

[webextractor.com](http://webextractor.com)

# Web Data Extractor - Extract URL, Meta Tag, Email, Phone, Fax from Web

4-6 minutos

---

## Quick Start

[Link to Glossary with hints](#)

## WDE Pro Performance

Web Data Extractor Pro is a web scraping tool specifically designed for mass-gathering of various data types. It can harvest URLs, phone and fax numbers, email addresses, as well as meta tag information and body text. Special feature of WDE Pro is [custom extraction of structured data](#).

This high-speed and multithreaded program works by using a keyword into [search engines](#), by spidering a [website](#) or a [list of URLs from a file](#). You can also allow it to follow [external links](#) from the original pages, with the capability to go as deep into the URL paths as you need and actually search the entire Internet.

Web Data Extractor is superior for harvesting structured information and specific data types related to the keywords you provide by searching through multiple layers of websites.

## Start using WDE Pro

- To experience full functionality of the program [download](#) the trial version, that is valid for 14 days.

Specification:

- |           |                            |
|-----------|----------------------------|
| OS        | Windows XP/2000+/Vista/7/8 |
| Framework | .Net Framework 4.0         |

- WDE Pro Setup Wizard will install the program in a few moments.

Discover the process of web data extraction with new features available in Pro version.

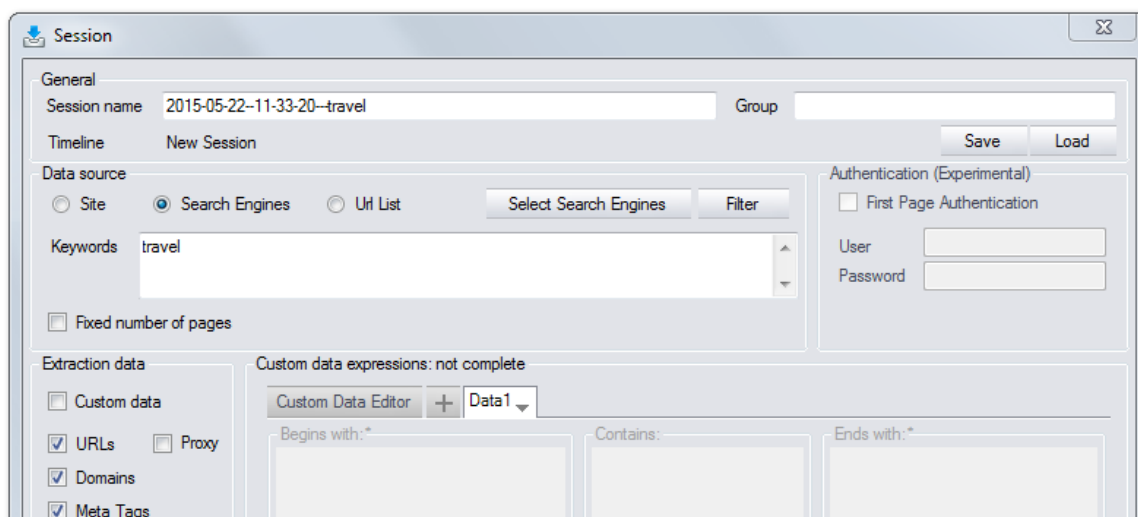
Though the trial version allows to extract any amount of information, mind that only 10 lines of extracted data can be saved.

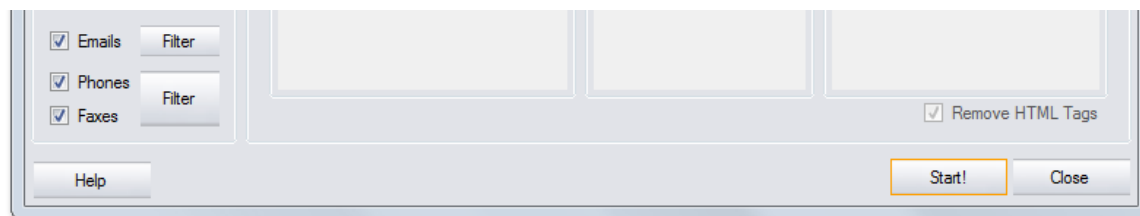
After [you purchase WDE Pro](#) your trial version will be converted into registered licensed version.

If you go through some problems or have ideas for improvement - don't hesitate to [communicate with us](#)!

