

**НАЦЫЯНАЛЬНА-КУЛЬТУРНЫ КАМΠΑЕНЕНТ
У ЛІТАРАТУРНАЙ І ДЫЯЛЕКТНАЙ МОВЕ**

Установа адукацыі
“Брэсцкі дзяржаўны ўніверсітэт імя А.С. Пушкіна”

**НАЦЫЯНАЛЬНА-КУЛЬТУРНЫ КАМПАНЕНТ
У ЛІТАРАТУРНАЙ І ДЫЯЛЕКТНАЙ МОВЕ**

Зборнік
навуковых артыкулаў

Брэст
БрДУ імя А.С. Пушкіна
2016

УДК 811.161.3(082)
ББК 81(4Бел)
Н 35

Рэцэнзенты:

кандыдат філалагічных навук, дацэнт,
дацэнт кафедры рускай і беларускай моў
УА “Брэсцкі дзяржаўны тэхнічны ўніверсітэт” **З. М. Заіка**,
кандыдат філалагічных навук, дацэнт,
дацэнт кафедры беларускага мовазнаўства і дыялекталогіі **І. У. Сацута**

Рэдакцыйная калегія:

кандыдат філалагічных навук, дацэнт **С. Ф. Бут-Гусаім**
кандыдат філалагічных навук, дацэнт **М. М. Аляхновіч**
кандыдат філалагічных навук, дацэнт **Л. В. Леванцэвіч**
кандыдат філалагічных навук, дацэнт **Н. Р. Якубук**

Н 35 **Нацыянальна-культурны** кампанент у літаратурнай і дыялектнай мове :
зб. навук. арт. / Брэсц. дзярж. ун-т імя А. С. Пушкіна ; рэдкал.: С. Ф. Бут-Гусаім
[і інш.]. – Брэст : БрДУ, 2016. – 231 с.
ISBN 978-985-555-480-7.

Зборнік навуковых артыкулаў складзены па матэрыялах рэгіянальнай навукова-
практычнай канферэнцыі “Нацыянальна-культурны кампанент кампанент у
літаратурнай і дыялектнай мове”, якая адбылася ў Брэсцкім дзяржаўным універсітэце
імя А. С. Пушкіна 18 сакавіка 2016 года. Змяшчае артыкулы па актуальных праблемах
мовазнаўства і лінгвадыдактыкі.

Адрасуецца навукоўцам, настаўнікам, студэнтам, усім, хто цікавіцца
нацыянальнай і рэгіянальнай культурай, мовай і літаратурай.

УДК 811.161.3(082)
ББК 81(4Бел)

ISBN 978-985-555-480-7

© УА «Брэсцкі дзяржаўны
ўніверсітэт імя А.С. Пушкіна», 2016

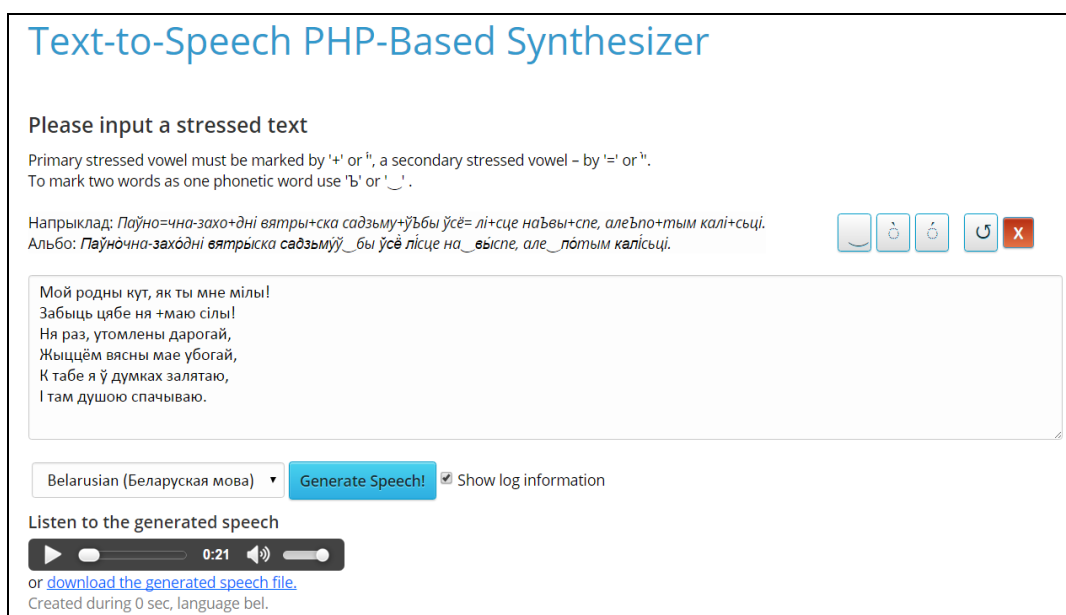
Касцючык В.М. Нацыянальна-культурная адметнасць анамастыкону прозы Уладзіміра Гніламёдава.....	88
Качан Я.С., Лысы С.І., Гецэвіч Ю.С., Станіславенка Г.Р., Гюнтар А.В. Камп'ютарна-лінгвістычныя сэрвісы www.corpus.by для аўтаматычнай апрацоўкі тэкстаў.....	93
Кісель Т.А. Назвы хваёвых дрэў у гаворках Брэстчыны.....	104
Клундук С.С. Словаскладанне: ад тэорыі да практыкі.....	108
Коцевіч С.С. Суггестивные особенности текстов колыбельных песен.....	113
Кузь Г.Т. Способи опису лінгвокультурологічнай інфармацыі у фразеологічных навчальных слоўніках.....	118
Купрыкова К.О. Фразеологические средства выразительности в художественно-публицистическом тексте областных газет Брестчины.....	123
Леваніцэвіч Л.В. Безэквівалентная лексіка ў гаворках Брэстчыны.....	126
Леванюк А.Я. Афарыстыка Я. Коласа як нацыянальна-культурны кампанент мовы.....	130
Лунакоў У.Э. Хімічныя тэрміны ва ўсходнеславянскіх мовах.....	133
Малахова О.О. Маркеры эстетичного коду білоруськоў пісенноў лірыкі (на матэрыялі пісень із репертуару І. Дорэфеевой).....	138
Пукала В. Нацыянальна-культурны кампанент у беларускіх парэміях з вобразамі жывёл.....	142
Пунько Т.М., Дзямчук Г.С. Гаваркія онімы ў творах Міколы Сянкевіча і Анатоля Крэйдзіча.....	148
Садоўская Т.Р., Струг В.І. Уласныя імёны ў кантэксте дзіцячай прозы Клаўдзіі Каліны.....	152
Самуйлік Я.Р. Гісторыя і сучаснасць, адметнасці гаворкі вёскі Дзятлавічы Лунінецкага раёна.....	155
Сацута І.У. Моўныя асаблівасці помнікаў старабеларускай перакладной літаратуры.....	160
Синкевіч Т.И. Репрезентация древних промыслов в лимнонимии Витебщины.....	164
Скрэбнева Т.В. Фаміліі витеблян суффіксальнага афаравання: статыстыка, мотывіруючыя асновы.....	171
Слесарева Т.П. Геортонімы в русской фразеологии: семантычэскі аспект.....	176
Солахаў А.В. Спосабы ўтварэння індывідуальна-аўтарскіх неалагізмаў у мове твораў Л. Дранько-Майсюка.....	181
Солахаў А.В., Бразінская Л.С. Складаныя сказы з чатырма прэдыкатыўнымі часткамі і рознымі відамі сувязі ў мове прозы У. Караткевіча.....	187

