

Үгзүйн орчин үеийн хандлага, түүнийг Монгол хэлэнд ашиглах боломжийн судалгаа

Б.Гүндсамбуу*, Б.Удвалцэцэг**, Б.Батзолбоо***

Компьютерийн ухааны салбар

ШУТИС, Мэдээлэл, Холбооны Технологийн Сургууль

{*gundsambuu, **udvaltsetseg, ***b.batzolboo}@must.edu.mn

Хураангуй— Энэ судалгааны ажлын хүрээнд үгзүйн орчин үеийн хандлагуудыг судалж, тэдгээрийг монгол хэлний үгзүйд хэрэгжүүлэх боломжийг эрэлхийлсэн юм. Монгол хэлний бүтэцтэй төстэй гэж үзэгддэг Турк хэлэнд туршигдсан байдлыг үндэслэн монгол хэлний үгзүйн онцлог, шинж чанаруудыг харгалзан удирдамжтай (supervised) болон удирдамжгүйгээр (unsupervised) сургах аргуудын гол төлөөлөл болсон Яровский ба Висентовский, Патрик Шоне ба Даниел Журавский нарын аргуудыг сонгож, хэрэгжүүлэв. Судалгааны ажлын үр дүнд боловсруулаагүй текст корпус (буюу хөмрөг) -ийг оруулахад үгийн үндсэнд залгах боломжтой залгавруудын олонлогийг гаргаж харуулахыг зорилло.

Тулхуур үгс— Үгзүй, хувирдаг үгзүй, удирдамжтай сургах арга, удирдамжгүй сургах арга, залгавруудын олонлог

I. ОРШИЛ

Дэлхий дээр бие даасан, тусгаар улс орон, ард түмэн оршоор байх цагт эх хэл бичгийн асуудал байсаар байх нь гарцаагүй зүйл билээ. Харин тухайн улсын болон хэлний том, жижиг зэрэг хүчин зүйлсээс шалтгаалж өөр өөрийн онцлогт тохирсон хэлний бодлоготой байдаг байна. “Монголын ард түмний түүх, соёлын дурсгалт зүйл, шинжлэх ухаан, оюуны өв төрийн хамгаалалтад байна” [1] гэж үндсэн хуульд бий.

Мэдээллийг автоматаар боловсруулах, хэрэглэхийн ач холбогдол нь мэдээллийн эрин зуунд шилжиж буй өнөө үед улам их өссөөр байна. Үүний тулд монгол хэлийг компьютерээр боловсруулах чиглэлд олон судалгаа хийх шаардлагатай бөгөөд одоогийн байдлаар манай улсын их дээд сургууль болон судалгаа шинжилгээний байгууллагуудад хийгдсэн ажлын төвшин нь автомат орчуулгын системийг бүтнээр нь хийхэд хүрэхгүй ч үгзүйн зарим төвшинд хангалттай хийгдсэн гэж үзэж байна [2].

Автоматаар үгзүйн шинжилгээ хийх анхны ажлууд 1960 -аад оны үеэс машин орчуулгын системд хийгдэж байв [3]. Эдгээрийн дийлэнх нь хэлний мэдээллийг хатуу кодоор оруулан бичиж, хэтэрхий энгийн түр аргацаасан аргуудыг ашигласан байдаг. Сүүлийн 15 гаруй жилд судлаачид удирдамжгүй сургах аргаас багахан хэмжээний удирдамжтай (supervised) үгзүйн статистик загвар бий болгох аргыг сонирхох болжээ. 1990 -ээд оноос өмнө үгзүйн зөв бичих дүрмийг хүний оролцоотой гар аргаар

оруулж байсан боловч 2000 оноос хойш машин сургах аргаар зөв бичгийн дүрмийг цуглуулах оролдлогууд хийгдэж байв.

