

А.Л. СТАНОВСКИЙ, д-р техн. наук, **Ю.С. БАРЧАНОВА**,
Ю.В. ШИХИРЕВА, Одесса, Украина

ФОРМАЛИЗАЦИЯ НЕЧЕТКИХ ПОНЯТИЙ ДЛЯ СИСТЕМ ОБРАБОТКИ СТИХОТВОРНОГО ТЕКСТА

Запропоновано деякі концепції комп'ютерної підтримки автоматизованого обробника віршованого тексту. Формалізовано основні атрибути віршованого тексту – риму та ритм. Розроблено алгоритм автоматизованого підбирання слів, що римуються. Запропонована електронна база даних (словник) мови, в якій впорядкованість тай добір слів виконуються, починаючи з останньої букви слова.

Предложены некоторые концепции компьютерной поддержки автоматизированного обработчика стихотворного текста. Формализованы основные атрибуты стихотворного текста – рифма и ритм. Разработан алгоритм автоматизированной подборки слов, которые рифмуются. Предложена электронная база данных (словарь) языка, в которой упорядочивание и подбор слов выполняются, начиная с последней буквы слова.

A.L. STANOVSKY, YU.S. BARCHANOVA, YU.V. SHICHIREVA

FORMALIZATION OF INDISTINCT CONCEPTS FOR SYSTEMS OF THE POETIC TEXT PROCESSING

Some concepts of computer support of the automated translator of the poetic text are offered. The basic attributes of the poetic text – a rhyme and a rhythm are formalized. The algorithm of the automated selection of rhymed words is developed. The electronic database (dictionary) of language in which ordering and selection of words are carried out, since last letter of a word is offered.

Стихотворный текст отличают, как правило, рифма, ритм, смысл, образность и новизна. Отметим, что при обработке такого текста (создании, переводе и т.п.) «обслуживание» последних трех атрибутов остаются за автором – человеком, и только первые два могут быть в необходимой степени формализованы. Для этого необходимо разработать новую, концептуальную, восходящую исключительно к проблемам информатики, основу системы обработки, так как, например, обычные переводчики оставляют «за кадром» те особенности текста, без которых не обходится стихотворная речь [1, 2].

Поэтому целью настоящей работы являлась разработка *информационной основы* системы обработки стихотворного текста (СОСТ), позволяющей в автоматизированном режиме подбирать необходимые элементы (слова, словосочетания).

Основной задачей в рамках этой цели явилась попытка формализации важнейшего атрибута стихотворного текста, которым в большинстве случаев является *рифма*. К сожалению, это понятие недостаточно определено в области информатики и не имеет в ней четкого смысла [3].

Для достижения цели художественное стихотворное произведение рассматривается с двух сторон – творческой и технической (информационной). Хотя такое разделение в значительной степени условно, оно приводит к неплохим результатам в качестве поддержки творчества в таких сугубо «человеческих» областях деятельности, как сочинение музыки, игра в шахматы и многое другое.

Рифма (греч. «*rhythmos*» – стройность, соразмерность) – созвучие концов стихотворных строк, создающее ощущение их единства и родства; регулярный звуковой повтор, несущий организационную функцию в метрической, строфической и смысловой композиции стихотворного произведения; один из основных элементов художественного стиля в системе выразительных средств поэзии; важнейший приём инструментовки стиха.

Первое слово в рифме – пассивное, второе – активное. Пассивное слово, идущее во временном отрезке первым, вносит в стих момент незавершённости, оставляя после себя напряжение, связанное с ожиданием гармоничной развязки. Это напряжение снимает активное слово, которое своим появлением заставляет вспомнить пассивное, и уже вместе они создают то ощущение законченности, гармонии и родства, которое и называется рифмой.

Таким образом, вся нагрузка в рифме лежит именно на втором слове, которое может оправдать (точная рифма), частично оправдать (приблизительная рифма) или не оправдать вовсе (диссонанс, белый стих) ожидание гармоничной развязки, созданное пассивным словом. Оправданное ожидание всегда усиливает эмоциональное воздействие стиха и заложено в самой природе поэзии.

