

ISSN 1999-9801



АУЭС

Алматы энергетика және
байланыс университетінің
ХАБАРШЫСЫ

ВЕСТНИК

Алматинского университета
энергетики и связи

2(49)

2020

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РЕСПУБЛИКИ КАЗАХСТАН
НЕКОММЕРЧЕСКОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО
«АЛМАТИНСКИЙ УНИВЕРСИТЕТ ЭНЕРГЕТИКИ И СВЯЗИ ИМЕНИ ГУМАРБЕКА
ДАУКЕЕВА»

В Е С Т Н И К

АЛМАТИНСКОГО УНИВЕРСИТЕТА ЭНЕРГЕТИКИ И СВЯЗИ

Учрежден в июне 2008 года

2 (49)

2020

Импакт-фактор - 0.154

Научно-технический журнал
Выходит 4 раза в год

Алматы

БЛОКЧЕЙН БІЛІМ САЛАСЫНДА: ЖОҒАРЫ ОҚУ ОРЫНДАРЫНДА СТУДЕНТТЕРДІҢ БАҒАСЫН БЛОКЧЕЙН АРҚЫЛЫ ҚАЛАЙ БАСҚАРУҒА БОЛАДЫ?

Б. Құмалақов¹, Я. Шақан^{2,*}

¹ Ақпараттық жүйелер кафедрасы, Халықаралық бизнес университеті, Алматы, Қазақстан

² Ақпараттық технологиялар факультеті, әл-Фараби атындағы Қазақ ұлттық университеті, әл-Фараби д-лы. 71, Алматы, Қазақстан

e-mail: b.kumalakov@gmail.com, shakan.yasin@gmail.com

Аңдатпа. Қазіргі уақытта технологиялар өте тез дамып келеді және ақпараттық қауіпсіздікке деген қажеттілік үнемі артып келеді. Осыған байланысты, Blockchain технологиясы сұранысқа ие болуда, бұл бізге ақпараттың қауіпсіздігі мен тұтастығын сақтауға мүмкіндік береді. Сонымен қатар, технология орталықтандырылмаған ортаны құруға мүмкіндік береді, онда транзакциялар мен мәліметтер қандай да бір тараптық қйымның қатысуынсыз жүзеге асырылады.

Біз студенттердің кредиттерін басқару үшін Ethereum платформасына негізделген орталықтандырылмаған веб-ресурстарды ұсындық. Орталықтандырылмаған бағдарлама (Dapp) студенттерге белгілі бір курстардан алған нәтижелерін білдіретін жетондарды өңдейді және басқарады. Несиелік жүйе - бұл университеттер мен студенттер несиелерді басқаруда қолдана алатын ашық және технологиялық тұрғыдан жетілдірілген формаға алғашқы қадам.

Бұл ғылыми зерттеудің жаңалығы - Блокчейн технологиясының негізінде ақпараттық веб - ресурстарды құру. Осы ресурстың арқасында студенттердің бағаларын бақылауға және жоғары білім туралы сенімді ақпарат алуға мүмкіндік туады. Бұл қолданыстағы жазбаларға өзгерістер енгізу мүмкіндігін толығымен жояды.

Түйін сөздер: блокчейн, жоғары білім, несиелер, жетондар.

БЛОКЧЕЙН В ОБРАЗОВАНИИ: КАК УПРАВЛЯТЬ СТУДЕНЧЕСКИМИ КРЕДИТАМИ ВЫСШЕГО ОБРАЗОВАНИЯ ЧЕРЕЗ БЛОКЧЕЙН?

Б. Кумалаков¹, Я. Шақан^{2,*}

¹ Кафедра информационных систем, Университет международного бизнеса, Алматы, Казахстан

² Факультет информационных технологий, Казахский национальный университет им. Аль-Фараби, пр. Аль-Фараби 71, Алматы, Казахстан

e-mail: b.kumalakov@gmail.com, shakan.yasin@gmail.com

Аннотация. В настоящее время технологии развиваются очень быстро и потребность в информационной безопасности постоянно растет. В связи с этим технология Blockchain становится востребованной, что позволяет нам сохранять безопасность и целостность информации. Кроме того, технология позволяет создавать децентрализованную среду, в которой транзакции и данные осуществляются без какой-либо сторонней организации.

Мы предложили децентрализованный веб-ресурс на основе платформы Ethereum для управления кредитами студентов. Децентрализованное приложение (Dapp) будет обрабатывать, управлять и контролировать токены, которые представляют собой кредиты, которые студенты получают за прохождение определенных курсов. Кредитная система является первым шагом к более прозрачной и технологически продвинутой форме, которая может использоваться университетами и студентами для управления кредитами.

