УДК: 004.8

1.8. Высокопроизводительный сетевой сервис для генерации словоформ

Гурин А.А.[1], Жуков Т.А.[1], Садыков Т.М.[1]
[1]РЭУ им. Г.В. Плеханова. Москва, Россия.

В данной статье описываются подходы и методы анализа морфологической изменчивости русского языка. Было разработано и опубликовано решение, способное производить лемматизацию русского текста, используя бессловарный подход

Введение

Одной из важнейших задач обработки естественного языка является лемматизация. Лемматизация позволяет переводить слова естественного языка в начальную форму. Это может использоваться в построении решений для анализа тональности текста, эмоциональной окрашенности, выявлению сущностей и эмоций, а также распознаванию сарказма и других задач машинной обработки текста. Для данной задачи уже существует множество способов решений. Одни выполнены в виде библиотек и словарей, которые уже содержат базу слов и все их формы, другие используют бессловарные алгоритмы видоизменения слов, существуют также генеративные модели, способные решить данную задачу.

На сегодняшний день существует множество генеративных моделей, которые способны генерировать текст и отвечать на вопросы пользователей в режиме диалога – взять, например ChatGPT. Однако у этих моделей имеются существенные недостатки. Помимо ограничений на исполнение запросов и скорости обработки, встречаются разного рода неточности в ответах и образовании словоформ.

Автоматическое изменение слов естественного языка необходимо для различных теоретических и прикладных целей, таких как синтаксический анализ, построение вопросов по темам [Chali, Hasan, 2015], распознавание и синтез речи, машинный перевод [Streiter, Iomdin, Sagalova, 2000], поиск информации [Iomdin, 2003], контент-анализ [Belonogov, Horoshilov, 2010], [Belonogov, Kotov, 1971] и генерация естественного языка [Cerutti, Toniolo, Norman, 2019], [Costa, Dolog, Ouyang, Lawlor, 2018], [Subramanian, Rajeswar, Dutil, Pal, Courville, 2017], [Tran, Nguyen, Tojo, 2017].

Различные подходы к автоматическому видоизменению использовались для решения конкретных аспектов видоизменения [Conway, 2001], [Зализняк, 1967] в заранее определенных языках [Foust, 1960], [Fuks, 2010], [Korobov, 2015], [Porter, 1980], [Raja, Rajitha, Lakshmanan, 2014] или на неопределенном языке [Faruqui, Tsvetkov, Neubig, Dyer, 2015], [Silberztein, 2016].

Несмотря на значительный недавний прогресс в работах [Buddana, Kaushik, Manogna, 2021], [d'Ascoli, Coucke, Caltagirone, Caulier, Lelarge, 2020], [Korobov, 2015], [Silberztein, 2016], [Sorokin, 2016], [Xiao, Zhu, Liu, 2013], автоматическое видоизменение и автоматическая генерация текста по-прежнему представляют собой проблему огромной вычислительной сложности для многих естественных языков мира. Большинство современных подходов используют обширные аннотированные вручную корпуса, которые в настоящее время существуют для всех основных языков [Segalovich, 2003]. Работа со словарем в реальном времени, содержащим миллионы словоформ и десятки миллионов отношений между ними, является непростой задачей [Goldsmith, 2001]. Кроме того, ни один словарь не может быть полным. По этим причинам алгоритмическое покрытие грамматики естественного языка важно при условии, что видоизменение в данном языке достаточно сложно.

Русский язык является сильно склоняемым языком, грамматика которого известна своей сложностью [Sorokin, 2016], [Зализняк, 1967]. В русском языке изменение слова может потребовать одновременного изменения его приставки, корня и окончания, а правила изменения слова очень сложны [Halle, Matushansky, 2003], [Зализняк, 1967]. Форма слова может зависеть от многих грамматических категорий, таких как число, род, лицо, время, падеж, залог, одушевленность и т.д. (см. рисунок 2). По оценке, основанной на [OpenCorpora, 2024], среднее количество различных грамматических форм – 11,716 для прилагательного, для глагола приходится в среднем 44,069 различных изменяемых форм, с учетом причастия всех видов и деепричастия (см. рисунок 2).

