

ТІЛ – ИНСТИНКТ

Сөйлейтін бастар





Жоспар:

- Есептеуіш машиналарды қолданысқа енгізу мәселелері;
- Менталды синтаксистік;
- Ағаштың құрылымы бойынша сөйлемдердің бұтақталуы;
- Сөйлемдерді дұрыс талдау түсіндірмелері.

Жүздеген жылдар бойы жасаған есептеуіш машиналар оларды жасап шығарған адамдарды мазалады. Бұның себебі де бар шығар. Мүмкін бұл машиналар адамдардың өзінен асып түсіп алдап соғып, жұмыс орнындарынан айыратын шығар деген оймен адамзатты күдіктендірді. Мұндай қорқыныш сезім орта ғасыр еврей аңыздарынан басталады. Алайда 1950 жылы «Artificial intelligence» – «Жасанды интеллект» сияқты қолданбалы ғылым саласы қалыптасқанда ғылыми фантастика үрейлі шындыққа айналатын сияқты әсер пайда болды.

Дегенмен үй роботтары әлі де болса фантастика күйінде қалған. Artificial intelligence құрастырушылары отыз бес жылғы еңбектен кейінгі алған басты сабағы – күрделі мәселелер қарапайым, ал қарапайым күрделі болып келетіндігіне көз жеткізу. Төрт жасар бала санасына жеткілікті және біз үшін өзінен-өзі түсінікті болатын – таныс адамды тану, қарындашты көтеру, бөлмеге өту, сұраққа жауап беру – шын мәнінде ең күрделі технологиялық мәселелердің бірі болып табылады.

Жақында біздің ортамызға компьютерлік бағдарлама пайдаланушыны адаммен тілдесіп тұрмын деп ойлауға мәжбүрлеп, оны ең көп жаңылыстыратын анықтауға арналған жыл сайын өтетін байқау өтті. Бұл байқаудың мақсаты Алан Тьюрингтің 1950 жылы әйгілі еңбегінде жасаған пайымдауын Лобнер (Loebner Prize) сыйлығына іске асыру болды. Ол «Машина ойлай ала ма?» деген философиялық сұрағының жауабына арбитр бір терминал арқылы адаммен және екінші терминал арқылы адам имитациясына бағдарланған компьютермен сөйлесетін ойынды ұсынды. Егер арбитр кімнің қайда екенін ажырата алмаса, онда Тьюрингтің ойы бойынша, компьютер де ойлай алады дегенді жоққа шығару негізсіз. Егер философияны алып тастаса, бір де бір бағдарлама 100 000 долларлық сыйақыны ұтпақ тұрмақ, маңына да бара алмайтыны байқау комиссиясы үшін анық болды. Сондықтан да оның көлемі 1500 долларға азайтылды, машина үшін бұл әділ болады.

Адам мен машинаның арасындағы сұхбат шамамен жеті минутқа созылды. Тілдесудің бағыты төменгідей болды.

Арбитр: Бұл күн қаншалықты ұзақ еді! Алайда енді мен кішкене көңіл көтеруге дайын сияқтымын. Күн райы бүгін онша емес. Бұлттарды айдайтын қандай да бір әзілдерді сен білесің бе?

Компьютер: Бәрі күн райы туралы айтып жатыр, дегенмен де ешкім оған ешнәрсе жасамайтын сияқты ғой. Неліктен сен маған енді кішкене көңіл көтеруге дайын екендігіңді міндетті түрде айтуың керек еді? О-о-о-ох...Осындай тұрғыда сұхбат жалғасты.

Машинамен адам арасындағы сұхбат аяқталған соң арбитрлер компьютерлердің кез келгеніне қарағанда адамдар барлығы «адамтекес» екеніне нұсқады. Бірақ шамамен арбитрлердің жартысы ұтқан бағдарламаны адаммен шатастырды.

Шын мәнінде 1965 ж. бағдарламаны енгізген компьютерші-ғалым Джозеф Вейценбаумның мақсаты ойлап тапқан технологияны жетілдіру болды.

Дегенмен сол кезде Вейценбаумның қорқынышына орай, ол бір күні оның хатшысы терминалға жабысып, бағдарламаға жан сырын ақтарып жатқанын көрген.

Лобнер сыйлығына байқау көптеген компьютерші-ғалымдардың ашуын тудырады. Олар оны компьютерді сөйлетуге үйрету емес, кәсіби маман еместі алдап соғу міндеті болып табылатын ғылыми құндылығы жоқ жарнамалық трюк ретінде қарады.