Новизной данного научного исследования заключается в создании информационного веб-ресурса на базе технологии Блокчейн. Благодаря этому ресурсу, появляется возможность сделать отслеживание оценок студентов и получать достоверную информацию о высшем образовании. При этом полностью исключается возможность вносить изменения в существующие записи.

Ключевые слова: блокчейн, высшее образование, кредиты, токены.



АВТОМАТИКА, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И КИБЕРБЕЗОПАСНОСТЬ

МРНТИ 20.23.17; 20.23.21; 20.23.25

ҚАЗАҚ ТІЛІНДЕГІ СӨЙЛЕМДЕР ҮШІН ТӘУЕЛДІЛІК АҒАШТАРЫН ҚҰРУ

М.Е. Мансурова, Н.Қ. Қадырбек, А.М. Мисебай, М.Е. Қырғызбаева*

Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан Республикасы
mansurova.madina@gmail.com, nurgaliqadyrbek@gmail.com, armanbek128@mail.ru,
marzhan.kyrgyzbaeva@gmail.com

Аңдатпа. Табиғи тіл – бұл фонетикалық, морфологиялық, лексикалық, синтаксистік, семантикалық деңгейлер секілді әртүрлі деңгейлер ерекшеленетін күрделі құрылым. Табиғи тілдегі мәтіндерді автоматты түрде өңдеуге арналған талдауыштың ішінде сөйлемнің негізгі семантикалық құрылымын анықтауға мүмкіндік беретін синтаксистік талдаушы ерекше орын алады. Бұл жұмыста семантикалық құрылымды анықтауға және сөйлемге синтаксистік талдау жасауға мүмкіндік беретін қазақ тілінде сөйлемдерге тәуелділік ағаштарын құру алгоритмі көрсетілген. Тәуелділік ағашы сөйлемдегі синтаксистік қатынастар ережелерімен құрылған бағдарланған граф арқылы ағаш қаңқасы негізінде жасалады.

Түйін сөздер: тәуелділік ағаштары, тәуелділік грамматикасы, корпуслық лингвистика, синтаксистік құрылым, тіркестердің формалды моделі.

Кіріспе

Грамматиканың құрамдас бөлігі саналатын синтаксистің негізгі зерттеу нысаны – тілдің синтаксистік құрылымы. Синтаксистік құрылым дегеніміз – әр түрлі тұлғада өзара тіркесіп келіп жұмсалатын сөздердің сөйлем құраудағы жүйелі ережелері мен заңдарының жиынтығы. Сөйлем – белгілі бір тілдегі грамматикалық заңдылықтар бойынша ұйымдасқан, ойды жеткізудің, хабарлаудың басты құралы болып табылатын біртұтас тілдік бірлік. Жалпы тіл білімінде сөз синтаксисіне қатысты зерттеулерде сөйлемнің тұлғалық жағынан ұйымдасуы немесе предикативті түзіліс (конструкция) ыңғайындағы құрылымы басты назарда болды. Бұл формалды құрылым деп аталады. Зерттеу мақаласына сөйлемнің осы синтаксистік сипаты назарға алынып, тәуелділік ағашын құруға негіз болды [1].

Тәуелділік ағаштары – сөйлемнің синтаксистік құрылымын бейнелеудің ең көрнекі және ең көп таралған тәсілі. Бұл жағдайда сөйлем бағдарланған ағашты (төбелер — сөйлемдегі токендер) құруға болатын элементтердің сызықтық реттелген жиыны (сөз тіркестері) ретінде ұсынылады. Түйіндердің жұбын байланыстыратын қабырға бағыты бағыныңқыдан басыңқыға сәйкес келеді, қосымша мәлімет ретінде қабырғаны синтаксистік қатынас арқылы маркерлеуге болады.

Тәуелділік ағаштарының тағы бір қолданылу мысалы ретінде [2] жұмысты келтіруге болады. Бұл жұмыста тәуелділік ағаштарын қолдана отырып сөйлемдер арасындағы ұқсастықты бағалау әдісі келтірілген.

Тәуелділік ағашы сөйлемнің синтаксистік құрылымын бейнелейді деуге негіз бар. Қазіргі таңда ережелерге сүйенетін екі түрлі әдіс қолданылады: *фулькрумдер әдісі* және *сүзгілер әдісі*. Біз қазақ тілі үшін сүзгілер әдісін таңдап алдық.

Фулькрумдер әдісі синтаксистік ағаштардың түбірін немесе кейбір тірек түйінді сөздерді іздестіруге негізделген, олар одан әрі тағайындалған ережелерге сәйкес басқа сөздермен байланысқа түседі. *Сүзгілер әдісі* синтаксистік қатынастың барлық ережелерін бір мезгілде қолдану арқылы, барлық мүмкін сөз тіркестері шығарып алынады, содан кейін алынған құрылым (бағдарланған граф) сүзгіден өтеді. Біз алдағы уақытта сүзгілер әдісін қолданамыз.

