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

Анужингийн Сайнзолбоо

АЖИЛ ОЛГОГЧДЫН ӨГӨГДӨЛД СУУРИЛСАН ЧАТ БОТ

(Chat bot based on employers' data)

Мэдээллийн технологи (D061303) Бакалаврын судалгааны ажил

Улаанбаатар

2022 он

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

АЖИЛ ОЛГОГЧДЫН ӨГӨГДӨЛД СУУРИЛСАН ЧАТ БОТ (Chat bot based on employers' data)

Мэдээллийн технологи (D061303) Бакалаврын судалгааны ажил

Удирдагч:	Б.Хуягбаатар доктор (Ph.D.)
Гүйцэтгэсэн:	А.Сайнзолбоо (18B1NUM1762)

Улаанбаатар

		ΓΑΡΥ	łИГ
УДИР	РТГАЈ	Ι	1
БҮЛГ	ΥΥД		2
1.	СЭД	ЦВИЙН ТАНИЛЦУУЛГА	2
	1.1	Оршил	2
	1.2	Зорилго	2
	1.3	Зорилт	2
	1.4	Алсын хараа	3
2.	СИС	СТЕМИЙН СУДАЛГАА	4
	2.1	Системийн судалгаа	4
	2.2	Ижил төстэй системүүд	7
	2.3	Ашиглах технологиуд	9
3.	СИС	СТЕМИЙН ШИНЖИЛГЭЭ	10
	3.1	Бизнесийн үйл ажиллагааны шинжилгээ	10
	3.2	Хэрэглэгч	11
	3.3	Функционал шаардлага	11
	3.4	Функционал бус шаардлага	12
	3.5	Use case диаграмм	13
4.	СИС	СТЕМИЙН ЗОХИОМЖ	14
	4.1	Өгөгдлийн сангийн диаграмм	14
	4.2	Өгөгдлийн элемент	15
	4.3	Өгөгдлийн сангийн холбоосын тайлбар	17
	4.4	Класс диаграм	18
5.	ХЭІ	РЭГЖҮҮЛЭЛТ, ҮР ДҮН	19
	5.1	Хөгжүүлсэн байдал	19
	5.2	Vn лvн	35

ДҮГНЭЛТ	41
ДҮГНЭЛТ	41
НОМ ЗҮЙ	42

ЗУРГИЙН ЖАГСААЛТ

2.1	Pizza Hut chat bot	7
2.2	WHO's chat bot	8
3.1	BPMN-1	10
3.2	BPMN-2	11
3.3	Use Case диаграмм	13
4.1	Өгөгдлийн сангийн диаграмм	14
4.2	Класс диаграм	18
5.1	Үндсэн процесс зураглал	20
5.2	Data set	21
5.3	Виртуал машин дээрх crontab	22
5.4	Өгөгдлийн статистик-1	23
5.5	Өгөгдлийн статистик-2	24
5.6	Өгөгдлийн статистик-3	25
5.7	Өгөгдлийн статистик-4	26
5.8	Өгөгдлийн статистик-5	27
5.9	Чатботын хариулж чадах асуултууд	28
5.10	Угтах тескт	35
5.11	<Байгууллага>-д <ажил> ажлын байрны мэдээлэл?	36
5.12	Бусад ажлын байрууд	37
5.13	<Байгууллага>-д нээлттэй ажлын байрууд?	38
5.14	<Ажиллах цаг> цагийн <ажил> ажлын байр?	39
5 15	<Папин> папинтай <ажил> ажпын байр?	40

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 сонирхлыг оновчтой 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 өгөгдөлд 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.

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 өгөгдөлд 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.

¹https://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 тодорхой дүрмийн дагуу боловсруулж 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 даалгавраа 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.

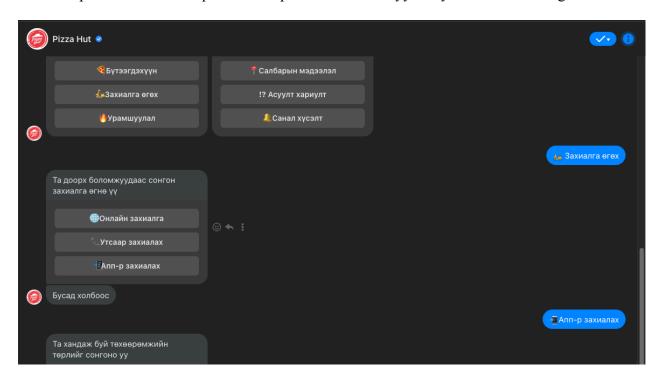
Ихэвчлэн энэхүү хоёр аргыг хослуулан ашигладаг бөгөөд шаардлагад нийцэх өгөгдлийг үлдээлгүй бүгдийг нь олоход data crawling-ийг ашиглах бол олсон өгөгдлийг задалж, шинжлэн өгөгдлийн санд хадгалах үйлдлийг data scraping хийдэг. Жишээлбэл, худалдааны сайтын бараа бүтээгдэхүүний өгөгдлийг цуглуулах гэж байгаа гэж үзвэл, барааны ангиллын хаягуудыг өөрчлөгдөх бүрд 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 СSV файл үүсгэн хадгалж цаашид ашигласан болно.

2.2 Ижил төстэй системүүд

Гадаад ба дотоодын байгууллагуудын үйл ажиллагаандаа хэрэгжүүлдэг чатбот системүүдээс, Domino's Pizza & Pizza Hut болон WHO's Chat bot гэсэн гурван чатботыг сонгон авч судалгаанд оруулав.

2.2.1 Domino's Pizza & Pizza Hut

Domino's pizza хоолны газар нь захиалгын алхмаас эхлээд бүх мэдээллийг ганцхан *Face-book 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.