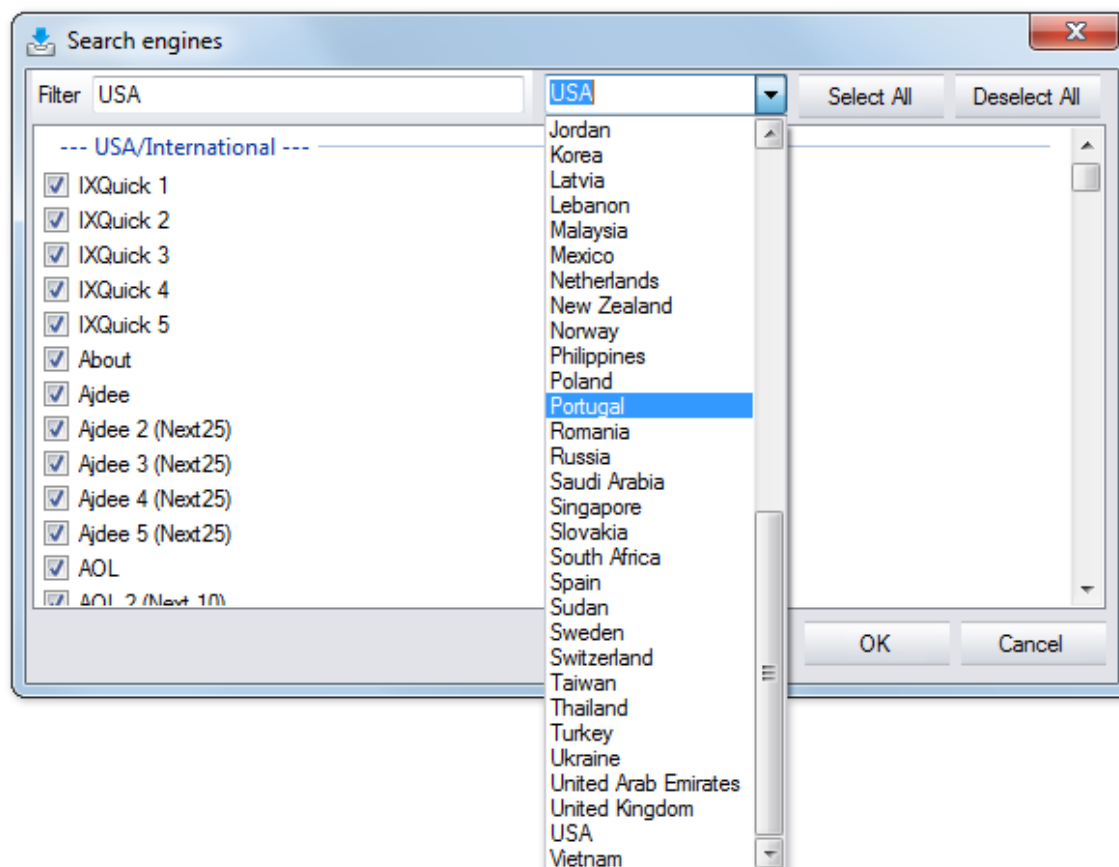
## Workflow review

Get started in a new [Session](#) by defining the search parameters.

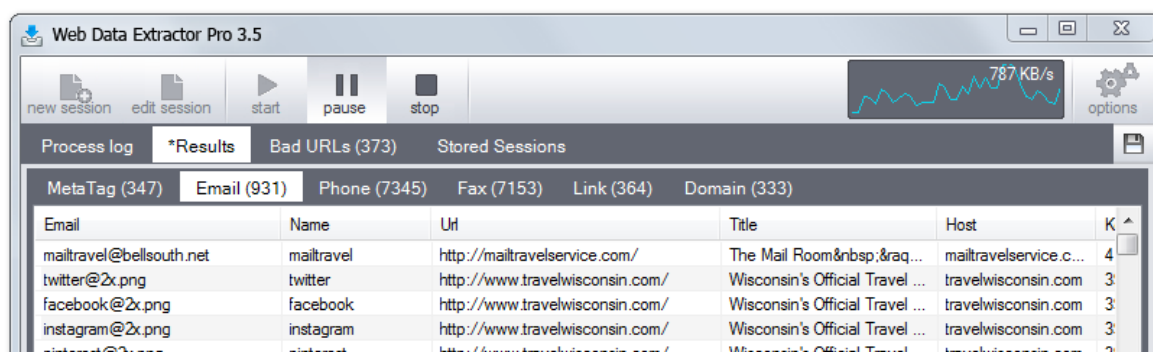




The geographical area you want to search is configurable (see [Search Engines](#)).



