

Anita OZOLINA

Latvijas Universitātes Matemātikas un informātikas institūts, Rīga

SENO TEKSTU DATORFONDA IZVEIDES PROGRAMMATŪRA

Mūsdienās datori tiek plaši izmantoti ļoti dažādu uzdevumu veikšanai. Tā tas ir, nodarbojoties arī ar valodnieciskiem pētījumiem, īpaši tajos gadījumos, kad apstrādājamās informācijas apjoms ir ļoti liels. Ir izveidojusies jauna zinātnes nozare – datorlingvistika, kuras uzdevums ir lingvistisko pētījumu automatizācija. Dators ir kļuvis par neaizstājamu palīgu valodas pētījumos. Lingvistisko pētījumu datorizācija ir arī kultūras un tehnikas attīstības sastāvdaļa, kas palīdz kultūras vērtību izpētē, saglabāšanā un papildināšanā. Rodas nepieciešamība veidot valodu datorfondus. Datorfondā apvienojas lingvistisko datu bāzes (vārdu krājuma reģistri, tekstu fondi, gramatikas fondi u. c.) un datora programmatūra.

Lingvistisko pētījumu automatizācija sekmīgi veicas arī Latvijā. Notiek interesanti pētījumi ar senajiem tekstiem.

Seno tekstu datorfonds, kas ir izveidots Latvijas Universitātes Matemātikas un informātikas institūtā (LU MII) un pašreiz satur 17. gs. latviski iespiests tekstu, ir nepieciešama sastāvdaļa latviešu valodas datorfondam. Šie teksti atbilst datorfonda vēsturiskajai daļai. Lingvistiskie pētījumi ar šiem tekstiem ļauj izsekot latviešu valodas attīstībai cauri gadsimtiem, un tas veicas daudz sekmīgāk, izmantojot datoru.

Kopš tika aizsākts darbs ar senajiem latviešu tekstiem, datorā ir ievadīti šādi teksti:

- E. Glikas Bībeles tulkojums (iespiests no 1685. līdz 1694. gadam), kura Ījaba grāmata bija seno tekstu datorizētas apstrādes pirmsākums. Viss teksts datorā aizņem aptuveni 6 Mb (megabaitus);
- G. Manceļa „Sprediķu grāmata“ (1654. gads). Datorā aizņem ap 2 Mb;
- pirmais Rīgā iespiestais latviešu Katehisms (1586. gads);
- pirmā latviski iespiestā baznīcas Dziesmu grāmata (1615. gads);
- Dziesmu grāmata (1685. gads);
- Dziesmu un lūgšanu grāmata (1686. gads);
- Dziesmu, sprediķu un lūgšanu grāmata (1685. gads);
- Kara tiesas likumi (1649. gads).

Tā kā visi teksti ir tā saucamajā „vecajā drukā“, tad to ievadīšanai datorā tika izveidota šo burtu atveidošanas jeb kodēšanas sistēma, lai, pārkodējot tekstu

uz gotu burtu fontu tabulu, iegūtu šo tekstu oriģinālam atbilstošu drukas variantu (skat. 1. pielik.). Līdz ar to programmas izmanto īpaši sastādītu alfabētu.

Programmas sastādītas programmēšanas valodā Turbo-Pascal, datoriem IBM PC.

Šobrīd programmatūra piedāvā šādas jau esošas, kā arī vēl izstrādājamas un papildināmas iespējas:

- vārdformu biežuma tabulu (indeksu) sastādīšana par atsevišķu teksta daļu, nodaļu, grāmatu (skat. 2. pielik.);
- vārdkopu meklēšanu biežuma (indeksa) tabulā;
- teksta apkārtnes meklēšanu no tabulas izvēlētam vārdam;
- vārdformu kārtošanu pēc pamatformām un pievienotās gramatiskās informācijas;
- tekstu konkordanču veidošanu;
- salīdzinājuma konstrukciju atpazīšanu.

