



# ENTITY EXTRACTOR

## *Manual*

## TABLE OF CONTENTS

<b>1</b>	<b>APP DOCUMENTATION.....</b>	<b>3</b>
1.1	HOW IT WORKS	3
1.2	Input data	4
1.3	Output data	4
1.4	Basic workflow example	5
<b>2</b>	<b>API DOCUMENTATION.....</b>	<b>7</b>

## 1 APP DOCUMENTATION

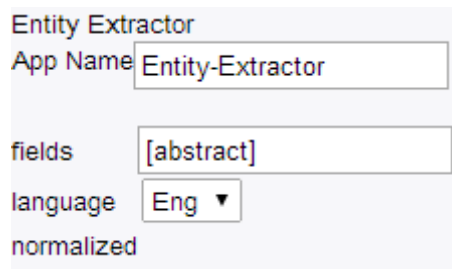
### 1.1 HOW IT WORKS

**Entity Extractor APP** is a normalization task that allows users to accurately extract valuable information from unstructured documents (e.g. plain text, pdf, doc) capturing several entities like person name, place, organization, company/brand and so on.

## 1.2 INPUT DATA

This APP allows you to accurately extract valuable information from unstructured documents (e.g. plain text, pdf, doc) capturing several entities like person, place, organization, and so on.

Input data shows in the following *Figure 1* return all entity specifications for persons, places, organizations and so on, contained in the “abstract” field given in output from the previous app in the workflow, by processing the English natural language. “abstract” field, in the next example, contains all article abstracts extracted from the home page of the Washington Post on line media.



The screenshot shows the 'Entity Extractor' application interface. It features four input fields: 'App Name' with the value 'Entity-Extractor', 'fields' with the value '[abstract]', 'language' with a dropdown menu set to 'Eng', and 'normalized' which is currently empty.

*Figure 1 – input data.*

## 1.3 OUTPUT DATA

This APP gives back, for each extracted entity, a set of four elements:

1. ent\_occurs: number of occurrences of this entity;
2. ent\_text: the text of the extracted entity;
3. ent\_type: entity typology;
4. id: unique number of the extracted entity.

This set of elements is given back in an excel spreadsheet or a data cube whose metadata depend on the data extracted and presented in the output.

You can see, in the following figures, an example of output data (*Figure 3*) based on the input data given (*Figure 2*):

