

Мета Өгөгдлийг Ашиглан Өгөгдлийн Чанарыг Сайжруулах

Цогбадрахын Гүндэгмаа
Мэдээллийн системийн менежментийн тэнхим.
Санхүү эдийн засгийн дээд сургууль
Улаанбаатар, Монгол
gundegmaa.ts@ife.edu.mn

Хураангуй - Хүний амьдралын алхам бүрд том, жижиг асуудлууд тулгарч байдаг. Эдгээр асуудлыг хэрхэн даван гарснаар байгууллага, хувь хүний амжилт тодорхойлогддог. Мэдээллийн эрин гэгдэж буй өнөө цагт бодит баримт, мэдээлэл дээр тулгуурлан шийдвэр гаргахыг эрмэлзэж, өгөгдлийн санг бүрдүүлж байна. Харин тэрхүү өгөгдлийн санд цуглуулагдаж байгаа өгөгдлийн чанарын асуудал анхаарал татах асуудлуудын нэг болоод байна. Иймд өгөгдлийн чанарыг сайжруулах судалгаанууд тасралтгүй хийгдсээр байгаа юм.

Эдгээр судалгаагаар өгөгдөл оруулах формын загварчилгаа, өгөгдөл оруулах үеийн системийн хариу үйлдлийн зэрэг нь өгөгдлийн алдааг илрүүлэх, бууруулах боломжийг анхдагч хэлбэрээр олгодог гэдгийг батласан. Үүндээр тулгуурлан мета өгөгдлийг ашиглан алгоритмчласан кодоор програм хангамжийн эцсийн хэрэглэгчийн өгөгдөл оруулах үйл явцыг удирдаж, өгөгдлийн чанарыг сайжруулах боломжтой юм гэж үзсэн. Судалгааны үр дүн нь энэ хэлбэрээр өгөгдлийн чанарыг сайжруулах боломжтой гэдгийг харуулж байна.

Түлхүүр үгс—өгөгдлийн чанар, мета өгөгдөл, мета загвар

I. УДИРТГАЛ

Манай бизнесийн байгууллагууд үйл ажиллагааны мэдээллийн системүүдийг нэвтрүүлэхэд багагүй хугацаа өнгөрч байна. Үйл ажиллагааны мэдээллийн системүүдийг нэвтрүүлэхдээ байгууллагын үр ашгийг нэмэгдүүлэх, зардлыг хэмнэх, хөдөлмөрийг хөнгөвчлөх зорилготойгоор нэвтрүүлж байсан бол одоо эдгээр системүүдээс цуглуулсан мэдээлэл дээрээ тулгуурлан шийдвэр гаргалтыг дэмжихийг зорьж байна. Мэдээллийн систем нэвтрүүлсэн байгууллага бүр их, бага хэмжээний өгөгдлийн санг бүрдүүлсэн ч энэ мэдээлэл нь шийдвэр гаргалтыг дэмжихэд учир дутагдалтай байна. Үүнд өгөгдлийн загварчлал, өгөгдлийн чанарын асуудлууд нэн түрүүнд тулгарч байгаа юм.

Өгөгдлийн чанарын стандартыг тодорхойлсон ч түүнд хүрсэн өгөгдлийг цуглуулах нь асуудалтай. Өгөгдөлд чанарын асуудал байгаа гэдэг нь бага хэмжээний өгөгдлийн сангуудад л мэдэгдэж, их хэмжээний өгөгдлийн сангуудад алдаа тэр бүр

илэрдэггүй. Иймд чанартай өгөгдөл цуглуулах хэрэгцээ дэлхий дахинд үүсээд байна [4]. Гэвч байгууллагууд цаасан болон цахим анкетаар чанартай өгөгдөл цуглуулах дадлага дутмаг байгаа юм [5] [6].

Чанартай өгөгдлийг цуглуулахын тулд нэгэнт цуглуулсан өгөгдөл дээрээ ажиллах биш, хэрэглэгчийн өгөгдөл оруулах шатанд анхаарлаа хандуулах ёстой гэж судлаачид үзсэн байдаг [2]. Иймд өгөгдөл оруулах үйл явцад өгөгдлийн агуулгын мета өгөгдлийг хэрэглэгчид таниулж, түүгээр системийн интерфэйс, өгөгдлийн сангийн түвшинд давхар шалгаснаар өгөгдлийн чанарыг сайжруулах боломжтой гэж таамаглаж байна.

