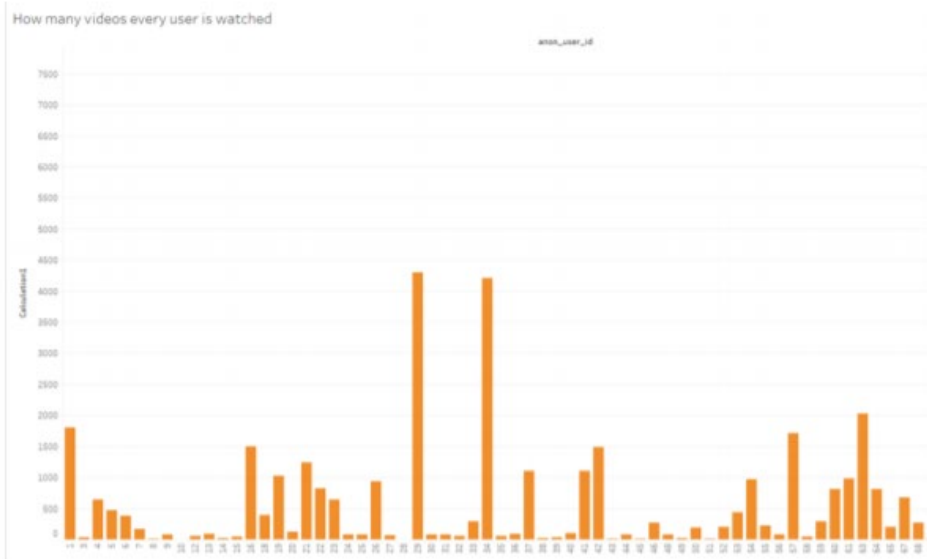


4.3. att. Video apturēšanas biežums pret video skatīšanās biežumu.



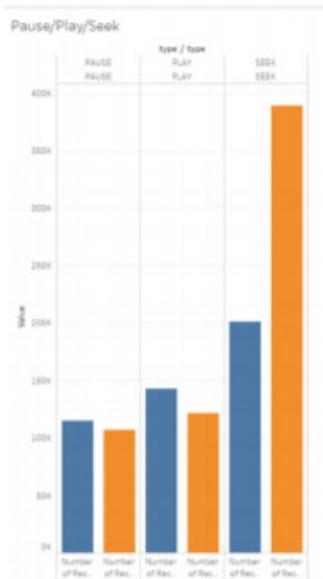
4.4. att. Video lekciju skatīšanās pa diennakts stundām.

4.5. attēlā sniegta ļoti interesanta informācija par studentu aktivitāti – cik reizes katrs students ir skatījis videoklipus.

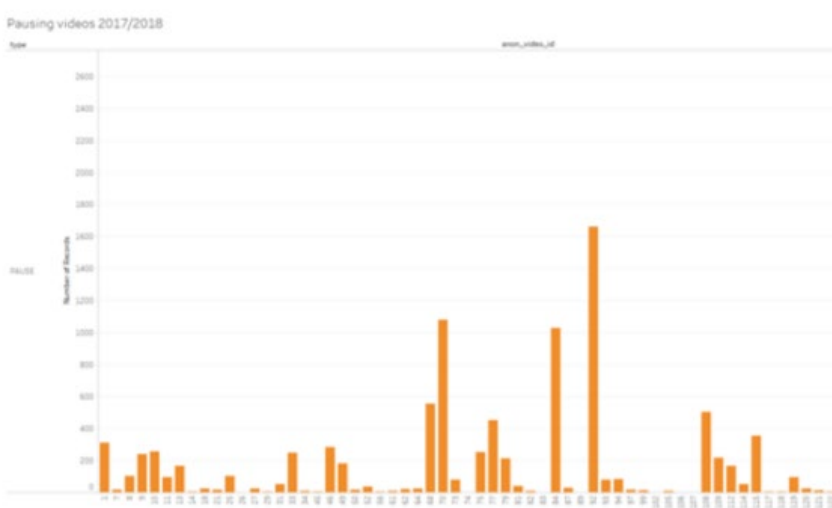


4.5. att. Katra lietotāja skatīto video lekciju skaits.

Pēc šī attēla redzams, ka ir studenti, kas izmanto iespēju skatīties video lekcijas ļoti bieži, un ir studenti, kas tās vispār neskatās. Protams, ņemot vērā šo informāciju, var secināt, cik pieprasīti ir lekciju video ieraksti, taču, lai pieņemtu konkrētus lēmumus, ir nepieciešama informācija arī par konkrēta studenta apmeklējumu un kursiem, kas tiek skatīti. Konkrētajā universitātē skatīties video lekcijas nav obligāti, tomēr var redzēt, ka studentu pieprasījums pēc tām ir milzīgs.



4.6. att. Pauzes, *Play* un meklēšanas pogu izmantošana.



4.7. att. Cik bieži apturēta katra video lekcija.



4.6. attēlā autore apkopojusi abus datu avotus par abiem semestriem, izfiltrējot darbības, kas veiktas ar videokliem – atskaņošana, apturēšana un meklēšana. Redzams, ka pēdējā semestra laikā studenti, skatoties video, aktīvāk izmantojuši meklēšanas iespēju. Autore uzskata, ka tas ir tāpēc, ka šo divu gadu laikā tehnoloģijas ir attīstījušās un meklēšana dod labāku rezultātu, kā arī tā ir laba informācijas mācībspēkiem, ja viņiem jāuzlabo videoklipa atslēgvārdi vai grāmatzīmes. E-studiju vides izstrādātāji var secināt, vai esošais meklētājs ir pietiekami attīstīts (Alksne, 2019).

### **Secinājumi**

1. Detalizēti izpētīta video lekciju skatīšanās Valensijas Politehniskajā universitātē.
2. Vislabāk video lekciju ir sadalīt 5–10 minūšu garos video, jo tas uzlabo lekcijas uztveri.
3. Visaktīvāk studenti meklē videoklipus otrajā semestrī, kas izskaidrojams ar meklēšanas rīku straujo attīstību.

## 5. Vadlīnijas video lekcijas radīšanai

Tālāk promocijas darbā izklāstītas autore apkopotās vadlīnijas video lekciju veidošanā un atbilstošo mācību materiālu filmēšanā papildus iepriekš minētajām tehnisko standartu vadlīnijām. Šīs vadlīnijas autore ir izstrādājusi, balstoties rūpīgā literatūras analīzē. Tālākā darba gaitā tiks aprēķināta video lekciju entropija un sintaktiskā informācija un analizēta vadlīniju ietekme uz šiem informatīvajiem raksturlielumiem.

