

Международный научно-популярный журнал

ISSN 2073-333X

# Наука и жизнь Казахстана Қазақстанның ғылымы мен өмірі

№ 3(58) 2018



Оңтүстік Қазақстан облысы,  
Мақтаарал ауданы, Жетісай қаласындағы  
«Сырдария» университетіне - 20 жыл

**Құрылтайшы:**  
«ҚҰҚЫҚТЫҚ МИССИЯ» ҚОҒАМДЫҚ ҚОРЫ  
ХАЛЫҚАРАЛЫҚ ҚАЗАҚСТАН КРИМИНОЛОГИЯЛЫҚ КЛУБЫ

**Учредитель:**  
ОБЩЕСТВЕННЫЙ ФОНД «ПРАВОВАЯ МИССИЯ»  
МЕЖДУНАРОДНЫЙ КАЗАХСТАНСКИЙ КРИМИНОЛОГИЧЕСКИЙ КЛУБ

**Founder:**  
PUBLIC FOUNDATION «LEGAL MISSION»  
INTERNATIONAL KAZAKHSTAN CRIMINOLOGY CLUB

**ҚАЗАҚСТАННЫҢ ҒЫЛЫМЫ МЕН ӨМІРІ**  
**НАУКА И ЖИЗНЬ КАЗАХСТАНА**  
**SCIENCE AND LIFE OF KAZAKHSTAN**  
**Халықаралық ғылыми-көпшілік журнал**  
**Международный научно-популярный журнал**  
**International popular-science journal**

**№3(58) 2018**

**ӨНЕРТАНУ - ЭКОНОМИКА**  
**ТЕХНИКА**

**Астана 2018**

**Тюлепбердинова Г.А.**

ф.м.ғ.к., доцент м.а., Әл-Фараби атындағы Қазақ Ұлттық Университеті,  
Алматы қаласы, tyulepberdinova@mail.ru

**Адилжанова С.А.**

аға оқытушы, Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қаласы,

**Ғазиз Г.Г.**

аға оқытушы, Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қаласы,

**Гусманова Ф.Р.**

ф.м.ғ.к., доцент, Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қаласы,

**Жумартов М.А.**

аға оқытушы, Әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қаласы,

### **ДЕРЕКТЕРДІ ИНТЕЛЛЕКТУАЛДЫ ТАЛДАУ**

**Түйін:** Бұл мақалада білім – адамзатпен жинақталған деректер, түсінік, ережелер, әдістер және т.б. мен принциптер жиыны ретінде немесе ойлау қызметінің нәтижелер жиыны ретінде немесе хабардар жағдай ретінде анықталатын ақпарат қарастырылады. Білім термині интелект ұғымымен тығыз байланысты. Көбінесе кең білімді адамдарды зияткерлер деп айтады. Білім, биологиялық ағзада өзара байланысқан нейрондардың күрделі құрылымы түрінде сақталады деп қарастырылады. Адам миының орташа салмағы шамамен 1,5 кг және ол  $10^{12}$  нейрондардан тұрады, бағалау бойынша жадының элементарлық сыйымдылығы  $10^{14}$  битті құрайды. Білім терминін деректер сөзімен шатастырмау керек. Сандық ақпараттарды жинаудың заманауи әдістері мен оны сақтау технологиясының дамуы үлкен деректер қорының тез таралуына әкеледі. Бұл адамзат қызметінің барлық салаларында, дәстүрлі салалардан бастап экзотикалық салаларға дейін пайда болады. Деректер қорын пайдаланушылар үшін маңызды мәні болуы мүмкін деректерді жүзілендіру мүмкіндіктерін оқуға және одан ақпарат алуға қызығушылығы арта түсті. Осындай зерттеулермен байланысты сала Data Mining атымен белгілі болуда. Яғни мақаламызда осындай жаңа деректермен бөліссек деп отырмыз.

**Кілт сөздер:** интелект, зияткерлер, нейрон, үлкен деректер қоры, Data Mining, деректер қорында білімді анықтау, болжайтын модель, сипаттайтын модель, жіктеу, регрессия, уақытша қатарлар талдауы, кластерлеу, модельдің жинақтылығы немесе жалпыламалылығы, ассоциация ережесі.

