

BiByte Butterfly 3.0

Beta 2



User Manual

How to find (quick start)

1. You select the appropriate scan profile, such as «Yandex»:

Источник ссылок:

Поиск ссылок Загрузка ссылок из файла

Профиль поиска: Yandex 2011-12-05 [icon] [icon]

2. Enter the keyword phrases separated by commas or new lines:

Ключевые слова:

Ключевые слова Загрузка ключевых слов из файла

Музыка mp3, Скачать музыку, mp3, фильмы, скачать фильмы

3. You select the appropriate profile filter (to filter results found):

Фильтр:

Профиль фильтра: Отбор Drupal 2011-12-07 [↕] [...]

4. Specify where to save the sites found:

Результаты:

Сохранить как: D:/results.txt [...]

5. Click the button:

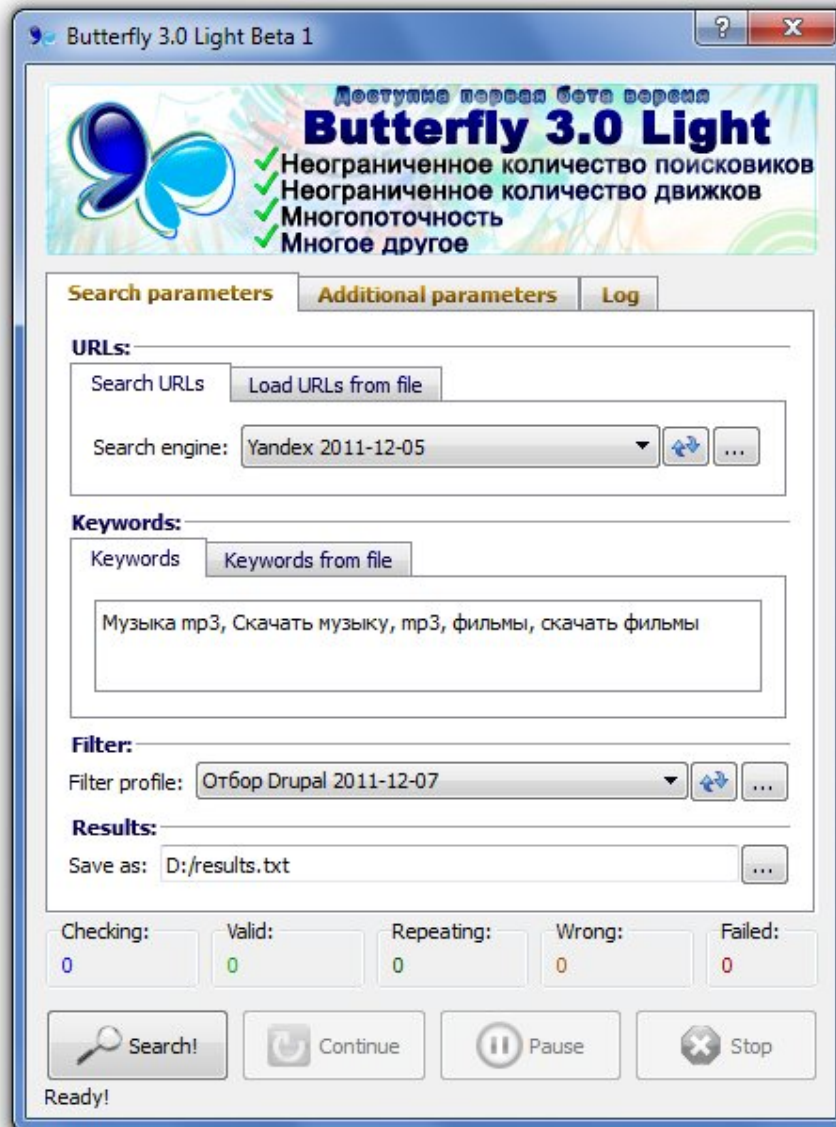


The program will begin the search. All found the results will be stored in the file specified.

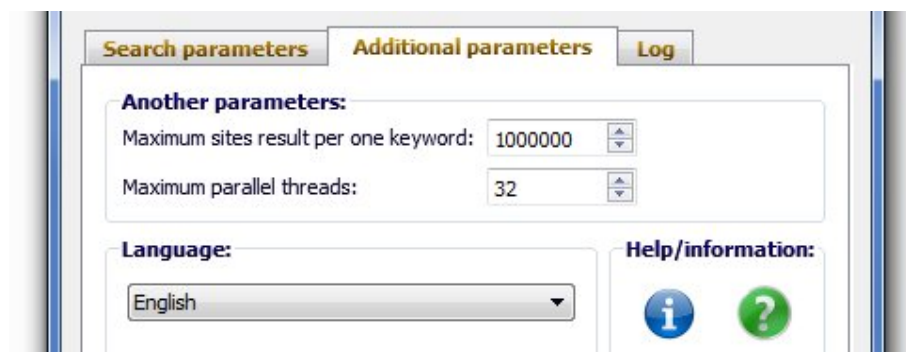
About the Program

The program is designed to parse the search results of search engines, directories, pages of the site with an arbitrary selection of links.

After starting the program opens the main window:

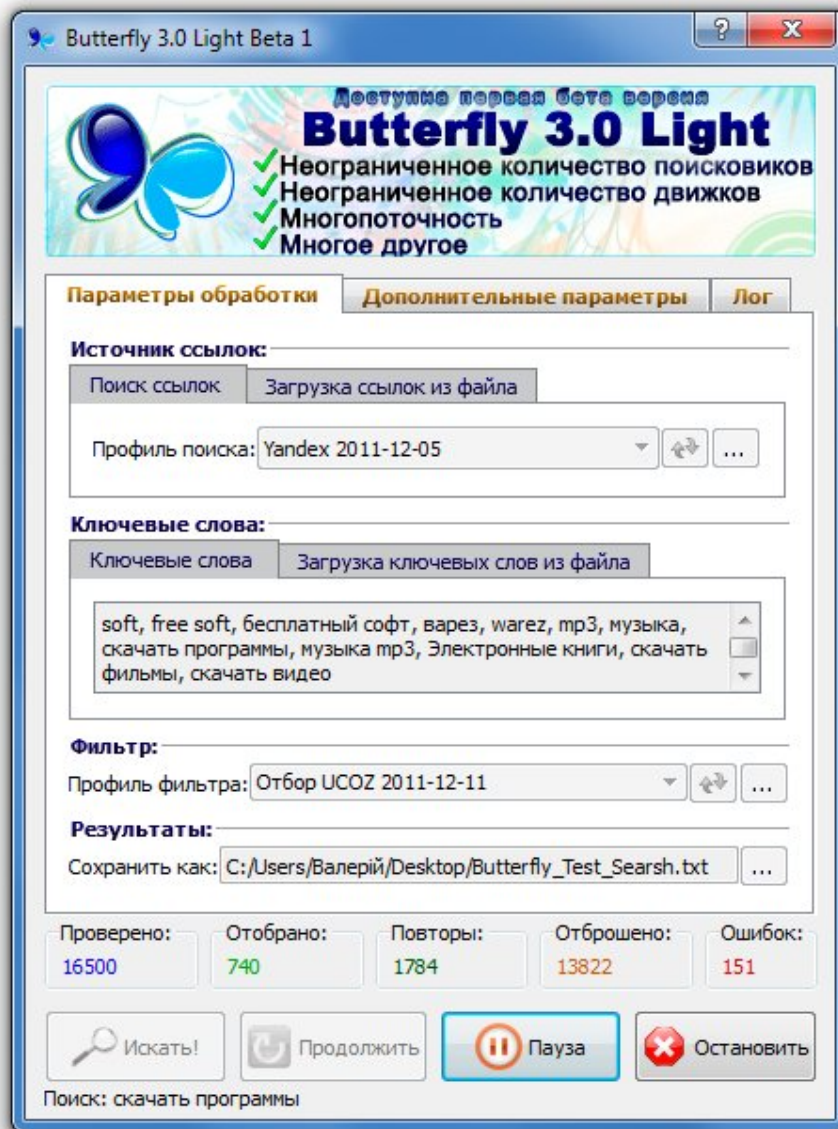


The main part of the window is divided into three tabs. Immediately you can go to the tab «Additional parameters» and switch the interface language:



After that, you must restart the program.