1. Pirms sākt video filmēšanu, jācenšas ievērot tehniskās vadlīnijas, kas rekomendē tehniskos standartus filmēšanai, rediģēšanai un uzstādīšanai.
2. Izmantotajai tehnoloģijai nevajadzētu traucēt vai novirzīt studentu uzmanību no lekcijas.
3. Izmantotajai video reģistrācijas sistēmai jāspēj pārvaldīt arī lielākus video ražošanas apjomus.
4. Videoklipiem jābūt pielāgotiem tādām formātām, lai tie būtu skatāmi pēc iespējas vairākās platformās.
5. Jāpievērš uzmanība videoklipu veidošanas formātam.
6. Visvairāk ilgstoši uzmanību piesaista “Balss kopā ar prezentāciju” video lekcijas.
7. Video veidotājiem vajadzētu mēģināt filmēt neformālā vidē, lai mācībspēks var izveidot labu acu kontaktu ar auditoriju.
8. Video ar labi izplānotu prezentācijas gaitu un elementu maiņas tempu studijās ir visefektīvākie.
9. Izkārtotuma krāsas nedrīkst būt pārāk daudzveidīgas.
10. Īsāki videoklipi ir saistošāki nekā garāki.
11. Veidojot mācību materiālu, jāņem vērā, ka *Khan* stila mācību video ir saistošāki nekā *PowerPoint* slaidi un/vai koda ekrānuzņēmumi.
12. Filmējot video, teksts jāraksta skaidri salasāmā rokrakstā, jāizmantojot labas zīmēšanas prasmes.
13. Visiem kadriem jābūt skaidri fokusētiem un labi kadrētiem; tālummaiņa jāizmanto tikai uzmanības fokusēšanai, pretējā gadījumā no tās lietošanas būtu jāizvairās.
14. Optimālais (ieteicamais) runas ātrums prezentācijām – 160 vārdu minūtē.
15. Mācībspēkiem pirms video uzņemšanas jāvelta laiks plānošanai. Plānošanas posmam ir vislielākā ietekme uz iegūto videoklipa rezultātu.
16. Pilnas klases lekciju saturs jāsadala 5–10 minūšu garos kopsavilkuma video.
17. Visu klasē stāstīto un rādīto materiālu vēlams parādīt arī video.
18. Kurša programmas ievada un satura teksta attēlojums nedrīkst būt pārāk garš un ilgs. (Alksne, 2016).

## 6. Video lekcijas maksimālā sintaktiskā informācija

Ar sintaktisko informāciju tiek raksturots tās saņēmēja zināšanu relatīvais pieaugums. Sintaktiskā informācija ir zināšanas, ko saņemam no trokšņaina datu kanāla. 1953. gadā Šenons pētīja jautājumu par sintaktiskās informācijas kvantitatīvo noteikšanu, lai analizētu sakaru sistēmas. Mūsdienās, 21. gadsimtā, ir radušies jaunu veidu dati, piemēram, bioloģiskie dati, tīmekļa dati, topogrāfiskās kartes un medicīniskie dati utt. Analizējot jaunus datus un atklājot jaunas zināšanas par jaunajiem datiem, ir radušās jaunas metrikas arī sintaktiskajai informācijai (*Li A. un Pan Y., 2016*).

Informācija var tikt uztverta vai pārraidīta kā mainīgais, kam var būt dažādas vērtības. Tehniski saņemam informāciju kā mainīgo, apskatoties, kāda ir tās vērtība, tāpat, kā saņemam informāciju e-pastā, kad iepazīstamies ar tā saturu. Tikai mainīgā gadījumā informācija ir process, kas stāv aiz tā (*Vajapeyam, 2014*).

Ir svarīgi atcerēties, ka sintaktiskā Šenona informācija nekādā veidā neparāda sniegtās informācijas kvalitāti vai patiesumu. Šim nolūkam kalpo semantiskā informācija, kas ietver tādus jēdzienus kā tēzauris un dinamiskā entropija. Šenona entropija ir informācijas vidējais daudzums, ko ietver ziņojums. Entropija ir lielums, kas atkarīgs tikai no ziņojuma avota statistiskā rakstura, kas izteikts ar varbūtību (*Carlson, 1986*).

Lai novērtētu video informatīvo saturu, tiek izmantotas dažādas entropijas definīcijas. Entropija ir skaitlis, ko var aprēķināt izlases veidā izvēlētas video lekcijas uz labu laimi izvēlētam laika intervālam, kas uzņemts auditorijā, un tikpat ilgu laika posmu video lekcijai, kas uzņemta, ievērojot izstrādātās vadlīnijas, kā izveidot labu, studentiem vieglāk uztveramu video lekciju (*Alksne, 2016*).

Visbiežāk entropija tiek izmantota, lai atrastu konkrētas vietas videoklipā. Piemēram, entropija ir daļa no *Hue* parametra *HSV* krāsu sistēmā, lai identificētu kadra reģionus, kas raksturo noteiktu darbību noteiktā periodā. Entropija ir labs veids, kā attēlot datu kopas nevienmērīgumu vai neparedzamību, kā arī tā ir atkarīga no mērījuma konteksta (*García-Rodríguez, 2013*).

Ir izstrādāts arī algoritms, kura pamatā ir cilvēka skeleta informācijas entropija, kas tiek izmantots, lai analizētu informāciju no *RGBD* video. Tiek analizēta cilvēka skeleta leņķu informācijas entropija, kuras vērtība ir ievērojami augstāka tādās video, kuros notiek kas ārpus normālā (nevis kā parastos videoklipos). Šādi var detektēt lielu daļu nenormālas uzvedības, piemēram, kautiņus, laupīšanas vai līdzīgu haosu (*Luo, 2016*).

Eksperimentālie pētījumu rezultāti liecina, ka panikas pārņemta pūļa kustības stāvoklim ir augstāka entropija, savukārt normālam pūļa stāvoklim – zemāka. Kad izceļas masu nekārtības, gājēji bieži pārvietojas steigā. Viņu pārvietošanās ir atkarīga no vairākiem atribūtiem, piemēram, vecuma un dzimuma. Atšķiras arī indivīdu kustības ātrums un ķermeņa daļu (roku, rumpja un kāju) kustību informācija. Kustības plūsma pūļa video atspoguļo kārtības vai nekārtību stāvokli (*Xuguang Zhang, 2019*).

Līdzīgs pētījums tika veikts, apkopojot atmiņā balstīto entropiju video. Autori prognozēja atmiņas jeb neaizmirstamības rādītāju un aprēķināja attēla entropijas vērtību. Lai izveidotu kāda video kopsavilkumu, tiek izvēlēts kadrs ar maksimālu neaizmirstamības rādītāju un entropijas vērtību katrā kadrā.

