



ივანე ჯავახიშვილის სახელობის თბილისის
სახელმწიფო უნივერსიტეტი

პროექტი

ქართული ხელნაწერის ამომცნობი

შემსრულებლები

იოსებ გობრონიძე, ილია კაპანაძე, იოსებ ლონგურაშვილი,
ბექა ბუხრაძე

ზუსტ და საბუნებისმეტყველო მეცნიერებათა ფაკულტეტი

ხელმძღვანელი

მაია არჩუაძე

თბილისი

2017 წელი

სარჩევი

ანოტაცია	3
Annotation	3
შესავალი	4
ქართული ხელნაწერის ამომცნობი.....	5
ფოტოებისდამუშავება (Image Processing)	5
სიმბოლოების ამოცნობა (Symbol Recognition)	7
სიტყვების დამთხვევა (Words Matching)	7
შედეგების განხილვა	8
სურათებისდაჭრისშეფასება	8
სიმბოლოებისამოცნობისშეფასება.....	8
სიტყვებისდამთხვევისშეფასება.....	9
დასკვნა.....	9
გამოყენებული ლიტერატურა	10

ანოტაცია

პროექტის - „ქართული ხელნაწერის ამომცნობი“ - მთავარი მიზანია ისეთი აპლიკაციის შექმნა, რომლის დახმარებითაც შევძლებდით ქართული ხელნაწერი დოკუმენტებიდან სიმბოლოების ამოცნობასა და ელექტრონულ ფორმატში გადაყვანას. მთავარი აქცენტი კეთდება ქართულ ხელნაწერზე, თუმცა მნიშვნელოვანია ის ფაქტიც, რომ პროექტის გადაწყობა შეიძლება ისე, რომ აპლიკაციამ ამოიცნოს სხვადასხვა ენებზე შესრულებული ხელნაწერები. პროექტი შედგება რამდენიმე ეტაპისაგან, ამ ეტაპებში წარმოქმნილი პრობლემები გადავწყვიტეთ სხვადასხვა ალგორითმების გამოყენებით, თუმცა მთავარი ამოსავალი წერტილი იყო რომ „ნეირონული ქსელის“ დახმარებით (კონკრეტული ალგორითმების გამოყენებით) მანქანურმა სისტემამ დაისწავლოს ქართული ენის ასოები, შემდეგ კი შეძლოს ეს ცოდნა გამოიყენოს დოკუმენტიდან სიმბოლოების ამოცნობისთვის.

Annotation

The Project is About Handwriting Recognition System, Which Can Recognize Georgian Handwriting from Photos. Application is not limited on Georgian, it can be rewritten in such forms that it can recognize other foreign language handwriting documents. The Project has several steps to be followed, these steps can be implemented with different algorithms but the main point is that we are going to use Neural Networks as main system which can learn character recognition and then use this knowledge for text recognition.

შესავალი

ტექნოლოგიები დღეს ძალიანსწრაფად ვითარდება, ახალი ტექნოლოგიების გამოყენება ბიზნესსა თუ წარმოებაში, ან თუნდაც განათლებაში ძალიან აქტუალური საკითხია. დღესდღეობით უამრავი ადამიანი თუ კომპანია ცდილობს შექმნას სისტემები და ინსტრუმენტები, რომლებიც გაამარტივებს სხვადასხვა ინფორმაციასთან მუშაობას, ასეთი კომპანიები საკუთარ პროდუქტებს ყიდვიან საერთაშორისო ბაზარზე, კლიენტებს სთავაზობენ გარკვეულ სერვისებს, მობილურ და „დესკტოპ“ აპლიკაციებს, ე.წ. ფასიანი სერვისები საკმაოდ ეფექტურია, თუმცა საშუალო და მცირე ზომის მომხმარებელ კომპანიებს არ აქვთ საამისო ფინანსური რესურსები და გარკვეული სახის ინფორმაციის დამუშავებას როგორც ადრე კვლავას ისეთივე დიდ დროს ანდომებენ.