**Я. С. Качан, С. І. Лысы, Ю. С. Гецэвіч, Г. Р. Станіславенка,
А. В. Гюнтар** (Аб'яднаны інстытут праблем інфарматыкі НАН
Беларусі, Мінск)

КАМП'ЮТАРНА-ЛІНГВІСТЫЧНЫЯ СЭРВІСЫ WWW.CORPUS.BY ДЛЯ АЎТАМАТЫЧНАЙ АПРАЦОЎКІ ТЭКСТАЎ

Сёння камп'ютарныя тэхналогіі выкарыстоўваюцца амаль ва ўсіх сферах дзейнасці чалавека. Адным з важных накірункаў іх выкарыстання з'яўляецца распрацоўка сістэм, здольных аўтаматычна апрацоўваць электронныя тэксты. Супрацоўнікамі лабараторыі распазнавання і сінтэзу маўлення Аб'яднанага інстытута праблем інфарматыкі Нацыянальнай акадэміі навук Беларусі быў распрацаваны інтэрнэт-рэсурс www.Corpus.by, які дапамагае вырашаць мноства задач, звязаных з апрацоўкай электронных тэкстаў і маўленчых сігналаў. Некаторыя з распрацаваных сэрвісаў апісаны ніжэй у артыкуле.

Адной з асноўных распрацовак лабараторыі з'яўляецца сістэма сінтэзу маўлення па тэксце (ССМТ) – сістэма, здольная пераўтвараць адвольны электронны тэкст у адпаведны маўленчы сігнал. Сярод распрацаваных версій ССМТ варта асабліва адзначыць інтэрнэт-версію, бо яна знаходзіцца ў вольным доступе ў інтэрнэце [1]. Знешні інтэрфейс інтэрнэт-сінтэзатара маўлення паказаны на малюнку 1.



Малюнак 1 – Знешні інтэрфейс інтэрнэт-версіі сінтэзатара маўлення па тэксце

У ходзе пераўтварэння электроннага тэксту ў маўленчы сігнал сінтэзатар генеруе мноства прамежкавых вынікаў. Сярод іх нармалізаваны тэкст, фанемны запіс тэксту, запіс тэксту ў алафонным выглядзе і інш. Гэтыя вынікі ў сваю чаргу могуць быць выкарыстаны для вырашэння шэрагу іншых камп'ютарна-лінгвістычных задач. Адной з такіх задач з'яўляецца атрыманне транскрыпцыі слоў і цэлых тэкстаў.

На сённяшні дзень ужо распрацаваны алгарытм генерацыі транскрыпцый розных фарматаў па электронным тэксце на беларускай мове [2]. Гэты алгарытм быў рэалізаваны ў выглядзе інтэрнэт-сэрвісу “Генератар транскрыпцый” [3]. У выніку працы дадзенага інтэрнэт-сэрвісу карыстальнік мае магчымасць атрымаць са звычайнага тэксту тэкст, сканвертаваны ў транскрыпцыі розных фарматаў. На малюнку 2 паказаны знешні інтэрфейс сэрвісу.