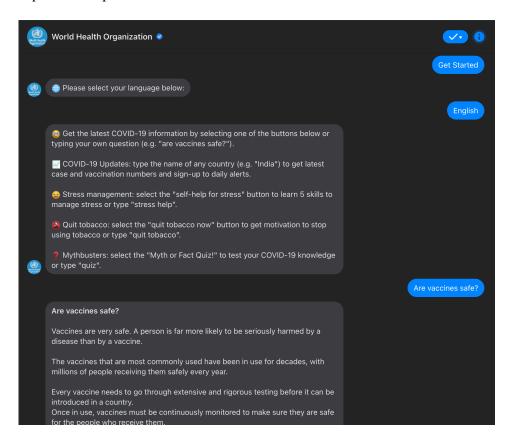


Зураг 2.1: Pizza Hut chat bot

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

2.2.2 World Health Orgazination's Chat bot

Цар тахал болох коронавирусын эрчимтэй тархаж байх үед дэлхийн өнцөг булан бүрд 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 зохицож 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 Ашиглах технологиуд

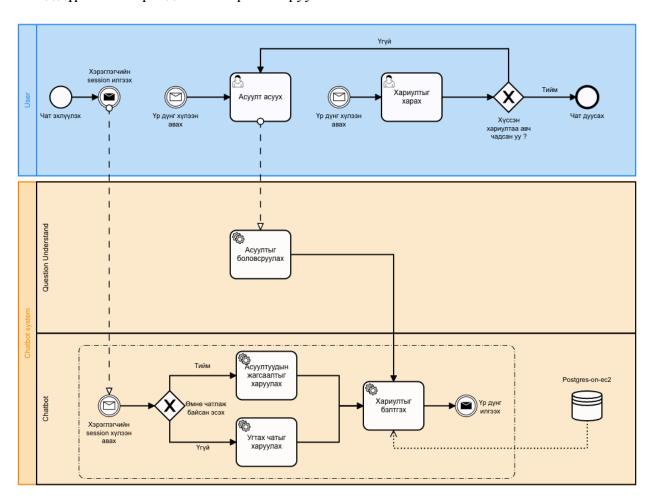
АЖИЛ ОЛГОГЧДЫН ӨГӨГДӨЛД СУУРИЛСАН ЧАТ БОТ-ыг хөгжүүлэхдээ өгөгдөл цуглуулгыг python хэлний сан болох BeautifulSoup HTML өгөгдөл задлах технологийг ашигласан бөгөөд чатбот системийн үндсэн хөгжүүлэлтийг Microsoft Bot Framework-ийн тусламжтайгаар javascript хэл дээр хөгжүүлэлтийг хийж гүйцэтгэсэн. Чатбот болон өгөгдлийн сантай холбогдох буюу back-end хөгжүүлэлтийг node.js-ийн framework болон express.js ашигласан. Харин цуглуулсан өгөгдлийг AWS-EC2 сүлжээнд байршуулсан бөгөөд дүн шинжилгээ хийхдээ CSV файлд хадгалан ашигласан.Ашигласан тенхологиудыг жагсаан бичвэл:

- BeautifulSoup
- Amazon-Web-Services
- PostgreSQL
- SQLAlchemy ORM
- Express.js
- Comma Separated Values CSV файл
- Microsoft Bot Framework

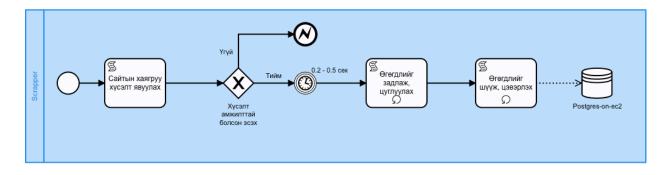
3. СИСТЕМИЙН ШИНЖИЛГЭЭ

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]. Диаграммд 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-1



Зураг 3.2: ВРММ-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 текст хэлбэрээр харагдана.
- ФШ 5 Чатботны хариулт нь текстэн хэлбэрээр хэрэглэгчид харагдана.
- ФШ 6 Чатбот нь зөвхөн Монголоор бичсэн асуултад хариулт өгнө.

ФШ 7 Чатбот нь дэлгэрэнгүй мэдээллийг карт хэлбэрээр харуулан эх сурвалжийн хаягийг мөн харуулна.

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)

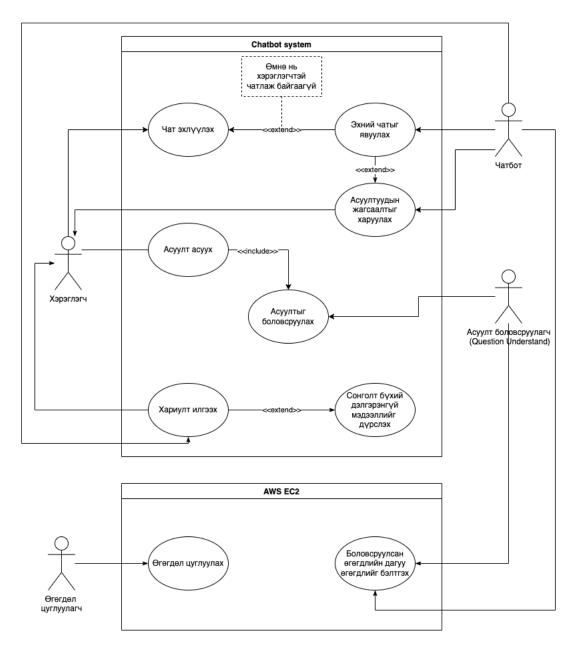
- ФБШ 06 Чатбот нь хэрэглэхэд хялбар, ойлгомжтой байна.
- ФБШ 07 Чатботны цэс нь ойлгомжтой цөөн үгээр илэрхийлэгдсэн байна.
- ФБШ 08 Чатботны цэсийн хэмжээ дарагдахуйц том байна.

