KONSEPTUAL FARAZLARNI SOLISHTIRISHGA ASOSLANGAN ANTIPLAGIAT TIZIMI

Egamberdiyev Elyor Hayitmamatovich

Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti, PhD katta o‘qituvchi
E-mail: elyor.egamberdiyev88@gmail.com

K E Y W O R D S

konseptual faraz, semantik o‘xshashlik, antiplagiat tizimi, plagiatni aniqlash, matnlarni tahlil qilish, ontologiya, kognitiv semantika

A B S T R A C T

Ushbu maqolada matnli hujjatlarda plagiat holatlarini aniqlash uchun konseptual farazlarni solishtirishga asoslangan antiplagiat tizimi taklif etiladi. An’anaviy antiplagiat tizimlari asosan leksik yoki sintaktik o‘xshashlikni aniqlashga yo‘naltirilgan bo‘lib, mazmuni saqlangan holda ifoda shakli o‘zgartirilgan, ya’ni parafrazlangan matnlarni aniqlashda cheklangan imkoniyatlarga ega. Mazkur tadqiqot ushbu muammoni bartaraf etish maqsadida inson tafakkurida shakllanadigan semantik tasavvurlarni (konseptual farazlarni) modellashtirishga asoslangan yondashuvni taklif etadi. Taklif etilgan metodda har bir gap yoki hujjatdan asosiy obyektlar, ularning xususiyatlari va o‘zaro munosabatlarini ifodalovchi konseptlar ajratib olinadi hamda ular yagona ontologik fazoda normallashtiriladi. Hosil qilingan konseptual farazlar to‘plami asosida hujjatlar orasidagi semantik o‘xshashlik darajasi aniqlanadi. Ushbu yondashuv matnlar o‘rtasida to‘g‘ridan-to‘g‘ri so‘z mosligi mavjud bo‘lmagan holatlarda ham mazmuniy yaqinlikni aniqlash imkonini beradi. Tajriba natijalari konseptual farazlarga asoslangan model parafrazlangan plagiatni aniqlashda an’anaviy usullarga nisbatan yuqori aniqlik ko‘rsatkichlariga ega ekanligini ko‘rsatdi. Taklif etilgan antiplagiat tizimi ilmiy tadqiqotlar, ta’lim muassasalari hamda elektron hujjatlar bilan ishlovchi axborot tizimlarida akademik halollikni ta’minlashda samarali qo‘llanishi mumkin.

Kirish

Zamonaviy ta’lim va ilmiy faoliyatda raqamli matnli hujjatlar hajmi keskin ortib bormoqda. Shu bilan birga, mualliflik huquqlarining buzilishi, ya’ni plagiat holatlarining ko‘payishi akademik hamjamiyat uchun dolzarb muammoga aylangan. Maxsus antiplagiat tizimlari yordamida hujjatlar orasidagi o‘xshashlik aniqlansa-da, an’anaviy yondashuvlar asosan leksik va sintaktik ko‘rsatkichlarga tayanganligi sababli, mazmuni saqlangan holda so‘z va gap tuzilishi o‘zgartirilgan, ya’ni parafrazlangan matnlarni aniqlashda yetarli samaradorlikni ta’minlay olmaydi [1, 2].

Ushbu muammoni hal etish uchun tadqiqotda inson tafakkurida shakllanadigan semantik tasavvurlar (konseptual farazlar)ni modellashtirishga asoslangan yondashuv taklif qilinadi. Bu yondashuv gap yoki hujjatdan asosiy obyektlar, ularning xususiyatlari va o‘zaro munosabatlarini ajratib olib, ularni ontologik fazoda normallashtirish orqali hujjatlar orasidagi semantik o‘xshashlikni aniqlash imkonini beradi [3]. Natijada, leksik jihatdan turli bo‘lgan, ammo mazmuni bir xil bo‘lgan matnlar ham samarali tarzda aniqlanishi mumkin.