Visos iepriekš minētajos gadījumos entropija meklē un izceļ atšķirīgus video kadrus. Iepriekš minētie autori ir pierādījuši, ka entropiju ir iespējams izmantot ne tikai tad, kad videoklips jau ir uzņemts, bet to ir iespējams ietekmēt, uzņemot videoklipu, ievērojot noteikumus vai vadlīnijas, kurām jau ir pierādīti rezultāti, pārraidot informāciju cilvēkiem. Iepriekš minētajos pētījumos ir pētīta arī video lekciju ietekme uz mācību rezultātiem (*Weber, 1980*).

Autore ir kvantitatīvi novērtējusi maksimālo sintaktisko informāciju, ko var sniegt noteikta garuma klātienes lekcija. Ja šī lekcija tiek filmēta, tad atbilstošā video lekcija ietver daudz mazāk informācijas uztverošās un raidošās videokameras tehnisko ierobežojumu dēļ. Tāpēc autores novērtējums klātienes lekcijai ir video lekcijas informatīvo parametru augšējā robeža.

Kā jau minēts, sintaktiskā informācija attiecas tikai uz negaidītu datu daudzumu, nevis uz to nozīmi. Tiek pieņemts, ka mācībspēks runā un rāda slaidus un demonstrējumus 10 minūtes (kā piemēru ņemot 10 minūtes). Tādējādi auditorija ir saņēmusi noteiktu daudzumu optiskās un akustiskās informācijas, izmantojot acis un ausis. Šī ir maksimāli iespējamā informācija. Ja lekcija tiek uzņemta video ar skaņu un vēlāk tiek producēta, šajā video lekcijā sniegtā informācija tiks samazināta videoreģistratora tehnisko ierobežojumu dēļ (ierobežots optiskais un akustiskais joslas platums utt.). Dabiska mācībspēka lasīta lekcija tiek uzskatīta par trokšņainu komunikācijas kanālu, kas ietver skaņas apakškanālu un gaismas apakškanālu. Katrs apakškanāls pārraida kadrus, kuriem tiek aprēķināta maksimālās Šenonas entropijas, pamatojoties uz informāciju kodējošo signālu dažādu parametru cilvēka sensoru – ausu un acu – izšķirtspēju. Kadru ātrumu nosaka cilvēka ausu un acu ierobežojošās īpašības. Tiek aprēķināts maksimālais pārsūtītais kopējais informācijas daudzums gan apakškanālos, gan visā kanālā, kā arī atbilstošā kanāla kapacitāte.

### Iegūtie rezultāti

Skaņas kanāla maksimālā entropija video lekcijai ir šāda (*Alksne, 2022*):

$$H_{s_{\max}} = \log_2 2^{N_s} = N_s. \quad (20)$$

Iepriekš aprēķināts, ka šūnu skaits kanālā  $N_s = 6583070$ , tāpēc  $H_{s_{\max}} = 6583070$  biti.

Maksimālā informācija pie video lekcijas garuma  $t = 600$  s ir šāda:  $Info_{s_{\max}} = 3,95 \cdot 10^{12}$  biti.

Beztrokšņa skaņas apakškanāla maksimālā informācijas caurlaides spēja:

$$C_{s_{\max}} = \frac{Info_{s_{\max}}}{t} = \frac{H_{s_{\max}}}{\Delta t_s}. \quad (21)$$

Ievietojot iepriekš minētās  $Info_{s_{\max}}$  un  $t$  vai  $H_{s_{\max}}$  un  $\Delta t_s$  vērtības 21. vienādojumā, tiek iegūts

$$C_{s_{\max}} = 6,58 \cdot 10^9 \text{ bits/s.}$$

Gaismas kanāla maksimālā entropija video lekcijai ir šāda:

$$H_{l_{\max}} = \log_2 2^{N_i} = N_i, \quad (22)$$

$$\text{tātad } H_{l_{\max}} = 7,72 \cdot 10^9 \text{ biti.} \quad (23)$$

Gaismas kanāla maksimālā informācija

$$Info_{l_{\max}} = \left[ \frac{t}{\Delta t} \right] \cdot H_{l_{\max}} \quad (24)$$

$$Info_{l_{\max}} = \left[ \frac{600}{5 \cdot 10^{-2}} \right] \cdot 5 \cdot 10^9 = 6,2 \cdot 10^{12} \text{ biti}$$

Gaismas kanāla maksimālā kapacitāte

$$C_{l_{\max}} = \frac{Info_{l_{\max}}}{t} = \frac{H_{l_{\max}}}{\Delta t_1} \quad (25)$$

25. vienādojumā ieliekot vērtības  $Info_{l_{\max}}$  un  $t$ , vai  $H_{l_{\max}}$  un  $\Delta t_1$ , tiek iegūts

$$C_{l_{\max}} \approx 1,54 \cdot 10^{11} \text{ biti/s.}$$

### 6.1. Kopējā maksimālā lekcijas informācija un lekcijas kā komunikācijas kanāla kapacitāte

Iegūtie skaņas un gaismas apakškanālu informācijas raksturlielumi ļauj atrast visas lekcijas maksimālo informāciju un maksimālo ietilpību, vienkārši summējot atbilstošos lielumus, jo var pieņemt, ka tie ir neatkarīgi. Tādā veidā aplūkotās video lekcijas sasniegtā maksimālā informācija ir

$$Info_{\max} = Info_{s_{\max}} + Info_{l_{\max}}, \quad (26)$$

un visa lekciju kanāla maksimālā informācijas caurlaides spēja ir

$$C_{\max} = C_{s_{\max}} + C_{l_{\max}}. \quad (27)$$

Informācijas apjomus var summēt, savukārt informācijas pārraides laiks abiem apakškanāliem ir vienāds. Ieliekot atbilstošos lielumus 26. un 27. vienādojumā, tiek iegūts

$$Info_{\max} = (3,95 \cdot 10^{12} + 9,26 \cdot 10^{13}) \text{ biti} \approx \mathbf{9,65 \cdot 10^{13} \text{ biti}}$$
 un

$$C_{\max} = (6,58 \cdot 10^9 + 1,54 \cdot 10^{11}) \text{ biti/s} \approx \mathbf{1,61 \cdot 10^{11} \text{ biti/s.}}$$

Ņemot vērā šos rezultātus, ir skaidri redzams, ka visa lekciju kanāla Šenonas (sintaktiskās) informācijas raksturlielumus gandrīz pilnībā nosaka gaismas apakškanāls, jo skaņas apakškanāla pārnēsātā informācija ir mazāka par vairāk nekā vienu lieluma kārtu. Gaismas un skaņas kadru entropiju attiecība ir vēl lielāka:

$$\frac{H_{l_{\max}}}{H_{s_{\max}}} = \frac{7,72 \cdot 10^9}{6583070} \approx 1173.$$

