



ПРЕДМЕТ НОМИНИ АТОВЧИ БИРЛИКЛАРНИНГ СИНОСЕТНИ ИШЛАБ ЧИҚИШ: ЁНДАШУВ ВА УСУЛЛАР

Нажмиддинов Муҳаммаджон

Филология фанлари бўйича фалсафа доктори, доцент,

Қўқон университети.

MAQOLA HAQIDA	ANNOTATSIYA
<p>Qabul qilindi: 24-dekabr 2024-yil Tasdiqlandi: 26-dekabr 2024-yil Jurnal soni: 13 Maqola raqami: 62 DOI: https://doi.org/10.54613/ku.v13i.1072</p>	<p>Мазкур мақолада синсетларни яратиш жараёни, уларнинг табиий тилга ишлов беришдаги аҳамияти ҳамда қўлланилиш доиралари таҳлил қилинади. Синсетларни қўлда ва автоматик равишда тузиш ёндашувлари ўрганилиб, уларнинг афзалликлари ва камчиликлари ёритилади. Шунингдек, замонавий технологияларнинг (Ворд Эмбеддингс, машинани ўргатиш) синсетлар яратишдаги роли кўрсатилиб, келажакдаги истиқболли тадқиқот йўналишлари белгилаб берилди. Ушбу мақола синсетлар соҳасидаги долзарб масалаларни ўрганаётган тадқиқотчилар ва дастурий таъминот ишлаб чиқувчилар учун фойдали бўлади.</p>
<p>KALIT SO'ZLAR/ КЛЮЧЕВЫЕ СЛОВА/ KEYWORDS</p>	
<p>Синсетлар, табиий тилга ишлов бериш, машинани ўргатиш, Ворд Эмбеддингс, онтология, кўп тиллилик, автоматлаштириш, технологик ютуқлар.</p>	

КИРИШ

Маъно жиҳатидан бир хил ёки ўхшаш объект (предмет, шахс, нарса, ҳодиса)ни билдирувчи сўз туркумлари предмет номларини атоқчи бирликлар синсетлардир. Бундай синсетларнинг яратилиши тезаурус ва онтологияларни ишлаб чиқишда асосий босқич бўлиб, ахборотни қидиришдан тортиб, машина таржимасигача бўлган кўплаб функцияларда аҳамиятли.

Синсетларни тузишнинг ўзига хос мураккаблик ва ва нюанслари мавжуд бўлиб, уларни қуйидагича изоҳлаш мумкин:

1. Кўп маънолик. Кўп сўзлар бир нечта маънога эга бўлиб, бир маъноли синсетларни яратишни қийинлаштиради. Масалан, банк сўзи молия муассасаси ёки моддий жамғармани сақлаш жойини англатиши мумкин. Бундай вазиятда ҳар бир синсетда синоним ва антонимлар фарқи бўлади.

2. Контекста боғлиқлик. Сўзнинг маъноси контекста қараб ўзгариши мумкин. Масалан, сичқонча сўзи ҳам ҳайвонни, ҳам компьютер қурилмасини англатади. Бу эса кўп маънолик ҳодисасидан фарқ қилади. Бундай синсетларни тузишда контекстуал маъно кўчиши ёки бошқа ҳодисаларни инобатга олиш талаб этилади.

3. Диалектал ибора ёки бирликлар. Турли минтақаларда ёки ижтимоий гуруҳларда бир мавзу учун турли номларнинг мавжудлиги синсетда қаторлар сонининг ошишига олиб келади.

4. Мавҳум тушунчалар. Айрим предмет ва объектларни ифодалаш мураккаб бўлиб, мавҳум хусусиятларга эга бўлиши мумкин.

Демак, синсетларни тузиш жараёнида юқоридагиларни инобатга олиш талаб этилади.

МЕТОДЛАР

Компютер лексикографиясида синсетларни яратишнинг турли усуллари мавжуд бўлиб, улар қатор афзаллик ва қулайликларга эга. Қуйида уларни бир нечасига тўхталамиз.