Lai gan programmas veidotas, balstoties uz senajiem tekstiem, tomēr tās ir universālas un izmantojamas arī, apstrādājot mūsdienu tekstus. Nepieciešams tikai pārlūkot tekstā sastopamās rakstu zīmes, kā arī pārlūkot teksta struktūru, lai konstatētu, vai teksts dalāms pantos, rindkopās vai rindās.

Seno tekstu analīzes programmlīdzekļu pielietojumu arī citiem tekstiem raksturo to sekmīga izmantošana E. Veidenbauma prozas analīzē un Latvju dainu (dravniecības un bišu dainu) analīzē.

1. pielikums.

E. Glika Bībeles (1685. – 1694.) Jaunās Derības teksta fragments.

Vahwīla ta Alpustuļa Grahmata uhs
Titu raķstīta.

Ta I..Nodaļa.

1. Vahwīls weens
Deewa Kalps/ un
Alpustuls IE
sus Kristus/ vēž
tahs Tizzibas to
Deewa Isre
dseto/ un Utsihshanas tahs Pa
teefibas/ kas irr pee tahs Deewa

Kalpoſchanas:

[w:1. 1.Tim: 3,16. un: 6,3.]

2. Eekſch Zerribas tahs muh
schigas Oſihwoſchanas/ ko
Deewas/ kas ne marr meloht/
preekſch Pasaules Laikem irr ap
fohlijis.

[w:2. 4.Gr: Mos: 23,19. Rom: 1,2.
un: 16,25. Efē: 1,9. un: 3,9.

Kol: 1,26. 2.Tim: 1,9,10. un:
2,13. 2.Petr: 1,20.]

3. Bet ſawā Laikā ſawu

Wahrdū irr sinnamu darrījīs
zaur to Pafluddinachanu/ kas
man irr ustizzeta pehz tahs Pa-
wehleschanas Deewa muhſo Pe-
titaja.

[w:3. Ap: Darb: 20,24. un: 9,6,15.
2.Kor: 2,12. un: 7,14. un: 8,6,16.
Gal: 1,1. un: 2,3. 2.Teff: 2,4.]

4. Titum (mannam) ihſtenam
Dehlam pehz tahs Tizzibas: kas
(mums) abbeem kohpā irr/ lai irr
Schehlastiba/ Apschehloschana/
Meers no Deewa ta Tehwa/ un ta
Kunga JEsus Kristus muhſo Pe-
titaja.

[w:4. Ewes: 1,2. Kol: 1,2. 1.Tim:
1,2. 2.Tim: 1,2. 1.Petr:1,2.]

5. Tadehl eſmu es tem Kreetā
pamettis/ ka tem bij fataiſiſt/
kas wehl atlizzis bija/ un pa Pils-
fateem Wezzajus eeſelt/ ka es tem
eſmu parwehleſis.

[5. Ap: Darb: 14,23. 2.Tim: 2,2.]

6. Ja kas irr beswainigs/ mee-
nas Seemas Wihrs/ kam tizzigi
Behrni irr/ kas ne marr leekas
Buhſchanas nedſ Pahrgalwibas
deht apſuhdſeti tapt.

[w:6. 1.Tim: 3,2.]

7. Jo weenam Pihsopam pee-
nahkahs nenoſeedsigam buht/ itt
ka Deewa Namma-turretajam/
ne pattgalwigam/ ne duſmigam/
ne Wihna-Plihtneekam/ ne Reh-
jeſam/ ne negohdigas Velnicha-
nas Oſinnejam.

[w:7. 3.Gr: Mos: 10,9. Matt:
24,45. 1.Kor: 4,1. Ewes: 5,18.

1.Tim: 3,3,15. 1.Pet: 5,2.]

8. Kas labraht mahjo/ to Lab-
bumu mihle/ gaddigam/ taſnam/
ſwehtam/ ſchēhſtam.

[w:8. 1.Tim: 3,2.]