Tādējādi lektora balss ieguldījums sintaktiskajā informācijā ir gandrīz niecīgs. Pirmajā acu uzmetienā šis rezultāts šķiet sagaidāms, jo redze ieņem augstāko vietu cilvēka maņu hierarhijā. Lekcijā šķiet, ka nepieciešama tikai slaidu demonstrēšana. No otras puses, tas ir paradoksāls rezultāts, jo praktiski ir zināms, ka mācībspēka loma ir primāra. Šis paradoksālais rezultāts ir

lekcijas jēgas neievērošanas sekas, aprēķinot sintaktisko Šenona informāciju. Liela nozīme ir ne tikai balsij, bet arī lektora intonācijai un žestiem, paužot viņa attieksmi pret saturu. Jāņem vērā arī tas, ka, ja mācībspēks izmantotu papildu skaņas pavadījumu, piemēram, mūziku, ar lielāku skaņas frekvenču joslas platumu līdz 20 kHz [cilvēka auss maksimālais joslas platums (Nave, 2016)], tad  $H_{s \max}$ ,  $Info_{s \max}$  un  $C_{s \max}$  palielinātos par vairāk nekā vienu kārtu, sasniedzot maksimālās iespējamās vērtības, kas ir salīdzināmas ar attiecīgajiem gaismas apakškanāla parametriem. Šāda situācija ir koncertos.

Iepriekš aprēķinātās skaņas un gaismas apakškanālu informācijas caurlaides spējas var salīdzināt ar zināmajām cilvēka dzirdes un redzes informācijas caurlaides spējām. Ir noskaidrots, ka  $C_{s \max} = 6,58 \cdot 10^9$  biti/s un  $C_{l \max} = 1,54 \cdot 10^{11}$  biti/s, turpretim cilvēka dzirdes kanāla kapacitāte ir attiecīgi aptuveni  $10^4$  biti/s, cilvēka redzes kanāla kapacitāte – aptuveni  $10^7$  biti/s, kā novērtējis Temņikovs (Temnikov, 1971). Jaunākie rezultāti šiem cilvēka sensoru kanāliem ir līdzīgi – aptuveni 105 biti/s un aptuveni 107 biti/s (Markowsky, 2017). Autores aprēķinātās jaudas vērtības ir par 4–5 kārtām lielākas.

Kā var izskaidrot tik lielu atšķirību. Pirmkārt, ir aprēķinātas dabiskās lekcijas maksimālās jaudas vērtības, kas kalpo kā skaņas un gaismas apakškanālu augšējās robežas. Tas nozīmē, ka visas šūnas un visi kadri bija vienlīdz ticami. Praktiski tas tā nav, jo ausu un acu jutība ir spektrāli selektīva. Piemēram, cilvēka auss ir visjutīgākā pret skaņām frekvenču diapazonā no 1500 līdz 4000 Hz, savukārt cilvēka acs ir visjutīgākā pie zaļi dzeltenas gaismas viļņa garuma 555 nm ( $5,4 \cdot 10^{14}$  Hz) (Jehonovičs, 1984). Arī lekcijas saturs var ietekmēt kadru varbūtības. Promocijas darba aprēķinos nav ņemta vērā trokšņa klātbūtne gan skaņas, gan gaismas apakškanālos. Visbeidzot, tiek ignorēta uztvertā gaismas un skaņas informācijas pārraide nervu sistēmā un tās apstrāde smadzenēs. Zināms, ka tajās notiek milzīga informācijas saspiešana (Markowsky, 2017). Acīmredzot cilvēku optiskās un akustiskās uztveres sistēmas nespēj uztvert visu fiziski pieejamo informāciju.

Neskatoties uz to, iegūtie dabiskās lekcijas rezultāti ir pārvērtēti. Taču tos var izmantot kā augšējās robežas atbilstošajiem video lekcijas daudzumiem, jo videokameras tehnisko ierobežojumu dēļ informācijas raksturlielumi būs daudz zemāki.

Piedāvātā Šenona informācijas raksturlielumu aprēķināšanas metode ir izmantojama ne tikai to maksimālo vērtību noteikšanai, bet arī vispārējā gadījumā, ieviešot šūnu varbūtību sadalījumus visos kadrus un arī mainot šūnu izmērus. Atbilstošās varbūtības sadalījuma un šūnu izmēru variācijas ļautu izpildīt optimālas lekcijas empīriskos nosacījumus. Tādējādi varētu aprēķināt optimālas lekcijas informācijas raksturlielumus. Protams, šajā gadījumā ir nepieciešama piedāvātās pieejas matemātiskā modifikācija (Alksne, 2016).

Metode, kas balsfīta nepārtrauktas komunikācijas kanāla kapacitātes formulā (3), kas tiek piedāvāta un izmantota, lai aprēķinātu dabiskās lekcijas maksimālos Šenonas informācijas raksturlielumus, nav precīza. Tomēr promocijas darba autore uzskata, ka metode ir loģiska un veiktie tuvinājumi rezultātus būtiski nemaina. Lai pierādītu tā praktisko lietojamību, ir nepieciešami turpmāki teorētiskie un eksperimentālie pētījumi.

Piedāvātā metode ir diezgan vispārīga. To var izmantot arī jebkura pasaules objekta, piemēram, ainavas un ielas ar cilvēkiem, kosmiskiem objektiem, sintaktiskās, optiskās un akustiskās informācijas raksturlielumu aprēķināšanai.

## 7. Video lekciju raksturojums un analīze

### 7.1. Video lekciju raksturojums

Analizēšanai jeb entropijas aprēķināšanai tika izvēlētas 11 dažādas lekcijas. Tās var iedalīt pēc veida, kur lekcija ir filmēta – auditorijā, *Zoom*, vai tā ir koriģēta, vai nē.

7.1. tabula

Video kārtas numurs un video lekcijas veids

Video Nr. 1	Zandas Rubenes video lekcija	Filmēts auditorijā, montēts
Video Nr. 2	Jura Blūma video lekcija	Filmēts auditorijā, nemontēts
Video Nr. 3	Kāna stila video lekcija – Jūlija Maksimkina	Prezentācija ar balsi ( <i>Voice over presentation</i> )
Video Nr. 4	Aleksandra Dolgicera video lekcija	Prezentācija ar balsi ( <i>Voice over presentation</i> )
Video Nr. 5	Ingusa Skadiņa video lekcija	Filmēts pēc scenārija
Video Nr. 6	Anša Jurģa Stabinga video lekcija	Filmēts auditorijā, montēts
Video Nr. 7	Paulas Freimanis video treniņš	Filmēts pēc scenārija
Video Nr. 8	Fizikas eksperimenta fragments	Prezentācija ar balsi ( <i>Voice over presentation</i> )
Video Nr. 9	Intas Volodko video lekcija	Filmēts auditorijā, montēts
Video Nr. 10	Andra Ozola video lekcija	Prezentācija ar balsi ( <i>Voice over presentation</i> )
Video Nr. 11	Zandas Rubenes otrā video lekcija	Filmēts auditorijā, montēts

### 7.2. Video lekcijas kā koda analīze

Dabiskā lekcija, ko lasa docētājs, tiek traktēta kā beztrokšņa komunikācijas kanāls, kas ietver skaņas apakškanālu un gaismas apakškanālu. Katrs apakškanāls pārraida kadrus, kuru Šenona entropijas aprēķina pēc 20. formulas:

$$H = \frac{\sum_{n=1}^N p(n) \log_2 p(n)}{\log_2 N}, \quad (28)$$

kur  $p$  – varbūtību sadalījums pa punktiem (pikseļiem);

$N$  – kopējais punktu skaits sadalījumā.

