

EL'MANUSCRIPT 2016: ШКОЛА

Расширенный поиск в НКРЯ

А.Ю. Станкевич

Гродненский государственный университет
им. Янки Купалы

Примеры запросов к РМ(Г)*: часть 1

*Корпусу СМИ Гродненщины в составе
модуля Региональной и зарубежной
прессы НКРЯ:

<http://www.ruscorpora.ru/search-regional.html>

**А. Использование масок и
оператора ИЛИ при
определении подкорпуса**

А.1. Определение подкорпуса текстов с заголовками, содержащими **любую словоформу указанной в фильтре леммы**

Основные параметры текста ?

Название

пособие

Синтаксис фильтра: ***лемма***
(без пробелов справа / слева от леммы).

лемма VS *словоформа*

пособие SYN *пособию*

пособие SYN *пособиям*

!ОДНАКО:

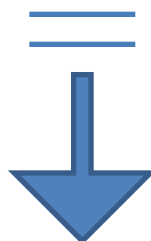
Результаты по фильтру ***семья*** \subset результаты по фильтру ***семью***

Фильтр ***семью*** подкорпус текстов, заголовки которых содержат словоформы леммы **семь** и / или словоформы леммы **семья**

Основные параметры текста ?

Название

пособие



Основные параметры текста ?

Название

пособиям

Результаты поиска (β)

[перейти на страницу поиска](#) [выбрать подкорпус](#) [версия без ударений](#) [настройки](#)

Объем всего корпуса: 27 982 документа, 1 085 671 предложение, 13 243 402 слова.

Найдено 38 документов общим объемом 1 422 предложения, 17 937 слов.

Сохранить подкорпус и перейти к странице поиска

Страницы: 1

1. [С начала 2013 года в Кирове предотвратили 28 попыток получить пособия по безработице обманным путем \(2013.04.27 не снята\)](#)
2. [БелаТА. Детские пособия на детей с января возрастут вдвое \(2012.12.19\) \[омонимия не снята\]](#)
3. [коллективный автор. В следующем году повысятся и детские пособия, и пенсии \(2012.12.12\) \[омонимия не снята\]](#)
4. [Мария Драпеза. Как исчисляется пособие по временной нетрудоспособности ? \(2012.11.02\) \[омонимия не снята\]](#)
5. [БелТА. Пособия и пенсии подрастут в ноябре \(2012.10.31\) \[омонимия не снята\]](#)
6. [Служба новостей «ВГ». Вырастут пенсии и детские пособия \(2012.10.29\) \[омонимия не снята\]](#)
7. [коллективный автор. Вырастут пенсии и детские пособия/Комментарии \(2012.10.29\) \[омонимия не снята\]](#)

<...>

А.2. Определение подкорпуса текстов с заголовками, содержащими **указанную в фильтре словоформу**

Основные параметры текста

Название

пособий

Синтаксис фильтра: **словоформа**
(без пробелов справа / слева от
леммы).

SYN

- фильтр "**словоформа**";
- фильтр *"**словоформа**";
- фильтр * **словоформа** * (с пробелами справа / слева от словоформы).

А.3. Определение подкорпуса текстов с заголовками, содержащими любую словоформу леммы₁, леммы₂, ..., леммы_i

Основные параметры текста

Название

кирмаш | *ярмарка*



<...>

Синтаксис фильтра:

лемма₁ | *лемма₂* | ... | *лемма_i*
(без пробелов справа / слева от леммы).

! Здесь и далее: с обязательными пробелами справа/слева от вертикального слеша!

<...>

/ кирмаш И/ИЛИ ярмарка

А.4. Определение подкорпуса текстов с заголовками, содержащими любую словоформу из указанных в фильтре словоформы₁, словоформы₂, ..., словоформы_i

Синтаксис фильтра:

словоформа₁ | словоформа₂ | ... | словоформа_i.

(без пробелов справа / слева от леммы).

SYN

- фильтр **"словоформа₁" | "словоформа₂" | ... | "словоформа_i «**
- фильтр ***"словоформа₁"* | *"словоформа₂"* | ... | *"словоформа_i"***
- фильтр *** словоформа₁ * | * словоформа₂ * | ... | * словоформа_i ***
(с пробелами справа / слева от словоформы).