Видоизменение в русском языке: алгоритмы и реализация

Сетевой сервис passare.ru предоставляет высокоскоростной пакетный доступ к функциям видоизменения отдельных слов, сопоставления слов и синтеза грамматически правильного текста. В частности, реализовано склонение существительного по числу и падежу, склонение прилагательного по числу, роду и падежу, склонение наречия производится по степеням сравнения. Глагол является частью речи, видоизменение которой является самым сложным в языке. Реализованные алгоритмы обеспечивают изменение глагола по времени, лицу, числу и роду. Эти алгоритмы также позволяют образовывать деепричастие и повелительные формы глагола. Кроме того, реализованы функции образования и изменения активных причастий настоящего и прошедшего времени. Пассивное причастие настоящего времени является единственной формой глагола, которая в настоящее время не поддерживается веб-сайтом из-

за крайнего уровня неравномерной сложности. Кроме того, для многих глаголов в русском языке страдательное причастие настоящего времени вообще не может образовываться.

Алгоритмическое покрытие русского языка, предоставляемое сетевым сервисом passare.ru, направлено на баланс грамматической точности и простоты использования. По этой причине было сделано несколько упрощающих предположений: буквы «е» и «ё» считаются идентичными; никакая информация об ударении в слове не требуется для образования его изменяемых форм; для функций видоизменения наличие входного слова в язык определяется пользователем. Кроме того, одушевленность существительного не рассматривается как переменная категория в функции изменения существительного, несмотря на существование 1037 существительных (около 1,4% существительных в базе данных OpenCorpora [OpenCorpora, 2024]) с неопределенной одушевленностью. Этот список существительных был проверен вручную авторами в каждом конкретном случае, и решение было принято в пользу той формы, которая чаще встречается в языке, чем другие. Другую форму можно получить, вызвав ту же функцию с другим параметром case (Nominative или Genitive вместо Accusative).

Точно так же вид глагола не реализован как параметр в функции изменения глагола, хотя по данным [OpenCorpora, 2024] в языке 1038 глаголов (около 3,2% глаголов в базе данных), вид которых не является однозначно определенным. Для таких глаголов функция образует формы, соответствующие как совершенному, так и несовершенному виду.

Форма видоизменения слова в русском языке, определяемая выбором грамматических категорий (таких как число, род, лицо, время, падеж, залог, одушевленность и т.д.), в целом не может быть определена однозначно. Это относится, в частности, ко многим существительным женского рода, женским формам прилагательных и многочисленным глаголам. Для таких слов, алгоритмы, реализованные в сетевом сервисе passare.ru, направлены только на поиск одной из форм видоизменения, обычно той, которая наиболее распространена в языке.

Таблица 1. Сравнение ПО для видоизменения форм слова или лемматизации

Часть речи	Общее количество слов	Время вычисления в, мин:сек	Количество вычисленных форм	Обработка слова, мсек	Согласованность с OpenCorpora
Существительное	74633	2:36	12	2	98,56%
Глагол	32358	5:49	24	10	98,68%
Прилагательное	42920	0:06	28	0,14	98,49%
Наречие	1507	<00:01	2	0,021	n/a
Порядковое числительное	10000 (диапазон 0-9999)	0:30	18	3	n/a
Количественное числительное	10000 (диапазон 0-9999)	0:23	24	2	n/a
Активное причастие настоящего времени	16946	4:55	28	17	98,96%
Активное причастие прошедшего времени	32358	10:19	28	19	99,15%
Пассивное причастие прошедшего времени	32358	10:32	28	19	94,80%
Деепричастие	32358	0:23	2	0,72	99,16%
Повелительное наклонение глагола	32358	0:42	2	1	95,33%

Ввиду богатой морфологии русского языка и высокой сложности его грамматики подробное описание алгоритмов видоизменения не может быть дано в краткой исследовательской работе. Алгоритм формирования формы деепричастия совершенного вида глагола представлен на рисунке 1. Используя на входе глагол «решать», алгоритм выводит деепричастие «решав». Большая часть обозначений, представленных на рисунке 1, совпадает с синтаксисом языка программирования C#. Более того, NF

обозначает входную нормальную форму (инфинитив) глагола, подлежащего обработке. GetPerfectness() — логическая функция, которая определяет, является ли глагол совершенного вида или нет. Verb() — это функция, которая изменяет данный глагол по отношению к лицу, числу, роду и времени. BF обозначает основную форму глагола, наиболее подходящую для построения совершенного деепричастия этого глагола. Было решено использовать одну из трех различных основных форм в зависимости от типа входного глагола, который нужно изменить. В список vowels входят все гласные русского алфавита.