Shuningdek, o‘zbek tilidagi matnlar uchun mo‘ljallangan antiplagiat tizimlarining

© Elyor Egamberdiyev

rivojlanmaganligi va tilning morfologik hamda sintaktik xususiyatlari ushbu tadqiqotni dolzarb qiladi. Shu sababli, konseptual farazlarga asoslangan semantik antiplagiat tizimi nafaqat akademik halollikni ta’minlashda, balki ta’lim va ilmiy faoliyatda elektron hujjatlar bilan ishlovchi tizimlarning samaradorligini oshirishda ham muhim ahamiyatga ega.

Mazkur maqolada taklif etilgan yondashuvning metodologiyasi, konseptual farazlarni hosil qilish va semantik o‘xshashlikni hisoblash algoritmlari batafsil tahlil qilinadi hamda uning samaradorligi tajriba natijalari asosida baholanadi.

Metodologiya

Mazkur tadqiqotda plagiatni aniqlash uchun konseptual farazlarni solishtirishga asoslangan semantik yondashuv qo‘llaniladi. Modelning asosiy g‘oyasi shundan iboratki, gap yoki hujjat inson ongida hosil bo‘lgan tasavvur (conceptual representation) orqali ifodalanadi va plagiat shu tasavvurlar o‘xshashligi bilan aniqlanadi.

Dastlab gapdan konseptual faraz hosil qilish jarayonini ko‘rib chiqamiz. Buning uchun har bir gap S quyidagi funksiyaga o‘tadi:

$$F(S) = \{C_1, C_2, \ldots, C_k\}$$

bunda:

$F(S)$ – gapdan hosil bo‘lgan konseptual farazlar to‘plami, C_i gapdagi asosiy konseptlar (obyekt, xususiyat, munosabat).

Masalan: S_1: “Mening uyumda oppoq itim bor”

$$F(S_1) = \{uy, it, oq, egalik\}$$

S_1: “Men har kuni uyga borganimda meni oppoq itim vov vov deb kutib oladi”

$$F(S_2) = \{uy, it, oq, egalik, harakat\}$$

S_1 va S_2 matnlar leksik jihatdan har xil bo‘lishiga qaramay, asosiy konseptlar bir xil. Bular *“uy”, “it”, “oq”, “egalik”.*

Keyingi qadamda konseptlarni ontologik normallashtirish uchun har bir so‘z yoki ibora mos keladigan ontologik konseptga transformatsiya qilinadi:

$$C = g(w)$$

bunda w so‘z yoki ibora, g(·) – ontologik xaritalash funksiyasi. Natijada turli ifodalar (masalan, “vov vov deb kutib oladi” → “HARAKAT”) yagona konsept sifatida ko‘riladi.

Keyingi qadamda tasavvurlar asosida semantik o‘xshashlikni hisoblash uchun iboralar yoki gaplar orasidagi semantik o‘xshashlikni Jaccard indeksi yordamida aniqlash mumkin:

$$SemSim_{concept}(S_1, S_2) = \frac{|F(S_1) \cap F(S_2)|}{|F(S_1) \cup F(S_2)|}$$

Ushbu formularni yuqorida keltirilgan misolda qo‘llab ko‘radigan bo‘lsak quyidagicha natija chiqishini kuzatish mumkin.

$$|F(S_1) \cap F(S_2)| = 4 \rightarrow \{uy, it, oq, egalik\}$$

$$|F(S_1) \cup F(S_2)| = 5 \rightarrow \{uy, it, oq, egalik, harakat\}$$

$$SemSim_{concept}(S_1, S_2) = \frac{4}{5} = 0.8$$

Ushbu natija yuqori semantik o‘xshashlikni ko‘rsatadi.

Ibora yoki gapdagi barcha konseptlar semantik o‘xshashlikni hisoblashda ishlatilishi mumkin, lekin ba’zi konseptlar mazmunan asosiy bo‘lmay, faqat qo‘shimcha ma’lumot sifatida qatnashadi (masalan, vaqt, holat yoki tafsilotlar). Shu sababli, plagiatni aniqlashda faqat asosiy konseptlar (yadro konseptlar)ni hisobga olish zarurati paydo bo‘ladi. Asosiy konseptlarni hisobga olish quyidagilar bilan asoslanadi.

