

Беларускамоўная галасавая пытальна-адказная сістэма

Юрсь Гецэвіч,

загадчык лабараторыі распазнавання і сінтэзу маўлення АІПІ НАН Беларусі, кандыдат тэхнічных навук, дацэнт

Яўгенія Зяноўка,

малодшы навуковы супрацоўнік лабараторыі распазнавання і сінтэзу маўлення АІПІ НАН Беларусі

Вольга Дыдо,

стажор малодшага навуковага супрацоўніка лабараторыі распазнавання і сінтэзу маўлення АІПІ НАН Беларусі

Максім Люціч,

аператар ПЭВМ лабараторыі распазнавання і сінтэзу маўлення АІПІ НАН Беларусі

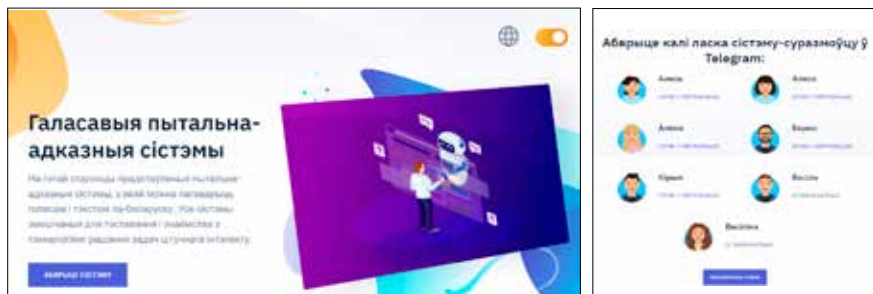
Марыя Павуціна,

стажор малодшага навуковага супрацоўніка лабараторыі распазнавання і сінтэзу маўлення АІПІ НАН Беларусі

У сучасным інфармацыйным асяроддзі найбольш перспектыўнымі напрамкамі навуковых даследаванняў штучнага інтэлекту з'яўляюцца аўтаматычная апрацоўка тэкстаў на натуральнай мове, сістэмы сінтэзу і распазнавання маўлення, машыннае навучанне. Камбінацыя дадзеных тэхналогій адкрывае магчымасць стварэння разнастайных праграм, у тым ліку пытальна-адказных, ці галасавых асістэнтаў (ад англ. Question Answering Systems). Гэта інфармацыйная сістэма, здольная прымаць пытанні і выдаваць асэнсаваны адказ. На сённяшні дзень рынак камп'ютарных праграм прадстаўляе вялікі шэраг галасавых асістэнтаў, якія запатрабаваны для розных мэт і шырокай сферы прымянення. Усім ужо вядомыя Siri, Google Assistant (Google Now), Amazon

Alexa, Microsoft Cortana, Vixby, Voice Mate, Аліса, Маруся, Дуся, Салют, якія падтрымліваюць розныя мовы, за выключэннем беларускай. Таму для беларускамоўных носьбітаў каманда лабараторыі распазнавання і сінтэзу маўлення Аб'яднанага інстытута праблем інфарматыкі (АІПІ) НАН Беларусі [1] распрацавала набор галасавых пытальна-адказных сістэм, якія змешчаны на платформе «АІ-асістэнт» (мал. 1) [2]. З дадзенымі галасавымі асістэнтамі можна паразмаўляць па-беларуску.

Мэтай распрацоўкі асістэнтаў з'яўляецца забеспячэнне эфектыўнага і простага ў выкарыстанні механізма прадстаўлення агульнай інфармацыі і рашэння пытанняў карыстальнікаў на беларускай мове. Пытальна-адказная сістэма «Галасавы АІ-асістэнт» дазваляе карыстальніку голасам ці ўводам тэксту з клавіятуры задаць пытанне на беларускай мове і атрымаць на яго гукавы ці надрукаваны адказ. За кошт выкарыстання штучнага інтэлекту яна дае магчымасць атрымліваць хуткія, якасныя і дакладныя адказы па розных тэмах. У выніку штодзённага навучання галасавы



Мал. 1. Інтэрфейс афіцыйнага сайту пытальна-адказных сістэм

асістэнт можа як весці гаворку па навуковых тэмах, так і зрабіць забаўляльныя прапановы.

