

XIII Международная конференция

**РАЗВИТИЕ ИНФОРМАТИЗАЦИИ
И ГОСУДАРСТВЕННОЙ СИСТЕМЫ
НАУЧНО-ТЕХНИЧЕСКОЙ
ИНФОРМАЦИИ**

РИНТИ-2014



20 ноября 2014 года, Минск

ДОКЛАДЫ

Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2014) : доклады XIII Международной конференции, Минск, 20 ноября 2014 г. – Минск : ОИПИ НАН Беларуси, 2014. – 360 с. – ISBN 978-985-6744-86-3.

Представлены доклады XIII Международной конференции «Развитие информатизации и государственной системы научно-технической информации» (РИНТИ-2014), Минск, 20 ноября 2014 г., в которых рассматриваются вопросы развития государственной системы научно-технической информации Республики Беларусь, стратегии информатизации и информационной безопасности системы образования в условиях развития информационного общества, создания облачного комплекса информационно-технологических систем для библиотек Беларуси, хода реализации Национальной программы ускоренного развития услуг в сфере информационно-коммуникационных технологий (ИКТ) на 2011–2015 гг., эффективности использования ИКТ для социально-экономического развития республики, электронного официального опубликования в Республике Беларусь, создания Национальной электронной библиотеки Беларуси, развития информатизации и государственной системы научно-технической информации в Национальной академии наук Беларуси в 2013–2014 гг., а также проектирования систем научно-технической информации в Беларуси, создания корпоративных библиотечно-информационных систем и технологий.

Материалы конференции будут полезны специалистам в области ИКТ, занимающимся разработкой и внедрением автоматизированных информационных систем управления и систем научно-технической информации, развитием информационной инфраструктуры Беларуси и других стран, реализацией проектов государственных программ в сфере информатизации.

Одобрены программным комитетом и печатаются по решению редакционной коллегии Объединенного института проблем информатики Национальной академии наук Беларуси в виде, представленном авторами.

Научные редакторы:

доктор физико-математических наук, профессор А.В. Тузиков;
кандидат технических наук, доцент Р.Б. Григянец;
кандидат технических наук В.Н. Венгеров

Чернявский Ю.А., Шемаров А.И., Белицкая И.В. Автоматизация оценки деятельности профессорско-преподавательского состава.....	252
Борискевич И.А. Выбор вейвлет-преобразования для пространственно-частотного ковариационного поиска цели	257
Веремейчик Д.В., Скудняков Ю.А. Исследование возможностей адаптации и встраивания отечественных криптографических алгоритмов с открытым ключом на основе эллиптических кривых в современные программные среды.....	262
Гецэвіч Ю.С., Лысы С.І. Праектаванне інтэрнэт-сэрвісаў для працэсараў сінтэзатара маўлення па тэксте з магчымасцю прадстаўлення бясплатных электронных паслуг насельніцтву	265
Коваленко Н.С. Построение оптимальной системы одинаково распределенных процессов в многопроцессорных вычислительных комплексах.....	270
3. КОРПОРАТИВНЫЕ БИБЛИОТЕЧНО-ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ	
Мыхлик В.В. Корпоративное сотрудничество в формировании национальных электронных ресурсов	278
Рожко А.В., Масякин В.Б., Надыров Э.А., Башилов А.В., Борисевич Н.Я., Иванов В.К., Максюттов М.А. Единый чернобыльский регистр России и Беларуси ...	283
Астапович Л.Л., Старовойтова О.И. Поиск и использование интернет-источников при создании авторитетных записей на «тематический предмет»	285
Ахремчик Р.В., Пинчук Т.В. Персональные указатели ученых НАН Беларуси в цифровую эпоху: опыт Центральной научной библиотеки НАН Беларуси	291
Бабарико Д.П., Важник М.Н. Применение интранет-технологий при обслуживании пользователей Белорусской сельскохозяйственной библиотеки	296
Грядовкина М.Ю. Форма авторитетной точки доступа имени лица на языке оригинала в сводном электронном каталоге системы корпоративной каталогизации библиотек Беларуси	301
Куранда В.Н., Кочеткова Е.П. Конкурентоспособность ИТ-отрасли Украины и стран мира.....	306
Сивурова О.А., Сивуров Д.В. Роль республиканских библиотек Беларуси в поддержании и развитии отраслевых систем НТИ на примере сельского хозяйства и здравоохранения.....	312
Сикорская О.Н., Бовкунович М.А. Показатели публикационной активности белорусских ученых на примере базы данных Scopus.....	318
Богданова И.Ф., Шемаров А.И., Богданова Н.Ф. Электронные книги: история и современность	322
Скалабан А.В., Юрик И.В. Роль научной библиотеки в формировании информационно-образовательного пространства вуза	327