Entropijas vērtība  $H$  tiek normēta, lai iegūtu relatīvu mērījumu, ko varētu salīdzināt ar video lekcijām.  $\log_2 N$  28. formulā apzīmē maksimālo entropiju.

Katrai video lekcijai tiek aprēķinātas trīs dažādas entropijas:

- video kadru entropija (video entropija);
- audio signāla intensitātes entropija (audio laika entropija);
- audio signāla spektra entropija (audio spektrālā entropija).

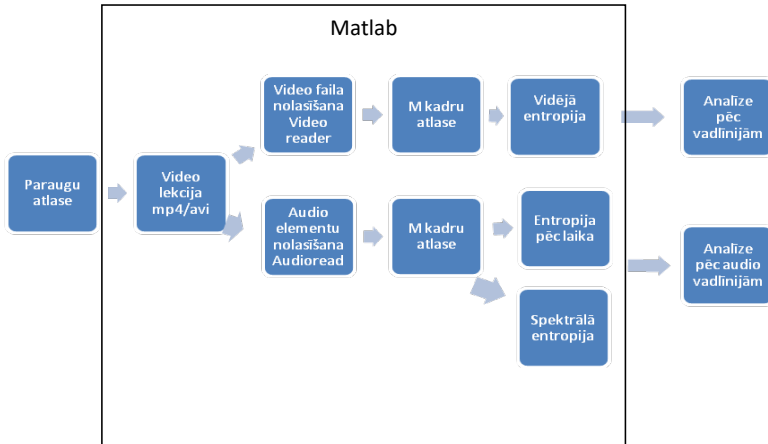
Video lekcijās ir lēni mainīgas ainas, tāpēc entropija tiek aprēķināta nevis katram kadrām, bet mazākam nejauši izvēlētu kadru skaitam. Autore pieņem, ka šie kadri uztver pietiekami daudz



informācijas, lai raksturotu visu lekciju. Audio kadri/paraugi tiek atlasīti tādā pašā veidā. Visbeidzot tiek iegūts vidējais entropijas līmenis, kas ir šāds:

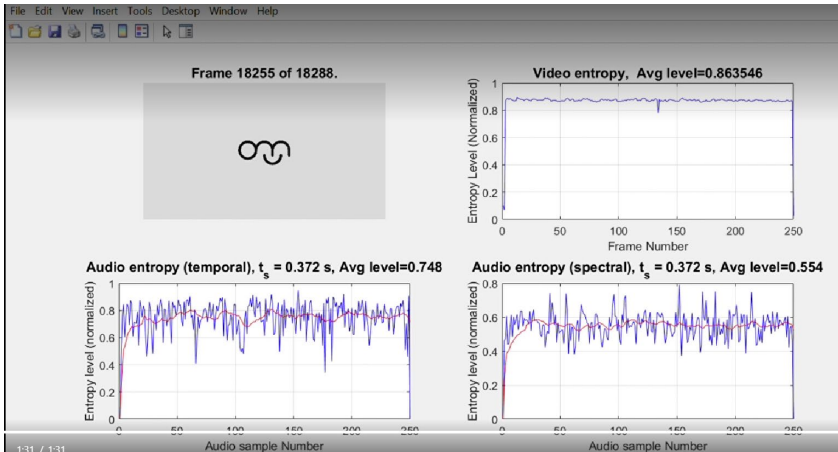
$$H = \frac{\sum_{m=1}^M H(m)}{M}, \quad (29)$$

kur  $M$  – video vai audio kadru skaits.



7.1. att. Process, pēc kura tika analizēta katra video lekcija programmā *Matlab*.

Tas, kā izskatās video lekcijas analīze programmā *Matlab*, redzams 7.2. attēlā.



7.2. att. Z. Rubenes video lekcijas analīzes entropiju rezultāti.

Visi iegūtie entropijas rezultāti un video lekciju atbilstība vadlīnijām ir parādīti 7.2. tabulā. Ņemot vērā, ka vadlīnijas ir radītas maksimāli kvalitatīvu lekciju filmēšanai, autore uzskata, ka video lekcijas kvalitāte ir atkarīga no zinātniskajā literatūrā publicēto vadlīniju skaita.

7.2. tabula

## Ievēroto vadlīniju skaits un entropijas rezultāti

	Video kārtas numurs									
	Nr. 1	Nr. 2	Nr. 3	Nr. 4	Nr. 5	Nr. 6	Nr. 7	Nr. 8	Nr. 9	Nr. 10
Prezentācija ar balsi	Nē	Jā	Jā	Jā	Nē	Nē	Nē	Nē	Jā	Nē
Labs acu kontakts	Jā	Nē	Nē	Nē	Jā	Nē	Nē	Nē	Jā	Nē
Spēcīga prezentācija, mainīgs temps	Nē	Jā	Jā	Nē	Nē	Nē	–	Nē	Jā	Nē
Nav pārāk daudz krāsu	Jā	Jā	Jā	Nē	Jā	Nē	Jā	Jā	Jā	Nē
Kāna stila video lekcija	Nē	Nē	Jā	Nē	Nē	Nē	Nē	Nē	Nē	Nē
Skaidrs rokraksts un zīmējumi	–	–	Jā	Nē	Jā	–	–	Jā	–	–
100 vārdu minūtē	99	117	94	40	76	146	–	55	104	116
Labs apgaismojums	Jā	Nē	Jā	Nē	Jā	Nē	Jā	Jā	Jā	Jā
Skaidri fokusēts un labi kadrēts	Jā	Jā	Jā	Nē	Jā	Jā	Jā	Jā	Jā	Jā
Jāizvairās no tālummaiņas	Jā	Nē	Jā	Nē	Jā	Jā	Nē	Nē	Jā	Jā
Nekustīga kamera	Nē	Nē	Jā	Jā	Jā	Jā	Jā	Jā	Jā	Jā
Skaidri saprotama runa	Jā	Nē	Jā	Nē	Jā	Jā	Jā	Jā	Jā	Jā
Mīnimāls troksnis	Jā	Nē	Jā	Nē	Jā	Jā	Jā	Jā	Jā	Nē
Standarta fons	Jā	Jā	Nē	Nē	Nē	Jā	Jā	Jā	Jā	Jā
Gludi efekti	Jā	Jā	Jā	Nē	Jā	Jā	Nē	Jā	Jā	Jā
Apstrādāts video	Jā	Nē	Jā	Nē	Jā	Nē	Jā	Jā	Jā	Nē
Fona audio balanss	–	–	–	–	–	–	Jā	–	–	–

