СЕМАНТИК ТАҲЛИЛ ДАСТУРИ ВА АСОСИЙ ЭЛЕМЕНТЛАРИ

Мўминов Баҳодир Болтаевич
Тошкент давлат иқтисодиёт университети т.ф.д., профессор
Аллаберганова Насиба Мурадовна
Тошкент ахборот технологиялари университети, докторант

Аннотация: Ушбу мақолада семантик таҳлил дастури ва унинг асосий элементлари ўрганилган. Семантик таҳлил ёрдамида қандай муаммоларни ҳал қилиш мумкинлиги, табиий тилга ишлов беришда семантик таҳлил дастури, унинг назарий асослари ва амалий элементлари ҳақида маълумотлар берилган. Хусусан, лексик семантика, синтактик семантика, сўзнинг маълум бир вазиятда қўлланишига қараб ва онтологик семантик элементлар изоҳланган, ҳамда уларнинг NLP тизимида қўлланиши таҳлил қилинган. Семантик таҳлилнинг долзарблиги, айниқса кўп маънолилик (polysemy), омонимлик (homonymy) ва семантик боғлиқликларни аниқлаш келтирилган. Ўзбек тилидаги сўзларнинг тўғри маъносини аниқлаш учун Naïve Bayes, Word2Vec ва Cosine Similarity, BERT моделлари қиёсий таҳлил қилинди.

Калит сўзлар: *онтология, NLP, семантик таҳлил, синтактик семантика, сўзнинг маълум бир вазиятда қўлланишига қараб семантика, лексик семантика, WordNet, Word Sense Disambiguation.*

Аннотация: В данной статье изучена программа семантического анализа и её основные элементы. Приведена информация о том, какие проблемы можно решать с помощью семантического анализа, об использовании программы семантического анализа в обработке естественного языка, её теоретических основах и практических элементах. В частности, разъяснены лексическая семантика, синтаксическая семантика, использование слова в определённом контексте и онтологические семантические элементы, а также проанализировано их применение в системах NLP. Отмечена актуальность семантического анализа, особенно в решении задач многозначности (polysemy), омонимии (homonymy) и выявления семантических связей. Для определения точного значения слов в узбекском языке проведён сравнительный анализ моделей Naïve Bayes, Word2Vec, Cosine Similarity и BERT.

Ключевые слова: онтология, NLP, семантический анализ, синтаксическая семантика, контекстуальная семантика, лексическая семантика, WordNet, определение значения слова (Word Sense Disambiguation).

Annotation: This article explores a semantic analysis program and its main components. It provides information on the types of problems that can be addressed using semantic analysis, the role of semantic analysis in natural language

ISSN 3030-3907

processing (NLP), its theoretical foundations, and practical elements. In particular, lexical semantics, syntactic semantics, contextual semantics based on the usage of a word in a specific situation, and ontological semantic elements are explained and their application in NLP systems is analyzed. The relevance of semantic analysis is highlighted, especially in identifying polysemy, homonymy, and semantic relationships. To determine the correct meaning of words in the Uzbek language, a comparative analysis of the Naïve Bayes, Word2Vec, Cosine Similarity, and BERT models has been conducted.

Keywords: Ontology, NLP, semantic analysis, syntactic semantics, contextual semantics, lexical semantics, WordNet, Word Sense Disambiguation.

1. Кириш.