Малюнак 2 – Знешні інтэрфейс інтэрнэт-сэрвісу “Генератар транскрыпцый”

Каб атрымаць неабходныя транскрыпцыі, карыстальнік мае магчымасць увесці тэкст у вызначанае для гэтага поле і націснуць кнопку “Атрымаць транскрыпцыі”.

Уваходны тэкст можа быць уведзены як у звычайным выглядзе (напрыклад, “звычайны тэкст”), так і ў алафонным выглядзе (напрыклад, “Z001, V002, Y212, CH001, A023, J’013, N004, Y222, /, T001, E022, K004, S002, T000, /, #P4”). У тэксце карыстальнік можа выкарыстоўваць наступныя пазнакі:

- плюс /+/ пасля націскай галоснай, каб пазначыць асноўны націск (напрыклад, “звыча+йны”);

- роўна /=/ пасля націскай галоснай, каб пазначыць пабочны націск (напрыклад, “тэ=леперада+ча”);

- цыркумфлекс /^/ паміж двума словамі, каб аб’яднаць іх у складанае фанетычнае слова (напрыклад, “на^стале+”, “сказа+ў^бы”).

Апроч таго, могуць выкарыстоўвацца і стандартныя знакі націскаў (напрыклад, “тэлеперада́ча”).

Сэрвіс дае магчымасць карыстальніку канвертаваць тэкст у транскрыпцыі 4 фарматаў: кірылічную, міжнародную (IPA), спрошчаную міжнародную і X-SAMPA. Больш падрабязна пра кожны з фарматаў можна даведацца, скарыстаўшыся прыведзенымі на старонцы сэрвісу спасылкамі на электронныя крыніцы, паводле якіх вялася распрацоўка.

Яшчэ адной функцыянальнай асаблівасцю сэрвісу з’яўляецца магчымасць наладкі выніковых даных. Адзначыўшы адпаведныя пункты ў наладках, карыстальнік можа атрымаць транскрыпцыі кожнага слова ў квадратных дужках альбо без іх; у транскрыпцыі кожнага слова фанемы могуць быць выведзены разам альбо раздзелены коскамі.

Калі адзначыць пункт “Паказаць інтанацыйныя пазнакі”, выніковы тэкст будзе разбіты на інтанацыйныя фрагменты – сінтагмы. Для гэтага будуць выкарыстаны спецыяльныя пазнакі (напрыклад, вертыкальная рыса ‘|’ ў выпадку коскі і падвоеная вертыкальная рыса ‘||’ ў канцы сказа ці абзаца). Пасля ўводу тэксту і адпаведных наладак патрэбна націснуць на кнопку “Атрымаць транскрыпцыі!”.

Прыклад вынікаў працы сэрвісу паказаны на малюнку 3.

Transcription in cyrillic letters

[γρύσα] [ц'в'іна] [απόσн'і] [γότ] ||
 [ус'э] [γал'іны] [іайэ] |
 [ус'э] [в'ал'ік'ііа] [рас'ох'і] |
 [даап'осн'ага] [пр'уц'іка] |
 [был'і] [ўс'іпаны] [б'урным] [б'эларуж'овым] [ц'в'этам] ||
 [іанэ] [к'іп'эла] |
 [мл'эла] [і] [раскашав'алас'а] [ўпчал'іным] [зв'он'э] |

Transcription in IPA

[ɣr'usʰa] [ts'v'il'a] [ap'osn'i] [ɣ'ot] ||
 [us'ɛ] [ɣal'iɲi] [jaɪ.ɛ] |
 [us'ɛ] [v'al'ik'ija] [ras'ox'i] |
 [daap'osn'aɣa] [pr'uts'ika] |
 [bɪl'i] [ws'ipani] [b'urnim] [b'ɛlaruz'ovim] [ts'v'etam] ||
 [jan.ə] [k'ip'ela] |
 [ml'ɛla] [i] [raskasav'alas'a] [wptʃal'inim] [zv'on'ɛ] |

Transcription in simplified IPA

[ɣrúʂa] [ts'v'ilá] [apósн'i] [yót] ||
 [us'è] [ɣal'iny] [jaɪɛ] |
 [us'è] [v'al'ik'ija] [ras'ox'i] |
 [daapósн'aɣa] [prúts'ika] |
 [bɪl'i] [us'ipany] [b'urnym] [b'ɛlaruz'ovym] [ts'v'etam] ||
 [janə] [k'ip'ela] |
 [ml'ɛla] [i] [raskasaválas'a] [wptʃal'inym] [zvón'e] |

Transcription in X-SAMPA

[Gr'us'a] [ts'v'il'a] [ap'Os'n'i] [G'Ot] ||
 [us'%E] [Gal''in1] [jaɪ%E] |
 [us'%E] [v'al''ik'ija] [ras'Ox'i] |
 [daap'O's'n'aGa] [pr'uts'ika] |
 [b1'l'i] [ws'1pan1] [b'urn1m] [b'%Elaruz''Ov1m] [ts'v''Etam] ||
 [jan%a] [k'ip''Ela] |
 [ml''Ela] [%i] [raskas'av'alas'a] [wptSal''in1m] [zv''On'E] |

