

Тетяна Стефанович, Сергій Щербовських
Національний університет “Львівська політехніка”

РОЗРОБЛЕННЯ НАБОРУ ПРАВИЛ ДЛЯ УКРАЇНОМОВНОГО МЕТОДУ АВТОМАТИЗОВАНОГО ТВОРЕННЯ ВІДМІНКОВИХ ФОРМ ІМЕН В ЕЛЕКТРОННОМУ ДОКУМЕНТООБІГУ

© Стефанович Тетяна, Щербовських Сергій, 2015

Удосконалено метод для автоматизованого відмінювання імен, який ґрунтується на створенні набору правил з врахуванням роду. Метод реалізовано на базі Google Apps, як складову тестового порталу адміністративних послуг, запропонованого для Львівської політехніки.

Ключові слова: автоматизована обробка даних, відмінювання імен, Google Apps

The method for first name automatic declension that based on the creating of a set of rules with taking into account grammatical gender is advanced. The method is implemented based on Google Apps, as part of a test portal for administrative services that proposed for Lviv Polytechnic

Key words: automated data processing, name declination, Google Apps.

Вступ та постановка проблеми

Інформатизація вищих навчальних закладів підвищує ефективність їх функціонування, що сприяє покращенню якості підготовки фахівців. Одним із етапів інформатизації є розроблення та впровадження систем автоматизованого формування та опрацювання електронних документів. Реалізація таких систем вимагає виконання низки науково-прикладних завдань. Наприклад, у Львівській політехніці існують бази даних, які містять понад 40 тисяч облікових записів. У таких базах прізвище, ім'я та по батькові осіб подано в називному відмінку. Під час автоматизованого формування різноманітних документів виникає потреба подавати ці дані у інших відмінкових формах. Наприклад, родовий відмінок використовується у заявах, давальний — у довідках та наказах, знахідний — у поданнях тощо. Тому автоматизація словотворення прізвищ, імен та по батькові є актуальною науково-прикладною проблемою. У статті автори зосереджуються на автоматизованому словотворенні імен українського походження.

Зазначимо, що загальне вирішення поставленої проблеми є відсутнім, тобто не можна розробити універсальний метод, який охопить абсолютно всі випадки словотворення прізвищ, імен та по батькові. У такому ракурсі завдання дослідників полягає у розробленні таких підходів, які охоплюватимуть доцільну з практичного погляду кількість випадків словотворення та мінімізують робоче навантаження на операторів, які опрацьовують винятки.

Аналіз останніх досліджень і публікацій

Теоретичне дослідження лінгвістичних методів аналізу власних назв проведено у [1], однак автори не розкрили правила для методів автоматизованого творення відмінкових форм, а також із матеріалу статті не зрозуміло, чи мала місце практична реалізація запропонованих ідей у вигляді закінченого програмного продукту. Як показує аналіз програмних продуктів, представлених на ринку, реалізувати автоматизоване відмінювання прізвищ, імен та по-батькові можна методом повного перебирання та евристичним методом. Як правило, в програмних продуктах застосовується комбінація цих методів, коли відмінювання більшості прізвищ, імен та по-батькові виконується за евристичним алгоритмом, а для винятків створені бази даних, до яких програмний продукт звертається за неможливості застосування евристичних підходів. Наведені продукти [2–4] є комерційними і, як правило, російськими розробками. До їх недоліків можна зарахувати закритість коду та неврахування особливостей словотворення прізвищ, імен та по-батькові для української мови, що перешкоджає їх інтегруванню в системи електронного документообігу Львівської політехніки.