Сунъий интеллект технологияси ривожланиши билан машина ҳамда инсон ўртасидаги мулоқот долзарб масалалардан бири бўлиб келмоқда. Сунъий интеллект соҳасининг муҳим йўналишларидан бири табиий тилга ишлов бериш (Natural Language Processing, NLP) бўлиб, компютер тизимларига табиий тилни чуқур тушуниш, табиий тилдаги матннинг маъносини аниқлаш ва унга мос жавоблар яратиш ва уни машина томонидан қайта ишлашга техник шаклда ифодалаш имконини беради. Бу таҳлил орқали сўзлар ўртасидаги боғланишлар, сўзнинг маълум бир вазиятда қўлланишига қараб маъно аниқлиги ва тушуниш таъминланади [1]. Ушбу жараённинг энг муҳим босқичларидан бири бу семантик таҳлилдир. Бироқ, ўзбек тилида семантик таҳлил бўйича дастурий ечимлар етарли эмас. Тил технологияларининг давлат тилининг ривожлантиришдаги роли ортиб бораётган бир пайтда, ўзбек тилининг семантик таҳлили дастурини ишлаб чиқиш илмий ва амалий аҳамиятга эга деб ҳисоблаймиз.

Семантик таҳлил – табиий тилдаги матн ёки гапнинг маъносини тушуниш ва уни формал ифодалаш жараёнидир [2]. Бу жараён табиий тилга ишлов бериш доирасидаги муҳим йўналишлардан бири бўлиб, сунъий интеллектга инсон тилини чуқурроқ тушуниш ва унга мос равишда жавоб қайтариш имкониятини яратади. Масалан, “Олма дарахтдан тушди” жумласида “олма” – мева, “дарахт” – ўсимлик, “тушди” – ҳаракатни билдирувчи феъл. Семантик таҳлил бунда ҳар бир элементнинг онтологик турини ва улар ўртасидаги боғлиқликни аниқлайди. Семантик таҳлил ёрдамида матннинг аниқ ва тўғри маъносини тушуниш, синоним ва омонимларни фарқлаш, сўзнинг маълум бир вазиятда қўлланишига қараб мос маънони аниқлаш, матндаги сўзлар ўртасидаги семантик боғланишларни аниқлаш, нутқни тушуниш ва автоматлаштирилган равишда жавоб қайтариш каби кўплаб муаммоларни ҳал қилади.

ISSN 3030-3907

2. Асосий қисм.

Семантик таҳлил — бу матн, сўз ёки ибораларнинг чуқур маъно жиҳатидан таҳлил қилиниши бўлиб, у табиий тил бирликларининг мазмунини аниқлаш, сўзнинг маълум бир вазиятда қўлланишига қараб қандай ишлатилишини тушуниш ва шу орқали янада аниқлаштирилган, тушунарли натижалар олиш имконини беради. Одатда, матн ёки тил бирликларининг сатҳини таҳлил қилиш синтактик (грамматик) ёки морфологик (сўз шакллари) жиҳатдан чекланган бўлса, семантик таҳлил мазмуний чуқурликни очиб беради. Бу, айниқса, машинага матнни тўғри тушунтириш, саволларга мантиқий жавоб қайтариш ёки автоматик таржима жараёнларида жуда катта аҳамият касб этади. Шунингдек, семантик таҳлил инсон миясининг тушуниш механизмини сунъий интеллект тизимларига яқинлаштиришга хизмат қилади. Бу орқали ахборотларни сўзнинг маълум бир вазиятда қўлланишига қараб тўғри англаш, маънодошлик, антонимлик, кўп маънолилик, ноаниқликни бартараф этиш мумкин бўлади.

Семантик таҳлил лексик семантика, синтактик семантика, сўзнинг маълум бир вазиятда қўлланишига қараб семантика, семантик муносабатлар, семантик тармоқлар, семантик анотация ва онтология каби асосий элементлардан иборат [3].



1-расм. Семантик таҳлил элементлари

2.1. ***Лексик семантика*** сўзларнинг маъноси, уларнинг ўзаро муносабатлари ва сўзнинг маълум бир вазиятда қўлланишига қараб қандай тушунилишини ўрганадиган тилшунослик ва табиий тилга ишлов бериш (NLP) бўлимидир. Лексик семантика сўзларнинг луғавий маънолари,

ISSN 3030-3907

уларнинг синонимия, антонимия, омонимия каби жиҳатларини таҳлил қилади.

