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

Анужингийн Сайнзолбоо

АЖИЛ ОЛГОГЧДЫН ӨГӨГДЛИЙН АНАЛИЗ СИСТЕМ ДЭЭР СУУРИЛСАН ЧАТ БОТ

(Chat bot based on sytem analysis of employers' data)

Мэдээллийн технологи (D061303)
Бакалаврын судалгааны ажил

Улаанбаатар

2022 оны 03 сар

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

АЖИЛ ОЛГОГЧДЫН ӨГӨГДЛИЙН АНАЛИЗ СИСТЕМ ДЭЭР СУУРИЛСАН ЧАТ БОТ

(Chat bot based on sytem analysis of employers' data)

Мэдээллийн технологи (D061303)
Бакалаврын судалгааны ажил

Удирдагч: _____ Б.Хуягбаатар доктор (Ph.D.)

Гүйцэтгэсэн: _____ А.Сайнзолбоо (18B1NUM1762)

Улаанбаатар

2022 оны 03 сар

Зохиогчийн баталгаа

Миний бие Анужингийн Сайнзолбоо ”АЖИЛ ОЛГОГЧДЫН ӨГӨГДЛИЙН АНАЛИЗ СИСТЕМ ДЭЭР СУУРИЛСАН ЧАТ БОТ” 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

ЗУРГИЙН ЖАГСААЛТ

ХҮСНЭГТИЙН ЖАГСААЛТ

Кодын жагсаалт

УДИРТГАЛ

Мэдээллийн технологи эрчимтэй хөгжиж буй өнөөгийн нийгэмд байгууллага үйл ажиллагаа явуулж эхэлсэн цагаасаа эхлэн өгөгдлийг үйлдвэрлэсээр байдаг. Тэдгээр өгөгдлийг байнга хадгалах нь өгөгдлийн сангийн нөөцөд хортой байдаг тул өгөгдөлд шинжилгээ хийж, тэдгээрээс шаардлагатай өгөгдлүүдийг түүвэрлэн хадгалах нь чухал юм.

Өнөөдөр бид дэлхий нийтээрээ хурдтай амьдралын хэмнэлд ажиллаж, амьдарч байна. Мөн зах зээлийн хөгжил, ажил олгогчийн эрэлт хэрэгцээ ажил хайгчийн хүсэл онирхлыг соновчтой бөгөөд хурдан холбож өгөх нь нэн шаардлагатай. Өнөөгийн байдлаар энэ эрэлт хэрэгцээг хангасан тодорхой шийдвэрлэсэн мэдээллийн систем хомс байна. Иймд энэхүү бакалаврын судалгааны ажлаар ажил олгогч болон ажил идэвхтэй хайгч хоёрыг түргэн шуурхай холбож өгөх чатбот системийг хөгжүүлж байна.

1. СЭДВИЙН ТАНИЛЦУУЛГА

1.1 Оршил

Энэхүү бакалаврын судалгааны ажлын хүрээнд ”Ажил олгогчдын өгөгдлийн анализ систем дээр суурилсан чатбот” сэдвийн дагуу ажил хайгчдыг ажлын байрны мэдээллээр хангах Чатбот системийг хөгжүүлнэ. Ажлын байрны мэдээллийг Data Scraping аргын тусламжтайгаар, системд шаардлагатай мэдээллийг өгөгдлийн сангийн хэлбэрт оруулан бүтэцтэйгээр нэгтгэн авах бөгөөд үүнээс ажил хайгчдын дунд байдаг түгээмэл асуултуудын хариултыг өгнө.
Мөн энэ системд машин сургалтын арга болох Language Understanding-ийг ашиглан хэрэглэгчийн асуултыг таамаглаж оновчтой хариулт өгөх боломжийг олгох юм.

1.2 Зорилго

Ажлын хайгчдын хэрэгцээт асуултад хариулж, ажлын байрны хүртээмжийг нэмэгдүүлэхэд энэхүү системийн гол зорилго оршино.

1.3 Зорилт

Дээрх зорилгод хүрэхийн тулд дараах зорилтуудыг тавьсан. Үүнд:

- Ашиглагдах технологиудыг сонгох, судлах
- Ижил төстэй системийн судалгаа хийх
- Системийн шинжилгээ хийх
- Системийг зохиомжлох
- Системийг хөгжүүлэх, сайжруулалт хийх

1.4 Алсын хараа

Ажлын байрны дэлгэрэнгүй мэдээллийг цуглуулснаар цаашид тэдгээрт шинжилгээ хийж хамгийн их эрэлттэй, өндөр цалинтай ажлын байр гэх зэрэг мэдээллүүдийг систем хэрэглэгчдэд хүргэх боломжтой юм.

2. СИСТЕМИЙН СУДАЛГАА

2.1 Системийн судалгаа

Сонгосон сэдэв болох ”Ажил олгогчдын өгөгдлийн анализ систем дээр суурилсан чатбот” сэдвийн хүрээнд судалгаа хийхдээ чатбот системийн талаар болон өгөгдөл цуглуулгын аргын талаар судалсан. Үүний дараа ижил төстэй системийн болон ашиглагдах технологийн талаар судалгааг хийсэн болно.

2.1.1 Чатбот систем

Чатбот систем нь ихэвчлэн хэрэглэгчийн асуултыг хиймэл оюун ухааны тусламжтайгаар ойлгож, хариултыг автоматжуулах үндсэн зорилготой компьютерийн програм хангамж юм. Орчин үед хэрэглэгчдэд туслах үндсэн үүргийн дагуу чатбот системийг байгууллагууд олон янзаар ашиглах болсон. Тэдгээрээс дурдвал,

- Цэс дээр суурилсан чатбот (Menu-based chatbot)
- Түлхүүр үгийг танихад суурилсан чатбот (Keyword recognition-based chatbot)
- Машин сургалтын чатбот (Machine learning chatbot)

Цэс дээр суурилсан чатбот

Өнөөгийн зах зээлд хэрэгжиж буй чатботуудын хамгийн энгийн бөгөөд түгээмэл хэлбэр юм.[1] [1] Хэрэглэгчийн асууж болох асуултуудыг урьдаас таамаглан хариултуудыг мод хэлбэртэйгээр бүтэцлэн хадгалдаг. Хэрэглэгч хүссэн хариултаа авахын тулд системийн хадгалсан хариултаар аялах хэрэгтэй болдог. Бусад чатботтой харьцуулбал, хариулт хязгаарлагдмал бөгөөд хэрэглэгчээс олон асуулт асууж цаг их шаарддагаараа сул талтай байдаг.

[1] https://www.engati.com/blog/types-of-chatbots-and-their-applications

Түлхүүр үгийг танихад суурилсан чатбот

Энэхүү чатбот нь хэрэглэгчийн бичсэнийг уншиж тохиромжтой хариултыг өгдөг. Ингэхдээ өгүүлбэрийг хиймэл оюун ухааны нэг хэсэг болох эх хэлний боловсруулалт (Natural Language Processing)-ын тусламжтайгаар шинжилж түлхүүр үгийг таньж хариултыг өгдөг. Ижил төстэй олон асуултад хариулах эсвэл түлхүүр үг дутуу үед амжилтгүй болдог. Мөн хэрэглэгч хүссэн хариултаа олж чадахгүй байх болон үр дүн муутай хариулт өгсөн тохиолдолд цэс дээр суурилсан чатботыг хослуулан ашиглах нь найдвартай болдог бөгөөд түгээмэл шийдлүүдийн нэг байдаг.

Машин сургалтын чатбот

Энэ төрлийн чатбот нь өмнө хэрэглэгчийн харилцан яриан дээр хиймэл оюун ухаан болон машин сургалтын тусламжтайгаар шинжилгээ хийж, хэрэглэгчийн зан төлөв, асуултын хэв маягаас суралцдаг. Ингэснээрээ чатботод хэрэглэгчийн зарцуулах цаг эрчимтэйгээр буурах буюу хариултаа авах алхам багасгах ба хэрэглэгчийн туршлага (UX) нь түүнийгээ даган өсөх нь энэхүү чатботын үндсэн зорилго болно.

Чатботыг сонгох

Машин сургалтын чатбот нь илүү уян хатан хэрэглэгчдэд ээлтэй чатботыг бий болгодог боловч хөгжүүлэхэд цаг хугацаа их шаардагдах ба машин өөрөө суралцахад мөн хугацаа шаардагддаг. Иймд системийн нөөц, шаардлагыг харгалзан үзэж энэхүү судалгааны ажлаар түлхүүр үг танихад суурилсан чатботыг хэрэгжүүлэхийг зориод байна.

2.1.2 Өгөгдөл цуглуулгын арга

Өгөгдөл цуглуулах (data scraping) нь хэрэглэгчдэд харагдаж буй өгөгдлийг олон янзын сувгаас цуглуулан хувийн орчинд хадгалан цаашид ашиглах боломжийг олгодог хамгийн үр дүнтэй автомат өгөгдөл олборлох арга юм. Ихэвчлэн өгөгдөл цуглуулах арга нь вэбсайтаас шаардлагатай өгөгдлийг цуглуулахад ашигладаг. Өгөгдөл цуглуулж буй хүнээс хамааран олборлосон өгөгдлийг таслалаар тусгаарлагдсан утгын (Comma-Separated Values) файл эсвэл

өгөгдлийн санд хадгалах боломжтой бөгөөд нэгэнт цуглуулсан их хэмжээний өгөгдөлд судалгаа шинжилгээ хийх, худалдаа, борлуулалтын хэрэгсэл болгох зэрэг олон төрлийн боломжийг олгодог.

Вебсайтаас өгөгдлийг олборлох хамгийн түгээмэл арга нь HTML parsing буюу HTML-ийг задлан шинжлэх юм. Энэ нь вебсайтын HTML болох сайтын үндсэн бүтцийг агуулгынх нь хамтаар хуулах бөгөөд авах гэж буй өгөгдлийн зан төрхийг нь зааж өгснөөр доторх агуулгыг хамгийн хялбар бөгөөд автомат байдлаар цуглуулдаг юм. Цуглуулга хийх 2 үндсэн арга байдаг. Үүнд:

- Өгөгдлийг цуглуулж, задлах (Data scraping)
- Өгөгдлийг олж илрүүлж, хаягийг цуглуулах (Data crawling)

Өгөгдлийг цуглуулж, задлах

Нэг үгээр хэлбэл өгөгдлийг цуглуулж, задлах нь зааж өгсөн хаягийн дагуу шаардлагатай өгөгдлийг задалж, хэрэгтэй агуулгыг хөгжүүлэгчдэд өгдөг бөгөөд хүссэн өгөгдлөө задлан авах боломжийг олгодгоороо давуу талтай. Өөрөөр хэлбэл өгөгдөл олборлох програм нь зорилго буюу даалгавараа мэдэж байгаа юм.

Өгөгдлийг олж илрүүлж, хаягийг цуглуулах

Энэхүү аргачлал нь хаяг тодорхой бус үед түүнийг олж илрүүлж шаардлагын дагуу хаягийг, зарим тохиолдолд өгөгдлийг цуглуулдаг. Системийн шаардлагын дагуу өгөгдлийг цуглуулах үед хаяг алгасах, дутуу өгөгдөл цуглуулахаас сэргийлдэг давуу талтай.