Працэс функцыянавання любой пытална-адказнай сістэмы адбываецца ў некалькі этапаў:

- *аналіз запыту, уведзенага карыстальнікам. Пры гэтым увадзіцца пытанне на натуральнай мове, пасля чаго ажыццяўляецца першаасная апрацоўка і фармалізацыя сказа рознымі аналізатарамі (сінтаксічным, марфалагічным, семантычным). Сістэма вызначае яго адпаведныя атрыбуты для далейшага іх выкарыстання;*
- *пошук і аналіз інфармацыі (сістэма шукае дакументы і іх фрагменты, у якіх можа знаходзіцца адказ на зыходнае пытанне);*
- *вывад адказу: сістэма выцягвае з тэкставых дакументаў або іх фрагментаў словы, сказы або ўрыўкі тэксту, якія патэнцыйна могуць быць адпавядаць запыту, і выдае адказ карыстальніку.*

Пералічаныя этапы рэалізаваны ў беларускамоўных пытална-адказных сістэмах, якія пабудаваны з дапамогай тэхналогій распазнавання і сінтэзу маўлення, машыннага перакладу і дыялогавых сістэм (мал. 2). Каб пачаць дыялог з асістэнтам, можна запісаць галасавое паведамленне ці ўвесці тэкст з клавіятуры. Для апрацоўкі вымаўленага пытання выкарыстоўваецца пабудаваная сістэма распазнавання беларускага маўлення (БСРМ) высокай якасці, заснаваная на end-to-end архітэктурой з выкарыстаннем глыбокага навучання. Яна змешчана на платформе Hugging Face [3], якая дазваляе карыстальнікам ствараць і абмень-

вацца мадэлямі машыннага навучання і наборамі даных. Для распрацоўкі БСРМ быў сабраны вялікі корпус начытаных тэкстаў на беларускай мове. Агульная працягласць аўдыязапісаў складае 987 гадзін, у іх агучванні прынялі ўдзел 6160 дыктараў. Гэта першы з падобных датасэтаў такога памеру для беларускай мовы. Высокая варыятыўнасць сабраных даных як адносна дыктараў (пол, узрост, тэмп маўлення, іншыя асаблівасці), так і адносна ўмоў запісаў (розныя мікрафоны, наяўнасць фонавага шуму і інш.) дазволіла навучыць сістэму распазнавання маўлення працаваць ва ўмовах, набліжаных да тых, з якімі ёй даведзецца мець справы ў штодзённым жыцці. Мадэль распазнавання маўлення была навучана на сучаснай глыбокай нейрасеткавай архітэктурой Whisper. Яе асаблівасцю з'яўляецца перадаванне на корпусе неанатаваных даных (у рэжыме без настаўніка) для вывучэння спосабаў якаснага вылучэння прыкмет па ўваходным аўдыязапісе. Атрыманыя дадзеныя выкарыстоўваюцца для далейшых падзадач: напрыклад, для давучвання мадэлі пераўтвараць маўленне ў тэкст. У якасці перадаванай мадэлі была абраная openai/whisper-small. Канчатковы вынік навучання сістэмы нейроннымі сеткамі складае WER 0,679 (або 6,79%), што з'яўляецца даволі добрым для мадэляў распазнавання [4]. Так, напрыклад, цяперашняе найлепшае значэнне test WER для нямецкага датасэту Common Voice роўнае 5,7%.

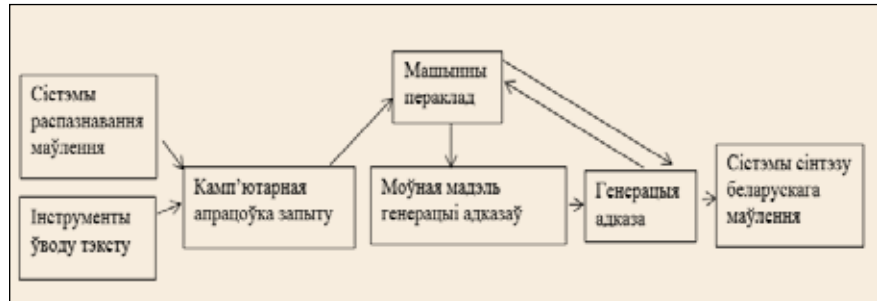
На наступным этапе адбываецца апрацоўка запыту тэкставым працэсарам, дзе сістэма аналізуе тэкст. Важна правільна