Малюнак 3 – Прыклад апрацоўкі тэксту на беларускай мове сэрвісам “Генератар транскрыпцый”

Зыходзячы з таго, што фанетычны працэсар інтэрнэт-сінтэзатара створаны для апрацоўкі як беларускамоўных тэкстаў, так і тэкстаў на рускай мове, было прынята рашэнне распрацаваць базу адпаведнасцяў “фанема – знак транскрыпцыі” і для рускай мовы. Распрацоўка спіса адпаведнасцяў для рускай мовы ішла па прыкладзе ўжо створанай базы для беларускай мовы. Былі абраны 72 словы, якія пакрылі ўсе фанемы рускай мовы і некаторыя іх варыяцыі (падоўжаныя зычныя, націскныя галосныя): агулам 94. У беларускай мове такіх варыяцый налічваецца 99. Падобраныя словы былі апрацаваны з дапамогай сінтэзатара маўлення па тэксце, у выніку чаго быў атрыманы іх алафонны запіс (табліца 1).

Табліца 1 – Фрагмент табліцы слоў, што пакрываюць усе фанемы рускай мовы, і іх алафоннага запісу ў фармаце, які выкарыстоўваецца ў сінтэзатары маўлення

Слова	Алафонны запіс
...	...
два+	D001,V001,A010,/,#P4,
желтова=то-бе+лый	ZH004,E321,L003,T002,A221,V012,A112,T002, A223,B'002,E041,L004,Y313,J'010,/,#P4,
оди+н	A203,D'002,I042,N000,/,#P4,
де+вять	D'002,E043,V'012,A343,T'000,/,#P4,
ше+сть	SH001,E023,S'002,T'000,/,#P4,
се=веро-за+пад	S'002,E143,V'012,E342,R002,A222,Z002,A021, P002,A312, T000,/,#P4,
миллио+н	M'004,I343,L'104,I241,O042,N000,/,#P4,
...	...

Для 94 фанем і іх варыяцый былі падабраны адпаведныя сімвалы транскрыпцый чатырох фармацатаў: кірылічнага, міжнароднага (IPA), спрошчанага міжнароднага і X-SAMPA, што адлюстравана на малюнку 4.

#СМТ	#Кірыл.	#Прыклады (кірыл.)	#Міжнародны	#Прыклады (Міжнар.)	#IPA	#IPA прыклады	#X-SAMPA	#X-SAMPA
A0	á	два́	á	dvá	'a	dv' a	"a	dv" a
A1	à	жэлтава́таб'э́лый'	à	zɛltavàtab'ɛ́lj'	ˌa	zɛltav_atab'ɛlj'	%a	z'Eltav%atab"Elj'
A2	a	ад'ін	a	ad'in	a	ad' in	a	ad""in
A3	a	д'эв'ат'	a	d'ev'at'	a	d'ɛvət'	a	d""Ev'at'
E0	é	шэ́с'т'	é	šés't'	'ɛ	š'ɛs'tj	"E	s""Es't'
E1	è	с'эв'эраза́пат	è	s'ev'erazápat	ɛ	s'ɛvɛraz'apat	%E	s'%Ev'Eraz"apat
E2	э	ч'эты́р'э	e	č'etýr'e	ɛ	čɛt'ir'ɛ	E	ts\Et"1r'E
E3	э	ч'эты́р'э	e	č'etýr'e	ɛ	čɛt'ir'ɛ	E	ts\Et"1r'E

Малюнак 4 – Фрагмент спісу адпаведнасцяў алафоннага і транскрыпцыйнага запісу фанем і іх варыяцый для рускай мовы

Параўнанне спісаў фанем і іх варыяцый для беларускай і рускай моў паказала, што беларуская мова мае на 6 алафонаў больш, чым руская, прычым у рускай мове ёсць 15 алафонаў, якіх няма ў беларускай (G1, G'1, D'0, D1, D'1, R'0, R1, R'1, T'0, T'1, CH'0, CH'1, SH'0, SH'1, J'1),

а ў беларускай – 19 алафонаў, якіх няма ў рускай (O2, O3, GH0, GH1, GH'1, DZ'0, D1, DZ'1, DZ0, DZH0, DZH1, R1, C'0, C'1, CH0, CH1, SCH0, SCH1, W0).

Сэрвіс “Генератар транскрыпцый” з’яўляецца карыснай прыладай, якая можа хутка пераўтварыць адвольны тэкст на беларускай ці рускай мове ў транскрыпцыі розных фарматаў. Яго можна выкарыстоўваць для навучання замежнікаў нормам беларускага і рускага вымаўлення, для знаёмства з асаблівасцямі беларускай і рускай фанетыкі, а таксама для навуковых мэт. Карыснай сферай выкарыстання распрацаванага сэрвісу з’яўляецца магчымасць хуткага аўтаматычнага стварэння спецыялізаваных фанетычных слоўнікаў беларускай і рускай моў.