ПРАЕКТАВАННЕ ІНТЭРНЭТ-СЭРВІСАЎ ДЛЯ ПРАЦЭСАРАЎ СІНТЭЗАТАРА МАЎЛЕННЯ ПА ТЭКСЦЕ З МАГЧЫМАСЦЮ ПРАДСТАЎЛЕННЯ БЯСПЛАТНЫХ ЭЛЕКТРОННЫХ ПАСЛУГ НАСЕЛЬНІЦТВУ

Ю.С. Гецэвіч, С.І. Лысы

Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі, Мінск

Апісана канцэпцыя праектавання і выкарыстання інтэрнэт-сэрвісаў у мэтах паляпшэння і ўдасканалення працы сістэмы сінтэзу маўлення па тэксце з адначасовым прадстаўленнем бясплатных электронных паслуг насельніцтву.

Уводзіны

Адным з асноўных сродкаў камунікацыі і ўзаемадзеяння ў чалавечым асяроддзі з'яўляецца маўленне. Таму з часоў стварэння першых ЭВМ і па сённяшні дзень важнай і актуальнай задачай застаецца ўсталяванне камунікацыі паміж чалавекам і кампутарам пры дапамозе натуральнай мовы. Дзеля вырашэння гэтай задачы навукоўцы працуюць над стварэннем сістэм сінтэзу і распазнавання маўлення. У апошні час сістэмы сінтэзу маўлення па тэксце (ССМТ) ужо дасягнулі пэўнай дасканаласці і шырока выкарыстоўваюцца. Аднак нельга сказаць, што задача сінтэзу маўлення з'яўляецца вырашанай, асабліва ў дачыненні да беларускай мовы [1, 2].

У сувязі з гэтым аўтарамі было прынята рашэнне стварыць шэраг інтэрнэт-сэрвісаў, кожны з якіх вырашаў бы пэўныя свае ўласныя задачы (напрыклад, прыкладныя лінгвістычныя задачы), але ў той жа час тым ці іншым чынам дапамагаў распрацоўшчыкам ва ўдасканаленні працы сінтэзатара маўлення па тэксце. Гэта дазволіць праз ужыванне сэрвісаў вялікай колькасцю знешніх карыстальнікаў правесці тэставанне і збор неабходнай інфармацыі, пры дапамозе якіх будзе ўдасканальвацца і паляпшацца якасць працы сінтэзатара маўлення па тэксце для беларускай мовы.

У дакладзе прадстаўлена канцэпцыя выкарыстання і праектавання інтэрнэт-сэрвісаў у мэтах паляпшэння працы ССМТ з прадстаўленнем бясплатных электронных паслуг насельніцтву праз інтэрнэт-сайт www.corpus.by.

1. Агульныя падыходы да праектавання інтэрнэт-сэрвісаў

Распрацоўка ССМТ з'яўляецца складанай задачай, якая ахоплівае вялікую колькасць неабходных падзадач, а таксама сутыкаецца з шэрагам пабочных прыкладных лінгвістычных задач, якія патрабуюць вырашэння. Задачы, якія паўстаюць пры стварэнні ССМТ, можна ўмоўна падзяліць на тры групы:

- перадапрацоўку ўваходных тэкстаў;
- непасрэдна сінтэз маўлення;
- апрацоўку і выкарыстанне выніковых дадзеных.

