

პროექტის ანგარიში

I-IV კვარტალი

ლევან შოშიაშვილი

ელექტრული და ელექტრონული ინჟინერიის დეპარტამენტი
ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი
ივ. ჯავახიშვილის თბილისის სახ. უნივერსიტეტი,
ჭავჭავაძის გამზ. №3, თბილისი, საქართველო
shoshia@hotmail.com

18 მარ. 2013

სარჩევი

1 cmr, cms და cmtt ფონტები	1
2 ავტომატური დამარცვლა და გადატანა	7
2.1 ავტომატური დამარცვლა- გადატანა ვებ გვერდისათვის	8
2.2 ქართული სიტყვების ბაზის შექმნა და დამარცვლა	8
3 საძიებელი-ტექსტის ინდექსირება	9
3.1 ინდექსირების მაგალითი	9

შესავალი

წარმოდგენილია პროექტის საბოლოო (I-IV კვარტალში) შესრულებული სამუშაოს ანგარიში

1 cmr, cms და cmtt ფონტები

cmr font

Normal:

მხედრული ანბანი:

ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,

წ,ჭ,ხ,ჯ,ჰ,ტ,ყ,დ, გ,მ,წ,ფ,გ,ც,ფ,ღ,ნ,:

ასომთავრული ანბანი:

Ⴀ, Ⴁ, Ⴂ, Ⴃ, Ⴄ, Ⴅ, Ⴆ, Ⴇ, Ⴈ, Ⴉ, Ⴊ, Ⴋ, Ⴌ, Ⴍ, Ⴎ, Ⴐ, Ⴑ, Ⴒ, Ⴓ, Ⴔ, Ⴕ, Ⴖ, Ⴗ, Ⴘ, Ⴙ, Ⴚ, Ⴛ, Ⴜ, Ⴝ, Ⴞ, Ⴟ,

Ⴀ, Ⴁ, Ⴂ, Ⴃ, Ⴄ, Ⴅ, Ⴆ, Ⴇ, Ⴈ, Ⴉ, Ⴊ, Ⴋ, Ⴌ, Ⴍ, Ⴎ, Ⴐ, Ⴑ, Ⴒ, Ⴓ, Ⴔ, Ⴕ, Ⴖ, Ⴗ, Ⴘ, Ⴙ, Ⴚ, Ⴛ, Ⴜ, Ⴝ, Ⴞ, Ⴟ

ტექსტის მაგალითი:

რომელმან შექმნა სამყარო ძალითა მით ძლიერითა,

ზეგარდმით არსნი სულითა ყვნა ზეცით მონაბერითა,

ჩვენ, კაცთა, მოგვცა ქვეყანა, გვაქვს უთვალავი ფერითა,

მისგან არს ყოვლი ხელმწიფე სახითა მის მიერითა

ფორტი	CMR	CMSS	CMTT
ნორმალური	ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ, ჟ,რ,ს,ტ,უ,ფ,ქ ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ ჳ,ჴ,ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ,ჳ,ჴ	ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ, ჟ,რ,ს,ტ,უ,ფ,ქ ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ ჳ,ჴ,ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ,ჳ,ჴ	ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ, მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ, ფ,ქ ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ, ჳ,ჴ,ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ,ჳ,ჴ
მუქი	ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ, ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ ჳ,ჴ,ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ,ჳ,ჴ	ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ, ჟ,რ,ს,ტ,უ,ფ,ქ ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ ჳ,ჴ,ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ,ჳ,ჴ	ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ, მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ, ფ,ქ ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ, ჳ,ჴ,ა,ბ,გ,დ,ე,ვ,ზ,თ,ი,კ,ლ,მ,ნ,ო,პ,ჟ,რ,ს,ტ,უ,ფ,ქ,ღ,ყ,შ,ჩ,ც,ძ,წ,ჭ,ხ,ჯ,ჰ,ჳ,ჴ

2 ავტომატური დამარცვლა და გადატანა

ავტომატური დამარცვლა გადატანა მუშაობს, მაგრამ ჯერჯერობით არაა დამატებული დისტრიბუტივებში. ავტომატური დამარცვლა გადატანა თანამედროვე ვერსიებში ხორციელდება hyph-utf8 პაკეტის საშუალებით. როცა MikTeX და TeXLive განაახლებენ ამ პაკეტს ქართული ავტომატურად იქნება. თუ გსურთ ავტომატური დამარცვლა გადატანა უნდა განაახლოთ hyph-utf8 პაკეტი. ბოლო ვერსია იხ. აქ: <http://tug.org/svn/texhyphen/trunk/hyph-utf8/>.