Аюулгүй байдал(Security)

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

3.5 Use case диаграмм

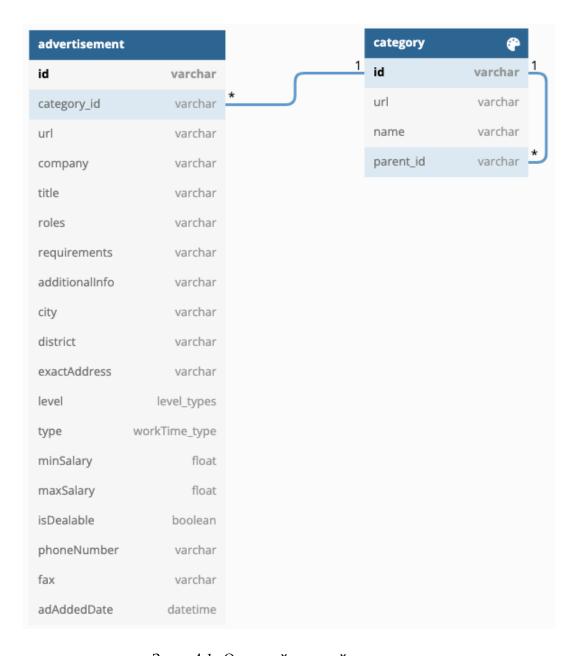
Чатбот системийн use-case диаграммыг байдлаар тодорхойлов [4].



Зураг 3.3: Use Case диаграмм

4. СИСТЕМИЙН ЗОХИОМЖ

4.1 Өгөгдлийн сангийн диаграмм



Зураг 4.1: Өгөгдлийн сангийн диаграмм

4.2 Өгөгдлийн элемент

Чатбот системийн өгөгдлийн сангийн диаграммд 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. Энд *level*

№	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 ангилалын дугаар
3	url		varchar	not null	Ажлын байрны зарын хаяг
4	company		varchar	not null	Ажил олгогч компани / хувь 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	city		varchar	null	Ажлын байрны харъяа хот
10	district		varchar	null	Ажлын байрны харъяа дүүрэг
11	exactAddress		varchar	null	Ажлын байрны бүтэн хаяг
12	level		level_types	null	Ажлын түвшин
13	type		workTime_type	null	Ажиллах цагийн төрөл

Table 4.1: advertisement хүснэгт

№	Баганын нэр	Түлхүүр	Өгөгдлийн	Хоосон	Тайлбар	
		өгөгдөл	төрөл	утга		
14	minSalary		float	null	Доод цалин	
15	maxSalary		float	null	Дээд цалин	
16	isDealable		boolean	null	Цалин тохиролцох эсэх	
17	phoneNumber		varchar	null	Холбоо барих утасны дугаар	
18	fax		varchar	null	Холбоо барих факсын дугаар	
19	publishedDate		datetime	not null	Зар нийтэлсэн огноо	

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
- topLevelManagemet Дээд 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
- seasonal Улирлаар

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

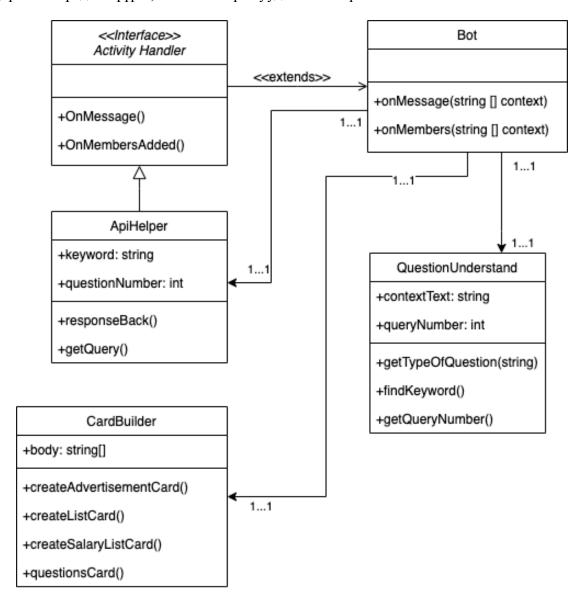
№	Баганын нэр	Түлхүүр	Өгөгдлийн	Хоосон	Тайлбар
		өгөгдөл	төрөл	Утга	
1	id	PK	varchar	not null	Ажлын байрны зарын
					ангиллын дугаар
2	url		varchar	not null	Ангиллын хаяг
3	name		varchar	not null	Ангиллын нэр
4	parent_id	FK	varchar	null	Эцэг ангиллын дугаар

4.3 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 категори байна.

4.4 Класс диаграм

Програмын класс диаграмд классууд болон тэдгээрийн хоорондын хамаарлыг дараах байдлаар тодорхойлж үндсэн үүрэг, ажиллах зарчмуудыг тайлбарлав.



Зураг 4.2: Класс диаграм

5. ХЭРЭГЖҮҮЛЭЛТ, ҮР ДҮН

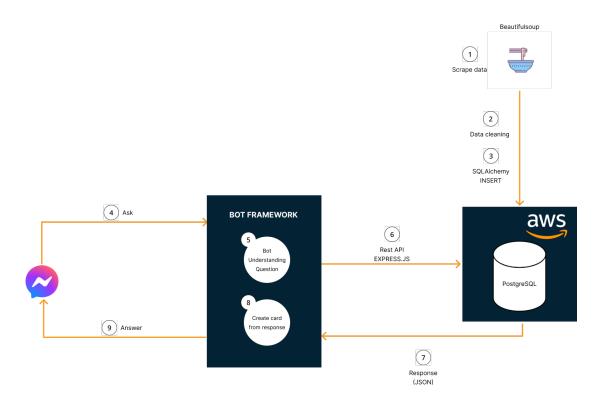
5.1 Хөгжүүлсэн байдал

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

- Өгөгдөл цуглуулах
- Өгөгдлийг нэгтгэх, цэвэрлэж өгөгдлийн санд хадгалах
- Системийн шаардлага, үйл ажиллагааг тодорхойлох
- Өгөгдөлд анализ хийх
- Чатбот хөгжүүлэх

гэсэн дарааллын дагуу хөгжүүлэлтийг хийсэн болно.

Доорх зурагт чатбот системийн үндсэн процессийн зураглал харагдаж байна.



Зураг 5.1: Үндсэн процесс зураглал

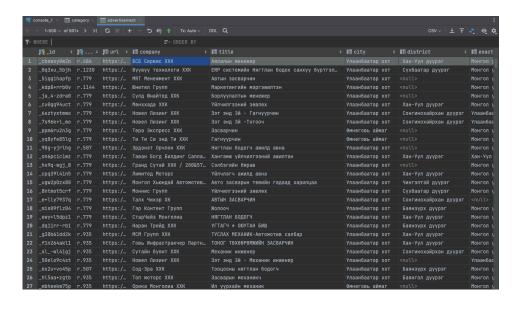
5.1.1 Өгөгдөл цуглуулах

Ундсэн 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.

