Document to Knowledge™
A CANON COMPANY

Text Finder Improvements

Multiple data extraction & Processing Speed Improvement

Revision History

Date	Description	Author
07/23/2019	V1.0 First version	EB
29/08/2019	V1.1 – Review for release	TF

I.R.I.S. SA/NV
Rue du Bosquet 10
B-1348 Louvain-la-Neuve

P: +32 (0) 10 45 13 64
F: +32 (0) 10 45 34 43

www.irislink.com

Canon
CANON GROUP

IRIS
Document to Knowledge™
A CANON COMPANY

Table of Contents

1. Introduction 3
2. Multiple values extraction 3
3. Improved Processing Time 6
4. Legal Notice 7
4.1 *Copyrights* 7

I.R.I.S. SA/NV
Rue du Bosquet 10
B-1348 Louvain-la-Neuve

P: +32 (0) 10 45 13 64
F: +32 (0) 10 45 34 43

www.irislink.com

Canon
CANON GROUP

IRIS Document to Knowledge™
A CANON COMPANY

1. Introduction

This technical note provides information about two new improvements of the Text Finder.

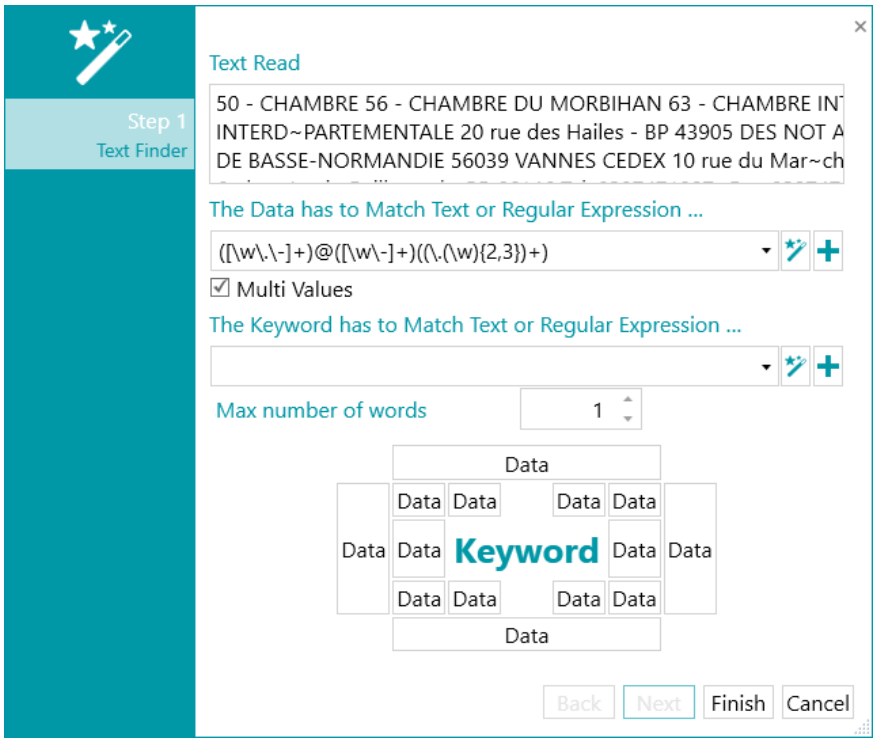
- Extract multiple values on a page
- Improved processing speed by providing the maximum number of words for Data and Keyword.

2. Multiple values extraction

In the first version of Text Finder, the search was automatically stopped when a text sequence was matching the search criteria.

