



TOPIC EXTRACTOR

Manual

TABLE OF CONTENTS

1	APP DOCUMENTATION.....	3
1.1	HOW IT WORKS	3
1.2	Input data	4
1.3	Output data	4
1.4	Basic workflow example	5
2	API DOCUMENTATION.....	6

1 APP DOCUMENTATION

1.1 HOW IT WORKS

Topic Extractor APP is a normalization task that enables to extract keywords from an input text becoming from a previous APP in the workflow. It uses natural language processing techniques and gives some interesting parameters as output.

1.2 INPUT DATA

This APP makes keywords extraction from plain text using the most common NLP tasks, such as tokenization, part-of-speech tagging, lemmatization. Rule patterns and statistic criteria support the keywords extraction.

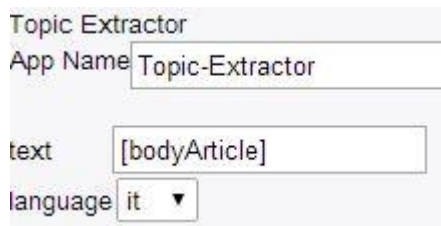


Figure 1 - input data.

1.3 OUTPUT DATA

This APP gives back a set of five elements:

- keyword: extracted keyword;
- language: used language;
- frequency: keyword frequency in the processed text;
- occurrences: keyword occurrences in the input corpus;
- score: this value measures the reliability of the keyword; higher is the score, the higher is the weight of keywords in the text with respect to the corpus.

You can see an example of output data in the following figure:

abstract	Prima al cimitero monumentale del Commonwealth di Cassino Condividi 80 condivisioni
bodyArticle	Vista a sorpresa di mezz'ora al Colosseo del principe Harry d'Inghilterra, il secondogenito di Carlo e Diana che si trova in Italia per la cerimonia commemorativa per ricordare i ca...
frequency	6
image	http://www.repstatic.it/content/nazionale/img/2014/05/19/165521593-709e9260-5c12-4005-a6b7-e476d0521a32.jpg
imageArticle	http://roma.repubblica.it/images/2014/05/19/112217354-3c5f5502-bc61-4297-ae35-07dee0aa4356.jpg
keyword	principe harry
language	it
link	http://roma.repubblica.it/cronaca/2014/05/19/news/cassino_il_principe_harry_al_cimitero_monumentale_del_commonwealth-86551095/
occurrences	6
score	36.31
subTitleArticle	Il secondogenito di Carlo e Diana, in Italia per la cerimonia commemorativa per i caduti nella battaglia di Montecassino, per la prima volta all'Anfiteatro Flavio che è rimasto aperto a...
title	Il principe Harry visita video a sorpresa il Colosseo foto
titleArticle	Il principe Harry visita a sorpresa il Colosseo: "E' incredibile". Prima al cimitero monumentale del Commonwealth di Cassino

Figure 2 - An example of output data.

In this example there are some extracted keywords becoming from some articles extracted from an online newspaper.

Keyword “*principe harry*” has 6 frequency in the text and 6 occurrences in the whole corpus with a total score valued 36.31.

1.4 BASIC WORKFLOW EXAMPLE

Topic Extractor APP needs an APP before in the workflow having a text field containing the text to process, so you can create a flow as shown in the following figure 3.

You can set input parameter and watch the results by the watcher button as shown in the figure 4:



Figure 3 - Simple twitter-follow example.

Execution Id	Name	Value
1	abstract	Prima al cimitero monumentale del Commonwealth di Cassino Condividi 80 condivisioni
2	bodyArticle	Visita a sorpresa di mezz'ora al Colosseo del principe Harry d'Inghilterra, il secondogenito di Carlo e Diana che si trova in Italia per la cerimonia commemorativa per ricordare i ca...
	frequency	6
	image	http://www.repstatic.it/content/nazionale/img/2014/05/19/166521593-709e9260-5c12-4005-a6b7-e476d0521a32.jpg
	imageArticle	http://roma.repubblica.it/images/2014/05/19/112217354-3c5f5502-bc61-4297-ae35-07dee0aa4356.jpg
	keyword	principe harry
	language	it
	link	http://roma.repubblica.it/cronaca/2014/05/19/news/cassino_il_principe_harry_al_cimitero_monumentale_de_commonwealth-86551095/
	occurrences	6
	score	36.31
	subTitleArticle	Il secondogenito di Carlo e Diana, in Italia per la cerimonia commemorativa per i caduti nella battaglia di Montecassino, per la prima volta all'Anfiteatro Flavio che è rimasto aperto a...
	title	Il principe Harry visita video a sorpresa il Colosseo foto
	titleArticle	Il principe Harry visita a sorpresa il Colosseo: "E' incredibile". Prima al cimitero monumentale del Commonwealth di Cassino

Figure 4 - The watcher panel.

2 API DOCUMENTATION

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