Алтай язгуурын хэлнүүд болох Турк, Монгол зэрэг залгамал хэлнүүдэд үг харьцангуй олон бүтээврийн дарааллаас бүтдэг. Иймээс удирдамжтай статистик аргыг хувирдаг үгзүйд хэрэглэхэд хэцүү байдаг [4]. Иймээс энэ ажлын зорилго нь үгзүйн шинжилгээнд багахан хэмжээний удирдамжтай ба бүрэн удирдамжгүй сургах аргуудыг Монгол хэлэнд хэрэгжүүлэх боломжийг судлах юм.

II. МАШИН СУРГАХ АРГА

Аливаа асуудлыг компьютерээр шийдэхийн тулд алгоритм шаардлагатай байдаг. Алгоритм бол оролтоос гаралт руу хувиргах үйлдлийг гүйцэтгэх алхмуудын дараалал юм.

Машин сургах арга нь жишээ өгөгдлүүд болон хуримтлуулсан мэдлэгтэй тулгуурлан гүйцэтгэлийн шалгуурыг оновчлох компьютерийн тооцоолол юм [5]. Машин сургах аргад хэд хэдэн параметрээр тодорхойлогдох загвар байх бөгөөд загвар нь ирээдүйн таамаглалуудыг тодорхойлохын тулд урьдчилсан эсвэл дүрсэлсэн шинжтэй байж болдог.

Удирдамжтай сургах аргын гол зорилго бол зөв үр дүнг агуулсан гаралтыг гаргахын тулд оролтын өгөгдлөөс зааварлагчийн тусламжтайгаар сурах явдал юм. Энэ аргын хүрээнд Яровский ба Висентовский нарын удирдамжтай сургах аргыг авч үзлээ.

Удирдамжгүй сургах аргад зааварлагч шаардлагагүй, зөвхөн оролтын өгөгдөл хэрэгтэй байдаг бөгөөд гол зорилго нь оролтын өгөгдлөөс зүй тогтлыг олох явдал юм. Энэ аргачлалын хүрээнд Патрик Шоне болон Даниел Журавскийн удирдамжгүй сургах аргыг авч үзлээ.

III. ЦАХИМ ХЭЛ ШИНЖЛЭЛ ДЭХ ҮГЗҮЙ

Хүн төрөлхтний хэл нь тодорхой тооны нэгж хэсгүүдээс бүрдэх бөгөөд тэдгээр нь нарийн зүй тогтлоор холбоотой бүхэл бүтэн дохионы тогтолцоо юм. Ийм нарийн тогтолцоотой хэлний нэгжүүд нь нэг нь нөгөөтэйгөө эсрэгцэхийн зэрэгцээ бие биеэ нөхөж, давхарлан шаталж тогтсон байдаг. Өөрөөр хэлбэл, дээд

төвшний нэгж нь доод төвшний нэгжээ багтаасан, доод төвшний нэгж нь дээд төвшний нэгждээ багтсан шинжтэй байна [6]. Хэлний нэгжүүдийн гол цөм нь үг бөгөөд хэлний аль ч салбарт үгийг олон талаас нь өөр өөр зорилгоор судалдаг.

XX зууны хагасаас эхлэн хэл шинжлэлийн ухаанд цахим хэл шинжлэл гэх шинэ салбар үүсэн бий болжээ. Компьютерийн хэл шинжлэл нь эх хэлийг компьютерээр хэрхэн боловсруулах арга техникийг судалдаг. Үг бол эх хэлний хамгийн бага нэгж юм [7]. Иймээс цахим үгзүй нь эх хэлийг компьютерээр боловсруулах эхний ажлуудын нэг билээ. Үгзүй нь үгсийн дотоод бүтцийг судалдаг ба үг нь бүтээвүүдээс тогтоно. Бүтээвэр гэдэг нь үгийн бүтцийн цааш үл задрах хамгийн бага утгат нэгжийг хэлнэ [8]. Үгийн язгуур, үндэс, залгавар нь бүтээврийн төрөл болно. Үгийн язгуур гэдэг нь үгийн гол цөм утга агуулдаг бөгөөд цаашид үл задрах бүтээвэр юм. Харин үгийн үндэс нь залгавар бүтээвэр авч хувилж болдог бүтээвэр юм. Залгавар бүтээвэр нь язгуур болон үндэс бүтээвэрт залгагдаж шинэ үг бүтээдэг [8].