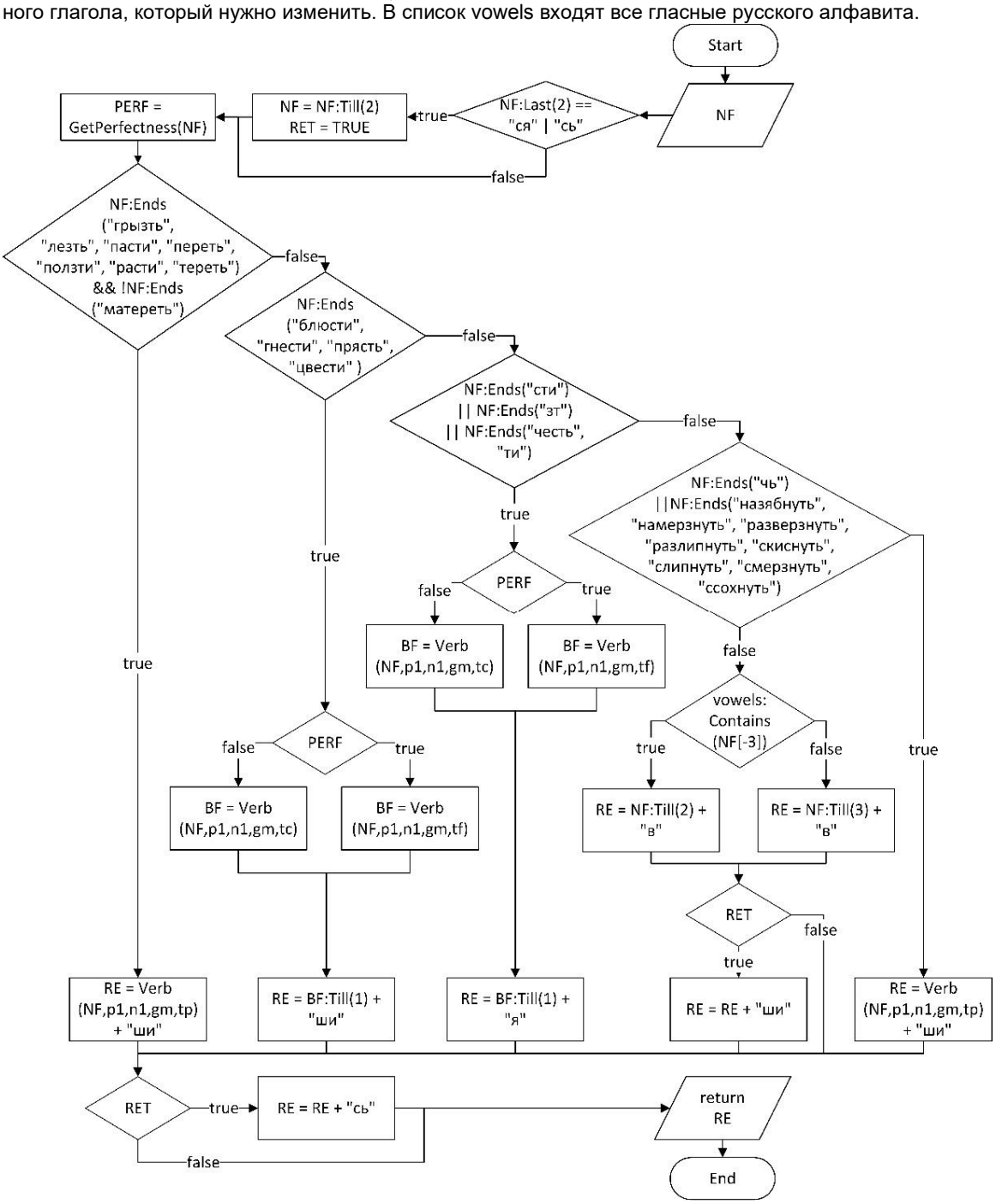

Рис. 1. Образование формы деепричастия совершенного вида глагола. Русские слова и буквы даны в общепринятой английской транслитерации. NF, BF, и RE обозначают нормальную форму глагола, основу глагола и результат вычисления, соответственно. PERF и RET являются логическими переменными, кодирующими совершенные и возвратные свойства глагола, соответственно. Список vowels содержит русские гласные. Остальные обозначения совпадают с синтаксисом языка программирования C# NF:Till(2) вместо строки NF будут удалены два последних символа. Соответствующий C# код доступен на портале https://github.com/passareru/PassareFunctions/