საქართველოში ქართულ ენაზე შესრულებული ხელნაწერების ამოცნობა ერთ-ერთი ყველაზე მნიშვნელოვანი საკითხია. ქართულ დამწერლობას საუკუნოვანი ისტორია აქვს, დამწერლობის თავისებურებებიდან გამომდინარე უკვე არსებული ხელნაწერის ამოცნობი სისტემები, რომლებიც სხვადასხვა უცხოურმა კომპანიებმა შეიმუშავეს, ქართული ენისთვის ვერ მუშაობს. უნდა მოიძებნოს გზა თუ როგორ ვასწავლოთ მანქანურ სისტემას ქართული ხელნაწერების ამოცნობა, შევინახოთ სხვადასხვა დოკუმენტები ელექტრონული სახით, ეს გაამარტივებს სხვადასხვა ინფორმაციასთან მუშაობას.

რა თქმა უნდა, ჩვენი პროექტის საბოლოო მიზანი არის ამოიცნოს ქართული ხელნაწერი, თუმცა მისი მიმდინარეობა შეიძლება დაიყოს ორ ნაწილად. პირველი ნაწილი აღწერს „დესკტოპ“ აპლიკაციის შექმნას, რომლის დახმარებითაც შესაძლებელია შევიმუშაოთ „ნეირონული ქსელი“, რომელიც შემდგომში გარკვეული ტექსტების ამოცნობას. მეორე ნაწილი კი ეხება სწორედ აპლიკაციის გამოყენებით ქართული ხელნაწერის ამოცნობი „ინტელექტუალური არსების“ შემუშავებას. ეს ერთი შეხედვით უფრო მარტივი ნაწილი სინამდვილეში საკმაოდ შრომატევადია, რადგან მოითხოვს ძალიან დიდი რაოდენობით მონაცემების შეგროვებასა და დამუშავებას, ასევე რთულ ანალიტიკურ მსჯელობას იმის შესახებ თუ როგორ შეიძლება ამ მონაცემების გამოყენება მაქსიმალურად ეფექტური შედეგის მისაღებად. ასევე ამ ნაწილში შეიძლება გაერთიანდეს მეორე აპლიკაციის შექმნაც, რომელიც არის ვებ-აპლიკაცია, იგი იყენებს ჩვენ მიერ შექმნილ და დამუშავებულ „ნეირონულ ქსელს“ და კლიენტს აძლევს საშუალებას გამოიყენოს ჩვენი აპლიკაცია. ვებ-აპლიკაცია არის ამ დიდი სამუშაოების ექსტერიერი, ყველაფერი დამუშავდა იმსითვის, რომ შექმნილიყო ეს საბოლოო აპლიკაცია და გამხდარიყო მოხმარებადი. ის არის საკმაოდ მოხერხებული, არის მხოლოდ ერთი ფუნქციონალით (ტექსტის ამოცნობა), ხოლო „დესკტოპ“ აპლიკაციის გამოყენება საკმაოდ რთულია თავისი მრავალფეროვანი ინსტრუმენტებიდან გამომდინარე და განკუთხნილია მხოლოდ შესაბამისი ცოდნის და გამოცდილების მქონე პირებისთვის. პროექტის მთავარი მიზანი არის რომ ეს ინსტრუმენტები გაგვეხადა მაქსიმალურად მოხერხებული და რაც მთავარია დაეფიქსირებინა სასურველი შედეგი, ანუ შეძლებოდა მაქსიმალური სიზუსტით ამოიცნოს ქართული ხელნაწერი.