The main window of the Russian interface is as follows:



In the "Processing Settings" are filled with all the necessary data for the search.

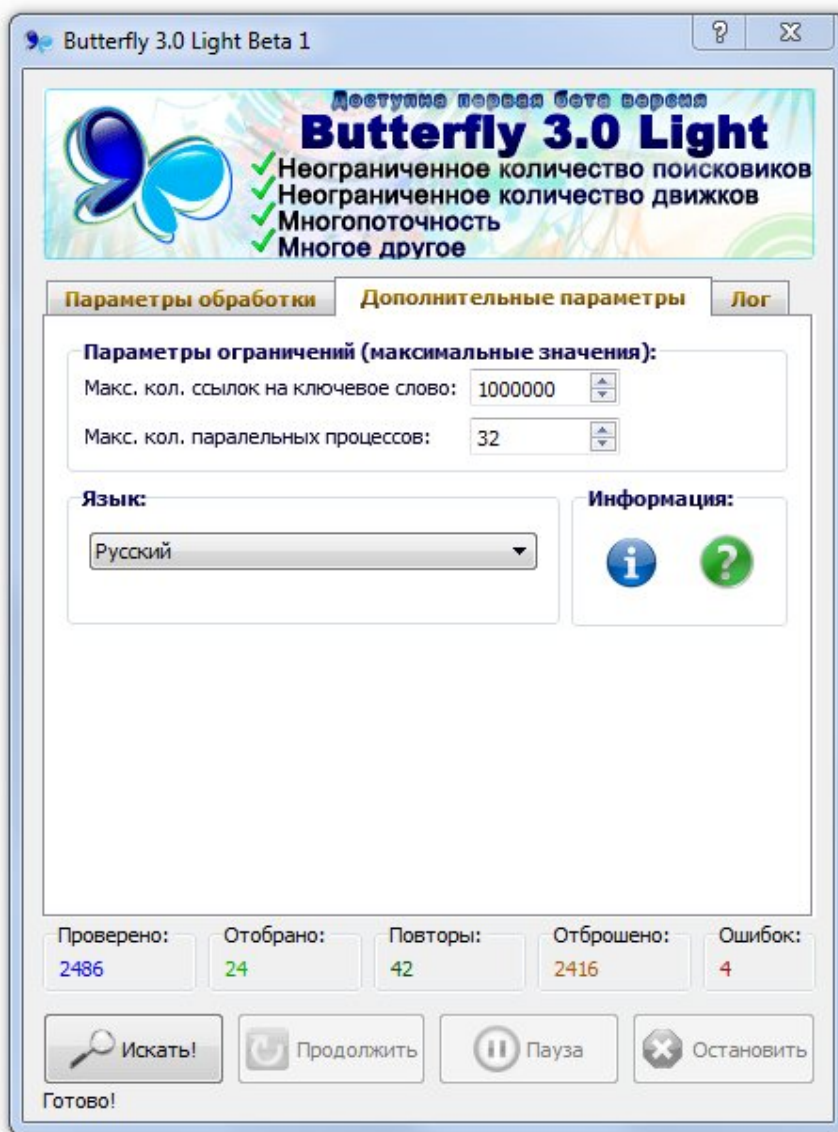
Source of reference - the source from which the program will receive a link for verification. It's either a search engine results, or specify a file ready reference. If the source has to be search engine - selects the appropriate scan profile.

In the Keywords box enter key phrases, separated by a comma or a new line. Also, keywords can be downloaded from the file. To do this, select the tab "Load keywords from file" and specify the appropriate file.

In the Filter field specifies an additional filter, which will be used to verify each site, derived from the source. To not use a filter, you need to select list "filter profile" select "No Selection".

In the "Results" is selected the file, which will be stored and filtered sites found. Once the program has found another site and filtered, it is immediately written to the output file, so you can not worry about that when the emergency close the program results will be lost.

In the "Additional parameters" are some other options that are used less frequently:



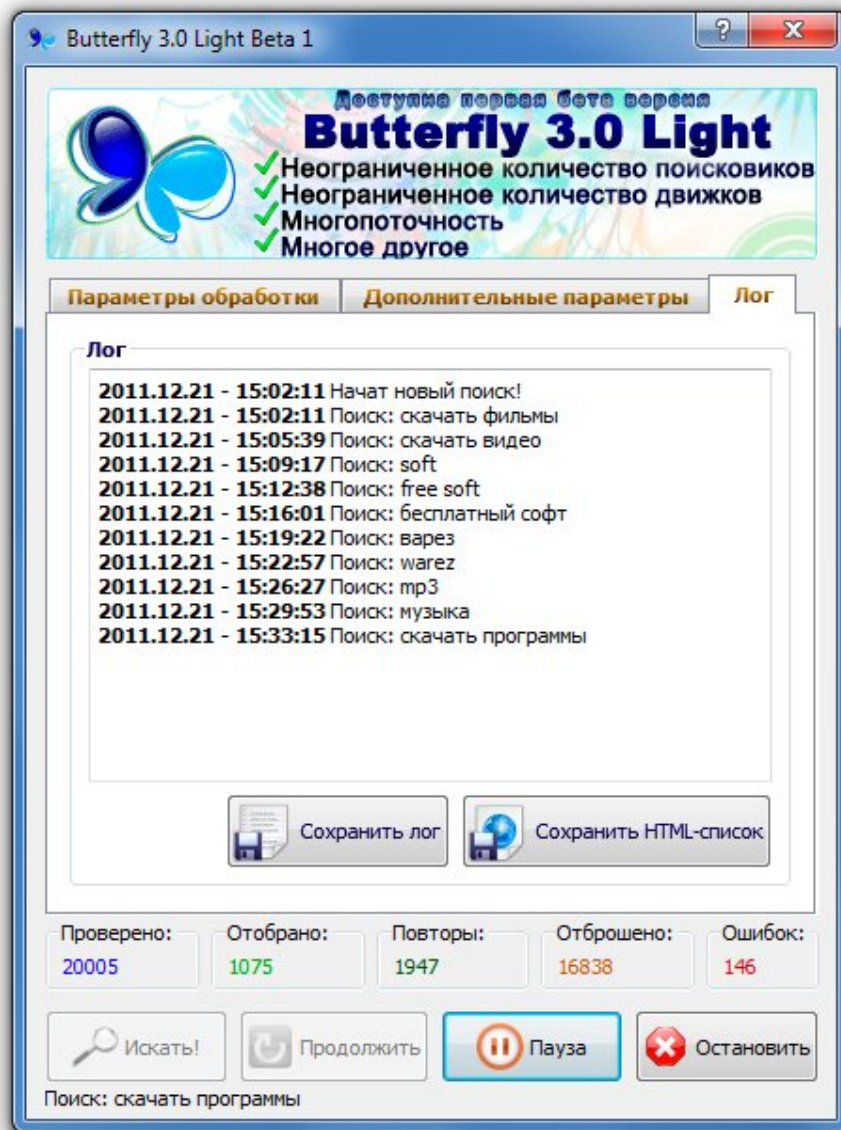
Max. number. references to the keyword - the maximum number of links, which will be tested the program for one key phrase. After the search, when the program checked the number of results equal to this parameter, the program sends the following keyword search engine, which also checks for less than the maximum number. If the reference count reaches the maximum, and search engine no longer has the results - the program continues to search for the next keyword.