Чатбот-д шаардлагатай мэдээллүүдийг HTML-ээс салган авсны дараа бүх өгөгдлөө AWS-EC2 сүлжээнд байршиж буй өгөгдлийн санд байршуулна. Ингэснээр чатбот нь бүрэн бүтэн ажиллах өгөгдлийг амжилттай цуглуулж байгаа хэрэг юм.

5.1.2 Ажлын байрны өгөгдөл

Доорх зурагт хамгийн сүүлд буюу 4 сарын 21нд өгөгдлийн цуглуулга хийж 8900 цуглуулж баазруу хийснийг dataGrip-ийн зургаас харж болж байна.



Зураг 5.2: Data set

Өгөгдөл цуглуулах нь цаг их шаардлагатай, тогтмол ажиллуулж байхад машиныг асаалттай ашиглаж байх шаардлага тулгарах бөгөөд энэхүү асуудлыг виртуал машин болох amazon-ec2-ээ ашиглаж linux машин дээр *crontab* ажиллуулж 7 хоног тутамд тогтмол өгөгдлийг цуглуулж өөр дээрх баазруу цуглуулах юм. Ингэхдээ эх кодоо github-аараа дамжуулан виртуал машин дээрээ ажиллуулж байгаа болно.

```
To define the time you can provide concrete values for minute (m), hour (h), day of month (dom), month (mon), and day of week (dow) or use '*' in these fields (for 'any').

# Notice that tasks will be started based on the cron's system daemon's notion of time and timezones.

# Output of the crontab jobs (including errors) is sent through email to the user the crontab file belongs to (unless redirected).

# For example, you can run a backup of all your user accounts at 5 a.m every week with:

# 0 5 * * 1 tar -zcf /var/backups/home.tgz /home/

# For more information see the manual pages of crontab(5) and cron(8)

# m h dom mon dow command

0 5 * * 1 python3 /home/ubuntu/bachelor/employementAnalysis/project/dataScrappi>

AG Get Help Owrite Out Ow Where Is Owe Cut Text Owe Cut Pos Owe Exit Owe Poste Cut Pos Owe Exit Owe Poste Cut Pos Owe Exit Owe Poste Cut Poste Owe Poste Owe Poste Cut Poste Owe Poste Cut Poste Owe Poste Owe Poste Owe Poste Owe Poste Cut Poste Owe P
```

Зураг 5.3: Виртуал машин дээрх crontab

5.1.3 Өгөгдлийг нэгтгэх, цэвэрлэх

Ихэнх цуглуулсан өгөгдөл нь өгөгдлийн сангийн диаграмын дагуу амжилттай цуглуулсан бөгөөд дүн шинжилгээ хийх боломжтой өгөгдлүүдийг тусад нь хадгалж ашигласан болно. Data scrape хийх явцад бүх өгөгдлийг *string* хэлбэрээр цуглуулсан бол өгөгдөлд дүн шинжилгээ хийх явцад энэ нь тохиромжгүй тул бүх өгөгдлийг эмхэтгэнэ.

5.1.4 АРІ ажиллуулах

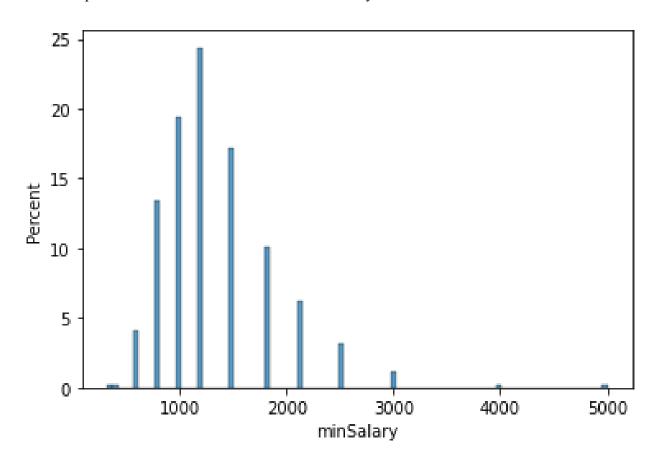
Чатбот системд шаардлагатай RestAPI-ийг express.js ашиглан хөгжүүлсэний дараа backend-ийг виртуал машин дээрээ ажиллуулсан. Ингэхдээ node.js -ийн сан болох *pm2*-ийг ашиглав.

5.1.5 Өгөгдлийн статистик

Өөрчлөлт хийсэн өгөгдөл буюу нийт 8202 зарын өгөгдөл дээрээ дүн шинжилгээ хийж, цаашид үүнийгээ чатбот хөгжүүлэлт, асуулт хариултын загварт тусгахыг зорив.

Статистик - 1

Энэхүү графикт нийт цалингийн давхардсан утгууд нийт өгөгдлийн хэдэн хувийг эзэлж байгааг харуулж байна. Үүнээс харвал дундаж цалин 1 сая төгрөгөөс 1.5 сая төгрөгийн хооронд байгааг харж болж байна. Энэ нь нийт цалингийн хувийн 50-аас 55-ийг эзэлж байна.

