

Кривальцэвіч, А.В. Апрацоўка колькасных выказаў у мабільным прыкладанні турыстычнай тэматыкі як сродак вывучэння культурна-маўленчых асаблівасцяў краіны / А.В. Кривальцэвіч, Я.С. Качан, Ю.С. Гецэвіч // Актуальныя праблемы сучаснай прыкладнай лінгвістыкі : сб. науч. статей / редкол. : Л.Н. Беляева, Т.В. Бусел, Р.В. Детскина [и др.]. – Минск : МГЛУ, 2017. – С. 150-154.

Journal of the International Phonetic Association

№ 2. – 2017

10.1017/S0022268917000011

А. В. Кривальцэвіч, Я. С. Качан, Ю. С. Гецэвіч

**АПРАЦОЎКА КОЛЬКАСНЫХ ВЫРАЗАЎ
У МАБІЛЬНЫМ ПРЫКЛАДАННІ ТУРЫСТЫЧНАЙ ТЭМАТЫКІ
ЯК СРОДАК ВЫВУЧЭННЯ КУЛЬТУРНА-МАЎЛЕНЧЫХ
АСАБЛІВАСЦЯЎ КРАІНЫ**

Інфармацыйныя тэхналогіі сучаснасці прадстаўляюць сабой дынамічную вобласць, развіццё якой паўплывала на ўсе сферы жыцця чалавека. Але нельга не заўважыць, што інфарматызацыя адбываецца не толькі праз распаўсюджванне персанальных камп'ютараў, але і праз павышэнне якасці працы мабільных тэлефонаў, а больш дакладна – смартфонаў.

Згодна заяве кіраўніка каманды Пошука Google, Аміта Сінгхала (Amit Singhal), летам 2015 года колькасць мабільных пошукаў у Google упершыню перавысіла колькасць дэсктопных [1]. З гэтага можна зрабіць выснову, што павялічылася колькасць карыстальнікаў мабільнымі прыстасаваннямі, а гэта, у выніку дае павелічэнне выкарыстання мабільных прыкладанняў.

Мабільнае прыкладанне – гэта праграмае забеспячэнне прыкладнога характару, якое выкарыстоўваецца для аптымізацыі працы карыстальніка на мабільных прыстасаваннях.

Таму важна далучыцца да распрацовак мабільных прыстасаванняў у сферы турызму, якія могуць быць карыснымі як для жыхароў Рэспублікі Беларусь, так і для замежных грамадзян.

На падставе зместу дзяржаўнай праграмы развіцця турызму ў Рэспубліцы Беларусь, вядома, што турызм з’яўляецца адной з важнейшых сфер эканомікі [2].

У якасці прыкладу ў дадзеным даследаванні разглядаецца мабільнае прыкладанне *KrokApp*, якое распрацоўваецца ў лабараторыі распазнавання і сінтэзу маўлення АПП НАН Беларусі.

Канцэпцыя мабільнага прыстасавання накіравана на тры асноўныя мэты: экскурсійная прэзентацыя гарадоў без удзелу чалавека-экскурсавода, фарміраванне цікавасці да гісторыі Беларусі і стварэнне станоўчага іміджа краіны. Мэтавай аўдыторыяй з’яўляюцца жыхары Рэспублікі Беларусь і турысты, якія валодаюць англійскай мовай.

Структура мабільнага прыкладання *KrokApp* уключае ў сябе сукупнасць наступных старонак: старонка-лагатып; галоўнае меню, у якім ёсць магчымасць выбару мовы (беларуская, англійская, руская); старонка выбару тыпу маршруту (экскурсійны, кафэ і рэстараны або шопінг); старонка-пералік гарадоў (Мінск, Гродна, Брэст, Віцебск, Магілёў, Гомель); старонка выбару тыпу экскурсіі (тэматычная экскурсія, аглядная экскурсія); старонка-пералік кропак агляду для кожнага горада адпаведна; старонкі-агляды кропак агляду для кожнага горада адпаведна.

