

ЭПІЧНАЯ ПРАЦА НА НІВЕ АРФАЭПІІ

07.08.2017/№32



Сёлета выйшаў першы арфаэпічны беларускі слоўнік.

Прэзентацыя прызначана на верасень, але ў СМІ яго ўжо актыўна абмяркоўваюць. Такі рэзананс на розных узроўнях сведчыць аб правільным выбары накірунку працы каманды з двух інстытутаў НАН Беларусі. Ад Інстытута мовазнаўства імя Якуба Коласа – доктар філалагічных навук Валянціна РУСАК (на фота) і кандыдат філалагічных навук Вераніка МАНДЗІК. Ад Аб'яднанага інстытута праблем інфарматыкі (АІПІ) – кандыдат тэхнічных навук Юрась ГЕЦЭВІЧ і магiстр фізіка-матэматычных навук Станіслаў ЛЫСЫ.

– Чым прадывітаваны менавіта такі склад? У чым ваш слоўнік – піянер?

В.Р.: Мы прадставілі поўную транскрыпцыю 117 тыс. слоў беларускай мовы. Складаныя ці неўсталяваныя выпадкі падаюцца з варыянтамі вымаўлення: «менеджмент», «мэнэджмент» і г.д. Таксама ў выданні сістэмна прастаўлены пачобныя націскі ў складаных словах. Аналагаў гэтай працы няма на ўсёй славянскай прасторы! Напрыклад, у арфаэпічных слоўніках рускай мовы да сённяшняга дня падаюцца толькі фрагменты транскрыпцыі слоў. Дарэчы, па такім прынцыпе быў пабудаваны і слоўнік беларускай мовы 1987 года пад рэдакцыяй М.Бірылы, які сумяшчаў словаўтварэнне, словазмяненне, акцэнтацыю і вымаўленне. У ім таксама падавалася частковая транскрыпцыя слоў, асобных сістэмных з'яў. Лексічнай базай для нашай працы стаў «Слоўнік беларускай мовы» 2012 года.

В.М.: У аспірантуры Інстытута мовазнаўства мне прапанавалі тэму, звязаную з беларускай арфаэпіяй. Многія праблемныя месцы – фанетычныя пазіцыі, апазіцыі, несістэмныя моманты – вырашаліся мной эксперыментальна. І вынік гэтых даследаванняў – на старонках выдання. Наш слоўнік – прыклад таго, як праца над дысертацыяй можа мець практычнае ўжыванне. Для мяне як навукоўцы гэта важна.

– Чаму ўзнік тандэм праграмістаў і лінгвістаў?

В.Р.: Для дасягнення мэты быў патрэбны базавы матэрыял, каб пакласці яго на тэхнічныя сродкі. Лінгвісты ўжо не спраўляліся з аб'ёмам, якія трэба было апрацоўваць «ручную». Да таго ж тэхнічныя сродкі забяспечылі аднастайнасць і сістэмнасць падачы фанетычных з'яў.

В.М.: Лінгвісты сабралі слоўнікавую мовазнаўчую базу, на гэтым «хрыбце» спецыялісты з АІПІ змаглі распрацаваць праграмны прадукт. Ён дазволіў аўтаматызаваць працу па пераўтварэнні арфаграфічнага запісу слова ў яго фанетычную транскрыпцыю.

– Што зрабілі менавіта ў АІПІ?

Ю.Г.: Мой навуковы кіраўнік Барыс Лабанаў «вучыць машыны гаварыць» на той ці іншай мове свету ўжо гадоў 50. Я і Станіслаў працягваем яго справу для беларускай мовы. У межах маёй кандыдацкай я рабіў частку сістэмы, якая перапрацоўвае беларускі тэкст для прамаўлення машынай. Атрымоўваліся праграмы для стацыянарнага камп'ютара. У апошнія 5 гадоў мы са Станіславам запусцілі інтэрнэт-версію сінтэзатара маўлення. Зараз час патрабуе палепшыць гук, але для гэтага патрэбны інвестыцыі. А пакуль мы вырашылі выкарыстоўваць часткі сінтэзатара для вырашэння сумежных задач камп'ютарнай лінгвістыкі і філалогіі.