Max. number. parallel processes - the maximum number of parallel threads checking sites. The more important - the faster the program will check the sites, but the greater burden on the resources of a PC. PC Intel Core2Duo 3Ghz, 2Gb RAM is not recommended to set the value of more than 400 (In the line of Butterfly Light This value is limited to 32).

Language - Changing the interface language. For the changes to take effect, you must restart the program.

Information - the information bar. There are buttons for opening this help and information windows "Pro" program.

In the "Log" is a log of all actions and conditions of the program:



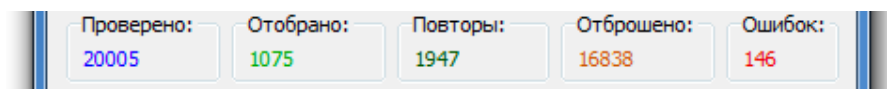
There are additional buttons:

Save the log - log is saved in the current. Txt file.

Save the HTML-List - This button allows you to save the last page of search results obtained from a search engine program for further analysis of the program with the search engine. This feature is useful when creating a profile search and filtering.

The lower part of the program, which is located under the main panel contains an information field that indicates the status of the search, as well as the panel of buttons control the program.

Panel of the information field is as follows:



It shows the number of tested options, including:

chosen - those who have passed the filter;

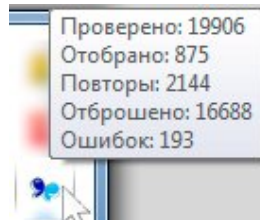
repeats - those who have passed the filter, but there are already results in a file;

dropped - those that do not pass the filter;

errors - number of sites that were unavailable at the time of inspection.

Always Checked = Selected Replays + + + Bugs discarded.

Also, this information is displayed we hover over the program icon in the tray:



Control panel program has the following buttons:

Search! - Start your search to specified parameters;

Pause - Pauses the search;



Stop - Stops the search by dropping the current state of the program;

Continue - continues to search after a pause, or an emergency stop.

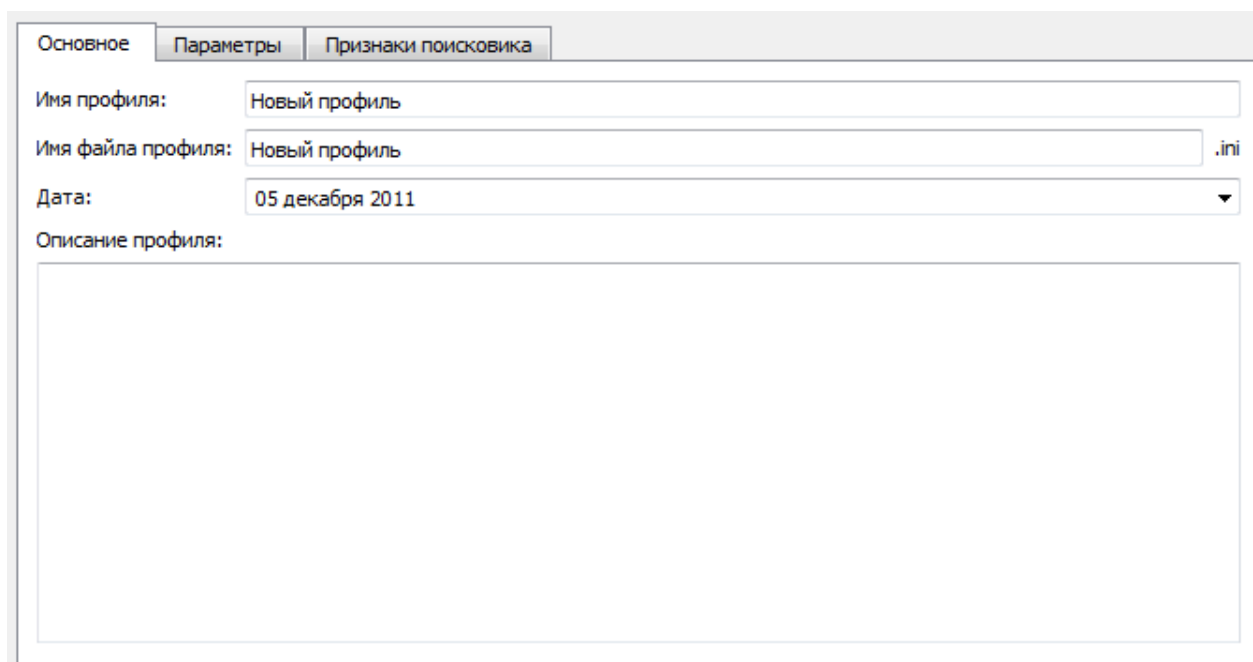
When the program if the program will go into alarm status (lost connection to the Internet, the search engine is not available, or others), as well as at the end of the search, the program goes into a pause. To be able to continue the search from its current location when it's favorable.

After the search program also goes into a pause. You need to press "Stop", then the field configuration will be available again and you can start a new search.

Creating and editing profiles, search

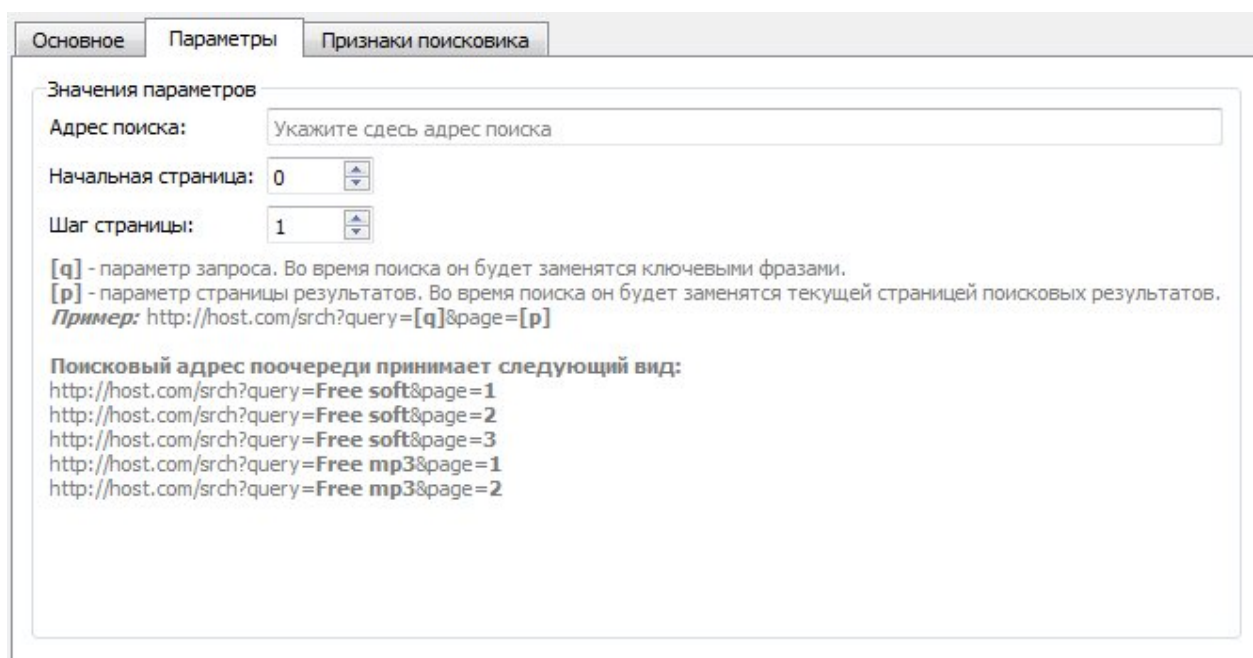
To create a new profile, click on . In order to create a profile by copying some of the existing ones, select an existing profile and click .