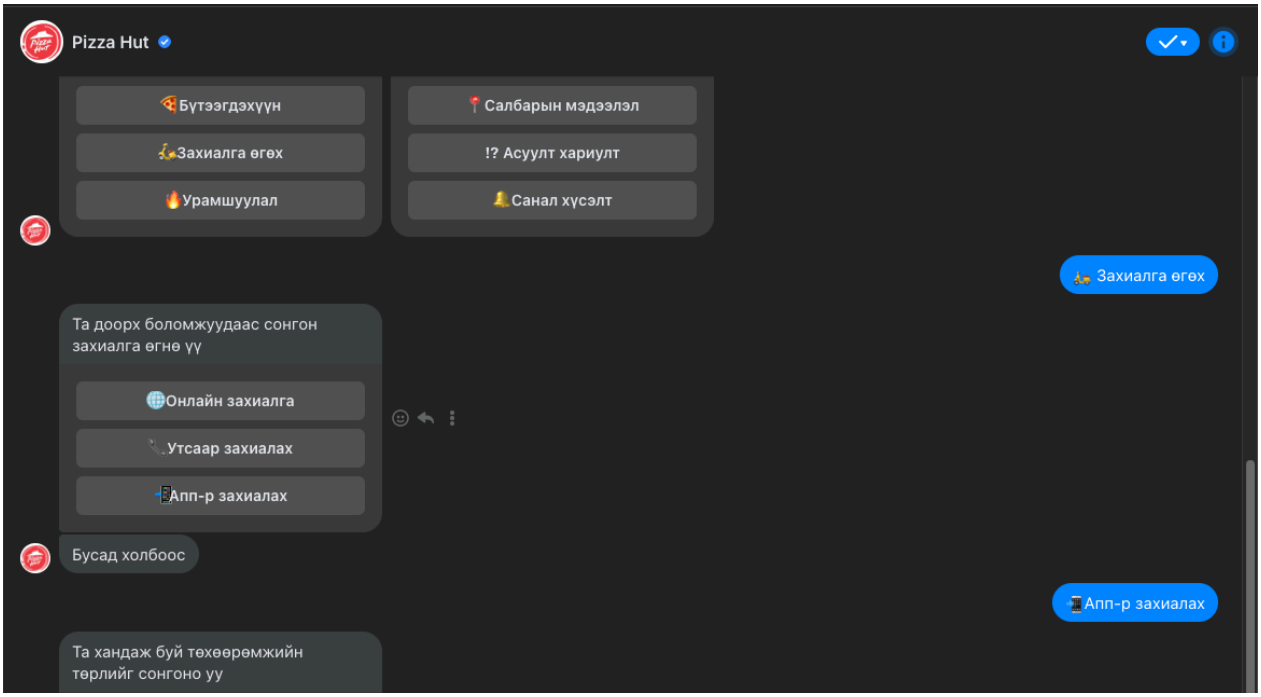
Ихэвчлэн энэхүү хоёр аргыг хослуулан ашигладаг бөгөөд шаардлагад нийцэх өгөгдлийг үлдээлгүй бүгдийг нь олоход *data crawling*-ийг ашиглах бол олсон өгөгдлийг задалж, шинжлэн өгөгдлийн санд хадгалах үйлдлийг *data scraping* хийдэг. Жишээлбэл, худалдааны сайтын бараа бүтээгдэхүүний өгөгдлийг цуглуулах гэж байгаа гэж үзвэл, барааны ангилалын хаягуудыг өөрчлөгдөх бүрт хадгалан өгөгдлийг цуглуулна. Өөрөөр хэлбэл нэг нь өөрчлөлт гарахыг ажиглаж вебсайтаар мөлхөж байх бол нөгөө нь шаардлагын дагуу бүх хэрэгтэй өгөгдлийг хэдийн цуглуулсан

байна. Энэхүү бакалаврын судалгааны ажлын хүрээнд өгөгдлийг CSV файл үүсгэн хадгалж цаашид ашигласан болно.

2.2 Ижил төстэй системүүд

2.2.1 Domino's Pizza & Pizza Hut

Domino's pizza хоолны газар нь захиалгын алхамаас эхлээд бүх мэдээллийг ганцхан *Facebook messenger chatbot* хангадаг. Чатбот эрчээ авч эхэлсэн шалтгаан нь хүмүүс, бусад хүмүүсийг хүлээлгүйгээр үйлчилгээ авах, тусламж авах зэрэг үйлчилгээг зэрэг нэвтрүүлсэнтэй холбоотой билээ. Үүний нэгэн адилаар Монголд үйл ажиллагаа явуулж буй Pizza Hut Mongolia юм.

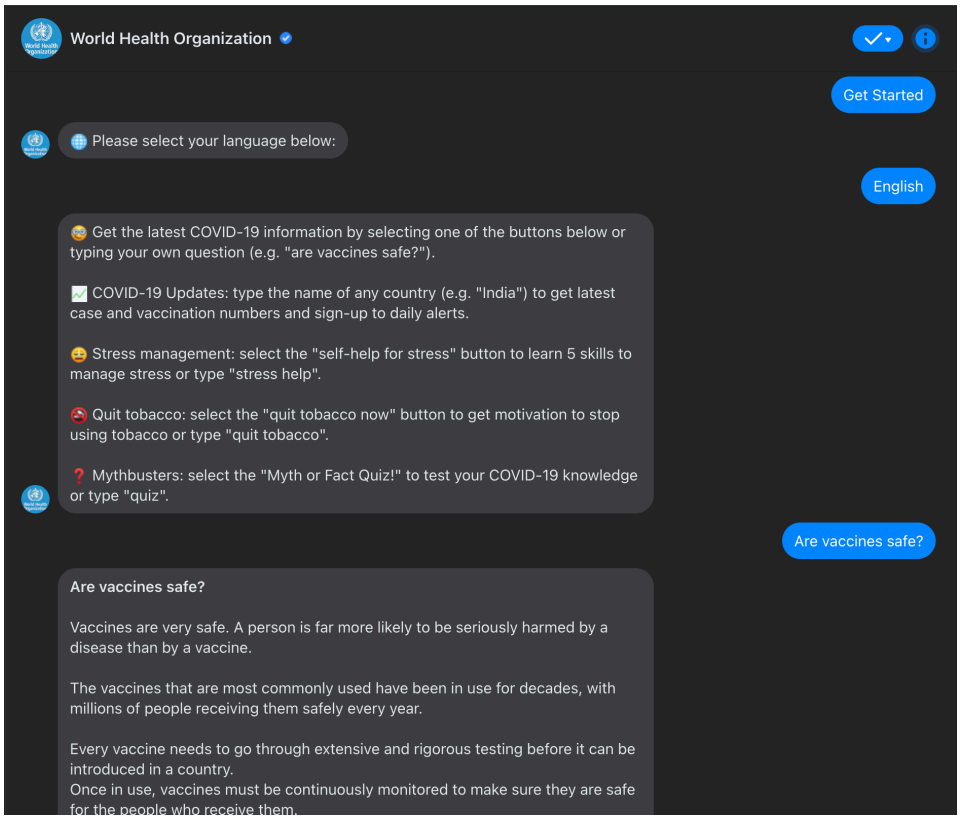


Зураг 2.1: Pizza Hut chat bot

Үйлчлүүлэгчдийн захиалга хүлээх хугацааг багасгахын тулд захиалгын үйл явцыг хурдасгаснаар тодорхой хэмжээнд нөлөөлж байгаа нь дээрх 2 жишээнээс харагдаж байна.

2.2.2 World Health Orgazination's Chat bot

Цар тахал болох коронавирусын эрчимтэй тархаж байх үед дэлхийн өнцөг булан бүрт оршин суугаа хүмүүст цар тахлын мэдээлэл, урьдчилан сэргийлэх арга, баталгаатай эх сурвалжийн мэдээллээр хангах зорилготой чатбот юм. Дэлхий нийтээр вакцинжуулалтын хөдөлгөөн өрнөж байх үеэр вакцины талаарх мэдээлэл, архаг хууч өвчинд нөлөөлөх талаар найдвартай, хамгийн сүүлийн үеийн албан ёсны мэдээллийг өгдөг. Хэдий халдварын тоо буурч, нийгэм өөрөө дасан зохьцож байгаа хэдий ч Дэлхийн Эрүүл Мэндийн байгууллага үүргээ гүйцэтгэж чухал эх сурвалжаар хангасаар байгаагийн шинж юм.



Зураг 2.2: WHO's chat bot

2.3 Технологийн судалгаа

АЖИЛ ОЛГОГЧДЫН ӨГӨГДЛИЙН АНАЛИЗ СИСТЕМ ДЭЭР СУУРИЛСАН ЧАТ БОТ-ыг хөгжүүлэхдээ өгөгдөл цуглуулгыг *python* хэлний сан болох *BeautifulSoup* HTML өгөгдөл задлах технологийг ашигласан бөгөөд чатбот системийн түлхүүр үг таних технологийг *Python* хэлний *Framework* болох *SentenceTransformers*-ийг сонгон хөгжүүлэлтийг хийсэн. Харин цуглуулсан өгөгдлийг *CSV* файлд хадгалан, *Microsoft Bot Framework*-ийг чатбот хөгжүүлэлтэд ашиглан судалгааг дараах байдлаар хийсэн болно.

2.3.1 Python

Python нь дээд түвшний маш олон төрлийн програмчлалыг өөртөө шингээсэн хэл юм. Хэлний сан болон *framework*-үүд нь тасралтгүй сайжирч, шинэчлэгдэж байдаг тул бүх л төрлийн програмчлалын аргуудыг гүйцэтгэж болдог. Орчин үед машин сургалт, хиймэл оюун ухаан болон эх хэлний боловсруулалтад(NLP) түгээмэл ашигладаг болсон бөгөөд веб хүртэл хийх боломжтойгоороо давуу талтай юм. Үүнээс гадна анхлан суралцаж буй хүмүүст ойлгоход хялбар *syntax*-ийн дүрэмтэй байдаг тул хэрэглэгчдийн тоо нь javascript, java хэлүүдтэй өрсөлдөхүйц байдаг.

Зураг 2.3: Python лого

```
x = 5
name = 'Sainzolboo'
print(x)
```

```
print(name)
```

Код 2.1: Python энгийн жишээ

Python програмчлалын хэл нь ойлгоход маш хялбар бөгөөд өөр дээрх функцууд нь шууд утгаараа ойлгомжтой байдаг. Syntax-ийн хувьд ; ашигладаггүй ба догол мөрөөр програмчлалын үндсэн схемийг гаргадагаараа онцлог хэл юм.

2.3.2 BeautifulSoup

Өгөгдөл цуглуулгын олон технологиудын нэг нь *BeautifulSoup* бөгөөд *python* програмчлалын хэлний сан юм. Энэ нь өгсөн вебсайтын хаяг (Url)-ийн дагуу бүх HTML өгөгдлийг агуулгын хамтаар нь хэрэглэгчид өгдөг. HTML хэл нь мод хэлбэртэй байдаг бөгөөд түүний хүүхэд элементүүдийн агуулгыг шаардлага болон түлхүүр үгийн дагуу цуглуулах зарчмаар ажилладаг.

Зураг 2.4: BeautifulSoup лого

Бакалаврын судалгааны ажлын сэдвийн дагуу ажлын байр олгогчдын мэдээлэл болон ажлын байрны мэдээллийг **zangia.mn**-ээс *BeautifulSoup* ашиглан цуглуулсан. Доорх кодын жишээнд бүх ажлын байрны ангилал болон шүүлтүүрийн агуулгыг цуглуулсан бөгөөд жишээнд зориулж зөвхөн эхний ангилалын мэдээллийг харуулав.

```
from bs4 import BeautifulSoup
import requests
```

```python
from urllib.error import HTTPError


url = 'https://zangia.mn/'
try:
    response = requests.get(url)
    response.raise_for_status()
except HTTPError as error:
    print(error)
soup = BeautifulSoup(response.text, 'html.parser')
navigatorList = soup.find_all('div', class_='filter')
print(navigatorList[0])
```

Код 2.2: BeautifulSoup жишээ өгөгдөл цугуулалт

```
<div class="filter">
 <h3>
  Онцлох
 </h3>
 <div>
  <a href="job/list/x.1">
   Удирдах албан тушаалын ажлын байр
  </a>
 </div>
 <div>
  <a href="job/list/x.3">
   Англи хэлний 100%-н мэдлэг шаардах ажлын байр
  </a>
 </div>
 <div>
  <a href="job/list/x.2">
   Ажлын туршлага шаардахгүй ажлын байр
  </a>
```

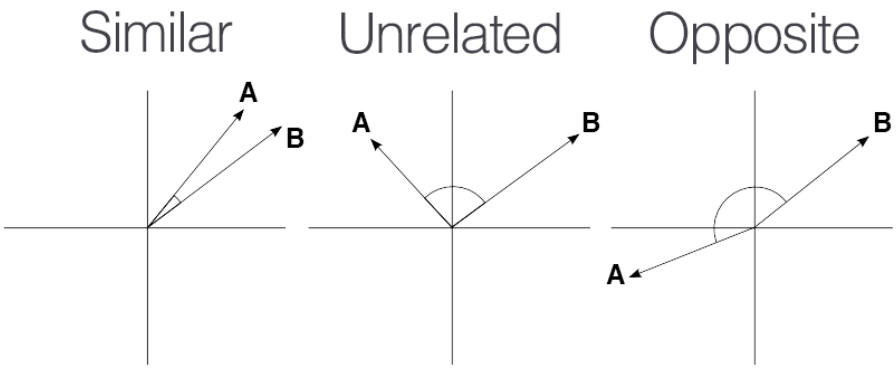
Зураг 2.5: Өгөгдөл цуглуулалтын жишээний үр дүн

2.3.3 SentenceTransformers

Python хэлний framework болох SentenceTransformers [2] буюу өгүүлбэр хувиргалт нь өгүүлбэр болон текстийн ижил төстэй байдал болон утгын хувьд адил байдлыг *cosine-similarity*[2] -ийн тусламжтайгаар тооцооолог. Энэхүү тооцооллыг цаашид өгүүлбэрийн ижил төстэй байдлыг

[2]Cosine-similarity нь өгөгдлийн шинжилгээнд 2 тооны ижил төстэй байдлыг вектор үржвэрээр илэрхийлдэг.