Інтэрнэт-рэсурс www.corpus.by (мал. 1) мае сваёй мэтай рашэнне як непасрэднай задачы сінтэзавання маўлення, так і шэрагу дадатковых задач перадапрацоўкі тэксту і апрацоўкі вынікаў працы сінтэзатара маўлення. Дзеля дасягнення дадзенай мэты было прынята рашэнне распрацаваць для гэтага рэсурса сістэму інтэрнэт-сэрвісаў, кожны з якіх вырашаў бы пэўную ўласную задачу, а ў спалучэнні з іншымі сэрвісамі – неабходныя падзадачы ССМТ. На сённяшні дзень распрацаваны і даступны бясплатна ў Інтэрнэце наступныя сэрвісы:

- сэрвіс канвертавання алафоннага тэксту ў розныя транскрыпцыі;
- генератар арфаэпічнага слоўніка;
- сэрвіс праверкі «у» і «ў»;
- частотнасць слоў;
- сэрвіс пошуку і сартыроўкі алафонаў;
- сэрвіс генеравання назваў сімвалаў (кадыроўка Windows-1251);
- іншыя.

Text-to-speech synthesis

Service of searching and sorting of allophones

Сэрвіс пошуку і сартыроўкі алафонаў

Service of converting allophonic texts into different transcriptions

Сэрвіс канвертавання алафоннага тэксту ў розныя транскрыпцыі

Service of generating names of characters (encoding Windows-1251)

Сэрвіс генеравання назваў сімвалаў (кадыроўка Windows-1251)

Orthoepic Dictionary Generator

Генератар арфаэпічнага слоўніка

Spell-check for Belarusian U / Сэрвіс праверкі "у" і "ў"

Frequency of Words / Частотнасць слоў

Service of getting page by URL / Сэрвіс узяцця старонкі па спасылцы

Spell-check form

Get Publication References / Генератар спасылкі на публікацыю

Specialized Dictionary

Мал. 1. Знешні інтэрфейс інтэрнэт-рэсурса www.corpus.by

2. Фармаванне патрабаванняў да праектуемых інтэрнэт-сэрвісаў

Мэтай распрацоўкі інтэрнэт-сэрвісаў з'яўляецца павышэнне якасці працы сінтэзатара маўлення па тэксце для беларускай мовы. Гэта можа быць дасягнута некалькімі спосабамі: 1) праз вынясенне пэўных элементаў існуючай ССМТ для вольнага карыстання ў Інтэрнэт у выглядзе інтэрнэт-сэрвісаў, якія б мелі пэўную мэтавую аўдыторыю; 2) праз распрацоўку пэўных інтэрнэт-сэрвісаў, якія б маглі быць у далейшым інтэграваныя ў ССМТ, а пакуль удасканаленыя і адтэставаныя ў Інтэрнэце. Абазначанымі спосабамі можна атрымаць:

1) непасрэдня вынікі працы кожнага сэрвісу на адвольна зададзеных карыстальнікам уваходных дадзеных. Мноства адпаведнасцяў «уваходныя дадзеныя – выніковыя дадзеныя» можа быць выкарыстана як для аналізу правільнасці працы таго ці іншага элемента ССМТ, так і для стварэння аўтаматызаванага тэставання працы сэрвісу і ССМТ;

2) адваротную сувязь ад мэтавай аўдыторыі і экспертаў, у супрацоўніцтве з якімі вядзецца распрацоўка;