Буни математик модель кўринишида қуйидагича ифодалаш мумкин:

$$S_T=\{L_S + S_S + C_S + S_R + S_N + S_A + O_N\}$$

Бу ерда: S_T — Семантик таҳлил, L_S — Лексик семантика, S_S — Синтактик семантика, C_S— Контекстуал семантика (сўзнинг маълум бир вазиятда қўлланишига қараб семантика), S_R — Семантик муносабатлар, S_N — Семантик тармоқлар, S_A — Семантик анотация, O_N — Онтология.

Лексик семантика қуйидаги асосий масалаларни ўрганади: сўзнинг асосий ва кенгайтирилган маънолари, сўзлар орасидаги семантик боғланишлар, синоним, антоним ва кўп маънолилик (polysemy) тушунчалари, сўзнинг маълум бир вазиятда қўлланишига қараб маъносини аниқлаш (Word Sense Disambiguation – WSD). Бу элементлар тилни аниқ англаш учун дастлабки таҳлил манбаи бўлиб хизмат қилади [4].

Лексик семантика сўзларнинг маъносини таҳлил қилиш билан шуғулланади. Лексик семантикада сўзнинг маъноси қуйидаги турларга бўлинади: **денотатив маъно** – сўзнинг луғавий ёки асосий маъноси. Масалан, *"китоб"* сўзи – босма ёки электрон шаклда мавжуд бўлган матнлар тўплами; **коннотатив маъно** – сўзнинг эмотсионал ёки қўшимча маънолари. Масалан, *"қоронғи"* сўзи баъзида негатив маънода (*"қоронғи фикрлар"*) ишлатилиши мумкин; **кўчма маъно (Метафора, Метонимия)** – сўзнинг ташқи ёки ички ўхшашлик асосида бошқа нарсани билдириши. Масалан, *"тилла қўл"* – маҳоратли одам деган маънода ишлатилади.

Синонимлар сўзларни яхши тушуниб олиш учун уни мисоллар орқали тушунтирамиз. Масалан, *катта* сўзини оладиган бўлсак, унинг бир нечта синонимлари бор. Булар: *улкан*, *йирик*, *баҳайбат* ва ҳоказо. Синонимлар бир-бирига яқин маъноли сўзлардир, лекин улар услубий ёки нозик семантик фарқларга эга бўлиши мумкин**.**

Тўлиқ синонимлар – маъноси бутунлай бир хил бўлган сўзлар. Масалан, *"шифокор"* ва *"доктор"*.

Қисман синонимлар – маъноси ўхшаш, лекин сўзнинг маълум бир вазиятда қўлланишига қараб фарқланиши мумкин. Масалан, *"катта"* ва *"улкан"*.

Антонимлар. Антоним сўзларга мисол қилиб, *иссиқ* ва *совуқ* сўзларини олишимиз мумкин. Яъни бу икки сўз бир-бирига зид маъноли бўлганлиги учун ҳам антоним сўзлар дейилади. Антонимлар бир-бирига зид маъноли сўзлардир. Улар икки хил бўлади:

ISSN 3030-3907

Градатсион антонимлар – оралиқ даражага эга бўлган антонимлар (*“иссиқ – совуқ”*, чунки “илиқ” оралиқ ҳолатда бўлиши мумкин).

Мутлақ антонимлар – бутунлай зид тушунчалар (*“тириклик – ўликлик”*, *“кун – тун”*).

Кўп маънолилик (Polysemy) – бир сўзнинг маълум бир вазиятда қўлланишига қараб ҳар хил маъноларга эга бўлиши. Кўп маъноли сўзга мисол қилиб *кўз* сўзини олишимиз мумкин. Кўз сўзи инсоннинг кўриш аъзоси сифатида эътироф этилади, лекин бу сўзни бошқа гапларда бошқа маънода ҳам қўллаш мумкин. Масалан, Бу бола ишни кўзини биладиган бола экан. Бу ерда ишни кўзини билади деганда унинг маҳоратли эканлиги эътироф этилмоқда.