ქართული ხელნაწერის ამომცნობი

ჩამოვაცალიბოთ პრობლემა და შევეცადოთ მისი გადაჭრის გზების განხილვას. გვაქვს ქართული ხელნაწერის დასკანირებული ვერსია ან ფოტო, ჩვენი მიზანია მაქსიმალური სიზუსტით ამოვიცნოთ სიტყვები და დოკუმენტი „გადავწეროთ“ ელექტრონულ ფორმატში. პრობლემის ასეთი ჩამოყალიბება არის ძალიან ზოგადი, სინამდვილეში აუცილებელია სხვადასხვა წინასწარ გარკვეული დაშვების გაკეთება, რომელის გათვალისწინებაც აუცილებელია პროექტის დასაწყისშივე. შემოვიღეთ რამდენიმე შეზღუდვა, რომელიც ამ მომენტისთვის აპლიკაციაში გათვალისწინებულია, თუმცა სამომავლოდ შესაძლებელია შეზღუდვების სიმრავლის შემცირება.

- ა) სატესტო სურათი უნდა იყოს საშუალო ხარისხის მაინც.
- ბ) სურათზე არ უნდა იყოს დამატებითი ნახატები.
- გ) აპლიკაცია განკუთვნილია მხოლოდ ქართულ სიმბოლოებისთვის და სასვენი ნიშნებისთვის.
- დ) ყველა აღმოჩენილი სიტყვა შემდგომ უნდა დაკორექტირდეს „ნამდვილი“ სიტყვების ბაზაში.

პრობლემის ასეთი სახით ჩამოყალიბების შემდეგ უკვე შესაძლებელია მის გადაჭრის გზებზე მსჯელობა. ამოცანა საკმაოდ კომპლექსურია, შესაბამისად იგი უნდა განვიხილოთ ერთმანეთისგან შედარებით დამოუკიდებელ სამ ეტაპად, რომელთაგან თითოეული, თავის მხრივ, შეიცავს გარკვეულ ეტაპებს :

- 1) ფოტოების დამუშავება (Image Processing)
- 2) სიმბოლოების ამოცნობა (Symbol Recognition)
- 3) სიტყვების დამთხვევა (Words Matching)

პირველ რიგში განვიხილოთ ამ სამი ეტაპის ერთობლიობაში მუშაობის შედეგი, შემდეგ კი ცალ-ცალკე თითოეული. აპლიკაციაში შემომავალი ფოტო უნდა დამუშავდეს ისე, რომ მარტივად შეიძლებოდეს კონტურების ერთმანეთისგან გამოყოფა, შესაბამისად დაიჭრას კონტურები და მივიღოთ ერთმანეთისგან გამოყოფილი სიმბოლოები. მიღებული მონაცემებით კი უკვე შესაძლებელია „ნეირონული ქსელის“ ამუშავება, რომელსაც შეეძლება მოგვაწოდოს ტექსტში არსებული სიმბოლოების ელექტრონული ვერსია. აღნიშნულ ვერსიაში აუცილებლად მოხდება შეცდომები, რომლის გამოსწორების საუკეთესო მეთოდია სიტყვების დამთხვევის ალგორითმი, ყოველი სიტყვისათვის უნდა მოვძებნოთ მასთან ყველაზე ახლოს მდგარი „რეალური“ სიტყვა და ჩავანაცვლოთ მისით.

ფოტოებისდამუშავება(Image Processing)