**Резюме:** В данной статье рассматриваются вопросы знания, накопленных человечеством данных, понятие, правила, методы и т.д. результаты деятельности и мышления, также рассматриваются принципы определяемые в качестве информации. Термин образования тесно связана с понятием интеллекта. Часто говорят, что у образованных людей высокий интеллектуал. Образование, рассматривается как сложная структура, сохраняется в организме в виде взаимосвязанных биологических нейронов. Средний вес человеческого мозга примерно 1,5 кг и он  $10^{12}$  состоит из нейронов, по оценке потенциальная емкость памяти  $10^{12}$  бит. Данные образования важно не путать с терминами. Развитие технологии хранения цифровой информации и современные методы ее сбора приводит к быстрому распространению большой базы данных. Это во всех сферах человеческой деятельности, сфер, отраслей, начиная с традиционных до экзотических. Обучение пользователей базы данных может иметь важное значение для систематизации данных и на получение информации. Последнее время появился интерес к возможностям электронных данных. Такие отрасли связаны с именем Data Mining. То есть в этой статье мы хотели рассказать об этом.

**Ключевые слова:** интелект, нейрон, большая база данных, Data Mining, выявление знаний в базах данных, модель, предполагающая, характеризующая модель, классификация, регрессия, анализ временных рядов, кластеризация, полнота, правила ассоциации.

**Summary:** This article discusses the issues of knowledge accumulated by mankind data, the concept, rules, methods, etc., the results of activity and thinking, also discusses the principles defined as information. The term education is closely related to the concept of intelligence. It is often said that educated people have a high intellectual. Education is considered as a complex structure, stored in the body in the form of interconnected biological neurons. The average weight of the human brain is about 1.5 kg and it is  $10^{12}$  consists of neurons, according to the assessment of the potential memory capacity  $10^{12}$  bit. Education data, it is important not to confuse the terms. The development of digital information storage technology and modern methods of its collection leads to the rapid spread of a large database. It is in all spheres of human activities, areas, sectors, ranging from traditional to exotic. User training database may be important for the systematization of data and information. Recently, there has been an interest in the possibilities of electronic data. Such industries are associated with the name Data Mining. That is, in this article we wanted to talk about it.

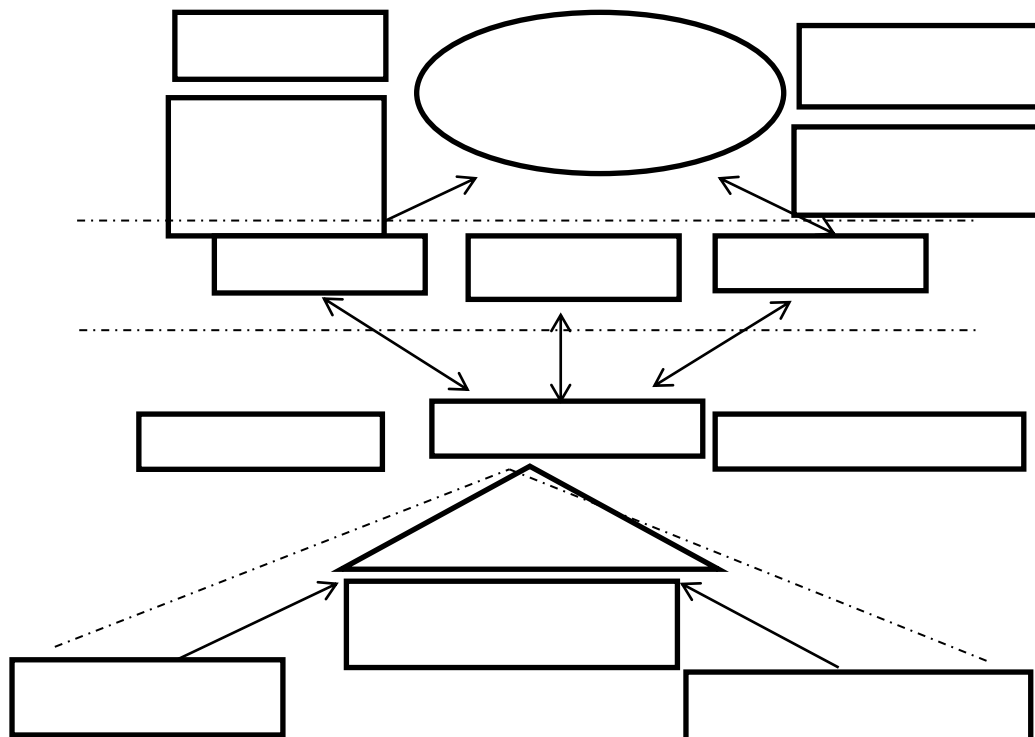
**Keywords:** intelligence, neuron, large database, Data Mining, detection of knowledge in databases model, which assumes a characterizing model, classification, regression, time series analysis, clustering, completeness and rules of the Association.