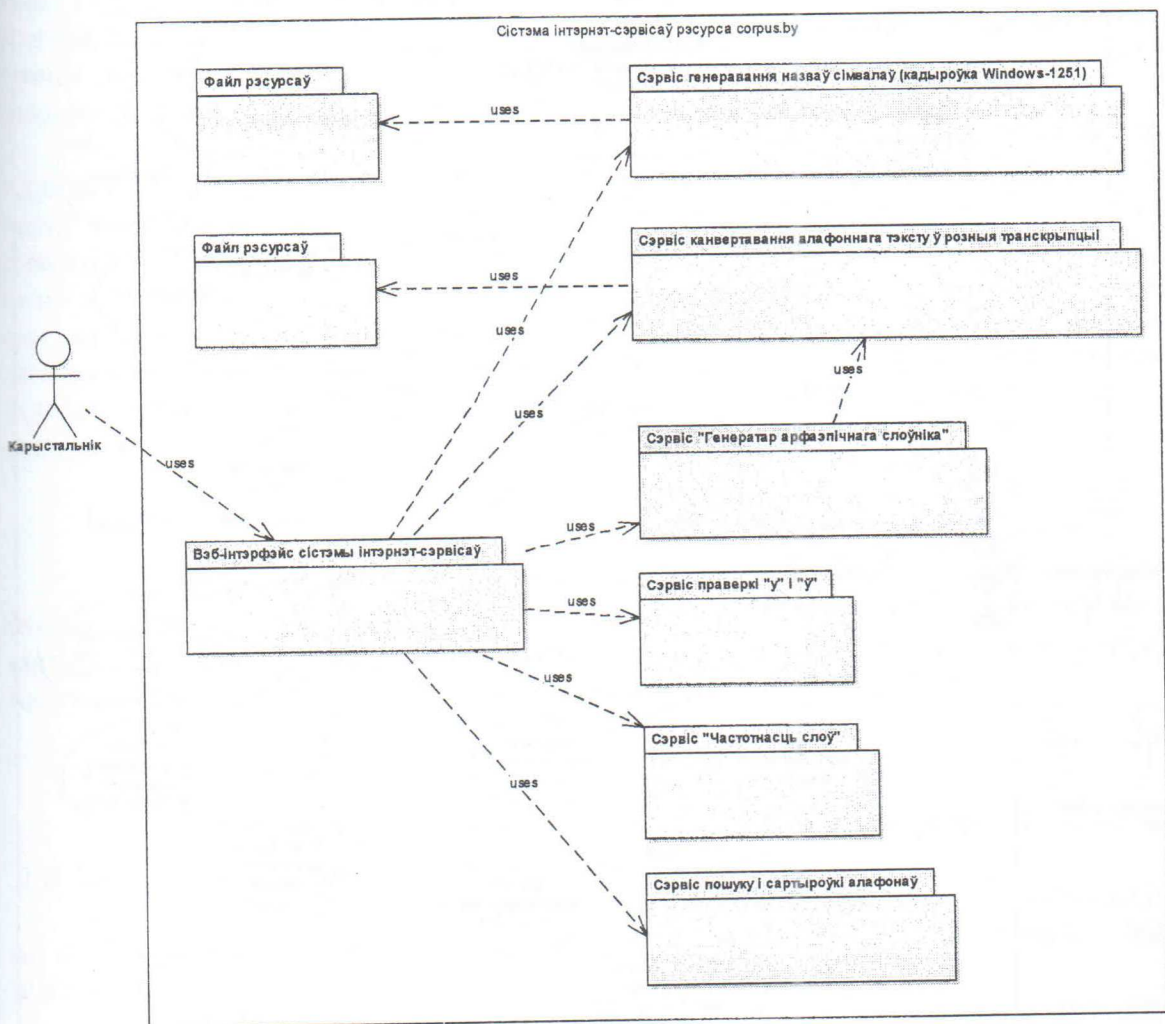
3) адваротную сувязь ад звычайных карыстальнікаў.

Прымаючы да ўвагі пералічаныя вышэй мэты, можна сфармуляваць наступныя асноўныя патрабаванні да праектуемых інтэрнэт-сэрвісаў:

- прастасць, зручнасць і інтуітыўная зразумеласць карыстальніцкага інтэрфейсу;
- захаванне ўсіх дадзеных, што пададзены на ўваход;
- захаванне ўсіх выніковых дадзеных;
- аўтаапавяшчэнне распрацоўшчыкаў аб памылках у працы сэрвісаў.

Праектаванне інтэрнэт-сэрвісаў у прылажэнне да ССМТ уключае ў сябе вызначэнне патрабаванняў, апісанне статычнай структуры і ўзаемадзеянне элементаў сістэмы (мал. 2). У якасці графічнай мовы дакументавання выкарыстоўваецца натацыя UML 2.0 (Unified Modeling Language) [3].

Пяройдзем да разгляду функцыянальных магчымасцяў праектуемых інтэрнэт-сэрвісаў.

