WEB EXTRACTOR

Manual

ALTILIA S.r.l.
Sede opertiva: Piazza Vermicelli – 87036 Rende (CS), Italy
Sede legale: Via A. Volta, 41 – 87036 Rende (CS), Italy

Tel & Fax: +39 0984 494277
E-mail: info@altiliagroup.com
Web: www.altilagroup.com

P. IVA e C. Fiscale: 10780921002 Reg. Imprese Cosenza
R.E.A. Cosenza n°209625 - Cap. Soc. versato € 17.395,56

TABLE OF CONTENTS

1 APP DOCUMENTATION

1.1 HOW IT WORKS

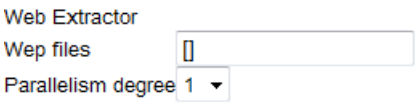
Web Extractor APP is an acquisition task that enables to extract information like product records, news, comments and other kind of information from web sites like e-commerce, online media and other web sites.

It requires as input one or more wep files properly created with the Web Extractor Modeler, an ALTILIA tool that enables to records action and rules for exploring a web site and its contents (see the Web Extractor Modeler User Guide to know more about this topic).

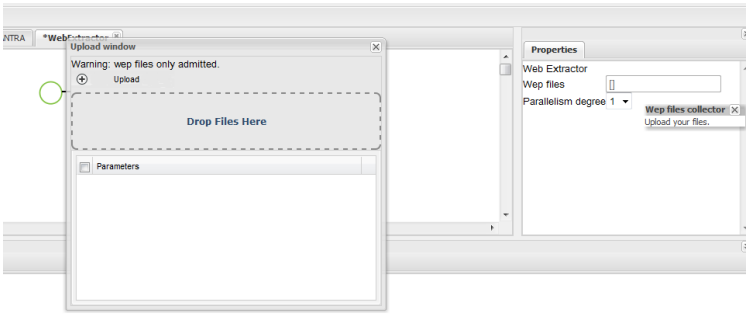
1.2 INPUT DATA

This APP allows You:

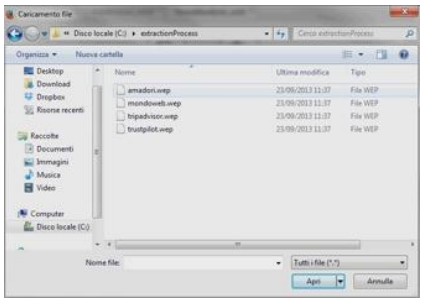
- ✓ to select one or more wep file that explain to MANTRA how and where capture data from a web site;
- ✓ specify the parallelism degree.



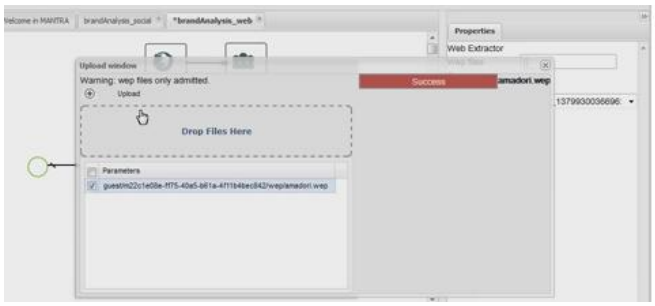
Step 1 – click in the wep files field



Step 2 – upload windows opening



Step 3 – selecting wep files



Step 4 – wep file selected

Figure 1 – how to select input data.

Once selected the wep files, you must select the parallelism degree.

1.3 OUTPUT DATA

This APP gives back a set of variable elements; they depends on the wep files you are using.

Then isn't possible describe the output data in this user guide.

But as example, if you made a wep file that capture data about products of a web site, you can have as output fields having information on product name or description or price and so one.

Those output fields will be available for the APPs that follow the Web Extractor APP.

These APPs can be Sentiment-Extractor APP for Brand Reputation Analysis, or MANTRA Language APP for transforming semi-structured data coming from the Web Extractor APP in structured data on which to make subsequent analysis like price comparison (Price Intelligence Analysis as example).

1.4 BASIC WORKFLOW EXAMPLE

Web Extractor needs some transformation and / or normalization APP after in the workflow that can process information captured, so you can create a flow as shown in the following figure left below.

In this example we have a market intelligence use case example, where APPs, starting from semi structured data, can compare product price from different web sites.