Тәуелділік ағашын құрудағы басты қиындық сөйлемдегі сөз тіркестерінің арасында сабақтас және салалас байланыстардың қатар болуы. Өйткені салалас сөздердің арасында тәуелділік қатынасын орнату қиын/мүмкін емес, олар әдетте тыныс белгілері арқылы яки шылаулар («және», «немесе...», «немесе...» т.с.с.) арқылы байланысып, бір деңгейде тұрады. Демек құрмалас сөйлемдерді, сөйлемнің бірыңғай мүшелерін, айқындауыш мүшелерді қарастыру өз алдына үлкен тақырып.

Тәуелділік грамматикасы таксономиялық бірліктермен жұмыс істейді. Тәуелділік грамматикасындағы барлық байланыстар бағынышты деп қарастырылады. Талғат Сайрамбайұлының ғылыми зерттеулерінде сөз тіркестері және жай сөйлем синтаксисіне қатысты өзекті мәселелерді қарастырған, демек сөйлемдегі токендердің орын тәртібінде белгілі бір қатаң заңдылық бар деп айтуға болады [3].

Біздің мәтіндік талдауыш тәуелділік ағашын құру үшін бірнеше сатыдан тұратын процесті жүзеге асырады: *сегментация – морфологиялық талдау – мүмкін сөйлем формаларын генерациялау – мүмкін сөз тіркестерін анықтау.*

Тәуелділік грамматикасы негізінде сөйлем синтаксисі:

1) сөйлемдегі сөздер арасындағы барлық байланыстар иерархиялық түрдегі тәуелділік арқылы анықталады,

2) сөйлемдегі барлық токендер талдау ағашының төбелерін қамтиды,

3) ағаштың түбірі ретінде предикат (ол көрсетілмеген жағдайда предикатқа нұсқайтын оның атрибуты) алынады. Субъект пен баяндауыш (қимыл) арасында қиысу синтаксистік қатынасы бар, бастауышқа баяндауыштың бағынуын А.Т. Аширова өз зерттеулерінде дәйектеген [4]. Дегенмен, ол бағынықтылық тек морфологиялық сөзжасам тұрғысынан болғандықтан және предикат семантикасы оған қолданылатын субъектіні анықтайтындығын ескердік.

Қазіргі заманғы тәуелділік грамматикасы көбінесе француз лингвисті Луи Теньердің идеяларына негізделеді. Тәуелділік ағашын құру барысында синтаксистік тәуелділікті анықтаудың бірнеше критерийлері бар:

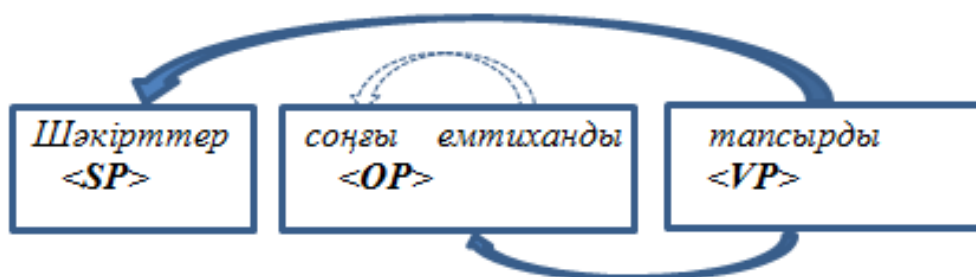
1) грамматикалық байланыс критеріі – екі сөз синтаксистік байланысқан деп саналады, олар бірге қандай да бір синтаксистік ережеге бағынса (қиысу, матасу, т. б.);

2) сызықтық позиция критеріі – екі сөз синтаксистік байланысқан болып саналады, егер бір сөз формасының сызықтық позициясы (сөйлемдегі позиция, сөз тіркестері) басқа сөз тіркестерінің сызықтық позициясымен анықталса;

3) күрделі сөйлемдерде субъектілік-объектілік тәртіп өзгеріссіз болады, ал барлық қосалқы тұрлаусыз элементтер предикатқа дейін болады [5].