харьцуулах, хайлт хийх, түүнд шинжилгээ хийх зэргээр ашиглаж болно. Доорх зурагт өгүүлбэрт хувиргалт хийж, шинжилгээний үр дүнгийн вектор хоорондын өнцгөөр хэрхэн тодорхойлогддог болох талаар харуулав.



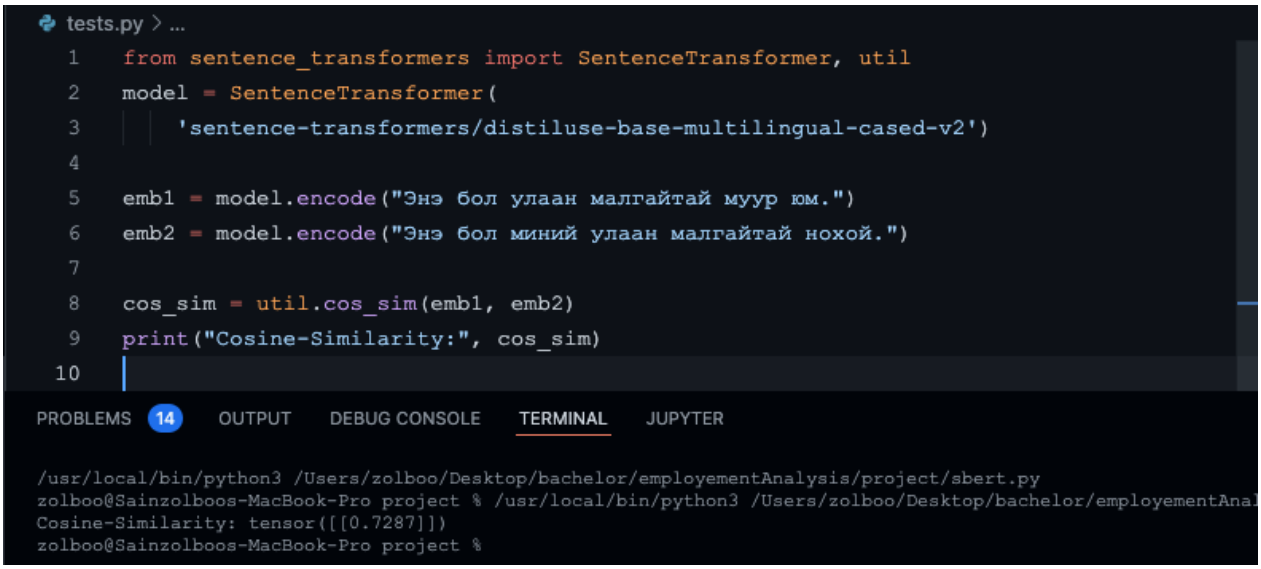
Зураг 2.6: Cosine similarity утгын график

SentenceTransformers-ийг дэлхийн 100 гаруй хэл дээр урьдчилан бэлтгэн, сургасан эх хэлний боловсруулалт (NLP)-ын загваруудыг ашиглаж болдогоороо давуу талтай. Чатбот системийн хувьд монгол хэлийг танин ашиглах боломжтой загвар болох *distiluse-base-multilingual-cased-v2*[3]-ийг ашиглан хийж гүйцэтгэв.

Хоёр өгүүлбэрийг *cosine-similarity* ашиглан ижил төстэй байдлыг илэрхийлэх жишээг доор харууллаа. Эх кодыг utf-8 формат танихгүй байсан тул зураг хэлбэрээр орууллаа.

2.3.4 Comma Separated Values - CSV файл

CSV нь өгөгдлийн утгуудыг тусгаарлахад таслал ашигладаг текст файл юм. Файлын мөр бүр нь өгөгдөл байдаг бөгөөд харгалзах утгуудад текст файлыг бичих энгийн өгөгдөл хадгалах технологи юм. Их өгөгдөлтэй хялбар харьцах боломжийг олгодгоороо давуу талтай.

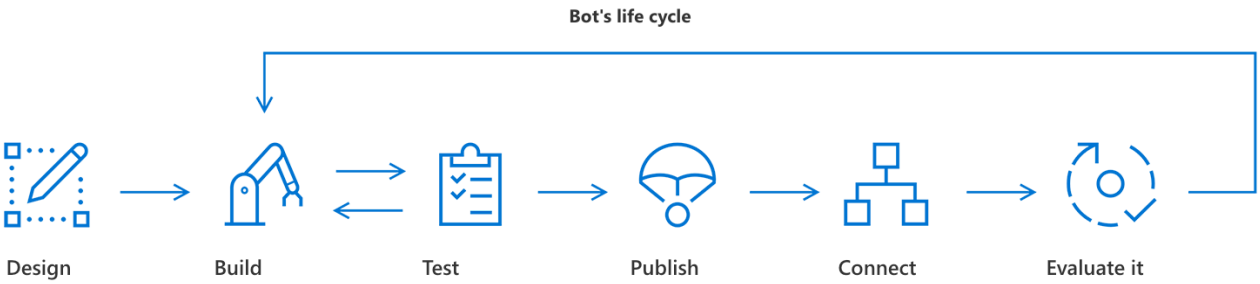


Зураг 2.7: cosine-similarity ашигласан жишээ

2.3.5 Microsoft Bot Framework

Bot Framework[3] нь Microsoft-ийн *Azure Bot Service*-ийн тусламжтайгаар чатботыг турших, үүсгэх, удирдах, хэрэглээнд нэвтрүүлэх гэх мэт боломжуудыг нэг дор хангаж өгдөг. Энэхүү боломжуудын хүрээнд асуулт хариултыг зохицуулах, хэрэглэгчид зориулсан User Interface бүтээх, Language Understanding аргыг ашиглах гэх мэт үйлдлүүдийг хийх боломжтой.

Bot бүтээх үйл явцыг Azure Bot Service болон Bot Framework нь ихэд хөнгөвчилж өгдөг бөгөөд доорх зурагт үзүүлсэн дарааллын дагуу Bot системийг бүтээдэг.



Зураг 2.8: Чатботын амьралын мөчлөг

[3]https://dev.botframework.com/

Design

Design буюу загварчлах нь төслийн төлөвлөгөөг гаргах юм. Өөрөөр хэлбэл, системийн зорилго, үйл явц, хэрэглэгчийн хэрэгцээг сайтар судлах нь амжилттай Bot систем бүтээх чухал хэсэг юм.

Build

Бот системийг угсрах буюу хөгжүүлэх үйл явц юм. Энэ алхамд хөгжүүлэгч хэрэглэгчийн харагдах хэсгийг загварчлах бөгөөд хөгжүүлэлтийн орчин нь *Azure Portal*, JavaScript, Python болон C програмчлалын хэлүүдээс сонгож хөгжүүлэлтийг гүйцэтгэх явц юм. Мөн системийн шаардлагыг тодорхойлсны дагуу бот системийг өргөжүүлж ашиглах боломжтой бөгөөд тэдгээрээс дурдвал:

- Эх хэлний боловсруулалт (NLP)
- Асуулт хариултыг сайжруулан мэдлэгийн сан үүсгэх
- Хэрэглэгчийн интерфейсийг сайжруулах

Test

Програм хангамжийн хөгжүүлэлтийн амьдралын мөчлөгийн адилаар тестийн үйл явцыг алгасаж болохгүй. Нэгэнт хэрэглэгчийн гарт бот системийг оруулахаас өмнө гарч болох алдаа дутагдлыг засан сайжруулах шаардлагатай. Иймд Bot системийг publish хийхээс өмнө заавал туршиж үзэх шаардлагатай. Энд Microsoft-ийн өөрсдийнх нь бие даасан програм болох *Bot emulator*-ийг ашиглан хөгжүүлэлтийн орчинд туршиж үзэх боломжийг олгодог.

Publish

Тестийн шатны дараа бот систем ашиглахад бэлэн болсон гэж үзсэн үед төсөл эсвэл чатботыг олон нийтэд ил болгох явдал юм.

Connect

Bot системээ Facebook messenger, Microsoft Teams, Telegram, Skype гэх мэт чат сувгуудыг өөрийн Bot-той холбоно.

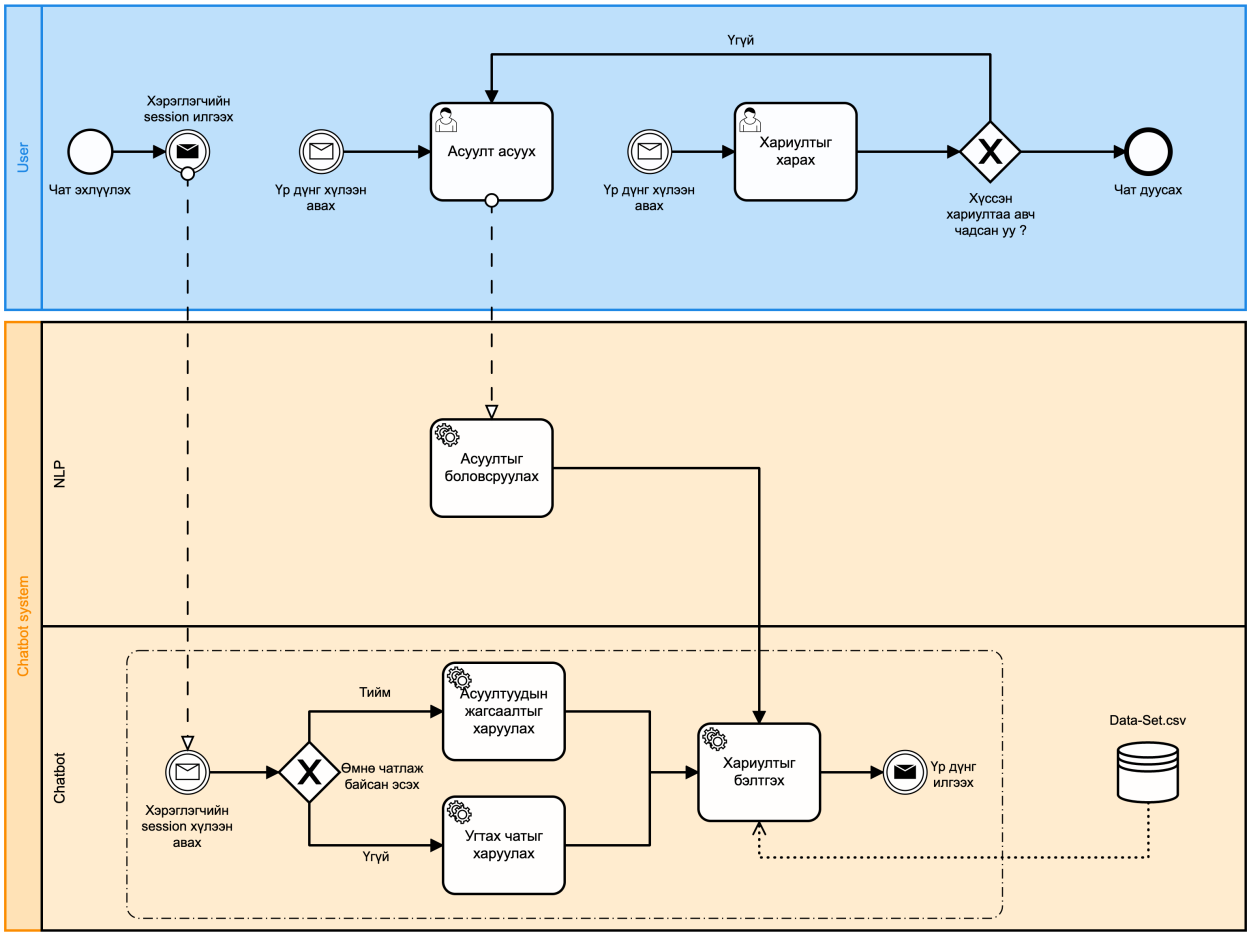
Зураг 2.9: Бот системийг холбож болох сувгууд

Ингэж бүх мөчлөгийг дууссаны дараа хөгжүүлэгч хэрэглэгчийн ашиглаж буй байдал дээр анализ хийж системийг дахин сайжруулах боломжтой бөгөөд буцаад угсрах үйл явцруу шилжих юм.