Білім – адамзатпен жинақталған деректер, түсінік, ережелер, әдістер және т.б. принциптер жиыны ретінде немесе ойлау қызметінің нәтижелер жиыны ретінде немесе хабардар жағдай ретінде анықталатын ақпарат десек болады. Білім термині интелект ұғымымен тығыз байланысты. Білім

терминін деректер сөзімен шатастырмау керек. Мысалы, дәрігер деректер – тексеру нәтижесі мен білімін пайдалана отырып емделушіні емдейді. Білім ұғымын берудің жалпы сұлбасы:

*Ойдың бейнесі* ⇒ *Жазу (символдар, мәтін, кескін)* ⇒ *компьютерде сақтау*

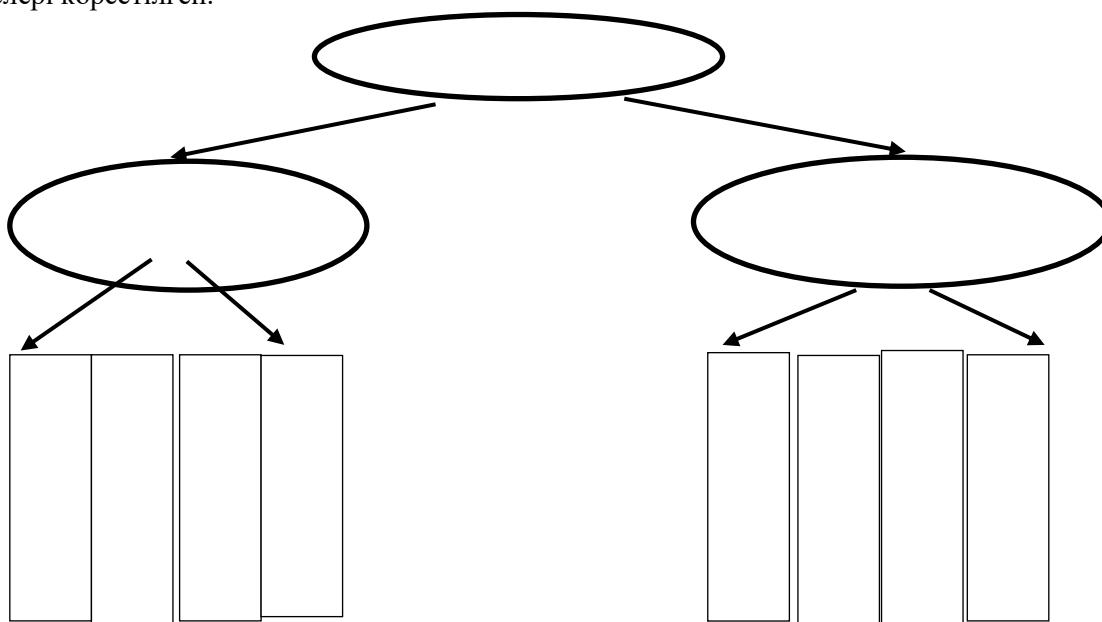
Сандық ақпараттарды жинаудың заманауи әдістері мен оны сақтау технологиясының дамуына үлкен (Деректер қоры) ДҚ-ның тез таралуына әкелді. Бұл адамзат қызметінің барлық салаларында, дәстүрлі салалардан (мысалы, кредиттік карточкамен сатып алулар туралы жазбалар, телефон қоңыраулары туралы ақпарат; тұрғындар туралы статистикалық деректері) бастап экзотикалық (мысалы, астрономиялық денелер кескіні, молекулалық ДҚ және медициналық құжаттамалар) салаларға дейін пайда болды. ДҚ-н пайдаланушылар үшін маңызды мәні болуы мүмкін деректерді жүйелендіру мүмкіндіктерін оқуға және одан ақпарат алуға қызығушылығы арта түсті. Осындай зерттеулермен байланысты сала (*ағылш.* Data Mining – деректерді интеллектуалды талдау, DM) атымен белгілі болды. DM үлкен ДҚ-нан болжамдық ақпараттарды генерациялау және автоматты түрде шығару үрдісін орындауды білдіреді. Ол пайдаланушыларға бір уақытта пайдалы және түсінікті болатын, бұрын белгісіз ара қатынастағы айқын емес деректерді табу мақсатында бақылаулар, деректер жиынын әдістерін талдауды қамтиды. DM көмегімен алынған қатынастар мен деректер үшін *модель* және *шаблон (прототип)* терминдері жиі пайдаланылады. Бұл сызықты тендеулер, ережелер, кластерлер, графиктер, ағаш тәрізді құрылымдар мен уақыттық қатарлардағы тізбектей қайталанатын шаблондар (мысалы, ауа температурасы мен ылғалдылығының сезондық өзгеруі) болуы мүмкін. DM саласы қандай да бір басқа мақсатпен (мысалы, олар клиенттердің банктік операциялар тарихын сақтау үшін жинақталған болуы ықтимал) бұрын жиналған деректермен жұмыс істейді. Бұл DM мәселелері деректерді жинау стратегиясында ешқандай рөл атқармайтынын білдіреді. Әрі бұл нақты сұрақтарға жауап беру үшін тиімді стратегиялардың көмегімен деректер жинаумен айналысатын көптеген статистикалық зерттеулерден DM-нің айырмашылығын білдіреді. DM-ды KDD (*ағылш.* Knowledge Discovery in Databases – Деректер қорында білімді анықтау) деп аталатын үлкен ДҚ-нда кездейсоқ сақталған немесе айқын емес білімдерді беретін шаблондарды автоматты немесе ыңғайлы тәсілдер ретінде шығарумен сипататуға болады. Олар деректер сақтау қоймасында, интернетте немесе деректер ағынында миллиондаған жолдардан тұруы мүмкін. Мысалы деректер сақтау қоймасындағы (*ағылш.* Data Warehouse – деректерді сақтау қоймасы, DWH) DM орны 6.1-суретте көрсетілген. Мұнда «деректер қоймасы» термині үлкен деректер қоймасын өңдеу мен сақтау жүйелерін толық ашып көрсетпейді [1].