Для стварэння арфаэпічнага слоўніка беларускай мовы быў выкарыстаны фанетычны працэсар сінтэзатара беларускага маўлення. Мы зразумелі, што калі падаць на ўваход фанетычнага працэсара электронныя масівы арфаграфічных слоў, то на выхадзе атрымаем электронныя масівы слоў у машынай транскрыпцыі. Далей машыныя коды транскрыпцый перакадоўваліся па спецыяльных табліцах, якія распрацоўвала Алена Гюнтар. Перад праграмістамі

ТОП-10 НАН БЕЛАРУСІ 2016

09.08.2017

В ПОИСКЕ КВАНТОВЫХ СОСТОЯНИЙ

09.08.2017

НОВЫЙ ТИП СВЕТОВЫХ ПОЛЕЙ

09.08.2017

ИНТЕРАКТИВНЫЕ МЕТОДЫ ДИАГНОСТИКИ

09.08.2017

ТОЧЕЧНАЯ ЗАЩИТА БИОМОЛЕКУЛ

09.08.2017

МАТЕРИАЛЫ ДЛЯ СОЛНЕЧНЫХ ЭЛЕМЕНТОВ

09.08.2017

УНИКАЛЬНЫЕ МАТЕРИАЛЫ ДЛЯ КОСМОСА

09.08.2017

НАНОГРУЗОВИК ДЛЯ ГЕНОВ

09.08.2017

ОСНОВА «ГЛЮКОСЕНА»

09.08.2017

НЕПОСЛЕДНИЙ КОРДОН

09.08.2017

«БЕЛАЯ ВЕЖА» ПРИВЕЛА К УСПЕХУ

паўстала задача, як зручна перадаваць вынікі апрацоўкі лінгвістам на праверку. Для гэтага Станіславам Лысым быў распрацаваны інтэрнэт-сэрвіс «Генератар арфаэпічнага слоўніка», які мы змясцілі на нашай бясплатнай платформе www.corpus.by.

Для нас гэтая сумесная праца важная таму, што пры абароне дысертацыі заўсёды задаюць пытанне «хто правяраў алгарытм?». Слоўнік як камертон ацэньвае нашу працу. Праверка лінгвістамі пацвердзіла на 98% правільнасць нашых алгарытмаў. Гэта нашмат больш неабходнага мінімуму. Алгарытмы паляпшаліся ад адной да другой літары ітэрацыйна.

– Супрацоўніцтва, аднак, не выглядае нечаканым у свеце тэндэнцый да сумеснай навуковай творчасці і агульнай аўтаматызацыі...

В.Р.: Здаецца, дзе слоўнікі, а дзе камп'ютарныя праграмы. Але без дапамогі калег з АІПІ мы чакалі б выхаду слоўніка яшчэ гадоў 15. Праца зусім новага фармату была падрыхтавана ўсяго за 4 гады! Ён запатрабаваны часам, бо свет убачылі новыя правілы арфаграфіі. Гэта было падмацавана выходам граматычных выданняў ад нашага інстытута. Вусная мова заставалася неахопленай.

– Акрамя папяровай версіі слоўніка транскрыпцыю слова можна паглядзець на спецыяльным электронным рэсурсе. Як працуе гэтая праграма?

С.Л.: Калі звонку, то вельмі проста. На гэта мы арыентуемся апошнім часам увогуле пры стварэнні сэрвісаў. Карыстальнік устаўляе слова ці тэкст у акенца і атрымлівае транскрыпцыю. Наша задача была сканвертаваць уяўленні машыны аб транскрыпцыі (пасля выканання алгарытмаў, распрацаваных цягам дзесяцігоддзяў) у выгляд, звычны для карыстальніка.

– На чарзе – слоўнік словаформаў усіх часцін мовы...

В.Р.: Так, мы заклалі падмурак нашага новага праекта. Будзем працаваць над агульным граматычным слоўнікам, які аб'яднае ўсе значымыя і службовыя часціны мовы. Будзе падрыхтаваны і электронны, яшчэ больш зручны, рэсурс.

Ю.Г.: Ну а пазней мы згенеруем усе арфаэпічныя нормы па 2 млн слоў. Але для гэтага нам яшчэ належыць адладзіць распазнаванне машынай марфем у адвольным слове. В.М.: Мы ўзялі пакуль толькі пачатковыя формы слоў для генеравання транскрыпцыі, так што надалей задача стварыць арфаэпічны слоўнік з вымаўленнем граматычных форм слоў, дзе адбываецца чаргаванне, асіміляцыя («плазма», «плазме»).

– З арфаграфіяй усё дакладна, а вось арфаэпічны слоўнік – гэта закон ці проста даведка?

В.М.: Безумоўна, вымаўленне – з'ява больш свабодная, чым пісьмовая мова. Накладаюць адбітак і эмоцыі, і рэгіянальныя асаблівасці. Але людзям, хто карыстаецца літаратурнай мовай, тым больш у афіцыйных зносінах, ведаць і прытрымлівацца арфаэпічных нормаў патрэбна.

Гутарыла Алена ЕРМАЛОВІЧ

Фота аўтара, «Навука»

© 2017 Навуковая, вытворча-
практычная газета Беларусі.
Все права захищены.

Адрес: г. Минск, ул. Академическая, 1, к.118, 122, 124
Телефоны: +375 17 284-16-12, +375 17 284-02-45

ГЕНШТАБ < РАЗРАБОТКА САЙТОВ
КОТОРЫЕ РАБОТАЮТ

40	121
18	48
18	17