© Elyor Egamberdiyev

- Mazmuniy aniqlikni oshiradi: faqat gapning asosiy g‘oyasi va obyektlari solishtiriladi.
- Shovqinni kamaytiradi: qo‘shimcha yoki tasodifiy tafsilotlar semantik o‘xshashlikni buzmaydi.
- Parafraz qilingan matnlarni to‘g‘ri aniqlashga yordam beradi: gap tuzilishi yoki so‘zlar o‘zgargan bo‘lsa ham, yadro konseptlar bir xil bo‘lsa, semantik o‘xshashlik yuqori bo‘ladi.

Buning uchun *Core(F)* to‘plamidagi asosiy konseptlar hisobga olinishi kerak.

$$SemSim_{core}(S_1, S_2) = \frac{|Core(F(S_1)) \cap Core(F(S_2))|}{|Core(F(S_1)) \cup Core(F(S_2))|}$$

Keyingi qadamda Semantik o‘xshashlik qiymati chegaraviy parametr λ bilan solishtiriladi:

$$P = \begin{cases} 1, & agar\ SemSim_{core}(S_1, S_2) \geq \lambda \\ 0, & aks\ holda \end{cases}$$

bunda P=1 – plagiat mavjud, P=0 – plagiat aniqlanmadi, λ – tajriba asosida tanlangan semantik chegaraviy qiymat.

Natijalar.

Tizimning samaradorligi turli turdagi matnlar bilan sinovdan o‘tkazildi. Sinovlarda to‘g‘ridan-to‘g‘ri plagiat holatlari, parafrazlangan matnlar va to‘liq o‘zgacha matnlar tizimga kiritildi. Natijalar ko‘rsatdiki, yadro konseptlar asosidagi semantik o‘xshashlik moduli tizimga yuqori aniqlik bilan plagiat holatlarini aniqlash imkonini beradi. To‘g‘ridan-to‘g‘ri plagiat matnlar uchun semantik o‘xshashlik qiymati 0,95 darajasida bo‘lib, tizim tomonidan plagiat mavjudligi to‘g‘ri belgilanadi. Parafrazlangan matnlar, ya’ni so‘zlar yoki jumla tuzilishi o‘zgartirilgan, lekin mazmuni saqlangan hujjatlar uchun SemSim_core qiymati o‘rtacha 0,85 bo‘lib, tizim ularni ham plagiat sifatida aniqlashda yuqori sezuvchanlik ko‘rsatdi.

Shuningdek, qisman o‘xshash matnlar uchun semantik o‘xshashlik 0,60 darajasida aniqlanib, tizim ularni plagiat sifatida belgilamadi, bu esa tizimning noaniq o‘xshashliklarni noto‘g‘ri aniqlamaslik qobiliyatini ko‘rsatadi. To‘liq o‘zgacha matnlar esa past SemSim_core qiymatlari (taxminan 0,20) bilan xarakterlanadi va tizim ularni plagiat mavjud emas deb to‘g‘ri baholaydi.

Natijalar shuni ko‘rsatadiki, konseptual farazlarga asoslangan semantik yondashuv an’anaviy leksik yoki sintaktik usullarga nisbatan parafrazlangan va mazmun jihatdan o‘xshash matnlarni aniqlashda ancha yuqori samaradorlikka ega. Shu tarzda ishlab chiqilgan tizim akademik matnlar, ilmiy maqolalar va elektron hujjatlar orasida plagiatni aniqlashda ishonchli vosita sifatida ishlatilishi mumkin.

Xulosa.

Taklif etilgan antiplagiat tizimi konseptual farazlarga asoslangan semantik yondashuvni qo‘llagan holda ishlab chiqilgan va tajriba sinovlaridan o‘tkazilgan. Sinov jarayonida turli darajadagi plagiat holatlari, shu jumladan parafrazlangan matnlar ham tizimga kiritildi. Natijalar ko‘rsatdiki, tizim leksik jihatdan farqli bo‘lgan, ammo mazmuni bir xil bo‘lgan matnlarni ham yuqori aniqlik bilan aniqlay oladi. Ayniqsa yadro konseptlar asosidagi semantik solishtirish mexanizmi tizimning aniqligini oshirishga katta hissa qo‘shdi, chunki asosiy mazmunni ifodalovchi konseptlar tafsilotlar va qo‘shimcha elementlardan mustaqil tarzda tahlil qilinadi.