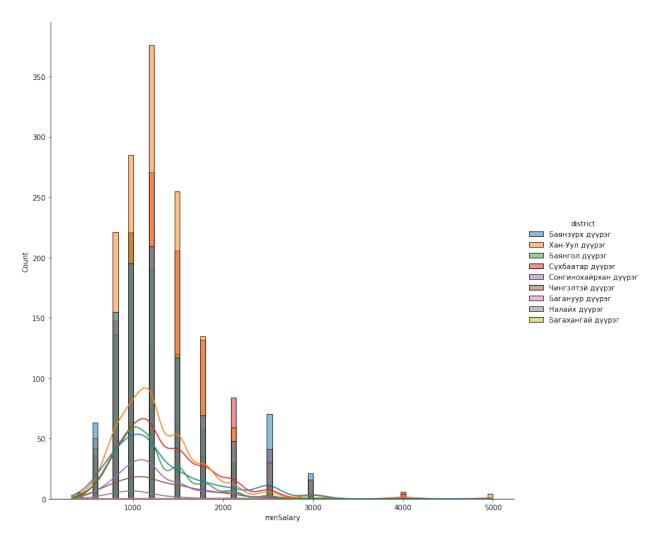


Зураг 5.4: Өгөгдлийн статистик-1

Статистик - 2

Доорх графикт нийт цалингийн давхардсан утгуудыг тоолж аль дүүрэгт хамгийн их байгааг өнгөөр нь дүрслэн харуулсан байна. Энэхүү графикаас харвал дундаж цалин буюу 1 сая

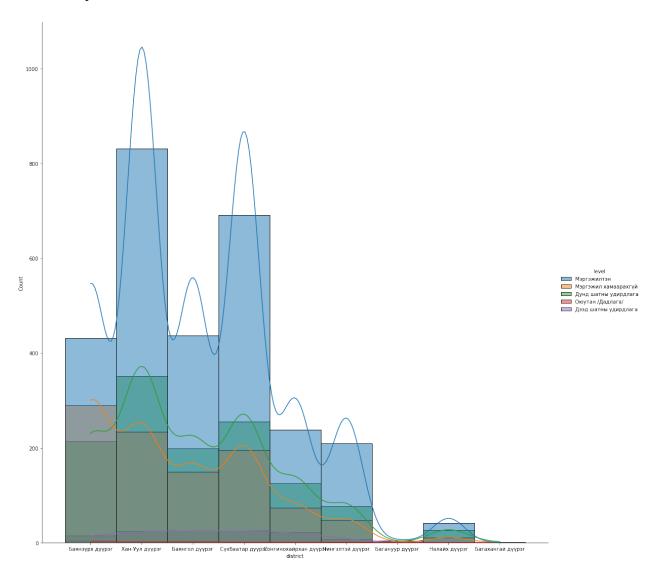
төгрөгөөс 1.5 сая төгрөгөөр цалинжуулах олон ажлын байр санал болгож байгаа дүүрэг нь Хан-Уул болон Сүхбаатар дүүрэг байна. Ажлын байрны төвлөрөл болон их хотын бүтээгдэхүүн үйлдвэрлэлийн цэгийг Хан-Уул, Сүхбаатар дүүрэг гэж дүгнэж болохоор байна.



Зураг 5.5: Өгөгдлийн статистик-2

Статистик - 3

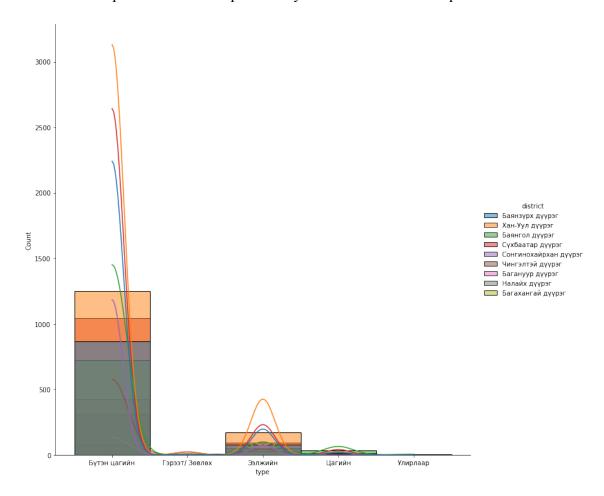
Өмнөх графикийн дүгнэлтийн адилаар хамгийн их ажлын санал болгож буй дүүрэг нь Хан-Уул, Сүхбаатар дүүрэг байх бөгөөд ажлын байрны шаардах түвшинг хамтад нь харуулсан байна. Үүнээс үзвэл, мэргэжилтэн болон дунд шатны удирдлагын орон тоо эрэлттэй байна гэж үзэж болно. Харин дадлагажигч болон дээд шатны удирдлагын эрэлт харьцангуй бага байгааг харж болж байна.



Зураг 5.6: Өгөгдлийн статистик-3

Статистик - 4

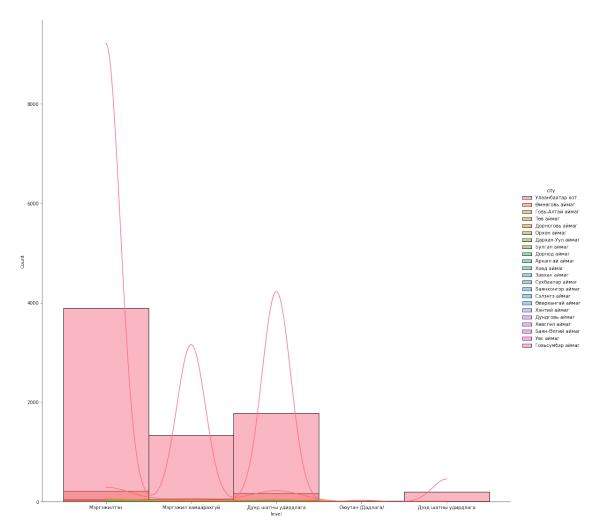
Энэхүү графикт дүүргүүд харгалзан ямар төрлийн цагийн хуваарьтай ажил санал болгож байгаа болон тэдгээрийн тоотой нь харьцуулан дүрсэлжээ. Эндээс бүтэн цагийн ажилтан болон ээлжийн төрлийн ажлын байр ихэнх хувийг эзэлж байгааг харлаа.



Зураг 5.7: Өгөгдлийн статистик-4

Статистик - 5

Доорх графикаас ажил олгогчид ямар төрлийн цагийн хуваарьтай, хаана ажил санал болгож байгааг 21 аймгаар бүсчлэн өнгөөр илэрхийлсэн байна. Үүнээс дүгнэвэл, Улаанбаатар хотод нягтаршил маш өндөр байгаа бөгөөд ажлын байрны эрэлт аймгуудтай харьцуулахад маш өндөр байна.