Омонимлар (Homonyms) – талаффузи ёки ёзилиши бир хил, лекин турли маъноларга эга бўлган сўзлар.

Талаффузи бир хил, лекин ёзилиши ҳар хил (*Homophones)* – *“гул” (ўсимлик) ва “ғул” (шовқин)*

Ёзилиши бир хил, лекин маъноси фарқли (*Homographs*) – *“чўқинди” (қушларнинг чўқиши) ва “чўкинди” (қолдиқ, қолдиқ модда)*

Мисол:

“Бош” – инсоннинг бош қисми

“Бош” – бирор ишнинг бошланиши

“ёз” – йил фасли,

“ёз” – ёзиш феъли.

Кўп маънолиликда сўз бир нечта маънога эга бўлади, лекин улар бир-бирига яқин бўлиши мумкин.

Мисол:

“Калит” – эшик очиш учун асбоб

“Калит” – муҳим ечим

Лексик семантика табиий тилни қайта ишлашнинг муҳим йўналиши бўлиб, сўзларнинг маъносини тушуниш, уларни сўзнинг маълум бир вазиятда қўлланишига қараб таҳлил қилиш ва автоматик қайта ишлаш имконини беради. Ушбу йўналиш машинали ўрганиш ва сунъий интеллект билан интегратсия қилиниб, NLP тизимларининг самарадорлигини оширади.

2.2. Синтактик Семантика

Бу элемент гап тузилишини ва сўзларнинг ўзаро боғлиқлигини таҳлил қилади.

Парсинг (Гап синтаксисни аниқлаш) – гапдаги сўзлар орасидаги грамматик муносабатларни топиш.

ISSN 3030-3907

Синтактик таҳлил воситалари ёрдамида матнда мавжуд бўлган грамматик тузилма семантик интерпретацияга тайёр ҳолга келтирилади. Хусусан, dependency tree (боғланиш дарахти) ва constituency tree (таркибий дарахт) каби синтактик тузилмалар асосида гап таркибидаги субъект, предикат ва объект каби грамматик функцияли компонентлар ажратиб олинади. Бу эса матннинг маъносини компьютер томонидан тушуниш ва семантик жиҳатдан талқин қилиш учун муҳим пойдевор ҳисобланади [5].

Грамматика асосида семантик таҳлил – сўзлар ва гаплар маъноси грамматик қоидалар асосида тушунтирилади

2.3. Сўзнинг маълум бир вазиятда қўлланишига қараб Семантика

Сўз ёки иборани аниқ сўзнинг маълум бир вазиятда қўлланишига қараб тушуниш учун ишлатилади. Сўзнинг маълум бир вазиятда қўлланишига қараб сўзнинг маъноси аниқланади. “Bank” сўзи молиявий муассаса ёки дарё соҳилини англатиши мумкин. Ушбу ҳолатни ечиш учун Word Sense Disambiguation (WSD) алгоритмлари, жумладан Леск алгоритми ёки машина ўрганиш асосидаги моделлар қўлланилади [6].

Матн сўзнинг маълум бир вазиятда қўлланишига қараб – сўзнинг яқин атрофидаги бошқа сўзлар унинг маъносини аниқлайди.

Маъно туманлиги (Word Sense Disambiguation, WSD) – турли сўзнинг маълум бир вазиятда қўлланишига қараб сўзнинг қайси маъноси мос келишини аниқлаш.

Word Sense Disambiguation (WSD) – Сўзнинг тўғри маъносини аниқлаш

Word Sense Disambiguation (WSD) – сўзнинг тўғри маъносини аниқлаш учун турли математик моделлар ва алгоритмлар мавжуд. Ушбу жараён табиий тилни қайта ишлашда муҳим бўлиб, турли сўз маънолари орасидан сўзнинг маълум бир вазиятда қўлланишига қараб мосини танлашни таъминлайди. Қуйида WSD учун асосий математик моделларни келтирилган [7].

