OSTIS-2015
(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

РАСПРАЦОЎКА КАМПАНЕНТА РАСПАЗНАВАННЯ МАЎЛЕННЯ ДЛЯ НАТУРАЛЬНА МАЎЛЕНЧАГА ІНТЭРФЕЙСУ

Нікалаенка К.А.[*], Кайгародава Л.І. [**], Гецэвіч Ю.С.[**]

[*] *Беларускі дзяржаўны ўніверсітэт інфарматыкі і радыёэлектронікі, Мінск, Рэспубліка Беларусь*

anak247@gmail.com

[**] *Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі, Мінск, Рэспубліка Беларусь*

lesia.piatrouskaya@gmail.com

yury.hetsevich@gmail.com

Апісваецца распрацоўка кампанента для аўтаматычнага распазнавання беларускага маўлення з мэтай кіравання рознымі мабільнымі прыстасаваннямі, у тым ліку робатамі. Сістэма распазнае абмежаваную колькасць каманд вызначаных дыктараў. Вынікі тэставання паказалі, што сістэма можа распазнаваць да 80% каманд. Прыведзены прыклад працы сістэмы распазнавання галасавых каманд, адаптаванай для PHP.
Ключавыя словы: кампанент; распазнаванне маўлення; інтэрфейс; натуральна маўленчы інтэрфейс.

Уводзіны

У аснове кожнай маўленчай тэхналогіі ляжыць так званы «engine» ці ядро праграмы — набор дадзеных і правіл, па якіх ажыццяўляецца апрацоўка дадзеных. У залежнасці ад прызначэння гэтага ядра адрозніваюць TTS і ASR engine. TTS (Text-to-Speech) engine дае магчымасць сінтэзу маўлення па тэксце, а ASR (Automatic Speech Recognition) engine прызначана для распазнавання маўлення. Існуе некалькі буйных вытворцаў, якія займаюцца стварэннем ASR ядзер. Сярод іх самымі даступнымі і папулярнымі з'яўлюцца Sphinx, HTK, Julius і Kaldi. Разгледзім іх.

CMU Sphinx складаецца з серыі распазнавальнікаў маўлення і трэніроўшчыка акустычнай мадэлі. Sphinx – гэта дыктаранезалежны распазнавальнік бесперапыннага маўлення, які выкарыстоўвае Схаваную Маркаўскую мадэль і n-грамную статыстычную моўную мадэль [Sphinx, 2014].

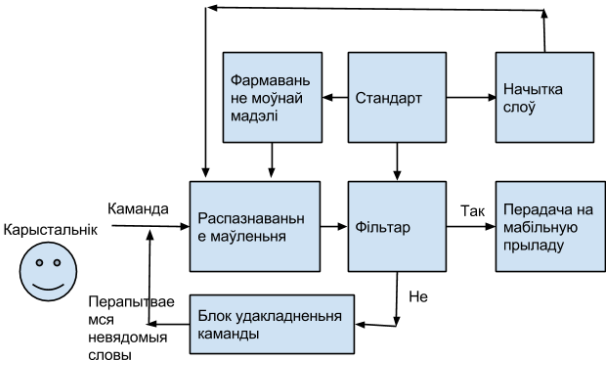
HTK — гэта інструментарый для распазнавання маўлення, што выкарыстоўвае Схаваную Маркаўскую мадэль. Праграмны пакет HTK — праграмнае забеспячэнне для апрацоўкі HMM мадэляў. HTK уяўляе сабою набор бібліятэк і інструментаў, якія могуць быць выкарыстаны ў аналізе і працы з маўленчым сігналам [HTK, 2014].

Julius — гэта распазнавальнік бесперапыннага маўлення з вялікім слоўнікам, дэкодар праграмнага забеспячэння для даследавання ў вобласці злучанага маўлення і распрацоўкі. Для запуску распазнавальніка маўлення Julius трэба падабраць моўную і акустычную мадэлі для мовы. Julius адаптуе акустычную мадэль кадаванага фармату HTK ASCII, базу дадзеных вымаўлення фармату HTK, і 3-х узроўневыя 2-грам пабудовы моўнай мадэлі стандарта [Julius, 2014].