9. Kas pastahwigī turrahſ pehz
tahs Mahzibas ta peetizziga
Wahrda/ ka tas warrens effus/
tit labb zaur to weſſeligu Mahzibū
pamahziht/ un tohs/ kas pretti
runna pahrmahziht.

[w:9. 1.Tim: 1,10. un: 6,3. 2.Tim:
1,13. un: 4,3. Tit: 2,1.]

10. Jo tur irr arriðsan daudſ
Nebehdneeki/ kas weltigas Leetas
runna/ un kas to Prahtu peewill/
wifswairak tee no tahs Upgraiſi-
ſchanas.

[w:10. Ap: Darb: 15,1. 1.Tim:
1,6.]

11. Kam peenahkahs to Mutti
aibahſt: teem/ kas Nammus
wiffai pahrgreeſch/ un mahza/ kas
ne peeklahjahs negohdigas Velnis-
ſchanas deht.

[w:11. Matt: 23,23. 1.Tim: 6,5.
2.Tim: 3,6.]

12. Weens no teem paſcheem
weens Prameets irr fazzijis: tee
Kreeteri irr allashin Melkuli/ niſ-
ni Svehtri/ flinti Wehderi.

13. Schi Leežiba irr pateefā:
tapehz pahrmahzi tohs bahrgi/ ka
tee weſſeligu effus Tizzibā.

14. Un ne dohdahs us Juddi-
ſkahm Paſſakahm un Zilweko
Bausleem/ kas no tahs Pateefi-
bas nogreeschahs.

[w:14. Esa: 29,13. Matt: 15,9.
Kol: 2,22. 1.Tim: 1,4. un: 4,7.
un: 6,20.]

15. Wiffas Leetas gann irr
teem Schékhsteem schékhstas: bet
teem Apgahniteem un Netizzi-
geem neneeka irr schékhstas/ bet
tur irr apgahnihts/ tit lab wianoo
Prahts/ ka sinnama Sirds.

[w:15. Matt: 15,11. Luk:
11,39.,41. Ap: Darb: 10,15. Rom:
14,14,20. 1.Kor: 6,12. un:
10,23,25. 1.Tim: 4,3,4.]

16. Tee fakka/ ka tee Deewu
pasihst/ bet ar Darbeem aisleeds
tee wianu/ neganti buhdami/ un
neklaufigi/ un pee wiffa labba
Darba nederrigi.

[w:16. 2.Tim: 3,5. Jud: w:4.]

Ta II..Nodalla.

1. BET tu runna/ ka tas tai ih
stenai Mahzibai peeklah-
jahs.

[w:1. 1.Tim: 6,3. 2.Tim: 1,13.]

2. Ka teem Wezzaeem buhs
mohdrigeem buht/ gohdigeem/
gaddigeem/ wesseligeem Tizzibâ/
Mihlestibâ/ Pazeefchanâ.

3. Itt ta patt wezahm
Seewahm/ ka tahm buhs turre-
tees itt ka teem Swehteem/ ka
tahm ne buhs Lazzeklehm buht/
kas us dauds Wihna ne mahzahs/
kas labbu Mahzibu dsenn.

[w:3. 1.Tim: 2,9. un: 5,13. 1.Pet:
3,3.]

4. Ka tahs marr tahs jaunas

Seewas mahziht sawus Wihrus
mihleht/ sawus Behrnus miheht.

5. Gaddigahm buht/ schékh-
stahm/ Namma-turrigahm/ lab-
bahm/ saweem Wihereem paklau-
sigahm/ ka Deewa Wahrd's sai-
mohsts ne tohp.

[w:5. 1.Gr: Mos: 3,16. 1.Kor:
14,34. Ewes: 5,22. Kol: 3,18.
1.Tim: 6,1. 1.Pet: 3,1.]

