

**МОНГОЛ УЛСЫН ИХ СУРГУУЛЬ  
ХЭРЭГЛЭЭНИЙ ШИНЖЛЭХ УХААН, ИНЖЕНЕРЧЛЭЛИЙН СУРГУУЛЬ  
МЭДЭЭЛЭЛ, КОМПЬЮТЕРИЙН УХААНЫ ТЭНХИМ**

Цэрэнбямбаагийн Хасбилэгт

**ГАР БИЧМЭЛ ТАНИЛТАНД ЗОРИУЛСАН  
МУИС-Н НЭЭЛТТЭЙ ӨГӨГДӨЛ БЭЛТГЭХ НЬ  
(Preparation of open NUM handwritten character dataset)**

Мэдээллийн технологи (D061304)  
Бакалаврын судалгааны ажил

Улаанбаатар

2020 он

**МОНГОЛ УЛСЫН ИХ СУРГУУЛЬ**  
**ХЭРЭГЛЭЭНИЙ ШИНЖЛЭХ УХААН, ИНЖЕНЕРЧЛЭЛИЙН СУРГУУЛЬ**  
**МЭДЭЭЛЭЛ, КОМПЬЮТЕРИЙН УХААНЫ ТЭНХИМ**

**ГАР БИЧМЭЛ ТАНИЛТАНД ЗОРИУЛСАН МУИС-Н**  
**НЭЭЛТТЭЙ ӨГӨГДӨЛ БЭЛТГЭХ НЬ**  
**(Preparation of open NUM handwritten character dataset)**

Мэдээллийн технологи (D061304)  
Бакалаврын судалгааны ажил

Удирдагч: \_\_\_\_\_ Др. Б.Сувдаа

Гүйцэтгэсэн: \_\_\_\_\_ Ц.Хасбилэгт (15B1SEAS0980)

Улаанбаатар

2020 он

# Зохиогчийн баталгаа

Миний бие Цэрэнбямбаагийн Хасбилэгт "ТАР БИЧМЭЛ ТАНИЛТАНД ЗОРИУЛСАН МУИС-Н НЭЭЛТТЭЙ ӨГӨГДӨЛ БЭЛТГЭХ НЬ" сэдэвтэй судалгааны ажлыг гүйцэтгэсэн болохыг зарлаж дараах зүйлсийг баталж байна:

- Ажил нь бүхэлдээ эсвэл ихэнхдээ Монгол Улсын Их Сургуулийн зэрэг горилохоор дэвшүүлсэн болно.
- Энэ ажлын аль нэг хэсгийг эсвэл бүхлээр нь ямар нэг их, дээд сургуулийн зэрэг горилохоор оруулж байгаагүй.
- Бусдын хийсэн ажлаас хуулбарлаагүй, ашигласан бол ишлэл, зүүлт хийсэн.
- Ажлыг би өөрөө (хамтарч) хийсэн ба миний хийсэн ажил, үзүүлсэн дэмжлэгийг дипломын ажилд тодорхой тусгасан.
- Ажилд тусалсан бүх эх сурвалжид талархаж байна.

Гарын үсэг: \_\_\_\_\_

Огноо: \_\_\_\_\_

## Гарчиг

ХУРААНГУЙ .....	1
Зорилго .....	1
Зорилт .....	1
Судалгаа .....	2
Аргын тодорхойлолт .....	2
Хөгжүүлэлт .....	2
Програм хангамж .....	3
Үр дүн .....	3

## **ХУРААНГУЙ**

Өнөөдөр хиймэл оюун ухаан болон машин сургалтын салбаруудын технологи, судалгаанууд хөгжихийн хэрээр, өмнө нь хүний төсөөлж байгаагүй олон шинэ боломжууд, зөвхөн технологийн салбаруудаар хязгаарлагдалгүй ар араасаа нээгдэж байгаагаас гадна өнөөдөр нийтэд ашиглагдаж буй технологиудын нарийвчлал, үр дүн, чанар ч мөн үүнийг дагаад сайжирч байгаа билээ. Энэ дундаас гар бичмэл танилттай холбоотой судалгаанууд олноор хөгжүүлэгдэх нь бидний мэдэх хүн компьютерийн харилцааг эергээр өөрчилж байна.

### **Зорилго**

Энэ чиглэлийн судалгаанууд, түүн дээр суурилсан програм хангамж, үйлчилгээ, бүтээгдэхүүн хөгжүүлэгдэхэд гар бичмэлийг таних, машин сургалтанд ашиглах өгөгдөл зайлшгүй хэрэгтэй, мөн Монгол хэл дээр энэ төрлийн нээлттэй, нэгдсэн өгөгдлийн сан хомс учир энэхүү судалгааны ажлаар өөрийн эх хэл дээрх нээлттэй гар бичмэлийн санг бий болгох, санг үүсгэхэд ашиглах арга, хэрэгслийг тодорхойлохыг зорьсон юм.

### **Зорилт**

Монгол хэл дээрх гар бичмэлийн нээлттэй санг үүсгэхдээ дараах үе шатын дагуу ажиллана.