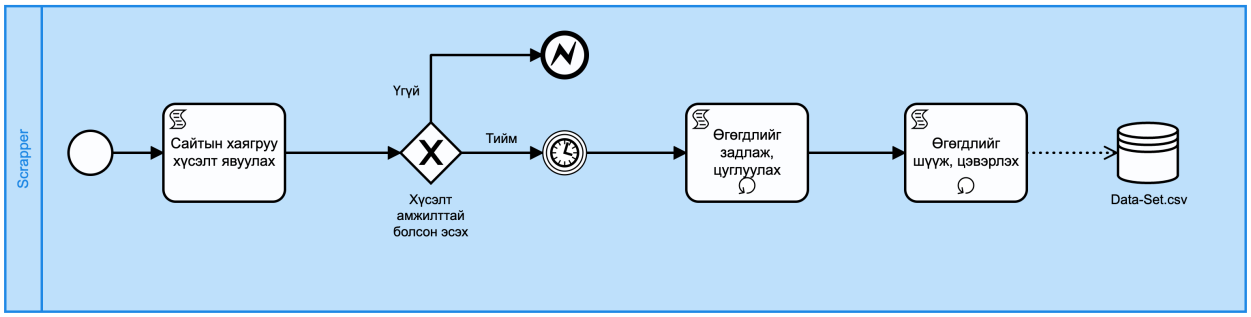
3. СИСТЕМИЙН ШИНЖИЛГЭЭ

3.1 Бизнесийн үйл ажиллагааны шинжилгээ

Бизнес процессийн модель нь чатбот системийн үндсэн процесс буюу үйл ажиллагааны явцыг BMPN-2.0 ашиглан дүрслэн харуулав [5]. Диаграмд дүрслэхдээ оролцогч талууд болох системүүдийг тус тусын *pool* дотор дүрсэлсэн бол дэд процесс буюу *subprocess*-ийг *lane*-д дүрсэлж хоорондын хамаарлыг харууллаа.



Зураг 3.1: BPMN 2.0-1



Зураг 3.2: BPMN 2.0-2

3.2 Хэрэглэгч

Чатбот системийг ямар ч хүн хэрэглэх боломжтой бөгөөд олон нийтэд нээлттэй байна. Системийн гол зорилго нь ажил хайж буй хэрэглэгчдэд ажлын байрны цогц мэдээллийг олгох зорилготой байх тул хэрэглэгчдийг дараах байдлаар тодорхойлж болно. Үүнд:

- Ажлын байр хайж буй хүн
- Хөгжүүлэгч

3.3 Функционал шаардлага

Дараах хэсэгт чатбот системд тавигдах функционал шаардлагуудыг харуулсан болно.

ФШ 1 Чатбот нь харилцан яриа эхэлмэгц хариу өгдөг байна.

ФШ 2 Чатбот нь ямар ч оролтод хариу өгнө.

ФШ 3 Хэрэв чатбот нь оролтод хариу өгч чадхааргүй байвал бусад асуултуудыг санал болгож ойлгомжгүй утга оруулсныг илэрхийлнэ.

ФШ 4 Чатботын санал болгох асуултууд нь цэс хэлбэртэй харагдана.

ФШ 5 Чатботын цэсэн дээр нэг товшилтоор асуултын хариултыг харуулдаг байна.

ФШ 6 Алхам бүрт үндсэн цэсрүү буцах сонголтыг харуулдаг байна.

ФШ 7 Чатботны хариулт нь текстэн хэлбэрээр хэрэглэгчид харагдана.

ФШ 8 Чатбот нь зөвхөн Монголоор бичсэн асуултад хариулт өгнө.

ФШ 9 Чатбот нь дэлгэрэнгүй мэдээллийг цэс хэлбэрээр сонгуулан харуулж чаддаг байна.

3.4 Функционал бус шаардлага

Бэлэн болон найдвартай байдал (Availability & Reliability)

ФБШ 01 Чатбот систем өдрийн аль ч цагт 99.999% ажиллагаатай байх ёстой.

ФБШ 02 Ямар ч хүсэлт ирсэн чатбот 100% хариу өгдөг байна.

Гүйцэтгэлтэй байдал (Performance)

ФБШ 03 Чатботын байршуулсан сувагт, шаардлагаас хамаарч ямар ч төхөөрөмжөөс хандаж болно.

ФБШ 04 Зарим тохиолдолд чатботын гүйцэтгэл нь хэрэглэгчийн интернет болон төхөөрөмжийн үйлдлийн системийн хувилбараас хамаарч болно.

Дэмжих чадвар (Supportability)

ФБШ 05 Чатботын эх кодыг *github* дээр нээлттэй эхийн систем хэлбэрээр байршуулна.

Хэрэгцээт байдал (Usability)

ФБШ 06 Чатбот нь хэрэглэхэд хялбар, ойлгомжтой байна.

ФБШ 07 Чатботны цэс нь ойлгомжтой цөөн үгээр илэрхийлэгдсэн байна.

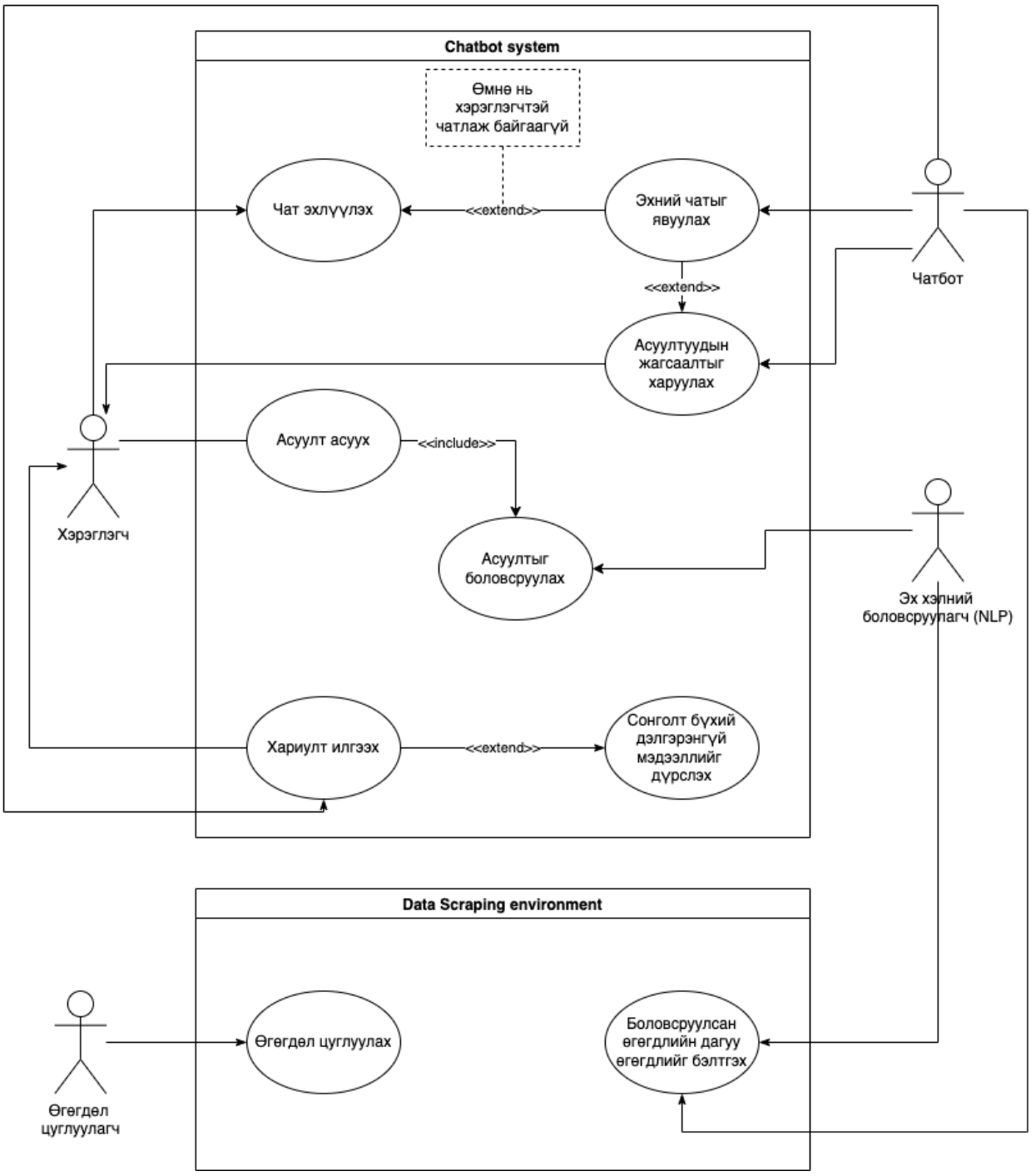
ФБШ 08 Чатботны цэсийн хэмжээ дарагдахуйц том байна.

Аюулгүй байдал(Security)

ФБШ 09 Чатбот системийн байршуулсан сувгийн стандартын дагуу хэрэглэгчийн мэдээллийг өгөгдлийн санд хадгалахгүй байна.

3.5 Use case диаграм

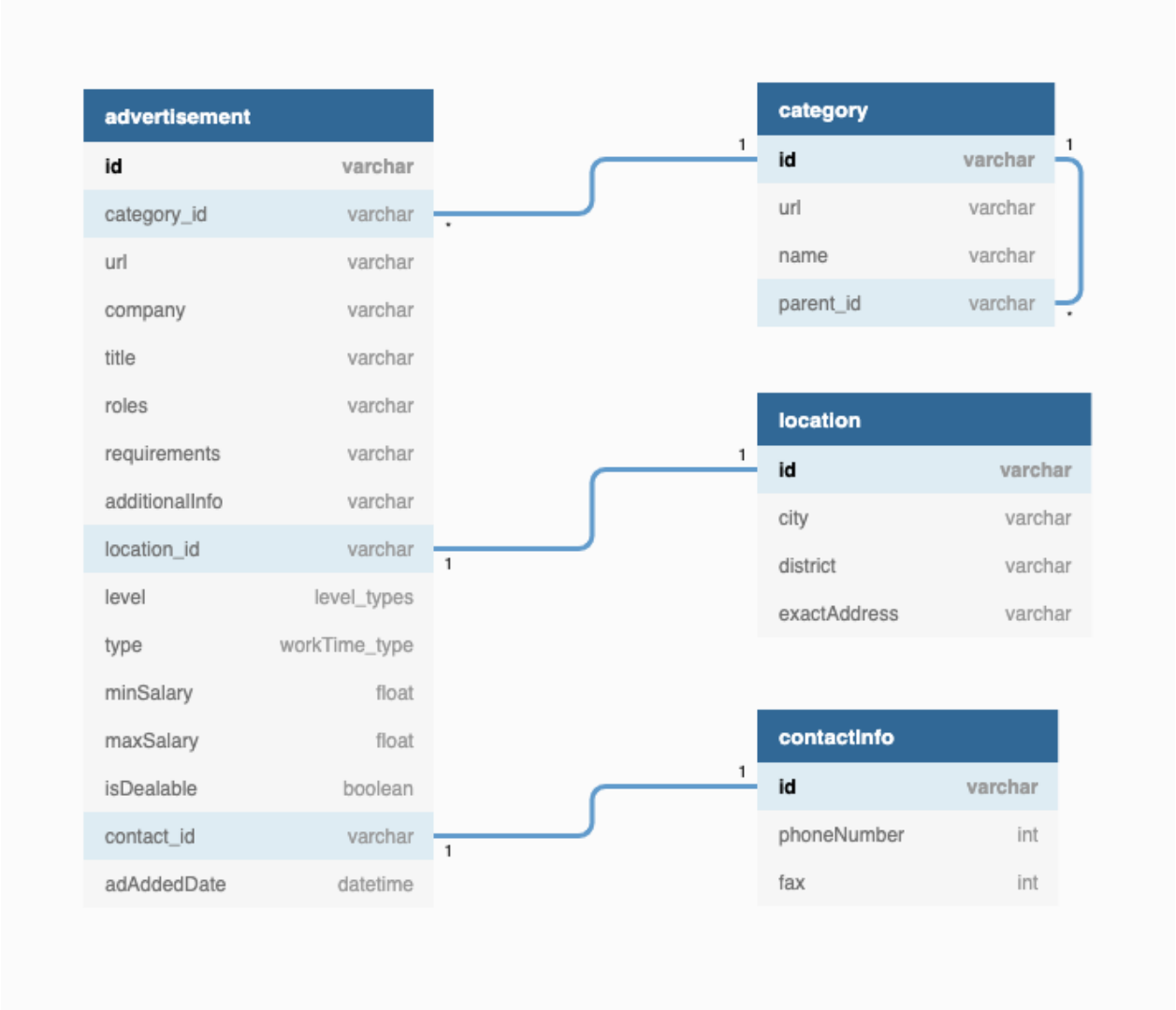
Чатбот системийн use-case диаграмыг байдлаар тодорхойлов [4]. Бараа материалын ажлын явцын диаграммыг дараах байдлаар тодорхойлов. Ажлын яв- цын диаграммын цар хүрээг хэрэгжүүлж буй гар утасны програм гэж тодорхойлсон. Үндсэн хэрэглэгчид болох салбар хариуцсан менежер, барааны нярав буюу ERP систем дээр бараа материалын модулийн эрх бүхий хэрэглэгч мөн дэлгүүрийн ажилтныг тоглогчоор дүрсэлсэн бөгөөд тэдгээрийн програм дээр хийх үйлдлийг дүрсэлж өгсөн болно.