When you create a new profile, it is filled with the following data:



In the "General" you must enter a profile name, profile name of the file, creation date and description. Date of the current profile is usually made to determine the relevance of the date of the profile. In the description of the profile information is entered on the profile.

In the "Settings" you should fill in the data that will permit an enumeration of pages a search engine:



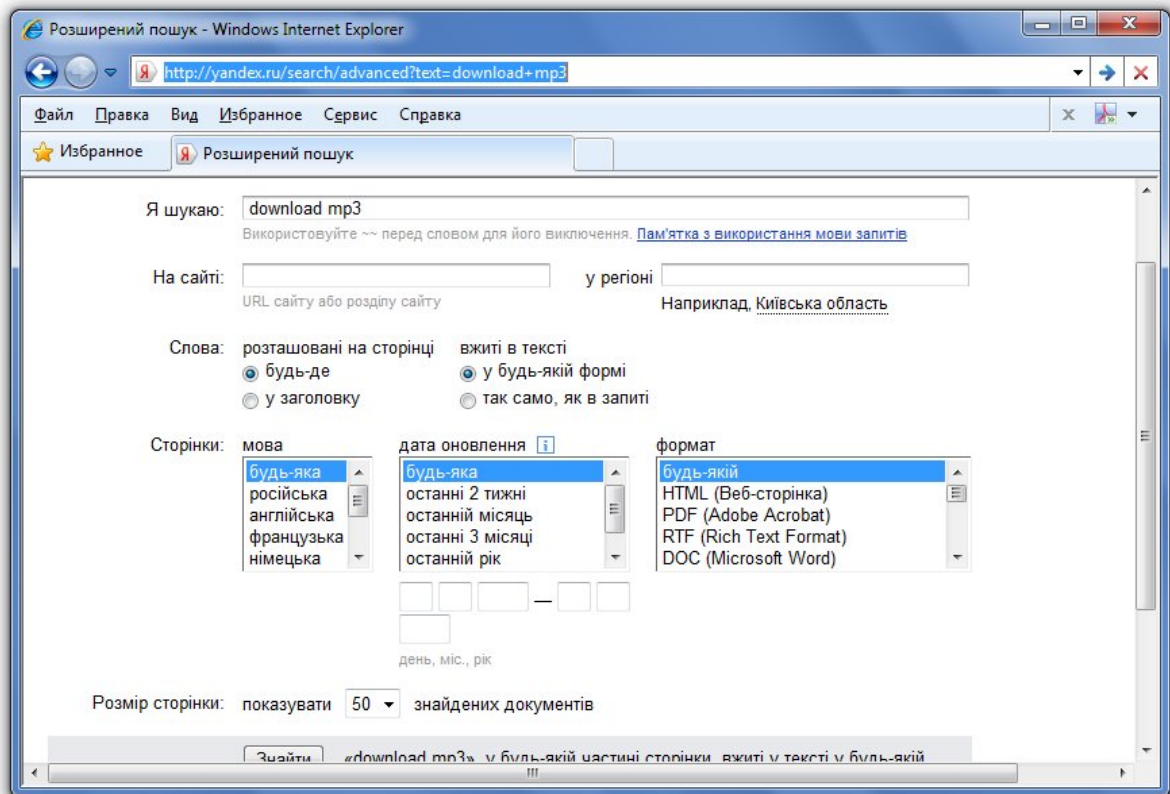
In the Address box, enter the parametric search to the address bar search engine, defining the parameters of the search query [q] and the current page [p]. During the search instead of the [q] will be replaced by real search query, and instead of [p] - the page number, starting position where oredelyaetsya value of "starting position", and a step value of the parameter "Step page."

Consider a real example of the filling of these parameters on the search engine Yandex.

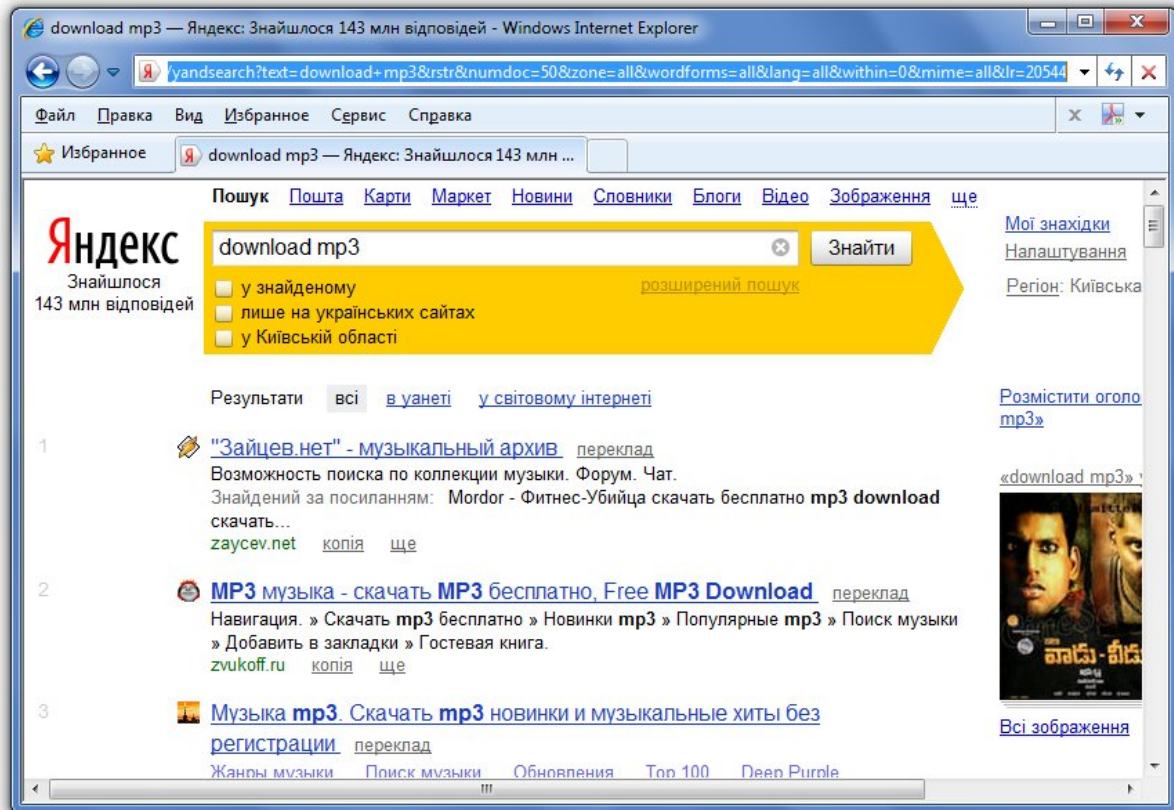
The main parameters of the search engine

First, we address the parametric search. It's enough to go to the search page, enter your search query and press "Search". Search engine will give results, and in the address field, we will form the desired line. This is the way it is now looking for search engine will find and the program, so the search engine, you can configure the advanced search options to search engine gave out as many as possible of the results, or results in English or another language. Thus, you can create a profile search in a specific language or from country to country.