Рифмы в русском языке классифицируются по происхождению, объёму (количеству слогов), звучанию и родственным отношениям, ощущению, положению в стихе и в строфе, а также по дополнительным признакам [4]. Кроме того, выделяют 103 вида рифмы, а также множество ее типов и родов (рис. 1) [5].

Перечисленные *качественные* характеристики рифмы не позволяют к сожалению сравнивать рифмы *количественно*. Поэтому была предложена классификация, направленная на использование в информационных системах.

В основу формализации понятия «рифма» положены следующие концепции (примеры приведены из стихотворений одного из авторов статьи).

В рифме участвуют два слова. СОСТ вычисляет коэффициент близости этих слов P – комплексное число, состоящее из двух частей – четкой и нечеткой.

Четкая часть C – суть *количественная* оценка меры полного совпадения букв из 42-буквенного алфавита (33 основных буквы + 9 ударных гласных) в словах рифмы на соответствующих местах, вычисленная по формуле:

$$C = \sum_{i=1}^L w_i p_i, \quad (1)$$

где i – номер позиции сравниваемой буквы, отсчитанный с конца слова; L – длина рифмы; p – наличие совпадения (1 – есть; 0 – нет); w_i – весовые коэффициенты, возрастающие по мере роста i .

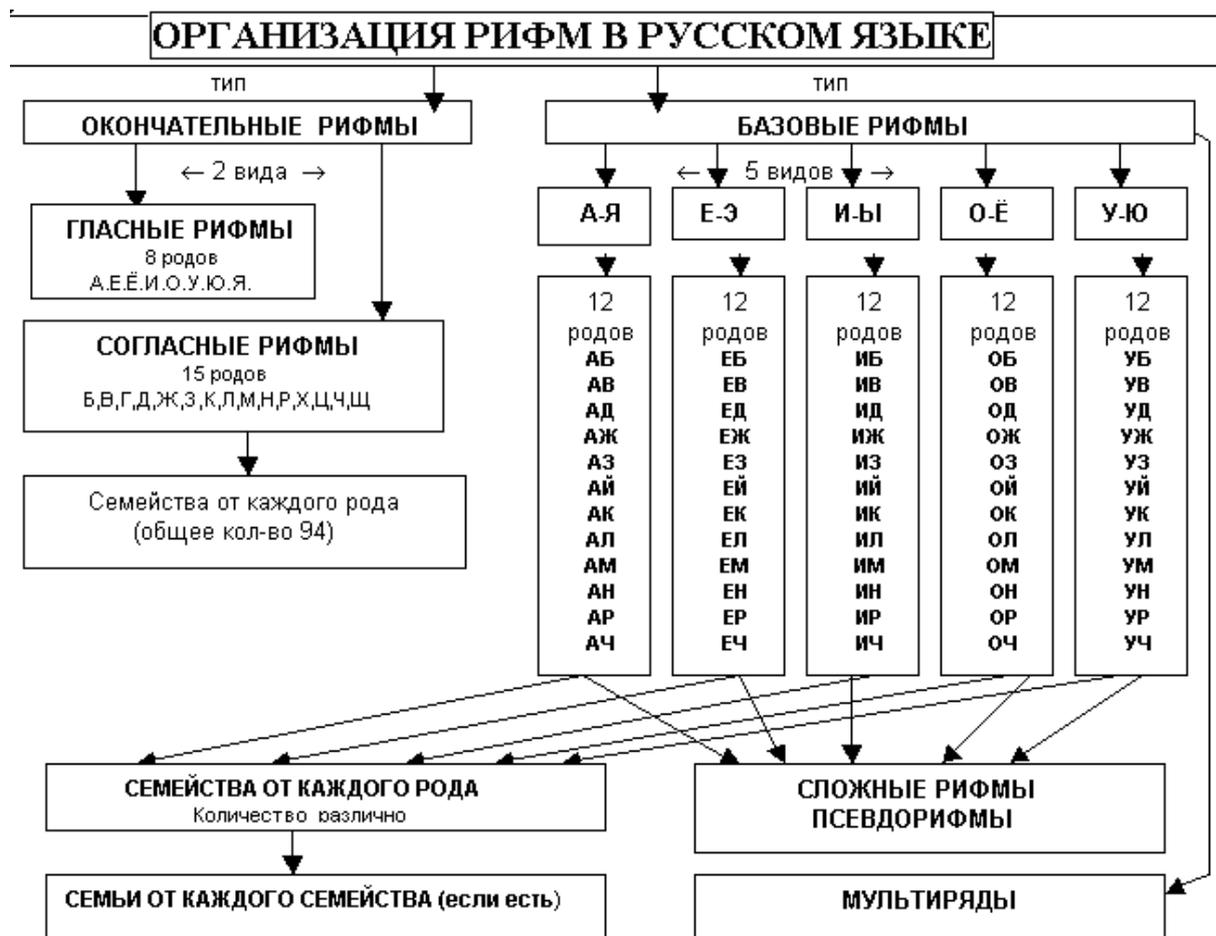


Рисунок 1 – Классификация рифм в русском языке по типам и родам