Зураг 3.3: Use Case диаграм

4. СИСТЕМИЙН ЗОХИОМЖ

4.1 Өгөгдлийн сангийн диаграм



Зураг 4.1: Өгөгдлийн сангийн диаграм

4.2 Өгөгдлийн элемент

Чатбот системийн өгөгдлийн сангийн диаграмд харуулсан хүснэгтүүдэд агуулагдах мэдээлэл болон үүргийн талаар дэлгэрэнгүй тайлбарласан болно.

4.2.1 advertisement - Ажлын байрны зар

Ажлын байрны зар нь ямар категори буюу ангилалд, ямар холбоо барих хаягийн хамтаар хадгалагдаж буй мэдээлэл болон бусад дэлгэрэнгүй мэдээллийг харуулсан байна.

Table 4.1: advertisement хүснэгт

№	Баганын нэр	Түлхүүр өгөгдөл	Өгөгдлийн төрөл	Хоосон утга	Тайлбар
1	**id**	PK	varchar	not null	Ажлын байрны зарын дахин давтагдашгүй дугаар
2	category_id	FK	varchar	not null	Ажлын байрны зард хамаарах ангиллын дугаар
3	url		varchar	not null	Ажлын байрны зарын хаяг
4	company		varchar	not null	Ажил олгогч компани / хүн
5	title		varchar	not null	Ажлын зарын гарчиг
6	roles		varchar	null	Гүйцэтгэхүндсэн үүрэг
7	requirements		varchar	null	Ажлын байранд тавигдах шаардлага
8	additionalInfo		varchar	null	Нэмэлт мэдээлэл
9	location_id	FK	varchar	not null	Ажлын байрны зард хамаарах ангиллын дугаар
10	level		level_types	null	Ажлын түвшин
11	type		workTime_type	null	Ажиллах цагийн төрөл

№	Баганын нэр	Түлхүүр өгөгдөл	Өгөгдлийн төрөл	Хоосон утга	Тайлбар
12	minSalary		float	null	Доод цалин
13	maxSalary		float	null	Дээд цалин
14	isDealable		boolean	null	Тохиролцох эсэх
15	contact_id	FK	varchar	not null	Ажлын байрны зард хамаарах холбоо барих хаягийн дугаар
16	adAddedDate		datetime	not null	Зар нийтэлсэн огноо

Энд *level* буюу ажлын түвшин, *type* буюу ажлын цагийн өгөгдлийн төрлийг тодорхойлохдоо дараах байдлаар зааж өгсөн.

Enum level_types буюу ажлын түвшний шаардлага нь дараах үндсэн 4 өгөгдлийн төрлөөс хамаарна:

- student - Оюутан / дадлагажигч
- professional - Мэргэжлийн
- occupasionDoesntRequire - Мэргэжил шаардахгүй
- intermediateManagemet - Дунд шатны удирдлага

workTime_type буюу ажиллах цагийн нөхцөл нь дараах үндсэн 4 өгөгдлийн төрлөөс хамаарна:

- shift - Ээлжийн
- fullTime - Бүтэн цагийн
- halfTime - Хагас цагийн
- contract - Гэрээт / зөвлөх

4.2.2 category - Ангилал

Ажлын байрны зарын бүх ангиллуудын хаяг болон нэрийн мэдээллийг хадгалах хүснэгт юм. Ангиллууд нь дэд ангилал байж болох учир түүнийг эцэг ангиллын дугаарыг хадгалах байдлаар зохиомжлов.

Table 4.2: category хүснэгт

№	**Баганын нэр**	**Түлхүүр өгөгдөл**	**Өгөгдлийн төрөл**	**Хоосон Утга**	**Тайлбар**
1	**id**	PK	varchar	not null	Ажлын байрны зарын ангиллын дугаар
2	url		varchar	not null	Ангиллын хаяг
3	name		varchar	not null	Ангиллын нэр
4	parent_id	FK	varchar	null	Эцэг ангиллын дугаар

4.2.3 location - Байршил

Ажлын байрны байршил болон хот, аймаг, дүүргийн дэлгэрэнгүй өгөгдлийг хадгална.

Table 4.3: location хүснэгт

№	**Баганын нэр**	**Түлхүүр өгөгдөл**	**Өгөгдлийн төрөл**	**Хоосон Утга**	**Тайлбар**
1	**id**	PK	varchar	not null	Ажлын байрны зарын хаягийн дугаар
2	city		varchar	null	Ажлын байрны зарын байрших хот, аймгийн нэр

№	Баганын нэр	Түлхүүр өгөгдөл	Өгөгдлийн төрөл	Хоосон Утга	Тайлбар
3	district		varchar	null	Ажлын байрны зарын байрших дүүрэг, сумын нэр
4	exactAddress		varchar	null	Дэлгэрэнгүй хаяг

4.2.4 contactInfo - Холбоо барих

Ажлын байр олгогчийн хаягийн дэлгэрэнгүй өгөгдлийг хадгалана.

Table 4.4: contactInfo хүснэгт

№	Баганын нэр	Түлхүүр өгөгдөл	Өгөгдлийн төрөл	Хоосон Утга	Тайлбар
1	**id**	PK	varchar	not null	Ажил олгогчтой холбоо барих хаягийн дугаар
2	phoneNumber		int	null	Ажил олгогчийн утасны дугаар
3	fax		int	null	Ажил олгогчийн факс дугаар

4.3 Өгөгдлийн сангийн холбоосын тайлбар

- Нэг ангилал буюу категорид олон ажлын байрны зар байж болно.
- Нэг ангилал буюу категорид олон категори байж болно.
- Нэг ажлын байрны зард нэг байршлын мэдээлэл байна.
- Нэг ажлын байрны зард нэг холбоо барих хаягийн мэдээлэл байна.

5. ХЭРЭГЖҮҮЛЭЛТ, ҮР ДҮН

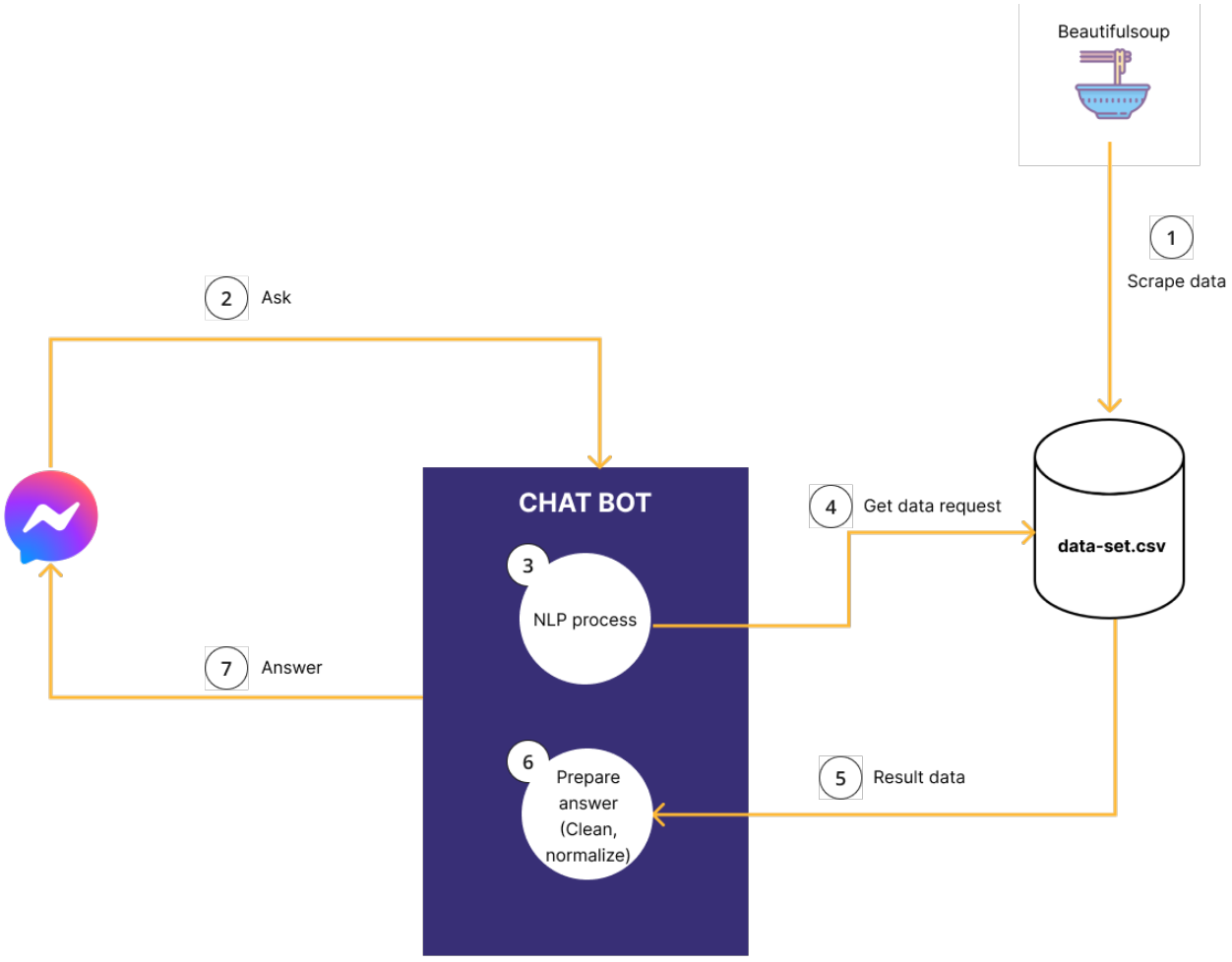
5.1 Хөгжүүлсэн байдал

Чатбот системийн хөгжүүлэлтийг хийхдээ шаардлагууд дээр үндэслэн, үечилсэн төлөвлөгөө болон шаардлагатай хөгжүүлэлтийг дэс дараалалтайгаар хийж гүйцэтгэсэн.

- Өгөгдөл цуглуулах
- Өгөгдлийг хадгалах, нэгтгэх, цэвэрлэх
- Системийн шаардлага, үйл ажиллагааг тодорхойлох
- Өгөгдөлд анализ хийх
- Эх хэлний боловсруулалт хийх
- Чатбот хөгжүүлэх

гэсэн дарааллын дагуу хөгжүүлэлтийг хийсэн болно.

Доорх зурагт чатбот системийн үндсэн процессийн зураглал харагдаж байна.