IV. ПАТРИК ШОНЕ БА ДАНИЕЛ ЖУРАВСКИЙН АРГА

Шоне болон Журавский нарын энэхүү арга нь боловсруулаагүй текст корпусаас утгазүй, үгийн зөв бичлэгийн дүрэм, өгүүлбэрзүйн мэдээллийг ашиглан үгзүйн загварчлалын шинжилгээг хийдэг бөгөөд тэдний 2000 оны хамтарсан ажлын үргэлжлэл юм. CELEX (1996 онд Англи, Голланд, Герман хэлнүүдийн үгсийн сангаар байгуулагдсан) корпусыг ашиглан Англи, Голланд, Герман хэлнүүдэд туршигдсан.

Шоне, Журавский нар уг ажилдаа үгзүйн загварчлалын зарим аргуудын асуудалтай талыг дурдсан байдаг. Тухайлбал, Голдсмитынх шиг аргуудад зөвхөн зөв бичих дүрмийн талаас анхаарлаа хандуулсан байна гэж үзсэн бөгөөд жишээ дурдвал утгазүйн мэдээллийг нэмж тооцоолохгүйгээр “ally” гэсэн үгийг “all+y” гэж боловсруулахад төвөгтэй байх юм. Голдсмитээс хойших үгзүйн загварчлалын аргуудад үгсийн дүрмийн өөрчлөлтөд нөлөөлөх ажлууд хийгдээгүй бөгөөд [11] дан ганц утгазүйг дагнан боловсруулалт хийх нь хангалтгүй юм. Үүсмэл үгзүй нь утгазүйн хувьд гол утгаасаа холдсон байж болно.

Тэдний энэхүү алгоритм нь дөрвөн алхамд хуваагддаг. Үүнд:

1. Залгавруудад нэр дэвшигчдийг таамаглах;
2. Нэр дэвшигчдийн хослолуудыг боломжит үгзүйн хувилбарууд бүхий залгавруудад тодорхойлох;
3. Нэр дэвшигдэж байгаа залгавруудын үгс дэх эзлэмжээр нь утгазүйн векторыг тооцоолох;
4. Эдгээр үгстэй ижил утгазүйн вектор бүхий бодит үгзүйн хувилбаруудыг сонгох;

A. Залгавруудад нэр дэвшигчдийг таамаглах

Энэ нь уг алгоритмын эхний алхамд боломжтой залгавруудыг олох, ижил үгсийг тодорхойлох юм. Хэрэв w_1 , w_2 гэсэн хоёр үг p -ижил бол:

- i. w_1 -ийн эхний p тэмдэгтүүд w_2 -ийн эхний p тэмдэгтүүдтэй адил байх
- ii. w_1 -ийн $p+1$ тэмдэгт w_2 -ийн тэмдэгттэй адил биш байх

Жишээлбэл, үз, үзлэг үгсийн хувьд эхний хоёр (2-ижил) тэмдэгт адил байна.

B. Боломжит үгзүйн хувилбарууд бүхий залгавруудыг тодорхойлох

Залгавруудад нэр дэвшигчдийг тодорхойлсны дараа нэр дэвшигч залгавруудын хослолуудыг тодорхойлно. Үүнийг бид хослолын дүрэм (rules) гэе. Хоёр үг нь ижил үндэстэй бөгөөд ижил залгаврын дүрэмтэй хэлбэрийг “Боломжит үгзүйн хувилбаруудын хослол” (pair of potential morphological variants (PPMVs)) гэе. Өгөгдсөн дүрэм болох нийтлэг дүрмүүдтэй бүх PPMV -н цуглуулгаас дүрмийн олонлогийг (ruleset) тодорхойлно. Бидний алгоритм нь өгөгдлөөс нэр дэвшүүлж буй дүрэм бүрд дүрмийн олонлогуудыг тодорхойлсон жагсаалтыг бий болгодог.