The program queries search engines using your desired keywords, fetches the matching URLs, removes duplicate URLs and searches through the links to extract specific data.



youtube@2x.png	youtube	http://www.travelwisconsin.com/	Wisconsin's Official Travel ...	travelwisconsin.com	3
google1@2x.png	google1	http://www.travelwisconsin.com/	Wisconsin's Official Travel ...	travelwisconsin.com	3
apple@1x.png	apple	http://www.theglobeandmail.com/life...	Travel - The Globe and Mail	theglobeandmail.c...	8
apple@2x.png	apple	http://www.theglobeandmail.com/life...	Travel - The Globe and Mail	theglobeandmail.c...	8
apple@2x.152.png	apple	http://www.theglobeandmail.com/life...	Travel - The Globe and Mail	theglobeandmail.c...	8
apple@3x.png	apple	http://www.theglobeandmail.com/life...	Travel - The Globe and Mail	theglobeandmail.c...	8
community@globeandmail.com	community	http://www.theglobeandmail.com/life...	Travel - The Globe and Mail	theglobeandmail.c...	8
bdc-sprite@2x.png	bdc-sprite	http://www.boston.com/travel		boston.com	2
tips@boston.com	tips	http://www.boston.com/travel		boston.com	2
TIPS@BOSTON.COM	TIPS	http://www.boston.com/travel		boston.com	2
plans@tripit.com	plans	https://www.tripit.com/		tripit.com	3
mailtravel@dmg-media.co.uk	mailtravel	https://plus.google.com/+MailTravel		plus.google.com	1
your_address@earthlink.net	your_address	http://travel.earthlink.net/	EarthLink - Travel	travel.earthlink.net	2
Travel@uconn.edu	Travel	http://travel.uconn.edu/	Home   Travel Services	travel.uconn.edu	1
site-masthead-logo@2x.png	site-masthead-logo	http://traveltips.usatoday.com/stop-m...	Stop Mail on Vacation   US...	traveltips.usatoday...	7
site-nav-logo@2x.png	site-nav-logo	http://traveltips.usatoday.com/stop-m...	Stop Mail on Vacation   US...	traveltips.usatoday...	7
footer-logo@2x.png	footer-logo	http://traveltips.usatoday.com/stop-m...	Stop Mail on Vacation   US...	traveltips.usatoday...	7
newsletters@utah.com	newsletters	http://www.utah.com/visitor/contact	Request a Utah Travel Gui...	utah.com	1

Processing time: 00:08:51.171    Sites processed: 1535 / 6188    Downloaded: 157 556 KB    Avg. Speed: 297 KB/s

There are also various filters and settings to help get the data you want: [URL filter](#), [Email filter](#), Phone/Faxes filter, [Connection settings](#), [Parser settings](#) etc.

Email Filter

☐ No duplicate domains  
☐ Except mail services

☐ Collect domains w/o e-mails  
☐ Remove address if it contains words

☐ Preference (keep if duplicate)

☒ Remove duplicates  
☐ No duplicate URL domains  
☐ Each email must match URL domain  
☐ Remove address in these domains