Например, рифма «художник – придорожных» при длине в шесть букв ($L = 6$) содержит совпадение букв «ожн», стоящих на позициях 3 – 5 с конца. Именно они участвуют в расчетах четкой части силы рифмы. Остальные буквы рифмы, стоящие на позициях 1, 2 и 6, не вносят вклад в величину \mathcal{C} ($w_1 = w_2 = w_6 = 0$) принимаются во внимание при расчете нечеткой части.

Нечеткая часть H – суть количественная оценка с помощью теории нечетких отношений меры близости несовпадающих букв из 42-буквенного алфавита в словах рифмы на соответствующих местах [7].

Коэффициент близости P , как и обычное комплексное число, может рассматриваться как векторная сумма \mathbf{P} четкого и нечеткого векторов $\mathbf{P} = \mathbf{C} + \mathbf{H}$ и рассчитываться по правилу скаляра:

$$P = |\mathbf{C}| \cdot |\mathbf{H}| \cdot \cos \angle \mathbf{C}, \hat{\mathbf{H}} . \quad (2)$$

К настройкам пользователя (НП) относится длина рифмы L .

Длина рифмы может быть определена «вручную» – прямым усмотрением (например, длина рифмы «и Тань я – скитанья» может быть оценена специалистом в шесть букв) или вычислена автоматически с учетом объединения и опускания согласных одним из следующих методов (рис. 2).

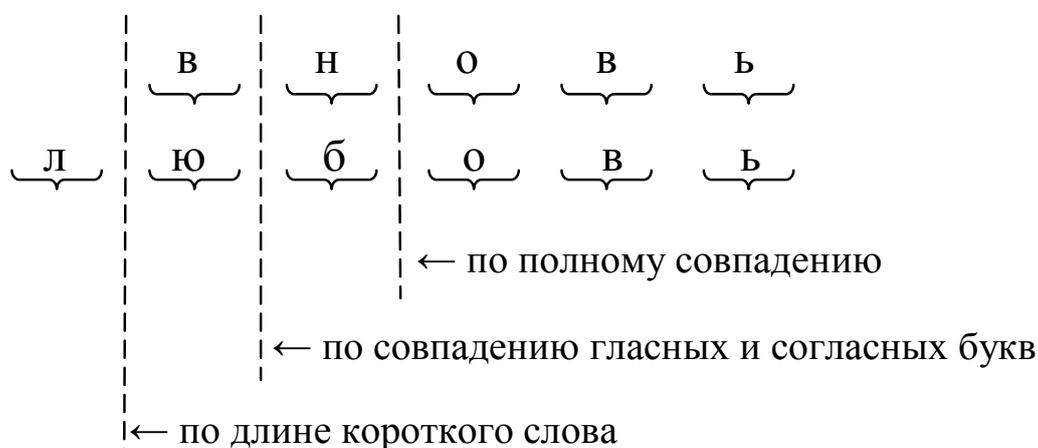


Рисунок 2 – Схема, иллюстрирующая методы определения длины рифмы

В соответствии с рис. 2 длина рифмы «вновь – любовь» равна трем буквам при использовании метода полного совпадения, четырем – метода совпадения гласных и согласных букв и пяти – метода определения по длине короткого слова.