- განაახლეთ hyph-utf8 პაკეტი სისტემურ დირექტორიაში. განაახლეთ ბაზა("Refresh FNDB" როგორც ზემოთ ან texhash კონსოლში)
- c:\Program Files\MiKTeX 2.9\tex\generic\config\ დირექტორიაში ნახეთ language.def language.dat ფაილები¹. language.def ფაილს აქვს შემდეგი სახე (ქართულის ჩამატების შემდეგ)

```
% Note: the first line of this file must match the corresponding line
% in "etex.src" and "etexdefs.lib", otherwise fallback will be used.

% This is "language.def", e-TeX's language definition file.

% It is explicitly permitted to augment this file by adding additional
% \addlanguage ... commands AFTER the first (\addlanguage {USenglish})
% which must remain as the first language added to ensure total document
% portability for pre-\language documents. The last line of the file
% MUST be left as \uselanguage {USenglish} for the sane reason.

\addlanguage {USenglish}{hyphen}{-}{2}{3} %% This MUST be the first
% % non-comment
%% line of the file

% Additional languages, patterns, exceptions and left- and right-hyphen
% minima may be added here, and an example is given for German.
\addlanguage {georgian}{hyp-ka}{-}{2}{2}
\def\l@georgian{\lang@georgian}

% The five parameters for \addlanguage are:
% language, patterns file, exceptions file, left- and right-hyphen minima
```

¹თუ არაა შექმნილი

```

%%% Next section is      E X A M P L E   O N L Y
%%% \addlanguage {German}{ghyph31}{-}{2}{2}  %%% further such lines
% %may be used;
%%% Previous section is  E X A M P L E   O N L Y

%%% No changes may be made beyond this point.

\uselanguage {USenglish} %%% This MUST be the last line of the file.

```

- language.dat ფაილში ჩაწერეთ:
georgian loadhyph-ka.tex
- განაახლეთ ბაზა("Refresh FNDB" როგორც ზემოთ ან texhash კონსოლში)
- ფორმატების განახლება initexmf --dump²

ამის შემდეგ სისტემას უნდა შეეძლოს ქართულის ავტომატურად დამარცვლა და გადატანა.

2.1 ავტომატური დამარცვლა- გადატანა ვებ გვერდისათვის

TEX სისტემისთვის გენერირებული მარცვლები გამოსადეგია ინტერნეტ ვებ გვერდების ავტომატურად დამარცვლა გადატანისათვის. ამისათვის მიღებული მარცვლები დაემატა <http://code.google.com/p/hyphenator/> პაკეტს. განახლებული კოდი იხ.<http://tex.tsu.ge/Hyphenator.js>. დამარცვლის მაგალითი კი აქ <http://tex.tsu.ge/WorkingExample.html>³ ვებ ჰიპენატორის გარატისათვის იხილეთ დოკუმენტაცია პაკეტის საიტზე. ქართულისათვის გამოიყენეთ Hyphenator.js ქართული TEX საიტიდან.

