

HOBBS ALGORITMINING O‘ZBEK TILI MATNLARIGA TATBIQI

Jumaqulova Nasiba Qurbon qizi
Alisher Navoiy nomidagi Toshkent davlat
o‘zbek tili va adabiyoti universiteti magistranti

APPLICATION OF HOBBS’S ALGORITHM TO UZBEK TEXTS

Jumakulova Nasiba Kurban kizi
Master’s degree student of
the Tashkent state university of the uzbek language
and literature named after Alisher Navoi



E-mail:
[nasiba.qurbonovna@
gmail.com](mailto:nasiba.qurbonovna@gmail.com)
ORCID: 0009-0003-
9150-2603

ПРИМЕНЕНИЕ АЛГОРИТМА ХОББСА К ТЕКСТАМ НА УЗБЕКСКОМ ЯЗЫКЕ

Жумақулова Насиба Курбон кизи, магистрант,
Ташкентский государственный университет
узбекского языка и литературы имени Алишера Навои

Annotatsiya: Ushbu maqolada olmoshlar bilan bog‘liq koreferentlikni hal etish muammosi, uni hal etishda Hobbs algoritmidan foydalanishning afzalliklari va kamchiliklari yoritilgan. Shuningdek, Hobbs algoritmining o‘ziga xos xususiyatlari, undan foydalanish yo‘l-yo‘riqlari bayon etilgan. Algoritm o‘zbek tilidagi matnlarga tatbiq etilib, uning ishlash prinsipi o‘rganilgan.

Kalit so‘zlar: *Hobbs algoritmi, olmosh, anafora, antisident, teg, sintaktik tahlil daraxti, birikma.*

Аннотация: В данной статье описана проблема разрешения кореференции, связанной с местоимениями, преимущества и недостатки использования алгоритма Хоббса при ее решении. Также описаны особенности алгоритма Хоббса и инструкции по его использованию. Алгоритм был применен к текстам на узбекском языке и изучен принцип его работы.

Ключевые слова: *алгоритм Хоббса, местоимение, анафора, антисидент, тег, дерево синтаксического анализа, сочетание.*

Abstract: This article discusses the problem of solving coreference related to pronouns, the advantages and disadvantages of using the Hobbs algorithm to solve it. It also describes the specific features of the Hobbs algorithm and instructions for its use. The algorithm is applied to Uzbek texts and its working principle is studied.

Key words: *Hobbs algorithm, pronoun, anaphora, antecedent, tag, parse tree, conjunction*

KIRISH. Hobbs algoritmi anaforani hal qilish uchun qoidaga asoslangan algoritmdir. U birinchi marta 1978-yilda Jerri Hobbs tomonidan taklif qilingan. Algoritm, avvalo, anaforani o‘z ichiga olgan gapning sintaktik tahlil daraxtini yaratish orqali ishlaydi. Tahlil daraxti gapdagi so‘zlarning o‘zaro grammatik munosabatlarini

ko‘rsatadi. Ya’ni Hobbs algoritmi anaforaga muqobil antisidentni aniqlash uchun tahlil daraxtidan foydalanadi. **Anafora** o‘zidan avval keltirilgan shaxs, narsa, holat kabilarni ko‘rsatuvchi, ishora qiluvchi, antisidentga bog‘liq so‘z bo‘lib [1], o‘zbek tilshunosligida anafora “olmosh” ekanligi ta’kidlangan [2]. Anafora bilan bir xil grammatik

shaklda va bir xil jinsda bo'lgan ot va otlashgan so'zlar **antisident** sifatida belgilanadi. Ushbu algoritmning o'ziga xos xususiyatlari quyidagilardir:

- Algoritm grammatik qoidalarga asoslangan;
- Faqat 3-shaxs olmoshlari uchun amal qiladi;
- Joriy gaplarda “to'g'ridan chapga” tamoyili bo'yicha qidiruvni olmoshga eng yaqin turgan ot yoki otlashgan so'zdan boshlaydi.
- Agar antisident topilmasa, oldingi gaplardan “chapdan to'g'riga” tamoyili asosida qadamma-qadam qidiruvni davom ettiradi.
- Sodda va amalga oshirish oson.