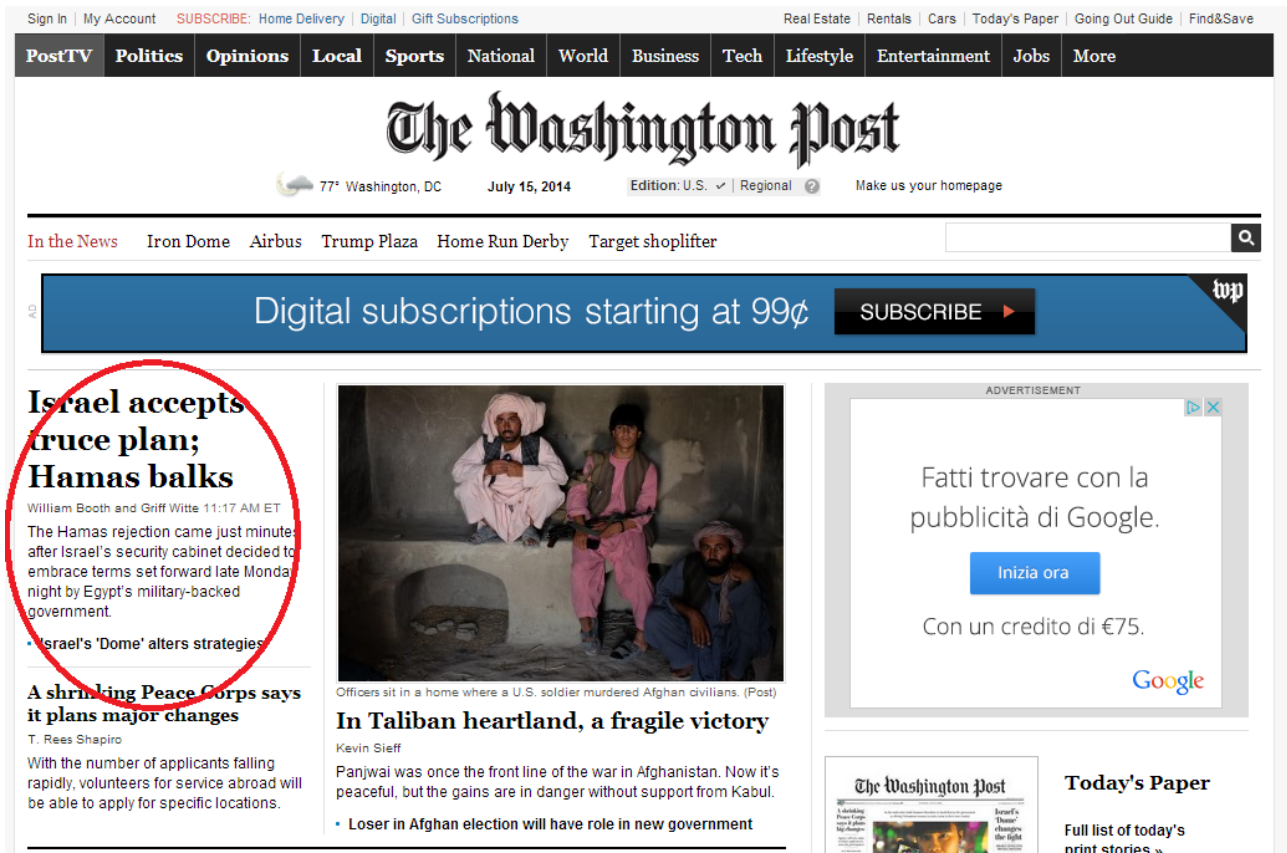


Figure 2 - example of input text (home page of The Washington Post)

⊟ (8 Items)	
abstract	William Booth and Griff Witte 9:17 AM ET The Hamas rejection came just minutes after israel's security cabinet decided to embrace terms set forward late Monday night by Egypt's ...
ent_occurs	1
ent_text	Israel
ent_type	Place
id	0000069
image	
link	http://www.washingtonpost.com/world/israel-accepts-truce-plan-hamas-balks/2014/07/15/04373008-0b15-11e4-8c9a-923ecc0c7d23_story.html
title	Israel accepts truce plan; Hamas balks
⊟ (8 Items)	
abstract	William Booth and Griff Witte 9:17 AM ET The Hamas rejection came just minutes after israel's security cabinet decided to embrace terms set forward late Monday night by Egypt's ...
ent_occurs	1
ent_text	Griff Witte
ent_type	Person name

Figure 3 - example of output data

This example is a sample of output where Entity Extractor APP extracts from the abstract of an article published on Washington Post the *Place* instance “Israel” and the *Person name* instance “Griff Witte”.

## 1.4 BASIC WORKFLOW EXAMPLE

Entity Extractor APP needs some extraction app before in the workflow that can get some text to process, so you can create a simple flow as shown in the following Figure 4.

You can set input parameter and watch the results by the watcher button as shown in the following Figure 5.

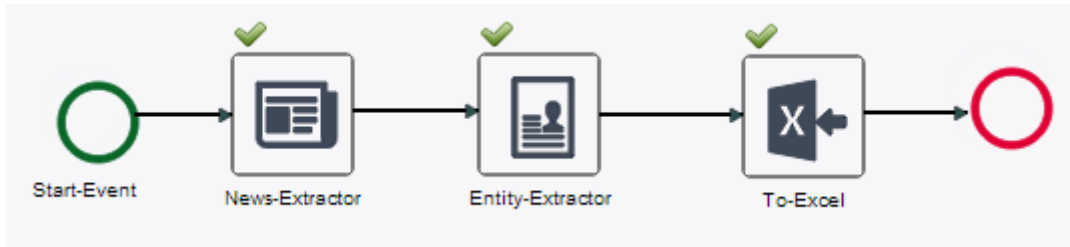


Figure 4 - simple entity extractor workflow example

**MANTRA** Smart Data Management Platform

EntityExtractor

Properties

Entity Extractor  
App Name: Entity-Extractor

fields: [abstract]

language: Eng

normalized

Execution Id	Name	Value
1	(8 Items)	
	abstract	William Booth and Griff Witte 9:17 AM ET The Hamas rejection came just minutes after Israel's security cabinet decided to embrace terms set forward late Monday night by Egypt's ...
	ent_occurs	1
	ent_text	Israel
	ent_type	Place
	id	0000069
	image	
	link	http://www.washingtonpost.com/world/israel-accepts-truce-plan-hamas-balks/2014/07/15/04373008-0b5f-11e4-8c9a-923ecc0c7d23_story.html
	title	Israel accepts truce plan; Hamas balks
	(8 Items)	
	abstract	William Booth and Griff Witte 9:17 AM ET The Hamas rejection came just minutes after Israel's security cabinet decided to embrace terms set forward late Monday night by Egypt's ...
	ent_occurs	1
	ent_text	Griff Witte
	ent_type	Person name

Figure 5 - the watcher panel.

## 2 API DOCUMENTATION

For information about how to use Entity Extractor API in your application, send us a message to [info@altiliagroup.com](mailto:info@altiliagroup.com).