6. Tohs jaunus Wihrus pa-
mahzi/ ta patt ka teem buhs gad-
digeem buht.

7. Parahdajis few paschu wif-
fâs Mallâs ka weenu Preefsch-
Sihmi/ ar labbeem Darbeem/
eeksch Mahzibas Tihribu/ Goâdu.
[w:7. 1.Tim: 4,12. 1.Pet: 5,3.]

8. To wesseligu un beswaini-
gu Wahrdu/ ka tas Prettineeks
Kaunâ tohp/ un tam ne neeka at-
leek/ ko tas no jums warr launu
rannaht.

[w:8. 1.Tim: 5,14. 1.Pet:
2,12,15. un: 3,16.]

9. Tohs Kalpus pamahzi/ ka
teem buhs saweem Kungeem pa-
dewegeem buht/ un wiffas Leetâs pa-
tizigeem buht/ un ne neeka pretti
rannaht.

[w:9. Ewes: 6,5. Kol: 3,22. 1.Tim:
6,1,2. 1.Pet: 2,18.]

10. Ne ko atraut/ bet wissu
labbu Peetizzibu parahdiht: ka
tee to Mahzibu Deewa muhso Pe-
stitaja eeksch wisseem istaisa.

[Lekzions us Seemas-Swehtkeem.]

11. ID ta Schehlastiba ta Dee-

ma muhsfo Pestitaja irr
atspihdejusi wisseem Zilwekeem.
[w:11. 1.Tim: 2,4. Tit: 3,4.]

12. Un pamahza muhs/ ka
mehs aisleedsam to Besdeewibu/
un tahs pafauligas Eekahroscha-
nas/ gaddigi un taifni un Deewa
bijigi dsihwojam schinni Pafaulē.
[w:12. Efes: 1,4. Kol: 1,22.
2.Tim: 1,9. 1.Jahn: 2,16.]

13. Gaſidami to Labbumu
tahs Zerribas/ un to Atspihdescha-
nu tahs Gohdibas ta leela Deewa/
un muhsfo Pestitaja IESus Kri-
stus.
[w:13. 1.Kor: 1,7. Wilip: 3,20.]

14. Kas few paschu par mums
dewees irr/ ka tas muhs atpirktu
no wiffas Netaishibas/ un schekih-
stithos few pattim ihpaschus Lau-
dis/ kas tschakli dsihthos uhs lab-
beam Darbeem.)
[w:14. 2.r: Mos: 19,5. Efes: 2,10
un: 5,2. Gal: 1,4. un: 2,20. Ebr:
9,14.]

15. Scho tunna/ un pamah-
zi/ un pamahzi ar wiffu Pameh-
leschanu; ne weenam buhs tew
nizzinah.
[w:15. 1.Kor: 16,11. 1.Tim: 4,12.
Ta III. Nodaka.

1. Atgahdinis tohs/ ka tee
tahm Wirsibahm un
Waldischanahm paſlau-
figi/ un us wiffu labbu Darbu
gattawi irr.
[w:1. Rom: 13,1. 1.Pet: 2,13.]

2. Ne weenu ſaimo/ ne Rahwe-

ji irr/ lehnigi/ wiffadu Laipnibu
parahda prett wiffeem Zilwekeem.
[w:2. Wilip: 4,5. 2.Tim: 2,24,25.]

3. Jo mehs arridsan zittkahrt
bijam ne prattigi/ paſrgalwigi/
allodami/ kalpodami tahm Eekah-
roschanahm un daschadahm Rah-
ribahm/ Blehdibā un Skaudibā
dsihwodami/ naibigi/ weens ohtru
eenihdedami.

[w:3. 1.Kor: 6,11. Efes: 2,1. un:
5,3,8. Kol: 3,7. 1.Pet: 4,3. Efes:
4,17,18.]