Hobbs algoritmiga ko'ra, anafora hamda antisident sifatida taxmin etiladigan birlik quyidagi jihatlarda teng bo'lishi lozim:

1. *Shaxs.*
2. *Jins.*
3. *Raqam.*
4. *Joylashuv.*
5. *Grammatik o'rni.*

Hobbs algoritmining afzalligi shundaki, keltirilgan qoidalar asosida har bir gapning sintaktik daraxtidan antisidentlarni topish va ularning aniqlik ko'rsatkichini qo'lda hisoblash mumkin. **Algoritmni o'zbek tili matnlariga tatbiq etishning birinchi bosqichi** o'zbekcha matnlarni *grammatik teglash* hisoblanadi. Bu jarayonda sintaktik birliklar formallashtirilishi talab etiladi, ya'ni ularning har biri ramzlanadi. Ana shunday maxsus qabul qilingan lingvistik belgilar, ramzlar **teg** deb nomlanadi. Teg – matn tahlilini amalga oshirish jarayonini tezlashtirish va osonlashtirishga xizmat qiluvchi shartli belgi yoki maxsus koddir [3]. Hobbs o'z algoritmidagi gap uchun S, o'tli birikmalar uchun NP, fe'lli birikmalar uchun VP, predlogli birikmalar uchun PP, o'rin-joy otlari uchun NP_loc, o'tli birikma tarkibidagi birliklar uchun N-bar kabi teglardan foydalangan. Ularning to'liq ro'yxati 1-jadvalda keltiriladi:

1-jadval. Hobbs tomonidan qo'llangan grammatik teglar

Teg	Izoh	Namuna
S	gap	<u>Bugun yomg'ir yog'adi.</u>
NP	otli birikma	<u>qiziqarli kitob</u>
VP	fe'lli birikma	<u>ta'tilda dam olmoq</u>
PP	predlogli birikma	<u>qalam bilan chizmoq</u>
Det	aniqlovchi	<u>ko'p mahsulot</u>
N-bar	otli birikmani tashkil etuvchi qism	<u>chiroyli qiz</u>
NP_nom	bosh kelishikdagi otli birikma	<u>go'zal tabiat</u>
NP_acc	tushum kelishigidagi otli birikma	<u>qarzni so'rash</u>
NP_dat	jo'nalish kelishigidagi otli birikma	<u>unga gapirish</u>
NP_abl	chiqish kelishigidagi otli birikma	<u>onadan ranjish</u>
NP_gen	qaratqich kelishigidagi otli birikma	<u>mening maoshim</u>
NP_loc	o'rin-joyini ifodalovchi otli birikma	<u>maktabda o'qimoq</u>

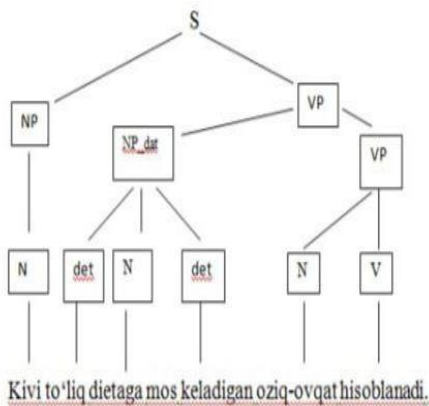
Keltirilgan grammatik teglar ingliz tili grammatikasi asosida ishlab chiqilgan. Shu bois tahlil jarayonida ushbu teglarga tayaniladi va ba'zi teglar o'zbek tili grammatikasi uchun moslanadi. Xususan, “predlog” termini “old ko'makchi” degan ma'noni bildiradi [4] va vazifasiga ko'ra o'zbek tilida yordamchi so'z turkumiga mos keladi. Biz sintaktik jihatdan **PP** sifatida yordamchi so'z orqali birikkan birikmalarni, **Det** sifatida sifatlovchi aniqlovchili birikmalarni, **NP_gen** sifatida qaratqich aniqlovchili birikmalarni tekshiramiz. Vaholanki, Hobbs *-ning* qo'shimchasiga teng keluvchi *of* (-ning) predlogini

PP doirasida, *'s* (-ning) qo'shimchasini Det doirasida, *his*, *her* kabi qaratqich kelishigida turlangan so'zlarni NP_gen doirasida tahlil qilgan. Teglar burchakli qavslarda (< >) aks ettiriladi [5]. < / > “ochuvchi”, < > “yopuvchi” teg belgisi sifatida qabul qilingan. Quyida o'zbekcha matnlarni sintaktik teglar yordamida teglash jarayonini ko'rish mumkin:

Asl matn: *Kivi to'liq dietaga mos keladigan oziq-ovqatdir (1). Uni dietangizga qo'shish tavsiya etiladi (2).*

Grammatik teglangan matn: </N> Kivi </N> </NP_dat> to'liq dietaga mos keladigan </NP_dat> </VP> oziq-ovqat hisoblanadi </VP> (1). </P> Uni </P> </NP_dat> dietangizga qo'shish </NP_dat> </VP> tavsiya etiladi </VP> (2).