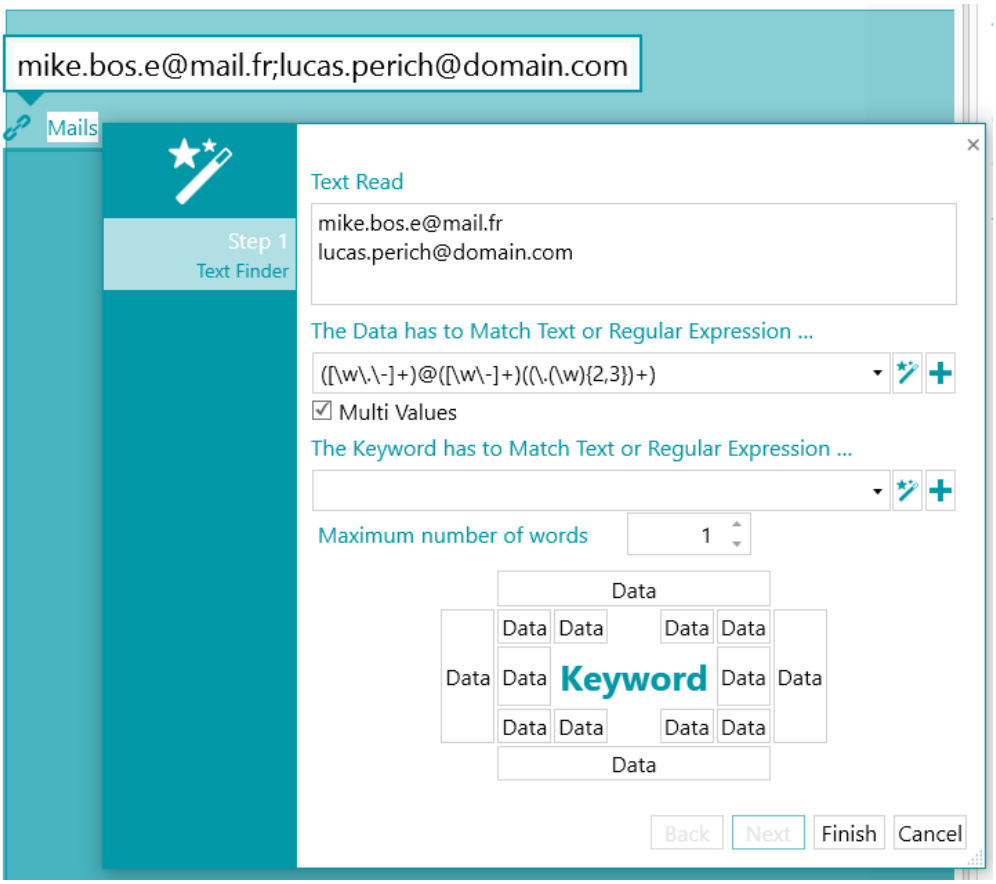
Now, a new checkbox “Multi Values” has been added.

It allows you to extract multiple text sequences if there is more than one that is matching your search criteria’s.