Бэкус формалды моделін қолдана отырып қазақ тіліндегі сөйлемдердің формальді моделін (ФМ) төмендегіше белгілеуге болады (1-сурет):

SENTENCE::=<SP><OP><VP>|<SP><VP><OP>|<OP><SP><VP>|<VP><SP><OP>|<VP><OP><SP>
>

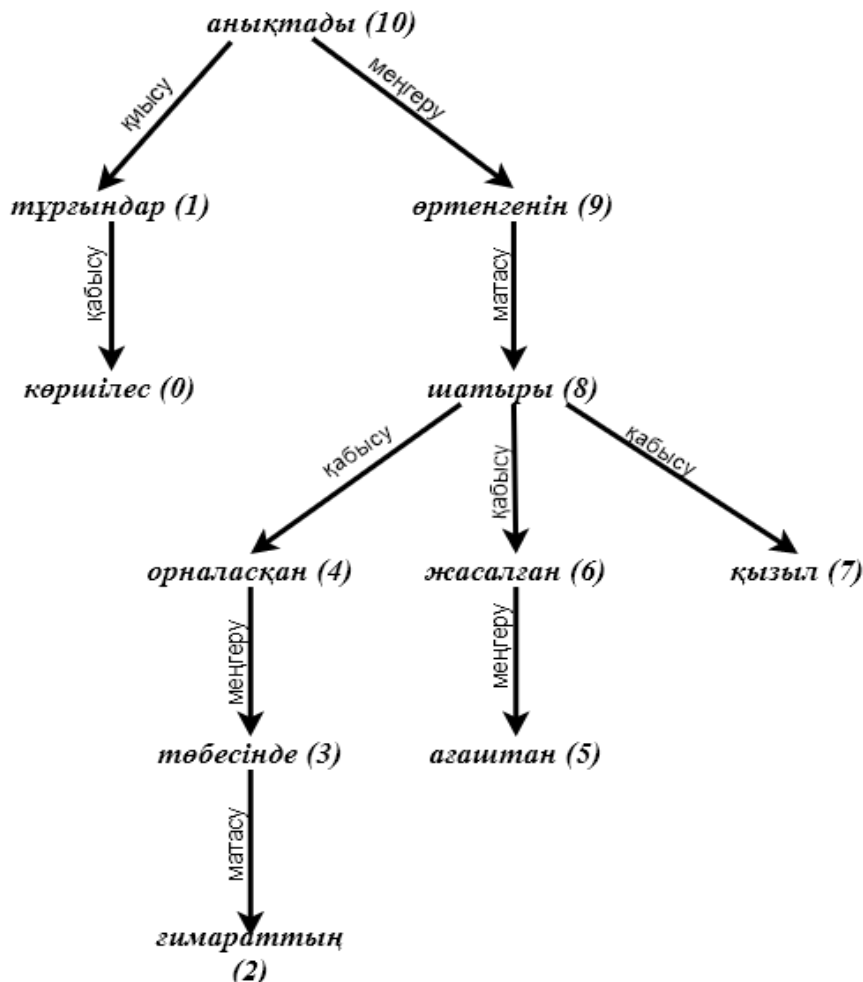


Сурет 1 – Бэкус формалды моделі

<SP><VP><OP> Шәкірттер тапсырды соңғы емтиханды
<OP><SP><VP> Соңғы емтиханды шәкірттер тапсырды
<VP><SP><OP> Тапсырды шәкірттер соңғы емтиханды
<VP><OP><SP> Тапсырды соңғы емтиханды шәкірттер

Мұндағы, *SP* – субъектілік фразалық құрылымы, *VP* – егістік/предикаттық фразалық құрылымы, *OP* – Объектілік фразалық құрылымы.

Келтірілген ережедегі фразалық құрылымды келесідей ашып жазуға болады:



<SP>::=<N>|<ADJ><SP>|<NUM><SP>|<N><SP>|<SP><CONN><SP>
<OP>::=<N>|<ADJ><OP>|<NUM><OP>|<OP><CONN><OP>

Сурет 2 – Тәуелділік ағашы

<VP>::=<VP >|<AUX><VP>|<ADV><VP>

Мұндағы: N – зат есім, NUM – сан есім, ADJ – сын есім, ADV – үстеу, CONN – шылау, AUX – көмекші егістік.

Бұл жерде, әрбір фразалық құрауыш (SP, OP, VP) күрделі сөз тіркестерінен тұруы мүмкін. 2-суретте «Көрішілес тұрғындар ғимараттың төбесінде орналасқан ағаштан жасалған қызыл шатыры өртенгенін анықтады» сөйлемінің тәуелділік ағашы келтірілген, әрбір төбеде сөздердің сөйлемдегі индексі де көрсетілген. Әуелі келесі қадамдар арқылы мүмкін синтаксистік қатынастарды анықтаймыз:

1) Сөйлем ұзындығына сәйкес 0-дік іргелестік матрицасы құрылады: Егер қандайда бір [i,j] қабырғасы белгіленсе мұнда i басты, ал j бағыныңқы сөздерге сәйкес келеді

2) Қосымша қатынасы ізделінеді: бүкіл шылаулармен көршілес сөздер анықталынады. Әдетте мұндай байланыс жанасуға жатқызылады. Дегенмен алдағы уақыттағы семантикалық талдаулар үшін мұндай байланыстарды ажырата білу мақсатқа жақынырақ

3) Қиысу қатынасындағы сөздер ізделінеді: Мейілінше сөйлем басына жақын, және ағаш түбірімен байланыса алатын сөз;

- 4) Көршілес сөздер арасынан қабысу қатынасындағы сөздер ізделінеді;
 5) Матасу ізделінеді;
 6) Жанасу байланысын ізделінеді.