Постановка завдань дослідження

Метою дослідження є

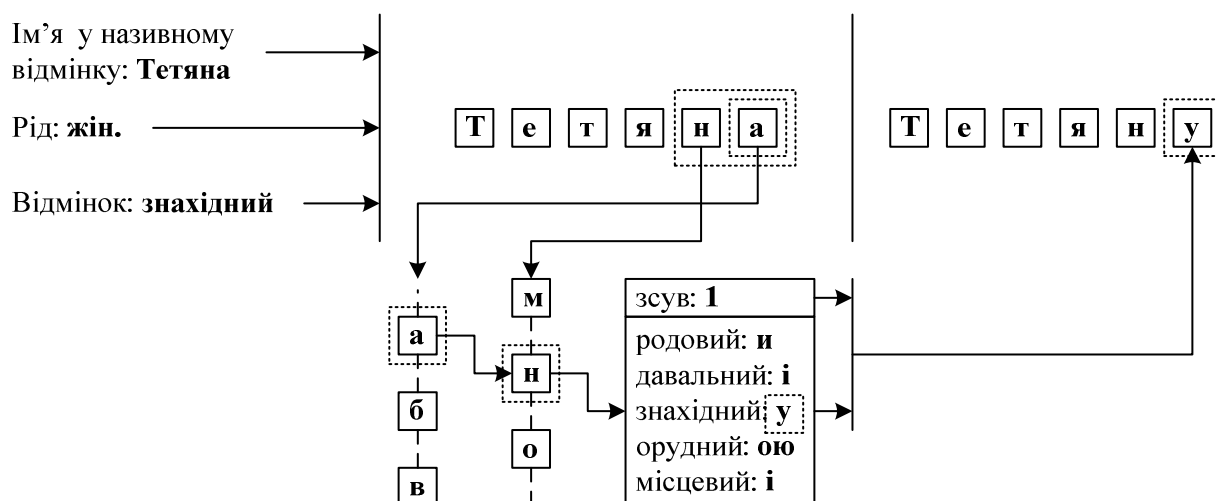
- вдосконалення теоретичних засад україномовного методу автоматизованого творення відмінкових форм імен;
- опрацювання імен та створення набору правил для їх коректного автоматизованого відмінювання;
- реалізація методу та набору правил засобами програмування.

Основна частина

Для автоматизованого відмінювання імен застосуємо метод, який виконує аналіз їх кінцевих букв. Вхідною інформацією для словотворення відмінкових форм є

- ім'я у називному відмінку,
- рід;
- потрібний відмінок.

Метод ґрунтується на тому, що виділяють останні букви імені, на підставі яких, а також на основі відомостей про рід, визначають, по-перше, скільки букв імені у формі називного відмінку необхідно відкинути і, по-друге, які букви треба додати до імені, щоб одержати його у заданій відмінковій формі. Кількість букв, яку необхідно відкинути у імені, назвемо *зсувом*. Набір букв, який необхідно додати до імені, щоб одержати задану відмінкову форму, назвемо *відмінковим закінченням*. Логічні умови, які на підставі останніх букв імені у називному відмінку та роду визначають зсув та відмінкове закінчення, називатимемо *правилами*. Покажемо суть методу відмінювання на прикладі імені «Тетяна» (див. рисунок).



Графічний опис автоматизованого методу творення відмінкових форм імен

Нехай необхідно визначити для жіночого імені у називному відмінку «Тетяна», його форму в знахідному відмінку. Така задача виникає, наприклад, під час автоматизованого формування подання на преміювання, в якому наявна фраза «преміювати Тетяну...». Визначаємо останню «а» і передостанню «н» букви імені. На основі даних про букви, а також про рід, вибираємо правило, відповідно до якого зсув становить 1 букву, а відмінкове закінчення для знахідного відмінка — «у». На підставі інформації про зсув та відмінкове закінчення виконуємо перетворення форми імені у називному відмінку в форму у знахідному: «Тетяна» (вхідні дані) — «Тетян» (зсув) — «Тетян» + «у» (вибір відмінкового закінчення) — «Тетяну» (вихідні дані).

У загальному випадку метод функціонує так. На першій ітерації виділяємо останню букву імені у називному відмінку і методом прямого перебору від «а» до «я» + «'» шукаємо правило, яке відповідає цій букві. Якщо таке правило знайдено, то виконуємо зсув та додавання відмінкового закінчення. Якщо правило відсутнє, то виділяємо передостанню букву імені і методом прямого перебору від «а» до «я» +

«'» шукаємо правило, яке відповідає цим двом буквам. Якщо таке правило знайдено, то застосовуємо його, інакше переходимо до третьої літери імені з кінця. У випадку, якщо усі букви імені перебрано, а правило не знайдено, то для такого імені вважаємо, що всі відмінкові форми збігаються із формою називного відмінку. Отже, ключовим завданням для вдосконалення цього методу є пошук та аналіз набору правил для творення відмінкових форм імен чоловічого та жіночого роду, а також формування бази винятків із правил та умов передчасного завершення пошуку.