Пример к А.4

Основные параметры текста ?

Название

кирмаш | ярмарка



<...>

11. Кирмаш без коней - не кирмаш (2012.09.19) [омонимия не снята]
12. Анненская ярмарка стала главным культурным мероприятием года в Зельве (2012.09.10) [омон
13. Ярмарка деловая (2012.09.07) [омонимия не снята]
14. обобщенный автор. В Гродно пройдет выставка-ярмарка «Еврорегион «Неман-2012» (2012.09.05)
15. Анненская ярмарка: листовая страницы (2012.08.31) [омонимия не снята]
16. В Зельве провели кирмаш (2012.08.29) [омонимия не снята]

<...>

/ кирмаш И/ИЛИ ярмарка

А.5. Определение подкорпуса текстов с заголовками, содержащими конструкцию из подряд* идущих элементов

ЭЛЕМЕНТ= лемма

Синтаксис фильтра:

* x_1 * x_2 *...* x_i *

Основные параметры текста ?

Название

*вырасти*пособие*



1. Служба новостей «ВГ». Вырастут пенсии и детские пособия (2012.10.29) [омонимия не снята]
2. коллективный автор. Вырастут пенсии и детские пособия/Комментарии (2012.10.29) [омонимия не снята]

<...>

ЭЛЕМЕНТ= словоформа

Синтаксис фильтра:

" x_1 "" x_2 "*...*" x_i "*

Смешанный фильтр:

вырасти"пособие"*

SYN

вырасти"пособие"

Основные параметры текста ?

Название

вырасти"пособие"*



1. Елена ТОЛКАЧЕВА. В Беларуси выросло пособие на детей (2011.05.05) [омонимия не снята]

**A.5'. Определение подкорпуса текстов с заголовками, содержащими
все элементы списка**

/ 2 элемента-леммы

Синтаксис фильтра:

$*x_1*x_2* \mid *x_2*x_1*$

И т.п.

**!Однако: число частей, объединенных знаком | :
 $P_n=n!$**

n	P_n
2	2
3	6
4	24
5	120
6	720

А.6. Определение подкорпуса текстов с заголовками, содержащими несколько конструкций из подряд идущих элементов

детский*пособие* | *пособие*на*ребенок*

SYN

*детский*пособие* | *пособие*на*"детей"* | *пособие*на*"ребенка"*



Имена текстов подкорпуса:

Детские пособия на детей с января возрастут вдвое (ОП); Вырастут пенсии и детские пособия (ВГ);

«Имеем ли право на пособие на детей старше 3 лет, если муж служит в армии?» (БГ);

Увеличились размеры пособий на детей (СГ); Районный отдел облуправления Фонда социальной защиты населения.

Пособия из средств государственного социального страхования, на которые имеет право семья, воспитывающая ребенка-инвалида (БГ);

Пособие на ребенка увеличили на 133 рубля (КП) и др.

! подряд* идущих элементов

При создании маски для словосочетания компоненты маски в запросе стоит набирать в соответствии с порядком слов в искомом словосочетании, однако стоит учесть, что потенциально в заголовке найдется не **ряд**, а **мешок** (неупорядоченное множество) компонентов маски. См. фрагмент выдачи заголовков на запрос ***Беларусь*Россия*** (словоформы лемм **Беларусь** и

1. Берестовицкий районный Совет депутатов, Берестовицкий районный исполнительный комитет. 2 апреля – День единения народов Беларуси и России // «Бераставіцкая газета», 2012.04.02 [омонимия не снята]
2. Россия не хочет кредитовать Беларусь из своего кармана // «Комсомольская правда» в Беларуси, 2011.05.11 [омонимия не снята]

Таким образом, в выдачу по запросу ***вырасти*пособие*** могли попасть и заголовки с инвертированным порядком слов в искомом словосочетании (например, *С нового года **пособия** на детей **выросли***). Кроме того, нужно учесть, что при отсутствии сильной лексико-грамматической связи между элементами-основами маски между элементами, найденными в заголовке, может отсутствовать непосредственная синтаксическая связь.

Б. Использование масок и операторов в запросе

!