☐ Include address only if it contains words

☐ Include address only if page title contains words

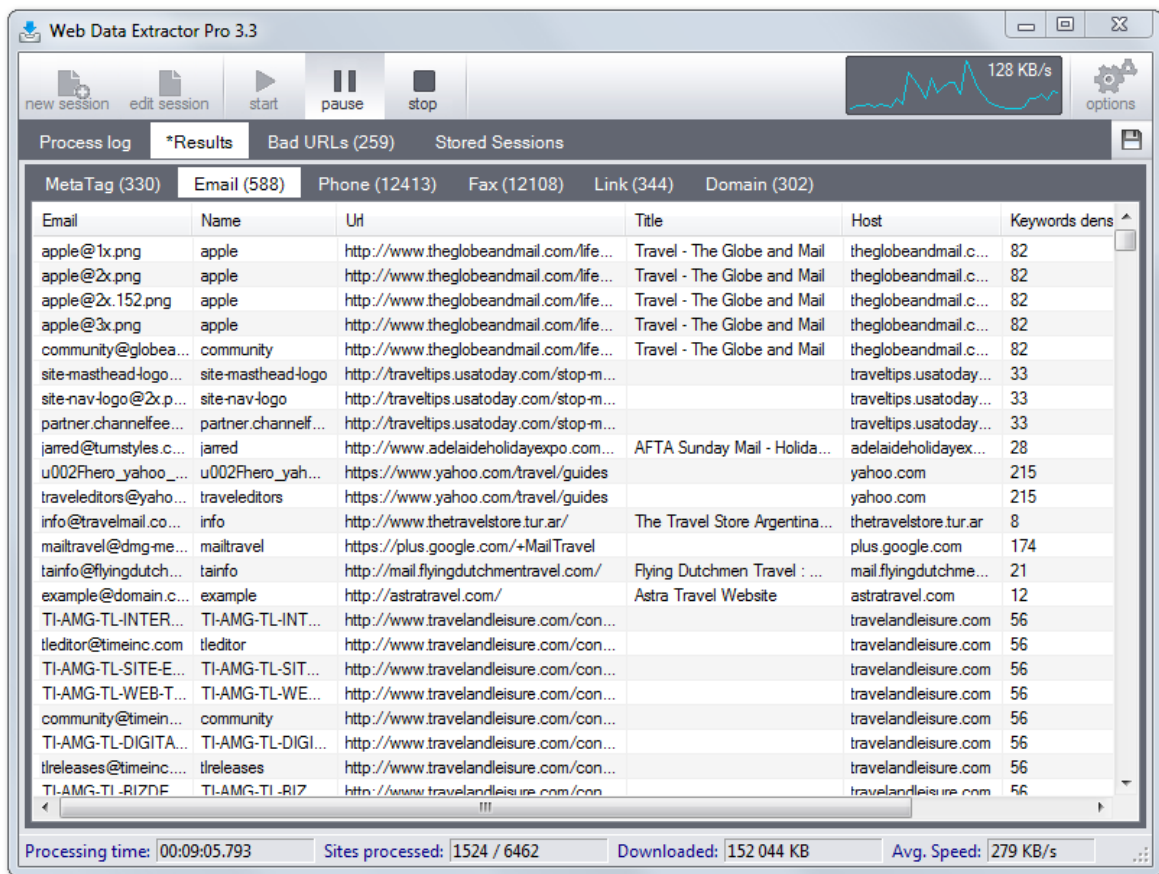
Spam

Help

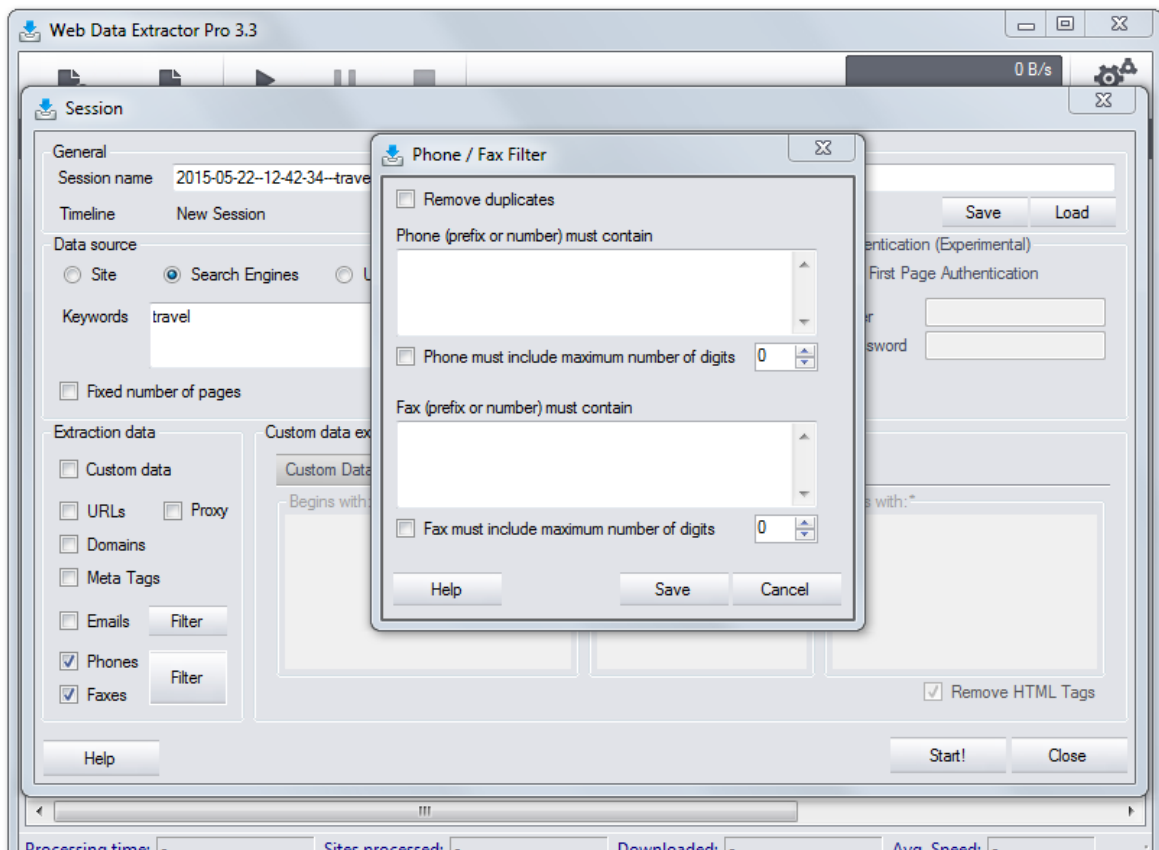
Save

Cancel

Email filter option gives you an opportunity to remove duplicate emails and do not show them in your results list.