Нәтижесінде тілді түсінетін компьютерлік жүйені жасау бойынша қарқынды жұмыс жүргізілді. Алайда бірде-бір байыпты инженердің адамның тілдік қабілетін қайталайтын жүйенің пайда болуы алыс емес деп айтуға батылы жетпес.

Шындығында, ғалымдардың көзқарасы бойынша, машина адамдардың сөйлемдерін соншалықты жақсы түсінуге құқы жоқ. Олар санаға кірмейтіндей күрделі есептерді жай ғана шешіп қоймайды, оны тез жасайды. Ұғыну, негізінен, «шынайы уақыт» режимінде іске асады.



Тыңдаушы сөйлеушімен кітапқа пікір жазған сыншы сияқты сөйлем бөлігінің соңына дейін соған сәйкес уақыттан кейін түсіну үшін күтпей, бірдей әрекет етеді. Сөйлеуші айтқан және тыңдаушы қабылдаған уақыт аралығы өте қысқа: жарты секундта бір немесе екі буын. Кейбір адамдар сөйлеуші сөйлеміне ілесе, секундтің төрттен бір бөлігіне қалып, сөйлемдерді түсініп, қайталай алады.

Әңгімелесуші машина жасаудан басқа ұғыну процесін түсіну өзге де практикалық қолданысқа ие. Адамдардың сөйлемдерді түсінуі тез әрі тиімді болады, алайда кемшіліксіз емес. Ол қабылданатын сөйлем немесе мәтін нақты бір құрылымданған болса, іске асады. Олай болмаған жағдайда процесс тірелеп қалуы, кері кетуі мүмкін және бұрыс түсіну қалыптасады.

Осындай келтірген деректермен Стивен Пинкер тілді ұғынуды зерттеп, қабылдаушы адамның санасында қандай сөйлемдер түрі шатасуы мүмкін екендігін анықтауға тырысты.

Екінші практикалық қолданыс құқық саласымен байланысты. Судьялардың орта статистикалық адам екі ойлы мәтін үзіндісін, мысалы, клиент келісімшартты оқығанда, алқабилер нұсқаулықты тыңдағанда және жұртшылық өкілі потенциалды жала сипаттаманы оқығанда қалай қабылдайтыны туралы жиі жорамалдауына тура келеді. Мәтінді түсіндірудің адамға тән әдеттерінің көбі зертхана зерттеулерінің арқасында анықталған, ал лингвист және заңгер Лоренс Солэн тіл мен құқық арасындағы байланысты өзінің 1993 жылы жарияланған «Судьялар тілі» атты қызықты кітабында түсіндірген.

Пинкердің келесі мәселесі – сөйлемді түсіну қалай іске асады? деген сұрақпен сипатталады. Бірінші кезең – бұл синтаксистік талдау. Ол сіз бастауыш мектепте тістеніп, жасаған жаттығуларға еш қатысы жоқ, олар туралы «Тіл мырзадан сұра» газет айдарын жүргізуші Дейв Барридің мынадай естелігі қалған:

Сұрақ: Өтінемін, түсіндіріңізші, сөйлем сызбасын қалай құруға болады?

Жауап: Ең алдымен сөйлемдерді таза, түзу тегістікке, мысалы, үтіктеу тақтасына жайғастырыңыз. Содан кейін, үшкірленген қарындаш пайдаланып, әрекеттің қайда болғанын көрсететін және негізінен желбезектердің артында орналасқан «баяндауышты» мүшелеңіз. Мысалы, «ЛаМонтты орманшылар ешуақытта тәртіпке келтіре алмас» сөйлемінде әрекет, сірә, орманда болып жатыр. Сонымен, сіздің сызбаңыз формасы жағынан бұтақты ағашқа ұқсас болар.

Дегенмен ұғыну шынында да санасыз түрде жасалатын, осыған ұқсас процесс бастауышты, баяндауышты, толықтауышты т.б. табудан тұрады.

Грамматиканың өзі код немесе хаттама, нақты бір тілде дыбыстардың қай түрі мағынаның қай түріне сәйкес келетінін анықтайтын мәліметтердің статикалық базасы болып табылады. Бұл сөйлеу және түсіну үшін рецепт не бағдарлама емес.

The dog likes ice-cream «Бұл ит балмұздақты жақсы көреді» сөйлемін алайық. Менталды синтаксистік талдауыш тап болатын сөз – бұл the. Синтаксистік талдауыш менталды сөздіктен ол туралы анықтама жинайды; бұл ереженің оң жағынан осы сөзді табу және сол жақта көрсетілген оның категориясын анықтау дегенмен бірдей.