Configure additional settings, press the search button:



Search engine results are issued, but also formed the address bar to be searched with advanced options:



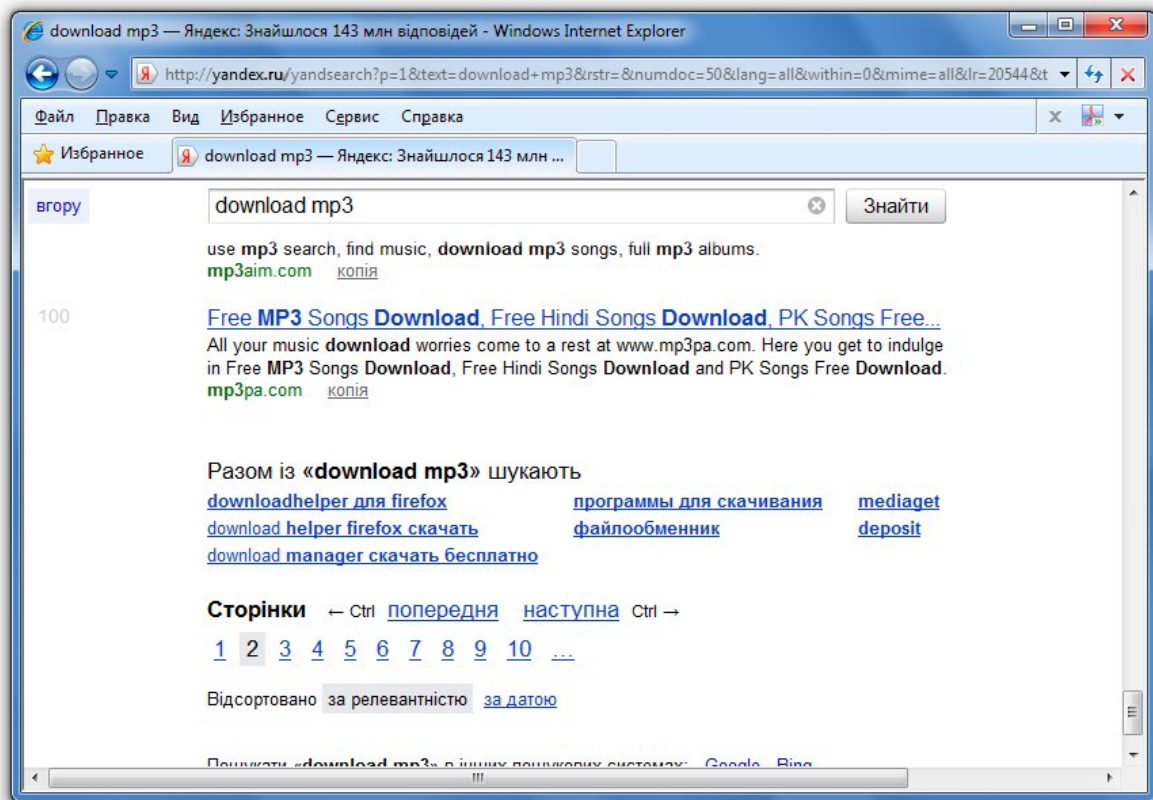
We got the address bar:

<http://yandex.ru/yandsearch?text=download+mp3&numdoc=50>

In the line we have left the most important parameters - the search query results and output parameter to 50 per page. Other options on the search result does not affect. Obviously, the search query parameter is «text», because instead of the search request, we substitute [q]:

[http://yandex.ru/yandsearch?text=\[q\]&numdoc=50](http://yandex.ru/yandsearch?text=[q]&numdoc=50)

However, our address does not have the option of a page number. If there is no need to just go to the next page of search and then it will appear:

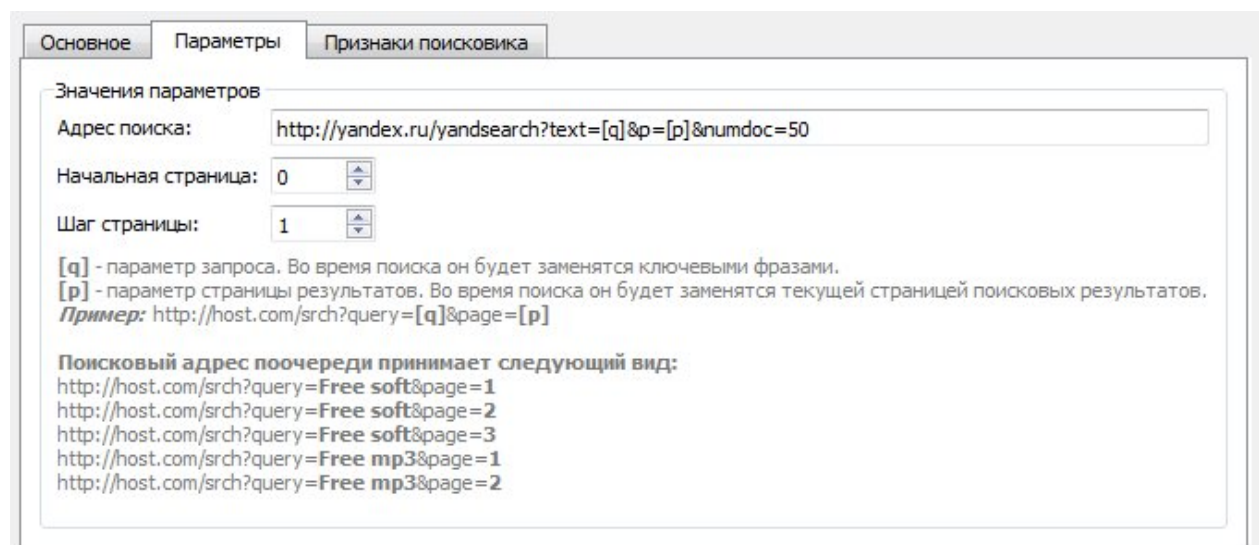


As we see, on the second page there now an option «p = 1.» This is the page parameter, because definitively address bar will look like this:

`http://yandex.ru/yandsearch?text = [q] & numdoc = 50 & p = [p]`

If we now turn to the third page, the option «p» takes the value 2. As we see, the page numbering starts at 0 with a step = 1. Because setting the home page set to "0", and move pages to "1".

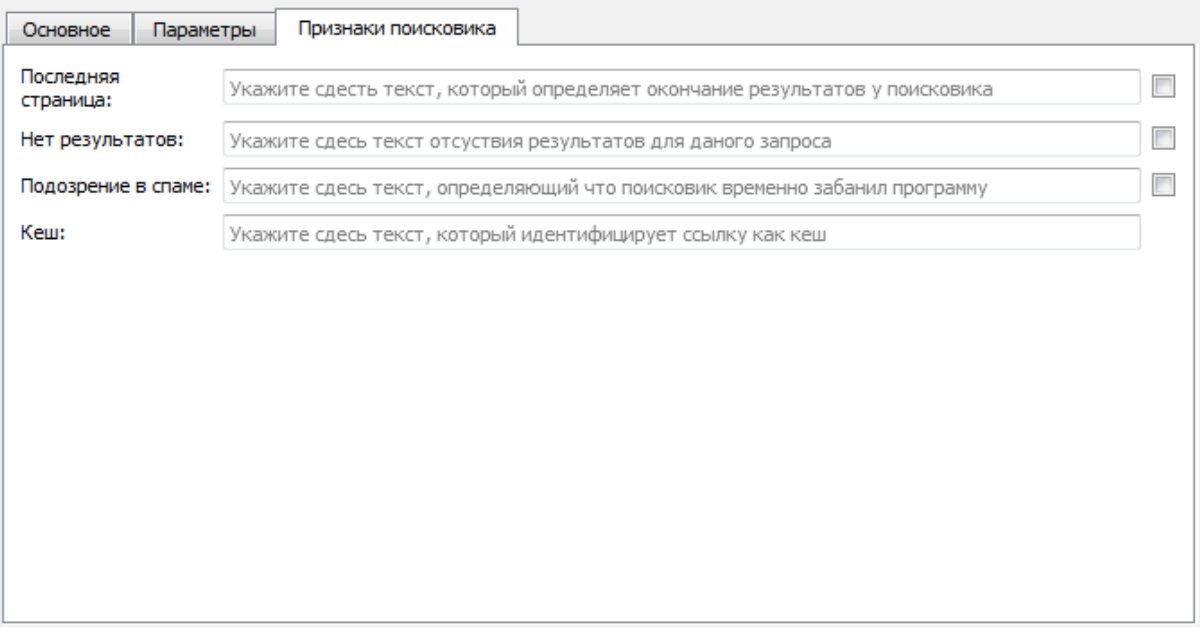
Thus, the settings page will look like this:



PS If we focus not only on the Russian-speaking users, it is better to use the domain name is not familiar to us «yandex.ru», but more general. For Yandex is «yandex.net»

Specifying a search engine

In addition to providing busting pages of search engine, you must provide a definition of what the results page is already finished, results are not at all or given a page notifying you that you are banned. In the first two cases, the program simply starts searching for the next keyword. In the third - suspends search.