4. Bet ta Laipniba un Mihle-
stiba Deewa muhsfo Pestitaja prett
teem Zilwekeem irr atspihdejusi/
[w:4. Tit: 2,11.]

5. Ne no teem Darbeem tahs
Taifnibas/ ko mehs effam darrig-
schi/ bet pehz faru Atschehloscha-
nu irr wiſch muhs fwehtus dar-
rijs/ zaur to Masgaschanu tahs
Atsfimfhanas un Atjaunafcha-
nas ta fwehta Garra.
[w:5. Jahn: 3,3,1. Ap: Darb:
15,11. Rom: 3,20,28. un: 4,2,6.
un: 9,11. un: 11,6. Efes: 1,4.
un: 2,4,9. un: 5,26. Gal: 2,16.
2.Tim: 1,9.]

6. Ko wiſch islehjis irr pahr
mums baggatigi zaur IESu Kri-
stu muhsfo Pestitaju.
[w:6. Efek: 36,25. Joel: 2,28.
Ap: Darb: 2,33. Rom: 5,5.]

7. Ka mehs zaur wiina Scheh-
lastibu taifnoti/ Mantneeki toh-
pam tahs muhschigas Dsihwo-
fhanas pehz tahs Zerribas.

[w:7. Rom: 8,23,24.]

8. Tas irr patees Wahrds:
Un scho grībū es/ ka tem buhs pa-
stiprinaht/ ka tee/ kas eeksch Dee-
na tizzigi tappuschi eeksch lab-
beem Darbeem pastahwigi darbo-
jahs: schis irr/ kas teem Zilwe-
keem labs un derrigs irr.

9. Bet no tahm gekēigahm
Tautaschanahm un Zilts-Rak-
steem/ un Reeschchanahm un Bahr-
schchanahm pahr tahs Bauslibas
atrujees/ jo tahs irr nederrigas
un neleetas:

[w:9. 1.Tim: 1,4. un: 4,7. un:
6,20. 2.Tim: 2,23. Tit: 1,14.]

10. Weenu sweschtizigu Zil-
weku astahj/ vež weenu un ohtru
Pakuhdinashchanu.

[w:10. Matt: 18,17. Rom: 16,17.
2.Tess: 3,6. 2.Tim: 3,5. 2.Jahn:
w:10.]

11. Sinnadams/ ka tahds irr
pahrgrohsights/ un Grehkōs buh-
dams tihscha Prahtā pats pasd-
dinajahs.

12. Kad es Artemu pee tem suh-
tischu/ jeb Tikkū/ tad steidsees pee
man naht Nitopoli/ jo tur esu
es apnehmees par Seemu mist.

[w:12. Ap: Darb: 20,4. Ewes: 6,21.
Kol: 4,6. 2.Tim: 4,12.]

13. Zenu to Bauslibas-Mah-
zitaju un Apollu parwaddi tikkū-
schī/ ka teem ne neeka truhēst.

[w:13. Ap: Darb: 18,24. 2.Kor:
1,12.]

14. Un lai arriðsan tee/ kas
mums peederr / labbōs Darbōs
pastahwigi darbojahs eeksch ta/
kas waijaga irr/ ka tee ne effus ne-
augligi.

15. Tur apsweizina tem wissi/
kas ar mannim irr: apsweizina
tohs/ kas muhs mihle eeksch Tizzi-
bas: Schehlastiba lai irr ar jums
wisseem/ Almen.

(Grahmata) uhs Titu to pír-
mu eezeltu Pihskopu tahs Kreete-
ro Draudsibas/ rakstita no Niko-
poli eeksch Makedonijs.

2. pielikums.

E. Glika Bibeles (1685. – 1694.) Jaunās Derības indeksa fragments.

[Ārons S o-c.]

Āarons [Sg Nom]

Ebr 5 4 1

Āaronu [(uz) Sg Acc]

ApDarb 7 40 1

Āarona [(no) Sg Gen]

Luhš 1 5 1

Ebr 7 11 1

Ebr 9 4 1

[abi Num]

abb [F Lok Pl]

JahnPar 1 16 1

Luhš 5 38 1

Luhš 6 39 1

Matt 9 17 1

Matt 13 30 1

Matt 15 14 1

[abpusgrīcīgs A oc.]

abpus-greefigs [NomSgM]

JahnPar 2 12 1

[Ābels S o-c.]