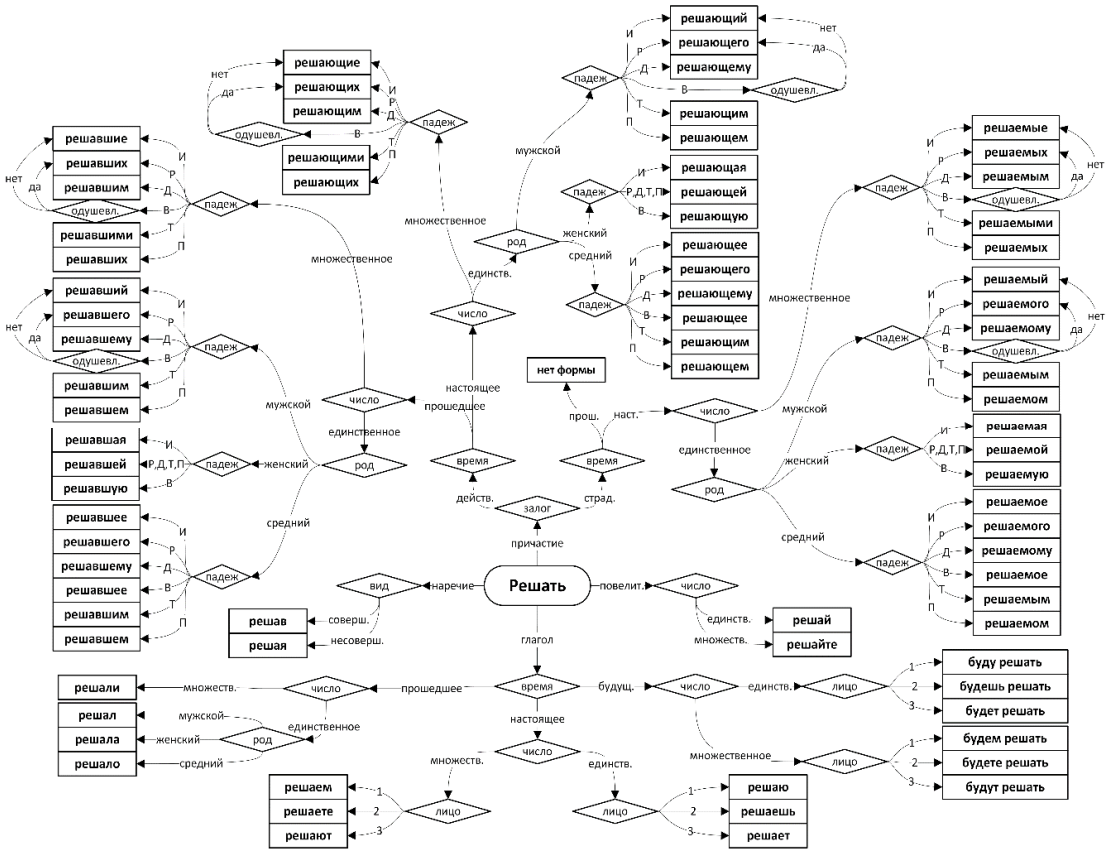

Рис. 2. Все формы глагола «решать» и их зависимость от времени (прошедшего, настоящего или будущего), числа (единственного или множественного числа), рода (женского, мужского или среднего рода), лица (1-го, 2-го или 3-го), залога (активного или пассивного), вида (совершенный или прогрессивный), падежи (именительный, родительный, дательный, винительный, творительный или предложный) и одушевленности (логическое значение).

Хотя морфология русского языка широко освещена в литературе, алгоритмы сетевого сервиса www.passare.ru в целом совершенно новые и сильно отличаются от других существующих алгоритмических подходов или хрестоматийных правил. Реализация включает около 35 000 строк кода и скомпилирована в исполняемый файл размером 571 КБ.

Тесты скорости программного обеспечения и проверка результатов

Представленное программное обеспечение протестировано на одном из крупнейших общедоступных корпусов русского языка, OpenCorpora. Был использован процессор Intel Core i5-2320 с тактовой частотой 3.00GHz с 16GB RAM под управлением Windows 10.