Kaldi — падобны да HTK з пункту гледжання мэты і сферы ўжывання прадукт. Асноўнай мэтай распрацоўшчыкаў з'яўляецца стварэнне сучаснага і лёгка пераноснага кода, які лёгка мадыфікаваць і пашыраць [Kaldi, 2014].

Існуюць і іншыя, больш спецыфічныя сістэмы распазнавання маўлення, такія як iATROS, RWTH ASR, Simon, а таксама больш марудныя воблачныя сэрвісы Google ASR і Yandex ASR.

Мэта гэтай працы — распачаць стварэнне сістэмы распазнавання беларускага маўлення для кіравання мабільнымі робатамі. Для гэтага патрэбна стварыць акустычную базу і моўную мадэль, якая будзе максімальна блізкай да натуральнай мовы. Для нас важна асобна даследаваць гэтыя дзве задачы, а потым аб'яднаць у адно. Прынцыповая архітэктура сістэмы распазнавання галасавых каманд паказана на малюнку 1.



Малюнак 1 – Прынцыповая архітэктура сістэмы аўтаматычнага распазнавання маўлення

Для стварэння акустычнай мадэлі нам падыходзіць праграмны пакет HTK, так як ён мае магчымасці настройкі і апрацоўкі HMM мадэляў, а таксама мае вельмі падрабязную дакументацыю. Нездарма на ім будуецца акустычная база шматлікіх распазнавальнікаў. Для стварэння моўнай мадэлі будзем выкарыстоўваць інструментарый NooJ — настройвальны лінгвістычны працэсар, які дазваляе будаваць спецыялізаваныя электронныя слоўнікі, правілавыя сінтаксічныя і марфалагічныя граматыкі для апрацоўкі электронных тэкстаў (корпусаў ці тэкставых запытаў) у рэальным часе.

1. Праект моўнай мадэлі сістэмы распазнавання

На пачатку стварэння моўнай мадэлі мы выкарыстаем глыбокі сінтаксічны аналіз, каб атрымаць мадэль, якая будзе простай для ўспрымання, але таксама будзе адлюстроўваць паўнату мадэлі натуральнай мовы. Будзем ужываць такія канцэпты як *‘Суб’ект’*, *‘Дзеянне’*, *‘Аб’ект’* і *‘Характарыстыка’*. Пры выкарыстанні інструментарыя *NooJ Syntactic Grammar* праектуем графавую мадэль для аб'яднання гэтых канцэптаў і лінгвістычных адзінак, з якіх будуць складацца гэтыя канцэпты. Далей мы генерыруем слоўнік для робатаў з дапамогай інструментарыя *NooJ Dictionary*. Некаторыя адзінкі з гэтага слоўніка будуць выглядаць наступным чынам:

Робат_Віцебск прынясі лыжку, GUID=R1+Action=take+Object=spoon

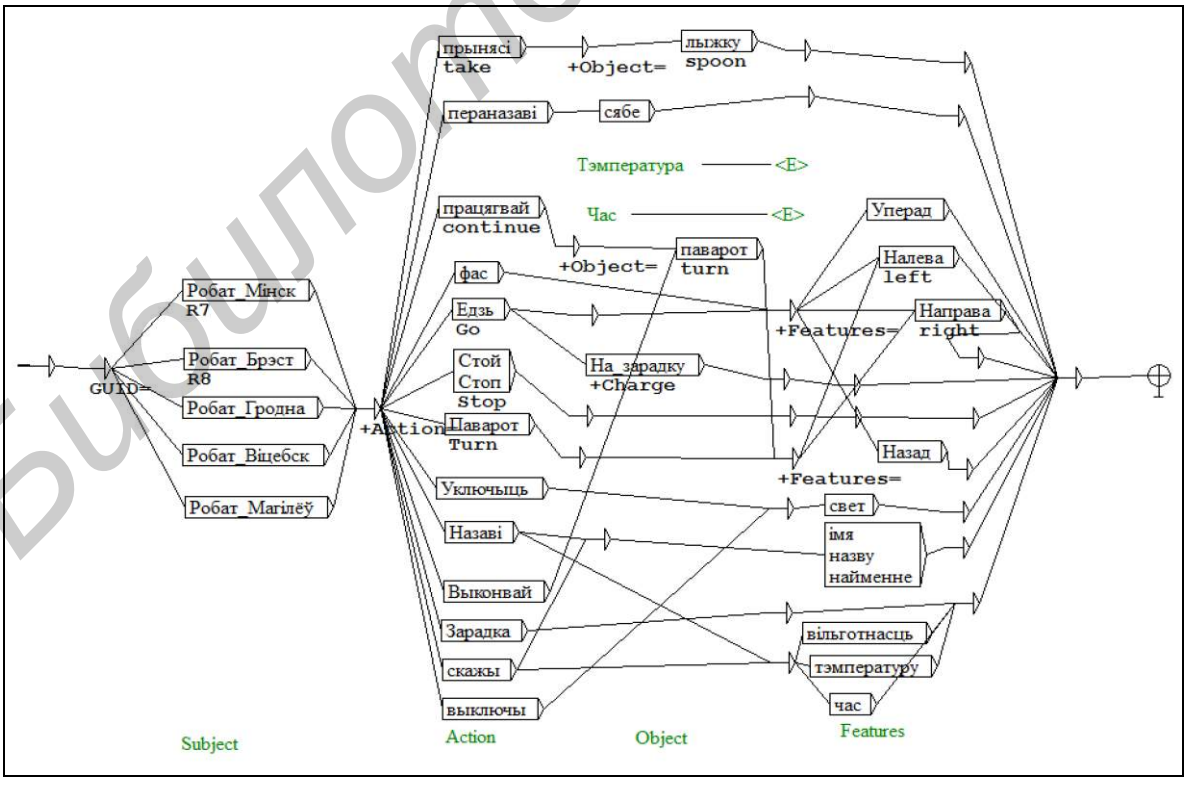
Робат_Гродна Едзь На_зарадку, GUID=R2+Action=Go+Charge

Робат_Брэст Выконвай паварот Направа, GUID=R3+Action=turn+Features=right

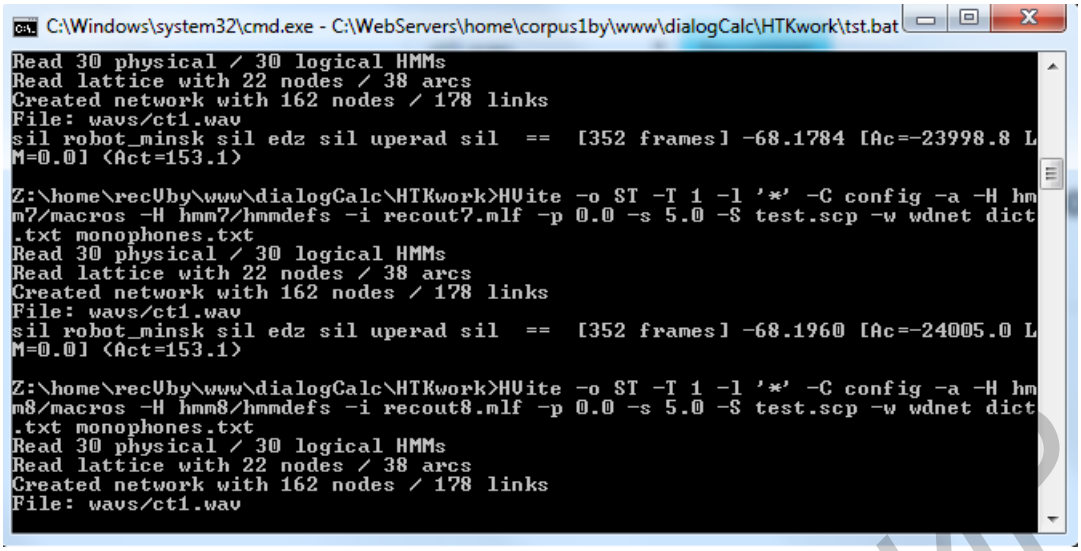
Робат_Брэст Выконвай паварот Налева, GUID=R3+Action=turn+Features=left

Тут канцэпт *‘Суб’ект’* прадстаўляе сабой імя робата, *‘Дзеянне’* — каманда, якую павінен выканаць робат, *‘Аб’ект’* — мэтанакіраванне для дзеяння, *‘Характарыстыка’* — удакладненне для канцэптаў *‘Аб’ект’* ці *‘Дзеянне’*.