Чанартай өгөгдөл бол шийдвэр гаргалтыг дэмжээд зогсохгүй байгууллагын үнэт хөрөнгө юм.

A. Судалгааны зорилго

Мета өгөгдлийн хэрэглээг өргөжүүлэх замаар цуглуулж буй өгөгдлийн чанарыг сайжруулах аргазүйг тодорхойлж, санал зөвлөмж боловсруулах.

II. ОНОЛ АРГАЗҮЙ

A. Судлагдсан байдал

Байгууллага, удирдлагууд өгөгдлийн санд хадгалагдаж байгаа буруу ташаа мэдээлэл дээр үндэслэн чухал шийдвэр гаргах магадлал их байдаг. Энэ нь эрүүл мэнд гэх мэт хариуцлага өндөртэй салбаруудад маш их гарз, хохирлыг ч дагуулж болно. Өгөгдлийн алдаа нь өгөгдлийг цуглуулах, хадгалах, цэвэрлэх, нэгтгэх аль ч шатанд үүсч болно [1]. Гэхдээ өгөгдлийг цуглуулах үе шатуудад өгөгдлийн чанарын асуудлыг шийдэх боломжтой. Гэвч өгөгдлийн санг үүсгэсэн хойноо өгөгдлийг цэвэрлэх гэх мэт үйл ажиллагаанд их цаг зарцуулж, цуглуулах үед өгөгдлийн чанар харьцангуй бага анхаарал хандуулж байна [1], [2]. Өгөгдөл оруулах үед өгөгдлийн чанарыг сайжруулахын тулд алдаатай өрсөн өгөгдлийг дахин оруулах аргыг л ашиглаж байна [3].

Ихэнх байгууллагад өгөгдлийн чанарыг шалгах процедур огт байхгүй байгаа юм. Иймд цуглуулах үед л чанартай өгөгдөл оруулах хэрэгтэй [7].

Байгууллагууд ухаалаг мэдээлэл оруулах маягтыг (програм хангамжийн интерфэйс)

боловсруулан чанартай өгөгдлийг цуглуулах боломжтой болох нь харагдаж байна. Иймд маягтыг шинэчлэн боловсруулахдаа Usher загварыг ашиглах хэрэгтэй [8]. Ингэж өгөгдөл оруулах маягт болон формын интерфейсийг сайжруулсан ч динамик интерфейс дээр тулгуурлан системээс хариу үйлдэл үзүүлэн(feedback Жишээ нь:cache) өгөгдөл оруулах үйл ажиллагааг дэмжиж өгснөөр өгөгдлийн чанарыг сайжруулах боломжтой. Энэ нь их хэмжээний өгөгдөл оруулах үед, мобайлаар өгөгдөл оруулах үед ихээхэн үр дүнтэй юм. [9].

Харин миний энэхүү судалгаа нь [9] ажлыг суурь болгон шийдэл нэмж буй юм. Энэ шийдэл нь одоогоор зарим нэг өгөгдөл оруулах формын нууц үг талбар дээр хийгдсэн байгаа ч судалгааны түвшинд хийгдээгүй байгаа юм.

В. Мета өгөгдлийн тухай

Мета өгөгдлийг янз янзаар тодорхойлдог ч “Мета өгөгдөл гэдэг бол өгөгдлийн тухай өгөгдөл юм” гэсэн тодорхойлолтыг илүүтэй ашигладаг. Мета гэдэг нь грек гаралтай үг бөгөөд шинж чанараар нь жагсаах, эмх цэгцэнд оруулах гэсэн утгыг илэрхийлдэг.

Мета өгөгдөл нь дараах агуулгаар ашиглагддаг.