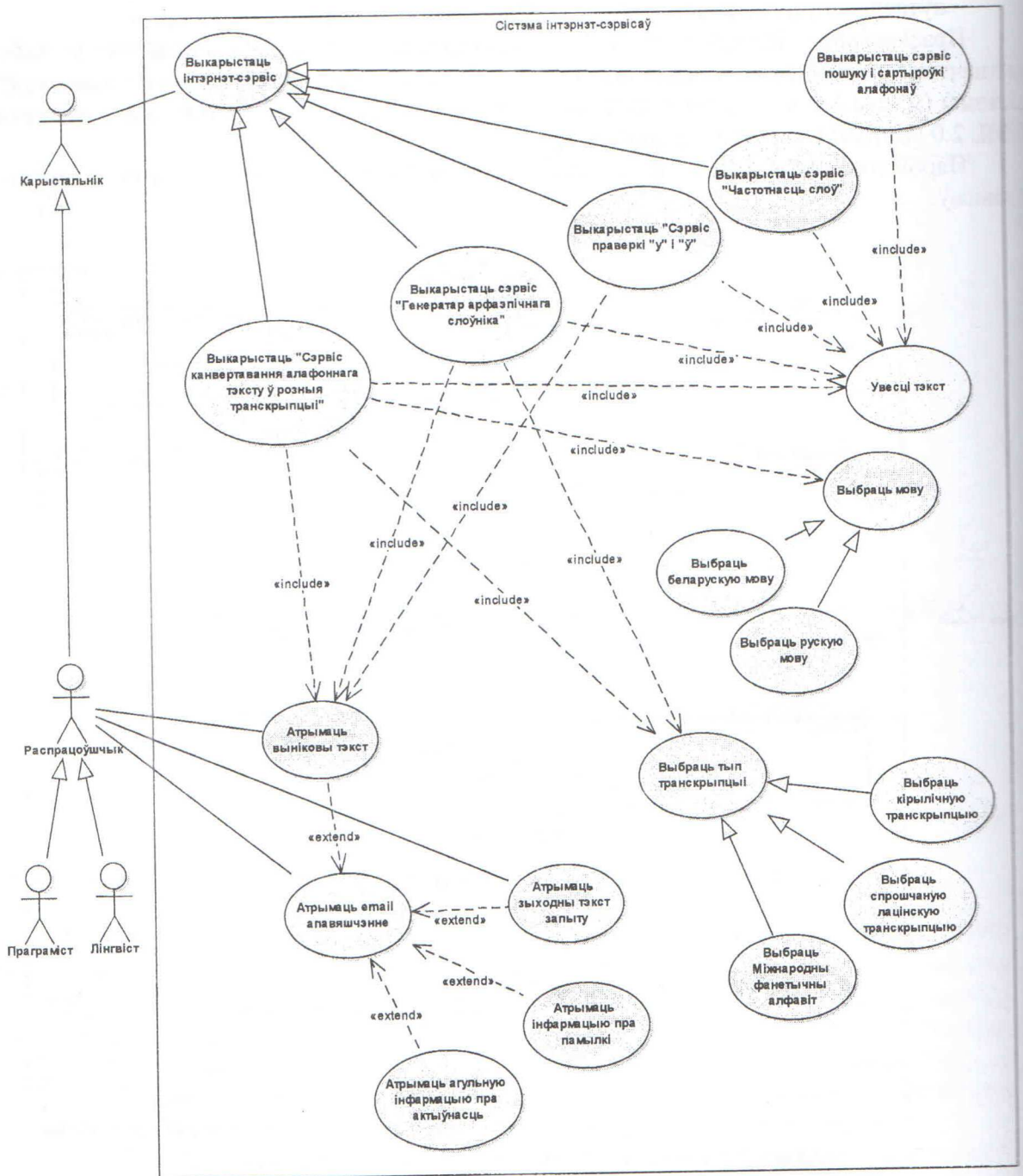


Мал. 2. Схема кампанентаў узаемадзеяння карыстальніка з сістэмай інтэрнэт-сэрвісаў www.corpus.by

3. Праектаванне варыянтаў выкарыстання інтэрнэт-сэрвісаў

Праектуемая сістэма інтэрнэт-сэрвісаў прадугледжвае выкарыстанне яе як знешнімі карыстальнікамі, так і распрацоўшчыкамі. Выкарыстанне інтэрнэт-сэрвісаў аднымі і другімі можа як супадаць, так і адрознівацца, бо шэраг распрацоўшчыкаў з'яўляюцца адначасова і карыстальнікамі ў сваіх працоўных мэтах.