На дадзены момант створаны прататып, які можна спампаваць у Google Play Market па наступнай спасылцы: <https://play.google.com/store/apps/details?id=by.ssrlab.krokapp&hl=be>.

У прыкладанні быў прапрацаваны беларускамоўны падраздзел. Для апісання кропак агляду дададзены фотаздымкі месцаў ці будынкаў, друкаваны тэкст і гукавы файл.

Для змяншэння памеру мабільнага прыкладання па меры дабаўлення тэкстаў апісанняў у будучым плануецца агучваць тэксты з дапамогай сэрвіса, распрацаванага на базе лабараторыі распазнавання і сінтэзу маўлення – “Сінтэз маўлення па тэксце” [3].

Адной з асноўных праблем для сінтэзатара беларускага маўлення з’яўляецца ўспрыманне і агучванне незразумелых для машыны сімвалаў (лікі, скарачэнні, даты і г.д.) [3]. Таму аўтары дадзенага артыкула лічаць неабходным распрацаваць спіс правяраных нармалізаваных мностваў колькасных выказаў, які будзе пакладзены ў аснову аўтаматызаванага працэсу нармалізацыі ўсіх невядомых сімвалаў.

Крыніцай матэрыялаў сталі тэксты апісанняў кропак агляду, падрыхтаваныя для агучвання. Сабраны матэрыял быў аформлены як спіс сказаў з ненармалізаванымі выразамі па наступных семантычных катэгорыях:

1. Час (г., гг., ст., -х).
2. Памер і габарыты (м, км, га).
3. Выпадковыя дадзеныя (скарачэнні, знак нумару, абрэвіятуры і інш.).

Важна адзначыць, што “Сінтэзатар маўлення па тэксце” (Text-to-Speech PHP-based Synthesizer) ажыццяўляе агучванне тэксту на беларускай і рускай мовах [3]. Але карэктная апрацоўка спецыфічных токенаў (лікаў, дат, скарачэнняў, адзінак вымярэння, абрэвіатур і інш.) на дадзены момант адсутнічае.

У працэсе даследвання былі вылучаны галоўныя правілы, паводле якіх адбывалася экспертная генерацыя лікаў у колькасных і парадкавых лічэбнікі. З правіламі можна азнаёміцца ў папярэдніх артыкулах па гэтай тэме: Гецэвіч Ю.С. і інш. “Алгарытм і лінгвістычныя рэсурсы для нармалізацыі тэкстаў геаграфічнага дамена” [4], Гецэвіч Ю.С. “Мадэляванне і распрацоўка сістэм пошуку колькасных выказаў з адзінкамі вымярэння ў электронных тэкстах на беларускай і рускай мовах” [5].

Паводле гэтага алгарытму атрымаўся спіс нармалізаваных дадзеных, размеркаваных па семантычных катэгорыях (час, памеры і габарыты, выпадковыя дадзеныя) (табл. 1):

Табліца 1

Фрагмент спіса нармалізаваных дадзеных па семантычных катэгорыях

Катэгорыя	Выраз па-беларуску	Нармалізаваны сказ на беларускай мове
Час	Мінскі чыгуначны вакзал (да 1928 г. – Віленскі вакзал) – пасажырскі тэрмінал чыгуначнай станцыі «Мінск-Пасажырскі».	Мінскі чыгуначны вакзал (да тысяча дзевяцьсот дваццаць восьмага года – Віленскі вакзал) – пасажырскі тэрмінал чыгуначнай станцыі «Мінск-Пасажырскі».
Памер і габарыты	Вышыня Фарнага касцёла дасягае 50 м, што робіць яго прыкметным з любой кропкі горада.	Вышыня Фарнага касцёла дасягае пяцьдзсят метраў, што робіць яго прыкметным з любой кропкі горада.

У будучыні плануецца дадаць лінгвадыдактычную функцыю на вывучэнне беларускіх слоў для замежнікаў, якія валодаюць англійскай мовай. Яе працу можна апісаць двума сцэнарыямі.