7.2. tabulas turpinājums

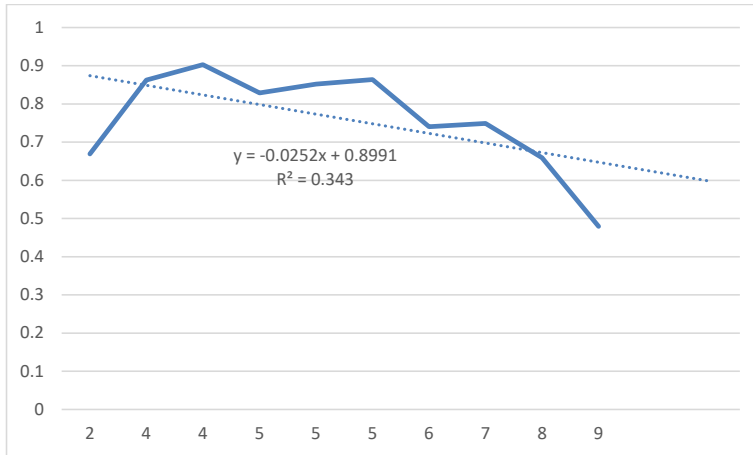
Plašs ievadmateriāls	Jā	Nē	Jā	Nē	Jā	Jā	Jā	Nē	Jā	Nē
Ieraksts studijā	Nē	Nē	Jā	Jā	Jā	Nē	Jā	Jā	Jā	Jā
Lektora balss vīrietis/sieviete	Sieviete	Vīrietis	Sieviete	Vīrietis	Vīrietis	Vīrietis	Mūzika	Sieviete	Sieviete	Vīrietis
Kopējā entropija	1,6115	1,845	1,2384	1,4601	1,5592	1,6498	1,7219	1,1993	1,4019	1,1837
“Jā” rezultāts video	5	4	9	2	7	4	5	6	8	5
“Jā” rezultāts skaņai	4	2	4	0	3	3	4	3	5	2

7.3. tabula

Vidējās entropijas rezultāti video lekcijām

	Video kārtas numurs									
	Nr. 1	Nr. 2	Nr. 3	Nr. 4	Nr. 5	Nr. 6	Nr. 7	Nr. 8	Nr. 9	Nr. 10
Vidējā video entropija, relatīvās vienībās	Video Nr. 1 0,8635 Video Nr. 11 0,8516	0,8625	0,4794	0,6691	0,7492	0,9028	0,8289	0,7403	0,6589	0,4067
Vidējā laika audio entropija, relatīvās vienībās	Video Nr. 1 0,748 Video Nr. 11 0,754	0,822	0,759	0,791	0,810	0,747	0,893	0,459	0,743	0,777
Vidējā spektrālā audio entropija, relatīvās vienībās	Video Nr. 1 0,554 Video Nr. 11 0,559	0,554	0,518	0,610	0,512	0,543	0,460	0,506	0,520	0,452

7.3. attēlā redzama vidējās video entropijas atkarība no vērā ņemto vadlīniju skaita. Var redzēt – jo vairāk vadlīniju ir aptvertas, jo zemāka ir entropija.



7.3. att. Vidējās video entropijas atkarība no ievēroto vadlīniju skaita (“Jā” skaita rezultāta).

Tuvināti var pieņemt, ka vidējās video entropijas un vadlīniju skaita korelāciju var raksturot ar lineārās regresijas vienādojumu. To nosaka pēc vismazāko kvadrātu metodes. Vienādojums un regresijas taisne redzami 7.5. attēlā.

Lai pārlicinātos par atrastās sakarības ciešumu starp vadlīniju skaitu video lekcijā un vidējo video entropiju, tiek izmantots Pīrsona korelācijas koeficients:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (30)$$

kur  $x_i$  – pazīmes variānte (šajā gadījumā izpildīto vadlīniju skaits jeb “jā” skaits);

$y_i$  – rezultatīvās pazīmes variānte (šajā gadījumā entropijas vērtības);

$\bar{x}$  – pazīmes vidējais aritmētiskais;

$\bar{y}$  – rezultatīvās pazīmes vidējais aritmētiskais;

$n$  – varianšu pāru skaits jeb paraugkopas apjoms (šajā gadījumā aplūkoto video lekciju veidu skaits  $n = 10$ ). Autore izmanto *MS Excel* funkciju, lai veiktu korelācijas koeficienta aprēķinu pēc 30. formulas. Pīrsona koeficients ir

$$r = -0,59546.$$

Korelācijas koeficients ir negatīvs, jo regresijas taisne ir dilstoša. Tā modulis atbilst vidēji ciešai lineārai korelācijai starp vidējo entropiju un vadlīniju skaitu. Ņemot vērā to, ka faktiskā sakarība 7.5. attēlā ir sarežģītāka par lineāru, rezultāts jāuzskata par labu.

To, kāda ir iegūtā Pīrsona korelācijas koeficienta ticamība  $\alpha$ , var atrast no Stjudenta sadalījuma (kas izsakāms caur gamma funkcijām) lielumam

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \quad (31)$$

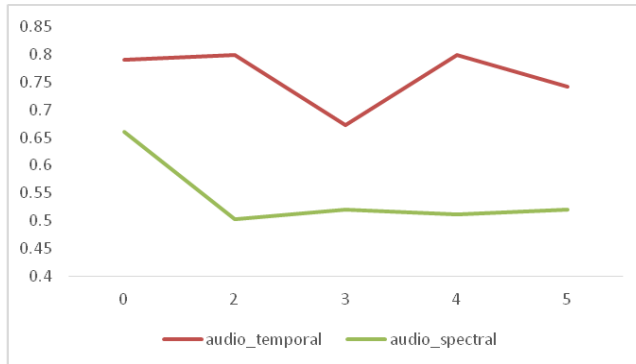
Tas tika izmantots *MS Excel* funkcijā arī  $r$  aprēķinam. *MS Excel* ir funkcija arī Pīrsona koeficienta  $r$  kļūdas varbūtības

$$p=1-\alpha \quad (32)$$

aprēķinam. To lietojot, autore ir atradusi kļūdas varbūtību  $p = 0,069323 \approx 0,0693$  un atrastā Pīrsona korelācijas koeficienta ticamības varbūtību jeb ticamību  $\alpha = 0,931$ . Tātad video lekciju entropijas negatīvo korelāciju ar vadlīniju skaitu var uzskatīt par pierādītu.