1. Қўлда таҳлил қилиш. Бу усул луғатларни синчковлик билан ўрганиш, матн корпуси ва эксперт маслаҳати асосида амалга ошириладиган усул бўлиб, сезиларли даражада вақтни талаб қилади, лекин юқори аниқлиқни таъминлайди.

2. Автоматик усуллар. Катта ҳажмдаги матн маълумотларини таҳлил қилиш ва синонимик муносабатларни аниқлаш учун машинали ўқитиш алгоритмларидан фойдаланиш мумкин. Бундай усуллар катта корпуслар учун самарали бўлиши мумкин, аммо натижаларни баҳолашни талаб қилади, чунки автоматик равишда шакллантирилган сўзлар рўйхатида тасодифий сўзлар ҳам учрайди.

3. Гибрид ёндашув. Қўлда ва автоматик таҳлилнинг комбинацияси аниқлик ва маҳсулдорлик ўртасидаги оптимал мувозанатга эришишга имкон беради.

НАТИЖАЛАР

Мавжуд тезауруслар тажрибасини ўрганиш шунга кўрсатадики, тезаурусларни ишлаб чиқишда юқорида келтирилган усулларнинг бир нетаси ёки айримини қўллаш мумкин. Бу, албатта, тузувчи имкониятига боғлиқ бўлиб, маҳсулот натижаси ҳар хил бўлади.

Синсетларнинг тузилиши

Предметни ифодаловчи бирликларнинг синсети қатор унсурларни ўз ичига олади. Уни қуйидаги чизмада ифодалаш мумкин (Қаранг: 1.1-чизма).

Бирлик	Вазифаси
Асосий синсет	асосий тушунчани билдирувчи марказий сўз ёки сўзлар гуруҳи
Синонимлар	ўхшаш маънога эга сўзлар
Гипернимлар	умумий атамалар “ит” – “ҳайвон”
Гипонимлар	Аниқроқ тушунчалар “ит” – “кўппак”
Мернимлар	бир бутуннинг қисмлари “машина” – “ғилдирак”
Ассоциациялар	маъноси бўйича ўзаро боғланган, лекин синоним ёки антоним бўлмаган сўзлар.
Семантик роллар	Объект(предмет)нинг турли вазиятларда ўйнайдиган роли “болға” – “асбоб”

1.1-чизма. Предметни ифодаловчи бирликлар синсети унсурлар

Қуйида синсет (тезаурусдаги сўз уяси) намуналарига мисол келтираемиз.

- Компютер: электрон ҳисоблаш машинаси, компютер, шахсий компютер, ноутбук, планшет, гаджет.
- Автомобил: автомобиль, авто, машина.
- Уй: уй, квартира, бино.

Синсетларнинг қўлланиши

Синсетлардан турли типдаги лексикографик ҳсулотлар ва табиий тилга ишлов беришни қатор вазиларини ечишда фойдаланилади. Қуйида уларни қисқача келтираемиз:

- Маълумот қидириш. Синонимларни киритиш қидирув натижаларини кенгайтиради. Қидирув тизимларида синонимларнинг мавжудлиги ахборот-қидирув жараёнида қидирилган сўзнинг синоними орқали қидиришни таъминлайди, натижада, қидирув натижасида кўпроқ материал топилади.
- Машина таржимаси. Синсетларнинг мавжудлиги таржиманинг аниқлигини таъминлайди.
- Матн таҳлил қилиш. Синсетларнинг мавжудлиги матн мавзуларини аниқлаш, асосий тушунчаларни аниқлашда аҳамиятли.

4. Онтологияларни яратиш. Фан соҳасига оид тузилган билимлар мажмуини шакллантиришга замин яратилади.

5. Чатбот ва виртуал ёрдамчиларни ишлаб чиқиш. Синсетлар фойдаланувчи сўровларини тушуниш ва тегишли жавобларни тақдим этишга ёрдам беради.

Синсет ишлаб чиқишнинг келажакда кутилаётган натижалари сифатида айтиш мумкинки, машинали ўрганиш ва табиий тилларни қайта ишлаш технологияларининг ривожланиши билан синсетларни яратиш тобора автоматлаштирилган ва самарали бўлиб бормоқда. Бироқ (айниқса) юқори ихтисослашган соҳалар учун қўлда таҳлил қилиш ва эксперт хулосаси ушбу жараённинг муҳим таркибий қисмлари бўлиб қолмоқда.