- Состав маски: буквенная часть и знак * (* = '0 и более символов словоформы/леммы')

Ср. при задании подкорпуса:

(* = '0 и более словоформ и/или разделителей и/или пробельных элементов')

- Оператор **НЕ**: знак минуса
- Оператор **ИЛИ**: вертикальный слеш (! пробелы справа/слева)
- Прямой ввод словоформы в поле СЛОВО: кавычки («бегу»)

Б.1. Поиск словоформ леммы, начинающейся на х

Синтаксис запроса: **х***

/ Пример-1 запроса: **пилигрим***

В выдаче: контексты со словоформами лемм *пилигримка, пилигрим*.

/ Пример-2 запроса (+грамм. метка):

В выдаче:

контексты с глаголами на *противо-*

Слово ? <input type="text" value="А Б В"/>	Грамм. признаки ?
<input type="text" value="противо*"/>	<input type="text" value="V"/>



поддержку друг друга, мы сможем **противостоять**

Однако это вовсе не **противоречит**

Противопоставим

Вместе **противостоять**

ератрицы Екатерины II и активно **противостоял**

Гости **противопоставили**

любым неурядицам.

тому, что при необходимости колокольня

снегопаду слаженность наших действий Вчера

правонарушениям «Прямая линия» с начальником

попыткам реформ в умиравшей Речи

хозяевам лишь один голевой удар

(Б.2. Поиск словоформы, начинающейся на х)

Синтаксис запроса: "х*"

Пример запроса: "пилигримк*".

В выдаче: контексты со словоформами на *пилигримк-* (но не *пилигримок-* и др.).

Тип запроса малоприменим.

Б.3. Поиск словоформ леммы, заканчивающейся на х

Синтаксис запроса: *х.

Пример запроса: *ировка.

В выдаче: контексты со словоформами лемм *антиблокировка*, *бомбардировка*, *грейдеровка* и др.

Запросы такого типа удобны для поиска грамматических подклассов.

Б.4. Поиск словоформ, заканчивающихся на х

Синтаксис запроса: **"*х"**

Пример запроса: **"*остию"**

В выдаче: контексты со словоформами *крепостию*, *милостию*, *ревностию* и др.

Б.5. Поиск словоформ с конфиксом/конфиксоидом

Синтаксис запроса: $x_1 * x_2$

Пример запроса: недо*ся

В выдаче: контексты для лемм *недооткрыться*,
недосчитаться, *недопоститься* и др.

Б.6. Поиск с исключением элемента *x*

Синтаксис запроса: **-*x***

Исключаемым элементом (*x*) может быть лемма, словоформа, тег грамматической, дополнительной, семантической разметки

Пример-1 запроса:

велo* -велосипед -велосипедный -велосипедист*

В выдаче: контексты со словоформами лемм, начинающихся на *велo-*, кроме лемм *велосипед*, *велосипедный*, *велосипедист* и *велосипедистка* (и любых других лемм, начинающихся на *велосипедист-*)

!

В запросах синтаксис типа *x* не поддерживается.

В фильтре имени текста (то есть при задании подкорпуса) такой синтаксис поддерживается.

!

Ожидается изменение интерфейса НКРЯ (см. здесь: http://www.ruscorpora.ru/new_design.html)

Примеры запросов к РМ(Г): часть 2

Ex-1

Слово ? <input type="text" value="А"/> <input type="text" value="Б"/> <input type="text" value="В"/> <input type="text" value="оказывать оказать"/>	Грамм. признаки <input type="text" value="v"/>
Доп. признаки ? выбрать <input type="text" value="-bmark"/>	Семант. признаки <input type="text"/>

Расстояние: от до

Слово ? <input type="text" value="А"/> <input type="text" value="Б"/> <input type="text" value="В"/> <input type="text" value="-помощь"/>	Грамм. признаки <input type="text" value="S,acc"/>
---	--

Ex-2. Поиск ДС структуры ЖИТЕЛЬ | ЖИТЕЛЬНИЦА + топоним (S, Gen)

Слово ? <input type="text" value="А Б В"/> <input type="text" value="житель жительница"/>	Грамм. признаки ? <input type="text"/>
Доп. признаки ? выбрать <input type="text" value="-bmark"/>	Семант. признаки ? <input type="text"/>

Расстояние: от до ?