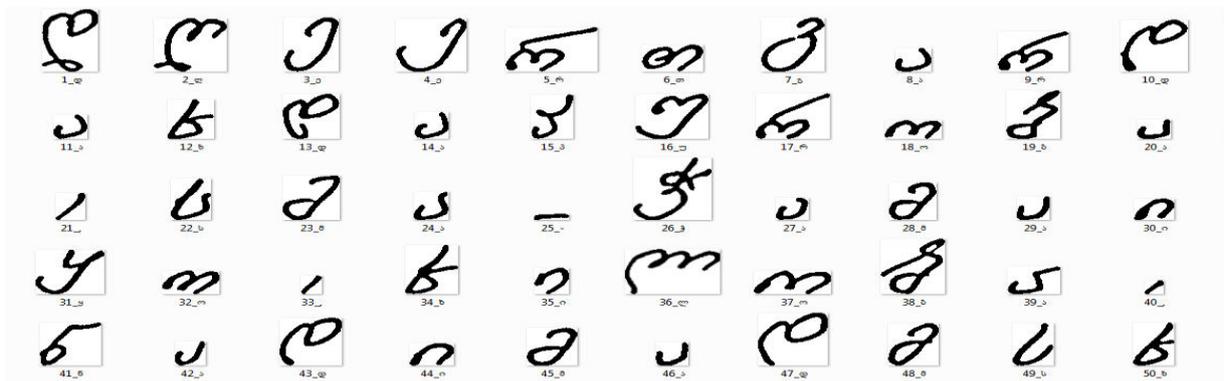
ფოტოს დამუშავებაში იგულისხმება, მისი გადაყვანა შავ-თეთრ ფორმატში. აღნიშნულის მისაღწევად საჭიროა რამდენიმე ეტაპის გავლა. მათ შორის შეიძლება გამოვყოთ პირველი ეტაპი - ფოტოს გადაყვანა რუხ ფორმატში და ე.წ. ხმაურის მოცილება. შემდეგ უკვე შესაძლებელია მისი გადაყვანა შავ-თეთრ ფორმატში.

ღღე ერთ ვარაზდა ზურობა, სმა-ჭამა იყო, ხეოობა,
ნარიძარ მხეობდა ცოშქართა მუნ რიდი შემოყრილობა.
მეფემან თავი რაშქირა, შექმნა რატრევილობა.
"ნეყარ, რა უძიძ, რა სჭირსო?" - შექმნეს ამისი ქილობა.

ღღე ერთ ვარაზდა ზურობა, სმა-ჭამა იყო, ხეოობა,
ნარიძარ მხეობდა ცოშქართა მუნ რიდი შემოყრილობა.
მეფემან თავი რაშქირა, შექმნა რატრევილობა.
"ნეყარ, რა უძიძ, რა სჭირსო?" - შექმნეს ამისი ქილობა.

ღღე ერთ ვარაზდა ზურობა, სმა-ჭამა იყო, ხეოობა,
ნარიძარ მხეობდა ცოშქართა მუნ რიდი შემოყრილობა.
მეფემან თავი რაშქირა, შექმნა რატრევილობა.
"ნეყარ, რა უძიძ, რა სჭირსო?" - შექმნეს ამისი ქილობა.

ამ ეტაპების წარმატებით გავლისათვის გამოყენებულია მზა ბიბლიოთეკა - OpenCV - რომელიც გამოირჩევა ფოტოებთან სამუშაო ალგორითმების მრავალფეროვნებით, შესაბამისად განსხვავებული ტიპის ფოტოებისათვის საჭიროა განსხვავებული მეთოდების გამოყენება. ამან შექმნა ახალი ეტაპის აუცილებლობა - ეს არის ფოტოს წინასწარი შეფასების ეტაპი, რათა დადგინდეს თუ რომელი მეთოდის გამოყენება იქნება ყველაზე ოპტიმალური. ამ ეტაპების გავლის შემდეგ კი უკვე შესაძლებელია კონტურების აღმოჩენა და სიმბოლოების გაცალკეება. ამისთვის გამოვიყენეთ ჩვენ მიერ შემუშავებული ალგორითმი, რომელიც სიღმეში ძებნის ალგორითმისა და რამდენიმე ევრისტიკის დახმარებით ახერხებს სიმბოლოების აღმოჩენასმათ გაცალკეებას.