В базе данных СОСТ хранятся:

– слова русского языка во всех склонениях, спряжениях и числах, составленные в 42-буквенном алфавите; система управления этой базой данных, была построена таким образом, чтобы она позволяла пользователю

подбирать слова не по первым буквам алфавита, как в обычных словарях, а по последним – необходимый атрибут при подборе соответствующих рифм;

– известные (применявшиеся в опубликованных стихотворениях) рифмы с указанием их длины, силы, новизны и неловкости.

Блок-схема алгоритма поиска второго слова рифмы из M слов в БД представлена на рис. 3.

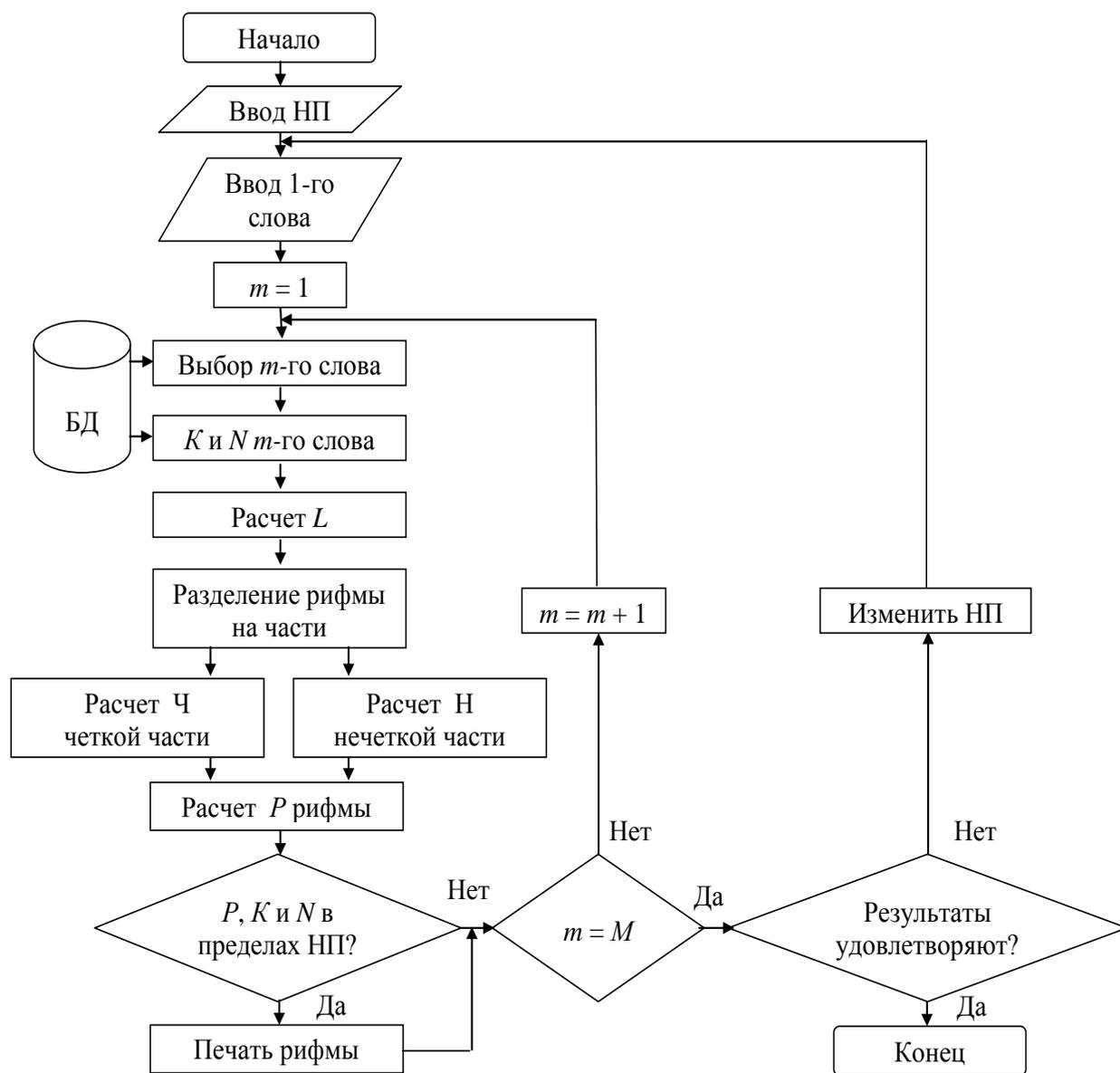


Рисунок 3 – Блок-схема алгоритма подбора вторых слов рифмы в пределах настроек пользователя

Практическое использование созданной СОСТ показало хорошие результаты, позволившие рекомендовать систему для применения ее при

«ручном» и полуавтоматическом (с предварительным использованием подстрочника) переводе стихотворений, а также для подбора рифм при стихосложении.

Список использованных источников: 1. *Шаляпина З.М.* Автоматический перевод: эволюция и современные тенденции // Вопросы языкознания. – 1996. – № 2. – С. 105 – 117. 2. *Лозинский М.А.* Искусство стихотворного перевода // Перевод – средство взаимного сближения народов. Сборник статей. – М.: Прогресс, 1987. – С. 23 – 33. 3. *Становский А.Л.* Концептуальная основа автоматизированного переводчика стихотворного текста / *А.Л. Становский, Е.А. Науменко, А.Н. Панченко* // Материалы XIV конференции «Моделирование в прикладных научных исследованиях». – Одесса: ОНПУ, 2007. – С. 78. 4. *Онуфриев В.В.* Словарь разновидностей рифмы. – Русские рифмы. – С. 1 [электронный ресурс]. – Режим доступа: <http://rifma.com.ru/Theory-0.htm>. 5. *Онуфриев В.В.* Словарь разновидностей рифмы. – Русские рифмы. – С. 9 [электронный ресурс]. – Режим доступа: <http://rifma.com.ru/SLRZ-09.htm>. 6. *Раскин Л.Г.* Нечеткая математика. Основы теории. Приложения / *Л.Г. Раскин, О.В. Серая*. – Харьков: Парус, 2008. – 352 с. 7. *Клименко И.* Сколько слов в русском языке. – Русская грамота [электронный ресурс]. – Режим доступа: http://www.igroslov.ru/index.php?option=com_content&task=view&id=37&Itemid=25.

Bibliography (transliterated): 1. Shal'japina Z.M. Avtomaticheskij perevod: jevoljucija i sovremennye tendencii // Voprosy jazykoznanija. – 1996. – № 2. – S. 105 – 117. 2. Lozinskij M.A. Iskusstvo stihotvornogo perevoda // Perevod – sredstvo vzaimnogo sbližhenija narodov. Sbornik statej. – M.: Progress, 1987. – S. 23 – 33. 3. Stanovskij A.L. Konceptual'naja osnova avtomatizirovannogo perevodchika stihotvornogo teksta / A.L. Stanovskij, E.A. Naumenko, A.N. Panchenko // Materialy XIV konferencii «Modelirovanie v prikladnyh nauchnyh issledovanijah». – Odessa: ONPU, 2007. – S. 78. 4. Onufriev V.V. Slovar' raznovidnostej rifmy. – Russkie rifmy. – S. 1 [jelektronnyj resurs]. – Rezhim dostupa: <http://rifma.com.ru/Theory-0.htm>. 5. Onufriev V.V. Slovar' raznovidnostej rifmy. – Russkie rifmy. – S. 9 [jelektronnyj resurs]. – Rezhim dostupa: <http://rifma.com.ru/SLRZ-09.htm>. 6. Raskin L.G. Nechetkaja matematika. Osnovy teo-rii. Prilozhenija / L.G. Raskin, O.V. Seraja. – Har'kov: Parus, 2008. – 352 s. 7. Klīmenko I. Skol'ko slov v russkom jazyke. – Russkaja gramota [jelektronnyj resurs]. – Rezhim dostupa: http://www.igroslov.ru/index.php?option=com_content&task=view&id=37&Itemid=25.