Выкарыстоўваючы гэтыя канцэпты і інструментарыі *NooJ*, у выніку мы атрымалі моўную мадэль для сістэмы ўзаемадзеяння чалавека і робата. Схема мадэлі бачна на малюнку 2.



Малюнак 2 – Схема моўнай мадэлі з выкарыстаннем інструментарыя NooJ



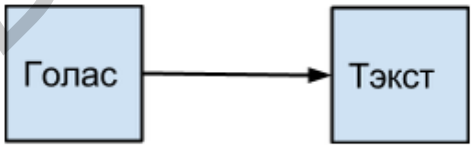
Малюнак 3 – Прыклад працы сістэмы распазнавання маўлення на базе пакета HTK

2. HTK як прылада для распазнавання

На базе HTK было створана ядро сістэмы распазнавання галасавых камандаў для кіравання мабільнымі робатамі на беларускай мове. У сістэме было 17 простых каманд кіравання і 5 эталонных галасоў розных дыктараў, у тым ліку і жаночых. Прыкладамі каманд могуць з'яўляцца словы: уперад, назад, налева, паварот і г.д. Назвы робатаў былі выбраныя адпаведна назвам абласным цэнтрам Рэспублікі Беларусь: Мінск, Гродна, Магілёў, Брэст, Віцебск, Гомель. Прыкладамі каманд-дзеянняў могуць з'яўляцца словы: едзь, стой, падымі, апусці. Пасля навучання і тэставання HMM мадэляў сістэмы HTK дакладнасць распазнавання склала каля 80% для мужчынскіх галасоў і каля 50-75% для жаночых. Прыклад працы сістэмы распазнавання галасавых камандаў на беларускай мове на базе HTK прыведзена на малюнку 3.

3. Прыклад распазнавання маўленчых каманд для мабільнага робата

Ідэя распазнавання маўленчых каманд заключаецца ў тым, каб атрымаць тэкставую інфармацыю з голасу карыстальніка сістэмы і вызначыць у ім каманду для кіравання мабільным робатам (малюнак 4).