Для апрацоўкі электронных тэкстаў інтэрнэт-рэсурс www.Corpus.by выкарыстоўвае шэраг слоўнікаў. Сярод іх ёсць два слоўнікі беларускай мовы (“Слоўнік беларускай мовы. Арфаграфія. Арфаэпія. Акцэнттуацыя. Словазмяненне” пад рэдакцыяй М.В. Бірылы; “Слоўнік беларускай мовы” пад рэдакцыяй А.А. Лукашанца, В.П. Русак), слоўнік рускай мовы (“Грамматический словарь русского языка: Словоизменение” А.А. Залізняка), слоўнік англійскай мовы (паводле “Carnegie Mellon University Pronouncing Dictionary” [4]), а таксама слоўнік, які папаўняецца словамі, што былі пададзены на апрацоўку ў сінтэзатар маўлення і не былі знойдзены ва ўбудаваных слоўніках.

Дзякуючы дадзеным слоўнікам, карыстальнік можа апрацоўваць тэксты не толькі на беларускай, але і на рускай і англійскай мовах. Адным з сэрвісаў, які выкарыстоўвае ў працы ўсе названыя вышэй слоўнікі, з’яўляецца “Агучаны электронны граматычны слоўнік” [5].

“Агучаны электронны граматычны слоўнік” дапамагае атрымоўваць граматычную і фанетычную інфармацыю аб словах паводле пералічаных вышэй слоўнікаў. Карыстальнік мае магчымасць выбіраць слоўнікі, паводле якіх будзе весціся пошук слоў, а таксама здзяйсняць наладкі (пошук па часціне мовы, пошук толькі пачатковых форм слоў, выбар максімальнай колькасці вынікаў) (малюнак 5). Пасля ўводу адвольнага слова ў поле “Слова для пошуку” і адзначэння неабходных наладак карыстальнік можа націснуць кнопку “Шукаць!” і атрымаць наяўную ў слоўніках інфармацыю пра зададзенае слова.

Пошук можа весціся непасрэдна па слове ці па пэўнай частцы слова (па шаблоне) з магчымай пазнакай асноўнага ці пабочнага націску. Спіс спецыяльных сімвалаў, якія могуць выкарыстоўвацца ў запыце, і прыклады іх выкарыстання пералічаны ў табліцы 2.

Табліца 2 – Спіс спецыяльных сімвалаў, якія могуць выкарыстоўваць ў полі “Слова для пошуку” сэрвісу “Агучаны электронны граматычны слоўнік”

Сімвал	Тлумачэнне ўжывання	Прыклад ужывання
. (кропка)	замест адной любой літары	к.т: <i>кат, кіт, кот, кут</i>
* (зорачка)	замест адвольнай колькасці любых літар	сло*ік: <i>слоік, слонік, слоўнік</i>
+ (плюс)	пасля націскай галоснай, каб знайсці словы з пэўным націскам	<i>музы+ка</i>
= (роўна)	пасля галоснай з пабочным націскам	<i>паўно=чна-захо+дні</i>

Апроч таго, сэрвіс прадстаўляе магчымасць праслухоўвання знойдзеных слоў. Іх агучка здзяйсняецца пры дапамозе інтэрнэт-сінтэзатара маўлення па тэксце. Словы англійскай мовы агучваюцца беларускімі алафонамі пасля выкарыстання алгарытму фанетызацыі англійскіх слоў на беларускую мову.

Слова для пошуку:

Часціна мовы: Толькі пачатковыя формы

Максімум вынікаў:

Шукаць у наступных слоўніках:

- SBM1987** (паводле публікацыі "Слоўнік беларускай мовы. Арфаграфія. Арфаэпія. Акцэнтацыя. Словазмяненне / пад рэд. М.В. Бірылы. – Мінск, 1987")
- SBM2012initial** (пачатковыя формы паводле публікацыі "Слоўнік беларускай мовы. / навук. рэд. А.А. Лукашанец, В.П. Русак. - Мінск : Беларус. навука, 2012")
- CMU** (паводле "[Carnegie Mellon University Pronouncing Dictionary](#)")
- ZALIZNIAK** (паводле публікацыі "Грамматический словарь русского языка: Словоизменение / А.А. Зализняк. — Москва : Русский язык, 1980. — 880 с.")
- NEW** (новыя словы для [CMT](#))

[Search! / Шукаць!](#)

Паводле слоўніка **sbm1987** (1)

↓ Слова	Лексема	Транскрыпцыя	IPA	Тэг	Катэгорыя	Агучка	
мама	мама	[мама]	[m'ama]	NNAFO	назоўнік	Voice!	0:00

Паводле слоўніка **sbm2012initial** (1)

↓ Слова	Транскрыпцыя	IPA	Катэгорыя	Агучка	
мама	[мама]	[m'ama]		Voice!	0:00

Паводле слоўніка **zalizniak** (1)

↓ Слова	Лексема	Транскрыпцыя	IPA	Тэг	Агучка	
мама	мама	[мама]	[m'ama]	NCAFSN	Voice!	0:00

Малюнак 5 – Пошук граматычнай і фанетычнай інфармацыі пра слова “мама” ў “Агучаным электронным граматычным слоўніку”

Акрамя граматычнай і фанетычнай інфармацыі, карыстальнік таксама можа атрымаць арфаграфічную інфармацыю, а менавіта, здзейсніць праверку правапісу слоў у адвольным тэксце. У гэтым яму дапаможа сэрвіс “Праверка правапісу” [6]. Каб пабачыць, як сэрвіс працуе, трэба ўвесці тэкст у поле для ўводу і націснуць кнопку “Атрымаць невядомыя словы!”. У акно з загаловак “Ігнараваць словы” можна ўвесці словы, якія не маюць патрэбы ў праверцы і будуць ігнаравацца сэрвісам. Пасля апрацоўкі тэксту карыстальнік атрымлівае статыстычныя даныя пра колькасць знойдзеных ва ўваходным тэксце слоў і колькасць слоў, не знойдзеных у выбраных слоўніках.