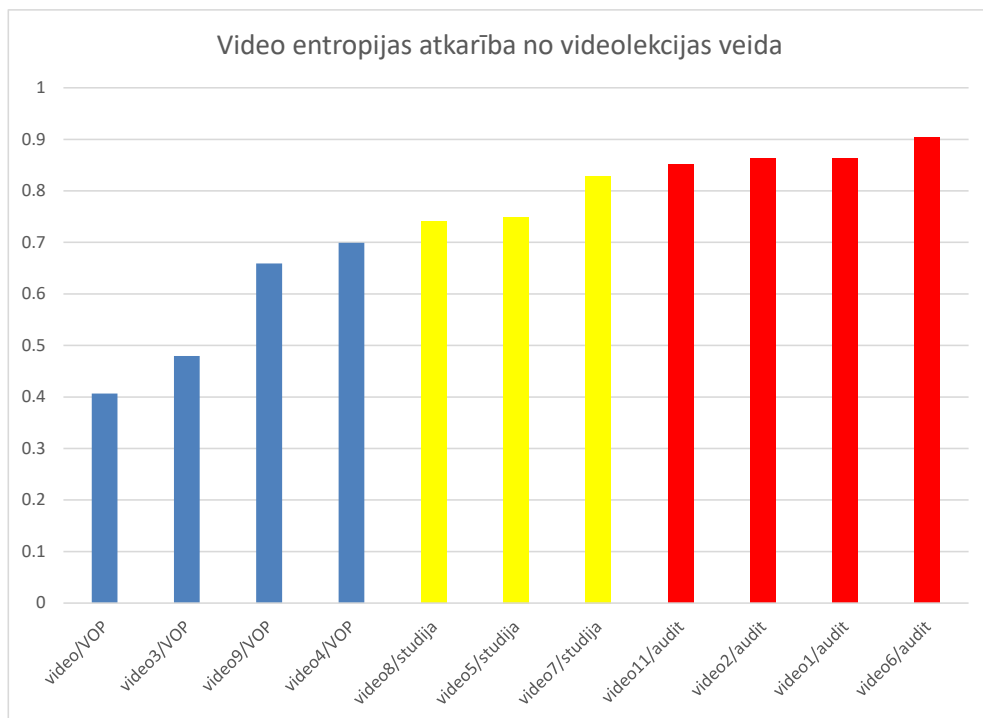
Turpinot video lekciju analīzi, jāatzīmē, ka ir arī izņēmumi, šajā gadījumā video Nr. 4, kā redzams, video lekcija saņem tikai divus “Jā”, jo, lai arī ir divi ekrāni, tomēr ir arī lielas melnas ekrāna daļas, kas nemainās visa video garumā, tāpēc šī videoklipa entropija ir zema.

Lai izdarītu dažus secinājumus par skaņas entropiju, jāaplūko 2. tabula. Tajā redzams – ja docētājas ir sievietes, vidējā laika audio entropija ir zemāka. Iespējamais cēlonis varētu būt tāds, ka lektores runā monotonākā balsī nekā vīrieši. Autore ir skaitījusi vārdus pirmajā eksperimenta minūtē, jo vienā no vadlīnijām ir minēts optimālais vārdu skaits minūtē, lai iegūtu aptuvenu runātāja aprakstu. To nevar izmantot, lai izdarītu secinājumus par visu lekciju, jo, piemēram, no 4. un 5. lekcijā redzams, ka skolotājs runā lēnāk, kad zīmē, un ātrāk, runājot kamerā. To var redzēt arī 8. video, kur audio entropija ir patiešām zema, jo sieviete visā video runā lēnā tempā.



7.4. att. Vidējo audio entropiju atkarība no ievēroto vadlīniju skaita (“Jā” skaita rezultāta).

Vidējā entropijā pret “Jā” skaitu (audio atbilstības vadlīnijas) var redzēt, ka nav būtiskas korelācijas starp atbilstību audio vadlīnijām un audio spektrālo un audio temporālo entropiju (7.4. att.), taču katras video lekcijas spektrālā un laika entropijas analīze sniedz lielu ieguvumu, jo var izdarīt secinājumus par runātāja/skolotāja balss un runas ietekmi uz entropiju. Jo vienmērīgāka runa, jo vidējā entropija ir zemāka un runa vieglāk uztverama.



7.5. att. Video entropijas atkarība no video lekcijas veida.

7.5. attēlā redzams, ka video lekcijas entropija ir atkarīga no tās veida. Ar zemāko entropiju, kas pierāda arī vienu no vadlīnijām, ir “Prezentācija ar balsi” (*VOP* jeb *voice over presentation*). Šajā grupā ir gan *Zoom* video, gan video, kur var redzēt mācībspēku un prezentāciju. Grafika vidū (dzeltenā krāsā) ir tās video lekcijas, kas filmētas pēc scenārija, montējot kadrus, bez auditorijas. Visaugstākā entropija ir visām auditorijā filmētajām lekcijām. Pēc tā var secināt, ka studentam, skatoties video lekciju, kas ietver prezentāciju un mācībspēka runu, video būs daudz vieglāk uztverams, salīdzinot ar auditorijā filmētām video lekcijām.

Video lekciju parametrus, piemēram, runas temps, apgaismojums, kameras fokuss un tālummaiņa, mācībspēks vai video lekcijas veidotājs var mainīt, lai ietekmētu entropiju – gan ar savu uzvedību, gan ar tehniskiem parametriem.

Galvenais šī promocijas darba secinājums ir šāds: pastāv korelācija starp lekcijas veidu un entropiju. Vismazākā entropija ir Kāna stila video lekcijām ar balss prezentāciju, un pētījumos ir arī pierādīts, ka šīs lekcijas studentiem uztvert ir vieglāk (*Chen, 2015*).



## Nobeigums un secinājumi

Pēc 11 dažādu video lekciju vidējo video entropiju un audio entropiju *Matlab* aprēķina un rezultātu analīzes ir iespējams izdarīt vairākus secinājumus.