2.3.1. Bayes Назарияси асосида Word Sense Disambiguation (Naïve Bayes Слассифиер)

Bayes классификатори шундай моделки, унда берилган сўзнинг тўғри маъносини танлаш учун **шартли эҳтимоллик** ҳисобланади [8].

Фараз қиламиз:

Агар S_i– сўзнинг i-маъноси бўлса ва C – сўзнинг маълум бир вазиятда қўлланиши бўлса, Bayes теоремаси асосида қуйидагича ифодаланади:

$$P(S_i|C) = \frac{P(C|S_i)P(S_i)}{P(C)}$$

Бу ерда:

ISSN 3030-3907

- $P(S_i|C)$ –C сўзнинг маълум бир вазиятда қўлланишига қараб сўзнинг S_i маъносига эга бўлиш эҳтимоли
- $P(C|S_i)$ – S_i маъноси учун ушбу сўзнинг маълум бир вазиятда қўлланишига қараб юзага келиш эҳтимоли
- $P(S_i)$ – сўзнинг ушбу маънода ишлатилиш эҳтимоли (одатда корпусдан олинади)
- $P(C)$ – сўзнинг маълум бир вазиятда қўлланишига қараб умумий эҳтимоли

Самарали ечим танлов:

$$\hat{S} = arg^{max}_{S_i} P(\boldsymbol{S_i}|\boldsymbol{C})$$

Бу усул юқори аниқликка эга бўлса-да, сўзнинг маълум бир вазиятда қўлланишига қараб маълумотлар сони ошган сайин самарадорлиги пастлашини кўрсатади [9].

2.3.2. Word2Vec va Cosine Similarity асосида Word Sense Disambiguation

Word2Vec моделида сўзлар **вектор** сифатида ифодаланади [10]. Ҳар бир маъно учун турли сўз векторларини ишлатиш мумкин ва сўзнинг маълум бир вазиятда қўлланишига қараб энг яқин вектор танланади.

Фараз қиламиз сўзнинг турли маънолари S_i учун *Word2Vec* векторлари мавжуд бўлсин:

$$\overrightarrow{w_i} = Word2Vec(S_i)$$

Сўзнинг маълум бир вазиятда қўлланишига қараб сўзлар учун ҳам ўртача вектор олинади:

$$\vec{C} = \frac{1}{|C|}\sum_{w \in C} \vec{w}$$

Cosine Similarity орқали энг мос келувчи сўз танланади:

$$\cos(\theta) = \frac{\vec{w}_i \cdot \vec{C}}{\|\overrightarrow{w_i}\|\|\vec{C}\|}$$

Энг катта cosine similarity қийматига эга бўлган S_i танланади.

Word2Vec семантик жиҳатдан яқинликни яхши ифодалаган бўлса-да, сўзнинг маълум бир вазиятда қўлланишига қараб ноаниқликларга заиф [11].

2.3.3. Трансформер Моделлар (BERT) асосида Word Sense Disambiguation

BERT (Bidirectional Encoder Representations from Transformers) сўзнинг маъносини танлаш учун жуда самарали модел бўлиб, сўзнинг маълум бир вазиятларда қўлланишига қараб динамик векторларни яратади.

BERT– bidirectional self-attention механизми ёрдамида сўзнинг маълум бир вазиятда қўлланишига қараб ҳар ҳил йўналишда таҳлил қилади [12]. Ҳар

ISSN 3030-3907

бир сўзнинг семантик вектори сўзнинг маълум бир вазиятда қўлланишига қараб динамик равишда яратилади.