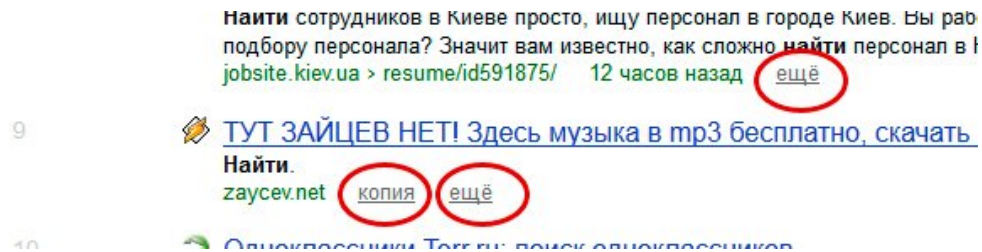
Label	Input Field	Checkbox
Последняя страница:	Укажите здесь текст, который определяет окончание результатов у поисковика	<input type="checkbox"/>
Нет результатов:	Укажите здесь текст отсутствия результатов для данного запроса	<input type="checkbox"/>
Подозрение в спаме:	Укажите здесь текст, определяющий что поисковик временно забанил программу	<input type="checkbox"/>
Кеш:	Укажите здесь текст, который идентифицирует ссылку как кеш	

If some of the parameters are optional, they do not check will be carried out. However, if the parameter definitions of the last page is not specified, the program will go through pages of search engine endlessly on a single keyword. In this case, you will notice that the program no longer go through links, or touches of 5-6 per minute.

In the same way, if you do not specify the definition of what the results on the keyword no, if such case happens, the program will behave as in the first case.

The "suspicion of spam" can be filled out if there was a ban (it comes with a few search engines). The program will behave the same way as in the first two cases. To determine the cause should stop the search, select the tab "Log" and click "Save HTML-list" - this is the page that the search engine tries to parse at this time. If you save you open it in your browser and would lead something like a captcha and similar output messages (or redirect to a page with a CAPTCHA), it is possible to find some admit it, according to which the program will identify this page. In the same way we can define the lack of results page and last page.

Most search engines, along with key references give references to the cache, which is useful to filter out:



To do this you need to fill in your profile sign "Cash" and enter the back piece of references to the cache. Typically, these links contain the name of the search engine.

For example, a link to the cache Yandex has the form:

<http://hghltd.yandex.net/yandbtm?fmode=inject&url=http% 3A% 2F% ...>

Thus, each such reference contains just a fragment of «. Yandex.net». It is this fragment to insert in the "cache", then the program will not consider these links in the search.

In each box features you need to fill the text that contains only one page of a search engine, which is the latest in search results, or does not have the results and set the check box. Or fill in the text, which is not only on the page, and the rest is. Then it is necessary to remove a tick.

For example, if the HTML-code search engine pages, which reports that the results found, contains the text of «div class =" msg_no_results "», and all other pages of search engine do not contain the text, then the text can be taken as an indication of the lack of results. It should fit in the "no results" and set the checkbox.

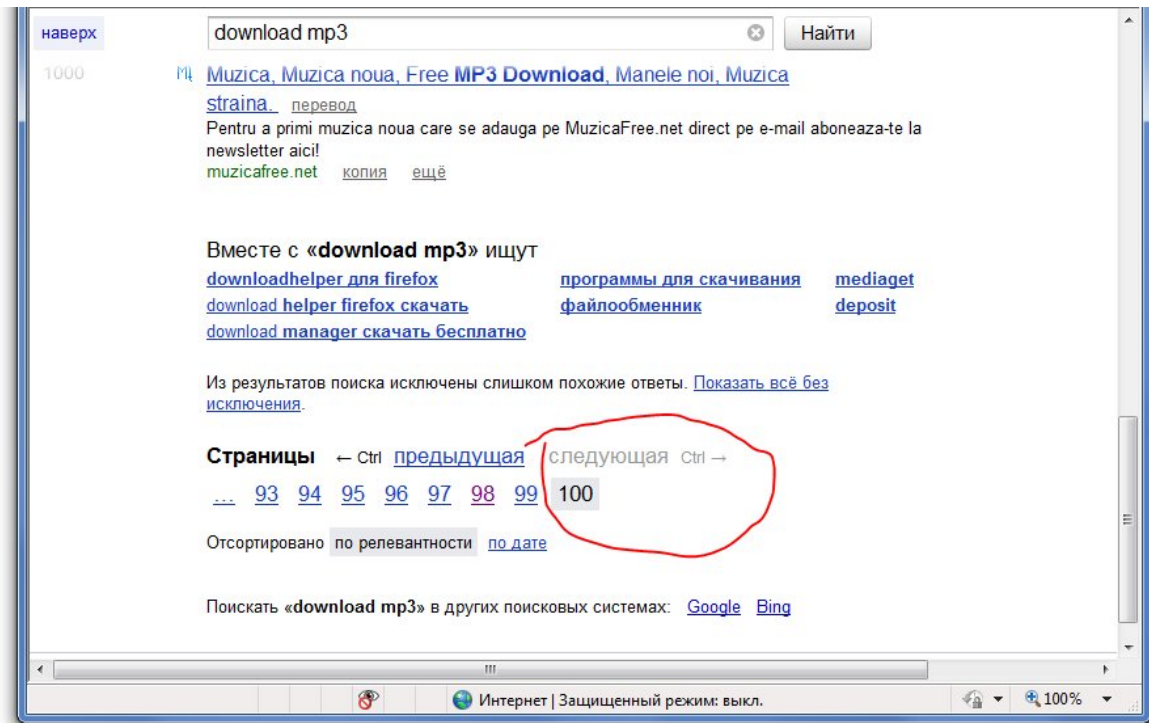
Another example: if the HTML-code of any page of search engine contains the text «div class =" result_list "», except that the page that says that the results found. This text can also be considered a sign of lack of results (if it is not detected). Then it should be written in the "no results" and remove the checkbox.

Check box options mean that the condition should be performed when there is a specified range of text on the page (check mark set) or not (tick off).

Determination of the end results of the search engine feature

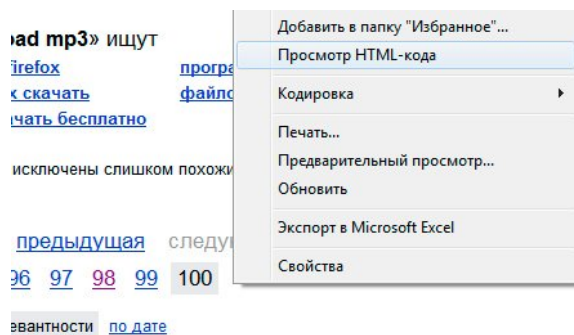
We find the sign of the end results for the search engine Yandex.