Маъно жиҳатидан бир хил ёки ўхшаш объект (предмет, шахс, нарсаси, ҳодисаси)ни билдирувчи сўз туркумлари предмет номларини атоқчи бирликлар синсетлардир. Бундай синсетларнинг яратилиши тезаурус ва онтологияларни ишлаб чиқишда асосий босқич бўлиб, ахборотни қидиришдан тортиб, машина таржимасигача бўлган кўплаб функцияларда аҳамиятли.

Синсетларни тузишнинг ўзига хос мураккаблик ва ва нюанслари мавжуд бўлиб, уларни қуйидагича изоҳлаш мумкин:

1. Кўп маънолилиқ. Кўп сўзлар бир нечта маънога эга бўлиб, бир маъноли синсетларни яратишни қийинлаштиради. Масалан, *банк сўзи молия муассасаси ёки моддий жамғармани сақлаш жойини* англатиши мумкин. Бундай вазиятда ҳар бир синсетда синоним ва антонимлар фарқли бўлади.

2. Контекста боғлиқлик. Сўзнинг маъноси контекста қараб ўзгариши мумкин. Масалан, *сичқонча* сўзи ҳам ҳайвонни, ҳам *компьютер қурилмасини* англатади. Бу эса кўп маънолилиқ ҳодисасидан фарқ қилади. Бундай синсетларни тузишда контекстуал маъно кўчиши ёки бошқа ҳодисаларни инobatга олиш талаб этилади.

3. Диалектал ибора ёки бирликлар. Турли минтақаларда ёки ижтимоий гуруҳларда бир мавзу учун турли номларнинг мавжудлиги синсетда қаторлар сонининг ошишига олиб келади.

4. Мавҳум тушунчалар. Айрим предмет ва объектларни ифодалаш мураккаб бўлиб, мавҳум хусусиятларга эга бўлиши мумкин.

Демак, синсетларни тузиш жараёнида юқоридагиларни инobatга олиш талаб этилади.

Синсетларни яратиш усуллари

Компьютер лексикографиясида синсетларни яратишнинг турли усуллари мавжуд бўлиб, улар қатор афзаллик ва қулайликларга эга. Қуйида уларни бир нечасига тўхталамиз.

1. Қўлда таҳлил қилиш. Бу усул луғатларни синчковлик билан ўрганиш, матн корпуси ва эксперт маслаҳати асосида амалга ошириладиган усул бўлиб, сезиларли даражада вақтни талаб қилади, лекин юқори аниқликни таъминлайди.

2. Автоматик усуллар. Катта ҳажмдаги матн маълумотларини таҳлил қилиш ва синонимик муносабатларни аниқлаш учун машинали ўқитиш алгоритмларидан фойдаланиш мумкин. Бундай усуллар катта корпуслар учун самарали бўлиши мумкин, ammo натижаларни баҳолашни талаб қилади, чунки автоматик равишда шакллантирилган сўзлар рўйхатида тасодифий сўзлар ҳам учрайди.

3. Гибрид ёндашув. Қўлда ва автоматик таҳлилнинг комбинацияси аниқлик ва маҳсулдорлик ўртасидаги оптимал мувозанатга эришишга имкон беради.

Мавжуд тезауруслар тажрибасини ўрганиш шуни кўрсатадики, тезаурусларни ишлаб чиқишда юқорида келтирилган усулларнинг бир нетаси ёки айримини қўллаш мумкин. Бу, албатта, тузувчи имкониятига боғлиқ бўлиб, маҳсулот натижаси ҳар хил бўлади.

Синсетларнинг тузилиши

Предметни ифодаловчи бирликларнинг синсети қатор унсурларни ўз ичига олади. Уни қуйидаги чизмада ифодалаш мумкин (Қаранг: 1.1-чизма).