Учитывая все неопределенные формы слов в базе данных OpenCorpora, результат, полученный с помощью сетевого сервиса www.passare.ru, был сверен с соответствующими формами в базе данных на предмет наличия несоответствий. Результаты сведены в таблицу 1.

Все слова, измененные формы которых не показали полного соответствия базе данных OpenCorpora, были вручную проверены авторами в каждом конкретном случае. В случае с существительными 26,76% всех входных слов, допускающих ошибки, относятся к классу русских существительных, одушевленность которых не может быть определена вне контекста, например «еж» или, в зависимости от контекста, «противотанковый еж», «жучок» или, в зависимости от контекста, «скрытый микрофон» и тому подобное.

Что касается глаголов, 11,26% расхождений обусловлены глаголами, совершенство которых невозможно определить вне контекста без дополнительной информации об ударении в слове. Например: «насыпать», «пахнуть», или (в зависимости от ударения) «посмаковать» и т.д.

Кроме того, был обнаружен ряд ошибок в OpenCorpora. Классификация недостатков в OpenCorpora выходит за рамки настоящей работы, и упомянем лишь, что склонение глагола «застелить», а также деепричастные формы глаголов «выместить», «напечь» и «перекиснуть» являются неправильными в этой базе данных на момент написания.

Кроме того, некоторые деепричастия класса возвратных глаголов, по-видимому, неправильно указаны в базе данных.

Заметим, что среднее время, необходимое для образования всех изменяемых форм прилагательного, более чем в десять раз короче, чем у существительного, несмотря на то, что число форм

прилагательного больше. Этот факт отражает высокую морфологическую регулярность прилагательных в русском языке, исключительное склонение которых встречается преимущественно в классе притяжательных прилагательных, происходящих от одушевленных существительных.

Используя описанные выше основные функции, можно реализовать автоматизированный синтез грамматически правильного русского текста на основе любых логических, числовых, финансовых, фактических или любых других точных данных. На сайте passare.ru представлены примеры таких метафункций, которые генерируют грамматически правильный прогноз погоды и отчеты о курсах валют на основе данных, доступных в режиме реального времени, доступных в Интернете. Кроме того, сайт предлагает функцию, преобразующую правильную арифметическую формулу в русский текст.

Сопоставление прилагательных с существительными по роду и числу, сопоставление глаголов с личными местоимениями по лицу, роду и числу, а также многочисленные подобные функции реализованы в разделе синтеза сайта. Эти функции также можно использовать для приведения компонентов предложения в грамматически правильную форму.

Количественный корпус анализа морфологической сложности русского языка

Были использованы алгоритмы, реализованные в сетевом сервисе www.passare.ru, для анализа сложности словоизменения разных частей речи в русском языке. В этом отношении интерес представляют только три части речи: прилагательные, существительные и глаголы (вместе с причастиями всех видов). Все остальные части речи в русском языке либо содержат очень ограниченное количество слов и их форм (например, личные и притяжательные местоимения, союзы, междометия и т. д.), либо имеют весьма регулярное склонение (например, наречия). Ни одна из этих частей речи не представляет интереса с точки зрения алгоритмического словоизменения, поскольку их неправильные формы видоизменения очень немногочисленны и их легко перечислить. Напротив, в русском языке словоизменение прилагательных, существительных и глаголов весьма сложное и часто нерегулярное (глаголы представлены на рисунке 2).

Для измерения морфологической изменчивости слова w введем функцию

$$\mathcal{L}(\mathrm{w}) := \sum_{\mathrm{i,j}} \mathrm{dist_L}(\mathrm{w_i}, \mathrm{w_j}), \qquad (1)$$

где w_i является списком всех форм слова w (с фиксированным порядком значений грамматических параметров, кодирующих эти формы) и $\mathrm{dist_L}$ расстояние Левенштейна [Levenshtein, 1966] между формами w_i и w_j. Например, для глагола *w:= "решать"*, список w_i его форм включает в себя 78 форм, представленных на рисунке 2.