Малюнак 4 – Агульны прынцып працы распазнавальніка маўлення

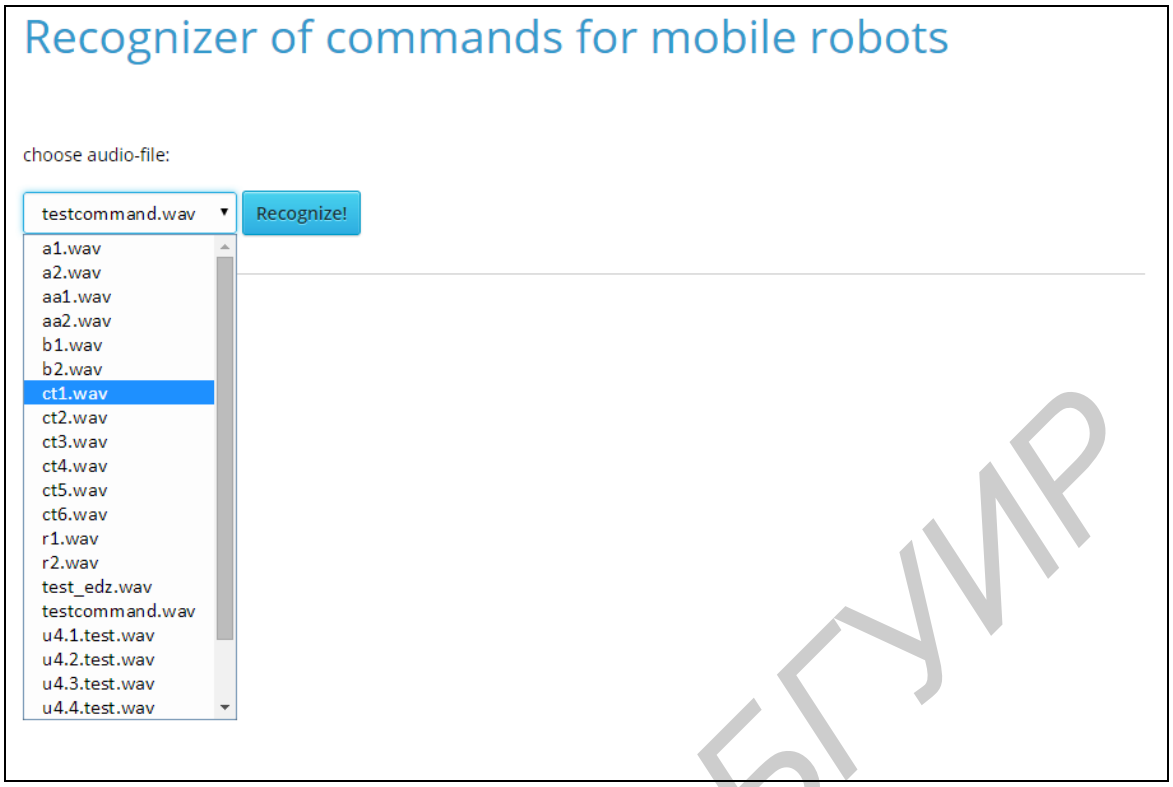
У працэсе працы натрэніраваны набор бібліятэк HTK быў інтэграваны ў мову праграмавання PHP. PHP – скрыптовая мова праграмавання агульнага прызначэння, якая інтэнсіўна ўжываецца для распрацоўкі вэб-прылад. У выніку быў распрацаваны распазнавальнік маўлення, які адаптуе вывад працы бібліятэк HTK да адмысловага фармату, з якім працягвае працаваць PHP – выдае карыстальніку.

Разгледзім прыклад працы створанага прататыпа тэставай сістэмы, якая распазнае каманды для мабільнага робата.

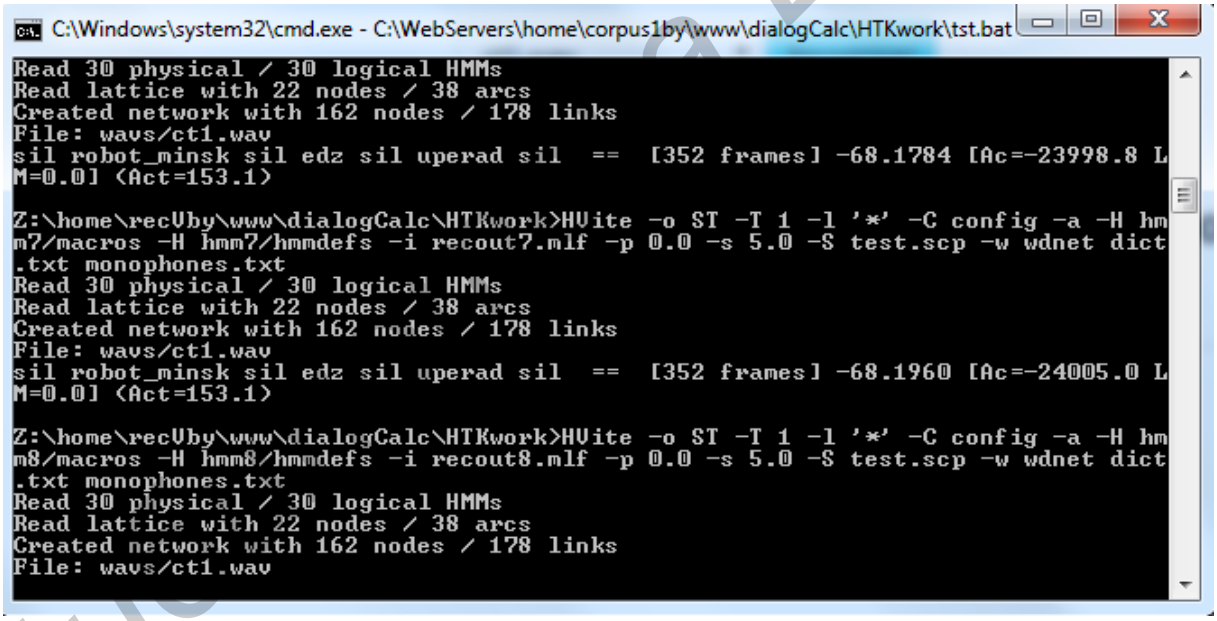
Крок 1. Выбар аўдыёфайла для распазнавання (малюнак 5). Праз прыладу для запісу голасу запісваецца аўдыёфайл з маўленчай камандай для сістэмы, якая разглядаецца. Потым гэты файл змяшчаецца ў папку з уваходнымі дадзенымі для сістэмы. У выніку, файл робіцца даступным для выбару ў спісе файлаў-прыкладаў каманд. Карыстальнік выбірае файл-прыклад і націскае на кнопку 'Recognize!'.