C. Утгазүйн векторыг тооцоолох

1990 оноос хойш компьютерийн хэл шинжлэлийн судлаачид корпус дахь бичвэр болон үгүүдийн хоорондын утгазүйн холбоосуудын ач холбогдлыг олохын тулд хүний оролцоо шаардлагагүй байж болно гэдгийг нотолж, харуулах болсон [12]. Үүнийг M матрицад ганц утгат задралыг ашиглаж үр дүнг харуулсан юм. M матрицийн $M(i, j)$ оролт бүр нь корпусын j бичвэр дэх i үгийн давтамжийг агуулна. Энэ аргачлал нь Далд утгазүйн шинжилгээ (LSA) -тэй хамааралтай бөгөөд Ландауэр (Landauer 1998); Мэннинг ба Шфитц (Manning and Schfitze 1999) бүтээлд тодорхой тайлбарлагджээ.

Далд утгазүйн шинжилгээ (Latent semantic analysis) нь аливаа бичиг баримт болон текстийн түлхүүр үгэнд нь суурилан өгөгдлийг олж авахын хажуугаар, өгөгдөл хоорондын утгын хамаарал ямар байгааг харуулсан статик үзүүлэлтэд үндэслэн дүгнэн шинжилдэг.

Энэ шинжилгээ нь үг болон текст хоёрын хоорондын харилцаа хамаарлыг олох асуудлыг шийддэг [13]. Энэхүү арга нь текстэд давхардаж орсон үгсээр матриц үүсгэдэг [14] [15]. Матриц үүсгэхдээ туслах болон чимэх үгс мөн цифр, тэмдэгтүүдийг матрицын өгөгдлөөс хасаж тооцдог. Далд утгазүйн шинжилгээ нь ойролцоолох загвар бөгөөд төсөөтэй объектуудыг хамтад нь орон зайгаар нь бүлэглэдэг математикийн нэгэн арга юм.

Бидний судалгааны ажлын хувьд энэ шинжилгээний явцад нөхцөлт нөхцөлт матрицыг байгуулсан бөгөөд уг матрицдаа ганц утгат задралыг (singular value decomposition) ашиглав. Энэ аргыг ашигласнаар матрицын хэмжээ багасах юм. Тодорхойлогч нь тэгтэй тэнцүү матрицыг сингуляр матриц гэнэ [16].

D. Ижил утгазүйн вектор бүхий үгзүйн хувилбаруудыг сонгох

PPMV бүрийг үг бүрийн хоорондох хамаарлыг тодорхойлно. Үүнийг тодорхойлохын тулд normalized cosine score (NCS) нэрлэгдэх аргыг ашиглах болно. NCS-г

олохын тулд эхлээд утгазүйн вектор бүрийн хоорондох Ω_w косинусийг тооцоолно.

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|} \quad (1)$$

Утгазүйн векторууд нь санамсаргүйгээр сонгогдсон 200 үгээс бүрдэнэ. Энэ утгыг олсноор бид w -ийн хамаарлын утга (μ_w) болон стандарт хазайлт (σ_w) -г олно. w_k үг бүр нь $k \in (1,2)$, векторын бусад 200 үг санамсаргүйгээр сонгогдоно. w_k болон тооцоолсон 200 үг бүрийн хооронд косинусийн утгууд болох утга (μ_k), вариаци (σ_k) байна. w_1, w_2 -ийн косинус нь нормальчлагдсан бөгөөд NCS дараах байдлаар тооцоологдоно.

$$NCS(w_1, w_2) = \min_{k \in (1,2)} \frac{\cos(\Omega_{w_1, w_2}) - \mu_k}{\sigma_k} \quad (2)$$

Санамсаргүй NCS таамаглахдаа $N(\mu_T, \sigma_T^2)$ зөв корреляцийн вариациуд, тархалтын утга адил байх $N(0,1)$ хэвийн тархалт хийнэ. Сэлгэлтүүдийн хослолын хоорондох дотоод өгүүлбэр зүйн орчны төсөөтэй талууд нь бас тооцоологдоно. Тухайн сэдвийн үгсийн анхны тооцоолол нь бага давталттай, их давталттай ч бай аль аль нь санамсаргүйгээр тооцоологддог. Үүнийг боломжит үг зүйн вариациудын хослол буюу PPMV гэх ба үгсэд зориулсан сэлгэлтийн тал бүрийн гишүүд юм.