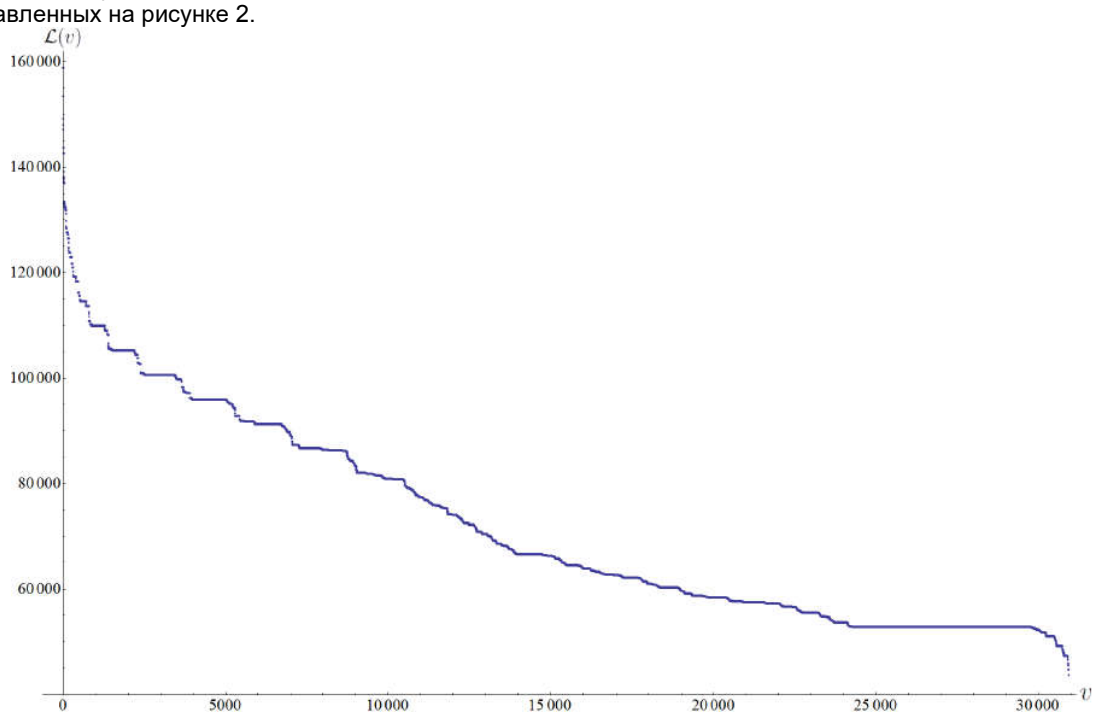

Рис. 3. Морфологическая изменчивость глаголов, глаголы сортированы по значениям $\mathcal{L}(v)$, общее расстояние Левенштейна (1) между спрягаемыми формами глагола *v*.

Глаголы. Глаголы обладают наибольшей морфологической изменчивостью среди всех частей речи в русском языке (рисунок 2). Алгоритмы изменения глаголов и образования различных глагольных форм (причастий и деепричастий) относятся к числу наиболее сложных в русской грамматике. Рисунок 3 отражает морфологическую изменчивость глаголов в русском языке. Горизонтальные оси

соответствуют 32358 русским глаголам, перечисленным в базе данных OpenCorpora. Высота $\mathcal{L}(v)$ вертикального отрезка, соответствующего глаголу v, рассчитана по формуле (1). Формы глагола рассчитаны с помощью алгоритмов видоизменения, реализованных на сайте www.passare.ru.

Прилагательные. Прилагательные — это часть речи, имеющая наиболее регулярное склонение в русском языке. (Здесь не были учтены малословные части речи, такие как личные местоимения, междометия и т.п.) Тем не менее алгоритмическое склонение русских прилагательных представляет собой задачу существенной вычислительной сложности.

Существительные. В русском языке существительные имеют среднюю сложность видоизменения по сравнению с прилагательными и глаголами. Несмотря на подавляющее большинство обычных случаев, существует множество исключений, к которым относятся, например, несклоняемые существительные иностранного происхождения.

Аналогичное исследование было проведено и для других частей речи в русском языке, что привело к ряду усовершенствований алгоритмов видоизменения.

Открытые вопросы

Существует несколько других подходов к автоматическому видоизменению русского языка и синтезу грамматически правильного текста, например [Kanovich, Shalyapina, 1994], [Korobov, 2015]. Кроме того, во многих программах предпринимаются попытки автоматического изменения определенной части речи или синтеза документа жесткой предопределенной структуры.