Ikkinchi bosqich berilgan matndagi har bir gapning sintaktik tahlil daraxtini shakllantirishdir. Sintaktik daraxt – tabiiy tilda gap yoki so‘z birikmasi sintaktik tuzilishining grafik tasviridir [6]. Ma'lumki, gap so‘z birikmalaridan tashkil topadi. Sintaktik daraxt dastlab gapni otli va fe'lli so‘z birikmalariga ajratadi va so‘z birikmalari zanjiridagi lingvistik birliklar munosabatini aniqlab beradi. Yuqorida grammatik teglangan matnning sintaktik tahlil daraxtlarini shakllantiramiz. 1-gap uchun sintaktik tahlil daraxti quyidagi ko‘rinishga ega:



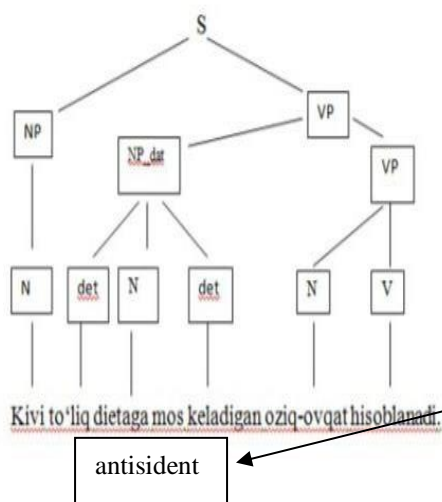
1-rasm. 1-gapning sintaktik tahlil daraxti

2-gap uchun sintaktik tahlil daraxtini quyida tasvirlaymiz:



2-rasm. 2-gapning sintaktik tahlil daraxti

Tahlil daraxtlari yaratilgach, **uchinchi bosqichda** qoidalarga tayangan holda antisident aniqlanishi zarur. Joriy gapda qidiruv olmoshdan oldinda joylashgan NP dan boshlanishi kerak. Olmosh gap boshida joylashgani sababli oldingi gapdagi NP larni “chapdan to‘g‘riga” tamoyili bo‘yicha qidiramiz. NP_dat tarkibidagi “dietaga” so‘zi ega vazifasida bo‘lmagani uchun keyingi NP – “kivi” so‘zi tekshirildi va barcha qoidalarni qanoatlantirgani uchun antisident sifatida tanlandi, qidiruv yakunlandi:



3-rasm. Antisidentni aniqlash jarayoni

Tahlillardan ko‘rinadiki, Hobbs algoritmini o‘zbekcha matnlarga tatbiq etganda qisqa va sodda gaplar tahlili orqali samarali natijaga erishish mumkin. Ammo nutqdagi murakkab gaplar tahlilida bu algoritmi qoidalarini to‘g‘ridan to‘g‘ri qo‘llab bo‘lmaydi. Demak, Hobbs algoritmi kamchiliklardan xoli emas:

- Ko‘p vaqt sarf lanadi;
- Murakkab gaplarda to‘g‘ri variantni aniqlay olmaydi.
- Matnning semantik ma‘nosini hisobga olmaydi.

Shunga qaramay, Hobbs algoritmi olmoshlarda koreferentlik muammosini hal etish uchun qimmatli vositadir. U ko‘plab tabiiy tillarni qayta ishlash dasturlarida ishlatilishi mumkin bo‘lgan oddiy va samarali algoritmi. Biroq uning grammatik qoidalarini bilish va uni boshqa usullar bilan birgalikda qo‘llash muhimdir.

Foydalanilgan adabiyotlar:

1. Elova D.Q. O‘zbek tili matnlarida koreferensiyani hal qilish bosqichlari. /

“Kompyuter lingvistikasi: muammolar, yechim, istiqbollari” xalqaro ilmiy-amaliy konferensiya materiallari, Vol 1, № 01 (2023), 156-161-b.

2. Hojiyev A. Tilshunoslik terminlarining izohli lug‘ati. – Toshkent, 2002.
3. Primova M. Til korpuslarida matnlarni annotatsiyalash: afzallik va kamchiliklari. / O‘zbekiston: til va madaniyat, 2023 Vol. 4 (6), 6-18-betlar.
4. <https://uz.m.wiktionary.org/wiki>
5. Abjalova M. Korpus birliklarini teglash va annotatsiyalash masalasi. / “Zamonaviy ilmfan va ta‘lim istiqbollari” Respublika ko‘p tarmoqli ilmiy-amaliy konferensiyasi to‘plami, Vol. 1, Issue 2, 2023. – 345-351-betlar.
6. Abdullayeva O. O‘zbek tili matnlarida sintaktik teg va teglash masalasi. / O‘zbekiston: til va madaniyat, 2023 Vol. 1 (6), 46-57-betlar.