Зураг 1. Патрик Шоне ба Даниел Журавскийн аргын үр дүн

ХҮСНЭГТ 1. ПАТРИК ШОНЕ БА ДАНИЕЛ ЖУРАВСКИЙН АРГЫН ХАРЬЦУУЛСАН ҮР ДҮН

Ангилал	Англи хэл	Герман хэл	Турк хэл	Монгол хэл
	Хувь	Хувь	Хувь	Хувь
Шоне ба Журавский	85.2	88.3	58.6	64.2

V. ЯРОВСКИЙ БА ВИСЕНТОВСКИЙН АРГА

Яровский ба Висентовский нар хувирдаг үгзүйн шинжлэгч бий болгох, бага хэмжээний удирдамжтай арга боловсруулжээ. Энэ арга нь хоёр алхмаас бүрдэнэ.

- Үгийн үндэс болон хувирсан үгийн хэлбэрийг зэрэгцүүлсэн хүснэгтийг корпусаар үнэлдэг. Жишээ нь: хүснэгт нь *морь/моринд* хосыг агуулбал *моринд* нь *морь* гэсэн үгний хувирсан

хэлбэр мөн эсэхийг корпус ашиглан танина. *морь/моринд* хосыг боломжит хос гэнэ.

- Удирдамжтай үгзүйн шинжлэгч хүснэгтийн жигнэсэн дэд олонлог дээр суралцдаг. Боломжит хосуудыг үгийн үндсийн өөрчлөлт болон залгах дагавраар илэрхийлэх боломжтой байна.

Хоёр дахь алхамд дараах нөөцүүд шаардлагатай байдаг. Үүнд:

- Хэлний үг хувирлын ангиллын хүснэгт: хувирлын ангиллуудад ангиллуудын *тогтсон дагаврын* жагсаалт харгалзана.
- Томоохон текст корпус
- Хэлний эгшиг болон гийгүүлэгчийн жагсаалт
- Толь бичгээс сонгон авсан хэлний нэр үг, үйл үг, тэмдэг нэр үгийн үндсүүдийн жагсаалт
- Хэрэв шаардлагатай бол утгвар үгсийн жагсаалт

Дагавруудыг тодорхойлж өгснөөр үндэс ба хувилсан хэлбэр гэсэн хосууд байх боломжтой гэдгийг хялбархан мэдэж болно. Гэвч ихэнх тохиолдолд энэ тогтсон дагаврууд нь хангалттай биш байдаг. Өөрөөр хэлбэл, тухайн хувирсан үг үгийн үндсээс ондоошин хэлбэржсэн бол энэ тохиолдол үүсдэг.

Яровский ба Висентовский нар энэ асуудлыг шийдэхдээ өгөгдсөн хувирсан хэлбэр/үндэс хос үгийн корпус дахь харьцааны тархалтыг ($\log \frac{\text{хувирсан хэлбэр}}{\text{үндэс}}$) тооцоолсон байна. Тархалтын утгын нормальчлал нь тус харьцааны хувьд магадлалын нягтын функц руу ойртуулна. Хувирлын байх боломжит хосууд тархалтын дунджид, буруу хосууд тархалтын захаар байрлана. Өөрөөр хэлбэл, бараг боломжгүй үзэгдэл болж байна.

Дээрх тохиолдолд хувилсан хэлбэр/хувилаагүй хэлбэрийн хос нь зөв хос уу, буруу хос уу гэдгийг мэдэх шаардлага гарна. Үүнийг шийдэхийн тулд дүрмийн болоод ондоошон хувирсан үгсийн тархалтын төсөөг гаргах хэрэгтэй.