- Тодорхойлох (Агуулга, утгын олонлог, өгөгдлийн эзэмшигч, дүрмүүд болон тооцооллууд, энгийн утгууд)
- Чиглүүлэх (Зориулалт, хэрэглээ, эх сурвалж, эргэх холбоо)
- Удирдах (Ашиглах хугацаа, нөлөөлөл, хяналт шалгалт)

Дублин Корийн тодорхойлсноор мета өгөгдөл доорх 15 элементийг агуулдаг . [10]

- Title – зохиогчийн өгөгдөлд өгсөн нэр
- Creator- нөөцийн агуулгыг хариуцдагч
- Subject- нөөцийн сэдэв, агуулгын түлхүүр үгнүүд, тодорхойлолтын бүлэглэл.
- Description- нөөцийн агуулгын тодорхойлолт
- Publisher- нөөцийг одоогийн хэлбэрт оруулагч
- Contributorts- зохиогчийн дараах бичлэгийг хариуцдагч
- Date- нөөцийг одоогийн хэлбэрт оруулсан огноо.
- Type- нөөцийн төрөл
- Format- нөөцийг илэрхийлж буй хэл. Жишээлбэл text/html гэх мэт
- Identifier- нөөцийг бусдаас ялгагч шинж.
- Source- эх сурвалж.

- Language- нөөцийг агууллагын хэл
- Relation- бусад нөөцтэй холбогдож буй холбоо
- Coverage- нөөцийн шинж чанар цаг хугацааны болон байршлын хувьд.
- Rights management- зохиогчийн эрхтэй бичлэг

Эдгээр элементүүдээс миний судалгаанд голчлон Description буюу нөөцийн агуулгын тодорхойлолтыг ашиглах юм.

С. Өгөгдлийн чанарын тухай

Өгөгдлийн чанарыг баталгаажуулахдаа тухайн өгөгдөл тодорхой хэмжээний чанарын шаардлагуудыг хангасан байх ёстой. Чанар өндөр байх тусам тухайн мэдээллийн ерөнхий байдлаа хадгалах боломж тэр хэмжээгээр өндөр байна.

Мэдээллийн чанар нь хэрэглэгчдэд хүрэх мэдээллийн чанарыг хэмжих хэмжүүр юм. "Чанар" нь тухайн мэдээллийг түгээгч болон хүлээн авагчдаас шалтгаалан янз бүр байдаг. Чанар нь тухайн мэдээллийн бодит байдлыг нэмэгдүүлж өгдөг.

Wang болон Strong нь Мэдээллийн чанарын хэмжүүрийг дараах байдлаар тодорхойлсон байна.

- Бодит өгөгдлийн чанар: Нарийвчлал, Бодит байдал, Үнэмшил, Нэр төр
- Хам өгөгдлийн чанар: Хамаарал, Нэмүү өртөг шингэсэн, Цаг үеэ олсон байдал, Гүйцэд, Мэдээллийн тоо хэмжээ
- Төлөөлөрсөн өгөгдлийн чанар: Тайлбар, Загвар, Уялдаа холбоо, Зохицол
- Хүртээмжтэй өгөгдлийн чанар: Хүртээмж, Хамгаалалт

Өөр өөр мэргэжилтнүүд ижил төрлийн гэхдээ арай ялгаатай мэдээллийн чанарын хэмжүүрүүдийг санал болгож байдаг. Larry English нь "шинж чанар" гэсэн тодорхойлолтыг авч үзсэн. Мэдээлэл бүр өөрийн гэсэн тодорхойлолттой байдаг боловч тухайн "үйл явдлын тодорхойлолт" бүхий дээрхтэй ерөнхий нэг ойлголт байдаг.

Тодорхой бус цагт, тодорхой бус үйл явдлыг дүрслэхэд элдэв нэмэлт хэтрүүлэг, чимэг нэмэгддэг учраас мэдээллийн тархацын бодит байдлын чанарыг тодорхойлоход хүндрэлтэй байдаг. Мэргэшсэн мэргэжлийн холбоод мэдээллийн чанарыг тодорхойлсон нэг ойлголттой болох хэрэгтэй.

Энэ судалгаандаа дээр дурьдсан Wang болон Strong –ны тодорхойлсон шинж чанаруудыг бүрэн хангасан бол чанартай өгөгдөл гэж үзье.