Сцэнарыі 1. На працягу экскурсіі карыстальніку прапаноўваецца азнаёміцца з беларускімі адпаведнікамі некаторых слоў (малюнак 1). Прыкладны сцэнар узаемадзеяння карыстальніка з мабільным прыстасаваннем наступны:

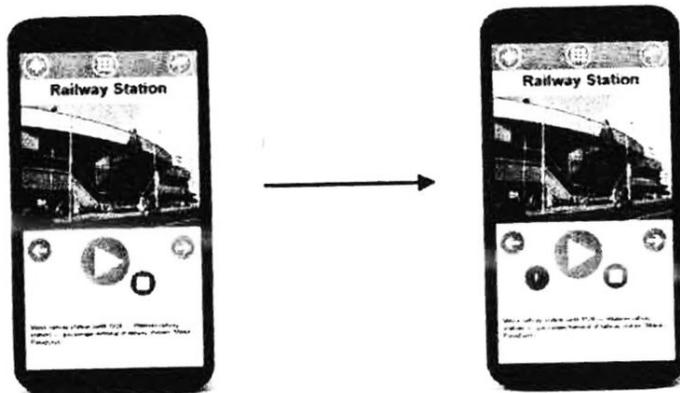
1. На экране паказваецца інфармацыя пра славутасць.
2. Прапаноўваецца націснуць на назву славутасці.
3. Пасля націскання карыстальнік слухае пераклад слова *Railway station* – ‘Вакзал’.

4. Прапаноўваецца паўтарыць за аўдыёгідам слова 'Вакзал'. На пра-
дальнае прагаворванне даецца 3 спробы.

5. Падсумоўваюцца балы за правільна прагавораныя словы.

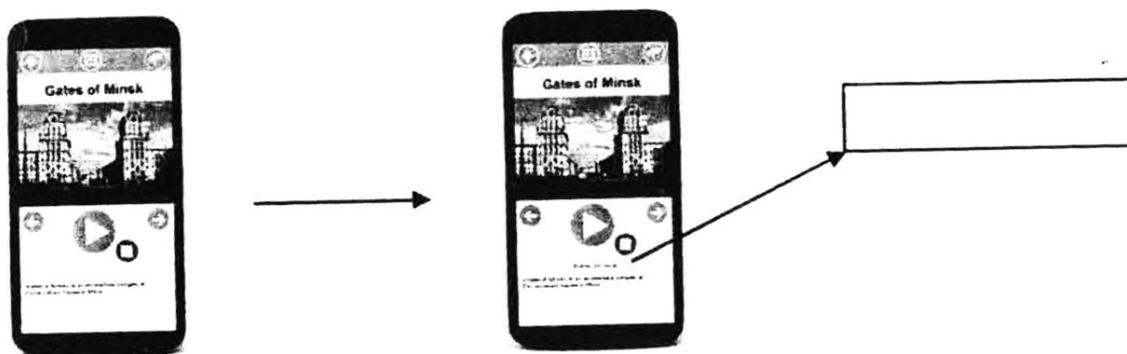
6. Карыстальніку выказваецца падзяка па заканчэнні экскурсіі, паказ-
ваецца агульная сума балаў, за якую ён можа атрымаць пэўны бонус.

Для аўтаматычнага агучвання беларускамоўных тэкстаў прапаноўваецца
выкарыстанне сервісаў, распрацаваных лабараторыяй распазнавання і сінтэзу
маўлення: Text-to-Speech PHP-Based Synthesizer ('Сінтэзатар маўлення па
тэксце') [6].



Малюнак 1. Прыклад першага варыянта ўзаемадзеяння
карыстальніка з мабільным прыкладаннем падчас
вывучэння назваў беларускіх славутасцяў

Сцэнары 2. Карыстальнік можа націснуць на назву славутасці
(Загаловак) і пабачыць транскрыпцыю гэтага слова ці фразы ў міжнародным
фармаце (International phonetic alphabet – IPA) (малюнак 2). Для гэтага будзе
выкарыстаны сервіс "Transcription Generator" ('Генератар транскрыпцый'),
які быў створаны на базе лабараторыі распазнавання і сінтэзу маўлення [7].