Зураг 5.8: Өгөгдлийн статистик-5

Дээрх статисткууд дээр үндэслэн, хэрэглэгчдэд хүртээмжтэй, нийтлэг асуултуудыг дараах байдлаар зохиомжлов.

Асуултын жагсаалт Чатбот нь доорх асуултуудад хариулж чадна. № Асуулт 1 Байгууллагын ажлын байрны дэлгэрэнгүй мэдээлэл. Жишээлбэл: Голомт банк-д дизайнер ажлын байрны мэдээлэл 2 Байгууллагын нээлттэй ажлын байруудын мэдээлэл. Жишээлбэл: Голомт банк-д нээлттэй ажлын байрууд 3 Бүх төрлийн цагийн хуваарьтай ажлын байрны дэлгэрэнгүй мэдээлэл Жишээлбэл: Бүтэн цагийн менежер ажлын байрууд 4 Цалингийн нөхцөлтэй ажлын байрны дэлгэрэнгүй мэдээлэл Жишээлбэл: 3000000 цалинтай хөгжүүлэгч ажлын байр

Зураг 5.9: Чатботын хариулж чадах асуултууд

5.1.6 Чатбот хөгжүүлэлт

Чатбот нь хэрэглэгчийн асуултаас онцлох түлхүүр үгшийг шүүж түүнд тохирох API-ын дагуу AWS дээр байрших өгөгдлийн санруу хүсэлт явуулна. API нь RESTful API бөгөөд express.js ажиллаж үр дүнд JSON өгөгдөл авна. Bot Framework нь тэрхүү өгөгдлийг *Adaptive-Card*-ын тусламжтайгаар хэрэглэгчдэд ойлгогдохуйц болгон харуулна. Үүний дараа хэрэглэгчид хариултыг буцаах зарчмаар чатбот систем нь ажиллах юм. Энэхүү процессыг алхам алхамаар хэрэгжүүлж тайлбарлая. Хэрэглэгч чатботтой холбогдох үед чатботын хариулж чадах асуултыг харуулна. Хэрэглэгч өөрийн шаардлагад нийцүүлэн асуултаас сонгож чатботоос асууна.

Эхлээд ашиглагдах REST API-г бэлдэж хэрэглэгчийн асуултад цаг алдалгүй хариулдаг байх шаардлагатай. Иймд express.js ашиглан PostgreSQL-ээс өгөгдлийг JSON-оор авах API бичиж өгсөн. АріНерler код дээр дурдсанчлан виртуал машин дээр ажиллах бөгөөд нь дараах байдалтай байна.

```
const categoryRoutes = require('./category/index')
const adRoutes = require('./advertisement/index')
```

```
app.use(express.json())

app.use('/api/v1/category', categoryRoutes)

app.use('/api/v1/ad', adRoutes)

app.listen(4000, () => {
    console.log('Server is listening in port 4000')
}
```

Код 5.1: REST API controller

Үүний дараа хэрэгэлгчийн асуусан асуултад хариулахад бэлэн болох бөгөөд дараах код нь хэрэглэгчийн асуултыг таньж query үүсгэхэд туслана. Асуултын төрлийг таньсаны дараа асуултаас шаардлагатай query-дэх түлхүүр үгийг олж авна.

```
getTypeOfQuestion(text) {
   if (this.contextText.search(' ') != -1) return 1
   else if (this.contextText.search(' ') != -1) return 2
   else if (this.contextText.search(' ') != -1) return 3
   else if (this.contextText.search(' ') != -1) return 4
   else return 404
}
```

Код 5.2: Question-understand классын кодын жишээ 1

```
findKeyWord() {
    switch (this.getTypeOfQuestion()) {
    case 1:
        this.queryNumber = 1
        var keyword = this.contextText.substring(0, this.contextText.
        search(' ')).trim().split('-')
```

```
return keyword
break
```

Код 5.3: Question-understand классын кодын жишээ 2

Ингэж query параметрийг түлхүүр үгийн тусламжтайгаар тодорхойлсны дараа API-аар хандаж өгөгдлийг авахад бэлэн болно. Үүний үр дүнд өгөгдлийн санруу явах query бэлэн болно.

```
getQuery() {
    switch (this.questionNumber) {
    case 1:
        return util.format('/ad/company=%s&title=%s', this.keyword[0].
            trim().replace(' ', '%20'), this.keyword[1].trim().replace(' ', '%20'))
        break
```

Код 5.4: query бэлтгэх кодын жишнээ

Дараах кодын хэсэг нь дотоод орчинд ажиллаж буй REST API controller-ыг ашиглан хүсэлт явуулж байна. Үр дүнд нь асуултын хариулт болох JSON объектыг авч байна.

```
async responseBack() {
   var query = this.getQuery()
   console.log('query:', query)

if (query === 404) {
   return undefined
}

var query = this.getQuery()

const host = 'http://3.228.127.116'

const port = ':4000'

const path = '/api/v1'
```

```
const url = host + port + path + query
       try {
         var res = await fetch(url)
13
      } catch (error) {
         console.error(error)
      }
16
      if (!res.ok) {
17
         throw res
      }
19
      var responseBody = await res.json()
20
      return responseBody
```

Код 5.5: JSON хүлээж авах кодын жишээ.