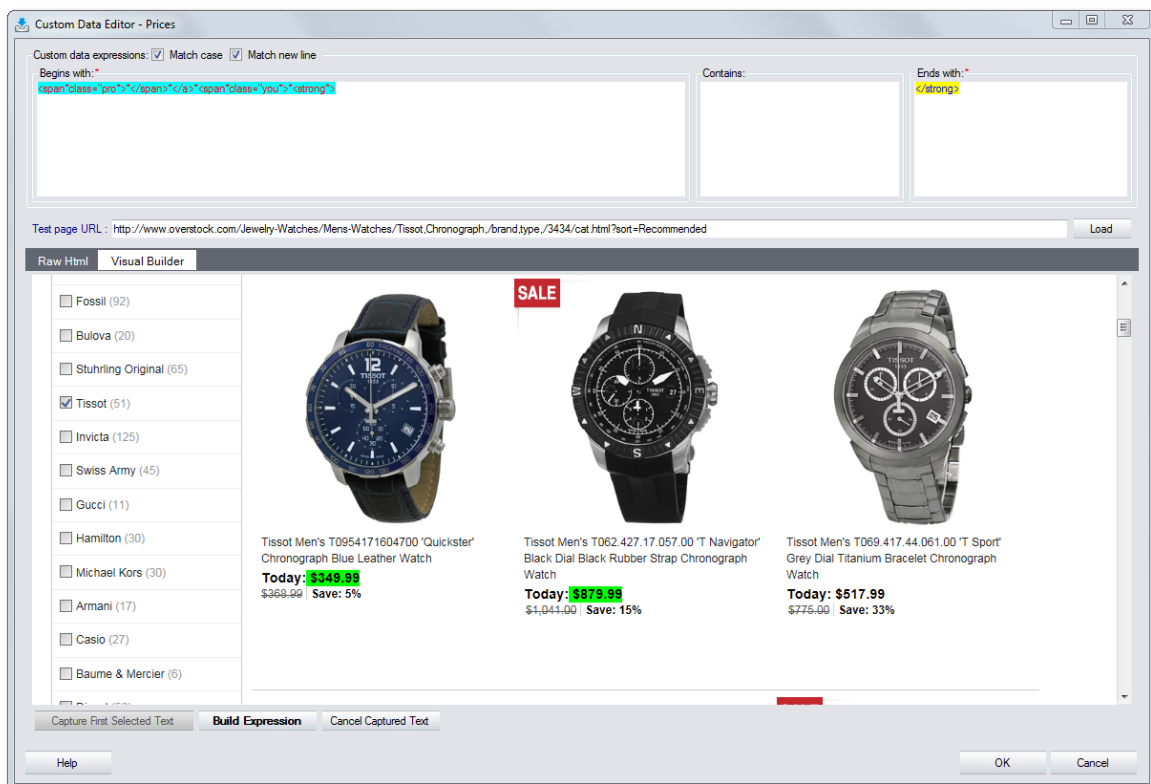
Phone/Fax filter gives you an option to remove duplicates and also to limit the number of digits in phone or fax