1-сурет. Деректер қоймасы архитектурасындағы DM

Оқырмандар (ДҚ жүйесінің құрылымымен таныс) деректер қоймасы архитектурасының Data Mining, Талдау (KDD бөлігі ретінде), OLAP, OLTP, сондай-ақ DWH немесе Marts (кәсіпорынның әртүрлі қызметтеріне арналған серверлер жиынтығы) серверлері енетін дәстүрлі ДҚ жүйесінің архитектурасынан негізгі айырмашылықтарын тануы мүмкін. Сонымен қатар, әдетте деректер қоймаларында келіп түскен деректерді алдын ала өңдейтін модульдер болады [2].

Терминдер тарихы туралы қысқаша мағлұмат айта кетейік. Григорий Питетский-Шапиро осы тақырып бойынша өткен бірінші семинарда (KDD-1989) ең алғаш рет «Knowledge Discovery in Databases» терминін енгізді және бұл термин ЖИ-ті зерттеу және машиналық үйрену қауымдастықтарында пайдаланыла бастады. Дегенмен, бизнес және медиялық қауымдастықтарда Data Mining (1990) термині кең таралды. Қазіргі кезде Data Mining және KDD терминдері синоним ретінде пайдаланылады. Шамамен 2007 жылдан бастап «Predictive Analytics» және 2011 жылдан бастап «Data Science» терминдері осы бағыттағы ғылымдарды белгілеу үшін пайдаланыла бастады. Біз DM-ды ДҚ-ндағы білімдерді айқындаудың әртүрлі әдістерінің алгоритмдеріне байланысты KDD үрдісінің бір бөлімі деп нақты айта аламыз, сондай-ақ, бұл шешім қабылдауды қолдау, болжау, алдын ала айту, бағалау, бейнені тану секілді проблемаларды зерттеуге арналған (статистикалық және математикалық әдістермен ұштастыратын) құрал. Деректерді талдаудың міндеттері мен модельдеріне тоқталайық. DM зерттеу жүргізуге қажетті көптеген әртүрлі алгоритмдерді қамтиды. Талдауға кіріспес бұрын, деректер өңдеудің сәйкес *болжайтындай* (predictive) немесе *сипаттайтындай* (descriptive) болып келетін моделін таңдау қажет. 2-суретте осы типтерге қатысты жиі қолданылатын DM мәселелері көрсетілген.