Зураг 5.1: Үндсэн процесс зураглал

5.1.1 Өгөгдөл цуглуулах

Үндсэн ашиглагдах өгөгдөл болох ажил олгогчид, ажлын байрны өгөгдлийг **zangia.mn**-ээс BeautifulSoup ашиглан авсан. Эхлээд вебсайтынхаа HTML бүтцийг нь судалж, авах өгөгдлийнхөө класс утгуудыг (className) олж авах нь зөв юм. Вебсайтаас өгөгдөл цуглуулах 2 үндсэн арга байдгаас өгөгдлийг олж илрүүлж, хаягийг цуглуулах (data crawling) аргаар бүх ангиллуудын хаяг (url)-уудын түүж авна. Харин data scraping нь тэр хооронд олсон бүх хаягуудаараа явж хэрэгтэй агуулгыг цуглуулна. [1]

[1] Кодын жишээг оруулахад utf-8 формат танихгүй байсан тул монголоос галиглаж бичсэн болно.

```
initialUrl = 'https://www.zangia.mn/'
today = str(date.today())
# all categories set
categorySet = set()
# all advertisement's link set
adUrlDict = {}
# all ads object set
adsSet = set()

# scrape initial links
soup = useScrape(initialUrl)
navigatorList = soup.find_all('div', class_='filter')
for navigator in navigatorList:
    if navigator.find('h3').text.strip() != 'Salbar, mergejil':
        continue
    # ALL CATEGORY LINKS
    categoryList = navigator.find_all('div')

for categoryItem in categoryList:
    categories = categoryItem.find('a')
    url = initialUrl + categories['href']
    tempCategory = Category(url, categories.text, '')
    soup = useScrape(url)
    subCategory = soup.find('div', class_='pros')
    # ALL SUBCATEGORY LINKS
    subCategoryList = subCategory.find_all('a')
    for subCategoryItem in subCategoryList:
        subCategoryUrl = initialUrl + subCategoryItem['href']
```

```
29          tempSubCategory = Category(
30              subCategoryUrl, subCategoryItem.text, tempCategory.name)
31          categorySet.add(tempSubCategory)
```

Код 5.1: Data Link crawling

Дээрх код нь эхлээд вебсайтруу орж ”filter” класс доторх ”Салбар, мэргэжил” гэсэн хэсгээс бүх эцэг категориудыг data crawling хийж авч байна. Үүний дараа хүүхэд категориудыг олж categorySet дотор бүх хаягуудыг хийж хадгалж байна. [2] Энд categorySet set-ийн элемент нь category төрлийн объект бөгөөд өгөгдлийн сангийн диаграм дээр тодорхойлж өгсөн байгаа. Ингэснээр data crawling-ийг зогсоож, цуглуулсан хаягаасаа өгөгдлөө цуглуулъя.

```
1  for categoryItem in categorySet:
2      if categoryItem.parentId == '':
3          continue
4      soup = useScrape(categoryItem.url)
5      hasPagination = soup.find('div', class_='page-link')
6      pagesUrl = []
7      if hasPagination != None:
8          pagesUrl = createLinkList(hasPagination, categoryItem.url)
9      else:
10         pagesUrl.append(categoryItem.url)
11     for pageUrl in pagesUrl:
12         soup = useScrape(pageUrl)
13         ads = soup.find_all('div', class_='ad')
14         # CREATE UNIQUE AD DICTIONARY
15         for ad in ads:
16             adUrl = initialUrl+ad.find('a', class_=None)['href']
17             adUrlDict[adUrl] = categoryItem
```

[2]Python хэлний set өгөгдлийн төрөл нь давхацахгүй утгуудын хүснэгт гэж хэлж болно.

```
    pagesUrl.clear()
```

Код 5.2: Өгөгдөл цуглуулах

Дээрх кодонд бүх хүүхэд категориудын дотор агуулагдаж буй зарын мэдээллийг цуглуулж байна. Ингэхдээ эхлээд категори доторх өгөгдлүүд нь хуудаслагдсан (pagination) байх боломжтой бөгөөд хэрэв олон хуудастай байвал хаягуудыг нь угсарч тэдгээрээс ч мөн өгөгдлийг нь цуглуулах ёстой юм.

```
from array import array
from .regex import a as useRegex
from .scrape import UseBeautifulSoup as useScrape


def createLinkList(pagination, url) -> array:
    linkList = []
    total = int(useRegex(pagination.find_all('a')[-1]['href']))

    for i in range(total + 1):
        if i == 0:
            continue
        link = url + '/pg.' + str(i)
        linkList.append(link)
    return linkList
```

Код 5.3: Хуудаслалтыг задлах

Энэ хэсэгт хуудаслан дугаарласан хэсгийн хамгийн сүүлийн тоог авч *createLinkList* функцруу дамжуулснаар тухайн категорийн бүх өгөгдлийг цуглуулах боломж үүсч байгаа юм. Ингээд дахин data crawling хийж бүх хаягуудыг цуглуулж энэ удаад dictionary үүсгэж зарын хаягуудыг хадгалсан. Энд disctionary үүсгэхдээ хаягийг нь түлхүүр(key) болгож категори объектыг нь

утга(value) болгож хадгалсан. Мэдээж хэрэг dictionary нь түлхүүр давхцахаас сэргийлдэг тул бид ямар нэгэн байдлаар нэг зарын өгөгдлийг 2 удаа цуглуулах эрсдэлгүй болж байна. [3]

Харин одоо үүсгэсэн dictionary-оо ашиглан өгөгдлөө CSV файлруугаа бичихэд ашиглаж болно.

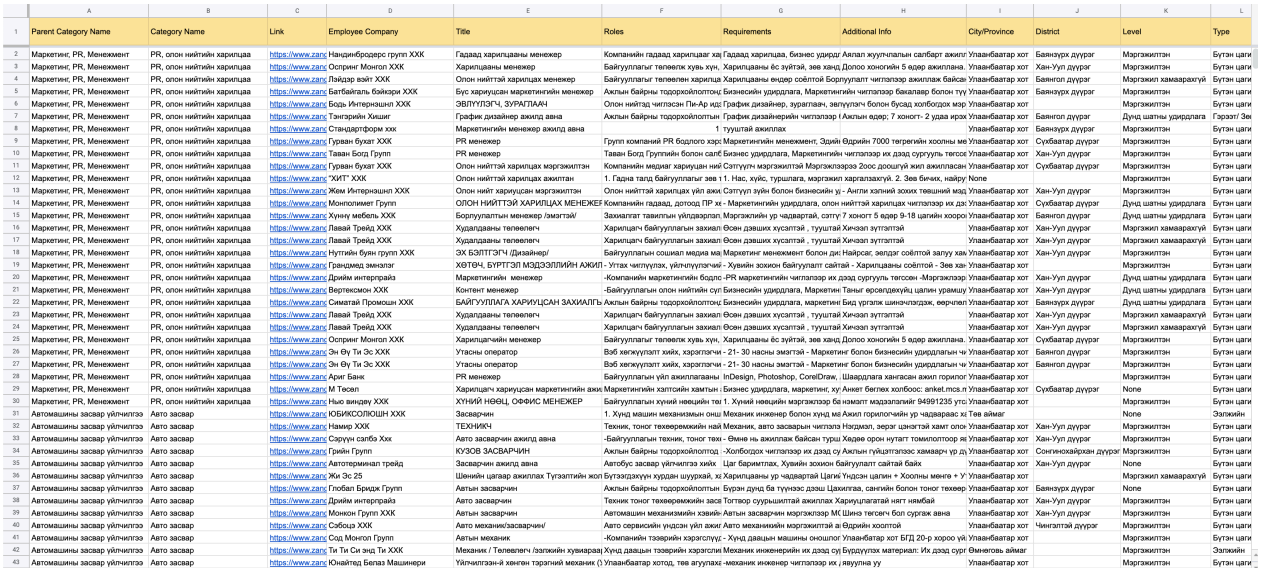
```
file = open(today+'adScrape.csv', 'w', encoding='utf-8')
file.write('Parent Category Name' + '\t' +
           'Category Name ' + '\t' +
           'Link' + '\t' +
           'Employee Company' + '\t' +
           'Title' + '\t' +
           'Roles' + '\t' +
           'Requirements' + '\t' +
           'Additional Info' + '\t' +
           'City/Province' + '\t' +
           'District' + '\t'
           'Level' + '\t' +
           'Type' + '\t' +
           'Min Salary' + '\t' +
           'Max Salary' + '\t' +
           'Is Dealable' + '\t' +
           'Address' + '\t' +
           'Phone' + '\t' +
           'Fax' + '\t' +
           'Ad Added Date' + '\n')

for adUrl in adUrlDict:
    print(adUrl)
```

[3]Нэг зарын өгөгдлийг цуглуулахад интернетийн хурдаас хамааран 0,2-оос 0.5 хугацаа зарцуулдаг

```
    try:
        tempAdItem = useAdScrape(adUrl)
        tempAdItem.setCategory(adUrlDict[adUrl])
        file.write(
            tempAdItem.category.parentId+'\t' +
            tempAdItem.category.name+'\t' +
            tempAdItem.url+'\t' +
            tempAdItem.company+'\t' +
            tempAdItem.title+'\t' +
            tempAdItem.roles+'\t' +
            tempAdItem.requirements+'\t' +
            tempAdItem.additionalInfo+'\t' +
            tempAdItem.city+'\t' +
            tempAdItem.district+'\t' +
            tempAdItem.level+'\t' +
            tempAdItem.type+'\t' +
            tempAdItem.minSalary+'\t' +
            tempAdItem.maxSalary+'\t' +
            tempAdItem.isDealable+'\t' +
            tempAdItem.address+'\t' +
            tempAdItem.phoneNumber+'\t' +
            tempAdItem.fax+'\t' +
            tempAdItem.adAddedDate+'\n')
        del tempAdItem
    except:
        print('Ad writing error')
file.close()
```