You can set input parameter and obtain the results by the to excel APP as shown in the following figure at the right side:

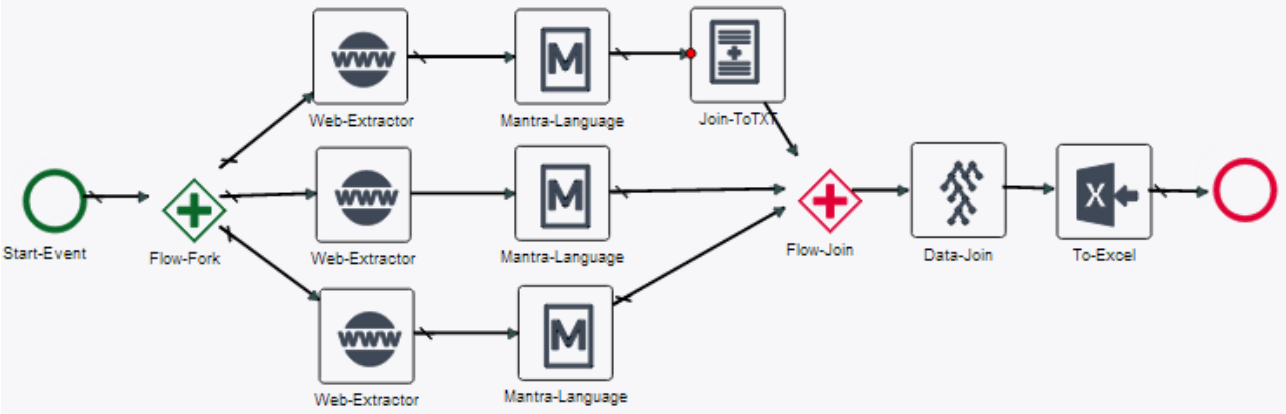


Figure 2 - web extractor workflow example.

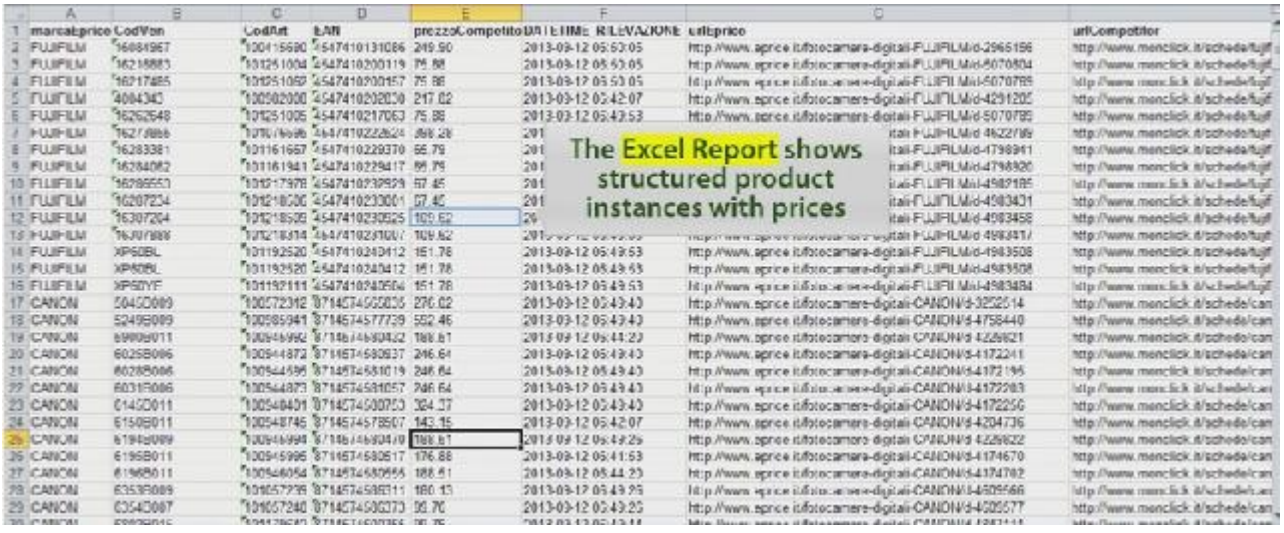


Figure 3 - the excel format of results.

2 API DOCUMENTATION

For information about how to use Article Extractor API in your application, send us a message to info@altiliagroup.com.