2-сурет. DM модельдері және олардың негізгі мәселелері.

*Болжайтын модель* (predictive models) деректердің басқа (үйреткіш) іріктеуінің белгілі нәтижелерін пайдалану жолымен деректер мәнін болжауға мүмкіндік беретін модель. Бұл модельдерге [3.4]:

- *жіктеу* (classification) алдын-ала анықталған класстар жиынын пайдаланып, ауқымды деректер банкінен деректерді топтастыруға мүмкіндік береді. Класстар деректерді оқығанға немесе зерттегенге дейін анықталады. Мәселелерді жіктеу бұрыннан белгілі деректерді сараптап, талдап қана қоймай, алдағы уақытта анықталатын таңдамалы деректерді де талдауға мүмкіндік береді. Мысалы, материалдық шығындардың алдын алу үшін кредиттік карточкалармен төлем жүргізу кезіндегі алаяқтарды анықтау немесе жұмысшының жоба аяқталмастан ұйымнан шығып кету ықтималдылығын бағалау – жіктеу әдісін пайдалана отырып шешуге болатын қандай да бір мәселелер;

- *регрессия* (regression) зерттеліп отырған атрибуттардың қазіргі және осыған дейінгі мәндерін қолданып, алдағы уақыттағы мәнін болжауға мүмкіндік беретін статистикалық әдістердің бірі болып табылады. Регрессияның негізгі міндеті деректердің мәндері арасындағы тәуелділікті анықтау және сол тәуелділіктің математикалық формуласын құру. Осы формуланы пайдаланып есептелінген нәтиже бар немесе өткізіліп жіберілген деректердің болашақ мәндерін болжауға мүмкіндік береді.

Регрессияның басты кемшілігі – бұл әдіс тек сандық деректер үшін қолданылады, мысалы, жылдамдық немесе салмақты пайдалана отырып, болашақтағы дененің әрекетін сипаттауға болады; - *уақытша қатарлар талдауы* (time series analysis) уақытқа тәуелді ағымдық уақыттағы және бұрынғы іріктеліп алынған деректер жиынтығын пайдаланып зерттелетін үрдістің болашақ мәнін болжауды зерттейін DM бөлігі. Уақытша қатарларды талдауға арналған пайдаланатын мәндер, әдетте сағат, күндер, апталар, айлар, жылдар т.с.с. бойынша бірқалыпты үлестірілген. Зерттеуші уақытша қатар бөлігінде нақты уақыт аралығындағы сандық өзгерістерді визуализациялау үшін график сыза алады. белгілі бір уақыт аралығында әртүрлі компаниялар үшін қор нарығында тенденцияларды меңгеру үшін уақытша қатарлар талдауын пайдалануға болады және сәйкес инвестицияны жасауға болады. Дербес жағдайда, өткен ғасырдың 80-жылдарында бұл әдісті регрессиялық талдаумен үйлестіріліп егістіктегі өнімді алдын-ала ертерек болжау үшін (наурыз-сәуір айларындағы қар қалыңдығы жөніндегі спутниктік деректерді пайдаланып ) және алдын ала жасалынатын шарттардағы астық бағасын болжау арқылы АҚШ-тағы астық өндірушілеріне үлкен пайда алып келді.

*Сипаттайтын модельдің* (descriptive model) негізгі мәселелеріне жиналған үлгілердегі, деректер жиынындағы шаблондар (pattern) мен қатынасты табу жатады.