Ескерту: *әрбір бағыныңқылық қатынас фразалық құрылым аясында локальды екеін ескеру қажет.*

Қазіргі уақытта сөз тіркестері арасындағы синтаксистік тәуелділіктер 1-кестеде көрсетілген ережелер бойынша анықталуда.

Кесте 1- Синтаксистік байланыс түрлері

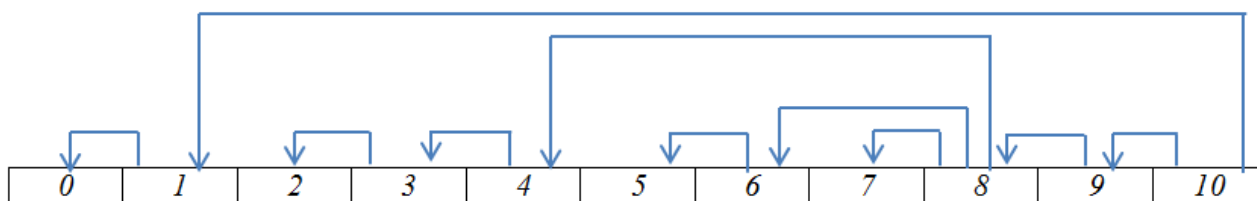
Байланыс түрі	Фразалық құрылым (мысал)	Ереже <бағыныңқы><басты> <i>қазіргі таңда бағдарламада қолданылатын ережелер</i>
Қиысу	<SP> – <VP>	Сөздердің жіктік және көптік жалғауы арқылы байланысуы
Матасу	<N> – <SP>	Ілік септігіндегі сөз бен тәуелдік жалғаулы сөздің байланысуы <сөз+IC><сөз+ТӘУЕЛДІК> <сөз><сөз+ТӘУЕЛДІК> <емістік+есімше+IC><сөз+ ТӘУЕЛДІК >
Меңгеру	<OP> – <VP>	Сөздердің ілік септіктен басқа септіктердің жалғаулары арқылы байланысуы <сөз+ТС>, <сөз+БС>, <сөз+КС>, <сөз+ЖС>, <сөз+ШС>
Қабысу	<ADJ> – <SP> <ADJ>– <OP>	Сөздердің ешбір жалғаулықсыз тек орын тәртібі арқылы қатар тұрып байланысуы. Әдетте Олардың арасына басқа сөз енгізуге келмейді. <зе><зе>, <сын><зе>, <сан><зе>, <есім><зе>, <зе+БС><есімше>, <сөз+ТС/ЖС/ШС/КС><есімше>, <үстеу><зе/сын/сан/есімдік >
Жанасу	<ADV>– <VP>	Сөздердің ешбір жалғаулықсыз бірде іргелес, бірде алшақ тұрып байланысуы. Негізінен үстеу мен етістік жанаса байланысады <үстеу><етістік> <көсемше><етістік>

Әрине, қазақ тілінде сөздердің интонация арқылы да, пунктуация арқылы да, көмекші сөздер арқылы да байланысуы мүмкін. Алайда, әзірше бұл жағдайларды қарастырмай, зерттеу нәтижелілігін тексеру үшін қарапайымырақ жағдаймен шектелеміз, яғни талдауыш әрқашан байланған бағдарланған граф қайтаратын жағдайды ғана қарастырамыз.

Егер барлық тәуелділік қатынасты көрсететін нұсқауыш сөйлем орналастырылған түзуде бір бағытта жүргізілсе, және бірде-бір нұсқауыш басқа нұсқауышты қиып өтпесе, сөйлем *проективті* деп аталады. 2-суретте қарастырылған сөйлемді екі түрлі фразалық модель арқылы проективтілікті бейнелейік (3, 4-суреттер):