Тухайн үр дүнгээ хэрэглэгчид ойлгогдохоор харуулах шаардлагатай. Үүний тулд Microsoft Bot Framework-ийн картын төрлүүдийг ашигласан. Карт нь олон төрлийн хувилбаруудтай бөгөөд тэр дундаас "Adaptive Cards"-ыг сонгосон.

```
const d = new Date()
d.toLocaleDateString

for (let index = 0; index < this.body.length; index++) {
    var cardData = {
        "type": "AdaptiveCard",
        "schema": "http://adaptivecards.io/schemas/adaptive-card.json"
        ,
        "version": "1.2",
        "body": [],
        "actions": []
}
const element = this.body[index]</pre>
```

```
cardData["body"].push({
           "type": "TextBlock",
           "size": "Medium",
14
           "weight": "Bold",
15
           "text": element['title'],
           "wrap": true,
           "style": "heading"
18
         }, {
19
           "type": "ColumnSet",
20
           "columns": [{
21
             "type": "Column",
22
             "items": [
               {
24
                  "type": "Image",
25
                  "style": "Person",
26
                  "url": "https://cdn-icons-png.flaticon.com
                     /512/622/622848.png",
                  "altText": element['company'],
28
                  "size": "Small"
29
               }
             ],
             "width": "auto"
32
           }, , {
33
             "type": "Column",
34
             "items": [{
               "type": "TextBlock",
               "weight": "Bolder",
37
               "text": element['company'],
```

```
"wrap": true
             }, {
40
               "type": "TextBlock",
41
               "spacing": "None",
42
               "text": "
                               " + new Date(element['publishedDate']).
                  toLocaleDateString('zh-Hans-CN'),
               "isSubtle": true,
               "wrap": true
45
             }],
             "width": "stretch"
47
           }]
48
         }, {
           "type": "TextBlock",
50
           "text": element['roles'],
51
           "maxLines": 3,
52
           "wrap": true
53
         }, {
54
           "type": "FactSet",
55
           "facts": [{
56
             "title": " , :",
             "value": (element['city']) ? element['city'] : ' ' + ' ' + (
58
                element['district']) ? element['distirct'] : ' '
           }, {
59
             "title": " :",
             "value": (element['maxSalary'] || element['minSalary']) ?
61
                element['minSalary'] + ' - ' + element['maxSalary'] : '
           }, {
```

```
"title": " :",
             "value": element['types']
64
           }, {
65
             "title": " :",
             "value": element['phoneNumber']
           }]
         }
         )
         cardData['actions'].push({
           "type": "Action.OpenUrl",
72
           "title": "
73
           "url": element['url']
         })
         ret.push(CardFactory.adaptiveCard(cardData))
       }
       return ret
    };
```

Код 5.6: Хэрэглэгчийн UI дүрслэх кодын жишээ.

Хэрэглэгчид харагдах байдлыг өөрийн хүссэнээр уян хатан зохион байгуулах боломжийг "Adaptive Cards" олгодог бөгөөд эмх цэгцтэй мэдээллийг хүргэхэд туслах юм. AdaptiveCard нь JSON хэлбэртэй өгөгдлийн хадгалдаг бөгөөд үндсэн 5 төрлийн өгөгдөлтэй байна. Үүнд:

- type: Microsoft Botframework нь өөр олон төрлийн картуудтай бөгөөд энд тэдгээрийг зааж өгнө.
- schema: Энд картын эх буюу схемийн хаягийг тодорхойлно.
- version: Сонгосон төрлийн картын ямар хувилбарыг ашиглах болохыг зааж өгнө.
- body: Энд карт доторх бүх контент буюу мэдээллүүд байна.

• actions: Хэрэв карт нь ямар нэг төрлийн товчтой байвал тэдгээр нь юу хийх талаар програмчлана. (OnClick() г.м)

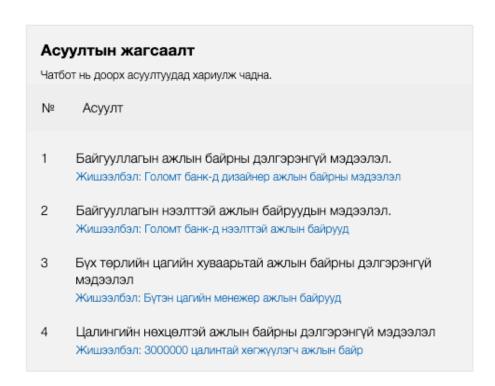
5.2 Ур дун

Энэхүү хэсэгт асуулт бүрийн хэрэгжүүлэлтийн хэрэглэгчдэд харагдах байдлыг дүрсэлсэн болно.

5.2.1 Угтах текст

Угтах текстийг шинэ хэрэглэгч болон чатлаагүй удсан хэрэглэгчдэд харуулна. Энэ нь хэрэглэгчийн session дээр суурилсан байна.

Сайн байна уу! Та ажлын байрны туслах чатботтой холбогдлоо. 🥮 Хүссэн ажлын байрны мэдээллийг цаг алдалгүй аваарай



Зураг 5.10: Угтах тескт

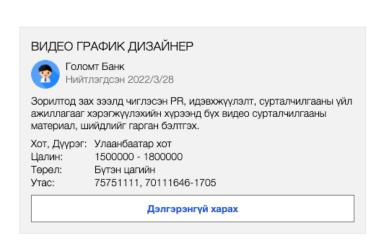
5.2.2 Байгууллагын ажлын байрны дэлгэрэнгүй мэдээлэл

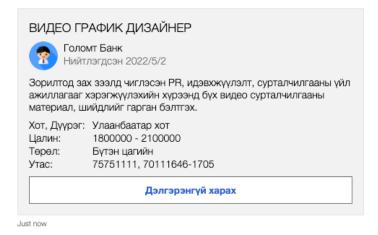
Ажиллахыг хүссэн байгууллын нэрийг ажлын байрны нэрийн хамтаар асуух нь хамгийн нийтлэг байх бөгөөд тухайн байгууллагын ажлын байрны нэрээр хайлт хийж хэрэглэгчдэд дэлгэрэнгүй мэдээллийг харуулна.

• Голомт Банк-д дизайнер ажлын байрны дэлгэрэнгүй мэдээлэл асуултын хариулт хэрэглэгчид харагдах байдал:

Голомт банк-д дизайнер ажлын байрны мэдээлэл

Just now

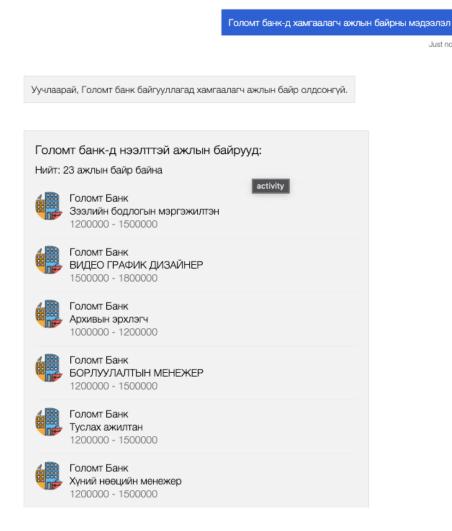




Зураг 5.11: <Байгууллага>-д <ажил> ажлын байрны мэдээлэл?