1. Үндсэн шаардлагуудыг, аргуудыг тодорхойлох
2. Тэмдэгтүүд бичүүлж авах маягтын загварыг гаргах
3. Бөглөгдсөн маягт дээр боловсруулалт хийх, тэмдэгтүүдийг ялгаж авах
4. Их хэмжээний оролтын үед боловсруулалт хийх боломжтой байдлаар програмыг хөгжүүлэх
5. Програмын нээлттэй авч ашиглах боломжийг хангахын тулд эх код, програмыг GitHub, PyPi дээр байршуулах

## **Судалгаа**

Машин сургалт тэр дундаа гар бичмэл танилттай холбоотой машин сургалтын өгөгдлийг хэрхэн бэлтгэх, зохион байгуулах тал дээр бодитой, албан ёсны судалгааны ажлууд хомс байсан тул өгөгдлийн хэлбэр, хэмжээ, санг хэрхэн зохиомжлох зэргийг энэ чиглэлийн өөр бусад хэл дээрх түгээмэл ашиглагддаг нээлттэй өгөгдлүүдийг судалсаны үндсэн дээр тэдгээрийн жишигт тааруулан гүйцэтгэхээр шийдсэн.

## **Аргын тодорхойлолт**

Энэхүү ажлын мөн чанар нь үндсэн судалгаа бөгөөд гэсэн хэдий ч үр дүн, түүнтэй хамт ашиглагдах програм болон санг үүсгэхдээ өөрийн тодорхойлсон дараах шаардлагуудад нийцүүлэн хөгжүүлэхээр зорьсон. Үүнд:

1. Өгөгдөл нь цэвэр гараар бичигдсэн ба маш сайн боловсруулагдсан байх
2. Тэмдэгтийн зургууд дахин давтагдаагүй, мөн тухайн тэмдэгтийн бичигдэж болох аль болох олон хувилбаруудыг агуулсан байх
3. Тэмдэгтүүд дээр шинэ зургууд нэмж өгөгдлийн хэмжээг томруулах, зөвхөн шаардлагатай тэмдэгтүүдээр өөр шинэ датасет үүсгэх боломжтой байх
4. Өгөгдлийг аль болох олон төрлөөр авч ашиглах, шаардлагатай статистик мэдээллийг авах боломжтой байх
5. Өгөгдөл нь өөрөө болон, өгөгдөл үүсгэх программ нь хэнд ч авч ашиглахад нээлттэй байх

## **Хөгжүүлэлт**

Өөрийн тодорхойлсон шаардлага, аргын дагуу хөгжүүлэлтийг хийхдээ дараах технологиудыг авч ашигласан.

- OpenCV — Анх 2000 онд танилцуулагдсан энэхүү нээлттэй эхийн сан нь өнөөдөр компьютерийн хараа<sup>1</sup> -тай холбоотой ажил, судалгаануудад хамгийн өргөн ашиглагддаг бөгөөд зураг боловсруулалтанд энэхүү сангийн санал болгосон функц, аргуудыг ашиглана.
- Python — OpenCV санг Python болон C++ хэл дээр ашиглах боломжтой байдаг бөгөөд зураг боловсруулах програм нь хөгжүүлэлтийн болон нийтэд нээлттэй тавих боломжийг үндэслэн Python хэлийг сонгон ашиглана.
- PyPi<sup>2</sup> — Зураг боловсруулж буй программаа энэхүү платформ дээр байршуулах ба хүссэн хэн нь ч Python багц удирдагч ашиглан амархан татан авах, ашиглах боломжтой болно.

## Програм хангамж

Өөрийн ажлын судалгаан дээр үндэслэн өгөгдөл боловсруулах, сан үүсгэхээр хөгжүүлсэн програмууд болон тэдгээрийг хэрхэн ашиглах талаар авч үзнэ. Энэхүү ажил нь програмтай судалгаа буюу судалгаа голлосон, шаардлагтай байдлаар туслах програмуудыг хөгжүүлсэн учир удирдагч багшийн зөвлөснөөр шинжилгээ, зохиомж шаардлагагүйгээр CLI буюу команд мөр интерфэйстэйгээр хийж гүйцэтгэсэн.

## Үр дүн

Энэхүү судалгааны ажлаар бусад хэл дээрх гар бичмэл танилттай холбоотой судалгаануудын ажлууд дээр тулгуурлан өөрийн хэлний гар бичмэл танилтанд зориулсан нээлттэй сан, түүнд ашиглах програм хангамжийг хөгжүүлэхээр зорьлоо. Энэ судалгааны ажил нь цаашид гар бичмэл танилт болон энэ чиглэлийн судалгаа, програм хангамж, үйлчилгээ зэргийг хөгжүүлэхэд суурь болж өгнө гэдэгт итгэлтэй байна.

Нийт өгөгдлийн тоо хэмжээ, төрлүүдийг нэмэх, өргөжүүлэхийн тулд python хэл дээрх

---

<sup>1</sup>Компьютерийн хараа (= Computer Vision)

<sup>2</sup>The Python Package Index (PyPI) — <https://pypi.org/>

програм хангамжийн<sup>3</sup> <sup>4</sup> боломжуудыг тогтмол нэмэх, сайжруулах, бичвэр агуулах маяг-  
тыг динамик байдлаар үүсгэх, ажиллагааг хурдасгах, нийтээс маягтын зургаа авч автома-  
таар боловсруулан сангаа шинэчлэх бие даасан вэбсайт, систем гэх мэт ажлууд нэмэгдэж  
хийх шаардлагатай.

---

<sup>3</sup>Эх код: <https://github.com/khasbilegt/numiner>

<sup>4</sup>PyPi дээрх сан: <https://pypi.org/project/numiner/>