Eksperimental sinovlar natijasida tizim odatiy leksik-sintaktik yondashuvlarga qaraganda sezilarli darajada yuqori sezuvchanlik va aniqlik ko‘rsatdi. Shu bilan birga, tizim tomonidan hosil qilingan semantik o‘xshashlik qiymatlari λ\lambdaλ chegarasi bilan solishtirilganda plagiat holatlari aniq belgilandi, bu esa tizimning amaliy samaradorligini ta’minladi.

Bundan tashqari, tizim o‘zbek tilidagi matnlar bilan ishlashda ham yuqori samaradorlikni ko‘rsatdi, chunki ontologik xaritalash va konseptual farazlarni aniqlash mexanizmi tilning morfologik va sintaktik murakkabliklarini hisobga oladi. Shu tarzda ishlab chiqilgan antiplagiat tizimi akademik matnlar, ilmiy maqolalar va elektron hujjatlar

(cc) BY © Elyor Egamberdiyev

orasida plagiatni aniqlashda samarali vosita sifatida ishlatilishi mumkin.

Adabiyotlar.

1. Maurer H., Kappe F., Zaka B. *Plagiarism — A Survey*. Journal of Universal Computer Science, 2006.
2. Clough P. *Plagiarism in Natural and Programming Languages: An Overview of Current Tools and Technologies*. Research Memorandum, 2000.
3. Jurafsky D., Martin J.H. *Speech and Language Processing*. Pearson Education, 2019.
4. Abduvaliyev A., Rahimov S. *O‘zbek tilini avtomatik qayta ishlash asoslari*. Toshkent: Fan, 2020.
5. Ilyas, M., Malik, N., Bilal, A., Razzaq, S., Maqbool, F., & Abbas, Q. Plagiarism Detection Using Natural Language Processing Techniques. *Technical Journal*, Vol. 26 No. 01, pp. 90–101, 2021.
– NLP asosida plagiatni aniqlash metodologiyasi va chuqur semantik tahlil haqida.
6. Al-Shamery, E. S., & Gheni, H. Q. Plagiarism Detection using Semantic Analysis. *Indian Journal of Science and Technology*, Vol. 9, Issue 1, 2016.
– Semantik tahlilga asoslangan plagiatni aniqlash usuli.
7. Weighted semantic plagiarism detection approach based on AHP decision model. *PubMed*, 2021.
– Semantik va strukturaviy o‘xshashlikni hisoblash orqali plagiatni aniqlash bo‘yicha tadqiqot.
8. On retrieving intelligently plagiarized documents using semantic similarity. *Engineering Applications of Artificial Intelligence*, Vol. 45, Oct 2015, pp. 246–258.
– Semantik o‘xshashlik asosida plagiat hujjatlarni aniqlash metodlari.
9. An external plagiarism detection system based on part-of-speech (POS) tag n-grams and word embedding. *Expert Systems with Applications*, 2022, Article 116677.
– NLP va embeddinglar yordamida sintaktik va semantik o‘xshashlikni aniqlash bo‘yicha usul.
10. Semantic Similarity/Relatedness for Cross Language Plagiarism Detection**.** *Indonesian Journal of Electrical Engineering and Computer Science*, 2021.
– Til kesimidagi semantik o‘xshashlikni o‘lchash orqali plagiatni aniqlash.
11. Plagiarism Detection in English Academic Documents using A Lexical-Semantic Hybrid and Support Vector Machine**.** INOVTEK Polbeng – Seri Informatika, 2026.
– Leksik va semantik xususiyatlar birgalikda SVM bilan plagiatni aniqlash usuli.

© *Elyor Egamberdiyev*