Код 5.4: CSV файлруу хадгалах

Дээрх код нь энгийн python програм файлтай харьцаж өөрт цуглуулсан өгөгдлөө хадгалж байна. Нийт өгөгдлийн хүснэгтийг энд [4] оруулав.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Parent Category Name	Category Name	Link	Employee Company	Title	Roles	Requirements	Additional Info	City/Province	District	Level	Type
2	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Нандинбродерс групп ХХК	Гадаад харилцааны менежер	Компанийн гадаад харилцаат хар	Гадаад харилцаа, бизнес удирд	Авлал жуулчлалын салбарт ажилт	Улаанбаатар хот	Баянзүрх дүүрэг	Мэргэжилтэн	Бүтэн цаг
3	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Оспринг Монгол ХХК	Харилцааны менежер	Байгууллагыг төлөөлж хувь хүн,	Харилцааны ёс зүйтэй, зөв хэнд	Долоо хоногийн 5 өдөр ажиллана	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
4	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Лайдэр вэйт ХХК	Олон нийттэй харилцах менежер	Байгууллагыг төлөөлөн харилца	Харилцааны өндөр сөөлтой Борлуулалт чиглэлээр ажилла	Байгаа	Улаанбаатар хот	Баянгол дүүрэг	Мэргэжил хамаарахгүй	Бүтэн цаг
5	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Батбайгаль байгэри ХХК	Бүс харицаан маркетингийн менежер	Ажлын байрны тодорхойлогдолт	Бизнесийн удирдлага, Маркетингийн чиглэлээр бакалавр болон түү	Улаанбаатар хот	Баянзүрх дүүрэг	Мэргэжилтэн	Бүтэн цаг	
6	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Бодь Интернэшнл ХХК	ЭВЛҮҮЛЭГЧ, ЗУРАГЛААЧ	Олон нийтэд чиглэсэн Пи-Ар ид	График дизайнер, зураглаач, эвлүүлэгч болон бусад холбогдох мэр	Улаанбаатар хот		Мэргэжилтэн	Бүтэн цаг	
7	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Тэнгэрийн Хишиг	График дизайнер ажилд авна	Ажлын байрны тодорхойлогдолт	График дизайнерийн чиглэлээр (	Ажлын өдөр: 7 хоногт- 2 удаа ирэх	Улаанбаатар хот	Баянгол дүүрэг	Дунд шатны удирдлага	Гэрээт/ Зө
8	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Стандартформ хөк	Маркетингийн менежер ажилд авна	1 туршлагатай ажилтан			Улаанбаатар хот	Баянзүрх дүүрэг	Мэргэжилтэн	Бүтэн цаг
9	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Гуравн бухат ХХК	PR менежер	Групп компаний PR бодлого хэр	Маркетингийнийн менежмент; Эдий	Өдрийн 7000 төгрөгийн хоолны ме	Улаанбаатар хот	Сүхбаатар дүүрэг	Мэргэжилтэн	Бүтэн цаг
10	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Таван Богд Груп ХХК	PR менежер	Таван Богд Группийн болон салб	Бизнес удирдлага, Маркетингийн чиглэлээр их дээд сургууль төгсөө	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг	
11	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Гуравн бухат ХХК	Олон нийттэй харилцах мэргэжилтэн	Компанийн медиаг харицаан нэ	Сэтгүүлч мэргэжилтэй Мэргэжлээрээ 2оос дээшгүй жил ажилласан	Улаанбаатар хот	Сүхбаатар дүүрэг	Мэргэжилтэн	Бүтэн цаг	
12	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	"ХИТ" ХХК	Олон нийттэй харилцах ажилтан	1. Гадна талд байгууллагыг зөв т	1. Нас, хүйс, туршлага, мэргэжил харгалзахгүй. 2. Зөв бичиг, найру	None		Мэргэжилтэн	Бүтэн цаг	
13	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Жем Интернэшнл ХХК	Олон нийт харицуцан мэргэжилтэн	Олон нийттэй харилцах үйл ажи	Сэтгүүл зүйн болон бизнесийн у	- Англи хэлний зохих төвшний мэд	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
14	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	ОЛОН НИЙТТЭЙ ХАРИЛЦАХ МЕНЕЖЕГ	Компанийн гадаад, дотоод ПР хэ	- Маркетингийн удирдлага, олон нийттэй харилцах чиглэлээр их дэ			Улаанбаатар хот	Сүхбаатар дүүрэг	Дунд шатны удирдлага	Бүтэн цаг
15	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Хүнөү мебель ХХК	Борлуулалтын менежер /эмэгтэй/	Захиалгат тавилгын үйлдвэрлэл	Мэргэжлийн ур чадвартай, сэтгү	7 хоногт 5 өдөр 9-18 цагийн хоорс	Улаанбаатар хот	Баянгол дүүрэг	Дунд шатны удирдлага	Бүтэн цаг
16	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Лавай Трейд ХХК	Худалдааны төлөөлөгч	Харилцагч байгууллагын захиал	Өсөн дэвших хүсэлтэй , туушай	Хичээл зүтгэлтэй	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжил хамаарахгүй	Бүтэн цаг
17	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Лавай Трейд ХХК	Худалдааны төлөөлөгч	Харилцагч байгууллагын захиал	Өсөн дэвших хүсэлтэй , туушай	Хичээл зүтгэлтэй	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжил хамаарахгүй	Бүтэн цаг
18	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Нүүгийн бүян групп ХХК	ЭХ БЭЛТГЭГЧ /Дизайнер/	Байгууллагын сошиал медиа ма	Маркетинг менежмент болон ди	Найрсаг, эелдэг сөөлтой залуу хам	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
19	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Грандмед эмнэлэг	ХӨТӨЧ, БҮРТГЭЛ МЭДЭЭЛЛИЙН АЖИЛ	- Үүтах чиглүүлэх, үйлчлүүлэгчий	- Хувийн зохион байгуулалт сайтай - Харилцааны сөөтгой - Зөв хэ		Улаанбаатар хот		Мэргэжилтэн	Бүтэн цаг
20	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Дрэйм интерграйз	Маркетингийн менежер	-Компанийн маркетингийн бодлс	-PR маркетингийнийн чиглэлээр их дээд сургууль төгссөн -Мэргэжлээр	Улаанбаатар хот	Хан-Уул дүүрэг	Дунд шатны удирдлага	Бүтэн цаг	
21	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Вертексмон ХХК	Контент менежер	-Байгууллагын олон нийтийн суг	Бизнесийн удирдлага, Маркетин	Таныг ёрхлэдвэхүйц цалин урамшу	Улаанбаатар хот	Хан-Уул дүүрэг	Дунд шатны удирдлага	Бүтэн цаг
22	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Сматай Промошн ХХК	БАЙГУУЛЛАГА ХАРИУЦСАН ЗАХИАЛПЪ	Ажлын байрны тодорхойлогдолт	Бизнесийн удирдлага, маркетинг	Бид үргэлж шинэчлэгдэж, өөрчлөл	Улаанбаатар хот	Баянхүрх дүүрэг	Дунд шатны удирдлага	Бүтэн цаг
23	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Лавай Трейд ХХК	Худалдааны төлөөлөгч	Харилцагч байгууллагын захиал	Өсөн дэвших хүсэлтэй , туушай	Хичээл зүтгэлтэй	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжил хамаарахгүй	Бүтэн цаг
24	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Лавай Трейд ХХК	Худалдааны төлөөлөгч	Харилцагч байгууллагын захиал	Өсөн дэвших хүсэлтэй , туушай	Хичээл зүтгэлтэй	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжил хамаарахгүй	Бүтэн цаг
25	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Оспринг Монгол ХХК	Харилцагчийн менежер	Байгууллагыг төлөөлж хувь хүн,	Харилцааны ёс зүйтэй, зөв хэнд	Долоо хоногийн 5 өдөр ажиллана	Улаанбаатар хот	Сүхбаатар дүүрэг	Мэргэжилтэн	Бүтэн цаг
26	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Эн Өү Ти Эс ХХК	Утасны оператор	Взб хөгжүүлэлт хийх, хэрэгтэчи	- 21- 30 насны эмэгтэй - Маркетинг болон бизнесийн удирдлагын чи		Улаанбаатар хот	Баянгол дүүрэг	Мэргэжилтэн	Бүтэн цаг
27	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Эн Өү Ти Эс ХХК	Утасны оператор	Взб хөгжүүлэлт хийх, хэрэгтэчи	- 21- 30 насны эмэгтэй - Маркетинг болон бизнесийн удирдлагын чи		Улаанбаатар хот	Баянгол дүүрэг	Мэргэжилтэн	Бүтэн цаг
28	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Арит Банк	PR менежер	Байгууллагын үйл ажиллагааны	InDesign, Photoshop, CorelDraw,	Шаардлага хангасан ажил горилог	Улаанбаатар хот		Мэргэжилтэн	Бүтэн цаг
29	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	М Төсөл	Харицацач харицуцан маркетингийн ажи	Маркетингийнийн хэлтэсийн хэмтэн	Бизнес удирдлага, маркетинг, ху	Анкет бөглөх холбоос: anket.mcs.n	Улаанбаатар хот	Сүхбаатар дүүрэг	None	Бүтэн цаг
30	Маркетинг, PR, Менежмент	PR, олон нийтийн харилцаа	https://www.zarg	Нью виндоу ХХК	ХУНИЙ НӨӨЦ, ОФФИС МЕНЕЖЕР	Байгууллагын хүний нөөцийн тө	1. Хүний нөөцийн мэргэжлээр бө	нэмэлт мэдээллийг 94991235 утс	Улаанбаатар хот		Мэргэжилтэн	Бүтэн цаг
31	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	ЮБИКСОЛЮШН ХХК	Засварчин	1. Хүнд машин механизмын онц	Механик инженер болон хүнд мэ	Ажил горилогчийн ур чадвараас х	Төв аймаг		None	Эрэлжийн
32	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Намир ХХК	ТЕХНИКЧ	Техник, тоног төхөөрөмжийн най	Механик, авто засварын чиглэлэ	Нэгдмэл, зорга цэнгэгтэй хамт олн	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
33	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Сарүүн салбэ Ххк	Авто засварчин ажилд авна	-Байгууллагын техник, тоног тех	- Өмнө нь ажилласан байсан турш	Хадөө орон нутагт томилолтоор яв	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
34	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Грийн Групп	КУЗОВ ЗАСВАРЧИН	Ажлын байрны тодорхойлогдо	-Холбоодох чиглэлээр их дээд су	Ажлын гүйцэтгэлээс хамаарч ур дү	Улаанбаатар хот	Сонгинохайрхан дүүрэг	Мэргэжилтэн	Бүтэн цаг
35	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Автотерминал трейд	Засварчин ажилд авна	Автобус засвар үйлчилгээ хийх	Цаг баримтлах, Хувийн зохион байгуулалт сайтай байх		Улаанбаатар хот	Хан-Уул дүүрэг	None	Бүтэн цаг
36	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Жи Эс 25	Шөнийн цагаар ажиллах Түгээлтийн жол	Бүтээгдэхүүн хүрдан шуурхай, хэ	Харилцааны ур чадвартай Цагий	Үндсэн цалин + Хоолны мөнгө + У	Улаанбаатар хот		Мэргэжил хамаарахгүй	Бүтэн цаг
37	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Глобал Бридж Групп	Автын засварчин	Ажлын байрны тодорхойлогдолт	Бүрэн дунд ба түүнээс дээш Цэвэрлэв, сангийн болон тоног техөөр		Улаанбаатар хот	Баянхүрх дүүрэг	None	Бүтэн цаг
38	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Дрэйм интерграйз	Авто засварчин	Техник тоног төхөөрөмжийн засв	Тэтгэр суурьшлтай ажилтах Харилуулагатай нягт нямбай		Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
39	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Монкон Групп ХХК	Автын засварчин	Автомашин механизмийн хэвийн	Автын засварчин мэргэжлээр МС	Шинэ төгсөгч болсон сургэж авна	Улаанбаатар хот	Хан-Уул дүүрэг	Мэргэжилтэн	Бүтэн цаг
40	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Сабоца ХХК	Авто механик/засварчин/	Авто сервисийн үндсэн үйл ажи	Авто механикийн мэргэжилтэй ai	Өдрийн хоолтой	Улаанбаатар хот	Чингэлтэй дүүрэг	Мэргэжилтэн	Бүтэн цаг
41	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Сод Монгол Групп	Автын механик	-Компанийн тээврийн хэрэглүүг	- Хүнд даацын машины оношлог	Улаанбаатар хот БГД 20-р хороо үй	Улаанбаатар хот		Мэргэжилтэн	Бүтэн цаг
42	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Ти Ти Си энд Ти ХХК	Механик / Тэлевлэгч /ээлжийн хувараа	Хүнд даацын тээврийн хэрэгслий	Механик инженерийн их дээд сур	Бүрдүүлэх материал: Их дээд сург	Өмнөговь аймаг		Мэргэжилтэн	Эрэлжийн
43	Автомашины засвар үйлчилгээ	Авто засвар	https://www.zarg	Юнайтед Белаз Машинери	Үйлчилгээн-й хөнгөн тэрэгний механик (\	Улаанбаатар хотод, тев агуулахд	-механик инженер чиглэлээр их / явуулна уу		Улаанбаатар хот		Мэргэжилтэн	Бүтэн цаг

Зураг 5.2: Data set

Энд хамгийн сүүлд буюу 3 сарын 31нд өгөгдлийн цуглуулга хийж 9000 өгөгдлийн excel хэлбэрт оруулсныг харж болж байна.

[4] https://docs.google.com/spreadsheets/d/1rtATUKhUlleIKaWgFGvqiUWMipsrv-aCWZk-tYmzezU/edit?usp=sharing

Bibliography

[1] Чатбот системийн тухай
https://www.engati.com/blog/types-of-chatbots-and-their-applications

[2] Өгүүлбэр хувиргалтын арга зүй
https://www.sbert.net/docs/quickstart.html

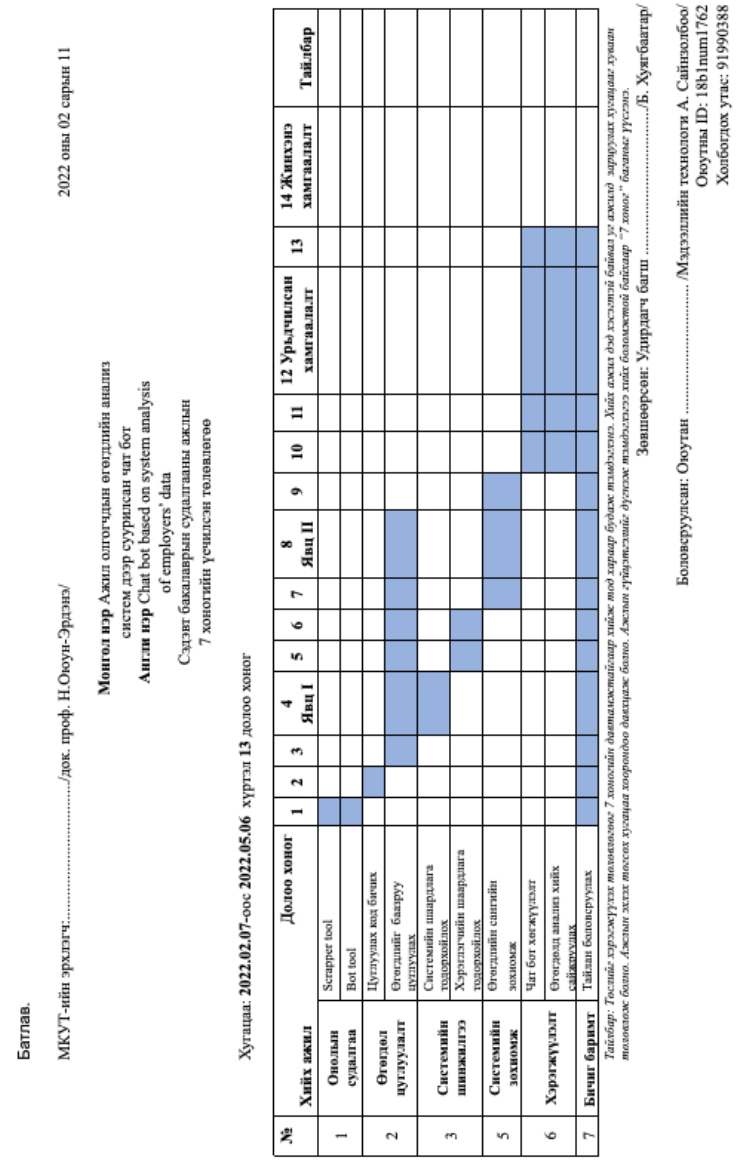
[3] Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks
https://arxiv.org/abs/1908.10084

[4] Use case diagram
https://app.diagrams.net/#G1jhom3sc_holt-X9XLALtQja_Gl_Eykhj

[5] Business Process Model Notation 2.0 диаграм
https://cawemo.com/diagrams/ea037ec0-c1c5-4ab6-8262-521657472803--bpmn-2-0?v=960,418,1

[6] Өгөгдлийн сангийн диаграм
https://dbdiagram.io/d/6249fb7cd043196e39e87451

А. ҮЕЧИЛСЭН ТӨЛӨВЛӨГӨӨ



Батлав.