Крок 2. Распазнаванне маўлення выбранага аўдыёфайла (малюнак 6). Пры націску па кнопцы 'Recognize!' назва выбранага файла перадаецца ў праграмны модуль, які рэалізаваны на PHP, які далей запускае пакет HTK. Гэта рэалізавана праз спецыяльны файл-сцэнар працы HTK, які выклікаецца з PHP камандай *exec*. З дапамогай атрыманай раней натрэніраванай сістэмы распазнавання маўлення на базе HTK у пункце 2 з аўдыяфайла атрымліваюцца распазнаныя каманды.

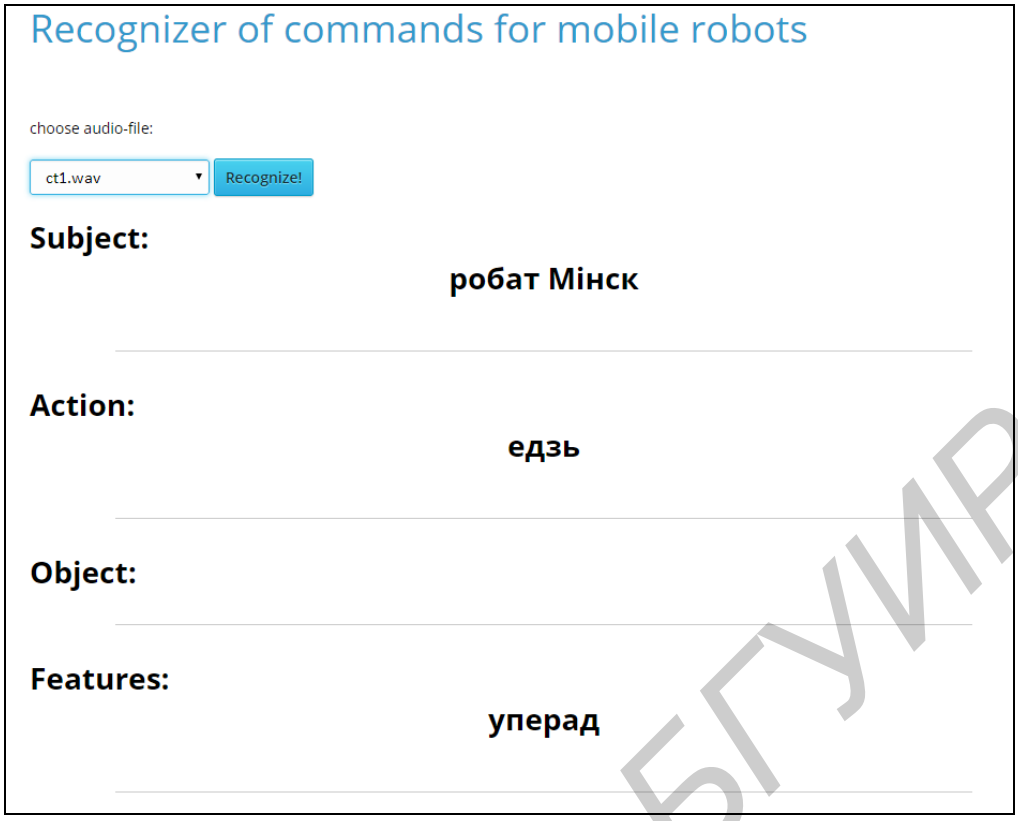
Крок 3. Вывад распазнаных каманд для мабільнага робата (малюнак 7). Сістэма выдае карыстальніку вынікі распазнанай маўленчай каманды ў выглядзе тэксту ў тэрмінах суб'ект, дзеянне, аб'ект, характарыстыка.