სიმბოლოების ამოცნობა(Symbol Recognition)

სიმბოლოების ამოცნობა პროექტის მთავარი ნაწილია, რომლის დამუშავებამაც ყველაზე დიდი დრო და რესურსი მოითხოვა. სიმბოლოების ამოცნობისთვის გამოყენებულია „ნეირონული ქსელი“, კერძოდ Convolution- ე.წ. „ხვეული“ - ქსელი. არსებობს აღნიშნული ქსელის უამრავი იმპლემენტაცია, მათ შორის Java-ში დაწერილიც, თუმცა ჩვენ გამოვიყენეთ საკუთარი ნეირონული ქსელი. ჩვენ მიერ შექმნილი ქსელი არის საკმაოდ მარტივი, იცავს „ხვეული“ ქსელის სტანდარტებს და ასევე გაკეთებულია რამდენიმე ოპტიმიზაცია მეხსიერების და დროის მოგების თვალსაზრისით. რა თქმა უნდა, აღნიშნული ქსელისათვის მთავარი პრობლემური საკითხია მონაცემების შეგროვება. მონაცემების შეგროვების წინ გავითვალისწინეთ რამდენიმე მნიშვნელოვანი საკითხი, რომელებიც აუცილებლად უნდა შეესრულებინა მონაცემთა ბაზას.

- 1) სიმბოლოები უნდა იყოს საკმარისად ბევრი.
- 2) სიმბოლოები უნდა იყოს რაც შეიძლება ბევრი ადამიანის ხელნაწერიდან.
- 3) მონაცემები უნდა იყოს მაქსიმალურად მარტივი დასამუშავებელი.

ქსელის სწავლებისათვის გამოყენებულია 250-მდე სხვადასხვა ადამიანის ხელნაწერი ტექსტი, რომელთა ჯამური სიმბოლოების რაოდენობა არის 40,000-ზე მეტი. აღნიშნული მონაცემები მოვიპოვეთ უნივერსიტეტში არსებული პროექტიდან - „გადაწერე ვეფხისტყაოსანი“ - ამ პროექტით შექმნილ ხელნაწერებიდან გადავიღეთ 500-მდე სტროფი, რომლებიდანაც მოვახერხეთ 250-მდე სტროფის გამოყენება. მოცემულ მონაცემებს გააჩნია რამდენიმე უპირატესობა. პირველ რიგში აღსანიშნავია მისი მრავალფეროვნება, რაც ნეირონული ქსელისთვის ერთ-ერთი მთავარი მახასიათებელია.

სიტყვების დამთხვევა(Words Matching)

სიტყვების დამთხვევის შესამოწმებლად, პირველ რიგში, საჭიროა „რეალური სიტყვების“ ბაზა. ამისთვის შევქმენით აპლიკაცია, რომელსაც შეუძლია ქართული ციფრული დოკუმენტებიდან სიტყვების ამორჩევა. შესაბამისად შევარგოვეთ 1000-მდე ქართული დოკუმენტი, რომელიც „გავატარეთ“ აპლიკაციაში და მივიღეთ 700,000-ზე მეტი უნიკალური სიტყვა. რამდენიმე ექსპერიმენტის ჩატარების შემდეგ დავრწმუნდით, რომ სიტყვები საკმარისად ბევრია, პრაქტიკულად შეუძლებელია ისეთი სიტყვის მოძებნა, რომ არ იყოს ბაზაში. შემდეგი ეტაპი უკვე იყო ისეთი ალგორითმის დამუშავება, რომელიც შეძლებდა გადმოცემული შესაძლო არარეალური სიტყვისთვის მასთან ყველაზე „ახლოს“ მდგარი რეალური სიტყვის ან სიტყვების მოძებნას. ასეთი საქმის საუკეთესო იმპლემენტაციად ჩავთვალეთ „ლევენშტეინის მანძილის“ ალგორითმი, რომელიც თვლის

თუ რა რაოდენობის ოპერაციაა საჭირო ერთი სიტყვიდან მეორე სიტყვის მისაღებად. აღნიშნული ალგორითმი მთლიანად ჭრის ჩვენს პრობლემას, შედეგად შესაბამისი მივლეთ, თუმცა აღსანიშნავია, რომ ამ ალგორითმის გამოყენება საგრძნობლად ზრდის მთლიანი დოკუმენტის ამოცნობისთვის საჭირო დროს. შესაბამისად მომხმარებელს ვთავაზობთ, რომ ეს ფუნქციონალი ჩართოს მხოლოდ საჭიროების შემთხვევაში.