Бирлик	вазифаси	
Асосий синсет	асосий тушунчани билдирувчи марказий сўз ёки сўзлар гуруҳи	
Синонимлар	ўхшаш маънога эга сўзлар	
Гипернимлар	умумий атамалар	“ит” – ҳайвон
Гипонимлар	Аниқроқ тушунчалар	“ит” – “кўппак”

Мернимлар	бир бутуннинг қисмлари	“машина” – “ғилдирак”
Ассоциациялар	маъноси бўйича ўзаро боғланган, лекин синоним ёки антоним бўлмаган сўзлар.	
Семантик роллар	Объект(предмет)нинг турли вазиятларда ўйнайдиган роли	“болға” – “асбоб”

1.1-чизма. Предметни ифодаловчи бирликлар синсети унсурлар

Қуйида синсет (тезаурусдаги сўз уяси) намуналарига мисол келтираемиз.

1. Компьютер: электрон ҳисоблаш машинаси, компьютер, шахсий компьютер, ноутбук, планшет, гаджет.
2. Автомобил: автомобил, авто, машина.
3. Уй: уй, квартира, бино.

Синсетларнинг қўлланиши

Синсетлардан турли типдаги лексикографик ҳсулотлар ва табиий тилга ишлов беришни қатор вазиларини ечишда фойдаланилади. Қуйида уларни қисқача келтираемиз:

1. Маълумот қидириш. Синонимларни киритиш қидирув натижаларини кенгайтиради. Қидирув тизимларида синонимларнинг мавжудлиги ахборот-қидирув жараёнида қидирилган сўзнинг синоними орқали қидиришни таъминлайди, натижада, қидирув натижасида кўпроқ материал топилади.

2. Машина таржимаси. Синсетларнинг мавжудлиги таржиманинг аниқлигини таъминлайди.

3. Матн таҳлил қилиш. Синсетларнинг мавжудлиги матн мавзуларини аниқлаш, асосий тушунчаларни аниқлашда аҳамиятли.

4. Онтологияларни яратиш. Фан соҳасига оид тузилган билимлар мажмуини шакллантиришга замин яратилади.

5. Чатбот ва виртуал ёрдамчиларни ишлаб чиқиш. Синсетлар фойдаланувчи сўровларини тушуниш ва тегишли жавобларни тақдим этишга ёрдам беради.

Синсет ишлаб чиқишнинг келажакда кутилаётган натижалари сифатида айтиш мумкинки, машинали ўрганиш ва табиий тилларни қайта ишлаш технологияларининг ривожланиши билан синсетларни яратиш тобора автоматлаштирилган ва самарали бўлиб бормоқда. Бироқ (айниқса) юқори ихтисослашган соҳалар учун қўлда таҳлил қилиш ва эксперт хулосаси ушбу жараённинг муҳим таркибий қисмлари бўлиб қолмоқда.

Юқорида айтилганидек, синсетларни ишлаб чиқишда қатор восита ва усуллардан фойдаланилади. Булар орасида автоматик ишлов бериш усуллари ҳам кенг тарқалган. Қуйида Word Embeddings функциясини қўллаш орқали синсет яратишга тўхталамиз.

Word Embeddings (ўхшаш сўзлар қатори) – бу кўп ўлчамли маконда сўзларни вектор сифатида кўрсатишга имкон берувчи кучли табиий тилни қайта ишлаш (NLP) воситаси. Бунда маъноси ўхшаш сўзлар бир-бирига яқинроқ жойлашади.

Синсетларни йиғишда “Word Embeddings”дан қуйидагича фойдаланиш мумкин:

1. Энг яқин ўхшаш сўзлар (валентлик қобилияти ўхшаш сўзлар) қидирилади.

2. Синоним/ўхшашини топилиши талаб қилинган мақсадли сўз танланади.

3. Вектор фазода унга энг яқин сўзлар топилади. Бу сўзлар рўйхати синоним ёки ўхшаш маънога эга бўлиш эҳтимоли юқори бўлган сўзлардан ҳосил қилинади.

4. Кластерлаш. Сўзлар вектор кўринишига қараб гуруҳланади. Худди шу кластердаги сўзлар синоним ёки ўхшаш маънога эга бўлиши мумкин.

5. Чизикли операциялар.