распазнаць не толькі вымаўленыя словы, але і лікі, абрэвіатуры, скарачэнні і тыя словы, якія карыстальнік прамовіў неадкладна ці невыразна. За гэта адказвае блок камп'ютарнай апрацоўкі запыту. Далей ён перадаецца ў блок моўнай мадэлі генерацыі адказаў, якая можа ўяўляць сабой калекцыю пытанняў і адказаў, дакументаў, у якіх адбываецца пошук, ці, напрыклад, web-пошук у інтэрнэт-прасторы. Для беларускамоўных пытална-адказных сістэм выкарыстоўваецца моўная мадэль ChatGPT-3,5, для трэніроўкі якой ужываюцца метады навучання з настаўнікам і з падмацаваннем. GPT з'яўляецца серыяй моўных мадэляў, распрацаваных кампаніяй OpenAI. Мадэль папярэдне навучана на велізарных наборах тэкставых даных. Дзякуючы гэтаму GPT можа генераваць тэкст, які мае сэнс, выкарыстоўвае правільную граматыку і структуру сказаў. Сістэма здольна адаптавацца да розных стыляў напісання і фарматаў (артыкулы, дыялогі, справаздачы і г.д.). ChatGPT адсочвае стан карыстальніка, запамінаючы папярэднія пытанні і адказы, прыведзеныя ў той жа размове. У спробе прадухіліць выдачу абразлівых выразаў у ChatGPT запыты фільтруюцца праз API-мадэрацыі, і ўсе непажаданыя падказкі адхіляюцца. Таму прадстаўленая сістэма з'яўляецца карысным інструментам пошуку і выдачы інфармацыі.

ChatGPT-3,5 падтрымлівае беларускую мову, але якасць адказаў не вельмі добрая. Для яе паляпшэння распрацаваны дадатковы блок машыннага перакладу, у якім ўсе запыты аўтаматычна перакладаюцца на англійскую мову з дапамо-

гай сістэмы Google Translate, потым перадаюцца ў моўную мадэль. Найбольш рэlevantны адказ, выдадзены моўнай мадэллю, зноў пераходзіць ў блок машыннага перакладу, дзе на выхадзе прадстаўляецца ўжо беларускамоўны варыянт.

Апошнім крокам прадстаўлення дадзеных карыстальніку з’яўляецца пераўтварэнне тэкставай інфармацыі ў аўдыяфайл. Для гэтага выкарыстоўваецца беларускамоўная сістэма сінтэзу маўлення па тэксце (БССМТ), размешчаная ў адкрытым доступе на платформе Hugging Face [5]. Для стварэння якаснай акустычнай мадэлі быў выбраны вялікі агучны тэкставы корпус з інтэрнэт-сайтаў donar.by і Common Voice (каля 20 гадзін). З дапамогай нейронных сетак ажыццяўлялася трэніроўка і навучанне акустычнай базы. Такім чынам яна створана аўтаматычна і мае вельмі дакладныя вынікі. Аднак дадзены сінтэзатар вялікі па памеры, што можа зніжаць хуткасць падрыхтоўкі тэкставай інфармацыі. Таксама вялікім недахопам лічыцца адсутнасць апрацоўкі лікаў, лічбаў, дат і абрэвіятур. Для яго вырашэння выкарыстоўваецца дадатковы блок беларускамоўнага сінтэзатара маўлення па тэксце [6, 7], прадстаўленага на платформе для апрацоўкі тэкставай і гукавой інфармацыі розных тэматычных даменаў cognus.by [8]. Асабліва гэта мадэль заключаецца ў прадстаўленні як сінтэзаванага маўлення, так і дадатковых прамежкавых вынікаў лінгвістычнага ўдасканалення ўваходнай інфармацыі. Так, напрыклад, карыстальнік можа пазнаёміцца з анатацыяй кожнага слова, знойдзенага сістэ-



Мал. 2. Структура пытальна-адказных сістэм «Галасавы AI-асістэнт»