Хэрэв тухайн байгууллагад хайсан ажлын байрны нээлттэй зарын мэдээлэл байхгүй тохиолдолд

байгууллагын нээлттэй ажлын байруудын мэдээллийг дүрслэнэ.



Зураг 5.12: Бусад ажлын байрууд

Just now

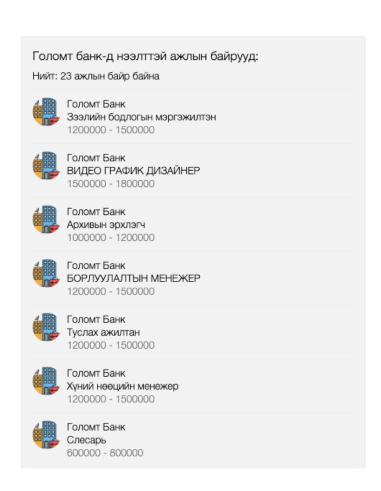
5.2.3 Байгууллагын нээлттэй ажлын байруудын мэдээлэл

Ажиллахыг хүссэн байгууллын нэрийн дагуу нээлттэй ажлын байруудыг харах боломжтой.

• Голомт Банк-д нээлттэй ажлын байруудын мэдээлэл асуултын хариулт хэрэглэгчдэд харагдах байдал:

Голомт банк-д нээлттэй ажлын байрууд

Just now



Зураг 5.13: <Байгууллага>-д нээлттэй ажлын байрууд?

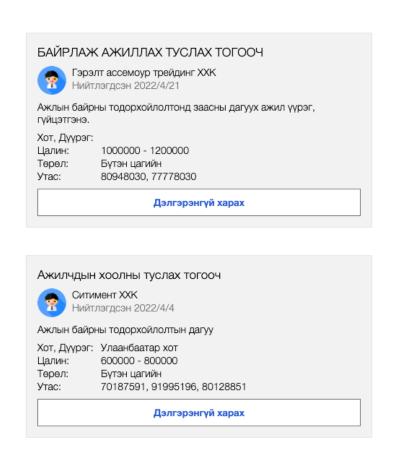
5.2.4 Бүх төрлийн цагийн хуваарьтай ажлын байрны дэлгэрэнгүй мэдээлэл

Ажиллахыг хүссэн ажлын байрны дагуу ажиллах цагийн хувиариар шүүж харах боломжтой байна. Бүх боломжит ажлын байруудын дэлгэрэнгүй мэдээллийг харуулна.

• Бүтэн цагийн тогооч ажлын байр асуултын хариулт хэрэглэгчдэд харагдах байдал:

Бүтэн цагийн тогооч ажлын байр

3 minutes ago



Зураг 5.14: <Ажиллах цаг> цагийн <ажил> ажлын байр?

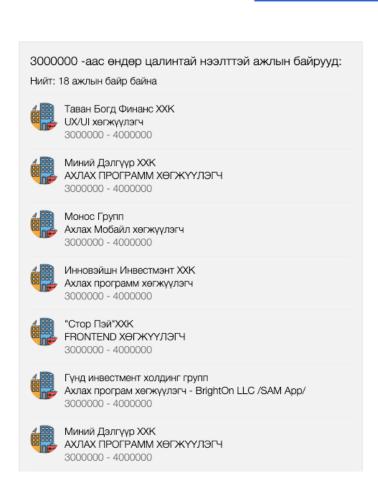
5.2.5 Цалингийн нөхцөлтэй ажлын байрны дэлгэрэнгүй мэдээлэл

Ажиллахыг хүссэн ажлын байрны дагуу цалингийн хэмжээгээр шүүж харах боломжтой байна. Бүх боломжит ажлын байрыг жагсаалт хэлбэрээр харуулна.

• Бүтэн цагийн тогооч ажлын байр асуултын хариулт хэрэглэгчдэд харагдах байдал:

3000000 цалинтай хөгжүүлэгч ажлын байр

Just now



Зураг 5.15: <Цалин> цалинтай <ажил> ажлын байр?

Дүгнэлт

Бакалаврын судалгааны ажлын хүрээнд ажил хайгчдын хэрэгцээнд нийцсэн, хялбар байдлаар ажил хайх боломжтой "Ажил олгогчдын өгөгдөлд суурилсан чатбот" -ыг хөгжүүлэхээр зорилго тавин зорилтуудын дагуу ажлуудыг хийж гүйцэтгэлээ.

Системийг хөгжүүлэхдээ уламжлалт холбоост өгөгдлийн сан(relational database) санг ашиглаж нийт өгөгдлийг zangia.mn-ээс цуглуулан ашиглав. Системийг Microsoft Azure платформын Bot Framework-д ашиглан хөгжүүлсэн бөгөөд хэрэглэгчдэд хүргэхэд олон төрлийн сувгийг ашиглах боломжийг олгон хөгжүүлэв.

Энэхүү програм нь хэрэглэгч ашиглаж эхэлснээс дуустал бүрэн хүний оролцоогүй бөгөөд өгөгдлүүдийг олборлох процесийг мөн автоматжуулав.

Цаашид хэрэглэхэд хялбар байдалд анхаарч, асуултад эх хэлний боловсруулалтын тусламжтайгаар хариулах, түгээмэл болон, шаардлагатай асуултуудыг хэрэглэгчийн туршлагад үндэслэн сайжруулалт, хөгжүүлэлт хийхэд анхаарал хандуулсаар байх болно.

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_G1_Eykhj
- [5] Business Process Model Notation 2.0 диаграмм https://cawemo.com/diagrams/ea037ec0-c1c5-4ab6-8262-521657472803--bpmn-2-0? v=960,418,1
- [6] Өгөгдлийн сангийн диаграмм https://dbdiagram.io/d/6249fb7cd043196e39e87451