6. Сўз векторлари устида арифметик амаллар бажарилади. Масалан, “қирол” - “эркак” + “аёл” ≈ “қиролича”. Бундай операциялар аналогларни топиш ва янги синонимларни аниқлаш имконини беради.

Синсетларни йиғишда “Word Embeddings”дан фойдаланишнинг ижобий ва салбий жиҳатлари бўлиб, уни қуйидагича изоҳлаш мумкин (1.2-чизма):

Word Embeddings функциясининг синсетларни тузишдаги афзалликлари	Word Embeddings функциясининг синсетларни тузишдаги камчиликлари
<p>1. <i>Автоматлаштириш.</i> Синонимларни йиғиш жараёни янада автоматлаштирилган бўлиб, бу вақт ва меҳнат харажатларини сезиларли даражада камайтиради.</p> <p>2. <i>Масштабlilik:</i> катта ҳажмдаги матнли маълумотлар билан ишлаш мумкин.</p> <p>3. <i>Кўп тилlilik.</i> Кўп тиллар учун олдиндан ўргатилган Word Embeddings моделлари мавжуд.</p> <p>4. <i>Контекстанд хабардорлик.</i> Замонавий Word Embeddings моделлари сўз ишлатилган контекстини ҳисобга олади, бу эса унинг маъносини аниқроқ аниқлаш имконини беради.</p>	<p>1. <i>Моделларнинг сифати.</i> Олинган синонимларнинг сифати тўғридан-тўғри ишлатиладиган Word Embeddings моделининг сифатига боғлиқ.</p> <p>2. <i>Сўзларнинг кўп маъноlilikи.</i> Кўп маъноли сўзлар учун уларнинг маълум бир контекстда қандай маънода ишлатилишини аниқлаш қийин бўлиши мумкин.</p> <p>3. <i>Кам қўлланувчи сўзларда натижанинг чегараланганlilikи.</i> Кам қўлланувчи сўзлар учун мос вектор кўринишларини топиш қийин бўлиши мумкин.</p>

1.2-чизма. Синсетларни йиғишда “Word Embeddings” дан фойдаланишнинг ижобий ва салбий жиҳатлари

“Word Embeddings”нинг энг кўп қўлланиладиган моделлари сифатида қуйидагиларни санаш мумкин:

1. Word2Vec¹ – Нейрон тармоқларга асосланган энг машҳур моделлардан бири.

2. GloVe² – сўзнинг ёнма-ён келиш эҳтимоли матричасида асосида тайёрланган модел.

3. BERT³ – жумладаги сўз контекстини ҳисобга оладиган чуқур ўрганиш (DeepLearnig) модели.

Gensim кутубхонасида фойдаланган ҳолда Python тилида қуйидаги код асосида энг яқин (кўшни) сўзларни топиш алгоритминини қуйидагича бериш мумкин:

```
from gensim.models import KeyedVectors
# Ўқитилган моделни юклаймиз:
model = KeyedVectors.load_word2vec_format('path/to/model')
# "яхши" сўзи учун ўхшаш сўзлар қаторини топамиз:
similar_words = model.most_similar('яхши')
print(similar_words)
```

Демак, Word Embeddings – бу синсетларни автоматик йиғишда юқори натижа берувчи восита. Бироқ юқори сифатли натижаларга эришиш учун моделни диққат билан танлаш ва унинг хатоларини ҳисобга олиш керак. Автоматлаштирилган усулларни қўлда таҳлил қилиш билан бирлаштириш юқори аниқлик ва ишончlilikни таъминлайди.

Юқорида айтилганидек, синсетларни ишлаб чиқишда қатор восита ва усуллардан фойдаланилади. Булар орасида гибрид ишлов бериш усуллари ҳам кенг тарқалган. Қуйида гибрид моделларни қўллаш орқали синсет яратишга жараёнини тавсифлаймиз.

Предметни атовчи бирликлар синсетларини тузиш учун код яратиш комплекс ёндашувни талаб қиладиган вазифадир, жумладан, қуйидагилар талаб этилади:

1. Маълумотларни йиғиш. Моделни ўргатиш учун фойдаланиладиган матн корпусини тайёрлаш талаб этилади. Бунда тайёр катта ҳажмли очик корпусдан фойдаланиш ёки мустақил равишда корпус тузиш талаб этилади.

