

Ajend ja eesmärgid

QWERTY klaviatuur ei ole eriti hea ning seda fakti on arutatud pikalt ja laialt. See loodi trükimasina eripära arvesse võttes ning inglise keele baasil on tehtud mitmeid uusi, arvuti jaoks mõeldud klaviatuuride paigutusi nt Colemak, Dvorak jt. Eesti keelele on ühel varasemal korral tehtud optimeeritud klaviatuur (<https://www.am.ee/Eesti-pimeklaviatuur>), ent selle loomise kohta pole välja toodud põhjalikku dokumentatsiooni, seega on raske öelda, kas klaviatuur täidab püstitatud eesmärgid. Küll aga on välja toodud klaviatuuri eelised. Meie klaviatuuri üks põhilistest ajenditest on luua klaviatuur, mis säilitaks osasid QWERTY omadusi, et uuel kasutajal oleks sellega lihtsam harjuda. Näiteks on varasemal eesti keelele optimeeritud klaviatuuril tähed ja kirjavahemärgid segamini.

Põhimõtted

Osade põhimõtete jaoks saime inspiratsiooni Dvoraki klaviatuurilt, osad on sellised, mis on meie jaoks olulised.

- Ühe sõna piires peaks kirjutamine toimuma suunaga väljast sissepoole
- Järjest kirjutatud tähed peaksid asetsema erinevatel kätel
- Tihti *shortcut*ides kasutatavad tähed ZXCV peaksid jääma sinna, kus nad on
- Punkt, koma, tühik jt mitte-täheklahvid peaksid jääma samadesse kohtadesse
- Suurem osa kirjutamist peaks toimuma parema käega, sest enamus inimesi on paremakäelised ja seega on nende parem käsi ilmselt paindlikum ja tugevam
- Enamus kirjutamisest peaks toimuma keskmisel real

Töö käik

Klaviatuuri loomise plaan on järgmine:

1. Kõigepealt tuleb leida kõige kasutatavamad eestikeelsed tähed ja need ridade vahel ära jagada: keskmisele reale kõige kasutatavamad, ülemisele järgmised ja alumisele kõige vähem kasutatavad. Ainuke erand on V, et kleepimise *shortcut* samasse kohta jääks.
2. Seejärel tuleb leida iga tähe keskmine positsioon sõnades. Kui tegemist on pigem sõna algusesse jääva tähega, saab selle paigutada klaviatuuri äärde ning kui see on sõna lõppu jääv täht, siis keskele. Niimoodi saab luua klaviatuuri, millega saab väljastpoolt sisse kirjutada.
3. Seejärel saab luua erinevaid klaviatuuri variante
4. Ning neid programmi abil testida
5. Ja kõige suurema skooriga klaviatuur osutub valituks.

1. Levinumate tähtede leidmine

Kasutades <https://www.sketchengine.eu/> võtsime etTenTen21 korpusest 10 000 suvalist rida. Otsingusõnaks kasutasime regulaaravaldist * ning funktsiooni Shuffle kasutades saime suvalised read. Iga rida koosneb võtmesõnast ning 100 eelnevast ja 100 järgnevast tähemärgist.

Seejärel puhastasime materjali: eemaldasime lingid (et vältida nt W ebaproportsionaalselt sagedust), muutsime kirjavahemärgid tühikuteks. Niimoodi jäid tekstiks alles ainult väikeste tähtedega kirjutatud sõned.

Linux käsurida:

```
cat suurkorpus.txt | sed 's@<s>@@g' | sed 's@</s>@@g' | tr '\t ' | tr 'A-ZŠŽÄÖÕÜ'
'a-zšžäöõü' | tr '-' ' ' | sed -E 's/(www|http)[^ ]+ /g' | sed 's/[^a-zšžäöõü ]/g' | tr -s ' ' >
suurkorpus_2kki_nnyd_on_norm.txt
```

Failis on kokku 10 000 rida ehk 299 117 sõna ehk 2 266 865 tähemärki.

2. Tähtede jagamine ridadesse

Jagasime tähed ridadesse populaarsuse alusel: keskmisele reale kõige kasutatavamad, ülemisele järgmised ja alumisele kõige vähem kasutatavad. Ainuke erand on V, et kleepimise shortcut samasse kohta jääks.

Ridadesisese jagunemise jaoks arvutasime iga tähe keskmise positsioon sõnades. Selleks jagasime tähe positsiooni sõnas tähtede koguarvuga.

Näiteks sõnas “jalutab” jagunevad skoorid järgnevalt:

J	A	L	U	T	A	B
1/7	2/7	3/7	4/7	5/7	6/7	7/7
0,142	0,285	0,428	0,571	0,714	0,857	1

Iga tähe kohta arvutasime kõikide positsioonide järgi aritmeetilise keskmise.

Keskmise rea puhul paigutasime tähed tähe keskmise positsiooni alusel. Need tähed, mis on tõenäolisemalt sõna alguses, paigutasime äärtesse ning need, mis on tõenäolisemalt sõna lõpus, paigutasime keskele.