D. Өгөгдлийн чанарын удирдлагын тухай

Өгөгдлийн чанарын удирдлага (DQM) нь хүчин төгөлдөр хэрэгтэй агуулга зөв хэлбэрээр урсаж байгаа эсэхийг шалгаж байгаа үйл ажиллагаа юм.

Ихэнх байгууллагуудад, нэгэн зэрэг олон мэдээлэл түгээх нь шийдвэр гаргах үйл явцад нөлөөлдөг. Ашиглах боломжгүй маш олон мэдээллийг хүлээж авах нь бизнест хүндрэл учруулдаг. Бизнесийн зорилгод хүрэхийн тулд мэдээллийн чанарын системчилсэн стратегийг ашиглах хэрэгтэй. Бизнесийн үйл явц, стратегийн дээд түвшинд хүрэх шилдэг хандлага, технологийг агуулдаг. Үүнийг мэдээллийн чанарын удирдлага гэдэг бөгөөд үр дүнд хүрэх арга замыг зааж өгдөг. Энэхүү программыг ашигласнаар бизнесийн шаардлагатай мэдээллийг байгууллагын дотоод болон гадаад хэрэглэгчдэд хүргэх боломжтой юм.

Мэдээллийн чанарын удирдлагыг амжилттай хэрэгжүүлсэнээр мэдээллийн менежмент болон мэдлэгийн менежментийг амжилттай хэрэгжүүлэх чухал ач холбогдолтой юм.

E. Өгөгдлийн чанарыг сайжруулах аргазүй

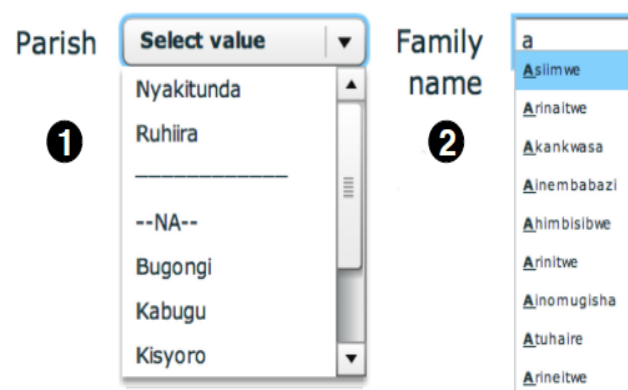
Одоогоор чанаргүй өгөгдлүүдийг задлан шинжлэх болон засварлах хэрэгсэл бүтээгдэж, тэгээрийг ашиглан өгөгдлийг цэгцэлж боловсруулж, цэвэрлэж байна. Жишээ нь: ETL системүүд.

Өгөгдөл сайжруулах хэрэгслүүд нийтлэг дараах функцуудыг заримыг эсвэл бүгдийг агуулсан байдаг. Үүнд:

- Өгөгдөл тодорхойлох- эхний ээлжинд тохиолдож болох асуудлыг өгөгдөлд таниулах
- Өгөгдлийг стандартжуулах- өгөгдөл нь өгөгдлийн журмыг дагаж байгаа эсэхийг баталгаажуулах бизнесийн дүрмийн хэсэг
- Геокодчиллол- нэр болон өгөгдлийг хаяглах. АНУ болон дэлхий дахины шуудангийн стандартын дагуу өгөгдлийг засварладаг.
- Тохируулах, холбох- өгөгдлийг хооронд нь харьцуулж, ойролцоонуудыг нь хооронд нь нэгтгэдэг.
- Хянах - өгөгдлийг хугацааны хувьд хянасны эцэст боломжит хувилбаруудыг буцааж мэдээллэх. Урьдах бизнесийн дүрмийн дагуу програм хангамж хувилбаруудыг автоматаар засварлаж чадна.
- Бодит хугацаа –нэгэнт өгөгдөл боловсруулагдаж дууссаны дараа (batch), компаниуд өгөгдлийг чанарыг алдагдуулахгүйн тулд боловсруулж, байгууллагын хэрэглээ болгохыг хүсдэг.