2. Модел танлаш. Вектор фазода сўзларни ифодалаш учун мос машинали ўрганиш моделини танлаш (масалан, Word2Vec, GloVe, BERT).

3. Моделни ўқитиш. Танлаб олинган ёки тайёрланган корпусда трейнинг (моделни ўқитиш) машғулотни амалга оширилади.

4. Синонимлар(ўхшаш сўзлар)ни қидириш. Маъноси яқин бўлган сўзларни қидириш учун ўргатилган моделдан фойдаланиш.

5. Дастурлаш тили ва кутубхоналарни танлаш. Ушбу вазифани амалга ошириш учун турли дастурлаш тиллари ва кутубхоналардан фойдаланиш мумкин. Энг кенг қўлланувчи вариантлар сифатида Python ва унинг кутубхоналари таъкидлаш мумкин:

1. NLT⁴: Табиий тилни қайта ишлаш учун асосий кутубхона.

2. Gensim⁵: Мавзуни моделлаштириш ва матнни таҳлил қилиш учун кутубхона.

3. spaCy⁶: Кўп тилларни қўллаб-қувватлайдиган табиий тилларни қайта ишлаш кутубхонаси.

4. TensorFlow/PyTorch⁷: BERT нисбатан мураккаб моделлар билан ишлаш учун чуқур ўрганиш (Deep Learning) тизими керак.

Python асосида Gensim кутубхонасини қўллашни қуйидагича тавсифлаймиз:

```
from gensim.models import KeyedVectors
# Ўқитилган моделни юклаш
model = KeyedVectors.load_word2vec_format('path/to/model')
# "компьютер" сўзига энг яқин бўлган 10 та сўзни топиш
similar_words = model.most_similar('компьютер', topn=10)
print(similar_words)
```

Python асосида spaCy кутубхонасини қўллашни қуйидагича тавсифлаймиз:

```
import spacy
nlp = spacy.load("en_core_web_lg") # ўқитилган тил моделини юклаш
# сўзнинг векторли кўринишларини оламиз
word = "computer"
doc = nlp(word)
word_vector = doc[0].vector
# ўхшаш сўзларни қидирамиз
similar_docs =
```

nlp.vocab.vectors.most_similar(positive=[word_vector], topn=10)

```
for sim_doc in similar_docs:
    print(nlp.vocab.strings[sim_doc[0]])
```

МУҲОКАМА

Юқоридаги натижалардан кўриниб турибдики, синсетларни тузиш ва улардан фойдаланиш жараёнида турли ёндашувлар ўзига хос ўстунлик ва чекловларга эга. Ҳозирги рақамли технологиялар даврида синсетлар табиий тилга ишлов бериш (НЛП) соҳасида муҳим аҳамият касб этади. Уларсиз ахборотни излаш, матнларни таҳлил қилиш, онлайн таржима каби кўплаб функциялар тўлиқ амалга оширилмайди. Айниқса, кўп маъноlilik ва контекстга боғлиқ муаммоларни ҳал қилиш синсетларнинг сифатли бўлишини таъминлайди. Бу эса нафақат илмий тадқиқотлар, балки амалий дастурларда ҳам муҳим аҳамиятга эга. Қўлда таҳлил юқори аниқликни таъминлайди, лекин кўп вақт ва меҳнат талаб қилади. Автоматик усуллар эса катта ҳажмдаги маълумотлар учун самарадор, аммо натижаларда нотўғри ёки иррелевант маълумотлар пайдо бўлиш эҳтимоли юқори. Шунинг учун, комбинацияланган (гибрид) ёндашув энг мақбул эчим сифатида қаралиши лозим. Бу ёндашувни жорий қилиш илмий изланишлар ва амалий дастурлар самарадорлигини оширади.