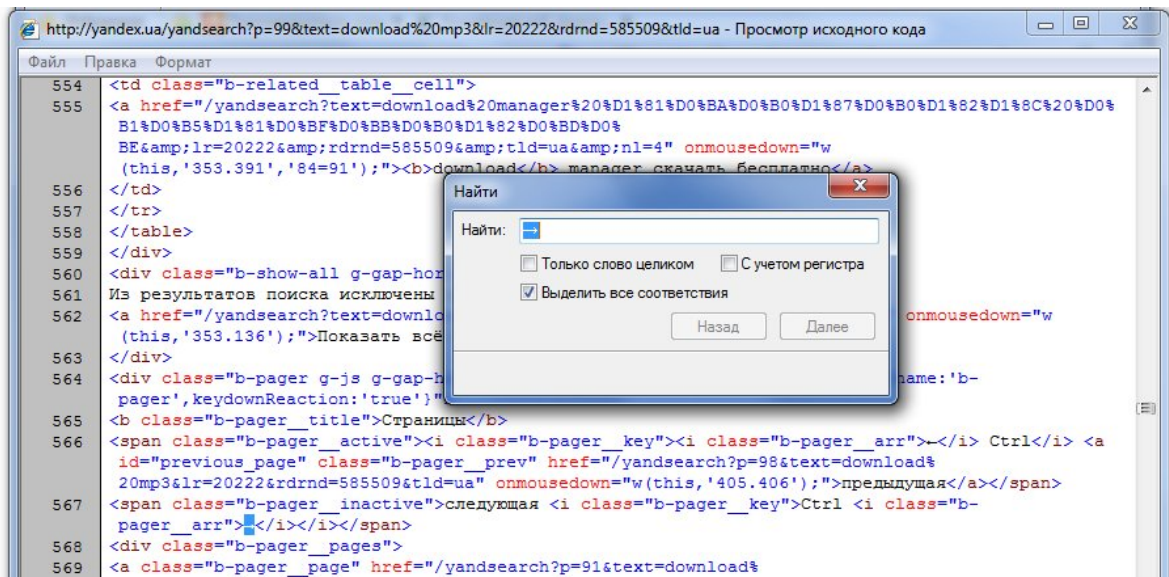
1. Let us turn to the search page, enter your search query and run the search.
2. Let us turn to the last page of search results:



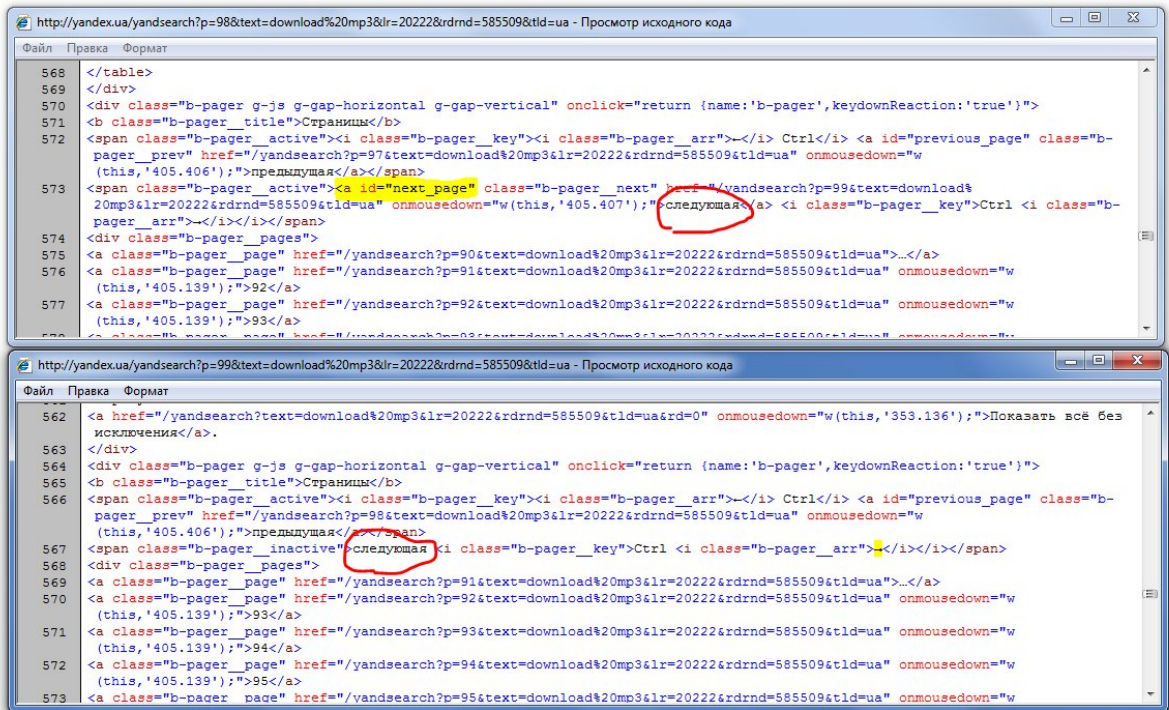
As you can see link "next" and the «Ctrl ->» not active, and page 101 does not (ie, no links to the right of the current page 100). They are no longer available only on the last page of search (and possibly on page 2 with 3 results.) It is in the code description contains the text of which is not on other pages, and which may be a sign the last page.

3. Let's go into the code of these links:





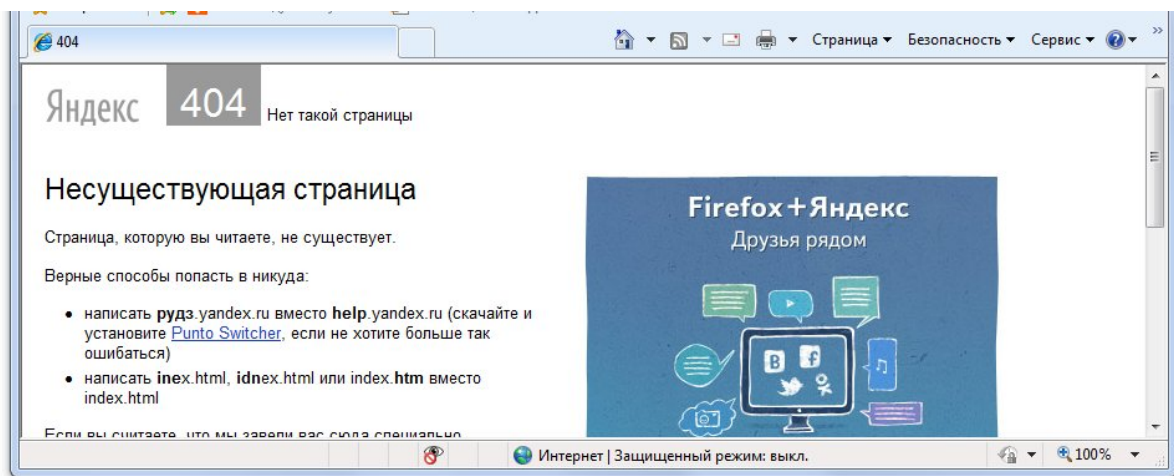
4. In the same way we find such code links to another page where they are active. For example in the past and compare the code:



As can be seen as a sign we can take the text "next </ a>". This text is not on the last page, but then lost the universality of the trait. After all, if someone is looking in a different language, for example, in Ukrainian, it will be "nastupna </ a>". Because it is better not to take as a sign of the text that is displayed to the user. We need to find a text that will be in text search engine ever, in all languages. You can take the link identifier on the following page: «<a id = "next_page" ».

5. Now you need to check whether the text is not found on the last page. Go to the code of the last page and are looking for. Do not find.

