

Московский государственный университет имени М.В.Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра алгоритмических языков

Ушивец Никита Алексеевич

Программные средства обработки именных словосочетаний со сложными числительными

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

к.ф.-м.н., доцент,

Е.И.Большакова

Москва, 2021

Содержание

**Элементы оглавления не найдены.**

1. Введение

***Аннотация***

В данной̆ работе описываются разработанные и реализованные на языке python3 средства обработки именных словосочетаний со сложными числительными. Программная реализация позволяет осуществлять проверку текстовых коллекций на правильность употребления словосочетаний с числительными. Работоспособность и корректность разработанной̆ программы продемонстрирована путём подачи заведомо искаженного текста с окончательной выдачей исправленного документа. Проведены эксперименты, благодаря которым подтвердилась верность выдачи.

1. ***Введение***

Количество слов, которые мы называем числительными, в русском языке очень невелико – всего несколько десятков единиц. Наши предки-славяне воспринимали число как определённое количество конкретных предметов: *"три дома", "пять быков".* И лишь постепенно число стало использоваться само по себе, независимо от предметов и лиц, просто как число *три* или число *пять*.

Когда-то давно существовало сочетание *"подай пясть*", то есть подай руку (на которой пять пальцев). Современное слово *"пять"* связано с существовавшим словом *"пясть"("рука")* и когда-то имело конкретное значение *пять пальцев пясти*.

Слова, обозначающие в русском языке числа, собирались постепенно, из разных частей речи. *"Один", "два", "три", "четыре"* – из прилагательных. *"Пять", "шесть", "семь", "восемь", "девять"* – из существительных. Более сложные счётные слова образовывались описательно. *"Шесть на десять"* – значило 16, а не 60 *("шесть за десятью").* Если нужно было обозначить ещё большее количество, прибегали к описательным словам или метафорам. Слово *"тысяча"* для обозначения больших чисел появилось раньше других слов. Когда потребовалось обозначить число 10000, использовали слово *"тьма"*. Дальнейшее увеличение числового ряда поставило славян в тупик, и, когда это потребовалось, они заимствовали у греков слова *"легион" (сто тысяч) и "леодр" (миллион).* Со временем понадобились слова для обозначения десяти, ста миллионов, миллиарда и т.д. Для этого стали употреблять слова со значением бесконечности: десять миллионов – *"ворон"* (вечная птица), сто миллионов – *"колода"* (гроб). Каждый раз, когда требовалось новое увеличение осознаваемого числа, наши предки надеялись, что уж теперь-то это действительно конечное число. Мир чисел они видели столь же замкнутым и конечным, как и всё вокруг.

Таким образом, потребовалось очень много времени, чтобы числа воспринимались как нечто отвлечённое от объектов и явлений конкретно-предметного мира. Долгое время на письме славяне передавали число не цифрой, а буквой: *А – один, Б – ничего не значит, В – два, Г – три, Д – четыре, Е – пять…* Цифровое обозначение чисел на письме появилось сравнительно поздно.

С течением времени счётные имена становились всё больше похожими друг на друга. У них исчезла категория рода и числа – на путях истории счётные имена растеряли эти важнейшие признаки и тем самым окончательно обособились от всех остальных имён. Они образовали самостоятельную часть речи – имя числительное. В середине 18 века М.В.Ломоносов в своей грамматике впервые назвал числительные самостоятельной частью речи.

Сейчас же имена числительные активно используются во всех сферах нашей жизни, но составляют в современном русском языке относительно немногочисленную и замкнутую по сравнению с другими частями речи группу слов. К примеру, существительные и прилагательные появляются с завидной частотой (сегодня эмодзи и косплей, завтра еще что-нибудь), а вот числительные все уже с нами.

Имя числительное — это самостоятельная знаменательная часть речи, объединяющая слова, которые обозначают числа, количество предметов или порядок предметов при счете и отвечают на вопрос *сколько? или какой?* Грамматические признаки числительных неоднородны и зависят от того, к какому разряду по значению принадлежит числительное. У разных числительных свои особенности словоизменения, поэтому нужно быть с ними очень внимательными.

Тем не менее, несмотря на частотность употребления и использование практически во всех сферах жизни, имена числительные, представленные в современном русском языке, вызывают определенные трудности. Общеизвестно, что русский язык — один из самых сложных языков в мире. Простое тому подтверждение — огромное количество ошибок в устной и письменной речи, связанных с падежными склонением, большим количеством правил и исключениями из них.

Как ни странно, но, по некоторым данным, около 90% носителей русского языка не замечают этих ошибок или таковыми их не считают. Посмотрев в интернете имеющиеся решения для исправления ошибок, стало понятно, что нет приложения с комплексным подходом к устранению повседневных упущений. В следствие чего было интересно решить данную трудность, хоть только на часть, связанную с числительными. Как никак это константная группа много, где применима.

Числительные в большинстве своем применяются в словосочетаниях. Это комбинация двух или более слов, которые имеют тенденцию к совместной встречаемости и отображают особенности языка. В русском, значительные трудности вызывает сочетание числительных с существительными: в одних случаях числительное выступает в роли синтаксически главного компонента, в других случаях – зависимого. Склонение числительных не имеет единого образца, поэтому придется ориентироваться по обстоятельствам, вернее, в зависимости от данного вам числительного.

Современные текстовые анализаторы не ориентированы на словосочетания с числительными. Они либо не умею компоновать имена числительные с другим звеном словосочетания, либо же соединяют их без учитывания рода, числа, падежа и т.д. Тем самым вопрос решения данных упущений крайне актуален.

Исходя из этого в данной выпускной квалификационной работе решается задача создания программного средства синтеза словосочетаний с существующими в русском языке именами числительными. На основе типизацииимён числительных, представленной в виде совокупности правил склонения с различного рода существительными русского языка, найден метод обработки вводимых словосочетаний. Используя имеющиеся данные реализован алгоритм, позволяющий автоматизировано склонять получаемые словосочетания.

Выполнение было разделено на три основные взаимосвязанные части:

* поиск и изучения необходимого теоретического материала,
* построение метода относительно созданных правил,
* доработка и тестирование алгоритмов.

Теоретический материал из различных источников позволив выявить и разделить характерные черты склонения существительных с разными группами имен числительных русского языка, что и легло в основу разработки базы правил. Основной задачей на этапе программной реализации алгоритмов было правильное перенесения словесного описания базы правил в структурированный вид, который можно использовать в коде и с легкостью обновлять или модифицировать. В рамках работы были написаны программные модули на языке Python3, на основе открытой библиотекой pymorphy2. Путём анализа вводимых данных и выбирается подходящий вариант склонения словосочетаний.

**Сложные числительные и словосочетания с числительными**

2.1 Основные понятия, виды числительных

В русском языке различают следующие виды числительных:

* **Порядковые**

Числительные обозначают порядок предметов при счете *(первый, второй, пятый, сто двадцать пятый)* и отвечают на вопрос *какой?*

* **Количественные**

Обозначают отвлеченные числа *(пять)* и количество предметов *(пять столов)* и отвечают на вопрос *сколько?*

Количественные числительные бывают **целые** *(пять),* **дробные** *(пять седьм*ых) и **собирательные** *(пятеро),* и **неопределенно-количественные** *(мало, много, несколько)*

**Целые** количественные числительные обозначают целые числа или количества. Целые количественные числительные сочетаются со счетными существительными, т. е. с такими существительными, которые обозначают предметы, которые можно посчитать штуками.

**Дробные** количественные числительные обозначают дробные числа или количества и сочетаются как со счетными существительными *(две третьих пирога),* так и с несчетными существительными *(две третьих воды),* но не могут сочетаться с одушевленными существительными в единственном числе.

**Собирательные** числительные обозначают количество предметов как целое. К собирательным числительным относятся слова: *оба, двое, трое, четверо, пятеро, шестеро, семеро, восьмеро, девятеро, десятеро*. Собирательные числительные имеют ограниченную сочетаемость; они сочетаются не со всеми существительными, а только с некоторыми:

1) с существительными, которые называют лиц мужского пола *(двое мужчин);* числительное оба сочетается также и с существительными, обозначающими лиц женского пола *(обе женщины);*

2) с существительными человек, лицо, ребенок *(пятеро людей, лиц, детей);*

3) с названиями детенышей животных *(семеро козлят);*

4) с существительными, имеющими формы только множественного *числа (двое саней)*; с этими существительными сочетаются преимущественно числительные *двое, трое и четверо*;

5) с существительными, называющими парные предметы: *двое носков* *(два носка — это два носка, а двое носков — это четыре носка, т. Е. две пары носков)*

6) с личными местоимениями мы, вы, они *(не было их двоих).*

**Неопределенно-количественные** числительные не обладают основным свойством количественных числительных: они не способны выражать точное количество предметов, а имеют лишь приблизительное или обобщенное значение *(мало времени, много людей).* Данная группа числительных сочетается со всеми разрядами существительных. В них так же выражены яркие признаки наречий: способность определять глагол *(мало двигаешься, много гуляешь)*; наличие форм сравнительной степени *(много говори — больше говори, мало танцуешь — меньше танцуй)*; наличие форм субъективной оценки *(немного отдохни — немножко отдохни, мало взаимопонимания — маловато взаимопонимания);* способность сочетаться с наречиями степени *(очень мало, очень много работаешь);* отсутствие падежных форм.

В некоторых источниках можно найти разделение имен числительных еще на несколько групп, которые не являются основными:

1. **Мультипликативное** числительное: *одиночный, двойной, тройной, четырёхкратный и т.д.*
2. **Счётное** числительное: *единичный, двоичный, шестнадцатеричный и т.д..*
3. **Дистрибутивное** **(распределительное)** числительное: *по одному, по двое, по трое и т. Д.*

**Мультипликативные** и **счетные** числительные часто относят к прилагательным или к порядковым числительным, так как они имеют схожие морфологические признаки.

**Дистрибутивные** числительные, в зависимости от контекста, могут отнести к местоимениям.

По структуре выделяют числительные простые и сложные:

1. **Простые** (в морфемной структуре только один корень: *один, двенадцать, тридцать)*
2. **Сложные** (в морфемной структуре два корня и два окончания: *пятьдесят, семьсот, двести)*
3. **Составные** (составлены из нескольких числительных: *сто сорок два, триста восемьдесят пятый, две третьих*).

**Целые** и **собирательные** числительные сочетаются с существительными следующим образом: в Им. П. (и В. П. при неодушевленных существительных) числительное является главным словом и управляет существительным, требуя его постановки в Род. П. единственного числа (при числительных *два, три, четыре*) или множественного числа (при числительных *пять и далее*). В остальных падежах главным является существительное, а числительное с ним согласуется, например:

*два* (Им. П.) *стола* (Род. П. ед.ч.)

*двух* (Род. П.) *столов* (Род. П. мн. Ч.)

*двум* (Д. п.) *столам* (Д. п. мн. Ч.)

*двумя* (Т. П.) *столами* (Т. П. мн. Ч.)

(о) *двух* (П. п.) *столах* (П. п. мн. Ч.)

**Дробные** количественные числительные всегда управляют Род. П. существительного, а число этого существительного зависит от смысла конструкции, ср.: *одна вторая конфеты — одна вторая конфет*.

В грамматическом отношении среди количественных числительных выделяются слова *один, тысяча, миллион, миллиард, триллион* и другие названия больших чисел.

Слово *один* изменяется по родам, числам и падежам, в которых согласуется с существительным *(один стол, одна парта, одно окно, одни сани).* В количественном значении форма множественного числа слова *один* сочетается с существительными, имеющими форму только множественного числа. Склоняется слово *один* по смешанному склонению: *один, одна, одно, одни, одного, одной, одного, одних и т. Д*. Другими словами, числительное *один* ведет себя, как относительное прилагательное.

Слова *тысяча, миллион, миллиард и др.* имеют постоянный морфологический признак рода *(первая тысяча —* жен. Род*, первый миллион —* муж. Род*)*, изменяются по числам и по падежам *(первые тысячи, первых тысяч).*

Склоняются эти слова, как существительные (*тысяча* — I склонение, *миллион* и др. — II склонение). При сочетании с существительными эти слова всегда управляют существительным, требуя его постановки в форме Род. П. мн. Числа:

Им. П. *тысяча тонн*

Род. П. *тысячи тонн*

Д. п. *тысяче тонн*

В. П. *тысячу тонн*

Т. П. *тысячей тонн*

П. п. (о) *тысяче тонн*

В украинском языке имена числительные имеют схожее деление по значению и грамматическим признакам. Соответственно, они бывают количественные и порядковые.

1. **Порядковые**: обоначают порядок предмета по счету *(десятий, перший, другий)*
2. **Количественные**: обозначают абстрактные числа *(п'ять, десять)* или количество *предметов (дві будівлі, три іграшки)*

**Количественные** числительные делятся соответствующим образом на целые *(три дівчини),* **дробные** *(одна третя шматка),* **собирательные** *(троє людей)* и **неопредленно-количественные** *(мало часу).*

По строению слова:

* **простые** *(два, три)*
* **сложные** *(шістнадцять)*
* **составные** *(двадцять шість).*

2.2 Проблемы автоматического синтеза и анализа именных словосочетаний с числительными

Для понимания проблем автоматического синтеза нужно изначально разобраться с трудностями перевода русского языка на украинский и наоборот. Потому опишем сложные случаи склонения имён числительных для двух языков, относительно их правил. Стоит начать с каждого разряда числительных по отдельности.

* **Порядковые числительные в обоих языках согласуются с существительными, т.е. склоняются как прилагательные.**

Порядковые числительные изменяются по падежам, числам, а в единственном числе и по родам. Они согласуются с существительными, как прилагательные, и склоняются по образцу прилагательных *(седьмой,седьмого,седьмому,седьмым и т.д.).*

Числительное **третий** склоняется, как притяжательные прилагательные на -ИЙ.

Винительный падеж мужского рода и множественного числа у числительных, согласованных с:

* неодушевлёнными существительными, совпадает с Им.п.
* одушевлёнными существительными, совпадает с Род.п.

При склонении составных порядковых числительных изменяется только последнее слово *(сто двадцать пятый, сто двадцать пятого, сто двадцать пятому и т.д.)* При склонении сложных порядковых числительных изменяется окончание только последней части *(пятидесятый,пятидесятого,пятидесятому,пятидесятым и т.д.)*

* **Дробные числительные**

В русском языке дробное числительное полтора имеет две формы именительного падежа — полтора (для мужского и среднего рода) и *полторы* (для женского рода): *полтора бревна, полтора литра, полторы жизни*. Данное числительное имеет собственные формы склонения и требует от существительного формы родительного падежа единственного или множественного числа. Форма всех косвенных падежей (кроме винительного) — полутора *(о полутора днях).* Числительное *полтораста* также имеет собственные форма склонения во всех косвенных падежах (кроме винительного) — полутораста*,* и требует от существительного форму родительного падежа множественного числа.

Все остальные дробные числительные имеют в себе числитель и знаменатель. Чтобы просклонять любую дробь, числитель нужно склонять как обычное количественное числительное, а знаменатель как порядковое *(одна восьмая, нет трех пятых).*

Все дробные числительные требуют от существительного формы родительного падежа, вне зависимости от падежа числительного, причем существительные могут употребляться как в единственном, так и во множественном числе): *две третьих спальни, одна десятая суток.*

В украинском языке все дробные числительные, которые имеют в своем составе числитель и знаменатель *(дві п ‘ятих, одна сьома)* склоняются по аналогии с русским языком, при этом требуя от существительного только форму родительного падежа единственного числа. Основное отличие в склонении дробных числительных приходится на слова *полтора, полторы, полтораста (півтора, півтори, півтораста).* В украинском языке эти числительные НЕ склоняются, следовательно во всех косвенных падежах имеют форму именительного падежа: *півтора аркушам, півтораста кілометрів, півтори години.*

* **Согласование количественных числительных с существительными.**
* В украинском языке числительное *"один"* изменяется по родам, числам и падежам, как местоимение *"той" (тот): один-ø / одна / одне (одно) / одні*. Это исключение среди количественных числительных.

В русском языке числительное *один (одна, одно, одни)* и все количественные числительные также грамматически зависимы от существительных и согласуются с ними в роде, числе и падеже, как прилагательные: *один год, одна неделя, одно окно, одни сутки*

* В украинском языке числительные *два, три, четыре* подчиняют существительные, которые стоят при них в форме именительного падежа множественного числа (часто с ударением родительного падежа единственного числа: *два столи, три карбованці, чотири мішки, дві сестри*).

В русском языке есть особенности в синтаксической сочетаемости с существительными, к которым они относятся. В Им.п. и В.п они требуют после себя существительные в форме Род.п., при этом числительные *полтора, два, три, четыре (два стола, три ночи, четыре девушки)* требуют существительного в ед. ч. Остальные числительные пять, *шесть, семь, восемь, девять*, а так же, *составные числительные*, которые оканчиваются на эти цифры - во мн. ч.

* В украинском языке у сложных числительных *от пятидесяти до восьмидесяти* склоняется только последняя часть *(п’ятдесяти, сімдесяти).*

В русском же языке при склонении подобных числительных изменяются обе части по образцу существительных третьего склонения.

Им. п. *Пятьдесят, семьдесят (домов)*

Род. п. *Пятидесяти, семидесяти (домов)*

Д. п. *Пятидесяти, семидесяти (домам)*

* **Собирательные числительные в обоих языках сочетаются с существительными так же, как обычные количественные числительные, т. е. согласуются в косвенных падежах (например: *нет троих студентов, подошли к троим студентам)* и не согласуются в именительном *(*например*, трое студентов).*** 
  + Однако для русского языка это не всегда так. В отличие от обычных количественных числительных собирательные сочетаются не со всеми существительными, поддающимися счёту, а только с некоторыми:

1. С существительными, которые употребляются только во множественном числе, например: *двое ворот, трое щипцов, четверо саней.* Но начиная с числительного пять возможно параллельное употребление собирательных и обычных числительных, например: *пятеро ворот и пять ворот, шестеро ножниц и шесть ножниц*; причём обычные числительные более употребительны. В косвенных падежах во всех случаях предпочтительнее употреблять количественные числительные, например: двух ворот *(а не двоих ворот*), двум воротам *(а не двоим воротам)*, двумя воротами *(а не двоими воротами)*, о двух воротах *(а не о двоих воротах)* и т. п

2. С существительными, имеющими парное значение, например: *двое перчаток, трое чулок (т. е. две пары перчаток, три пары чулок)*. В косвенных падежах, кроме винительного, вместо собирательных числительных употребляется сочетание обычного числительного с существительным пара, например: *нет двух пар перчаток, трёх пар чулок и т. д.*

В настоящее время и в именительном падеже вместо *двое перчаток, трое чулок* чаще говорят *две пары перчаток, три пары чулок и т. п.*

3. С существительными, одушевлёнными мужского рода, обозначающими лиц, параллельно с обычными числительными, например: *двое учеников (и два ученика). пятеро мальчишек (и пять мальчишек)*. (Обычное числительное предпочитается при существительном, обозначающем лицо, имеющее высокое воинское звание, учёное звание, степень и т. д., например: четыре генерала, пять маршалов, шесть профессоров; не принято говорить: *«пятеро генералов», «шестеро профессоров», «четверо маршалов»*.)

4. С существительными дети, ребята, люди, с существительным лицо в значении «человек»: *двое детей, трое ребят, трое молодых людей, четверо незнакомых лиц*

5. Кроме сочетания с указанными существительными, собирательные числительные сочетаются с личными местоимениями *мы, вы, они*. например: *мы, всё пятеро, послали вас четверых; не было их троих*

С существительными, одушевлёнными женского рода, обозначающими как лиц, так и животных, собирательные числительные употреблять не принято, т. е. только: *две девочки, три женщины, четыре козы и т. п.*

* + В украинской языке для собирательных числительных сочетаются такие же группы существительных, тем не менее и здесь есть отличия:

1. Собирательные числительные сочетаются с одушевленными лицами мужского рода *(троє друзів, шестеро випускників)*

2. С одушевлёнными существительными среднего рода, которые обозначают малое существо *(двоє кошенят, семеро каченят)* или с предметами среднего рода *(четверо відер, восьмеро вікон);*

3. С существительными, которые имеют только форму множественного числа *(троє саней, четверо дверей)*.

В украинском языке, так же, как и в русском, нельзя согласовать собирательное числительное с существительным женского рода или с неодушевленным предметом мужского рода *(три дівчини, два олівці).*

**Самым главным отличием** является то, что что в украинском языке можно употреблять собирательное числительное с неодушевленным предметом среднего рода, что нельзя сделать в русском языке *(четверо полотен, двоє відер)*