Харин [8],[9]-д өгөгдлийн чанарыг сайжруулах илүү үр дүнтэй аргачлалыг гаргаж өгсөн. Энэ нь өгөгдлийн чанарыг хангахуйц динамик

хэрэглэгчийн интерфейс үүсгэж өгөгдөл оруулах формд ашиглах, мөн түүндээ системийн хариу үйлдлийг Зураг 1-т үзүүлснээр кодчилсон юм.



Зураг 1. [9] судалгаанд дурдсан системийн хариу үйлдлийн аргачлал

Энэ арга, өгөгдлийн чанар, хүртээмжийг сайжруулах зорилуулалтаар шинэ технологиудад өргөн хэрэглэж байгаа юм.

III. СУДАЛГАА

Мета өгөгдөл ашиглан өгөгдлийн чанарыг сайжруулах аргачлалыг СЭЗДС-ийн нэгдсэн мэдээллийн системийн өгөгдөл оруулах зарим формүүд дээр туршиж үзлээ.

A. Мета өгөгдөл ашиглах аргачлал

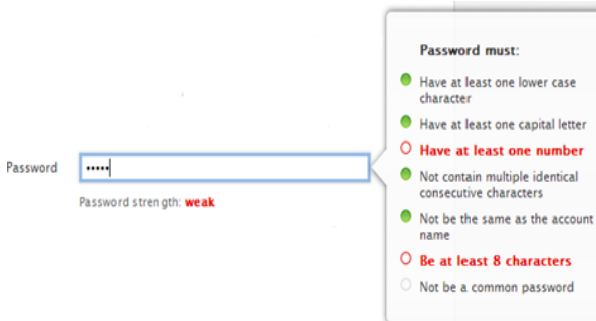
Өгөгдлийн нөөцийн агуулгын тодорхойлолтын мета өгөгдлийг хэрэглэгчид ойлгуулснаар зөв өгөгдөл оруулах боломжийг бүрдүүлнэ гэсэн үндсэн логикуыг баримтлах юм. Үүнийг алхамаар тайлбарлавал:

- Алхам 1. Хэрэглэгчийн интерфейст өгөгдлийн агуулгын мета өгөгдлийг өгөгдөл оруулах талбарын тайлбар хэлбэрээр харуулах (Зураг 2.)



Зураг 2. Мета өгөгдлийг тайлбар хэлбэрээр харуулж буй байдал

- Алхам 2. Өгөгдөл оруулж эхлэмэгц мета өгөгдлийн дагуу шалгах
- Алахам 3. Хэрэглэгчийн интерфейст өгөгдлийн зөв, бүрдсэн эсэхийг харуулах



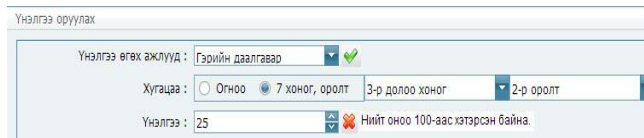
Зураг 3. Мета өгөгдлийн дагуу шалгаад, өгөгдлийн бүрэн, зөв эсэхийг харуулж буй байдал

Зураг 4-т үзүүлсэн үйл явцын дагуу алгоритмчлана.



Зураг 4. Мета өгөгдөл ашиглан хэрэглэгчийн интерфейсийг загварчлах үйл явц

Дээрх процессыг хэрэгжүүлсэн өгөгдөл оруулах формыг Зураг 5-аар харуулав.



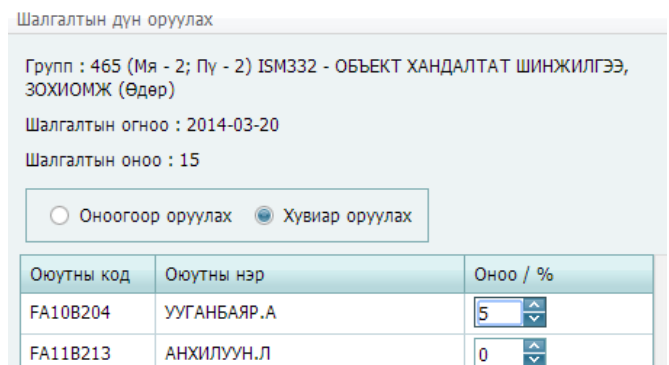
Зураг 5. Мета өгөгдөл ашиглан өгөгдөл оруулах алхам бүрийг шалгаж, хэрэглэгчийн интерфейст харуулсан форм