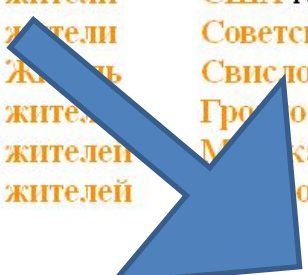
Слово ? <input type="text" value="А Б В"/> <input type="text"/>	Грамм. признаки ? <input type="text" value="S"/>
Доп. признаки ? выбрать <input type="text"/>	Семант. признаки ? <input type="text" value="r:propn & t:topon"/>

Точнее: S, Gen

Ex-2. Результат (KWIC)

Страницы: [1](#) [2](#) [3](#) [4](#) [следующая страница](#)

грустное — ушла из жизни старейшая **жительница** **Нового** Света, Тамара Иосифовна Тишкова ←...→
Житель **Нового** Двора Иван Владимирович Савко ←...→
Жителя **Свислочи** интересовал вопрос, касающийся охраны ←...→
– В прошлом месяце **жители** **Свислочи** вместе со счетом-извещением ←...→
Например, многие **жители** **США** готовятся к этому дню ←...→
по недавней истории великого государства, **жители** **Советска** используют неприхотливое название своего ←...→
жителя, нарушающего трудовое законодательство? **Житель** **Свислочи**. – В первую очередь, как ←...→
тели власти, руководители крупных предприятий, **жители** **Гродно**. ←...→
Для **жителей** **Минска**, конечно, удобно, а вот ←...→
только эта «начинка» здания для **жителей** **Свислочи**, которые с нетерпением ждали ←...→



Житель **Нового** Двора И
Жителя **Свислочи** интер
Свислочи вмес
США готовятся
Советска испол
Свислочи. – В
Гродно. ←...→
Минска, конечн
Свислочи, кото

Ex-3. СВОБОДА + И + СУЩЕСТВИТЕЛЬНОЕ

Слово ? <input type="text" value="А Б В"/>	Грамм. г
<input type="text" value="свобода"/>	<input type="text"/>
Доп. признаки ? выбрать	Семант.
<input type="text" value="-bmark"/>	<input type="text"/>

Расстояние: от до ?

Слово ? <input type="text" value="А Б В"/>	Грамм. г
<input type="text" value="и"/>	<input type="text"/>
Доп. признаки ? выбрать	Семант.
<input type="text" value="-bmark"/>	<input type="text"/>

Расстояние: от до ?

Слово ? <input type="text" value="А Б В"/>	Грамм. г
<input type="text"/>	<input type="text" value="S"/>
Доп. признаки ? выбрать	Семант.
<input type="text"/>	<input type="text"/>

Ex-3. Результат

свобода -bmark

на расстоянии 1 от и -bmark

на расстоянии от 1 до 2 от \$

Найдено 7 документов, 8 вхождений.

об известных случаях нарушения прав, **свобод** и законных **интересов** детей; о ←...→
кто собственной жизнью заплатил за **свободу** и **независимость** Родины. ←...→
Отечества свои жизни, подарил нам **свободу** и **независимость**, говорили выступающие на ←...→
штамп в паспорте лишает человека **свободы** и **возможности** быть счастливым. ←...→
Штамп в паспорте лишает человека **свободы** и **возможности** быть счастливым." ←...→
узнают о героях, сражавшихся за **свободу** и **счастье** людей. ←...→
спожил голову в сражениях за **свободу** и **независимость** нашей Родины! ←...→
самоотверженность, воинская доблесть солдат, отстаивших **свободу** и **независимость**. ←...→



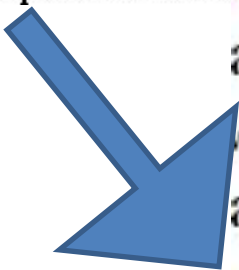
з, **свобод** и законных **интересов** детей; о ←...→
а **свободу** и **независимость** Родины. ←...→
м **свободу** и **независимость**, говорили высту
а **свободы** и **возможности** быть счастливым
а **свободы** и **возможности** быть счастливым
а **свободу** и **счастье** людей. ←...→
а **свободу** и **независимость** нашей Родины!
х **свободу** и **независимость**. ←...→

Ex-3' (be). Результат

- **свабода -bmark**
- на расстоіанні 1 от і -bmark
- на расстоіанні от 1 до 2 от S

Найдено 9 вхождений.