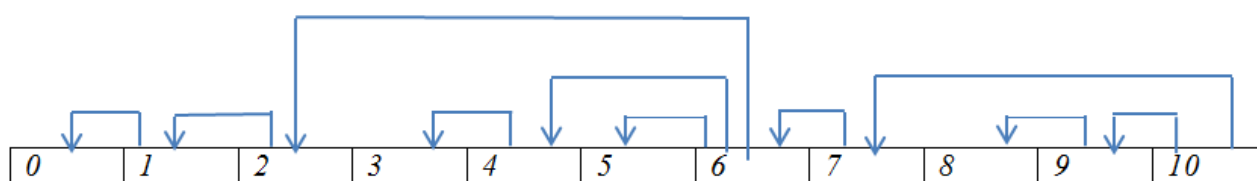
а) *Көршілес(0) тұрғындар(1) гимараттың(2) төбесінде(3) орналасқан(4) ағаштан(5) жасалған(6) қызыл(7) шатыры(8) өртенгенін(9) анықтады(10).*

(Фразалық модель: <SP><OP><VP>)



Сурет 3 – а) мысалы үшін синтаксистік байланыс бағытын көрсететін сөйлемнің фразалық моделі

б) *Фимараттың(0) төбесінде(1) орналасқан(2) ағаштан(3) жасалған(4) қызыл(5) шатыры(6) өртенгенін(7) көршілес(8) тұрғындар(9) анықтады(10).*
(Фразалық модель : <OP><SP><VP>)



Сурет 4 – б) мысалы үшін синтаксистік байланыс бағытын көрсететін сөйлемнің фразалық моделі

Көріп отырғанымыздай, сөйлемнің фразалық моделі әртүрлі ретпен орналасса да проективтілік қасиеті сақталды. Бұл өз кезегінде, сөйлемдегі сөздердің синтаксистік байланысы сызықты тәуелділік арқылы жүзеге асуымен түсіндіріледі. Жоғарыда келтірілген ФМ сөйлемнен синтаксистік байланыстарды іздеу барысында пайда болатын жалған байланыстарды болдырмау үшін қолданылады. Себебі әрбір синтаксистік байланысты ФМ аясында локализациялайды.

Тәуелділік ағаштарын құруда мүмкін синтаксистік қатынастағы сөз тіркестері ішінен негізгі/негізгіге жақын тәуелділік ағашын іріктеп алуда екі түрлі тәсілді қолданамыз: біріншісі, ағаш қаңқасы арқылы, екіншісі, Бэкус формальді моделін қолдана отырып сөйлемді негізгі SP, VP, OP құрауыштарға ажыратып, ережелерге сүйену;

Ағаш қаңқасы (каркасы) деп – $G \langle V, E \rangle$ графының барлық төбелерін қамтитын ағашты айтамыз, мұндағы V – граф төбелері, сөйлемдегі токендер жиыны, E – граф қабырғалары, сөз тіркестері арасындағы тәуелділікті көрсетеді. Takeaki Uno өз жұмыстарында ағаш қаңқаларын және бағдарланған графтан оны алу жолдарын көрсеткен, сонымен қатар бір графтан бірнеше ағаш қаңқасы алынуы мүмкін. Себебі графтың кейбір төбелері арасында бірнеше тәсілмен жол құруға болады [6].

Ағаш қаңқасын табу алгоритмін келтірейік:

1) G графынан ішкі граф құрайтын ағаш бола алатын u_1 төбесін таңдап аламыз, мысалы оны $i=1$ деп ұйғарамыз (әдетте қиысу қатынасын құрайтын сөз тіркестерінің басыңқы жұбы таңдап алынады);

2) Егер $i=n(G)$ болса, онда есеп аяқталады және G_i ізделген ағаш қаңқасы болады. Басқа жағдайда 3-қадамға көшеміз;

3) айталық G графының ішкі графы болатын $\{u_1, u_2, \dots, u_n\}$ төбелерін қамтыған G_i графы құрылған болсын, мұндағы $1 \leq i \leq n - 1$. Әрі қарай G_i графындағы u_i төбесіне іргелес жаңа $u_{i+1} \in V$ төбені қосып G_{i+1} графын құрамыз, осылайша $\{u_{i+1}, u_i\}$ қабырғасы қосылады. Бұлай жасауға еш кедергі жоқ, себебі біздің ұйғарымымызда граф – байланған. Сондай-ақ, G_{i+1} – ағаш, себебі G_i графы цикл қамтымаған болатын. $i:=i+1$ орындай отырып әрі қарай жалғастырамыз, яғни 2) қадамға көшеміз.

Нәтижелер

Жоғарыда мысалда қарастырылған сөйлем үшін біздің талдауыш тәуелділік графын құрды және оның іргелестік матрицасы келесідей болады:

```

0 0 0 0 0 0 0 0 0 0 0
5 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 0
0 0 3 0 0 0 0 0 0 0 0
0 0 0 2 0 0 0 0 0 0 0
0 0 0 0 5 0 0 0 0 0 0
0 0 0 2 0 2 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 5 3 0 0 0
0 0 0 0 0 2 0 0 2 0 0
0 0 0 0 0 0 0 0 0 2 0
    
```

Мұндағы, матрицадағы көрсетілген сандар келесі мағынаны білдіреді: 1 – қиысу, 2 – меңгеру, 3 – матасу, 4 – жанасу, 5 – қабысу, 6 – қосымша.