Таблица 2. Сравнение ПО для видоизменения форм слова или лемматизации

Программная среда	Функциональность	Поддерживаемые языки	Зависимость от словаря	Распостраняется как	Реализация
passare.ru	видоизменение, подбор слов, данные в текст	русский	слабая	Free web service	алгоритм извлечения из языка
morpher.ru	Видоизменение (Существительные, Числительные), простые предложения подбор	русский, украинский	высокая	коммерческий веб сервис / автономные библиотеки	поиск по словарю
phpmorphy	Морфологический анализ, лемматизация, видоизменение	английский, русский, немецкий, украинский, эстонский и другие	высокая	библиотека (php)	поиск по словарю
pymorphy2	морфологический анализ, лемматизация, видоизменение	русский, украинский	высокая	библиотека (python)	поиск по словарю
NooJ	развитие грамматики среда, лингвистический анализ	произвольный	высокая	фреймворк	грамматика базовое производство
MARu	морфологический анализ, лемматизация (использование pymorphy2)	русский	высокая, через pymorphy2 лемматизация	библиотека (python)	различные методы машинного обучения: линейная модель, МСП, глубинные нейронные сети
natasha	сегментация, вложения, морфология, лемматизация, синтаксис, NER, извлечение фактов	русский	зависимость в обучении данных, зависимость в обученных моделях	несколько библиотек (python)	razdel и yargy системы на основе правил; navec и slovnet нейронные сети

Судя по общедоступной информации, большинство таких программ широко используют аннотированные вручную корпуса, что может привести к сбою, если слово, которое нужно изменить, достаточно отличается от элементов в базе данных. Результаты сравнения подхода, представленного в настоящей статье, с другими программными средами, предлагающими функциональные возможности русской модуляции или лемматизации, обобщены в таблице 2.

Данное исследование выполнено в рамках государственного задания в сфере научной деятельности Министерства науки и высшего образования РФ на тему «Модели, методы и алгоритмы искусственного интеллекта в задачах экономики для анализа и стилизации многомерных данных, прогнозирования временных рядов и проектирования рекомендательных систем», номер проекта FSSW-2023-0004.

Литература

1. Белоногов Г.Г., Котов Р.Г. Автоматизированные информационно-поисковые системы. — М.: Советское радио, 1968. 184 с.
2. Зализняк А. А., «Русское именное словоизменение» с приложением избранных работ по современному русскому языку и общему языкознанию. - М: Наука, 1967. - 752 с.
3. Belonogov, G., Horoshilov, A., Horoshilov, A.: Automation of the English-Russian bilingual phraseological dictionaries based on arrays of bilingual texts. Automatic Documentation and Mathematical Linguistics 44(3), 103–110 (2010).
4. Buddana, H., Kaushik, S., Manogna, P., P.s., S.: Word level lstm and recurrent neural network for automatic text generation. 2021 International Conference on Computer Communication and Informatics, ICCCI 2021 (2021).
5. Cerutti, F., Toniolo, A., Norman, T.: On natural language generation of formal argumentation. CEUR Workshop Proceedings 2528, 15–29 (2019).
6. Chali, Y., Hasan, S.: Towards topic-to-question generation. Computational Linguistics 41(1), 1–20 (2015).
7. Chernikov, B., Karminsky, A.: Specificities of lexicological synthesis of text documents. Procedia Computer Science 31, 431–439 (2014).
8. Conway, D.: An algorithmic approach to English pluralization. In: Second Annual Perl Conference. COPE (2001).
9. Costa, F., Dolog, P., Ouyang, S., Lawlor, A.: Automatic generation of natural language explanations. International Conference on Intelligent User Interfaces, Proceedings IUI (2018).
10. d'Ascoli, S., Coucke, A., Caltagirone, F., Caulier, A., Lelarge, M.: Conditioned text generation with transfer for closed-domain dialogue systems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).
11. Faruqui, M., Tsvetkov, Y., Neubig, G., Dyer, C.: Morphological inflection generation using character sequence to sequence learning. CoRR abs/1512.06110 (2015).
12. Foust, W.: Automatic English inflection. In: National Symposium on Machine Translation. pp. 229–233. UCLA (1960).
13. Fuks, H.: Inflection system of a language as a complex network. CoRR abs/1007.1025 (2010).
14. Goldsmith, J.: Unsupervised learning of the morphology of a natural language. Computational Linguistics 27(2), 153–198 (2001).
15. Halle, M., Matushansky, O.: The morphophonology of Russian adjectival inflection. Linguistic Inquiry 37(3), 351–404 (2006).
16. Iomdin, L.: Natural language processing as a source of linguistic knowledge. pp. 68–74 (2003).
17. Kanovich, M., Shalyapina, Z.: The RUMORS system of Russian synthesis. COLING pp. 177–179 (1994).
18. Korobov, M.: Morphological analyzer and generator for Russian and Ukrainian languages. Communications in Computer and Information Science 542, 330–342 (2015).
19. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady 10(8), 707–710 (feb 1966).
20. OpenCorpora: An open corpus of Russian language, http://www.opencorpora.org/. (2024)
21. Porter, M.: An algorithm for suffix stripping. Program 14(3), 130–137 (1980).
22. Raja, S., Rajitha, V., Lakshmanan, M.: Computational model to generate case-inflected forms of masculine nouns for word search in Sanskrit e-text. J. Comput. Sci. 10(11), 2260–2268 (2014).
23. Segalovich, I.: A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. Proceedings of the International Conference on Machine Learning; Models, Technologies and Applications pp. 273–280 (2003).
24. Silberztein, M.: Formalizing Natural Languages: The NooJ Approach. John Wiley and Sons Limited (2016).
25. Sorokin, A.: Using longest common subsequence and character models to predict word forms. Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, SIGMORPHON 2016 at the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 pp. 54–61 (2016).
26. Streiter, O., Iomdin, L., Sagalova, I.: Learning lessons from bilingual corpora: Benefits for machine translation. International Journal of Corpus Linguistics 5(2), 199–230 (2000).
27. Subramanian, S., Rajeswar, S., Dutil, F., Pal, C., Courville, A.: Adversarial generation of natural language. Proceedings of the 2nd Workshop on Representation Learning for NLP, Rep4NLP 2017 at the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017 pp. 241–251 (2017).
28. Tran, V.K., Nguyen, L.M., Tojo, S.: Neural-based natural language generation in dialogue using rnn encoder-decoder with semantic aggregation. SIGDIAL 2017 - 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference pp. 231–240 (2017).
29. Xiao, T., Zhu, J., Liu, T.: Bagging and boosting statistical machine translation systems. Artificial Intelligence 195, 496–527 (2013).