МКУТ-ийн эрхлэгч:...../док. проф. Н.Оюун-Эрдэнэ/ 2022 оны 02 сарын 11

Монгол нэр Ажил олгогчын өгөгдлийн анализ
систем дээр суурилсан чат бот
Англи нэр Chat bot based on system analysis
of employers' data
Сэдэт бакалаврын судалгааны ажлын
7 хоногийн үечилсэн төлөвлөгөө

Хугацаа: **2022.02.07**-оос **2022.05.06** хүртэл **13** долоо хоног

№	Хийх ажил	Долоо хоног	1	2	3	4 Яац I	5	6	7	8 Яац II	9	10	11	12 Урьдчилсан хамгаалалт	13	14 Жинхэнэ хамгаалалт	Тайлбар
1	Онолын судалгаа	Scrapper tool															
		Bot tool															
2	Өгөгдөл цуглуулалт	Цуглуулах код бичих															
		Өгөгдлийг баяжуу цуглуулах															
3	Системийн шинжилгээ	Системийн шаардлага тодорхойлох															
		Хэрэглэгчийн шаардлага тодорхойлох															
5	Системийн зохиомж	Өгөгдлийн сангийн зохиомж															
		Чат бот хөгжүүлэлт															
6	Хэрэгжүүлэлт	Өгөгдөл анализ хийх сайжруулах															
7	Бичиг баримт	Тайлан боловсруулах															

Тайлбар: Тестийг хэрэгжүүлэх төлөвлөгөө 7 хоногийн давтамжтайгаар тийм тоо харгар будаж тэмдэглэнэ. Хийх ажил дуусгахаар байвал ус ажилд зарцуулах хугацаг хувиар төлөвлөж болно. Ажлын эхэл төгсөх хугацаа хоорондоо давхцаж болно. Ажлын гүйцэтгэлийг дүгнэж тэмдэглэсэн хийх боломжтой байхаар "7 хоног" баганас үүсгэнэ.

Зөвшөөрсөн: Удирдагч багш ...../Б. Хүягбаатар/

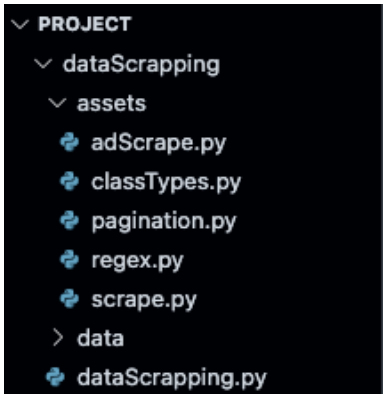
Боловсруулсан: Оюутан ...../Мэдээллийн технологи А. Сайнзолбоо/
Оюутны ID: 18b1num1762
Холбогдох утас: 91990388

Зураг А.1: Бакалаврын судалгааны ажлын үечилсэн төлөвлөгөө

В. КОДЫН ХЭРЭГЖҮҮЛЭЛТ

В.1 Өгөгдөл цуглуулалт

Өгөгдөл цуглуулах програм нь дараах бүтэцтэй байх бөгөөд assets доторх кодууд нь үндсэн кодыг ажлуулахад туслах функцууд байна.



Зураг В.1: Фолдерийн бүтэц

В.1.1 Үндсэн өгөгдлийг цуглуулах эх код

```
from datetime import date
import time
from assets.classTypes import Category
from assets.scrape import UseBeautifulSoup as useScrape
from assets.adScrape import advertisementScrape as useAdScrape
from assets.pagination import createLinkList as createLinkList

start_time = time.time()
initialUrl = 'https://www.zangia.mn/'
today = str(date.today())
# all categories set
categorySet = set()
# all advertisement's link set
adUrlDict = {}
# all ads object set
adsSet = set()

# scrape initial links
soup = useScrape(initialUrl)
navigatorList = soup.find_all('div', class_='filter')
for navigator in navigatorList:
    if navigator.find('h3').text.strip() != 'Salbar, mergejil':
        continue
    # ALL CATEGORY LINKS
    categoryList = navigator.find_all('div')
```

```
26
27 for categoryItem in categoryList:
28     categories = categoryItem.find('a')
29     url = initialUrl + categories['href']
30     tempCategory = Category(url, categories.text, '')
31     soup = useScrape(url)
32     subCategory = soup.find('div', class_='pros')
33     # ALL SUBCATEGORY LINKS
34     subCategoryList = subCategory.find_all('a')
35     for subCategoryItem in subCategoryList:
36         subCategoryUrl = initialUrl + subCategoryItem['href']
37         tempSubCategory = Category(
38             subCategoryUrl, subCategoryItem.text, tempCategory.name)
39         categorySet.add(tempSubCategory)
40
41 for categoryItem in categorySet:
42     if categoryItem.parentId == '':
43         continue
44     soup = useScrape(categoryItem.url)
45     hasPagination = soup.find('div', class_='page-link')
46     pagesUrl = []
47     if hasPagination != None:
48         pagesUrl = createLinkList(hasPagination, categoryItem.url)
49     else:
50         pagesUrl.append(categoryItem.url)
51     for pageUrl in pagesUrl:
52         soup = useScrape(pageUrl)
53         ads = soup.find_all('div', class_='ad')
54         # CREATE UNIQUE AD DICTIONARY
55         for ad in ads:
56             adUrl = initialUrl+ad.find('a', class_=None)['href']
57             adUrlDict[adUrl] = categoryItem
58     pagesUrl.clear()
59
60 file = open(today+'adScrape.csv', 'w', encoding='utf-8')
61 file.write('Parent Category Name' + '\t' +
62            'Category Name ' + '\t' +
63            'Link' + '\t' +
64            'Employee Company' + '\t' +
65            'Title' + '\t' +
66            'Roles' + '\t' +
67            'Requirements' + '\t' +
68            'Additional Info' + '\t' +
69            'City/Province' + '\t' +
70            'District' + '\t'
71            'Level' + '\t' +
72            'Type' + '\t' +
73            'Min Salary' + '\t' +
74            'Max Salary' + '\t' +
75            'Is Dealable' + '\t' +
76            'Address' + '\t' +
77            'Phone' + '\t' +
```

```
            'Fax' + '\t' +
            'Ad Added Date' + '\n')

for adUrl in adUrlDict:
    print(adUrl)
    try:
        tempAdItem = useAdScrape(adUrl)
        tempAdItem.setCategory(adUrlDict[adUrl])
        file.write(
            tempAdItem.category.parentId+'\t' +
            tempAdItem.category.name+'\t' +
            tempAdItem.url+'\t' +
            tempAdItem.company+'\t' +
            tempAdItem.title+'\t' +
            tempAdItem.roles+'\t' +
            tempAdItem.requirements+'\t' +
            tempAdItem.additionalInfo+'\t' +
            tempAdItem.city+'\t' +
            tempAdItem.district+'\t' +
            tempAdItem.level+'\t' +
            tempAdItem.type+'\t' +
            tempAdItem.minSalary+'\t' +
            tempAdItem.maxSalary+'\t' +
            tempAdItem.isDealable+'\t' +
            tempAdItem.address+'\t' +
            tempAdItem.phoneNumber+'\t' +
            tempAdItem.fax+'\t' +
            tempAdItem.adAddedDate+'\n')
        del tempAdItem
    except:
        print('Ad writing error')
file.close()
print("--- %s seconds ---" % (time.time() - start_time))
```

Код В.1: Бүх өгөгдлийг цуглуулах - dataScrapping.py

В.1.2 Нэг зарын шаардлагатай бүх мэдээллийг цуглуулах код

```
import re
from .classTypes import Advertisement
from .scrape import UseBeautifulSoup as useScrape


def listScraper(sections, key) -> str:
    content = []
    for section in sections:
        subTitle = section.find('h2', class_=None).text
        if key != subTitle:
            continue
        div = section.find('div', class_=None)
        children = div.next_element

```

```
        while(children != None):
            try:
                content.append(textStrip(children.text))
                children = children.next_sibling
                continue
            except:
                print('An error occured')
            children = children.next_sibling
        content = [s for s in filter(listFunc, content)]
    if not content:
        return ''
    return ' '.join(content)


def textStrip(text) -> str:
    pattern = re.compile('[\r\n\xa0\t ]+', re.MULTILINE | re.IGNORECASE
        )
    return pattern.sub(' ', text.strip())


def listFunc(e):
    return len(e) != 0


def singleItemScraper(sections, key, subKey) -> str:
    for section in sections:
        subTitle = section.find('h2', class_=None).text
        if key != subTitle:
            continue
        div = section.find_all('div', class_=None)
        for item in div:
            if item.next_element.text == subKey:
                return textStrip(item.find('span').text)
    return 'None'


def salaryScraper(salary):
    isDealable = ''
    k = re.split(r'[^\d,]+', salary, 2, re.IGNORECASE)
    if len(k) < 2:
        [a] = k[0:1]
        return a, a
    [a, b] = k[0:2]
    if len(k) > 2:
        isDealable = '     '
    return a, b, isDealable


def locationScrapper(location):
    city = ''
    district = ''
    k = location.split(',')
```

```
    if len(k) < 2:
        city = k[0]
        return city, district
    [city, district] = k[0:2]
    return city, district


def advertisementScrape(url) -> Advertisement:
    soup = useScrape(url)
    advertisement = Advertisement(url, soup.find('h3').text.strip())
    companyTitle = soup.find('div', class_='nlp').find('td')
    for item in companyTitle:
        try:
            if item.name == None:
                advertisement.company = textStrip(item.text)
        except:
            print('Company name scrape error')
    # advertisement.company = textStrip(company)

    # all items
    sections = soup.find_all('div', class_='section')
    advertisement.roles = listScraper(
        sections, 'Guitsetgeh undsen uurerg')
    advertisement.requirements = listScraper(
        sections, 'Ajliin bairnii shaardlaga')
    advertisement.additionalInfo = listScraper(
        sections, 'Nemelt medeelel')
    advertisement.level = singleItemScraper(sections, 'Busad', 'Tuvshin
        ')
    advertisement.type = singleItemScraper(sections, 'Busad', 'Turul')
    minSalary, maxSalary, isDealable = salaryScraper(
        singleItemScraper(sections, 'Busad', 'Tsalin'))
    city, district = locationScrapper(
        singleItemScraper(sections, 'Busad', 'Bairshil'))
    advertisement.minSalary = minSalary
    advertisement.maxSalary = maxSalary
    advertisement.isDealable = isDealable
    advertisement.city = city
    advertisement.district = district
    advertisement.address = singleItemScraper(sections, '      ', '  ')
    advertisement.phoneNumber = singleItemScraper(
        sections, 'Holboo barih', 'Utas')
    advertisement.fax = singleItemScraper(
        sections, 'Holboo barih', 'Fax')
    advertisement.adAddedDate = singleItemScraper(
        sections, 'Zariin hugatsaa', 'Zar niitelsen ognoo')
    print(advertisement.additionalInfo)
    print('SINGLE AD SCRAPPING DONE!!!', url)

    return advertisement
```

Код B.2: Нэг зарын өгөгдлийг цуглуулах - adScrape.py

B.1.3 Цуглуулах өгөгдлийн төрөл

```
class Category:
    url = ''
    name = ''
    parentId = ''

    def __init__(self, url, name, parentId='None') -> None:
        self.url = url
        self.name = name
        self.parentId = parentId

    def getUrl(self) -> str:
        return self.url


class Advertisement:
    category = Category
    url = ''
    company = ''
    title = ''
    # ListInfo
    roles = ''
    requirements = ''
    additionalInfo = ''
    # OtherInfo
    city = ''
    district = ''
    level = ''
    type = ''
    minSalary = ''
    maxSalary = ''
    isDealable = ''
    # ContactInfo
    address = ''
    phoneNumber = ''
    fax = ''
    adAddedDate = ''

    def __init__(self, url, title) -> None:
        self.url = url
        self.title = title

    def setCategory(self, category) -> None:
        self.category = category
```

Код B.3: Өгөгдлийн төрөл - classTypes.py

B.1.4 BeautifulSoup scraper

```
from bs4 import BeautifulSoup
import requests
```

```
3  from urllib.error import HTTPError
4
5
6  def UseBeautifulSoup(url):
7      try:
8          response = requests.get(url)
9          response.raise_for_status()
10     except HTTPError as error:
11         print(error)
12     soup = BeautifulSoup(response.text, 'html.parser')
13     return soup
```

Код B.4: Scrape хийх функц - scrape.py