Фараз қиламиз: BERT **self-attention** механизми ёрдамида сўзларни таҳлил қилади:

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Бу ерда:

Q, K^T, V – **query, key, value** *векторлари*

d_k– вектор ҳажми

A – сўзлар ўртасидаги боғланишларни ҳисоблайдиган матритса

BERT моделининг натижаси орқали ҳар бир **сўзнинг маълум бир вазиятда қўлланишига қараб мос келувчи** семантик вектори олинади.

Трансформер архитектураси Word Sense Disambiguation учун энг илғор ёндашувлардан бири ҳисобланади [13].

2.4. Семантик Тармоқлар

Бу усул семантик боғланишларни визуал шаклда кўрсатишга имкон беради.

WordNet – сўзларни маъно гуруҳларига ажратиб, уларнинг семантик боғлиқлигини аниқловчи тармоқ.

WordNet – инглиз тилидаги лексик маълумотлар базаси бўлиб, сўзларни синонимлар тўплами (synset) кўринишида ташкил этади. Ўзбек тили учун бу каби ресурслар чекланган бўлса-да, сўнгги йилларда UzWordNet лойиҳаси устида ишлар олиб борилмоқда [15].

FrameNet – сўзлар ва уларнинг семантик ролларини таҳлил қилувчи модел.

2.5. Семантик Анотация ва Онтологиялар

Онтологиялар маълум бир предмет соҳаси бўйича тушунчалар тўплами ва уларнинг ўзаро боғлиқликларини белгиловчи моделлар.

Онтология – бу тушунчалар ва уларнинг муносабатларини формал структурада тасвирлайдиган моделдир. Семантик аннотация эса матнларга маълумотларни боғлаб, машина томонидан таҳлил қилиш имконини яратади. Масалан, “John Apple компаниясида ишлайди” жумласида “John” – инсон, “Apple” – ташкилот сифатида белгиланиши мумкин [16].

Мисол: Тиббиёт онтологияси: “касаллик → симптом → даволаш усули” занжири

Семантик аннотация — бу матн бирликларини расмийлаштирилган семантик маълумотлар билан белгилаш жараёни бўлиб, сўзлар ва уларнинг контекстдаги маънолари сунъий интеллект тизимлари томонидан қайта ишлаш учун техник тайёр ҳолга келтирилади.

ISSN 3030-3907

Семантик аннотация – ҳужжатларга ёки матнларга семантик маълумотларни қўшиш.

3. Семантик таҳлилнинг амалий қўлланилиши

Семантик таҳлил NLP ва сунъий интеллектнинг турли соҳаларида қўлланилади:

– *Маълумотларни қидириш тизимлари* – сўровларни аниқ тушуниш ва энг мос жавобни топиш

– *Маълумотларни экстракция қилиш* – матндан муҳим фактларни ажратиб олиш

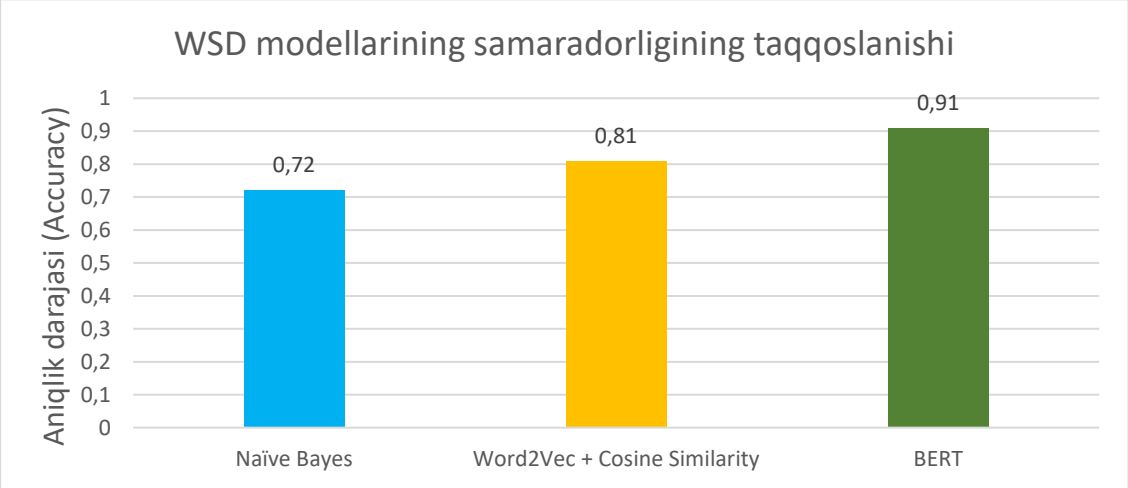
– *Машинали таржима* – сўзларнинг маъносига асосланган сифатли таржима яратиш