Осы кезеңде сөйлемнің мағынасын қандай да бір деңгейде болжауға болады. Атауыштық топта бір зат есім өзегі – басы болып табылатынын еске түсіріңіз (барлық топ айтып жатқан нәрсе), ал атауыштық топ құрамындағы қалған синтаксистік топтар бұл зат есімді анықтай алады.

Мұндай операцияларды орындауға компьютерлерді бағдарлау нәтижесінен соншалықты қиын? Бюрократиялық құжаттарды және стилі нашар басқа да үлгілерді оқығанда адамдарға да ол неге кенеттен қиын бола қалады? Сөйлемдерді синтаксистік талдауыш рөлінде қарастырғанда, біз нәтижеге жету жолында екі қиындыққа тап болдық. Біріншісі – бұл жады: бізге жадымызда жасақталу үшін сөздің нақты түрлерін талап еткен жасақталмаған топтарды сақтап тұруымыз керек болды. Екіншісі – бұл шешім қабылдау: синтаксистік топ екі түрлі ереженің оң жағында болғанда, ағаштың келесі бұтағын өсіру үшін олардың қайсысын пайдалану жөнінде шешім қабылдау керек болды. Күрделі мәселелердің жеңілдігі және жеңіл мәселелердің күрделілігі туралы жасанды интеллекттің бірінші ережесіне сәйкес жадыға қатысты бөлігі компьютер үшін жеңіл және адамдар үшін ауыр, ал шешім қабылдауға қатысты бөлігі адамдар үшін жеңіл (тым болмағанда, егер сөйлем дұрыс құрастырылса) және компьютерлер үшін ауыр.



Сөйлемнің синтаксистік талдауышы жады түрлерінің көп болғанын талап етеді. Алайда бәрінен да айқын қажеттілік – бұл жасақталмаған құрылымдарға жады, синтаксистік талдауға ұшырайтынды есте сақтау. Компьютерлер өз жадында осы мақсатқа арналған «сөре» деп аталатын ұяшықтар жиынтығын бөлу керек; тап осы синтаксистік талдауышқа жай ғана сөз тізбегінің генераторы болуға емес, тікелей құрастырушылар құрылымының грамматикасын пайдалануға мүмкіндік береді. Адамдар да тура сол сияқты өздерінің жедел жадының біршама бөлігін жасақталмаған синтаксистік топ үшін бөлуі керек. Алайда жедел жады – бұл ақпаратты адамның өңдеу процесіндегі негізгі «тар жері». Тек бірнеше бірлік негізінен – жеті, плюс-минус екі – бір уақытта санада сақталуы мүмкін; сонымен қатар, олар бірден «сіңе» бастайды немесе оларды басқалары ығыстыра бастайды. Келесі сөйлемдер мысалында сіз жасақталмаған топты жадыда тым ұзақ ұстағаннан не болатынын көресіз:

Адамдардың белгілі жағдайда ісі нашар және компьютерлерде жақсы болып тұрған жадымен салыстырғанда, шешім қабылдау – бұл адамдардың күшті, ал компьютерлердің әлсіз болатын тұсы.

Осы жерде Жедел жады деген мәселеге тоқталған жөн.

Жедел жады – компьютерлік есептеулерге арналған мәліметтерді белгілі бір уақытқа дейін сақтайтын құрылғы. Мәліметтер 0,1 түрінде жедел жадыға жазылады. Процессор мәліметтерді жедел жадыдан алып, оларды өндейді.

Жедел жады – оперативті жады, орысша баламасы – оперативное запоминающее устройство – ОЗУ, Random Access Memory-RAM) немесе жедел жадтайтын құрылғысы – ЖЖҚ аббревиатурасымен айтылады, ақпараттың кез келген бөлігіне лезде қатынауды қамтамасыз етеді. Жедел жады, жүйелік логика жиынтығымен және орталық процессорымен бірге, кез келген дербес компьютердің негізін құрайды.

Жедел жады – компьютердің тез әрекет етуі, көлемі мен жұмыс жылдамдығына байланысты болатын құрылғылардың бірі. Бірақ компьютерді өшірген кезде жедел жадының барлық ақпараты бірден жойылады. Дербес компьютерлердің жедел жадының өлшемі жылдан жылға өсіп келеді. Компьютерлердің жедел жадының көлемі 32 Мбайттан 2Гбайтқа жетеді. ЖЖҚ-ның құрылысы оны үнемі ұлғайтып отыратындай етіп жасалынған. Компьютердің жедел жадының көлемі өскен сайын оның есептеу жылдамдығы артады.

Жедел жад – ерікті жету жады деп айтылады. Бұл процессормен тікелей байланысты және программалар өңдейтін мәліметтерді жазу, оқу және сақтауға арналған аса үлкен емес көлемді жылдам сақтау құрылғысы.