Ингэж өгөгдөл оруулах формын алхам бүрт өгөгдлийг шалгаснаар хэрэглэгчийн буруу өгөгдөл оруулах магадлалыг бууруулж, чанартай өгөгдөл цуглуулна гэж таамаглал дэвшүүлсэн. Гэхдээ үүнд өгөгдлийн агуулга нь тодорхой дүрэмтэй, алдаатай орох магадлалтай талбаруудыг хамруулж авч үзнэ. Бүх талбарыг үүнд хамруулбал програмын

ажиллагааны хурд саарах, кодчилол ихсэх зэрэг хүндрэл үүснэ.

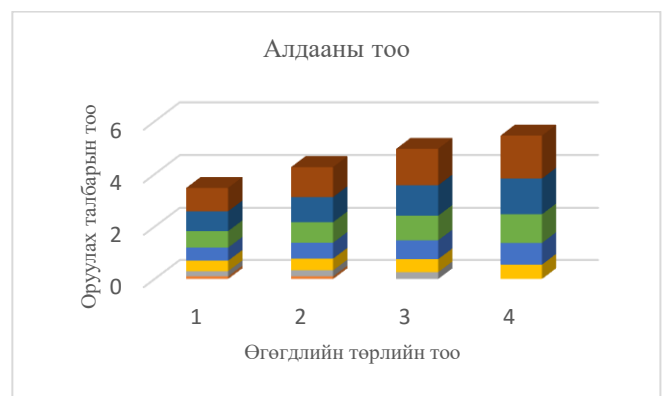
В. Судалгааны гүйцэтгэл

Энэ аргыг кодчилолдоо ашиглахаас өмнө СЭЗДС-ийн дүнгийн бодит тоон өгөгдөл оруулах 156435 тохиолдлоос 2354 тохиолдолд алдаатай өгөгдөл оруулж зассан, ирцийг бүртгэх бит төрлийн өгөгдлөөс 564786-ийн 1653 тохиолдлыг буруу оруулж зассан байна. Үүнээс 100 дүнгийн өгөгдлийн алдааг түүвэрлэн шалтгааныг тодруулахад 43 нь үйл ажиллагааны засвар, 57 нь механик алдаа байсан юм. Механик алдааны ихэнх нь хувь сонголтыг сонгочихоод, оноо оруулсантай холбогдон гарсан байсан. Уг дүн оруулах формыг Зураг 6-т харуулав.



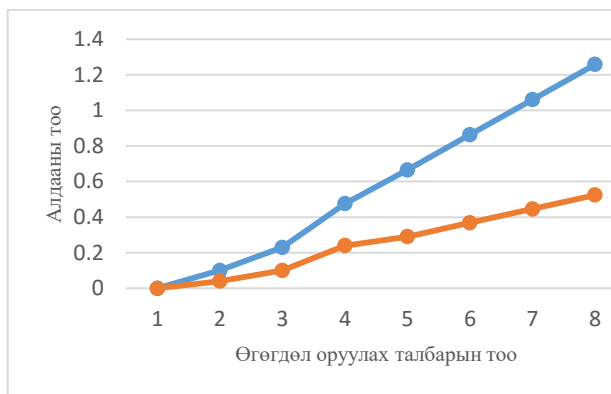
Зураг 6. Аргачлал ашиглаагүй энгийн форм

Энэ маягаар өгөгдөл оруулах формын өгөгдөл оруулах талбарын тоо, оруулах өгөгдлийн төрлийн тоо нэмэгдэх тутамд алдааны магадлал шугаман өсч байгаа нь өгөгдлийн сангийн өгөгдөлд хийсэн судалгаанаас Зураг 7-д үзүүлснээр нэмэгдэж байсан юм.