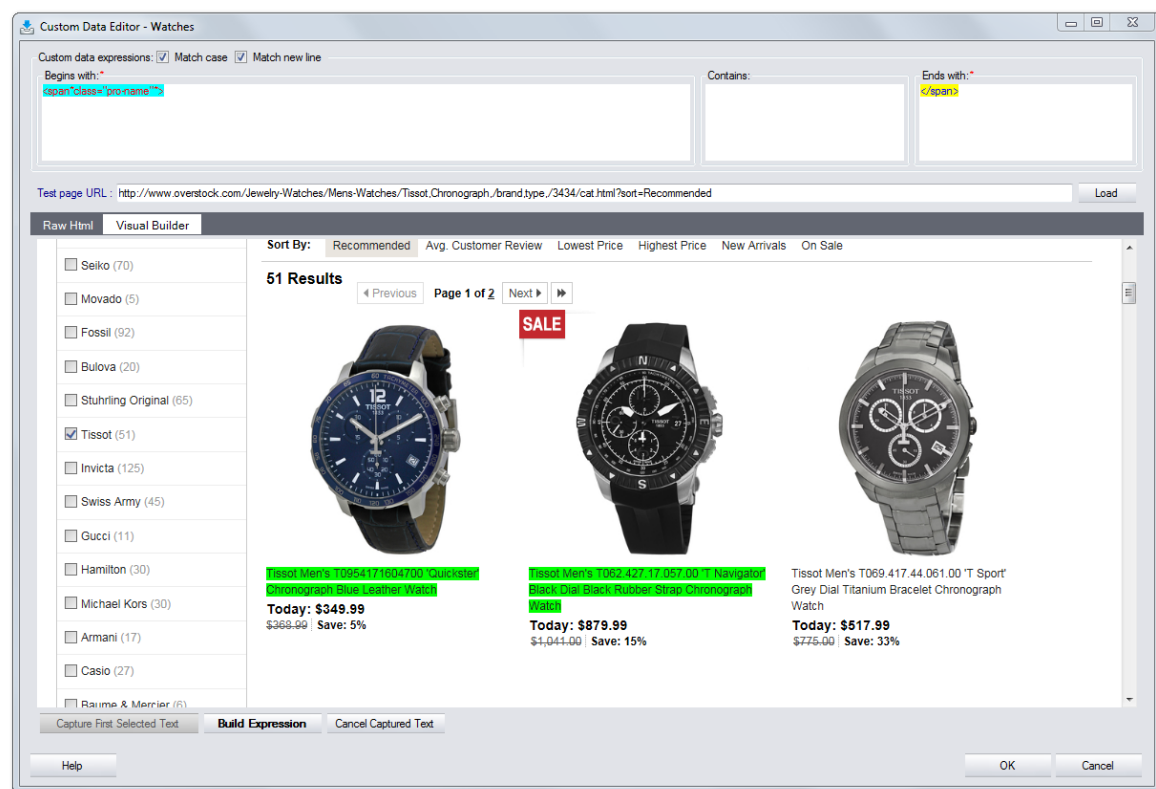
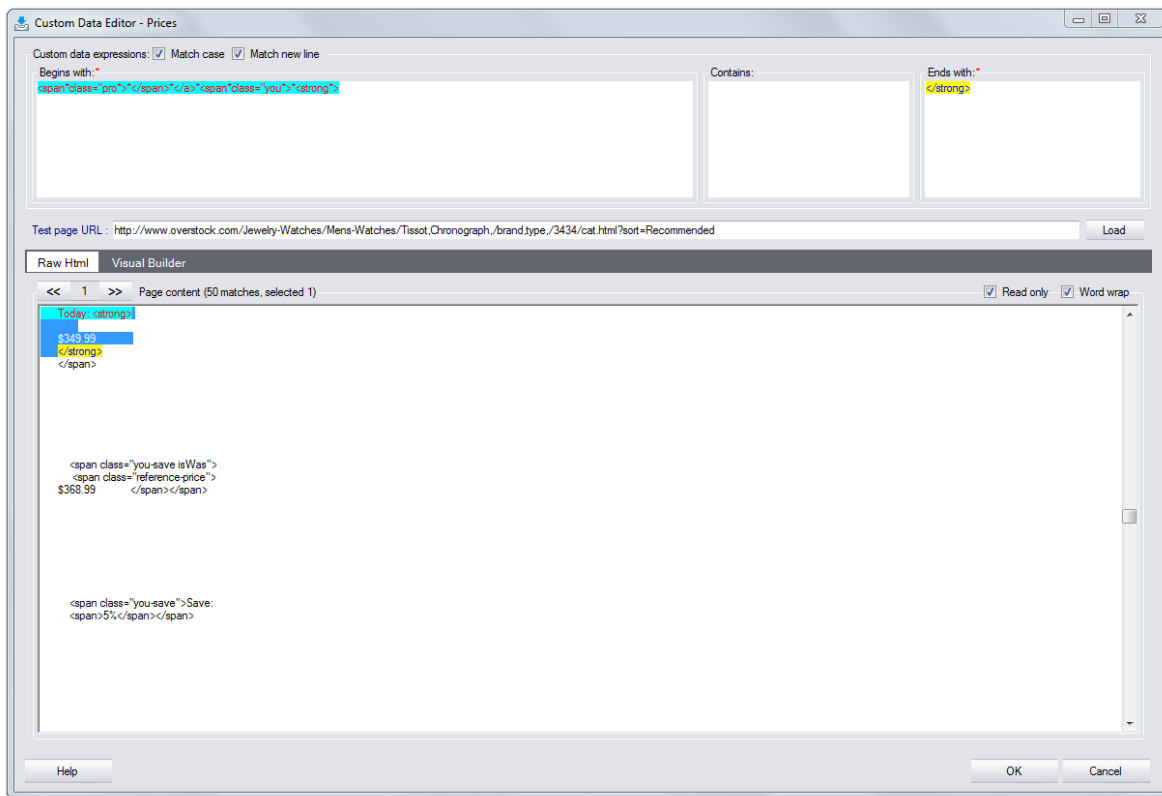


There is a special feature of WDE Pro - [custom extraction of structured data](#).

[Visual expression builder](#) tool will help you to get all items that have similar structure from a specified website. You just need to choose two items of information.