Формування набору правил для імен виконано у два етапи. На першому етапі дослідження було сформовано базу, яка містить 1000 найуживаніших українських імен, із них 550 чоловічих та 450 жіночих. Шляхом аналізу цієї бази було встановлено закономірності їх відмінювання та сформовано набір правил для методу автоматизованого словотворення (див. таблицю).

Правила методу автоматизованого творення відмінкових форм імен

№	Рід	Кінцева буква	Передкінцева буква	Зсув	Відмін. закінчення	Приклад	
1	ч, ж	а	г	2	ги, зі, гу, гою, зі	Ольга	
2			к	2	ки, ці, ку, кою, ці	Параска	
3			х	2	хи, сі, ху, хою, сі	Домаха	
4			ж, ч, ш, щ	1	і, і, у, ею, і	Малуша	
5			й, ь, «'»	1	ї, ї, ю, єю, ї	Майя	
6			г, б, в, д, з, л, м, н, п, р, с, т, ф, ц	1	и, і, у, ою, і	Ярема Тетяна	
7		я	б, в, г, г, д, ж, з, к, л, м, н, п, р, с, т, ф, х, ц, ч, ш, щ	1	і, і, ю, ею, і	Ілля, Маруся	
8			а, е, є, и, і, ї, о, у, ю, я, й, ь, '	1	ї, ї, ю, єю, ї	Манасія Марія	
9	ч	б, в, г', г, д, з, к, л, м, н, п, р, с, т, ф, х, ц	–	0	а, у, а, ом, ові	Володимир	
10			Ігор, Лазар	0	я, ю, я, ем, еві	–	
11			Антін, Сидір, Федір тощо	2	о?а, о?у, о?а, о?ом, о?ові	–	
12		ж, ч, ш, щ	–	0	а, у, а, ем, еві	Пугач	
13			Тиміш, Тихіш тощо	2	о?а, о?у, о?а о?ем, о?еві	–	
14		ж	о	к, л, р, н, д, т	1	а, у, а, ом, ові	Петро
15			й	а, е, є, и, і, ї, о, у, ю, я	1	я, ю, я, ем, еві	Андрій
16			ь	д, т, з, с, ц, л, н	1	я, ю, я, ем, еві	Василь
17	в		О	0	і, і, –, 'ю, і	Любов	
18	ь		д, т, з, с, ц, л, н	1	і, і, –, ?ю, і	Нінель	

В одержаному наборі правил умовно виділяємо три множини. До першої множини (рядки 1–8) входять чоловічі та жіночі імена, які закінчуються на «а» та «я». Ця множина відповідає першій відміні. До другої множини (рядки 9–16) — лише чоловічі імена, які закінчуються на приголосний, на «о» та «ь». Ця множина відповідає другій відміні. І, до третьої множини (рядки 17–18) — тільки жіночі імена. Вона відповідає третій відміні. У правила внесено винятки, такі як «Ігор» та «Лазар» (рядок 10). Передбачено можливість чергування букв «і»/«о» у таких іменах як «Антін», «Сидір», «Федір», «Тиміш», «Тихіш» тощо (рядки 11 та 13). Відмінкове закінчення для такого типу імен формуємо за шаблоном, в якому символ «?» позначає останню букву імені у називному відмінку. Забезпечено подвоєння передостанньої букви в іменах типу «Нінель» в орудному відмінку (рядок 18). Одержані емпіричним шляхом правила були зіставлені із правописом [5], в результаті чого було підтверджено їх узгодженість. Слід зазначити, що правила, подані у таблиці, є чіткішими.