Талдауыш қарастырылған сөйлем үшін 4 түрлі ағаш қаңқасы нұсқаларын қайтарады (5-сурет):

1)	2)	3)	4)
the graph is:	the graph is:	the graph is:	the graph is:
vertex 0 : []	vertex 0 : []	vertex 0 : []	vertex 0 : []
vertex 1 : [10, 0]	vertex 1 : [10, 0]	vertex 1 : [10, 0]	vertex 1 : [10, 0]
vertex 2 : []	vertex 2 : []	vertex 2 : []	vertex 2 : []
vertex 3 : [2]	vertex 3 : [2]	vertex 3 : [2]	vertex 3 : [2]
vertex 4 : [3]	vertex 4 : [3]	vertex 4 : []	vertex 4 : []
vertex 5 : []	vertex 5 : []	vertex 5 : []	vertex 5 : []
vertex 6 : [5]	vertex 6 : []	vertex 6 : [5, 3]	vertex 6 : [3]
vertex 7 : []	vertex 7 : []	vertex 7 : []	vertex 7 : []
vertex 8 : [4, 6, 7]	vertex 8 : [4, 6, 7]	vertex 8 : [4, 6, 7]	vertex 8 : [4, 6, 7]
vertex 9 : [8]	vertex 9 : [8, 5]	vertex 9 : [8]	vertex 9 : [8, 5]
vertex 10 : [9]	vertex 10 : [9]	vertex 10 : [9]	vertex 10 : [9]

Сурет 5 – 2-суреттегі мысал үшін ағаш қарқасы

Мұндағы 1-нұсқа ақиқат мәндегі күтілген нәтижемен 100%, ал 2-нұсқада 90%, 3-нұсқада 90%, 4-нұсқада 80% сәйкес келеді.

Сонымен қатар, тәуелділік ағашын құру арқасында сөйлемді талдау барысында пайда болатын омонимияны шешуде оң нәтижеге жетуге болады. Себебі бұл таңдап алынған сөйлем үшін жалпы генерацияланған сөйлемдер саны 24 болатын. Күтілген нәтижені іріктеп алуда тәуелділік ағашы ішінара қолданылды.

Синтаксистік құрылымдарды тәуелділік ағаштары түрінде ұсынудың кемшіліктері:

1) сөйлемнің әрбір формальды түрде бөлінген кіруін сөйлемнің жеке элементі ретінде қарастырудың қатаң талабы;

2) фразалар арасындағы барлық байланыс бағынышты деп түсіндіріледі.

Қорытынды

Қазақ тіліндегі сөздер арасындағы байланысты автоматты түрде дәл көрсетуде бірқатар қиындықтар бар: сөйлемдегі сөз тіркестерінің арасында сабақтас және салалас байланыстардың болуын, қыстырма сөздердің, ауыспалы мағынадағы сөздердің, фразеологизмдердің қолданылуын айта аламыз.

Сөйлемді бағдарланған байланған граф ретінде кескіндеп, ағаш қаңқаларын анықтау арқылы сөйлем үшін ақиқат немесе ақиқатқа жақын ағаштарды аламыз. Сөйлемді граф түріне аударуда алдын-ала тағайындалған ережелер арқылы мүмкін синтаксистік қатынастар құрылады. Тәуелділік

ағаштары сөйлемдегі сөздердің нақты орын тәртібін көрсетпейді, тек сөздердің байланыс иерархиясын кескіндейді және қазақ тілінің орыс тілінен басты ерекшелігі ретінде сөйлемнің проективті болуын айта аламыз.

Бұл жұмыс ҚР БҒМ О.0856 BR05236340 «Қазақстан Республикасының цифрлы экономикасын қалыптастыру шеңберінде «логистикалық-агломерациялық» жүйесінің талдау және шешім қабылдау жоғары өнімді зияткерлік технологияларын құру» және АР05132933 «Шешім қабылдау сапасын жақсарту үшін деректердің гетерогенді көздерінен білімді алу жүйесін құру» жобалары шеңберінде жасалды.