– *Чатбот ва овозли ёрдамчилар* – инсон билан табиий мулоқот қилиш

– *Автоматлаштирилган хулоса чиқариш* – катта ҳажмдаги матнларни қисқа ва мазмунли хулосаларга ажратиш

4. Хулоса

Bayes klassifikatori – статистик услуб, лекин катта маълумотлар талаб қилади. **Word2Vec ва Cosine Similarity** – семантик жиҳатдан мос келадиган маънони танлайди. **BERT ва Transformer моделлар** – энг замонавий ва кучли модел. **Натижада, Word Sense Disambiguation NLP**да муҳим рол ўйнайди ва турли методлар билан амалга оширилиши мумкин. Агар ресурслар мавжуд бўлса, **BERT каби Transformer моделларини ишлатиш энг самарали бўлади.**



1-расм. Word Sense Disambiguation (WSD) бўйича турли моделларнинг самарадорлик даражалари (аниқлик)

Графикдан кўриниб турибдики:

Naïve Bayes модели содда ва асосий статистик ёндашувга эга бўлиб, аниқлиги нисбатан паст – *0.72*.

ISSN 3030-3907

Word2Vec + Cosine Similarity ёндашуви семантик боғланишни яхши ифодалайди ва аниқликни оширади – *0.81.*

BERT модели сўзнинг маълум бир вазиятда қўлланишига қараб модел сифатида энг юқори аниқликка – *0.91* – эришмоқда, бу уни WSD учун энг самарали ёндашув сифатида кўрсатади.

Сўз маъноларини ажратиш — табиий тилни қайта ишлашда (NLP) ҳал қилувчи босқичлардан бири ҳисобланади. Word Sense Disambiguation (WSD) методлари сўзнинг тўғри маъносини аниқлаш орқали семантик таҳлил сифатини оширади. Натижада WSD учун қўлланиладиган асосий ёндашувлар: эҳтимоллик моделларига асосланган Bayes классификатори, семантик вектор моделларига асосланган Word2Vec ва Cosine Similarity ёндашуви ҳамда чуқур ўрганиш усулларига асосланган BERT модели таҳлил қилинди.

Bayes классификатори содда ва тушунарли математик асосга эга бўлсада, сўзнинг маълум бир вазиятда қўлланишига қараб сўзлар ўртасидаги мураккаб семантик боғланишларни чуқур таҳлил қилишда бу моделнинг имкониятлари бироз чегараланган. Word2Vec эса семантик алоқадорликни яхши ифодалаши билан ажралиб туради, бироқ сўзнинг маълум бир вазиятда қўлланишига қараб динамик жавоб бера олмайди. BERT модели эса селф-аттентион механизми орқали ҳар бир сўз учун сўзнинг маълум бир вазиятда қўлланишига қараб боғлиқ бўлган динамик векторларни ҳосил қилади, бу эса WSD жараёнида юқори аниқликни таъминлайди.