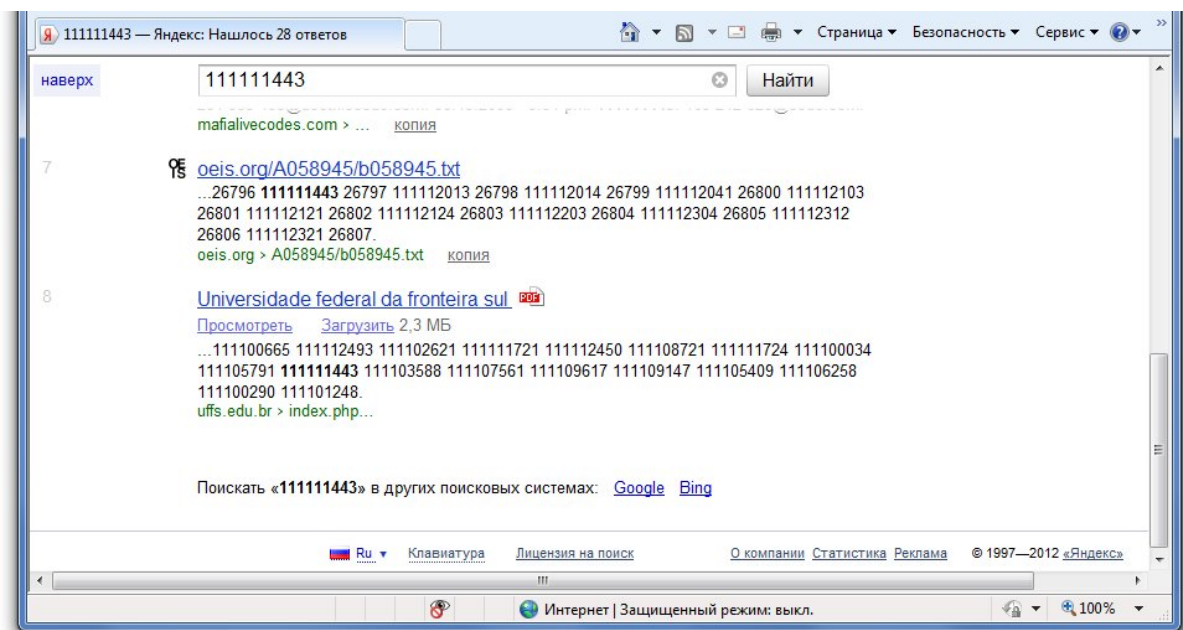
By the way, the text will not be present not only on the last page, but all subsequent ones, the parameter p = 101, 102, etc. For such requests, Yandex page will look like this:



You can select the sign of the last page, focusing on this page.

- Just need to check whether this meets the text on the page with a small amount of results (where there is no reference to the following page).

Enter a search term phrase for which there is probably not a lot of options that will fit on one page. For example, "111111443":



Search engine has given us a total of 8 results. Reference to the following pages are not visible. Let's see HTML-code, whether the text is not found «<a id = "next_page"» there. As it turned out - not found.

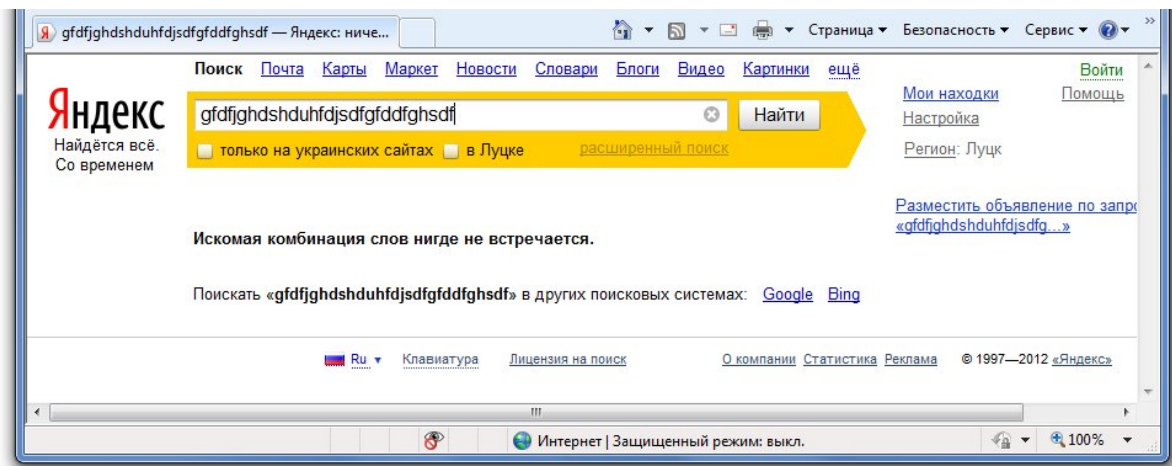
Thus, we determined that the text «<a id = "next_page"» is present only on the last page, because we can take it as a sign of the last page, removing the check box feature.

If we had found a text that was present only on the last page, check box would need to install it.

Defining characteristic lack of results

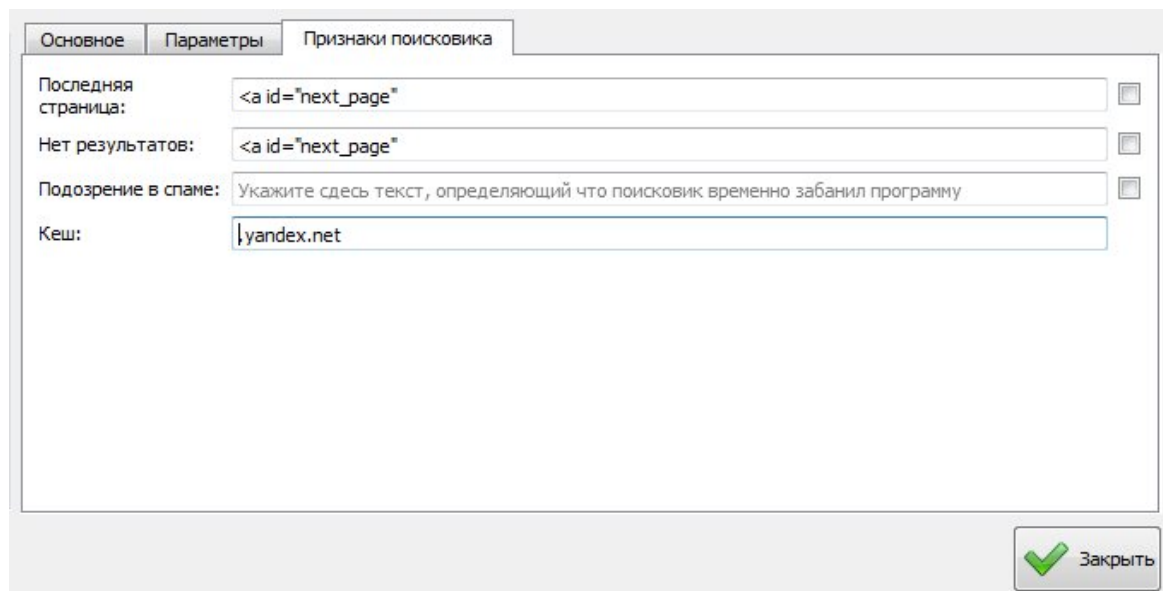
Determination of the sign is the same as the sign of the last page.

Necessary to form a search query, which does not show any search results, for example:



Check whether met our previously found the text «<a id = " next_page "» in HTML-code of this page. Does not occur. Because this text is suitable to us, and as a sign of lack of results. Brings it back and remove the check mark next to the field sign the last page.

Thus the profile tab fill characters will look like this:



Profile is ready, you can close the window and use the profile.

For each profile, you can find many options for characteristics that are or will be only on the last pages, or just there and will not, but if these conditions are fulfilled and properly exposed to ticks, the profile is stable to work in any form.