На другому етапі дослідження, на основі аналізу знайдених закономірностей було додано нові правила, які виходять за межі 1000 найуживаніших українських імен, проте відповідають українському словотвору. Наприклад, перелічено увесь набір шиплячих звуків «ж», «ч», «ш» та «щ» (рядки 4, 7, 12), перелічено стандартну групу «д», «т», «з», «с», «ц», «л», «н», які вживаються разом із м'яким знаком, утворено повні набори приголосних за винятком шиплячих (рядки 6, 7 та 9) та голосних (рядки 8 та 15) букв. Дискусійним питанням є формування правил, які відповідають

чоловічим іменам на «о» (рядок 9). Це правило включає вживання поширених українських імен «Павло», «Петро», українських імен, які вийшли із вжитку «Жито», «Ладо», а також імен закордонного походження «Альберто», «Давиччо», «Антоніо». Аргументовано про це питання можна буде говорити після дослідження словотвору імен іноземного походження. Крім того, встановлено, що для автоматизованого словотвору відмінкових форм українських імен достатньо обмежитись лише двома кінцевими буквами. Всі інші імена, які не підпадають під жодне із перелічених вище правил, зараховуємо до таких, що не відмінюються.

Результати дослідження впроваджені у системі автоматизованого творення відмінкових форм прізвищ, імен та по батькові, реалізованої засобами Google Apps [6]. Текст основної програми написаний мовою Google Script, а база українських імен створена засобами Google Sheets. З результатами роботи можна ознайомитися за адресою <https://sites.google.com/a/lpnu.ua/e-adm>. За цією адресою знаходиться тестова версія порталу адміністративних послуг, запропонована для Львівської політехніки. У межах цього порталу забезпечено автоматизоване формування документів, в яких прізвище, ім'я та по батькові подано у потрібному відмінку. Також сервіс відмінювання прізвищ, імен та по батькові винесено на окрему вкладку «Відмінювання ПІБ», яка знаходиться за адресою: <https://sites.google.com/a/lpnu.ua/e-adm/name>. Результат роботи сервісу показано на рис. 2.

Перелік послуг Особистий кабінет **Відмінювання ПІБ** Новини та оголошення

Прізвище	Ім'я	По батькові	Надіслати
Дубок	Іван-Петро	Андрійович	

Називний відмінок: **Дубок Іван-Петро Андрійович**
Родовий відмінок: **Дубка Івана-Петра Андрійовича**
Давальний відмінок: **Дубку Івану-Петру Андрійовичу**
Знахідний відмінок: **Дубка Івана-Петра Андрійовича**
Орудний відмінок: **Дубком Іваном-Петром Андрійовичем**
Місцевий відмінок: **Дубкові Іванові-Петрові Андрійовичу**
Коментар:

Рис. 2. Результат роботи сервісу відмінювання прізвищ, імен та по батькові, реалізованого засобами Google Apps

Висновки

За результатами дослідження розроблено набір правил для україномовного методу творення відмінкових форм імен. Удосконалено набори правил для відмінювання чоловічих та жіночих імен та встановлено не відображені в правописі закономірності їх словотворення. Україномовний метод творення відмінкових форм імен реалізований засобами Google Script як складова тестової версії порталу адміністративних послуг, запропонованого для Львівської політехніки. Подальші дослідження скеровані на розроблення набору правил для україномовного методу автоматизованого творення відмінкових форм імен іноземного походження.

Література

1. *Войтко В. В. Розробка та реалізація лінгвістичних методів аналізу власних назв / В. В. Войтко, М. П. Боцула, Г. Л. Луцишин // Оптико-електронні інформаційно-енергетичні технології. – 2008. – № 1. – С. 81–85.*
2. *Морффер / Програма склонення по падежам Сергея Слепова. – Режим доступу: <http://morpher.ru/DemoUA.aspx>.*
3. *Відмінювання ПІБ української та російської мови / NameCaseLib. – Режим доступу: <http://namecaselib.com/ua/case/>.*
4. *Склонение фамилий, имен и отчеств по падежам / Королевство Delphi. Виртуальный клуб программистов. – Режим доступу: http://www.delphikingdom.com/padeg_online.asp.*
5. *Український правопис. – К.: Наукова думка, 2015. – 288 с. – Режим доступу: <http://izbornyk.org.ua/pravopys/pravopys2015.htm>.*
6. *Жежнич П. І. Використання Google Apps в організації навчального процесу / П. І. Жежнич // Інноваційні комп'ютерні технології у вищій школі: наук.-практ. конф., 18–20 листопада 2014 р.: тези. доп. – Львів, 2014. – С. 35–38.*