1. Balss prezentācijas un Kāna stila video lekcijām ir viszemākā video entropija, ir arī pierādīts, ka šīs lekcijas studentiem uztvert ir vieglāk (*Chen, 2015*).
2. Studentam, skatoties video lekciju, kas ietver prezentāciju un mācībspēka balsi, video būs daudz vieglāk uztverams, salīdzinot ar auditorijā filmētām video lekcijām.
3. Tālummaiņas video, mainīgs kameras fokuss un vienmērīgi efekti maina video entropiju.
4. Montētām lekcijām, kas uzņemtas studijā, ir zemāka video entropija.
5. Lekcijām, kas filmētas labi apgaismotā vietā, bez spilgtuma izmaiņām un pārāk daudzveidīgām krāsām video entropija ir mazāka.
6. Ja runātājs runā lēnāk – audio entropijas svārstības ir lielākas.
7. Ja runātājs runā lēnāk – vidējā audio entropija ir zemāka.
8. Video lekcijās ar sieviešu balsīm audio entropija ir zemāka.
9. Video ar ievada ekrānu ir zemāka vidējā skaņas entropija.
10. Audio entropija ir lielāka, ja runātājs ieraksta brīdī sāk runā ātrāk un skaļāk.

## Izmantotā literatūra

1. Alksne, L. (2016). How to produce video lectures to engage students and deliver the maximum amount of information. Rēzekne: Proceedings of the International Scientific Conference. doi:<http://dx.doi.org/10.17770/sie2016vol2.1424>.
2. Alksne, L., Jansone, A., & Bērzkalne, Z. (2019). Benefits from analyzing video lecture logs with leading business analytics tools. *Baltic Journal of Modern Computing*, 7 (3), 393–404. doi:10.22364/bjmc.2019.7.3.06.
3. L. Alksne, A. Ozols. Maximum Shannon information delivered in a lecture. *Latvian Journal of Physics and Technical Sciences*, 2022, No. 2, pp. 12–22. DOI:10.2478/lpts-2022-0008. SCOPUS.
4. Association for recorded sound collections. Available from Internet [http://www.arscaudio.org/committees/video\\_production\\_guidelines.html](http://www.arscaudio.org/committees/video_production_guidelines.html).
5. Bajarune, L. & Ozols A. (2015). Latvian language as a code in different communication channels. Liepaja: [www.scopus.com](http://www.scopus.com). doi:10.17770/etr2015vol3.182.
6. Blaumanis, R., 'Salna pavasarī', Egle, K. (sast.), *Kopoti raksti*, 2. sējums, Rīga, Latvijas Valsts izdevniecība, 1958, 225.–238. lpp.
7. Carlson, Bruce A. (1986). *Communication Systems. An Introduction to Signals and Noise in Electric Communication*. New York etc.: McGraw-Hill Book Company.
8. Chandler, Paul & Sweller, John. (1991). Cognitive Load Theory and the Format of Instruction. *Faculty of Education – Papers*. 8. 10.1207/s1532690xci0804\_2.
9. Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, 80, 108–121. <http://doi.org/10.1016/j.compedu.2014.08.015>.
10. Diena, Ivo Leitāns Svētdiena, 7. decembris (2008).
11. Garcez, A., & Eisenberg, Z. (n.d.). Production and analysis of video recordings in qualitative research. *Qualitative Research*, 249–260.
12. Guo, P., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. *Proceedings of the First ACM Conference on Learning at Scale Conference*, 41–50. <http://doi.org/10.1145/2556325.2566239>.
13. Jehonovičs, A. (1984). *Handbook of physics and technics*. Rīga: Zvaigzne. In Latvian.
14. Kaudzītes R. un M. romāns “Mērnīeku laiki”. Jelgava, H. Allunans, 1879. gads.
15. Kulkarni R Sanjeev. “Lecture Notes for ELE201 Introduction to Electrical Signals and Systems”. Princeton, 2002:
16. Kozłowski, L. Shannon entropy calculator. [www.shannonentropy.netmark.pl](http://www.shannonentropy.netmark.pl).
17. Latvian Academy of Sciences “Terminology Commission database AkadTerm” [Online] Available <http://termini.lza.lv/term.php?term=redundance&list=&lang=LV&h=yes> [Accessed: Jan. 20, 2015].

18. LETA, Sestdiena, 6. decembris (2008).
19. Li A. and Pan Y., "Structural Information and Dynamical Complexity of Networks," in IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3290–3339, June 2016, doi: 10.1109/TIT.2016.2555904.
20. Likumi.lv.
21. Luo, Z. B. (2016, June). Human Abnormal Behavior Detection Based on RGBD Video's Skeleton Information Entropy. Lecture Notes in Electrical Engineering. doi:[https://doi-org.resursi.rtu.lv/10.1007/978-3-662-49831-6\\_74](https://doi-org.resursi.rtu.lv/10.1007/978-3-662-49831-6_74).
22. Mackay, J. D. C., 2006. Information Theory, Inference, and learning Algorithms. Cambridge University Press.
23. Mika V. "Datu kompresija kļūst arvien aktuālāka", – R., Sakaru pasaule 2 (50) 2008.
24. Markowsky, G. (2017). Information Theory. Encyclopaedia Britannica. Available at <https://www.britannica.com/science/information-theory>.
25. Ozols A. Lekciju konspekts "Signālu pārraides teorijā".
26. Papadimitriou C., K. Karamanosa, F. K. Diakonou, V. Constantoudis, H. Papageorgiou. "Entropy analysis of natural language written texts, Contents lists available at ScienceDirect Physica A.Greece, 2010 [Available: journal homepage]: [www.elsevier.com/locate/physa](http://www.elsevier.com/locate/physa).
27. Pumpurs A. "Lāčplēsis", Rīga: B. Diriņa un biedru apgādībā, 1888. gads.
28. Rainis "Tālas noskaņas zilā vakarā", Rīga, 1903.
29. Rubene Z: X, Y, Z un A paaudzes – LU Open Minded, 2020.
30. Shannon, C.E., A Mathematical Theory of Communication. The Bell System Technical Journal, Vol. 27, 1948. pp. 379–423, 623–656.
31. Underwood J. M., "Cultsock", available <http://ponto-kom.blogspot.com/2006/11/redundancy-and-entropy.html>.
32. Vajapeyam, S. (2014). Understanding Shannon's Entropy metric for Information.
33. web.hc.lv, 2008.
34. Xuguang Zhang, X. S. (2019). Crowd panic state detection using entropy of the distribution. Elsevier.



**Linda Alksne** dzimusi 1985. gadā Rīgā. Rīgas Tehniskajā universitātē (RTU) ieguvusi bakalaura grādu elektrozinātnē (2006) un maģistra grādu telekomunikācijās (2009). Strādājusi RTU IT departamentā. Kopš 2018. gada ir Liepājas pilsētas pašvaldības administrācijas IT daļas IT pakalpojumu pārvaldības procesu vadītāja. Patlaban ir arī Liepājas Universitātes lektore. Zinātniskās intereses saistītas ar video lekciju analīzi, informācijas tehnoloģijām un viedo pilsētu attīstību.