შედეგების განხილვა

პროექტის შედეგების განხილვა შესაძლებელია ორი კუთხით. ერთი მხრივ მთლიანობაში, ხოლო მეორე მხრივ, ისე როგორც ზემოთ დავსვით ამოცანა - რამდენიმე ეტაპად. პირველ რიგში განვიხილოთ ნაწილობრივი შეფასება.

სურათებისდაჭრისშეფასება - ამ ნაწილის შეფასება შეუძლებელია ზუსტი პარამეტრის გამოყენებით, თუმცა შესაძლებელია გარკვეული დაკვირვებების გაკეთება. პირველ რიგში, დავახარისხოთ სურათები რამდენიმე ნაწილად. შემდეგ, უფრო სიმარტივისთვის, განვიხილოთ ის ნაწილები რომლის დამუშავებაც აპლიკაციას „უჭირს“, შესაბამისად სხვა დანარჩენ ნაწილებს ადვილად უმკლავდება.

1) თუ ფოტოზე ჩანს რომ გადაღებულია გარკვეული ფურცელი რაიმე ფონზე, ამ შემთხვევაში აპლიკაციას შეუძლია ფოტოდან ამოჭრას მხოლოდ ფურცელი და ეს მოხდება მხოლოდ იმ შემთხვევაში თუ ფონი არ არის მრავალფეროვანი, ასევე ფურცელი უნდა ჩანდეს მთლიანად მართკუთხედი კონტურებით.

2) თუ ფოტოს ხარისხი არის ცუდი, აპლიკაცია ვერ შეუძლებს მის გამოყენებას, აქ საუბარია ფოტოებზე, რომელებიც შეიძლება იყოს „გადაღბნილი“ ან ცუდი განათებით გადაღებული ან ზედმეტად დაბალი გაფართოების და ა.შ.

3) ფურცელზე არსებული „ცუდი“ ფონი აპლიკაციისთვის არ არის პრობლემა, ასევე პრობლემას არ ქმნის უკანა ფონზე არსებული სწორი ხაზები, თუმცა მნიშვნელოვანია, რომ მელნის ფერი ოდნავ მაინც განსხვავდებოდეს ფურცლისბადეების ფერისაგან.

4) ფოტოზე არსებული ნახატები ან ცხრილები ამ მომენტისთვის წარმოადგენს პრობლემას.

სიმბოლოებისამოცნობისშეფასება - ამ შემთხვევაში უკვე ფასდება მხოლოდ „ნეირონული ქსელის“ მუშაობის პროცესი. შესაბამისად, საწყის ეტაპად ჩავთვალოთ, რომ შემომავალი ინფორმაცია უკვე დაჭრილი სიმბოლოებია და ვაფასებთ „ნეირონული ქსელის“ ამოცნობის ხარისხს. რა თქმა უნდა, პროექტის მსვლელობისას შეიქმნა უამრავი ნეირონული ქსელი განსხვავებული პარამეტრებით და გარკვეული ანალიზის და ტესტირების საფუძველზე ამოვარჩიეთ საუკეთესო ქსელი. სწორედ ეს ქსელია გამოყენებული აპლიკაციის ამოცნობის რეჟიმში. ახლა კი განვიხილოთ ამ ქსელის მიერ დაფიქსირებული შედეგები ორ განსხვავებულ სატესტო მონაცემზე.

