

УДК 004.912

DOI 10.24412/2413-7383-2024-3-45-54

С. А. Большакова

Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», г. Донецк,
283048, г. Донецк, ул. Артема, 118 б

СИСТЕМА АВТОМАТИЧЕСКОЙ АДАПТАЦИИ РУССКОЯЗЫЧНЫХ ТЕКСТОВ И ЕЕ ПРАКТИЧЕСКАЯ ЗНАЧИМОСТЬ

S. A. Bolshakova

Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»
283048, Donetsk, Artema str, 118-b

THE SYSTEM OF AUTOMATIC ADAPTATION OF RUSSIAN-LANGUAGE TEXTS AND ITS PRACTICAL SIGNIFICANCE

В статье описана система адаптации, использующая способ лексического упрощения русскоязычного текста с помощью специально размеченной базы синонимов и набора правил, которые позволяют автоматически осуществлять лексическую замену слов и словосочетаний с сохранением семантики текста и восстановлением правильного синтаксиса. Для формирования размеченной базы синонимов использовались словари, находящиеся в открытом доступе. Предложенные методы и алгоритмы, а также база синонимов могут быть использованы для задач адаптации, поисковой оптимизации и автоматического реферирования текстов, а также при автоматическом переводе.

Ключевые слова: обработка естественного языка, автоматический анализ текста, синонимия, автоматическая адаптация.

The article describes an adaptation system using a method of lexical simplification of the Russian-language text using a specially marked database of synonyms and a set of rules that allow automatic lexical replacement of words and phrases while preserving the semantics of the text and restoring the correct syntax. Dictionaries that are in the public domain were used to form a marked-up database of synonyms. The proposed methods and algorithms, as well as a database of synonyms, can be used for adaptation tasks, search engine optimization and automatic text abstraction, as well as for automatic translation.

Keywords: natural language processing, automatic text analysis, synonymy, automatic adaptation.

Введение

В настоящее время существует довольно много программных средств перефразирования текста, в частности использующих нейросети. Они предназначены для преодоления программ антиплагиата и увеличения оригинальности текста [1]. В то же время одной из актуальных и общественно-значимых задач в области обработки естественного языка является преобразование сложных текстов в более простые и понятные. Процесс изменения текста с целью упрощения сложных языковых конструкций известен как адаптация или симплификация.

Эта проблема особенно важна для людей, уровень владения языком которых не позволяет полностью усвоить сложную текстовую информацию, включая иностранцев, находящихся на стадии изучения языка [2], людей с начальными признаками когнитивных расстройств из-за возрастных изменений или черепно-мозговых травм, а также детей с нарушениями речевого развития.

Автоматизированные средства адаптации могут также быть использованы в качестве компонентов в приложениях по обработке естественного языка: системах машинного перевода [3], поисковых системах и приложениях для автоматического реферирования текстов.

Эффективным методом в этом отношении может быть использование синонимии, так как один и тот же смысл может быть передан различными словами и словосочетаниями, среди которых можно найти наиболее простую форму выражения.

В широком понимании синонимы — это слова, которые имеют одинаковое или близкое значение. Синонимы могут быть представлены как одной частью речи, так и разными. Также в качестве синонимов могут выступать одно или несколько слов. Например, «атака» и «наступление», «большой» и «огромный», «рассердиться» и «дойти до белого каления», «идеализировать» и «смотреть сквозь розовые очки» [4]. В русском языке синонимы зачастую обладают различными морфологическими характеристиками, что создает трудности при автоматической замене, связанные с соблюдением правил синтаксиса в адаптированном тексте.

Описание системы автоматической адаптации

В данной работе предлагается система обработки и анализа текстовой информации использующая метод лексической адаптации текста с помощью синонимических замен. Для автоматического упрощения текста используется метод обратного использования синонимических рядов. Был проанализирован ряд словарей синонимов, из которых для данной работы был выбран словарь З.Е. Александровой [5]. Он представляет собой набор синонимических рядов, каждый из которых представлен доминантой и последующими синонимами, среди которых могут быть как отдельные слова, так и синонимичные словосочетания. Схематично смысл предлагаемого метода состоит в замене встретившегося в тексте члена синонимического ряда соответствующей доминантой, как наиболее общим и стилистически нейтральным представителем ряда. Такую замену нельзя делать абсолютно формально, ибо это может привести к потере смысла, т.к. не все члены синонимического ряда могут быть заменены на доминанту. Например: *телевизор! телик, ящик, телевизионный ящик*. Нельзя каждый раз «ящик» заменять на «телевизор», т.к. у этого слова более широкое значение. В связи с этим был проведен анализ и сокращение предлагаемых синонимических рядов.

Кроме того, в русском языке синонимы обычно обладают различными морфологическими характеристиками, что может создавать сложности при автоматической замене, связанные с сохранением правильного синтаксиса в упрощенном тексте [6].

Для более корректной работы системы адаптации все скорректированные синонимические ряды были разделены на категории, и создана база, состоящая из 3-х частей: База слов, База словосочетаний и База неизменяемых словосочетаний. Работа с каждой категорией имеет свои особенности. Так База слов и База словосочетаний содержит только леммы (начальные формы) каждого члена синонимического ряда, а в Базе неизменяемых словосочетаний лемматизация не применяется, т.к. она по большей части содержит синонимические эквиваленты наречий.

Метод синонимических замен опирается на базу продукционных правил, необходимых для соблюдения синтаксиса при заменах. Ниже приведены две базы продукций: синонимической замены отдельных слов и словосочетаний.

База правил для восстановления синтаксиса

При синонимических заменах отдельных слов используется База слов. Система адаптации находит в тексте член синонимического ряда и заменяет его соответствующей доминантой, восстанавливая при этом в предложении–результате правильный синтаксис. Рассмотрим более подробно первоначальные правила порождения нужной формы для слова-замены при замене одиночных слов.

Правила для слов-замен, которые являются **существительными (или местоимениями-существительными)**:

– если слово-замена и заменяемое слово является существительным (местоимением-существительным), то грамматическая форма слова-замены согласуется с заменяемым словом («мы столкнулись с *чинушей*» → «мы столкнулись с *бюрократом*»).

– если заменяемое слово в именительном падеже и в предложении есть глагол, то он согласуется по роду и числу со словом–заменой. Если есть необходимость, род также изменяется у предикатива («*капитал должен работать на страну*» → «*богатство должно работать на страну*»). Если глагола в предложении нет, но есть краткое прилагательное (причастие), то род изменяется у него;

– если перед заменяемым словом стоит прилагательное или местоимение-прилагательное, то его грамматическая форма согласуется со словом-заменой («он купил *новую автомашину*» → «он купил *новый автомобиль*»);

Правила для слов-замен, которые являются **прилагательными, местоимениями-прилагательными, причастиями**:

– если заменяемое слово является прилагательным, местоимением-прилагательным или причастием, то грамматическая форма слова-замены согласуется с формой заменяемого слова («нам нужны *предприимчивые люди*» → «нам нужны *активные люди*»). В противном случае грамматическая форма слова-замены согласуется с формой ближайшего существительного в предложении («пейзажи там – *загаденье*» → «пейзажи там *красивые*»).

Правила для слов-замен, которые являются **глаголами**:

– если заменяемое слово является глаголом или деепричастием, то грамматическая форма слова-замены согласуется с формой заменяемого слова («он *лоботрясничал весь месяц*» → «он *бездельничал весь месяц*»).

Правила для слов-замен, которые являются **числительными**:

– если заменяемое слово является существительным или местоимением-существительным, то слово-замена по падежу согласуется с ним («я выбрал *первые две фотографии из дюжины*» → «я выбрал *первые две фотографии из двенадцати*»);

– если после заменяемого слова стоит существительное, то оно склоняется в соответствии с числительным («*четверка лошадей*» → «*четыре лошади*»);

– если перед заменяемым словом стоит прилагательное (местоимение-прилагательное), то оно согласуется по падежу со словом-заменой и преобразуется в форму множественного числа («*дадим сена нашей четверке лошадей*» → «*дадим сена нашим четырьмя лошадям*»);

– если в предложении есть глагол и слово-замена стоит в именительном падеже, то глагол преобразуется в форму множественного числа («*едет четверка лошадей*» → «*едут четыре лошади*»).

Правило для слов-замен, которые являются **нпредикативами**: если заменяемое слово является глаголом в прошедшем времени, то перед заменяющим его предикативом ставится глагол «быть» в форме «было» («*ему нетерпелось увидеть родных*» → «*ему было нетерпещ увидеть родных*»).

База правил синонимических замен словосочетаний

При синонимических заменах изменяемых словосочетаний используется База словосочетаний. Обнаружив в тексте словосочетание из базы, система заменяет его соответствующей доминантой. При замене словосочетаний восстановление синтаксиса в полученном предложении намного сложнее, чем при замене отдельных слов. В этом случае необходимо определить, от какого слова в исходном тексте будет зависеть грамматическая форма слова-замены, то есть найти опорное слово. В случае, если замена состоит из одного слова, для нахождения опорного слова и восстановления правильного синтаксиса после замены прежде всего используются следующие правила.

Рассмотрим более подробно первоначальные правила порождения нужной грамматической формы для слова-замены при замене словосочетаний.

Правила для слов-замен, которые являются **существительными** (или **местоимениями-существительными**):

– если слово-замена является существительным, то в качестве опорного слова выбирается первое по порядку существительное, местоимение-существительное или числительное в заменяемом словосочетании. Грамматическая форма слова-замены согласуется с опорным словом по числу и падежу («*у всех есть свои слабые стороны*» → «*у всех есть свои недостатки*», «*я нашел десять копеек*» → «*я нашел гривенник*»);

– если слово-замена не совпадает по роду с опорным словом, опорное слово в именительном падеже, и в предложении есть глагол, не входящий в заменяемое словосочетание, то он согласуется по роду со словом – заменой («*нападающая сторона дезориентировала противника*» → «*агрессор дезориентировал противника*»). Если глагола в предложении нет, но есть краткое прилагательное или причастие, то род изменяется у него;

– если перед заменяемым словосочетанием стоит прилагательное или местоимение-прилагательное, его грамматическая форма согласуется со словом-заменой («*мы гордимся нашим воздушным флотом*» → «*мы гордимся нашей авиацией*»);

– если заменяемое словосочетание содержит глагол в прошедшем времени и перед словом заменой стоит существительное, то перед словом-заменой вставляется глагол «быть» в прошедшем времени, соответствующий по грамматической форме слову-замене («*ему птичьего молока не хватало*» → «*у него было изобилие*»).

Правила для слов-замен, которые являются **прилагательными (местоимениями-прилагательными) или причастиями:**

– если слово-замена является прилагательным (или причастием), то в качестве опорного слова выбирается ближайшее к нему существительное, местоимение-существительное или местоимение прилагательное, которое не входит в заменяемое словосочетание («старик гол как сокол» → «старик бедный»). Если в предложении не нашлось такого слова, то опорным словом выбирается первое по порядку прилагательное (причастие) из заменяемого словосочетания, либо первое по порядку существительное, если прилагательных нет. Грамматическая форма слова-замены согласуется с опорным словом.

– если заменяемое словосочетание содержит глагол в прошедшем или будущем времени (кроме глагола «быть»), то перед словом-заменой вставляется глагол «быть» в прошедшем (будущем) времени, соответствующий по роду и числу слову-замене («вопрос стоял на повестке дня» → «вопрос был актуален»).

Правила для слов-замен, которые являются **глаголами:**

– если слово-замена является глаголом, то в качестве опорного слова выбирается глагол либо краткое причастие в заменяемом словосочетании. Если опорный глагол не в инфинитиве, то глагол-замена ставится в той же грамматической форме («преступника посадят в тюрьму» → «преступника арестуют», «он прикован к постели» → «он болеет»);

– если опорный глагол в неопределенной форме, то глагол-замена также ставится в неопределенной форме («нужно поднять настроение публике» → «нужно ободрить публику»);

– если в заменяемом словосочетании нет глагола, то глагол-замена согласуется по роду, числу и лицу с существительным или местоимением-существительным в именительном падеже. Если нет вспомогательных глаголов «быть» в прошедшем или будущем времени, то для глагола-замены выбирается настоящее время («они в волнении» → «они волнуются»). Если есть глагол «быть» в прошедшем времени, то для глагола замены выбирается прошедшее время, а глагол «быть» удаляется («они были в волнении» → «они волновались»). Если есть глагол «быть» в будущем времени, то глагол-замена ставится в инфинитиве («они будут в волнении» → «они будут волноваться»).

Правила для слов-замен, которые являются **наречиями:**

– для слова-замены, которое является наречием, никаких преобразований не выполняется. Исключением является наречие «сам», которое ставится в том же роде и числе, что и глагол, стоящий в предложении перед данным наречием («они видели это своими глазами» → «они видели это сами»). Если в заменяемом словосочетании есть глагол в прошедшем времени, то перед наречием-заменой добавляется глагол «было» («у него кровь стыла в жилах» → «ему было страшно»), если в будущем времени – добавляется глагол «будет» («у него кровь застынет в жилах» → «ему будет страшно»).

Правила для слов-замен, которые являются **предикативами:**

– для слова-замены, которое является предикативом, никаких преобразований не выполняется. Если в заменяемом словосочетании есть глагол в прошедшем времени и в предложении есть существительное в именительном падеже, то перед предикативом-заменой добавляется глагол «быть», в форме, совпадающей по роду и числу с этим существительным («он имел основания сомневаться» → «он был вправе сомневаться»). Если в именительном падеже местоимение первого или второго лица единственного числа, то глагол «быть» выбирается того же рода что и глагол в заменяемом словосочетании («я имела основания сомневаться» → «я была вправе сомневаться»).

Правила для замен, состоящих из нескольких слов (если нет метки <N>):

1) **Глагол + существительное.** В заменяемом словосочетании ищется глагол, и если он есть, то глагол в словосочетании-замене ставится в той же форме («*отряд покрыл расстояние в шестьдесят километров*» → «*отряд проделал путь в шестьдесят километров*»). Если глагола в заменяемом словосочетании нет, но есть краткое прилагательное или причастие, то глагол в словосочетании-замене ставится в форме настоящего времени, а по числу согласуется с прилагательным/причастием. Если перед заменяемым словосочетанием стоит существительное, то глагол ставится в третьем лице, если там стоит местоимение, то глагол согласуется с ним по лицу («*они сыпали соль на рану*» → «*они бредили душой*»).

2) **Глагол + прилагательное/причастие.** В заменяемом словосочетании ищется глагол, и если он есть, то глагол в словосочетании-замене согласуется с ним по роду. Прилагательное в словосочетании-замене ставится в ту же форму, что и прилагательное в заменяемом словосочетании («*он вышел сухим из воды*» → «*он остался безнаказанным*»). Если в заменяемом словосочетании нет прилагательного, то прилагательное в словосочетании-замене ставится в творительном падеже и по числу согласуется с существительным, которое стоит перед заменяемым словосочетанием («*факты получают огласку*» → «*факты становятся известными*»).

3) **Прилагательное + существительное** или **существительное + прилагательное.** В заменяемом словосочетании ищется существительное, и если оно есть, оба слова в словосочетании-замене согласуются с ним по падежу и числу («*он приглашает артистку из погорелого театра*» → «*он приглашает плохую актрису*»). Если перед заменяемым словосочетанием стоит числительное либо прилагательное, перед которым стоит числительное, оба слова в словосочетании-замене согласуются по падежу с этим числительным.

4) **Частица «не» + глагол.** Если словосочетание состоит из двух слов, первое из которых частица «не» а второе глагол, то первое слово остается неизменным, а форма второго слова определяется по правилам замены в случае, когда слово-замена является глаголом. («*он пустил на самотек производство*» → «*они не заботился о производстве*»)

5) **Глагол + другие части речи.** Если словосочетание состоит из двух или нескольких слов, одно из которых глагол, и оно не подпадает под перечисленные выше правила, то в заменяемом словосочетании ищется глагол, и если он есть, то глагол в словосочетании-замене ставится в той же форме. Остальные слова не изменяются («*узников вывели на свет божий*» → «*узников вывели наружу*»).

Изменение синтаксического управления

Зачастую для получения в замененном предложении правильного синтаксиса данных правил может быть недостаточно и требуется изменить синтаксическое управление. База слов и База словосочетаний снабжены метками для восстановления правильного синтаксиса в тексте после проведения замены:

| - означает, что последующая запись относится к заменяемому слову;

/ - означает, что последующая запись относится к слову, которое непосредственно предшествует заменяемому;

\ - означает, что последующая запись относится к слову, которое непосредственно следует за заменяемым [7].

Пример:

\ (с) тв-дат

творительный падеж с предлогом «с» должен быть заменен на дательный падеж без предлога.

Пример: «Раскланялся с соседями» система адаптации превратит в «Поклонился соседям». [8]

Более полный список данных меток приведен в работах [7], [8]. Система адаптации текста использует специально размеченную базу синонимов и набор правил соблюдения синтаксиса, что позволяет успешно осуществлять лексическую замену слов и словосочетаний с соблюдением правильного синтаксиса и сохранением семантики текста.

Тестирование работы системы адаптации на материалах Национального корпуса русского языка

Для тестирования работы системы адаптации использовались материалы Национального корпуса русского языка (НКРЯ) [9], т. к. он представляет современный русский язык во всём многообразии жанров, стилей, территориальных и социальных вариантов. Приведем ряд примеров работы системы адаптации для Базы слов:

1. Эта НЕРАЗБЕРИХА вокруг вашего вида спорта, наверное, мешает готовиться к стартам? → Этот БЕСПОРЯДОК вокруг вашего вида спорта, наверное, мешает готовиться к стартам?

2. На Руси считалось, что с этого дня начинается весенняя РАСПУТИЦА. → На Руси считалось, что с этого дня начинается весеннее БЕЗДОРОЖЬЕ.

3. Однажды к полковнику Тихомирову НАГРЯНУЛ дальний родственник - Сучков. → Однажды к полковнику Тихомирову ЯВИЛСЯ дальний родственник - Сучков.

4. Даже Станислав не нашел в себе ни сил, ни желания БАЛАГУРИТЬ. → Даже Станислав не нашел в себе ни сил, ни желания ШУТИТЬ.

5. Но, к сожалению, у нас нет необходимых служб, которые поддержали бы благородство и АЛЬТРУИЗМ людей, согласных отдать частицу себя ради спасения других. → Но, к сожалению, у нас нет необходимых служб, которые поддержали бы благородство и БЕСКОРЫСТИЕ людей, согласных отдать частицу себя ради спасения других.

6. Раз бросившись в ПРАЗДНОШАТАНИЕ, никак нельзя с ним скоро покончить. → Раз бросившись в БЕЗДЕЛЬЕ, никак нельзя с ним скоро покончить.

Наиболее ярко предлагаемый метод синонимических упрощений проявляется при использовании Базы словосочетаний. Ниже приведены соответствующие примеры на материале Национального корпуса русского языка (НКРЯ) [9]:

1. Но и таким людям он скажет: невелика радость будет им пережить бунт; пусть приготовятся ПОЛОЖИТЬ ЗУБЫ НА ПОЛКУ. → Но и таким людям он скажет: невелика радость будет им пережить бунт; пусть приготовятся ГОЛОДАТЬ.

2. Оптимистов много и сейчас; они продолжают на что-то надеяться, а у меня КОШКИ СКРЕБУТ НА СЕРДЦЕ. — Оптимистов много и сейчас; они продолжают на что-то надеяться, а я ТРЕВОЖУСЬ.

3. Как будто у тебя ДЕНЕГ КУРЫ НЕ КЛЮЮТ. → Как будто ты БОГАТЫЙ.

4. Ну конечно, куда этим перечницам Юдину и Константинову?! Из них же ПЕСОК СЫПЕТСЯ. А ты, Алешка, наша молодая надежда. → Ну конечно, куда этим перечницам Юдину и Константинову?! Они же СТАРЫЕ. А ты, Алешка, наша молодая надежда.

5. В общем-то, Роман – парень отнюдь не робкого десятка, но сейчас у него не то, чтобы ПОДЖИЛКИ ТРЯСЛИСЬ, нет, конечно, однако некоторая неуверенность наблюдалась. → В общем-то, Роман – парень отнюдь не робкого десятка, но сейчас он не то, чтобы ИСПУГАЛСЯ, нет, конечно, однако некоторая неуверенность наблюдалась.

6. Ты останешься не один, тебе будет весело. У меня КРОВЬ ЗАСТЫЛА В ЖИЛАХ. Бедная Надина! → Ты останешься не один, тебе будет весело. Я ИСПУГАЛСЯ. Бедная Надина!

7. Но я решил повезти его в Москву, немыслимо было его ОСТАВЛЯТЬ НА ПРОИЗВОЛ СУДЬБЫ. → Но я решил повезти его в Москву, немыслимо было о нем НЕ ЗАБОТИТЬСЯ.

8. И приходится через силу улыбаться и делать вид, что все это – ЧЕПУХА НА ПОСТНОМ МАСЛЕ. → И приходится через силу улыбаться и делать вид, что все это – ВЗДОР.

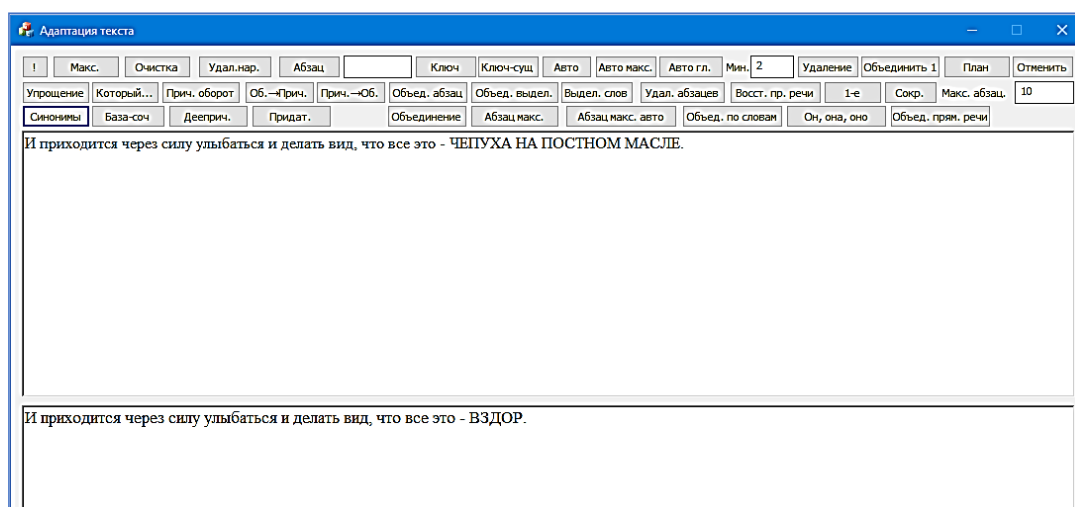


Рисунок 1 – Пример работы системы адаптации

Одним из эффективных примеров практического применения системы автоматической адаптации русскоязычных текстов является предобработка текста перед применением машинного перевода текста.

Так текст, содержащий словосочетание для замены, при дословном машинном переводе теряет смысл. Пример:

И приходится через силу улыбаться и делать вид, что все это – ЧЕПУХА НА ПОСТНОМ МАСЛЕ. — And you have to force yourself to smile and pretend that all this is NONSENSE IN VEGETABLE OIL.

В то время как упрощенный текст после машинного перевода сохраняет смысл и логику. Например:

И приходится через силу улыбаться и делать вид, что все это – ВЗДОР. — And you have to force yourself to smile and pretend that all this is NONSENSE.

Заключение

Перспектива практического применения системы автоматической адаптации русскоязычных текстов состоит в возможности применения полученных результатов при разработке программных приложений для машинного перевода, информационного поиска, автоматического реферирования, классификации текстов. А также для под-

готовки текстов для детей или взрослых, изучающих русский язык как иностранный, для людей, страдающих различными нарушениями восприятия, препятствующими пониманию лексически сложных текстов.

Список литературы

1. Павлов, А. А. О применении авторами научных текстов технических приемов, искажающих результаты проверки уникальности текстов. Обзор проблемы, опыт выявления и анализ подобных текстов / А. А. Павлов // Научная периодика: проблемы и решения. – 2020. – Т. 9, № 3-4. – EDN ZDFDNE.
2. Шатохина, С. А. Адаптация текстов в системе обучения русскому языку как иностранному (на материале онлайн-курса "Твои сибирские каникулы") / С. А. Шатохина, Т. К. Веренич, О. В. Вязовикина, В. С. Срмикян // Казанская наука. – 2019. – № 9. – С. 65-67. – EDN KVFZQB.
3. Некрасова, С. Д. Технические средства подготовки текста к переводу (на примере программ автоматического распознавания текста) / С. Д. Некрасова // Проблемы языка и перевода в трудах молодых ученых. – 2020. – № 19. – С. 121-127. – EDN RFQRDW.
4. Полицына, Е. В. Анализ подходов к автоматическому выделению контекстных синонимов из текстов на русском языке. / Е. В. Полицына, С. А. Полицын, А. С. Поречный, Е. Е. Милованова // Вестник ВГУ. Серия: Системный анализ и информационные технологии – 2020. – №3. – С. 120-132. <https://doi.org/10.17308/sait.2020.3/3046>
5. Александрова, З.Е. Словарь синонимов русского языка: Практический справочник [Текст] / З. Е. Александрова. - М.: Рус. яз., 2001. - 568 с.
6. Большакова, С.А. Об автоматизированных системах адаптации русскоязычных текстов [Текст] / С.А. Большакова // Материалы Донецкого международного научного круглого стола «Искусственный интеллект: теоретические аспекты и практическое применение» ИИ-2020. – Донецк: ГУ ИПИИ, 2020. – С. 27–32.
7. Ниценко А.В., Шелепов В.Ю., Большакова С.А., Ивашко К.С. Русское синтаксическое управление при словесных заменах. Слова с функциями наречия и существительного // Проблемы искусственного интеллекта. – 2020. – № 2(17). – С. 46–57.
8. Ниценко А.В., Шелепов В.Ю., Большакова С.А., Ивашко К.С. О словесных заменах, сохраняющих смысл русского предложения // Проблемы искусственного интеллекта. – 2020. – № 1(16). – С. 63–74.
9. Национальный корпус русского языка. [Электронный ресурс] // URL: <http://ruscorpora.ru/new/index.html>. (дата обращения: 10.05.2024).

References

1. Pavlov, A. A. On the use by the authors of scientific texts of technical techniques that distort the results of checking the uniqueness of texts. Review of the problem, the experience of identifying and analyzing such texts / A. A. Pavlov // Scientific periodicals: problems and solutions. - 2020. – Vol. 9, No. 3-4. – EDN ZDFDNE.
2. Shatokhina, S. A. Adaptation of texts in the system of teaching Russian as a foreign language (based on the material of the online course "Your Siberian holidays") / S. A. Shatokhina, T. K. Verenich, O. V. Vyazovikina, V. S. Srmikyan // Kazan Science. – 2019. – No. 9. – pp. 65-67. – EDN KVFZQB.
3. Nekrasova, S. D. Technical means of preparing text for translation (on the example of automatic text recognition programs) / S. D. Nekrasova // Problems of language and translation in the works of young scientists. - 2020. – No. 19. – pp. 121-127. – EDN RFQRDW.
4. Politsyna, E. V. Analysis of approaches to automatic selection of contextual synonyms from texts in Russian. / E. V. Politsyna, S. A. Politsyn, A. S. Porechny, E. E. Milovanova // Bulletin of the VSU. Series: System analysis and Information Technologies – 2020. – No.3. – pp. 120-132. <https://doi.org/10.17308/sait.2020.3/3046>
5. Alexandrova, Z.E. Dictionary of synonyms of the Russian language: A practical reference [Text] / Z. E. Alexandrova. - M.: Rus. yaz., 2001. - 568 p.
6. Bolshakova, S.A. On automated systems for adapting Russian-language texts [Text] / S.A. Bolshakova // Materials of the Donetsk International Scientific round table "Artificial intelligence: theoretical aspects and practical application" AI-2020. – Donetsk: GU IPII, 2020. – pp. 27-32.

7. Nicenko A.V., Shelepov V.Yu., Bolshakova S.A., Ivashko K.S. Russian syntactic control in verbal substitutions. Words with adverb and noun functions // Problems of artificial intelligence. – 2020. – № 2(17). – Pp. 46-57.
8. Nicenko A.V., Shelepov V.Yu., Bolshakova S.A., Ivashko K.S. On verbal substitutions preserving the meaning of a Russian sentence // Problems of artificial intelligence. – 2020. – № 1(16). – С. 63-74.
9. The National corpus of the Russian language. [Electronic resource] // URL: <http://ruscorpora.ru/new/index.html> . (date of application: 05/10/2024).

RESUME

S. A. Bolshakova

The system of automatic adaptation of russian-language texts and its practical significance

The article describes an adaptation system using a method of lexical simplification of the Russian-language text using a specially marked database of synonyms and a set of rules that allow automatic lexical replacement of words and phrases while preserving the semantics of the text and restoring the correct syntax. Dictionaries that are in the public domain were used to form a marked-up database of synonyms. The proposed methods and algorithms, as well as a database of synonyms, can be used for adaptation tasks, search engine optimization and automatic text abstraction, as well as for automatic translation.

РЕЗЮМЕ

С. А. Большакова

Система автоматической адаптации русскоязычных текстов и ее практическая значимость

В статье описана система адаптации, использующая способ лексического упрощения русскоязычного текста с помощью специально размеченной базы синонимов и набора правил, которые позволяют автоматически осуществлять лексическую замену слов и словосочетаний с сохранением семантики текста и восстановлением правильного синтаксиса. Для формирования размеченной базы синонимов использовались словари, находящиеся в открытом доступе. Предложенные методы и алгоритмы, а также база синонимов могут быть использованы для задач адаптации, поисковой оптимизации и автоматического реферирования текстов, а также при автоматическом переводе.

Большакова Светлана Анатольевна – младший научный сотрудник, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта». *Область научных интересов:* компьютерные информационные технологии, искусственный интеллект, обработка естественного языка, компьютерная лингвистика, интеллектуальный анализ информации, эл. почта svetlako@yandex.ru, адрес: 283048, г. Донецк, ул. Артема, д. 118 б, телефон +7949 431 53 58.

Статья поступила в редакцию 14.06.2024.