Құрастырған сөйлемдегі грамматика қарапайым – «ойыншық» болды, ал біз бірте-бірте өткен сөйлемнің өзі – «баланікі», онда әр сөз үшін тек бір ғана сөздік мақала болды (ал сөздің өзі тек бір ереженің оң жағында болды). Алайда сөздікті тек ашқаның өзінде – қаншама зат есімдердің бұл сөздер етістік және керісінше қарастырылатын екінші сөздік мақала бар екенін сіздер көресіздер. Мысалы, Scandals dogged the administration all year «Жыл бойы жанжалдар әкімшілікті шайқалтты» сөйлемінен көрініп тұрғандай dog «ит» сөзінің екінші – етістікті – сөздік мақаласы бар. Сол сияқты шынайы өмірде hot dog «ыстық шұжықшалы тоқаш» – бұл тек зат есім ғана емес, «селтеңдеу» дегенді білдіретін етістік те.

Стивен Пинкердің грамматикасында әр етістік те зат есім мағыналы екінші сөздік мақалаға ие болуы керек. Себебі ағылшын тілінде сөйлеушілер арзан eats «жейтін тамақ», біреудің likes және dislikes «ұнататын және ұнатпайтын нәрселері» және бірнеше bites «тілімді» іліп әкету туралы сөз етуі мүмкін.

Сөздер синтаксистік топтарға біріккенде жағдай күрделене түседі. Себебі соңғылары әртүрлі тәсілдер арқылы үлкен синтаксистік топтарға кіруі мүмкін.

Мұндай екі ұштылық бірыңғай жағдай емес, ереже болып табылады; сөйлемнің әр нүктесінде (жерінде) тексеруді қажет ететін ондаған жүз нұсқасы кездесуі мүмкін.

Грамматикалық тұрғыдан болатын, алайда табиғи емес баламаға көңіл бөлмей, адамдардың қолынан сөйлемдерді дұрыс талдау қалай келеді екен? Мүмкін болатын екі түсіндірме бар. Біріншісі: біздің миымыз артқы жағында болашағы жоқ және қабылданбайтын, санаға жеткенше басында әлдеқалай сүзгіден өтіп кететін ондаған ағаш кесінділерін өңдейтін компьютердің синтаксистік талдауыш үлгісі сияқты орналасқан.



Екінші түсіндірме: адам басындағы синтаксистік талдауыш әр деңгейде қандай да бір жолмен қай балама барынша мүмкін болатынын таразылайды да, осы жалғыз түсініктеме үшін мүмкіндігінше алға қарай жолды ашады. Компьютерші-ғалымдар мұндай мүмкіндіктерді «екінші ізденіс» және «бірінші ізденіс» деп атайды. Ұзақ уақыт болмаса да, дұрыс болмауы мүмкін екіұшты сөздер үшін мүмкін болатын бірнеше сөздік мақаланы назарға алып, жеке сөздер деңгейінде ми алдымен екінші ізденісті жүргізетін сияқты.

Адам қандай да бір сөзді естіген кезде, басқа сөзді, менталды сөздік тақырыптық сөздік қағидат бойынша бір сөзді тапқан кезде, мағынасы жақын басқалары жылдамырақ табылатын – ұйымдастырылғаны сияқты. Егер ол бірінші сөзге қатысты болса, тану жеңіл болатыны жақсы белгілі. Күткеніміздей, адамдар онымен ешқандай байланысы жоқ «тігу» сөзіне қарағанда, «қоңыз» сөзіне мағынасы жағынан жақын «құмырсқа» сөзін танып, тетікті жылдамырақ басқан. Таңырқарлық болмаса да, адамдар мәнмәтінде ешқандай мәні жоқ, басқа мағынада қоңызбен, әрине, байланысы бар «тыңшы» сөзін де сондай жылдамдықпен таныған. Бұл басында біреуін қабылдамау ақылға қонымды екеніне қарамастан, мидың бірден «қоңыз» сөзі үшін екі сөздік мақаласын беретінін болжайды. Үйлеспейтін мағына санада ұзақ сақталмайды: «қоңыздан» кейін, бірден одан кейін емес, үш буын арқылы экранда бақыланатын сөз шыққан кезде, тек «құмырсқа» сөзі тез танылды, «тыңшы» тану жылдамдығы бойынша одан соң «тігуден» озбады.

Сонымен, менталды сөздікте іздеу шапшаң әрі мұқият жүреді, алайда аса мағыналы емес. Себебі соңынан сұрыптаудан өтетін мағынасыз сөздік мақалалардан алынады.