самаахвярными барацьбітамі за ўсталяванне **свабоды і дэмакратыі** ў арабскім свеце ←...→
хто палажыў сваё жьццё за **свабоду і незалежнасць** Беларусі. ←...→
ахвяр і разрушэнняў дасталіся нам **свабода і незалежнасць**, колькі жьццяў пакладзена ±
дзе выбіты імёны загінуўшых за **свабоду і незалежнасць** Радзімы землякоў, сабраліся
Беларусі ад нямецка-фашысцкай навалы, бо **свабода і незалежнасць** нашай краіны — непадзельны
для кожнага з нас паняцці — **Свабода і Незалежнасць**. ←...→
за наша мірнае жьццё, за **свабоду і незалежнасць** Радзімы праграмамі зашты ±
каленняў абаронцаў Айчыны, што адста
з'яўляюцца выдатным узорам змагання **свабоды і дэмакратыі** ў арабскім свеце
а **свабоду і незалежнасць** Беларусі. ←...→
і **свабода і незалежнасць**, колькі жьццяў
а **свабоду і незалежнасць** Радзімы земля
о **свабода і незалежнасць** нашай краіны -
— **Свабода і Незалежнасць**. ←...→
а **свабоду і незалежнасць** Радзімы прагр
і **свабоду і незалежнасць** роднай зямлі.
а **свабоду і незалежнасць**. ←...→



Ex-4. ПРАВО+ И + существительное

- -водительский
- на расстоянии 1 от право -bmark
- на расстоянии 1 от и -bmark
- на расстоянии от 1 до 2 от S

Найдено 44 вхождения.

прекращаются **определенные права и обязанности.** ←...→
е связывалось **никаких прав и обязанностей** для супругов ←...→
гоняли отходы **на право и на лево!** ←...→
обеспечивает **защиту прав и законных интересов** граждан ←...→
равительство и **защиту прав и интересов** определенной социальной
На **защите права и справедливости** Председатель суда
енство закона, **защищать права и свободы** граждан, обеспечивать ←...→
ают на защиту **своих прав и законных интересов**, здесь ←...→
зять на защиту **своих прав и безусловное выполнение** норм ←...→
обеспечивают **соблюдение права и установление** истины. ←...→
ете надежную **защиту прав и законных интересов** граждан ←...→
юза офицеров **отстаивают права и интересы** офицеров запаса ←...→
лучить защиту **своих прав и законных интересов.** ←...→
х обеспечения **защиты прав и законных интересов** детей ←...→
р по вопросам **защиты прав и интересов** детей прошел ←...→
кб с семьей по **защите прав и законных интересов** детей ←...→
осуществлять **защиту прав и законных интересов** детей ←...→
ую очередь на **соблюдение прав и интересов** детей. ←...→

Ex-4' (be).

- права -bmark
- на расстоянии 1 от i -bmark
- на расстоянии от 1 до 2 от S

Найдено 15 вхождений.

праў і **інтарэсаў** грамадзян, паколькі дакументы
правоў і **свабод** грамадзян, эканамічных пераўтвар
правоў і **свабод** грамадзян, а таксама $\leftarrow \dots \rightarrow$
правоў і законных **інтарэсаў**, не змяншаецца $\leftarrow \dots \rightarrow$
Правоў і **Абавязкаў**, Пушкіна, Асіпка, Турыстычна:
правах і **гарантыях**, якія ім забяспечвае $\leftarrow \dots \rightarrow$
правы і **абавязкі** грамадзян Рэспублікі Беларусь
правамі і **абавязкамі**, з гэтага моманту $\leftarrow \dots \rightarrow$
правы і **свабоды**, і няўхільна выконваць $\leftarrow \dots \rightarrow$
правы і **гарантыі**, якія дае ім $\leftarrow \dots \rightarrow$
правоў і законных **інтарэсаў** непаўналетніх, якія \leftarrow
правоў і законных **інтарэсаў** непаўналетніх Артура
правах і **абавязках** інспектараў грамадскага кантр
правоў і **інтарэсаў** людзей. $\leftarrow \dots \rightarrow$

Ех-5а. Канцелярит

"вопросы"

на расстоянии от 1 до 2 от N, s, Gen

на расстоянии от 1 до 2 от N, s, Gen

Найдено 28 вхождений.