Таҳлил натижаларига кўра, замонавий сўзнинг маълум бир вазиятда қўлланишига қараб моделлар — хусусан, трансформер архитектурасига асосланган BERT ва GPT сингари моделлар — WSD вазифаларида юқори натижаларни кўрсатмоқда. Шунингдек, ушбу моделларни ўзбек тили учун мослаштириш ва локал корпусларга мос синовдан ўтказиш келгусидаги истиқболли йўналишлардан бири сифатида белгиланиши мумкин.

Шундай қилиб, турли математик моделлар ва чуқур ўрганиш асосидаги ёндашувларнинг WSD вазифасида қўлланилиши табиий тилни семантик таҳлил қилишда самарадорликни ошириш имконини беради ва ўзбек тили учун ҳам бу ёндашувлар кенг имкониятлар яратади.

Адабиётлар рўйхати:

1. Jurafsky D., Martin J. H. Speech and Language Processing. – 3-ed. – Stanford University, 2023.
2. Y. -S. Li, T. M. Cao, S. -T. Wang, X. Huang, C. -H. Hsu and P. -C. Lin, "A Resource-Constrained Asymmetric Redundancy Elimination Algorithm,"

ISSN 3030-3907

in IYEYEYE/ACM Transactions on Networking, vol. 23, no. 4, pp. 1135-1148, Aug. 2015, doi: 10.1109/TNET.2014.2322889

3. Navigli R. Word Sense Disambiguation: A Survey // ACM Computing Surveys. – 2009. – Vol. 41, No. 2. – P. 1–69. DOI: https://doi.org/10.1145/1459352.1459355

4. Cruse D. A. Lexical Semantics. – Cambridge: Cambridge University Press, 1986. – 328 p.

5. Manning C. D., Schütze H. Foundations of Statistical Natural Language Processing. – Cambridge, MA: MIT Press, 1999. – 620 p.

6. Lesk M. Automatic sense disambiguation using machine readable dictionariyes // SIGDOC '86 Proceyedings. – 1986. – P. 24–26.

7. Navigli R. Word Sense Disambiguation: A Survey // ACM Computing Surveys. – 2009. – Vol. 41, No. 2. – P. 1–69. https://doi.org/10.1145/1459352.1459355

8. Jurafsky D., Martin J.H. Speyech and Language Processing. – Pearson, 2023. – Chapter 20: Naïve Bayes and WSD.

9. Pedersen T. A simple approach to building ensembles of Naive Bayesian classifiyers for word sense disambiguation // NAACL. – 2000. – P. 63–69. https://aclanthology.org/A00-2010

10. Mikolov T., Chen K., Corrado G., Dean J. Efficiyent Estimation of Word Representations in Vector Space // arXiv preprint. – 2013. – arXiv:1301.3781. https://arxiv.org/abs/1301.3781

11. Camacho-Collados J., Pilehvar M.T. From Word to Sense Embeddings: A Survey on Vector Representations of Meaning // Journal of Artificial Intelligence Research. – 2018. – Vol. 63. – P. 743–788. https://www.jair.org/index.php/jair/article/viyew/11259

12. Devlin J., Chang M.-W., Leye K., Toutanova K. BERT: Pre-training of Deyep Bidirectional Transformers for Language Understanding // arXiv preprint. – 2019. – arXiv:1810.04805. https://arxiv.org/abs/1810.04805

13. Rogers A., Kovaleva O., Rumshisky A. A Primer in BERTology: What We Know About How BERT Works // Transactions of the Association for Computational Linguistics (TACL). – 2020. – Vol. 8. – P. 842–866. https://doi.org/10.1162/tacl_a_00349

14. Fellbaum C. WordNet: An Electronic Lexical Database. – Cambridge, MA: MIT Press, 1998.

15. Karimov B., Muminov K. Developing WordNet-like lexical database for Uzbek // Proceyedings of the 13th Global WordNet Conference. – 2022.

ISSN 3030-3907

16. Gruber T. R. A Translation Approach to Portable Ontology Specifications // Knowledge Acquisition. – 1993. – Vol. 5(2). – P. 199–220.

ISSN 3030-3907