Зураг 7. Энгийн үед гарч байгаа алдааны тоог янз бүрийн формд харуулсан график

Туршилтаар МСМТ-ийн 5 багшид 14 хоногийн хугацаанд мета өгөгдөл ашиглан тайлбар оруулж, өгөгдөл оруулах алхам бүрийг шалгасан формуудыг ашиглуулахад алдааны магадлал 43%-иар буурсан үзүүлэлт гарсан. Үүнийг өмнөх алдааны тоотой жишиж Зураг 8-д үзүүлэв.



Зураг 8. Энгийн форм болон аргачлал ашиглан програмчласан формд оруулсан өгөгдлийн алдааны тооны харьцуулалт

С. Судалгааны тайлбар

Судалгаа хэт ерөнхий хийгдсэн нь анхан шатны судалгаа хийгдэж байгаатай холбоотой. Онолын хэсэгт гарсан мета өгөгдлүүдийн элементүүдийг бүрэн, зохих хэмжээнд ашиглан өгөгдлийн төрөл бүр дээр шалгах алгоритмыг тодорхойлж болох юм. Гэвч хэдий ингэж тодорхойлсон ч утгын хувьд шалгагдах боломжгүй тохиолдлууд маш олон гэдгийг хүлээн зөвшөөрч байна. Судалгаа нь өгөгдлийн оруулах үеийн алдааны магадлалыг бууруулахад үр дүнтэй гэж үзэж байна.

IV. ДҮГНЭЛТ

Энэ аргачлал нь практикт хэрэглэгдэх боломжтой ба өгөгдлийн чанарыг сайжруулах боломжтой юм. Үүнийг нэвтрүүлснээр програм хангамжийн эцсийн хэрэглэгчийн өгөгдөл оруулах хариуцлагыг чангатгах боломж үүсгэж болох юм.

Эцэст нь энэ аргачлалыг СЭЗДС-ийн нэгдсэн систем дээр хэрэгжүүлэх төлөвлөгөөтэй байна.

V. ЦААШДЫН СУДАЛГАА

Дараагийн удаад энэхүү судалгааг өргөжүүлж, өгөгдлийн төрлүүд дээрх зохицсон шалгах алгоритм, мөн энэ аргачлал нь онлайн орчинд хэрхэн зохицох гэх мэт олон зүйлсийг тодорхойлох шаардлагатай.

ТАЛАРХАЛ

Судалгаанд идэвхитэй оролцсон СЭЗДС-ийн багш, оюутнууддаа болон судалгааг явуулах дэд бүтцээр хангаж өгсөн СЭЗДС-ийн Мэдээллийн системийн албаны хамт олондоо талархал илэрхийлье.

НОМ ЗҮЙ

[1] T. Dasu and T. Johnson, "Exploratory Data Mining and Data Cleaning". Wiley Series in Probability and Statistics, 2003.

[2] C. Batini and M. Scannapieco, Data Quality: Concepts, Methodologies and Techniques. Springer, 2006.

[3] R. M. Graves, F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau, Survey Methodology. Wiley-Interscience, 2004.

[4] J. Lee. (2003) Address to WHO staff. [Online]. Available: http://www.who.int/dg/lee/speeches/2003/21_07/en

[5] J. V. D. Broeck, M. Mackay, N. Mpontshane, A. K. K. Luabeya, M. Chhagan, and M. L. Bennish, "Maintaining data integrity in a rural clinical trial," Controlled Clinical Trials, 2007.

[6] R. McCarthy and T. Piazza, "Personal interview," University of California at Berkeley Survey Research Center, 2009.

[7] S. Patnaik, E. Brunskill, and W. Thies, "Evaluating the accuracy of data collection on mobile phones: A study of forms, sms, and voice," in *ICTD*, 2009.

[8] K.Chen, H.Chen, N.Conway, J.M.Hellerstein,T.S.Parikh, "Usher:Improving data quality with dynamic form" University of California at Berkeley, 2010

[9] K.Chen, J.M.Hellerstein,T.S.Parikh, "Designing adaptive feedback for improving data entry accuracy" University of California at Berkeley, 2010

[10] "Understanding metadata" National information standards organization, USA, 2004