Обсуждались **вопросы** организации **условий** труда, **поощрения** рабочих ←...→

необходимо на строгом контроле держать **вопросы** **кормления**, **поения** и содержания животных ←...→

пристальным его вниманием находятся и **вопросы** **улучшения** бытовых **условий** работников. ←...→

Все **вопросы** организации **питания** **детей** обсуждаются на ←...→

Особо были подчеркнуты **вопросы** неукоснительного **соблюдения** **правил** пожарной безопасности
в ОАО «Хоневичи» будут обсуждены **вопросы** **окончания** **строительства** объекта. ←...→

Затронуты **вопросы** **кормления** **животных**, организации пастьбы, благоустройства ±

В основном **вопросы** касались **разъяснений** **требований** трудового законодательства

населенных пунктах, а также на **вопросы** **обеспечения** **населения** топливом. ←...→

местах массового отдыха населения, **вопросы** **проведения** профилактических **мероприятий** по безопасности ±

Ex-5b.

- ***ОСТЬ**

на расстоянии от 1 до 3 от **составить** | **составлять**
Найдено 47 вхождений.

Товарность молока **составляет** 90%. Председатель ра

Стоимость уникальной ткани **составляет** полмиллис

безопасности, приоритетная **составляющая** народных

урожайность **составила** 19,8 ц/га. ←...→

успеваемости учащихся **составил** 7,45. ←...→

стоимость одного прыжка **составляла** 380 000 ←...→

Мощность Островецкой АЭС **составит** 2,4 тысячи

урожайность **составила** 50,5 ц/га. ←...→

урожайность по району **составляет** 41,8 ц ←...→

преступности основную массу **составляют** кражи – 47

Ех-5с

по

на расстоянии 1 от "**линии**"

Найдено 19 вхождений.

*Анализ оперативной обстановки **по линии** фальшивомонетничества за последние годы свидетельствует ...*

*В семье говорят, что любовь к физкультуре у Даши от отца. Ведь он – неоднократный участник различных соревнований **по линии** районного узла электросвязи, где работает электромехаником.*

Ex-5d

- **труженик**

на расстоянии **1** от **gen**

Найдено 43 документа, 53 вхождения.

*В упорной борьбе первое место завоевали теннисисты из РУП ЖКХ, второе — из ОАО «Акр-Агро», а «бронзовыми» призерами стали **труженики УСП «Совхоз «Порозовский»**.*

*От имени руководителя открытого акционерного общества и его 11-тысячного коллектива **тружеников сельского хозяйства Ивьевщины** <...>*

*Желаю вам, уважаемые **труженики сельхозотрасли Ивьевщины** <...>*

*Нынче **труженики аграрной** отрасли района праздновали <...>*

*Затронув тему прошедшего профессионального праздника **тружеников сельского хозяйства** <...>*

***Тружеников села** всегда отличали природная мудрость <...>*

*Все **труженики** поощены доброй славой и формами их*

Ex-6. Отсев омонимичных форм

<p>Слово ? <input type="text" value="А"/> <input type="text" value="Б"/> <input type="text" value="В"/></p> <p><input type="text" value="межа -" меж"=""/></p>	<p>Грамм. признаки ? выбрать</p> <p><input type="text"/></p>
--	--

Ex-7. Поиск элементов словообразовательного гнезда (смешанная маска)

Слово ?

мяжа | замежн* | замежны | абмежава* | абмяжоў* | абмежак | адмежава* | адм.

Грамм. признаки ? [выбрать](#)

! При заданном подкорпусе (язык=белорусский)

Представленность элементов словообразовательного гнезда лексемы

мяжа

25 лексем, **503** с/у ^в РМ(Г): **ВЕ**

мяжа (193; 242,2),
замежны (159; 199,6),
абмежаванне (32; 40,2),
абмежаваць (21; 26,4);
замежжа (18; 22,6);
памежжа (15; 18,8);
бязмежны (12; 15,1);
абмяжоўвацца (11; 13,8);
прамежак (7; 8,8);
абмяжоўваць (6; 7,5);
абмежавацца (4; 5);
неабмежаваны (4; 5);
смежны / *сумежны* (4; 5);

Спасибо за внимание:)