Ніжэй будзе прыведзены спіс слоў, якія былі знойдзены ў тэксце, але адсутнічаюць у абраных слоўніках (малюнак 6). Кожнае са слоў гэтага спіса карыстальнік можа пазначыць як “правільнае” / “вядомае” альбо пакінуць “няправільным” / “невядомым”, выбраўшы адпаведную пазнаку “так” ці “не”. Для зручнасці працы са спісам ёсць магчымасць у любы момант разбіць спіс на дзве паловы па прыкмеце “вядомасці” ці “невядомасці” слова карыстальніку.

Калі ласка, увядзіце тэкст
↻ ✕
Ігнараваць словы:
↻ ✕

Груша цвіла апошні год. Усе галіны яе, усе вялікія расохі, да апошняга пруціка, былі ўсыпаны буйным бела-ружовым цветам. Яна кіпела, мела і раскашавалася ў пчаліным звоне, цягнула да сонца сталья лапы і распускала ў яго зяніні маленькія, кволяы пальцы новых парасткаў. І была яна такая магутная і свежая, так утрапёна спрачаліся ў яе ружовым раі пчолы, што, здавалася, не будзе ёй зводу і не будзе ёй канца.

сонца
пчолы

Выкарыстаць наступныя слоўнікі:

- SBM1987** (паводле публікацыі "Слоўнік беларускай мовы. Арфаграфія. Арфаэпія. Акцэнтауацыя. Словазмяненне / пад рэд. М.В. Бірылы. - Мінск, 1987")
- SBM2012initial** (пачатковыя формы паводле публікацыі "Слоўнік беларускай мовы. / навук. рэд. А.А. Лукашанец, В.П. Русак. - Мінск : Беларус. навука, 2012")
- ZALIZNIAK** (паводле публікацыі "Грамматический словарь русского языка: Словоизменение / А.А. Зализняк. — Москва : Русский язык, 1980. — 880 с.")
- CMU** (паводле "[Carnegie Mellon University Pronouncing Dictionary](#)")
- NEW** (новыя словы для [CMT](#))

Паказаць вынікі працы тэкставага працэсара

Get unknown words! / Атрымаць невядомыя словы!

СТАТЫСТЫКА АПРАЦОЎКІ:

Знойдзена ўнікальных напісанняў слоў: **57**
 Пасля ўзгаднення са спісам ігнаравання слоў: **55**
 Невядомых сэрвісу слоў: **2**
 Невядомых сэрвісу і карыстальніку слоў: **2**

ЦІ ВЕДАЕЦЕ ВЫ ГЭТАЕ СЛОВА? (спампаваць спіс)

пруціка	<input type="radio"/> Так <input checked="" type="radio"/> Не	... расохі, да апошняга пруціка , былі ўсыпаны буйным ...
расохі	<input type="radio"/> Так <input checked="" type="radio"/> Не	... яе, усе вялікія расохі , да апошняга пруціка, ...

Recheck with your labels / Пераправерыць з улікам вядомых Вам слоў

Малюнак 6 – Прыклад працы інтэрнэт-сэрвісу “Праверка правапісу”

Адной з найбольш распаўсюджаных памылак у беларускай мове з’яўляецца памылка ў напісанні літар “у” і “ў”. Сэрвіс “**Праверка правапісу Ў**” [7] распрацаваны для хуткай праверкі вялікіх электронных тэкстаў на наяўнасць такіх памылак. На ўваход сэрвісу падаецца адвольны тэкст на беларускай мове. Калі карыстальнік націсне кнопку “Праверыць!”, сэрвіс выведзе спіс фрагментаў уваходнага тэксту, дзе верагодна ёсць памылка, а таксама невялікі каментар (малюнак 7).

У працэсе пошуку магчымых памылак сэрвіс не толькі вызначае, галосная ці зычная літара знаходзіцца перад “у” / “ў”, але таксама аналізуе знакі прыпынку, якія могуць уплываць на правапіс. На сённяшні дзень сэрвіс не ўлічвае выключэнні. Убудоваць у сэрвіс спіс выключэнняў плануецца.

Калі ласка, увядзіце тэкст
Refresh! / Аднавіць!
Clear! / Ачысціць!

Напрыклад,

Мама ў краме. (Сказ без памылкі).

Мама у краме.

Кот у ботах. (Сказ без памылкі).

Кот ў ботах.

А ч.іх ёсць пчолы.

Абрэвіятуры ў тэксце:

Праверыць!