Яровский ба Висентовский нар үгийн хэлбэрүүдийн хоорондын хамаарлыг дан ганц үнэлгээгээр хязгаарлалгүйгээр Шоне, Журавский нарын ашигласан нэгэн төрлийн агуулгын нотолгоог авч үзжээ. Утга агуулгын жигнэсэн векторуудын хоорондын косинуслаг төсөөний хэмжээ нь утгын холбоотой үгсийн хоорондын хамаарлыг тодорхойлох үзүүлэлт болно.

Яровский ба Висентовский нар тухайн үндэс болон түүний хувилсан хэлбэрүүдийн нарийвчилсан харилцан хамаарлын хүснэгтийг үүсгэхээс гадна шинэ тохиолдлуудад өргөтгөх боломжтой үгзүйн шинжлэгч бий болгох зорилго тавьсан. Үүний тулд үндэс, тухайн үндсийн өөрчлөлт, дагавар, POS –ийн магадлалыг тооцоолсон.

$$P(\text{streamchange}|\text{root}, \text{suffix}, \text{POS})$$

Энэ нь дараах магадлалуудын жигнэсэн нийлбэр бүхий төгсгөлөөс нь салгах загвараар үнэлнэ.

- POS, залгавар, үндсийн төгсгөлийн 3 үсгийг өөрчилж залгагдах магадлал,

- POS, залгавар, үндсийн төгсгөлийн 2 үсгийг өөрчилж залгагдах магадлал,
- POS, залгавар, үндсийн төгсгөлийн 1 үсгийг өөрчилж залгагдах магадлал,
- POS, залгавар, үндсийн төгсгөлийг өөрчлөхгүйгээр залгагдах магадлал,
- Орчноос хамаарахгүй α үндэс β болж өөрчлөгдөх магадлал.

$$\begin{aligned}
 P(\text{streamchange}|\text{root}, \text{suffix}, \text{POS}) &= P(\alpha \rightarrow \beta|\text{root}, \text{suf}, \text{POS}) \\
 &\approx \lambda_1 P(\alpha \rightarrow \beta|\text{last}_3(\text{root}), \text{suf}, \text{POS}) + \\
 &(1-\lambda_2)(\lambda_2 P(\alpha \rightarrow \beta|\text{last}_2(\text{root}), \text{suf}, \text{POS}) + \\
 &(1-\lambda_3)(\lambda_4 P(\alpha \rightarrow \beta|\text{suf}, \text{POS}) + \\
 &(1-\lambda_4)P(\alpha \rightarrow \beta))) \quad (3)
 \end{aligned}$$

Үүнд λ нь алгоритмын алхам бүрд дахин тооцоологдох хувьсагч байна.

ХҮСНЭГТ 2. ЯРОВСКИЙ БА ВИСЕНТОВСКИЙН АРГЫН ХАРЬЦУУЛСАН ҮР ДҮН

Ангилал	Англи хэл		Турк хэл		Монгол хэл	
	Сур. хувь	Үр дүн	Сур. хувь	Үр дүн	Сур. хувь	Үр дүн
Яровский ба Висентовский	75	97.6	75	91.4	50	68.7

VI. ДҮГНЭЛТ

Энэхүү судалгааны ажлын хүрээнд удирдамжтай, удирдамжгүй сургах үгзүйн аргуудын төлөөлөл болсон аргуудыг судалж, хэрэгжүүлэв. Ингэснээр дараах дүгнэлтэд хүрэв. Үүнд:

1. Дээрх өгүүлсэн аргууд нь үгзүйн дүрмийн олон ажилд иш татагдсан байдаг. Хэл шинжлэлд гарч буй шинэ аргачлалууд төдийлөн хурдан хугацаанд гараад байдаггүй ч шинэ гарч буй аргуудыг үнэлэх нэгэн төрлийн стандарт болж чадсан.
2. Монгол хэлтэй төстэй залгамал хэлнүүдэд туршигдсан. Нөгөөтэйгүүр, туршигдаж буй хэлнүүдийнхээ хувьд бусад аргуудтай харьцуулбал харьцангуй өндөр үр дүнг үзүүлдэг.
3. Залгамал хэлнүүдийн онцлог болох хувирдаг үгзүйн асуудал нь статик үгзүйн загварчлалаас татгалзах нэгэн шалтгаан болж байна. Тиймээс дээрх аргуудыг залгамал хэлнүүдийн хувьд туршиж үзэх нь илүү үр дүнг үзүүлэх боломжтой.
4. Яровский ба Висентовский нарын арга нь удирдамжтай бөгөөд үгийн боломжит хосуудыг үнэлэх нарийн аргачлалтай.

5. Патрик Шоне ба Даниел Журавскийн алгоритм нь хувирдаг хэлнүүдэд төвлөрч байгаа боловч алгоритмд хүний зүгээс мэдээлэл оруулахгүй, зөвхөн утгазүйн нөлөөллөөр нь тооцоолж байгаа учраас үгийн хуваалтын хувьд онцгой тохиолдол гарч байгаа.
6. Уг алгоритмд далд утгазүйн шинжилгээ, ганц утгат задралын аргуудыг давхар ашигласан нь эдгээр аргуудын монгол хэлний үгзүйн хувьд анхны туршилтууд болж байна.
7. Ижил үгзүйн бүтэцтэй үгс ижил утгыг хуваалцаж байдаг. Тиймээс уг алгоритмын үр дүнд PPMV буюу боломжит үгзүйн хувилбаруудны хослолын векторуудын хамааралтай байдалд анхаарлаа хандуулав.
8. Судалгааны ажлын дагуу хийсэн програмд 1000 үгтэй текст файлыг оруулан гаралтын үр дүнг авсан.

НОМ ЗҮЙ

[1] Монгол улсын үндсэн хууль, Улаанбаатар, 1992.

[2] А.Хүдэр, Өгүүлбэр дэх үгийн аймгийг марковын гинжээр тодорхойлох компьютерийн загвар боловсруулах нь, Улаанбаатар, 2013.

[3] R, Roack; R, Sproat, Computational Approaches to Morphology and Syntax, New York: Oxford University Press Inc, 2007.

[4] Б. Батзолбоо, “Кирилл, Монгол бичгийг хөрвүүлэх ба Монгол хэлний үг хувиллыг загварчлах судалгаа,” Докторын зэрэг горилсон нэг сэдэвт бүтээл, pp. 25-29, 2012.

[5] E. Alpaydin, Introduction to Machine Learning, London: The MIT Press, 2009.

[6] Ү. Ариунболд, Д. Төмөртоогоо, Г. Цэцэгдарь, Ж. Санжаа ба Ц. Өнөрбаян, Орчин цагийн монгол хэл, Улаанбаатар, 2004

[7] S.R.Spiegler, Machine Learning for the Analysis of Morphologically Complex Languages, United Kingdom: University of Bristol, 2011.

[8] Ц.Өнөрбаян, Орчин цагийн Монгол хэлний үгзүй, Улаанбаатар, 2004.

[9] R, Roack; R, Sproat, Computational Approaches to Morphology and Syntax, New York: Oxford University Press Inc, 2008.

[10] J. Goldsmith, Unsupervised Learning of the Morphology of a Natural Language, Chicago: MIT Press, 2001.

[11] R. Brian ба S. Richard, Computational Approaches to Morphology and Syntax, New York: Oxford University Press, 2007.

[12] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer ба R. Harshman, Indexing by latent semantic analysis, Journal of the American Society for Information Science, 1990.

[13] T. K. Landauer, D. S. McNamara, S. Dennis ба W. K. (eds.), Handbook of Latent Semantic Analysis, Lawrence Erlbaum, 2007.

[14] Association for Computational Linguistics, Advantages in WSD, 2005.

[15] B. Chen, Latent Semantic Analysis (LSA).

[16] Trucco, Singular Value Decomposition (SVD), Appendix A.6.