If you want to extract, for example, the list of products and their prices, you need to build two rules for products and for prices separately

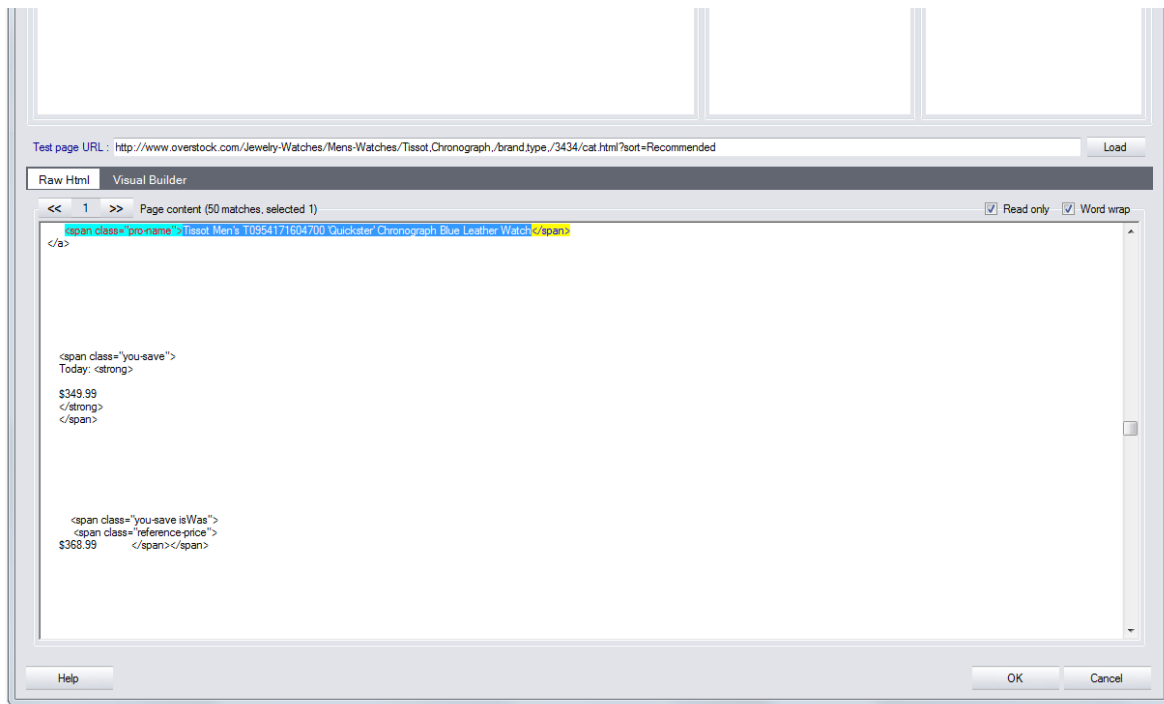




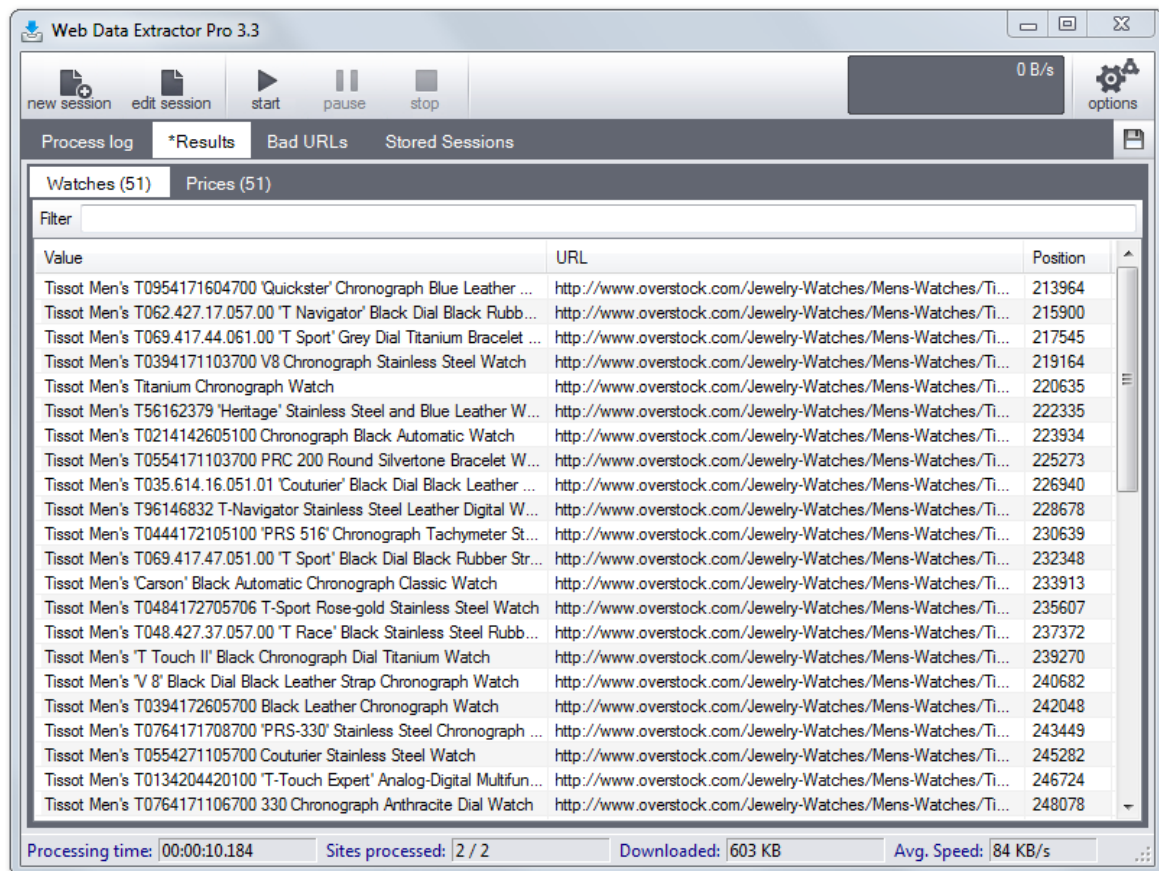
The program builds a pattern expression on account of the two items you've selected.