Вынікі праверкі

Магчыма, патрэбна пісаць “Ў” ці “ў”:

Сустрэлася “а у”:	“... Мама у краме. ...”	<i>(“у” пасля галоснай “а” без знакаў прыпынку)</i>
Сустрэлася “А у”:	“... А у іх ёсць пчолы. ...”	<i>(“у” пасля галоснай “А” без знакаў прыпынку)</i>
Сустрэлася “а» у”:	“... «Рама» у краме. ...”	<i>(“у” пасля галоснай “а” без знакаў прыпынку)</i>
Сустрэлася “а-у”:	“... На ўкраіне паўднёва- у сходні вецер. ...”	<i>(“у” пасля галоснай “а” і злучка)</i>
Сустрэлася “ау”:	“... Сястра есць ау сянку. ...”	<i>(“у” пасля галоснай “а”)</i>

Магчыма, патрэбна пісаць “У” ці “у”:

Сустрэлася “т ў”:	“... Кот ў ботах. ...”	<i>(“ў” пасля зычнай “т” без знакаў прыпынку)</i>
Сустрэлася “т» ў”:	“... «Брат» ў краме. ...”	<i>(“ў” пасля зычнай “т” без знакаў прыпынку)</i>
Сустрэлася “ў”:	“... На ўкраіне паўднёва- у сходні вецер. ...”	<i>(ВЯЛІКАЯ “Ў” СУСТРАКАЕЦЦА ТОЛЬКІ Ў ТЭКСТАХ, ДЗЕ ЎСЕ СЛОВЫ ПІШУЦЦА ВЯЛІКІМІ ЛІТАРАМІ)</i>
Сустрэлася “м-ў”:	“... Усім- ў сім пра ўсё распавядзем! ...”	<i>(“ў” пасля зычнай “м” і злучка)</i>
Сустрэлася “бў”:	“... Тата любіць буль ў”	<i>(“ў” пасля зычнай “б”)</i>

Малюнак 7 – Прыклад працы інтэрнэт-сэрвісу “Праверка правапісу “ў”

Для пошуку амографаў можа быць выкарыстаны сэрвіс пад назвай “Ідэнтыфікатар амографаў” [8]. “Ідэнтыфікатар амографаў” разглядае кожнае слова ўваходнага тэксту на наяўнасць некалькіх спосабаў іх прачытання паводле інфармацыі прадстаўленай у слоўніках. Для беларускай і рускай мовы адбываецца пошук слоў з аднолькавым напісаннем і рознымі націскамі, для англійскай – як слоў з рознымі пазіцыямі націску, так і слоў, якія маюць варыятыўнае вымаўленне. Пасля апрацоўкі тэксту карыстальнік можа пабачыць табліцу, у якой адлюстраваны знойдзеныя словы, варыянты націску, тып амографа, частата сустракання, прыклад кантэксту і слоўнік, паводле якога вызначаны амограф.

Амограф	Варыянты націску	↓ Тып амографа	Кольк.	Кантэксты	Слоўнік
нары	нары на́ры	адна часціна мовы	1	... паклічуць мяне з нары як музыка. I ...	SBM1987
спазнаю	спазнао спазна́ю	адна часціна мовы	1	... і непакоіцца. Я спазнаю кошт шчасця! А ...	SBM1987
людскія	лю́дскія людскія	адна часціна мовы	1	... Калі я чую людскія крокі, я ўцякаю ...	SBM1987
музыка	музы́ка му́зыка	адна часціна мовы	1	... з нары як музыка. I потым – паглядзі! ...	SBM1987
шчаслівейшы	шчаслівейшы шчасліве́йшы	адна часціна мовы	1	... часу, тым я шчаслівейшы. А чацвёртай гадзіне ...	SBM1987
лісу	лісу лісу́	адна часціна мовы	1	... З радасцю, – адказаў Лісу Маленькі прынц, – але ...	SBM1987
куры	ку́ры куры́	адна часціна мовы	1	... на мяне. Усе куры падобны адна на ...	SBM1987
хапае	хапае хапа́е	адна часціна мовы	1	... У людзей не хапае часу ўведаць што-небудзь. ...	SBM1987
прыкладу	прыкла́ду прыкладу́	адна часціна мовы	1	... ты прыходзіш, да прыкладу, а чацвёртай папалудні, ...	SBM1987
гадзіне	гадзі́не гадзі́не	адна часціна мовы	2	... ўжо а трэцяй гадзіне я адчуваю сябе шчаслівейшы. А чацвёртай гадзіне я ўжо стану ...	SBM1987
адказаў	адка́заў адказаў́	розныя часціны мовы	2	... ён. – З радасцю, – адказаў Лісу Маленькі прынц, – быць вельмі цяглівым, – адказаў Ліс. – Спачатку ты ...	SBM1987
сказаў	ска́заў сказаў́	розныя часціны мовы	1	... як прыручаеш іх, – сказаў Ліс. – У людзей ...	SBM1987

Малюнак 8 – Прыклад працы інтэрнэт-сэрвісу “Ідэнтыфікатар амографаў”

Стварэнне бібліяграфічных спасылак – неад’емная частка навуковай працы. Каб спрасіць гэту задачу для карыстальнікаў, быў створаны сэрвіс “Генератар спасылак на публікацыі” [9]. Дадзены сэрвіс распрацаваны для аўтаматычнай генерацыі спасылак паводле пэўных стандартаў. Зайшоўшы на старонку сэрвісу, карыстальніку неабходна пазначыць

фармат спасылкі, тып крыніцы (кніга, артыкул часопісу, канферэнцыя, электронны рэсурс) і мову афармлення публікацыі (беларуская, руская ці англійская) (малюнак 9). Праца з генератарам бібліяграфічных спасылак дазваляе даведацца больш дакладна пра іх змест і структуру, што з’яўляецца добрай асновай для распрацоўкі сістэмы распазнавання спасылак у тэкстах беларускай і рускай моў. Інфармацыя, якую можна атрымаць з бібліяграфічных спасылак пры распазнаванні, можа быць карыснай для бібліятэк, выдавецтваў ці навуковых устаноў, бо яна змяшчае звесткі пра аўтараў і рэдактараў, назвы публікацый, назву і месца выдавецтва, год выдання і г.д.