Малюнак 2. Прыклад другога варыянта ўзаемадзеяння карыстальніка
з мабільным прыкладаннем пры вывучэнні назваў беларускіх славутасцяў

Такім чынам, замежнік будзе мець магчымасць не толькі праслухаць
экскурсію па горадзе, але і вывучыць назвы беларускіх рэалій. Экскурсія на
англійскай мове можа быць прапанавана таксама і беларускім студэнтам
адпаведных спецыяльнасцей для павышэння навыкаў аўдыявання ці падрых-
тоўкі ўласнай экскурсіі для сваёй групы.

1. *Zakrzewski, Cat.* Mobile Searches Surpass Desktop Searches At Google For The First Time [Electronic resource] / Cat Zakrzewski. – Mode of access : <http://techcrunch.com/2015/10/08/mobile-searches-surpass-desktop-searches-at-google-for-the-first-time/?ncid=rss#.54rwqr:fdAV>. – Date of access : 12.02.2016.
2. Государственная программа развития туризма в Республике Беларусь на 2011-2015 год [Электронны рэсурс]. – Рэжым доступу : <http://mst.gov.by/ru/programma-razvitiya-turizma>. – Дата доступу : 8.02.2016.
3. Text-to-Speech PHP-Based Synthesizer [Электронны рэсурс]. – Рэжым доступу : <http://www.corpus.by/tts3/>. – Дата доступу : 15.01.2016.
4. *Гецэвіч, Ю. С.* Алгарытм і лінгвістычныя рэсурсы для нармалізацыі тэкстаў геаграфічнага дамена / Ю. С. Гецэвіч, Я. С. Качан, С. І. Лысы [і інш.] // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2016) : матер. VI Междунар. науч.-техн. конф., Минск, 18–20 февр. 2016 г. / БГУИР; пад рэд. В. В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР, 2016. – С. 397–400.
5. *Гецэвіч, Ю. С.* Мадэляванне і распрацоўка сістэм пошуку колькасных выказаў з адзінкамі вымярэння ў электронных тэкстах на беларускай і рускай мовах / Ю. С. Гецэвіч, А. М. Скопінава, А. Ф. Есіс // Развитие информатизации и государственной системы научно-технической информации (РИНТИ-2013) : докл. XII Междунар. конф., Минск, 20 нояб. 2013 г. – Минск : ОИПИ НАН Беларуси, 2013. – С. 282–287.
6. Лабараторыя распазнавання і сінтэзу маўлення. Сэрвіс Text-to-Speech PHP-Based Synthesizer [Electronic resource]. – Mode of access : <http://corpus.by/tts3/>. – Date of access : 9.02.2016.
7. Лабараторыя распазнавання і сінтэзу маўлення. Сэрвіс “Генератар транскрыпцый” [Электронны рэсурс]. – Рэжым доступу : <http://ssrslab.by/archives/3286>. – Дата доступу : 10.02.2016.

А. В. Молоткова

СРЕДСТВА ВЫРАЖЕНИЯ ИМЕНОВАННЫХ СУЩНОСТЕЙ И ФАКТОВ В ПИСЬМЕННОМ ТЕКСТЕ

Преобразование неструктурированной текстовой информации в хорошо организованные и структурированные данные и знания является одной из основных задач современных технологий *Data Mining, Text Mining, Information Extraction*. Указанные технологии положены в основу создания компьютерных систем, способных правильно извлекать из текстовых массивов различные объекты и факты. Например, даты каких-либо событий, имена, адреса и телефоны персон, наименования товаров и услуг, названия компаний, описания событий в новостных сводках, мнения, оставленные пользователями социальных сетей, гостевых книг или форумов, отзывы о разнообразных продуктах и услугах и т.д. Успешность функционирования подобных систем зависит как от самой извлекаемой информации, т.е. количества и типов извлекаемых объектов, фактов и связей между ними, так и от способа представления и средств обработки полученных данных и знаний,