*Кластерлеу* (clustering) берілген деректер жиынтығындағы заңдылықтарды, шаблондар мен мәндердің арасындағы өзара байланысты зерттеуге негізделген жаңа кластар және топтарды құруға мүмкіндік беретін қандай да бір мағынада жіктеуге келетін деректерді өңдеу. Ол жіктеуге ұқсас болады, бірақ алдын ала топтарды немесе кластарды анықтауды қажет етпейді. Кластерлеу ғылыми әдебиетте өзін-өзі үйрету немесе сегментация (unsupervised learning or segmentation) ретінде белгілі. Кластерлеу кезінде деректердің бір-біріне ұқсас келетін барлық нұсқалары кластер деп аталатын бір топқа жиналады. Мысал ретінде бірдей өнімді шығаратын кәсіпорын топтарын; бірдей қасиетке ие өнімдер; бірдей талғамға ие адамдар топтарын алуға болады және т.б.

*Жинақтылығы немесе жалпыламалылығы* (summarization / generalization) Web-беттерінде немесе құжатта орналасқан деректер бөліктерін кіші көлемді қамтитын құжатқа қосуға мүмкіндік беретін әдістер. Басты идея, бүкіл жиынтықтың басты ақпаратын қамтитын, деректердің көрнекті бөліктерін табудан тұрады. Мысал ретінде оның негізгі күйін, түйін сөздерін қамтитын үлкен мақаланың андатпасын қарастыруға болады. Жинақтылық технологиясы бүгінде өндірістің көптеген секторларында пайдаланылады. Тағы бір мысал ретінде осы технологияның Google тәрізді іздеу жүйелерінде пайдаланылуын жатқызуға болады. Басқа мысалдарға видео материалдар немесе суреттер топтамасының мазмұндарын жалпылау жатады. Жинақталған құрама деректерді зерттеу бүкіл Web-парақты немесе бүкіл құжатты анықтауға мүмкіндік береді. Осылайша, жинақтылықты басты сипаттамаларды анықтау немесе үлкен деректер жиынтығын қорытындылау ретінде де түсінуге болады.

*Ассоциация ережесі* (Association rules) анықталған белгілер және сипаттамаларға негізделіп жүйелендірілмеген деректер тобының арасындағы қарым қатынасты және белгіні тағайындауға мүмкіндік береді. Бұл әдіс деректер элементтерінің арасындағы кейбір ассоциативті ережені анықтауға көмектеседі, содан кейін осы ережелер қарым қатынастарды орнатуға пайдаланады, мысалы, үлкен және реттелмеген деректер банкісінде болуы мүмкін реттіліктің типтік бейнесін айқындауда пайдаланылады. Банкте реттіліктің пайда болуы кезінде уақыт факторы есепке алынады, яғни деректер тобының арасындағы байланыс олар алынған уақытқа сәйкес келеді. Мысалы, қауіпсіздік ұйымдарына оқиғаның сипатына байланысты реттілігін зерттеу және талдау, сонымен қатар, полиция қызметкерлеріне қылмыстарды ашуға және олардың болашақта қайталануын болдыртпауға көмектеседі. Осыған ұқсас, ерекше және белгісіз аурулардың пайда болуы, нақты мезгілге байланысты келеді. Осындай зерттеулер биологтарға профилактикалық препараттарды ойлап табу және алдын-алу шараларды ұсыну кезінде көмектеседі.

#### **Әдебиеттер тізімі**

1. Газуль С.М., Ананченко И.В., Кияев В.И. Совершенствование образовательного процесса в вузе: активные методы обучения и гибридные информационные системы на основе виртуализации // *Совр. проблемы науки и образования*. – 2015.
2. Тюлепбердинова Г.А., Унайбаева Р.К., Шанлаякова А.С. Процесс компьютеризации и информатизации общества // *ТРУДЫ Международной научно-практической конференции «Информационные и телекоммуникационные технологии: образование, наука, практика»*, - Алматы, Казахстан, 5-6 декабрь, 2012 стр. 204-207.
3. Б.А. Урмашев, Ф.Р. Гусманова, Г.Г. Газиз, Г.А. Тюлепбердинова, М.Ж. Сақыпбекова, А. Алтыбай. *Ақпараттық-коммуникациялық технологиялар: Оқу құралы*. Алматы, КазНУ, 2017.
4. Сейтбекова Г.О., Тюлепбердинова Г.А. *Ақпараттық технологиялар: оқу құралы / Г.О.Сейтбекова, Г.А.Тюлепбердинова - Алматы: «Эвер» баспасы, 2015. - 256 б.*