Калі ласка, увядзіце дадзеныя публікацыі

Фармат спасылкі

Тып крыніцы

Мова афармлення публікацыі

Звярніце ўвагу!

— [Інструкцыя па афармленню дысертацыі, аўтарэферата і публікацый](#)

— [ГОСТ 7.1-2003. Бібліяграфічны запіс. Бібліяграфічнае апісанне](#)

Erase All Data Entered in the Fields Below and Show Our Examples / Ачысціць усё і паказаць прыклады

Clear All Fields Below! / Ачысціць усё палі!

Аўтар адсутнічае

* Аўтары

Ініцыялы / Firstname	Прозвішча / Lastname
<input style="width: 95%;" type="text" value="Ю.С."/>	<input style="width: 95%;" type="text" value="Гецэвіч"/>
<input style="width: 95%;" type="text"/>	<input style="width: 95%;" type="text"/>

Малюнак 9 – Знешні інтэрфейс сэрвісу “Генератар спасылак на публікацыі”

Акрамя акрэсленых вышэй магчымасцяў, трэба адзначыць яшчэ адзін прыярытэтны напрамак выкарыстання сістэм сінтэзу маўлення па тэксце – дыялог паміж машынай і чалавекам. Дзякуючы гэтаму, сёння лінгвісты таксама запатрабаваны ў тэхнічнай сферы дзейнасці. Так, напрыклад, на выставе “ТІБО-2013” сектар робататэхнікі лабараторыі мадэлявання самаарганізавальных сістэм сумесна з лабараторыяй распазнавання і сінтэзу маўлення АПП НАН Беларусі паказаў некалькі прататыпаў мадэляў універсальных мабільных робатаў на гусенічных шасі. З дапамогай інфрачырволага датчыка робаты ўмеюць не толькі распазнаваць наяўнасць перашкоды, але і тып перашкоды (сцяна, прыступка, калона) і, дзякуючы галасавой сістэме, могуць паведамляць аб тыпе перашкоды на беларускай мове [10].

Разгледжаныя ў артыкуле інтэрнэт-сэрвісы з'яўляюцца карысным і эфектыўным сродкам аўтаматычнай апрацоўкі тэкстаў на беларускай, рускай і англійскай мовах. Паасобку кожны сэрвіс дае магчымасць вырашыць пэўную камп'ютарна-лінгвістычную задачу, а ў сукупнасці яны дазваляюць атрымаць якасны вынік апрацоўкі электроннага тэксту.

Літаратура

1. Text-to-Speech PHP-Based Synthesizer [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/tts3/>. – Дата доступу: 25.03.2016.
2. The system of generation of phonetic transcriptions for input electronic texts in belarusian / Yu. Hetsevich [et al.] // Pattern Recognition and Information Processing : Proceedings Of The 12th International Conference ; Minsk, Belarus, 28–30 May. – Minsk : УІР НАСБ, 2014. – С. 81–85.
3. Генератар транскрыпцый [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/TranscriptionGenerator/>. – Дата доступу: 25.03.2016.
4. Carnegie Mellon University Pronouncing Dictionary [Электронны рэсурс]. – Рэжым доступу: <http://svn.code.sf.net/p/cmuspinx/code/trunk/cmudict/cmudict.0.7a>. – Дата доступу: 25.03.2016.
5. Агучаны электронны граматычны слоўнік [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/VoicedElectronicGrammaticalDictionary/>. – Дата доступу: 25.03.2016.
6. Праверка правапісу [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/SpellChecker/>. – Дата доступу: 27.03.2016.
7. Праверка правапісу “ў” [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/ShortUSpellChecker/>. – Дата доступу: 26.03.2016.
8. Ідэнтыфікатар амографаў [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/HomographIdentifier/>. – Дата доступу: 27.03.2016.
9. Генератар спасылак на публікацыі [Электронны рэсурс]. – 2016. Рэжым доступу: <http://corpus.by/PublicationReferenceGenerator/>. — Дата доступу : 28.03.2016.
10. TUT.BY На “ТІБО-2013” паказалі ўніверсальнага мабільнага беларускамоўнага робата // Лабараторыя распазнавання і сінтэзу маўлення [Электронны рэсурс]. – 2016. Рэжым доступу: <http://ssrlab.by/archives/802>. – Дата доступу: 29.03.2016.

Т. А. Кісель (Брэсцкі дзяржаўны ўніверсітэт імя А.С. Пушкіна, Брэст)

НАЗВЫ ХВАЁВЫХ ДРЭЎ У ГАВОРКАХ БРЭСТЧЫНЫ

Самым распаўсюджаным відам дрэў у нашай краіне з'яўляецца хвоя (сасна) звычайная, а каля 70 % плошчы лясных масіваў займаюць менавіта хваёвыя лясы. Таму невыдкова, што ў маўленні кожнага жыхара адзначаюцца лексемы дадзенай тэматычнай групы.