Ābels [Nom Sg]

Ebr 11 4 1

Ebr 12 24 1

[Aba S o-c.]		Abela [(no) Gen Sg]	
Abba [Voc Sg]		Luhk 11 51 1	
	Salat 4 6 1	Matt 23 35 1	
	Mark 14 36 1		
	Reem 8 15 1		
[Abadons S o-c.]		[Abija S o-c.]	
Abbadons [Nom Sg]		Abias [Nom Sg]	Matt 1 7 1
	TahnPar 9 11 1	Abiahu [Acc Sg]	Matt 1 7 1
[abi Num]		Abijasa [Gen Sg]	Luhk 1 5 1
abbeem [Dat Pl]		[Abilene S ē-c.]	
	Eves 2 14 1	Abilenes [Gen Sg]	Luhk 3 1 1
	Eves 2 18 1		
	Wilipp 1 23 1	[Abihuds S o-c.]	
	Luhk 7 42 1	Abiuds [Nom S]	Matt 1 13 1
	Titam 1 4 1	Abiudu [Acc Sg]	Matt 1 13 1
[abas Num]		[Abjatars S o-c.]	Matt 2 26 1
abbei [Dat Pl]		Abjatara [Gen Sg]	Tahn 8 33 1
	TahnPar 22 2 1	Abrahams [Gen Sg]	Matt 12 26 1
[abejādi Adv]			Matt 1 1 1
abbejādi			Matt 1 17 1
	TahnPar 13 15 1		Matt 8 11 1
[abēji Num]			Matt 22 32 1
abēju [Gen Pl]		Abraham [Dat Sg]	Matt 3 9 1
	Tahn 19 18 1	Abrahams [Nom Sg]	Matt 1 2 1
	ApDarb 23 8 1		
[abi Num]			
abbi [Nom PlF]			
	ApDarb 8 38 1		
	ApDarb 27 41 1		
	Eves 2 16 1		
	Tahn 4 36 1		
	Luhk 1 6 1		
	Luhk 1 7 1		
	Luhk 5 7 1		

Gauta 1998 02 20

SENUJU TEKSTU KOMPIUTERIZUOTOS DUOMENU BAZES SUDARYMO PROGRAMINĖ ĮRANGA

Santrauka

Pirmieji latvių kalba išspausdinti tekstai (nuo XVII a.) saugomi ir tyrinėjami Latvijos universiteto Matematikos ir informatikos instituto (Ryga) kompiuteriuose. Tai pirmas E. Gliko atlirkas biblijos vertimas į latvių kalbą (išspausdintas 1685–1694 m.), G. Mancelio Pamokslų knyga (1654 m.), pirmasis giesmynas (1685 m.), Giesmių ir maldų knyga (1686 m.) ir kiti senieji tekstai. Sudarytos kompiuterinės programos žodžių dažnumo analizei, konkretaus žodžio pavartojoimo teksto paieškai ir pan. Visi šie latvių senųjų raštų tyrinėjimai – svarbi latvių kalbos diachroninių duomenų bazės dalis.

SOFTWARE FOR LATVIAN ANCIENT TEXT ANALYSIS

Summary

The first Latvian printed texts from 17th century are saved electronically and processed by computer in the Institute of Mathematics and Computer Science in Riga. These texts are the first Bible translation in Latvian by E. Gliks (published in 1685–1694), Sermon Book of G. Mancelis (published in 1654), Song Book, Song and Prayer Book (published in 1685 and 1686) and other ancient texts. Software (computer programs) were made for word frequency analyses, for finding of neighbourhood of any word in the text etc. Same experiments with computerised concordance were made. All these researches of Latvian ancient texts are important part of historial database of Latvian language.