References in Cyrillics

1. Belonogov G.G., Kotov R.G. Avtomatizirovanny`e informacionno-poiskovy`e sistemy`. — M.: Sovetskoe radio, 1968. 184 s.
2. Zaliznyak A. A., «Russkoe imennoe slovoizmenenie» s prilozheniem izbranny`x rabot po so-vremennomu russkomu yazy`ku i obshhemu yazy`koznaniyu. - M: Nauka, 1967. - 752 s.
3. Belonogov, G., Horoshilov, A., Horoshilov, A.: Automation of the English-Russian bilingual phraseological dictionaries based on arrays of bilingual texts. Automatic Documentation and Mathematical Linguistics 44(3), 103–110 (2010).

Гурин Анатолий Анатольевич,
лаборант-исследователь учебно-научной лаборатории искусственного интеллекта, нейротехнологий и бизнес аналитики РЭУ им. Г.В. Плеханова,
Gurin.AA@rea.ru

Жуков Тимур Алекперович,
лаборант-исследователь учебно-научной лаборатории искусственного интеллекта, нейротехнологий и бизнес аналитики РЭУ им. Г.В. Плеханова,
Zhukov.TA@rea.ru

Садыков Тимур Мрадович,
профессор кафедры информатики РЭУ им. Г.В. Плеханова,
Sadykov.TM@rea.ru

Ключевые слова
Обработка естественного языка, автоматическая генерация текста, алгоритмы видоизменения слов Русского языка, морфологическая изменчивость.

Anatoly Gurin, Timur Zhukov, Timur Sadykov, High-performance network service for generating word forms

Keywords
Natural language generation, automatic text synthesis, algorithmic inflection of Russian, morphological variability

DOI: 10.33276/DE-2024-02-08
JEL classification: M15 Управление информационными технологиями

Abstract
We present a set of deterministic algorithms for Russian inflection and automated text synthesis.These algorithms are implemented in a publicly available web-service www.passare.ru. This service provides functions for inflection of single words, word matching and synthesis of grammatically correct Russian text. Selected code and datasets are available at https://github.com/passare-ru/PassareFunctions/

Performance of the inflectional functions has been tested against the annotated corpus of Russian language OpenCorpora, compared with that of other solutions, and used for estimating the morphological variability and complexity of different parts of speech in Russian.