2.2 ქართული სიტყვების ბაზის შექმნა და დამარცვლა

მიუხედავად იმისა, რომ დამარცვლა გადატანის ალგორითმი დიდი ხანია რაც არსებობს TEX-ში არის ხოლმე შემთხვევები როცა ის არ მუშაობს ისე როგორც საჭიროა. ინგლისური ენის შემთხვევაშიც კი პერიოდულად ხდება მარცვალთა ნიმუშების განახლება/შესწორება. ქართულისთვისაც იგივე პრობლემა იქნება. ჩვენ მიერ შექმნილი იქნა ორი პროგრამა *indexwords* და *splitword*. პირველი მათგანი ნებისმიერი ტექსტურ ფაილიდან იღებს უნიკალურ ქართულ სიტყვებს და წერს გამოსავალ ფაილში. მეორე პროგრამა *splitword* როგორც შესავალ ფაილს იღებს ფაილს სადაც წერია ქართული სიტყვები და მარცვლავს მას ქართულად დამარცვლა გადატანის წესის შესაბამისად.

მას შემდეგ რაც გვაქვს დამარცვლილი სიტყვები შეიძლება გამოყენებულ იქნას პროგრამა *patgen* დამარცვლის ნიმუშების გენერირებისათვის. სწორედ ეს გაკეთდა მოცემული პროგრამის ფარგლებში. ინტერნეტზე არსებული ქართულენოვანი ტექსტებიდან გენერირებულ იქნა უნიკალური სიტყვების ბაზა (150 000 მდე სიტყვა) *indexwords* პროგრამის საშუალებით და დამარცვლა *splitword* საშუალებით.

²Linux/Unix სისტემებში ანალოგიური ბრძანებაა *fmtutil - all,fmtutil - sys*

³შეცვალეთ ფაილს რომა. დამარცვლა ავტომატურად შეიცვლება

3 საძიებელი-ტექსტის ინდექსირება

TEX ბაზური პაკეტი *makeidx* მუშაობს ქართული ტექსტებისათვის, მაგრამ გამო-სავალი ფაილი არის ე.წ. ტექსის შენაგან წარმოდგენაში, ანუ მაკროსების ენაზე და არა utf-8 ტექსტური ფაილი. უნიკოდ ტექსტებისთვის (და ქართულისთვისაც ცხადია) იდეალურად მუშაობს ინდექსატორი xindy, რომელსაც აქვს დამატე-ბული ქართულის მხარდაჭერა. მაგრამ ეს პაკეტი არ არსებობს Windows OS გაგრძელებულ MikTeX დისტრიბუტივში. მაგრამ აქაც შესაძლებელია ინდექ-სირების გამოყენება index პაკეტის საშუალებით. ჩვენს მიერ შექმნილი კო-დირების ფაილები შექმნილია ისე რომ ინდექსატორით მიღებული ინდექსი ავტომატურად დალაგებულია ანბანურად.

ქვემოთ მოყვანილია ინდექსირების მაგალითი.

3.1 ინდექსირების მაგალითი

ინდექსირებულია ქვემოთ მოყვანილი ანბანი და ასევე რამდენიმე სიტყვა/ტერმინი. ინდექსის გვერდი შეგიძლიათ იხილოთ დოკუმენტის ბო-ლოს.

ანბანი: აბგდევზთიკლმნოპჟრსტუფქღყმჩცძწჭხჟჰჳდგძწჯღწ:

ზოგიერთი ტერმინი: რადიუს ვექტორი, მაქსველის განტოლებები

საძიებელი

ა, 9	ფ, 9
ავტომატური დამარცვლა, 7	ქ, 9
ბ, 9	ღ, 9
გ, 9	ყ, 9
დ, 9	შ, 9
ე, 9	ჩ, 9
ვ, 9	ც, 9
ზ, 9	ძ, 9
თ, 9	წ, 9
ი, 9	ჭ, 9
კ, 9	ხ, 9
ლ, 9	ჯ, 9
მ, 9	ჰ, 9
მაქსველის განტოლებები, 9	ჳ, 9
ნ, 9	დ, 9
ო, 9	ჯ, 9
პ, 9	წ, 9
ჟ, 9	ძ, 9
რ, 9	ჭ, 9
რადიუს ვექტორი, 9	ხ, 9
ს, 9	ღ, 9
ტ, 9	ყ, 9
უ, 9	შ, 9