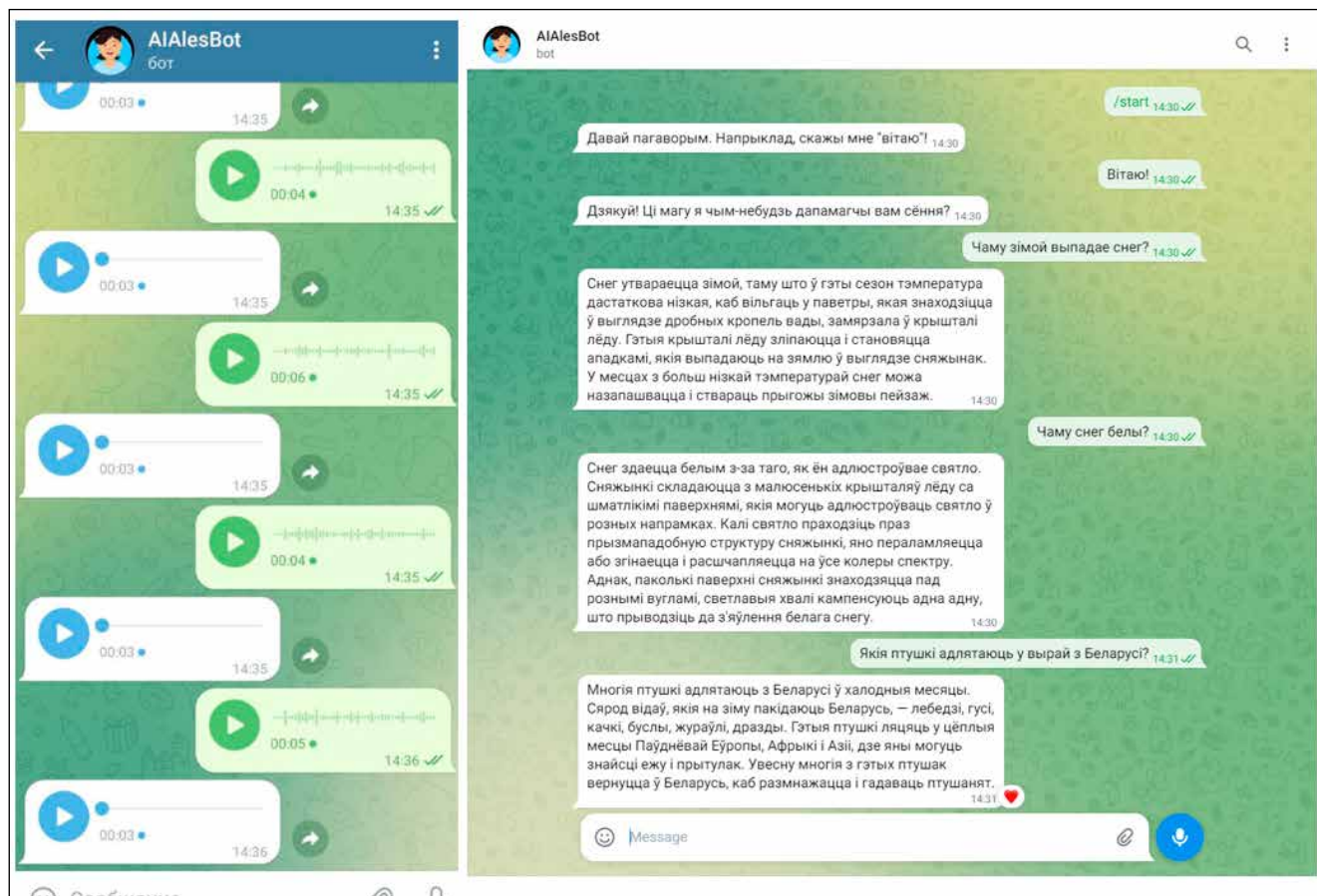
май у тэксце; са спісамі слоў з прастаўленымі націскамі і тых, якія адсутнічаюць у базе даных сістэмы; амаграфамі, якія сустраляся ў тэксце; спісам слоў у фанемным і алафонным выглядзе. Дадатковая функцыя сінтэзатара – генерацыя транскрыпцыі ўваходнага тэксту ў 4 фарматах: кірылічна, міжнародная (IPA), спрошчаная міжнародная і X-SAMPA. Дадзены сінтэзатар – мадэль кампілятыўнага сінтэзу, які ўяўляе сабой інтэграцыю маўлення шляхам канкатэнацыі запісаных узораў асобных алафонаў, вымаўленых дыктарам. Акустычная база даных складаецца з гукавых фрагментаў, якія ў працэсе апрацоўкі ўваходнай тэкставай інфармацыі аб'ядноўваюцца ў склады/словы/словазлучэнні і пераўтвараюцца ў гукавыя файлы. Гэты спосаб забяспечвае высокую якасць сінтэзуемага маўлення, паколькі дазваляе аднавіць форму натуральнага маўленчага сігналу. Галоўным недахопам апісанай сістэмы з’яўляецца штучнае, рэбатызаванае гучанне. Таму перавага аддадзена БССМТ, прадстаўленай на платформе Hugging Face з дабаўленнем дадатковага блоку тэкставай апрацоўкі лікаў, лічбаў, дат і скарачэнняў.

На бягучы момант галасавыя асістэнты даступны ў фармаце тэлеграм-ботаў (сем гала-

соў). Карыстальнік можа абраць любога асістэнта з прапанаваных на сайце asistent.by і пачаць з ім гутарку. Кожная пытальна-адказная сістэма мае імя і прыемны голас, што ачалавечвае яе, і ў карыстальніка з’яўляецца ўражанне размовы з рэальным суразмоўцам. Таксама галасавы асістэнт «Алесь» даступны праз тэлеграм бот @AIAlesBot. Дадаткова вядуцца працы па распрацоўцы web-версіі.

На мал. 3 прадстаўлены дыялог чат-бота Алеса з карыстальнікам. Адметнасць сістэмы заключаецца ў магчымасці як вуснай камунікацыі, так і перапіскі з дапамогай клавіятуры. Сістэма якасна распазнае галасавыя і тэкставыя паведамленні і дае адказы на простыя абстрактныя пытанні, якія тычацца агульнапрынятых і вядомых фактаў. Часам узнікаюць праблемы падчас адказу на складаныя пытанні. Аднак дыялог сістэма вядзе без адхіленняў (падтрымлівае гутарку, адказвае на паслядоўныя пытанні). Таксама бот задае дадатковыя пытанні, каб працягнуць размову, напрыклад: «Чым я яшчэ магу дапамагчы?», «Калі ласка, звяртайцеся» і г.д. На кожнае паведамленне сістэма выдаткоўвае каля 10–20 сек., што даволі хутка.

Праект знаходзіцца ў актыўнай распрацоўцы,



Мал. 3. Дыялог з чат-ботам «Алесь» у Тэлеграме

у межах якой плануецца рэалізаваць наступныя пункты:

- *распрацаваць версіі асістэнта для Web, IOS і Android дзеля больш зручнага карыстання;*
- *стварыць новую версію дызайну і лагатыпа афіцыйнага сайта;*
- *дадаць англійскую, рускую і кітайскую мовы інтэрфейса афіцыйнага сайта asistent.by;*
- *прапрацаваць некалькі «тыпажоў» для асістэнта згодна прадметнай накіраванасці, напрыклад «вучоны», «дызайнер», «бізнесмен» і інш.;*
- *сбраць водгукі ад карыстальнікаў пра вынікі працы сістэмы.*

Такім чынам, пабудова і выкарыстанне пытална-

адказных сістэм адыгрывае важную ролю ў развіцці мадэляў зносінаў чалавека і камп'ютара. Мадэляванне падобных праграм дазваляе пашырыць магчымасці зручнага прымянення камп'ютарных тэхналогій у паўсядзённым жыцці праз дыялог з машынай. Вялікая колькасць галасавых асістэнтаў сведчыць аб зацікаўленасці распрацоўшчыкаў новых метадаў і алгарытмаў для паляпшэння якасці сістэм, у тым ліку галасавых памочнікаў, на беларускай мове, што вельмі актуальна для нашага асяроддзя. Платформа «AI-асістэнт» прадстаўляе шэраг беларускамоўных галасавых асістэнтаў, якія з'яўляюцца якаснымі суразмоўцамі і маюць шырокі патэнцыял прымянення. ■

СПІС ВЫКАРЫСТАНЫХ КРЫНІЦ

1. Лабараторыя распазнавання і сінтэзу маўлення // <http://ssrlab.by/>.
2. Галасавыя пытална-адказныя сістэмы // <https://asistent.by/>.
3. Ales/whisper-small-belarusian. Hugging Face // <https://huggingface.co/ales/whisper-small-belarusian>.
4. Трафімаў А.С. Аўтаматычнае пераўтварэнне беларускага маўлення ў тэкст / А.С. Трафімаў, Ю.С. Гецевіч // XXI Міжнар. навук.-тэх. канф. «Развіццё інфарматызацыі і дзяржаўнай сістэмы навукова-тэхнічнай інфармацыі РІНТІ-2022» / АПІ НАН Беларусі; пад навук. рэд. С.В. Круглікава і інш. – Мінск, 2022. С. 241–245.
5. Bel-tts / Hugging Face // <https://huggingface.co/jhlfrfufyfn/bel-tts>.
6. Сінтэзатар маўлення па тэксце // <http://corpus.by/TextToSpeechSynthesizer/?lang=be>.
7. Ю.С. Гецевіч. Компьютерная платформа для обработки электронного текста и речи на белорусском, русском и английском языках / Ю.С. Гецевич и др. // Речевые технологии. 2021. №1. С. 37–46.
8. Платформа для апрацоўкі тэкставай і гукавай інфармацыі для розных тэматычных даменаў corpus.by. 2022 // <http://corpus.by/>.