На мал. 3 прадстаўлена схема варыянтаў выкарыстання сістэмы інтэрнэт-сэрвісаў карыстальнікамі і распрацоўшчыкамі.



Мал. 3. Схема варыянтаў выкарыстання сістэмы інтэрнэт-сэрвісаў www.corpus.by

На схеме варыянтаў выкарыстання прыведзены наступныя пяць сэрвісаў: канвертавання алафоннага тэксту ў розныя транскрыпцыі, генератар арфаэпічнага слоўніка, праверкі «у» і «ў», частотнасці слоў, пошуку і сартыроўкі алафонаў. У кожным з іх карыстальнік можа падаць на ўваход пэўны тэкст альбо паслядоўнасць сімвалаў і атрымаць на выхадзе неабходны выніковы тэкст у адпаведнасці з задачай, якую вырашае дадзены сэрвіс. У «Сэрвісе канвертавання алафоннага тэксту ў розныя транскрыпцыі» і сэрвісе «Генератар арфаэпічнага слоўніка» карыстальнік можа абраць

адзін з трох тыпаў транскрыпцыі: кірылічную, спрошчаную лацінскую ці Міжнародны фанетычны алфавіт. У «Сэрвісе канвертавання алафоннага тэксту ў розныя транскрыпцыі» таксама магчымы выбар мовы ўваходнага тэксту: беларускай альбо рускай.

Дадзеныя сэрвісы праектуюцца як для звычайных карыстальнікаў, так і для адмысловай мэтавай аўдыторыі: лінгвістаў-экспертаў і ўласна распрацоўшчыкаў дадзеных сэрвісаў і ССМТ. Такім чынам, распрацоўшчыкі (лінгвісты і праграмісты) з'яўляюцца і прыватным выпадкам карыстальніка, але між іншага валодаюць і сваімі ўласнымі спецыфічнымі магчымасцямі выкарыстання.

Распрацоўшчыкі могуць выкарыстоўваць інтэрнэт-сэрвісы для аналізу працы сэрвісаў і прыняцця рашэнняў па дапрацоўцы і выпраўленню памылак іх працы, а таксама для вырашэння шляхоў развіцця, пошукаў сродкаў і спосабаў павышэння эфектыўнасці працы праз назіранне за пэўнымі тэндэнцыямі і заканамернасцямі ў карыстальніцкіх запытах да сэрвісаў.

Звычайныя карыстальнікі маюць магчымасць бясплатна карыстацца інтэрнэт-сэрвісамі сайта www.corpus.by у любых мэтах. Прывядзём некалькі прыкладаў. Сінтэзатар маўлення па тэкстце (www.corpus.by/tts3) дазваляе згенераваць гукавы файл і дазволіць яго ці скачаць, ці зрабіць даступным праз глабальную інтэрнэт-спасылку любому карыстальніку Інтэрнэта. Адвольны тэкставы радок на беларускай мове можна сканвертаваць паводле Міжнароднай фанетычнай транскрыпцыі і даслаць тым, каму складана чытаць кірыліцу (<http://www.Corpus.by/convertAllophonesToDifPhonemes/>). Адвольны тэкст карыстальнік можа правесці на арфаграфічную правільнасць напісання нескладовага ў (<http://www.Corpus.by/convertAllophonesToDifPhonemes/>).

Заклучэнне

У выніку прыведзенай працы была распрацавана канцэпцыя стварэння сістэмы інтэрнэт-сэрвісаў www.corpus.by у мэтах паляпшэння і ўдасканалення працы сістэмы сінтэзу маўлення па тэкстце з адначасовым прадстаўленнем бясплатных электронных паслуг насельніцтву ў вырашэнні розных прыкладных задач.

Спіс літаратуры

1. Лобанов, Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск : Белорусская наука, 2008. – 344 с.
2. Гецэвіч, Ю.С. Алгарытмы лінгвістычнай апрацоўкі тэкстаў для сінтэзу маўлення на беларускай і рускай мовах: дыс. ... канд. тэхн. навук / Ю.С. Гецэвіч. – Мінск, 2012. – 193 с.
3. Фаулер, М. UML. Основы / М. Фаулер, К. Скотт. – СПб. : Символ-плюс, 2002. – 192 с.