Малюнак 5 – Спіс гукавых файлаў-прыкладаў для падачы на ўваход сістэмы распазнавання маўлення на базе PHP і HTK



Малюнак 6 – Дэталізаваная справаздача працы функцый распазнавання HTK



Малюнак 7 – Вынік распазнавання каманды ў тэрмінах суб'ект, дзеянне, аб'ект, характарыстыка для мабільнага робата на галасавы запыт карыстальніка

4. Сэрвіс запісу голасу

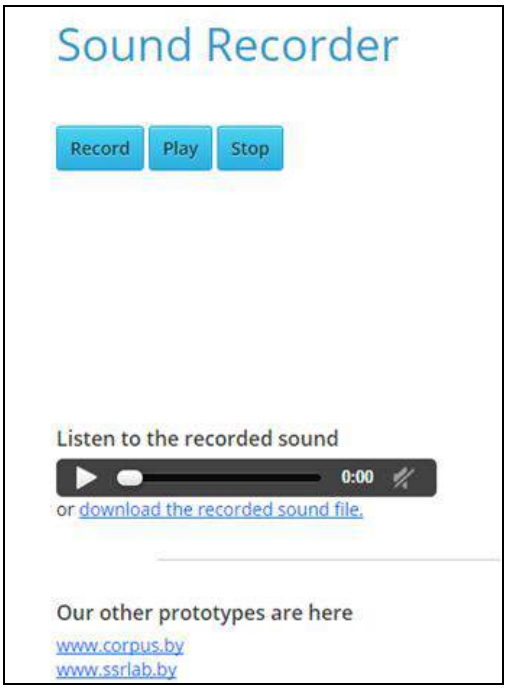
Прылада запісу голасу была распрацавана як Інтэрнэт-сэрвіс запісу голасу. Гэтая прылада не патрабуе ўстаноўкі ніякіх іншых праграм акрамя Adobe Flash Player і дазваляе наўпрост запісаць голас на сервер праз мікрафон. Дадзеная прылада была рэалізавана з дапамогай Флэш-прылады Wami, якая была інтэгравана ў PHP з дапамогай JavaScript. Знешні інтэрфейс прылады на малюнку 8.

Прылада складаецца з двух частак. Першая частка прадстаўлена трыма кнопкамі, праз якія карыстальнік мае магчымасць пачаць запіс голасу, прайграць запісаны гук і спыніць запіс ці праслухванне. Другая частка інтэрфейсу прадстаўлена фрэймам, які дазваляе прайграць запісаны аўдыёфайл, а таксама скачаць яго.

Заўважым, што лінк на назву файла будзе генеравацца ўнікальна. Для фарміравання ўнікальнасці імя выкарыстоўваецца імя, дата, ай-пі адрас карыстальніка і рандомны лік. Такім чынам, любы карыстальнік зможа выкарыстоўваць такі аўдыёфайл для сваіх іншых мэт па спецыяльнай унікальнай спасылцы.

У будучым плануецца выкарыстоўваць гэты сэрвіс для начыткі галасавых каманд рознымі дыктарамі. Гэтыя дыктары будуць заходзіць на старонку сэрвіса праз свае хатнія камп'ютары і начытваць пэўную паслядоўнасць каманд. Праз гэта будут будавацца ў паўаўтаматычным рэжыме шматлікія акустычныя мадэлі, што будзе набліжаць нашу сістэму распазнавання маўлення да дыктаранезалежнай.