შემომავალი ინფორმაციის ტიპი	შემომავალი	სწორად ამოცნობილის
-----------------------------	------------	--------------------

	ინფორმაციის რაოდენობა	პროცენტული წილი
ქსელის სავარჯიშო და სატესტო მონაცემები აღებულია ერთი და იმავე ადამიანების ხელნაწერ დოკუმენტებიდან, თუმცა რა თქმა უნდა არა ზუსტად ერთი დოკუმენტიდან	1000	90%
სატესტო მონაცემები - აღებულია სხვადასხვა ადამიანების ხელნაწერი დოკუმენტებიდან	500	75%

სიტყვების დამთხვევის შეფასება - სიტყვების დამთხვევა არის აპლიკაციის დამხმარე შესაძლებლობა, რითიც შეგვიძლია უკვე დაფიქსირებული შედეგის კორექტირება. შესაბამისად მისი არსებობაც ნაკლებ გავლენას ახდენს პროექტის მთლიან შეფასებაზე და წარმოადგეს მხოლოდ დამატებით ფუნქციონალს. ამის მიუხედავად, მაინც შეიძლება გარკვეული დასკვნის გაკეთება. ამ საკითხის შეფასება შესაძლებელია გაიყოს ორ ნაწილად დასაკორექტირებელი სიტყვის სიგრძის მიხედვით: პატარა სიტყვების შეთხვევაში, დამთხვევა ფიქსირდება საკმაოდ ბევრ სიტყვასთან, რის გამოც რთულდება იმის დადგენა თუ სინამდვილეში რომელ სიტყვას შეესაბამება. ამ შემთხვევაში საჭიროა დამატებითი ევრისტიკული მსჯელობები და მხოლოდ მის შემდეგ არის შესაძლებელი გარკვეული დასკვნის გამოტანა. საშუალო და დიდი სიგრძის სიტყვების შემთხვევაში აშკარა ხდება ყველაზე „მსგავსი“ სიტყვა და კორექტირებაც უფრო ზუსტ შედეგს იძლევა. საბოლოო ჯამში შეიძლება ითქვას, რომ ეს ფუნქციონალი საკმაოდ დიდ როლს თამაშობს საბოლოო შედეგში, თუმცა მაინც ითვლება დამატებით საშუალებად და არა ძირითად მეთოდად.

დასკვნა

ჩვენ უკვე აღწერეთ აპლიკაციის მუშაობის შედეგები გარკვეულ ეტაპებზე, ახლა კი შეგვიძლია აპლიკაციის მუშაობის ეფექტურობის შეფასება. „ხელნაწერის ამომცნობი“ აპლიკაცია საკმაოდ კარგად ასრულებს მასზე დაკისრებულ მოვალეობებს. პირველი-მას შეუძლია დიდი რაოდენობით მონაცემების დამუშავებაში დახმარება და საჭირო ქსელის შექმნა. ამ კუთხით აპლიკაციის გამოყენება ქმნის მრავალგზის შესწავლილ ქსელს, რომელიც შემდეგ ეტაპზე გამოიყენება და რაც მთავარია ეს ყველაფერი არის საკმაოდ გამარტივებული. მეორე მხრივ კარგი შეფასება შეიძლება მიეცეს ჩვენს მიერ შექმნილ (ნავარჯიშევ) ქსელს, რომელიც გაწვრთნილია ქართული სიმბოლოების ამოცნობაზე. რა თქმა უნდა, ქსელის უკეთესად ამუშავება შეიძლება, რადგან საჭიროა უფრო მეტი მონაცემების შეგროვება და ქსელის უფრო მეტი ვარჯიში. აპლიკაციის გამოყენება თავისუფლად არის შესაძლებელი გარკვეული სტილის შედარებით სტანდარტული დოკუმენტებისთვის და ასევე შესაძლებელია მისი გადაწყობა ისე, რომ პრაქტიკულად ნებისმიერი სახის დოკუმენტს მოერგოს.

გამოყენებული ლიტერატურა

http://docs.opencv.org/2.4/opencv_tutorials.pdf

<http://www.dkriesel.com/media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf>