**ZMD\_092 «Лемматизация названий материалов**»

**Таблицы:**

MAKT «Краткие тексты к материалам»

**Функциональный модуль:**

Z\_MD\_SEARCH\_IN\_DICTIONARY «Поиск слова и его признаков в словаре»

Z\_MD\_PROCESS\_WORDS «Получение леммы слова»

Z\_MD\_GENERATE\_COMBINATIONS «Создание комбинаций полученных слов»

Z\_MD\_GENERATE\_COMBINATIONS\_HELPER «Вспомогательная функция для создания комбинаций»

**На выходе:**

ZMD\_092 «Лемматизация названий материалов»

**Алгоритм:**

Создать локальную таблицу LT\_WORDS «Слова»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | WORD\_ID | Идентификатор слова | - | NUMC | 3 | Первичный ключ.  Счётчик. |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |
| 3 | ORDER | Порядок слова в записи | - | INTEGER | 2 | - |
| 4 | INPUT\_WORD | Содержание слова | - | CHAR | 40 | - |
| 5 | LEMMA | Лемма слова | - | CHAR | 40 | - |
| 6 | PART\_OF\_SPEECH | Часть речи | - | CHAR | 30 | - |
| 7 | CASE | Падеж слова | - | CHAR | 40 | - |
| 8 | ANIMATION | Одушевленность/неодушевленность | - | CHAR | 15 | - |
| 9 | KIND | Пол слова | - | CHAR | 15 | - |
| 10 | MEASURE\_OF\_NUMBER | Мера численности | - | CHAR | 19 | - |
| 11 | SPECIAL\_MARKS | Особые метки | - | CHAR | 3 | - |

Создать локальную таблицу LT\_WORDS\_INTERMEDIATE «Промежуточная таблица»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | WORD\_ID | Идентификатор слова | - | NUMC | 3 | Первичный ключ.  Счётчик. |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |
| 3 | INPUT\_WORD | Содержание слова | - | CHAR | 40 | - |
| 4 | PART\_OF\_SPEECH | Часть речи | - | CHAR | 4 | - |
| 5 | CASE | Падеж слова | - | CHAR | 4 | - |
| 6 | ANIMATION | Одушевленность/неодушевленность | - | CHAR | 4 | - |
| 7 | KIND | Пол слова | - | CHAR | 4 | - |
| 8 | MEASURE\_OF\_NUMBER | Мера численности | - | CHAR | 4 | - |

Создать локальную таблицу LT\_INPUT\_MATERIALS «Полученный материал»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | WORD | Текст слова | - | CHAR | 40 | - |

Создать локальную таблицу LT\_TAGS «Промежуточная таблица тэгов слова»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | TAGS | Тэги слова | - | CHAR | 40 | - |

Создать локальную таблицу LT\_ ID\_PER\_ORDER «Промежуточная таблица ID лемм, соответствующих порядковому номеру»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | ORDER | Порядок слова в записи | - | INTEGER | 2 | - |
| 3 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

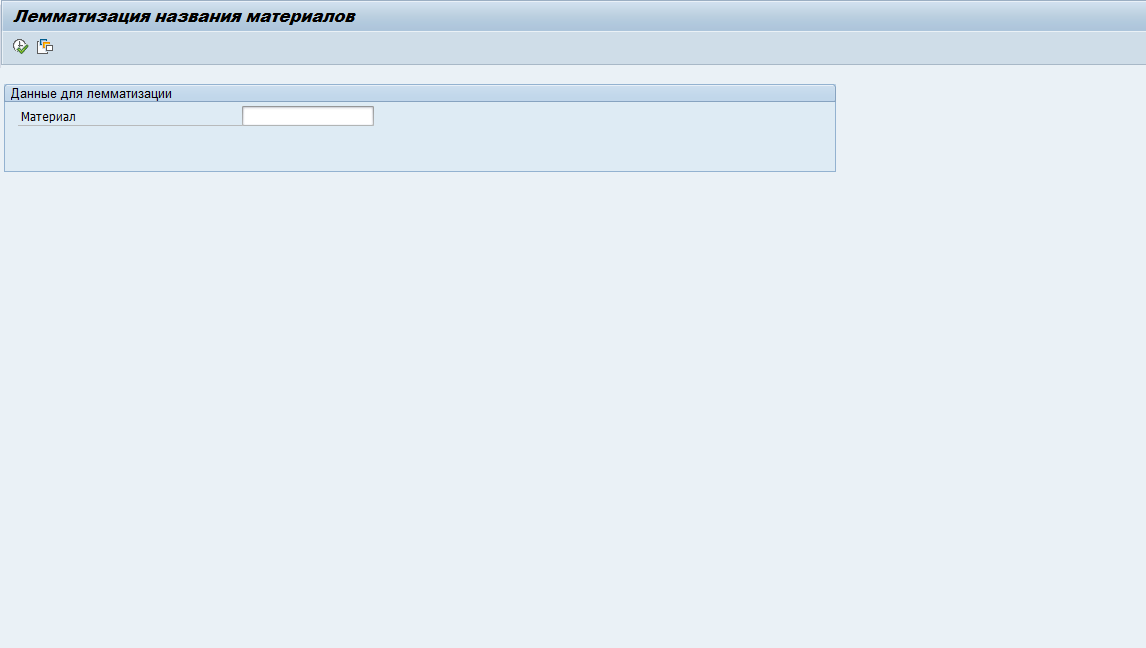
Создать локальную таблицу LT\_ LEMMAS\_PER\_ORDER «Количество уникальных лемм на порядковый номер»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | ORDER | Порядок слова в записи | - | INTEGER | 2 | - |
| 3 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Создать локальную таблицу LT\_ RESULT\_LEMMAS\_ORDER «Итоговые комбинации лемм»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Создать пользовательскую форму для заполнения поля «Материал»:



На форме разместить:

* Текстбокс:
  + S\_MATNR
* Кнопки:
  + Выполнить 
  + Сохранить вариант 

При нажатии на кнопку «Выполнить» :

Создать переменную LV\_INPUT\_WORDS\_REPLACE(CHAR, 40) = “”.

Найти запись в таблице MARA по условию:

* MAKT - MATNR = S\_MATNR

Сохранить название материала в переменной:

* LV\_INPUT\_WORDS \_REPLACE = MAKT – MAKTX

Создать переменную LV\_INPUT\_WORDS (CHAR, 40) = “”.

Записать данные в переменную LV\_INPUT\_WORDS из переменной LV\_INPUT\_WORDS\_REPLACE до символа “\”, не включая его.

Заменить все символы “.” в переменной LV\_INPUT\_WORDS на символы “. ”(точка и пробел).

Заменить все символы “)” в переменной LV\_INPUT\_WORDS на символы “ ”(пробел).

Заменить все символы “(” в переменной LV\_INPUT\_WORDS на символы “ ”(пробел).

Разбить переменную LV\_INPUT\_WORDS в таблицу LT\_INPUT\_MATERIALS по « »(пробел).

Обработать все записи таблицы LT\_INPUT\_MATERIALS по условию:

* Если LT\_INPUT\_MATERIALS-WORD содержит “-”:
  + Создать переменную LV\_INTERMEDIATE\_WORD(CHAR, 40) = LT\_INPUT\_MATERIALS-WORD
  + Вызвать ФМ Z\_MD\_SEARCH\_IN\_DICTIONARY «Поиск слова и его признаков в словаре» Условия запуска:
    - IMP\_WORD= LV\_INTERMEDIATE\_WORD
    - IMP\_WORD\_CHECK= “X”
  + На выходе получить LT\_WORDS
  + Если количество записей в LT\_WORDS ==”0”
    - То:
      * Создать переменную LV\_AFTER\_HYPHEN (CHAR,40)
      * Записать в переменную LV\_AFTER\_HYPHEN значение записи LV\_INTERMEDIATE\_WORD после знака «-», не включая его
      * LT\_INPUT\_MATERIALS-WORD = значение LV\_INTERMEDIATE\_WORD до знака «-»
      * Добавить запись после LT\_INPUT\_MATERIALS-WORD= LV\_INTERMEDIATE\_WORD в таблицу LT\_INPUT\_MATERIALS:
        + LT\_INPUT\_MATERIALS-WORD LV\_AFTER\_HYPHEN

Создать переменную LV\_ INPUT\_MATERIALS\_COUNT (INTEGER, 3) = количеству записей в LT\_INPUT\_MATERIALS.

Создать переменную LV\_CURRENT\_POSITION\_NUMBER (INTEGER, 3) = 0

Создать переменную LV\_CURRENT\_TAG (INTEGER, 3) = 0

Создать переменную LV\_DEVIDED\_HYPHEN\_FLAG (CHAR,1) = ””

Создать переменную LV\_CURRENT\_WORD (CHAR, 40) = ””

Создать переменную LV\_RESULT\_WORD\_TAGS (CHAR, 40) = ””

Создать переменную LV\_TAGS\_INTERMEDIATE\_COUNT (INTEGER, 3) = 0

Создать переменную LV\_CURRENT\_LEMMA\_ID (CHAR, 8) = ””

Создать переменную LV\_CURRENT\_WORD\_LEMMA (CHAR, 40) = ””

Выполнить LV\_ INPUT\_MATERIALS\_COUNT раз:

{

LV\_DEVIDED\_HYPHEN\_FLAG=””.

Получить запись из таблицы LT\_INPUT\_MATERIALS по условию:

* LT\_INPUT\_MATERIALS-POSITION\_ID = LV\_CURRENT\_POSITION\_NUMBER

Для полученной записи выполнить:

* LV\_CURRENT\_WORD = LT\_INPUT\_MATERIALS-WORD
* Заменить в LT\_INPUT\_MATERIALS-WORD все символы «-» на «@»
* Если в LV\_CURRENT\_WORD содержатся символы «Y»
* То:
  + Удалить в LT\_INPUT\_MATERIALS-WORD все символы «Y»
  + LV\_DEVIDED\_HYPHEN\_FLAG = «Х»

Добавить запись в LT\_WORDS с помощью ФМ Z\_MD\_SEARCH\_IN\_DICTIONARY «Поиск слова и его признаков в словаре». Условия запуска:

* IMP\_WORD= LV\_CURRENT\_WORD
* IMP\_WORD\_CHECK= “”

На выходе получить:

Добавить в LT\_WORDS\_INTERMEDIATE запись:

* LT\_WORDS\_INTERMEDIATE- LEMMA\_ID = LT\_ OUT- LEMMA\_ID
* LT\_WORDS\_INTERMEDIATE-INPUT\_WORD= LT\_ OUT -INPUT\_WORD
* LT\_WORDS\_INTERMEDIATE-PART\_OF\_SPEECH= LT\_ OUT -PART\_OF\_SPEECH
* LT\_WORDS\_INTERMEDIATE-CASE= LT\_ OUT -CASE
* LT\_WORDS\_INTERMEDIATE-ANIMATION= LT\_ OUT -ANIMATION
* LT\_WORDS\_INTERMEDIATE-KIND= LT\_ OUT -KIND
* LT\_WORDS\_INTERMEDIATE- MEASURE\_OF\_NUMBER= LT\_OUT- MEASURE\_OF\_NUMBER

Очистить таблицу LT\_TAGS.

Разбить переменную LV\_RESULT\_WORD\_TAGS в таблицу LT\_TAGS по символу « ».

LV\_WORD\_PLUS\_TAGS \_COUNT = количество записей в LT\_WORDS\_INTERMEDIATE.

Обнуляем счётчик LV\_CURRENT\_TAG = 0

Выполнить LV\_WORD\_PLUS\_TAGS \_COUNT раз:

{

Выбрать запись из таблицы LT\_WORDS\_INTERMEDIATE по условию:

* LT\_WORDS\_INTERMEDIATE-WORD\_ID= LV\_CURRENT\_TAG

Сохраним найденное значение в переменной:

* LV\_CURRENT\_WORD= LT\_WORDS\_INTERMEDIATE- INPUT\_WORD
* LV\_CURRENT\_LEMMA\_ID = LT\_WORDS\_INTERMEDIATE- LEMMA\_ID

Получить лемму слова с помощью ФМ Z\_MD\_PROCESS\_WORDS «Получение леммы слова» Условие запуска:

* IMP\_LEMMA\_ID = LV\_CURRENT\_LEMMA\_ID
* IMP\_WORD\_TEXT = LV\_CURRENT\_WORD

На выходе:

* IMP\_LEMMA

Сохраним полученную лемму в переменной:

* LV\_CURRENT\_WORD\_LEMMA= IMP\_LEMMA

Если LV\_DEVIDED\_HYPHEN\_FLAG = «Х»

То:

* Найти запись в таблице LT\_WORDS\_INTERMEDIATE LV\_CURRENT\_TAG по условию:
  + LT\_WORDS\_INTERMEDIATE-WORD\_ID= LV\_CURRENT\_TAG
* Добавить запись в LT\_WORDS по условию:
  + LT\_WORDS-LEMMA\_ID= LT\_WORDS\_INTERMEDIATE-LEMMA\_ID
  + LT\_WORDS-INPUT\_WORD= LT\_WORDS\_INTERMEDIATE-INPUT\_WORD
  + LT\_WORDS-ORDER= LV\_CURRENT\_POSITION\_NUMBER
  + LT\_WORDS-LEMMA= LV\_CURRENT\_WORD\_LEMMA
  + LT\_WORDS-PART\_OF\_SPEECH= LT\_WORDS\_INTERMEDIATE-PART\_OF\_SPEECH
  + LT\_WORDS-CASE= LT\_WORDS\_INTERMEDIATE-CASE
  + LT\_WORDS-ANIMATION= LT\_WORDS\_INTERMEDIATE-ANIMATION
  + LT\_WORDS-KIND= LT\_WORDS\_INTERMEDIATE-KIND
  + LT\_WORDS- MEASURE\_OF\_NUMBER= LT\_WORDS\_INTERMEDIATE- MEASURE\_OF\_NUMBER
  + LT\_WORDS- SPECIAL\_MARKS=”Y”

Иначе:

* Найти запись в таблице LT\_WORDS\_INTERMEDIATE LV\_CURRENT\_TAG по условию:
  + LT\_WORDS\_INTERMEDIATE-WORD\_ID= LV\_CURRENT\_TAG
* Добавить запись в LT\_WORDS по условию:
  + LT\_WORDS-LEMMA\_ID= LT\_WORDS\_INTERMEDIATE-LEMMA\_ID
  + LT\_WORDS-INPUT\_WORD= LT\_WORDS\_INTERMEDIATE-INPUT\_WORD
  + LT\_WORDS-ORDER= LV\_CURRENT\_POSITION\_NUMBER
  + LT\_WORDS-LEMMA= LV\_CURRENT\_WORD\_LEMMA
  + LT\_WORDS-PART\_OF\_SPEECH= LT\_WORDS\_INTERMEDIATE-PART\_OF\_SPEECH
  + LT\_WORDS-CASE= LT\_WORDS\_INTERMEDIATE-CASE
  + LT\_WORDS-ANIMATION= LT\_WORDS\_INTERMEDIATE-ANIMATION
  + LT\_WORDS-KIND= LT\_WORDS\_INTERMEDIATE-KIND
  + LT\_WORDS- MEASURE\_OF\_NUMBER= LT\_WORDS\_INTERMEDIATE- MEASURE\_OF\_NUMBER
  + LT\_WORDS- SPECIAL\_MARKS=”0000”

Увеличить счётчик LV\_CURRENT\_TAG на 1.

}

Увеличить счётчик LV\_CURRENT\_POSITION\_NUMBER на 1.

}

Обработать все записи таблицы LT\_WORDS по следующему условию:

* Для поля LT\_WORDS-PART\_OF\_SPEECH:
  + Если LT\_WORDS-PART\_OF\_SPEECH== «NOUN»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «имя существительное »
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «ADJF»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «имя прилагательное (полное)»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «ADJS»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «имя прилагательное (краткое)»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «COMP»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «компаратив»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «VERB»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «глагол (личная форма)»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «INFN»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «глагол (инфинитив)»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «PRTF»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «причастие (полное)»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «PRTS»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «причастие (краткое)»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «GRND»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «деепричастие»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «NUMR»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «числительное»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «ADVB»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «наречие»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «NPRO»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «местоимение-существительное»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «PRED»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «предикатив»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «PREP»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «предлог»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «CONJ»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «союз»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «PRCL»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «частица»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «INTJ»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «междометие»
  + Иначе если LT\_WORDS-PART\_OF\_SPEECH== «0000»
  + То:
    - LT\_WORDS-PART\_OF\_SPEECH= «»
* Для поля LT\_WORDS-CASE:
  + Если LT\_WORDS-CASE == «nomn»
  + То:
    - LT\_WORDS-CASE = «именительный падеж»
  + Иначе если LT\_WORDS-CASE == «gent»
  + То:
    - LT\_WORDS-CASE = «родительный падеж»
  + Иначе если LT\_WORDS-CASE == «datv»
  + То:
    - LT\_WORDS-CASE = «дательный падеж»
  + Иначе если LT\_WORDS-CASE == «accs»
  + То:
    - LT\_WORDS-CASE = «винительный падеж»
  + Иначе если LT\_WORDS-CASE == «ablt»
  + То:
    - LT\_WORDS-CASE = «творительный падеж»
  + Иначе если LT\_WORDS-CASE == «loct»
  + То:
    - LT\_WORDS-CASE = «предложный падеж»
  + Иначе если LT\_WORDS-CASE == «voct»
  + То:
    - LT\_WORDS-CASE = «звательный падеж»
  + Иначе если LT\_WORDS-CASE == «gen1»
  + То:
    - LT\_WORDS-CASE = «первый родительный падеж»
  + Иначе если LT\_WORDS-CASE == «gen2»
  + То:
    - LT\_WORDS-CASE = «второй родительный (частичный) падеж»
  + Иначе если LT\_WORDS-CASE == «acc2»
  + То:
    - LT\_WORDS-CASE = «второй винительный падеж»
  + Иначе если LT\_WORDS-CASE == «loc1»
  + То:
    - LT\_WORDS-CASE = «первый предложный падеж»
  + Иначе если LT\_WORDS-CASE == «loc2»
  + То:
    - LT\_WORDS-CASE = «второй предложный (местный) падеж»
  + Иначе если LT\_WORDS-CASE == «0000»
  + То:
    - LT\_WORDS-CASE = «»
* Для поля LT\_WORDS-ANIMATION:
  + Если LT\_WORDS- ANIMATION == «anim»
  + То:
    - LT\_WORDS- ANIMATION = «одушевлённое»
  + Если LT\_WORDS- ANIMATION == «inan»
  + То:
    - LT\_WORDS- ANIMATION = «неодушевлённое»
  + Если LT\_WORDS- ANIMATION == «0000»
  + То:
    - LT\_WORDS- ANIMATION = «»
* Для поля LT\_WORDS-KIND:
  + Если LT\_WORDS- KIND == «masc»
  + То:
    - LT\_WORDS- KIND = «мужской род»
  + Если LT\_WORDS- KIND == «femn»
  + То:
    - LT\_WORDS- KIND = «женский род»
  + Если LT\_WORDS- KIND == «neut»
  + То:
    - LT\_WORDS- KIND = «средний род»
  + Если LT\_WORDS- KIND == «ms-f»
  + То:
    - LT\_WORDS- KIND = «общий род (м/ж)»
  + Если LT\_WORDS- KIND == «0000»
  + То:
    - LT\_WORDS- KIND = «»
* Для поля LT\_WORDS- MEASURE\_OF\_NUMBER:
  + Если LT\_WORDS- MEASURE\_OF\_NUMBER == «sing»
  + То:
    - LT\_WORDS-MEASURE\_OF\_NUMBER = «единственное число»
  + Если LT\_WORDS- MEASURE\_OF\_NUMBER == «plur»
  + То:
    - LT\_WORDS-MEASURE\_OF\_NUMBER = «множественное число»
  + Если LT\_WORDS- MEASURE\_OF\_NUMBER == «0000»
  + То:
    - LT\_WORDS-MEASURE\_OF\_NUMBER = «»
* Для поля LT\_WORDS- SPECIAL\_MARKS:
  + Если LT\_WORDS- SPECIAL\_MARKS == «Y»
  + То:
    - LT\_WORDS- SPECIAL\_MARKS = «разделённый дефис»
  + Если LT\_WORDS- SPECIAL\_MARKS == «0000»
  + То:
    - LT\_WORDS-SPECIAL\_MARKS = «»

Обнуляем счётчик:

* LV\_CURRENT\_POSITION\_NUMBER =0.

Создать переменную LV\_ORDER\_COUNT (INTEGER, 3) =0;

Создать переменную LV\_WORDS \_COUNT (INTEGER, 3) =количество записей в таблице LT\_WORDS;

Создать переменную LV\_ORDER\_STORAGE(INTEGER,3) =0;

Выполнить LV\_WORDS \_COUNT раз:

{

* Выбрать из LT\_WORDS запись по условию:
  + LT\_WORDS-WORD\_ID= LV\_CURRENT\_POSITION\_NUMBER
* Для полученной записи:
  + Если LT\_WORDS-ORDER != LV\_ORDER\_STORAGE
  + То:
    - Увеличиваем счётчик LV\_ORDER\_COUNT на 1.
    - Сохраняем значение ORDER
      * LV\_ORDER\_STORAGE= LT\_WORDS-ORDER
* Увеличить счётчик LV\_CURRENT\_POSITION\_NUMBER на 1.

}

Создать переменную LV\_ID\_CURRENT\_NUMBER(INTEGER, 3)=0;

Создать переменную LV\_ORDER\_CURRENT\_NUMBER(INTEGER, 3)=0;

Создать переменную LV\_ID\_STORAGE(CHAR, 8)= «»;

Создать переменную LV\_ID\_COUNT(INTEGER, 3)=0;

Создать переменную LV\_ID\_UNIQUE\_COUNT(INTEGER, 3)=0;

Выполнить LV\_ORDER\_COUNT раз

{

* Выбрать из LT\_WORDS запись по условию:
  + LT\_WORDS-ORDER= LV\_ORDER\_CURRENT\_NUMBER
* Сохранить полученные записи в таблице LT\_ ID\_PER\_ORDER:
  + LT\_ ID\_PER\_ORDER- ID\_PER\_ORDER = LT\_WORDS-LEMMA\_ID
* Обнулить счётчик LV\_ID\_CURRENT\_NUMBER;
* LV\_ID\_COUNT = количество записей в таблице LT\_ID\_PER\_ORDER
* Выполнить LV\_ID\_COUNT раз

{

* Выбрать из LT\_ID\_PER\_ORDER запись по условию:
  + LT\_ID\_PER\_ORDER-POSITION\_ID= LV\_ID\_CURRENT\_NUMBER
* Если LT\_ID\_PER\_ORDER - LEMMA\_ID!= LV\_ID\_STORAGE
* То:
  + Увеличиваем счётчик LV\_ID\_UNIQUE\_COUNT на 1.
  + Сохраняем значение ID
    - LV\_ID\_STORAGE= LT\_ID\_PER\_ORDER - LEMMA\_ID
  + Добавить в таблицу LT\_ LEMMAS\_PER\_ORDER запись по условию:
    - LT\_ LEMMAS\_PER\_ORDER-ORDER =LV\_ORDER\_CURRENT\_NUMBER
    - LT\_ LEMMAS\_PER\_ORDER-LEMMA\_ID =LV\_ID\_STORAGE
* Увеличить счётчик LV\_ID\_CURRENT\_NUMBER на 1.

}

}

Вызвать Z\_MD\_GENERATE\_COMBINATIONS «Создание комбинаций полученных слов». Условия запуска:

* IMP\_ LEMMAS\_PER\_ORDER = LT\_ LEMMAS\_PER\_ORDER

На выходе получить:

* IMP\_RESULT

Добавить в LT\_ RESULT\_LEMMAS\_ORDER все записи из таблицы IMP\_RESULT

Создать переменную LV \_POSITIONS\_COUNT (INT, 3)= Количество записей в LT\_WORDS.

Создать переменную LV \_INPUT\_WORD (CHAR, 40)= “”

Создать переменную LV \_INPUT\_WORD\_OLD (CHAR, 40)= “”

Создать переменную LV \_PART\_OF\_SPEECH (CHAR, 30)= “”

Создать переменную LV \_CASE (CHAR, 40)= “”

Создать переменную LV \_ANIMATION (CHAR, 15)= “”

Создать переменную LV \_KIND (CHAR, 15)= “”

Создать переменную LV \_MEASURE\_OF\_NUMBER (CHAR, 19)= “”

Создать переменную LV \_SPECIAL\_MARKS (CHAR, 3)= “”

Обнулим счётчик LV\_CURRENT\_POSITION\_NUMBER.

Выполнить LV \_POSITIONS\_COUNT раз

{

Выбрать запись из таблицы LT\_WORDS по условию:

* LT\_WORDS-POSITION\_ID= LV\_CURRENT\_POSITION\_NUMBER

Для полученной записи:

* LV \_INPUT\_WORD= LT\_WORDS-INPUT-WORD
* LV \_PART\_OF\_SPEECH= LT\_WORDS- PART\_OF\_SPEECH
* LV \_CASE= LT\_WORDS- CASE
* LV \_ANIMATION= LT\_WORDS- ANIMATION
* LV \_KIND= LT\_WORDS- KIND
* LV \_MEASURE\_OF\_NUMBER= LT\_WORDS- MEASURE\_OF\_NUMBER
* LV \_SPECIAL\_MARKS= LT\_WORDS- SPECIAL\_MARKS

Если

Если LV \_INPUT\_WORD == LV \_INPUT\_WORD\_OLD:

То:

* Вывести на экран: «/»+LV \_PART\_OF\_SPEECH+ « » + LV \_CASE+ « »+LV \_ANIMATION+ « » + LV \_KIND + « »+ LV \_MEASURE\_OF\_NUMBER « » + LV \_SPECIAL\_MARKS

Иначе если LV \_INPUT\_WORD != LV \_INPUT\_WORD\_OLD и LV\_CURRENT\_POSITION\_NUMBER!=0:

* Вывести на экран: «) »+ LV \_INPUT\_WORD + « (»+LV \_PART\_OF\_SPEECH+ « » + LV \_CASE+ « »+LV \_ANIMATION+ « » + LV \_KIND + « »+ LV \_MEASURE\_OF\_NUMBER + « » + LV \_SPECIAL\_MARKS

Иначе:

* Вывести на экран: LV \_INPUT\_WORD + « (»+LV \_PART\_OF\_SPEECH+ « » + LV \_CASE+ « »+LV \_ANIMATION+ « » + LV \_KIND + « »+ LV \_MEASURE\_OF\_NUMBER

}

Очистить переменную LV\_ID\_STORAGE.

Обнулить счётчик LV\_CURRENT\_POSITION\_NUMBER.

LV \_POSITIONS\_COUNT = количество записей в LT\_ RESULT\_LEMMAS\_ORDER

Вывести на экран «\n»

Выполнить LV \_POSITIONS\_COUNT раз

{

Выбрать запись из LT\_ RESULT\_LEMMAS\_ORDER по условию:

* LT\_ RESULT\_LEMMAS\_ORDER-POSITION\_ID= LV\_CURRENT\_POSITION\_NUMBER.

Сохранить полученную запись:

* LV\_ID\_STORAGE = LT\_ RESULT\_LEMMAS\_ORDER- LEMMA\_ID

Выбрать запись из LT\_WORDS по условию:

* LT\_WORDS- LEMMA\_ID = LV\_ID\_STORAGE

Вывести на экран полученную запись:

* “ ”+LT\_WORDS- LEMMA

}

Вывести на экран «|»

**Приложение**

Создать класс ZMD\_LEMMATIZER «Лемматизатор»

Создать в классе ZMD\_LEMMATIZER «Лемматизатор» метод Z\_MD\_SEARCH\_IN\_DICTIONARY «Поиск слова и его признаков в словаре».

**На входе:**

Переменная LV\_WORD (CHAR, 40)

Переменная LV\_WORD\_CHECK (CHAR, 3)

В методе Z\_MD\_SEARCH\_IN\_DICTIONARY предусмотреть следующий алгоритм действий:

Создать локальную таблицу LT\_ ROOTS \_OF\_DICTIONARIES «Пути словарей»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | DICTIONARY | Найденные словари | - | CHAR | 11 | - |

Создать локальную таблицу LT\_ EXTRACTED\_STRINGS «Полученные строки»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | STRING | Строки | - | CHAR | 4000 | - |

Создать локальную таблицу LT\_WORDS «Промежуточная таблица»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | WORD\_ID | Идентификатор слова | - | NUMC | 3 | Первичный ключ.  Счётчик. |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |
| 3 | INPUT\_WORD | Содержание слова | - | CHAR | 40 | - |
| 4 | PART\_OF\_SPEECH | Часть речи | - | CHAR | 4 | - |
| 5 | CASE | Падеж слова | - | CHAR | 4 | - |
| 6 | ANIMATION | Одушевленность/неодушевленность | - | CHAR | 4 | - |
| 7 | KIND | Пол слова | - | CHAR | 4 | - |
| 8 | MEASURE\_OF\_NUMBER | Мера численности | - | CHAR | 4 | - |

Добавить в LT\_ ROOTS \_OF\_DICTIONARIES запись по условию:

* LT\_ ROOTS \_OF\_DICTIONARIES- DICTIONARY = “lemmas.xml”

Если LV\_WORD содержит «е» или «ё»

То:

* Добавить в LT\_ ROOTS \_OF\_DICTIONARIES запись по условию:
  + LT\_ ROOTS \_OF\_DICTIONARIES- DICTIONARY = “lemmasE.xml”
* Заменить в переменной LV\_WORD все символы «ё» на «е»

Создать переменную LV\_DICTIONARIES\_COUNT (INT, 3) = количество записей в LT\_ ROOTS \_OF\_DICTIONARIES.

Создать переменную LV\_DICTIONARIES\_CURRENT\_NUMBER (INT, 3) = 0.

Создать переменную LV\_ CURRENT\_DICTIONARY (CHAR, 11) = 0.

Создать переменную LV\_STRINGS\_COUNT (INTEGER, 3) =0.

Создать переменную LV\_CURRENT\_STRING\_NUMBER (INT, 3) =0.

Создать переменную LV\_CURRENT\_STRING (CHAR, 4000) =0.

Создать переменную LV\_NUMBER\_OF\_MATCHES (INTEGER, 3) = 0.

Создать переменную LV\_WORDS\_COUNT (INTEGER, 3)=0.

LV\_WORD=«”» +LV\_WORD+ «”».

Выполнить LV\_DICTIONARIES\_COUNT раз:

{

Выбрать запись из LT\_ ROOTS \_OF\_DICTIONARIES по условию:

* LT\_ ROOTS \_OF\_DICTIONARIES-POSITION\_ID = LV\_DICTIONARIES\_CURRENT\_NUMBER

Сохраним значение полученного словаря:

* LV\_ CURRENT\_DICTIONARY= LT\_ ROOTS \_OF\_DICTIONARIES- DICTIONARY

Получить данные из файла, путь которого LV\_ CURRENT\_DICTIONARY по условию:

* Строка содержит LV\_WORD

Для полученных строк:

* Добавить в LT\_ EXTRACTED\_STRINGS запись по условию:
  + LT\_ EXTRACTED\_STRINGS- STRING = полученная строка

LV\_STRINGS\_COUNT=Количество записей в LT\_ EXTRACTED\_STRINGS.

Если LV\_STRINGS\_COUNT==0 и LV\_WORD содержит «-» и LV\_WORD\_CHECK== «Х»

То:

* **Дать на выход таблицу LT\_WORDS.**
* **Выйти из ФМ**

Иначе если LV\_STRINGS\_COUNT!=0 и LV\_WORD содержит «-» и LV\_WORD\_CHECK== «Х»

То:

* Добавить в LT\_WORDS запись:
  + LT\_WORDS-LEMMA\_ID= «0000»
  + LT\_WORDS-INPUT\_WORD= «0000»
  + LT\_WORDS-PART\_OF\_SPEECH= «0000»
  + LT\_WORDS-CASE= «0000»
  + LT\_WORDS-ANIMATION= «0000»
  + LT\_WORDS-KIND= «0000»
  + LT\_WORDS-MEASURE\_OF\_NUMBER= «0000»
* **Дать на выход таблицу LT\_WORDS.**
* **Выйти из ФМ**

LV\_CURRENT\_STRING\_NUMBER = 0.

Очистить LV\_WORD от символов «”».

Если LV\_STRINGS\_COUNT==0

То:

* Добавить в LT\_WORDS запись:
  + LT\_WORDS-LEMMA\_ID= «9999999»
  + LT\_WORDS-INPUT\_WORD= LV\_WORD
  + LT\_WORDS-PART\_OF\_SPEECH= «0000»
  + LT\_WORDS-CASE= «0000»
  + LT\_WORDS-ANIMATION= «0000»
  + LT\_WORDS-KIND= «0000»
  + LT\_WORDS-MEASURE\_OF\_NUMBER= «0000»
* **Дать на выход таблицу LT\_WORDS.**
* **Выйти из ФМ**

Выполнить LV\_STRINGS\_COUNT раз:

{

* Выбрать запись из LT\_ EXTRACTED\_STRINGS по условию:
  + LT\_ EXTRACTED\_STRINGS- POSITION\_ID = LV\_CURRENT\_STRING\_NUMBER
* Сохранить значение полученной строки:
  + LV\_CURRENT\_STRING= LT\_ EXTRACTED\_STRINGS- STRING
* LV\_NUMBER\_OF\_MATCHES = количество найденных LV\_WORD
* Выполнить LV\_NUMBER\_OF\_MATCHES раз:
  + Распарсить полученную запись по следующему условию:
  + Добавить запись в LT\_WORDS по следующему условию:
    - LT\_WORDS-LEMMA\_ID= lemma\_id-tag
    - LT\_WORDS-INPUT\_WORD= LV\_WORD
  + LV\_WORDS\_COUNT= количество записей в LT\_WORDS.
  + Выполнить для LV\_CURRENT\_MATCH\_NUMBER в полученной записи
    - Если part\_of\_speech-tag существует:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-PART\_OF\_SPEECH= part\_of\_speech-tag
    - Иначе:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-PART\_OF\_SPEECH = «0000»
    - Если case-tag существует:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-CASE= case-tag
    - Иначе:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-CASE= «0000»
    - Если animation-tag существует:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-ANIMATION= animation-tag
    - Иначе:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-ANIMATION= «0000»
    - Если kind-tag существует
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-KIND= kind-tag
    - Иначе:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-KIND= «0000»
    - Если measure\_of\_number\_tag существует:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-MEASURE\_OF\_NUMBER= measure\_of\_number\_tag
    - Иначе:
      * Выбрать в LT\_WORDS запись по условию:
        + LT\_WORDS-POSIYION\_ID = LV\_WORDS\_COUNT.
      * Изменить полученную запись:
        + LT\_WORDS-MEASURE\_OF\_NUMBER= «0000»
* Увеличить счётчик LV\_CURRENT\_STRING\_NUMBER на 1.

}

}

**Дать на выход таблицу LT\_WORDS.**

Создать в классе ZMD\_LEMMATIZER «Лемматизатор» метод Z\_MD\_PROCESS\_WORDS «Получение леммы слова»

**На входе:**

Переменная LV\_LEMMA\_ID(CHAR, 8).

В методе Z\_MD\_PROCESS\_WORDS предусмотреть следующий алгоритм действий:

Создать переменную LV\_LEMMA(CHAR, 40)=””.

Найти в файле «lemmas.xml» строку, в которой тэг lemma\_id-tag = LV\_LEMMA\_ID.

Выбрать в соответствующей строке тэг lemma-tag

LV\_LEMMA= lemma-tag.

**Дать на выход LV\_LEMMA.**

**Выйти из ФМ.**

Создать в классе ZMD\_LEMMATIZER «Лемматизатор» метод Z\_MD\_GENERATE\_COMBINATIONS «Создание комбинаций полученных слов»

**На входе:**

Таблица LT\_ LEMMAS\_PER\_ORDER «Количество уникальных лемм на порядковый номер»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | ORDER | Порядок слова в записи | - | INTEGER | 2 | - |
| 3 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

В методе Z\_MD\_GENERATE\_COMBINATIONS предусмотреть следующий алгоритм действий

Создать таблицу LT\_ CURRENT\_COMBINATION «Текущая комбинация»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Создать локальную таблицу LT\_ RESULT «Количество уникальных лемм на порядковый номер»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Вызвать ФМ Z\_MD\_GENERATE\_COMBINATIONS\_HELPER «Вспомогательная функция для создания комбинаций» Условия запуска:

* IMP\_LEMMAS\_PER\_ORDER = LT\_ LEMMAS\_PER\_ORDER
* IMP\_CURRENT\_POSITION\_NUMBER = LV\_CURRENT\_POSITION\_NUMBER
* IMP\_CURRENT\_COMBINATION = LT\_ CURRENT\_COMBINATION
* IMP\_RESULT = LT\_ RESULT

На выходе:

* **Дать на выход IMP\_RESULT.**
* **Выйти из ФМ.**

Создать в классе ZMD\_LEMMATIZER «Лемматизатор» метод Z\_MD\_GENERATE\_COMBINATIONS\_HELPER «Вспомогательная функция для создания комбинаций»

**На входе:**

Таблица LT\_ LEMMAS\_PER\_ORDER «Количество уникальных лемм на порядковый номер»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | ORDER | Порядок слова в записи | - | INTEGER | 2 | - |
| 3 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Переменная LV\_CURRENT\_POSITION\_NUMBER (INT, 3)=0;

Таблица LT\_ CURRENT\_COMBINATION «Текущая комбинация»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Таблица LT\_ RESULT «Количество уникальных лемм на порядковый номер»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

В методе Z\_MD\_GENERATE\_COMBINATIONS\_HELPER предусмотреть следующий алгоритм действий:

Создать локальную таблицу LT\_NEW\_COMBINATION «Промежуточная таблица комбинаций»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№ п/п** | **Поле** | **Краткое описание** | **Элемент данных** | **Тип данных** | **Длина** | **Комментарий** |
| 1 | POSITION\_ID | Идентификатор позиции | - | NUMC | 3 | Первичный ключ. Счетчик |
| 2 | LEMMA\_ID | Id-индекс леммы | - | CHAR | 8 | - |

Вызвать ФМ Z\_MD\_GENERATE\_COMBINATIONS\_HELPER. Условия запуска:

* IMP\_ LEMMAS\_PER\_ORDER = LT\_ LEMMAS\_PER\_ORDER

Создать переменную LV\_LEMMAS\_PER\_ORDER\_COUNT (INTEGER, 3)= количество записей в LT\_ LEMMAS\_PER\_ORDER.

Если LV\_CURRENT\_POSITION\_NUMBER == LV\_LEMMAS\_PER\_ORDER\_COUNT

То:

* Добавить в LT\_ RESULT все записи из таблицы LT\_ CURRENT\_COMBINATION
* **Дать LT\_ RESULT на выход**
* **Выйти из ФМ.**

Выбрать записи из LV\_LEMMAS\_PER\_ORDER\_COUNT по условию:

* LV\_LEMMAS\_PER\_ORDER\_COUNT-ORDER= LV\_CURRENT\_POSITION\_NUMBER

Выполнить для каждой полученной записи:

* Очистить LT\_NEW\_COMBINATION
* Добавить в LT\_NEW\_COMBINATION все записи из LT\_CURRENT\_COMBINATION
* Вызвать ФМ Z\_MD\_GENERATE\_COMBINATIONS\_HELPER «Вспомогательная функция для создания комбинаций» Условия запуска:
  + IMP\_LEMMAS\_PER\_ORDER = LT\_ LEMMAS\_PER\_ORDER
  + IMP\_CURRENT\_POSITION\_NUMBER = LV\_CURRENT\_POSITION\_NUMBER+1
  + IMP\_CURRENT\_COMBINATION = LT\_NEW\_COMBINATION
  + IMP\_RESULT = LT\_ RESULT