Зараз гэты сэрвіс даступны ў рэжыме on-line на сайце www.corpus.by [Sound Recorder, 2015].



Малюнак 8 – Знешні інтэрфейс сістэмы запісу гуку

Заключэнне

Такім чынам, быў створаны першы прататып сістэмы распазнавання маўлення галасавых камандаў для кіравання мабільнымі робатамі на беларускай мове. Дакладнасць распазнавання дастатковая для выкарыстання сістэмы ў рэальных мабільных прыстасаваннях. Заўважым, што колькасць каманд для распазнавання можна змяняць у залежнасці ад мэт той ці іншай сістэмы.

У перспектыве два сэрвісы – сістэма па распазнаванню і сістэма запісу голасу з мікрафона будуць аб'яднаны ў адзіную сістэму, якая павінна стаць інтэрфейсам для галасавога ўводу дадзеных. Такую сістэму можна будзе выкарыстаць як модуль галасавога ўводу ў розных сістэмах, у тым ліку, для сістэмы OSTIS [Гецевич, 2011] .

Далей плануецца дадаць у распрацаваны прататып моўную мадэль, якую можна распрацаваць, напрыклад, з дапамогай прылады NooJ, для вырашэння наступных задач:

- вызначэнне набору каманд для кіравання мабільнымі робатамі;
- распрацоўка правіл адсячэння неадпаведных каманд для выканання мабільнымі робатамі.

Спіс літаратуры

[www.Corpus.by, 2012] Text-to-Speech PHP-Based Synthesizer [Electronic resource]. – 2012. – Mode of access : http://corpus.by/. – Date of access : 12.12.2015.

[Sound Recorder, 2015] Запіс гука // [Электронны рэсурс]. — 2014. Рэжым доступу : http://corpus.by/soundRecorder/. — Дата доступу : 10.01.2015.

[HTK, 2014] HTK [Электронны рэсурс] – Рэжым доступу: http://htk.eng.cam.ac.uk – Дата доступу: 22.06.2014.

[Sphinx, 2014] CMU Sphinx [Электронны рэсурс] – Рэжым доступу: http://cmusphinx.sourceforge.net/ – Дата доступу: 26.10.2014.

[KALDI, 2014] KALDI [Электронны рэсурс] – Рэжым доступу: http://kaldi.sourceforge.net/ – Дата доступу: 26.10.2014

[Julius, 2014] Julius [Электронны рэсурс] – Рэжым доступу: http://julius.sourceforge.jp/en_index.php – Дата доступу: 26.10.2014

[PHP, 2014] PHP [Электронны рэсурс] – Рэжым доступу: http://php.net/ – Дата доступу: 26.10.2014

[Гецевич, 2011] Естестевенно-языковые интерфейсы интеллектуальных вопросно-ответных систем / В.М. Вяльцев, Ю.С. Гецевич, В.А. Житко, А.А. Кузьмин // Доклады БГУИР. – 2011. – № 8 (62). – С. 80–86.

COMPONENT DESIGN FOR SPEECH RECOGNITION OF NATURAL LANGUAGE INTERFACE

Nikolaenko K.A.[*], Kaigorodova L.I. [**], Hetsevich Y.S. [**]

**Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

anak247@gmail.com

***United Institute of Informatics Problems, National Academy of Sciences, Minsk, Republic of Belarus*

lesia.piatrouskaya@gmail.com

yury.hetsevich@gmail.com

INTRODUCTION

This work is the start for further design of recognition system for robots-human interaction. The goal of the project is to interact with some number of robots in order to make them perform commands.

MAIN PART

In this article we present the design of some building blocks of the recognition system.

Using deep syntactic analysis and NooJ tools we design the language that would be common and close to every-day language of the humans and that it would be able for machines to 'understand' it.

HTK toolkit was integrated with PHP programming language for our recognition system. In the result syntactic analyzer has been created to gain text information out of voice data.

CONCLUSION

In the result the prototype of recognition system of robots-human interaction has been designed. This prototype can be the basis for two services to be combined: recognition system and dictation system with the use of microphones.