2.5 Ülemise ja alumise rea eripära

Ülemise rea puhul kasutasime teistsugust strateegiat kui keskmise rea puhul. Nimelt on ülemises reas keskmiste tähtedeni raske ulatuda ning kõige paremad kohad jäävad vasakult ja

paremalt 3. ja 4. positsioonile. Sinna paigutasime kõige kasutatavamad tähed. 2. ja 5. positsioonile jäävad populaarsuselt keskmised ja kõige vähem populaarsed on keskel ja äärtes. Alumisel real on ZXCV samas kohas kui QWERTY-l ning ülejäänud kolm klahvi paigutasime populaarsuse alusel keskest väljapoole.

Kõige vähem kasutatavad tähed eesti keeles on Š, Ž ja Q, seega otsustasime, et samamoodi nagu eesti QWERTY klaviatuuril, ei saa Š ja Ž oma klahvi, vaid neid saab kirjutada S ja Z klahvide abil.



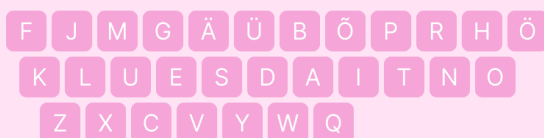
3. Klaviatuuride loomine

Lõime 10 erinevat klaviatuuri, et neid hiljem algoritmiga testida. Kasutasime eelnevalt leitud potentsiaalseid asukohti, aga võtsime vabaduse neid ka muuta ja katsetada.

ONTIA: eelnevalt loodud paigutusreeglite alusel loodud klaviatuur ehk keskmisel ja ülemisel real suvalised valikud vastavalt kahe või nelja tähe vahel



BÕPR: eelnevalt loodud paigutusreeglite alusel loodud klaviatuur, kuid võrreldes eelmise klaviatuuriga on keskmisel real tähed peegeldatult



HMP: eelnevalt loodud paigutusreeglite alusel loodud klaviatuur, kuid keskmisel real on tähed suvaliselt paigutatud (pole arvesse võetud tähtede asetsemise tõenäosust sõnas) ning ülemisel real suvalised valikud nelja tähe vahel

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

FÄRG: eelnevalt loodud paigutusreeglite alusel loodud klaviatuur, kus täishäälikud keskmisel real on paigutatud kõik ühele poolele teljest ning ülemisel real suvalised valikud nelja tähe vahel

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

KLOUNDSAITE: eelnevalt loodud paigutusreeglite alusel loodud klaviatuur, kus ülemisel real tehakse suvalised valikud nelja tähe vahel, kuid keskmisel real on tähed suvaliselt paigutatud (pole arvesse võetud tähtede asetsemise tõenäosust sõnas)

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

ONU: suvalised valikud kahe tähe vahel

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

LONT: kõik tähed ridades suvaliselt paigutatud

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

FKNTEADSI: biigrammide abil loodud paigutus

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

PÕBÜ: eelnevalt loodud paigutusreeglite alusel loodud klaviatuur, kus täishäälikud on paigutatud parema käe alla nii palju kui võimalik

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

ETNO: suvalised valikud kahe tähe vahel

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

4. Klaviatuuride testimine

Lõime programmi, mille abil klaviatuure testida.

Punktide saamise põhimõtted:

- Keskmisel real: kaks järjestikust tähte kirjutatakse suunaga väljast sissepoole +1
- Keskmisel real: kaks järjestikust tähte kirjutatakse suunaga seest väljapoole -1
- Vahetatakse ainult kätt või kätt ja rida +1
- Vahetatakse ainult rida, aga mitte kätt -1

Lisaks loeb programm vasaku ja parema käega kirjutatavate tähtede arvu.

Klaviatuuride skoorid on järgmised.

Klaviatuur	Skoor	Vasak käsi	Parem käsi
ONTIA	573	1571	1599
BÕPR	629	1463	1707
HMP	518	1398	1772
FÄRG	1030	1835	1335
KLOUNDSAITE	544	1196	1974
ONU	704	1627	1543
LONT	450	1443	1727
FKNTEADSI	733	1585	1585
PÕBÜ	590	1367	1803
ETNO	743	1450	1720

Ilmselgelt parimaks klaviatuuriks osutus FÄRG, aga kuna sellel on vasaku ja parema käe vajutuste arvud tasakaalust väljas, lõime klaviatuuri FÄRG2, mille keskmine rida on peegelpildis.

FÄRG2

1006

1309

1861

5. Klaviatuur

Seega on meie eesti keelele optimeeritud klaviatuur selline.



Võrreldes QWERTYga on FÄRGi *heatmap*ilt selgelt näha, et meie eesmärgid on täidetud. Mõlema puhul on võetud sama suvaline 100-sõnaline tekst ning klaviatuuridele kantud tähtede kasutus skaalal sinine (vähe kasutatud) - roosa (palju kasutatud).