RULE SET DEVELOPMENT FOR UKRAINIAN LANGUAGE METHOD OF FIRST NAME CASE FORMS AUTOMATED CREATING IN ELECTRONIC WORKFLOW

Informatization of higher education increases the efficiency of their operation that promotes teaching quality improving. One of the stages of informatization is development and implementation of a system for automated creation and processing of electronic documents. Implementation of such systems requires solving a number of scientific problems. For example, in Lviv Polytechnic National University there are databases that contain more than 40,000 accounts. In such databases last name, first name and patronymic of persons are given in the nominative case. During the automated creation of various documents there is a need to give this data in other case form. Therefore automation forming last names, first names and patronymic is actual scientific and applied problems. In this paper, the authors focus on automated forming Ukrainian origin first names. For automatic first name declension is applied a method that analyzes their last letters. Input information for case form creation is first name in the nominative case, the grammatical gender and desirable case. Based on this information, firstly, it is determined how many last letters for first name nominative case should be reject (that called shift) and, secondly, what letter should be add to first name to get it in desirable case form (that called case end). Logical conditions that based on the nominative case fist name last letters and grammatical gender determine the shift and case end we call the grammatical rules. Rule set forming for first names we made in two stages.

On the first stage of the study the base which contains 1000 the most usable Ukrainian origin names is created. The base includes 550 men and 450 women first names. By analyzing this database declension patterns and rule set for method of first name case form automated creating was set. In the obtained rule set three groups are distinguished. The first group consists of male and female names that end on “а” and “я”. This group corresponds to the first variation. The second group consists of only male first names that end on consonant, “о” and “ь”. This group corresponds to the second variation. And the third group includes only female names that correspond to the third variation. The empirical rules were matched with Ukrainian spelling. It was confirmed its consistency. It should be noted that obtained rules are more well-defined. On the second stage of the study, based on analysis of found patterns new rules were added. These rules are beyond of 1000 the most usable Ukrainian names, but they fully agreed with Ukrainian word formation. For example, all sizzling letters “ж”, “ч”, “ш” and “щ” are took into account. Or other example all letters used with soft sign such as “д”, “т”, “з”, “с”, “ц”, “л”, “н” are took into account too and so on. The results of the study are implemented in the system of automated creation of last name, first name and patronymic case form that realized by means of Google Apps. Program is written by Google Script and Ukrainian origin names base is created by Google Sheets. The results of the work are available by link <https://sites.google.com/a/lpnu.ua/e-adm>.

References

1. Voytko V. V. *Development and realization of linguist methods for proper name analysis* / V. V. Voytko, M. P. Botsula, H. L. Lutsishyn // *Optoelectronic snformation and power technologies*. – 2008. – № 1. – C. 81–85.
2. Morfer / *The program for case declension by Sergey Slepov*. – Access mode: <http://morpher.ru/DemoUA.aspx>.
3. *Last name, first name and patronymic declension in Ukrainian and Russian language / NameCaseLib*. – Access mode: <http://namecaselib.com/ua/case/>.
4. *Case declension of Last name, first name and patronymic / Delphi Kingdom. Programmer virtual club*. – Access mod: http://www.delphikingdom.com/padeg_online.asp.
5. *Ukrainian spelling*. – K.: *Scientific thought*, 2015. – 288 p. – Access mode: <http://izbornyk.org.ua/pravopys/pravopys2015.htm>.
6. Zhezhnych P. I. *Google Apps using in studying process organization / P. I. Zhezhnych // Innovation computer technologies in height education: scientific and practical conference, 18–20 November 2014 p.: proceedings*. – Lviv, 2014. – P. 35–38.