Синсетларнинг қўлланилиши фақат матн таҳлили ёки ахборот излаш билан чекланмайди. Улар онтологияларни тузишда, чат-ботларни ривожлантиришда ва машина таржимасини такомиллаштиришда ҳам муҳим роль ўйнайди. Айниқса, кўп тилlilikни қўллаб-қувватлашда синсетлар тил ўрганиш, халқаро мулоқот ва маданиятлараро алоқаларни кучайтиради. Шу сабабли, синсетларни турли тил ва маданиятларга мослаштириш масаласи долзарблигини сақлаб қолади. Сўнги йилларда Ворд Эмбеддингс каби технологиялар синсетларни яратишда катта ўзгаришларга сабаб бўлди. Ушбу технологиялар соддалаштирилган ва автоматлаштирилган ёндашувларни тақдим этмоқда. Шунга қарамай, юқори даражада мураккаб ёки тор доирадаги соҳага оид синсетларни яратишда эксперт маслаҳати ва қўлда таҳлил қилиш зарурати ҳали ҳам мавжуд.

Келгусида машиналарни ўргатиш ва табиий тилни қайта ишлаш технологиялари ривожланиб, синсетларни янада автоматлаштириш ва оптималлаштириш кутилмоқда. Шу билан бирга, бу технологияларни ривожлантириш учун миллий ва

¹ <https://www.tensorflow.org/text/tutorials/word2vec>

² <https://nlp.stanford.edu/projects/glove/>

³ [https://en.wikipedia.org/wiki/BERT_\(language_model\)](https://en.wikipedia.org/wiki/BERT_(language_model))

⁴ <https://www.nltk.org/>

⁵ <https://gensim2.github.io/>

⁶ <https://spacy.io/>

⁷ <https://www.tensorflow.org/>

халқаро миқёсда кўпроқ тадқиқотлар ўтказиш лозим. Айниқса, кам ишлатиладиган тиллар учун синсетлар яратиш, уларнинг тез-тез янгиланиб туриши ва реал ҳаётда фойдаланиш имкониятлари бўйича ишлар олиб борилиши керак.

ХУЛОСА

Хулоса сифатида айтиш мумкинки, синсетларни ишлаб чиқиш қилиш учун код яратиш табиий тилни қайта ишлаш усулларини ва машинали ўрганиш тамойилларини чуқур тушунишни талаб қиладиган мураккаб вазифадир. Муайян ечимни танлаш қўйилган

Фойдаланилган адабиётлар:

1. Mamatov, A. E. (2019). Zamonaviy lingvistika. Toshkent: Noshir, 135.
2. Abjalova, M., & Sharipov, E. (2021). O'ZBEK TEZAURUS LUGATI UCHUN SIFAT TURKUMI BAZASINI YARATISH MASALASI. COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS, 1(1).
3. Suyunov, B. (2021). Tezaurus va kompyuter texnologiyalariga doir. Computer Linguistics: Problems, Solutions, Prospects, 1(1).

вазифа(лар), мавжуд ресурс(лар)га ва қутилаётган аниқлик даражасига боғлиқ.

Синсетларни яратиш жараёни долзарб бўлиб, замонавий технологиялар билан уйғунлашган ҳолда янада такомиллаштирилиши мумкин. Уларнинг муваффақияти ва кенг қўлланилиши кўп ма'нолилик, контекстга боғлиқ маъно ва мураккаб терминология муаммоларини ҳал қилишга боғлиқ. Синсетлардан самарали фойдаланиш учун ҳар бир тилнинг ўзига хос хусусиятларини ҳисобга олган ҳолда аниқ стратегия ва ёндашувлар ишлаб чиқиш талаб этилади.

4. Abdullayeva, N. (2023, April). Tezaurus Tushunchasi, Etimologiyasi, Va Til O'rganishda Qo'llanilishi. In Conference on Applied and Practical Sciences (pp. 117-119).
5. <https://www.tensorflow.org/text/tutorials/word2vec>
6. <https://nlp.stanford.edu/projects/glove/>
7. [https://en.wikipedia.org/wiki/BERT_\(language_model\)](https://en.wikipedia.org/wiki/BERT_(language_model))
8. <https://www.nltk.org/>
9. <https://gensim2.github.io/>
10. <https://spacy.io/>
11. <https://www.tensorflow.org/>