ӘДЕБИЕТТЕР ТІЗІМІ

1. Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов : учеб. пособие / Новосибир. гос. ун-т. – Новосибирск: РИЦ НГУ, 2016. – 166 с.
2. Saziye Betul Ozates, Arzucan Ozgur, Dragomir R. Radev. Sentence Similarity based on Dependency Tree Kernels for Multi-document Summarization // Қолжетімділік режимі: <https://www.aclweb.org/anthology/L16-1452.pdf>, еркін [13.01.2019].
3. Сайранбаев Т. Қазақ тіл білімінің мәселелері. Алматы, Абзал ай: 2014. – 640 б.
4. Аширова А.Т. Сөздердің тіркесу тәсілдері мен байланысу формалары / Herald of journalism № 37(1). – 2016.
5. Ж. Мейрамбекқызы, А.А. Хорошилов. Синтактико-семантические представления для развития казахско-русского фразеологического машинного перевода / Труды V международной конференции по компьютерной обработке тюркских языков «Turklang 2017», Казань, Россия, 2017.
6. Takeaki UNO. An Algorithm for enumerating all directed spanning trees in a directed graph // Қолжетімділік режимі: <http://research.nii.ac.jp/~uno/papers/isaac96web.pdf>, еркін [20.01.2019].

LIST OF REFERENCES

1. Batura, T. V. Matematicheskaya lingvistika i avtomaticheskaya obrabotka tekstov: ucheb. posobiye / Novosib. gos. un-t. – Novosibirsk: RITS NGU, 2016. – 166 s.
2. Saziye Betul Ozates, Arzucan Ozgur, Dragomir R. Radev. Sentence Similarity based on Dependency Tree Kernels for Multi-document Summarization // Availability Mode: <https://www.aclweb.org/anthology/L16-1452.pdf>, еркін [13.01.2019].
3. Sairanbaev T. Kazak til biliminin maseleleri. Almaty, Abzal ai: 2014, - 640 b.
4. Ashirova A.T. "Sozderdin tirkesu tasilderi men bailanysu formalary Herald of journalism № 37(1). - 2016.
5. J. Meyrambekkyzy, A.A. Khoroshilov. Sintaktiko-semanticheskije predstavleniya dliya razvitia kazakhsko-russkogo phraseologyycheskogo mashinnogo perevoda / V international conference on computer processing of Turkic languages "Turklang 2017", Kazan, Russia, 2017.
6. Takeaki UNO. An Algorithm for enumerating all directed spanning trees in a directed graph // Availability Mode: <http://research.nii.ac.jp/~uno/papers/isaac96web.pdf>.

СОЗДАНИЕ ДРЕВА ЗАВИСИМОСТЕЙ ДЛЯ ПРЕДЛОЖЕНИЙ НА КАЗАХСКОМ ЯЗЫКЕ

М.Е. Мансурова, Н.К. Кадырбек, А.М. Мисебай, М.Е. Кыргызбаева*

Казахский национальный университет имени аль-Фараби, Алматы,
Республика Казахстан

e-mail: mansurova.madina@gmail.com, nurgaliqadyrbek@gmail.com, armanbek128@mail.ru,
marzhan.kyrgyzbaeva@gmail.com

Аннотация. Естественный язык является сложной структурой, в которой выделяются различные уровни: фонетический, морфологический, лексический, синтаксический, семантический. Особое место среди анализаторов для автоматической обработки естественно-языковых текстов занимает синтаксический парсер, который позволяет определить основную смысловую структуру предложения. В данной работе представлен алгоритм построения дерева зависимостей предложений на казахском языке, который позволяет выявить смысловую структуру и выполнить синтаксический анализ предложения. Дерево зависимостей создается на основе остова дерева из ориентированного графа, построенного посредством правил синтаксических отношений в предложении.

Ключевые слова: дерево зависимостей, грамматика зависимостей, корпусная лингвистика, формальная модель фраз.

DEVELOPMENT OF THE DEPENDENCY TREES FOR THE KAZAKH LANGUAGE SENTENCES

M.E. Mansurova, N.K. Kadyrbek, A.M. Misebay, M.E. Kyrgyzbayeva*

Kazakh national university named after al-Farabi, Almaty, Republic of Kazakhstan
e-mail: mansurova.madina@gmail.com, nurgaliqadyrbek@gmail.com, armanbek128@mail.ru,
marzhan.kyrgyzbaeva@gmail.com

Abstract. Natural language is a complex structure in which various levels are distinguished: phonetic, morphological, lexical, syntactic, semantic. A special place among analyzers for the automatic processing of natural language texts is occupied by the syntactic parser, which allows you to determine the main semantic structure of the sentence. In this paper, an algorithm for constructing a sentence dependency tree in the Kazakh language is presented. It allows you to identify the semantic structure and perform syntactic analysis of the sentence. A dependency tree is created on the basis of a spanning tree from a directed graph constructed by the rules of syntactic relations in a sentence.

Keywords: Dependency tree, dependency grammar, corpus linguistics, syntactic structure, phrase structure rule.