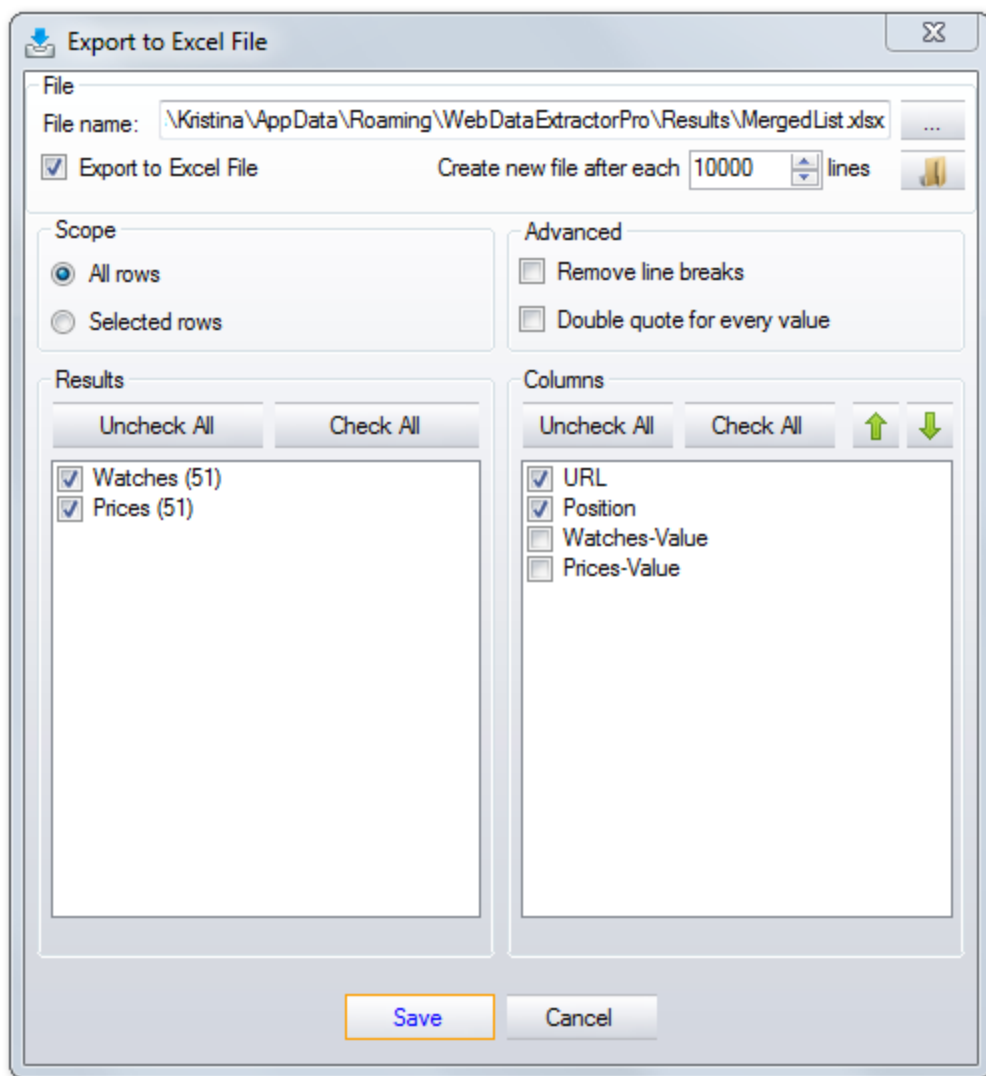
And on the basis of this expression WDE searches for further information.



Now you are ready to experience your own projects of web data



extraction.



When saving results you can choose rows you want to save, format of the file, also you have an option to save the list of results in different files creating new file after chosen number of results.