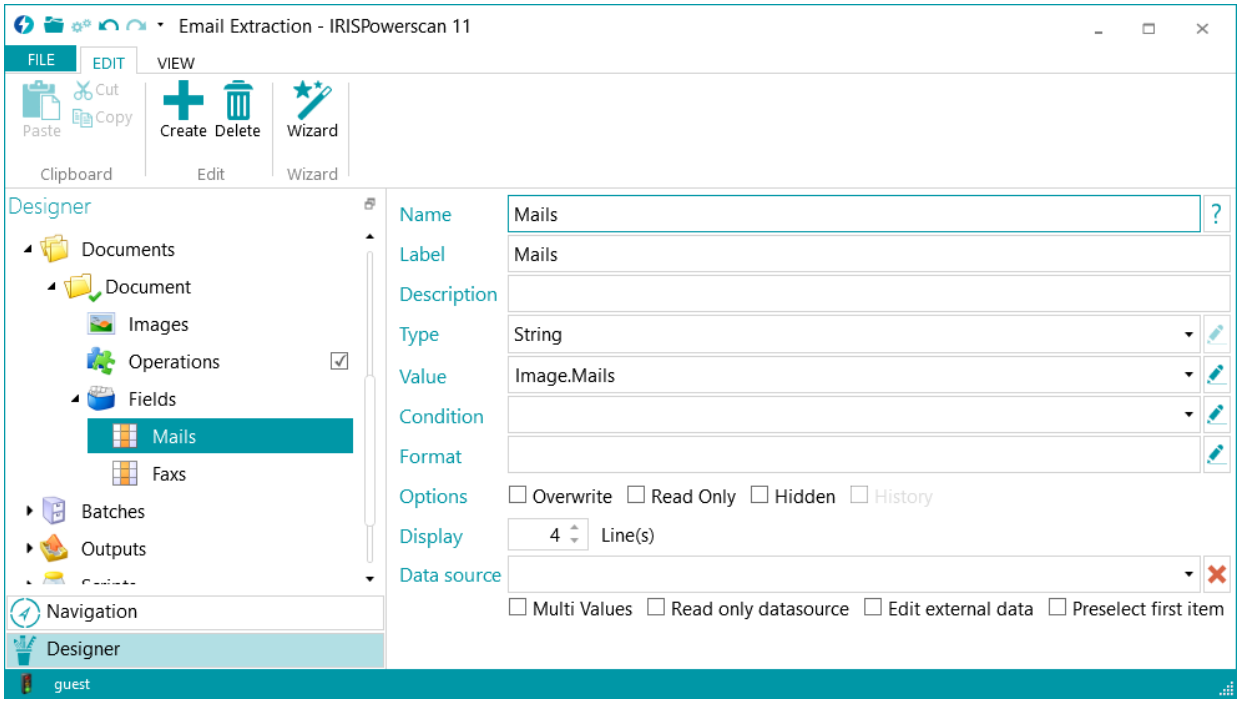
Canon
CANON GROUP

For instance, if you want to extract multiple emails that are on the same page, this what you will see in the designer.

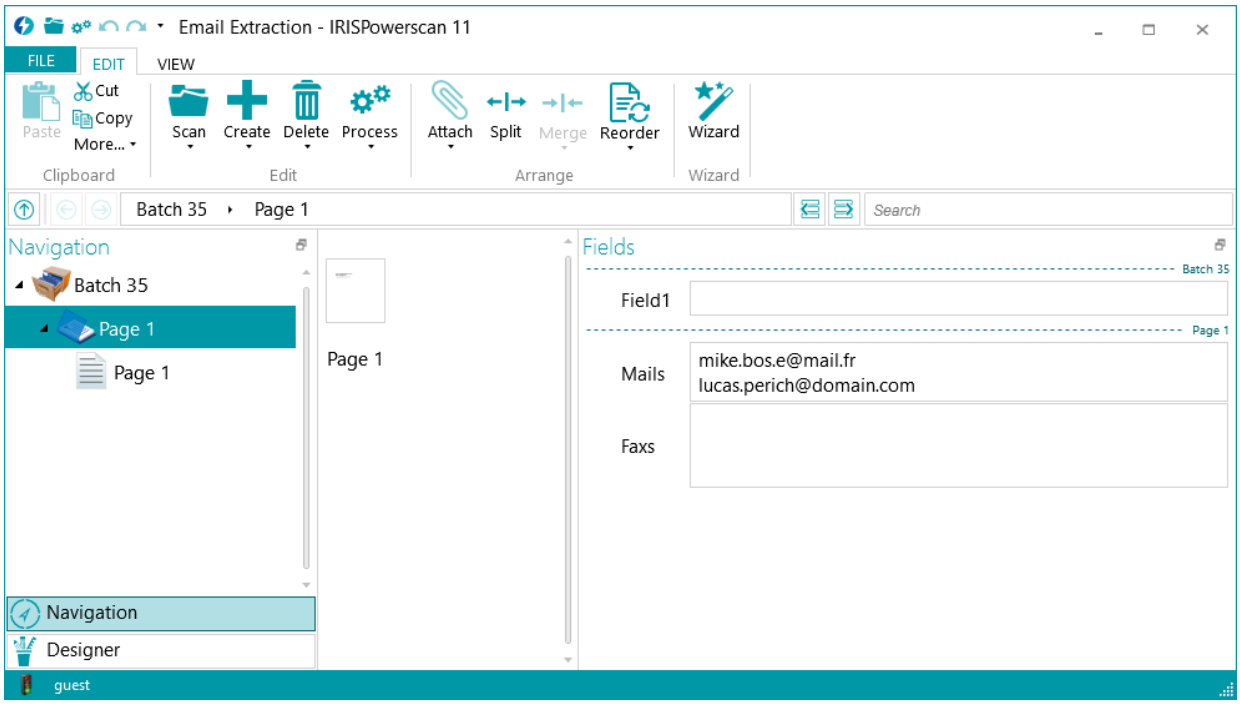


The extracted text sequences are separated by “;”.

When you link a field with the extraction area, if you want to visually hide the ";" and show the result on multiple lines, just set at least 2 lines for the Display property.



And this is what you will see in Navigation.



3. Improved Processing Time

A new parameter “Maximum number of words” has been added.

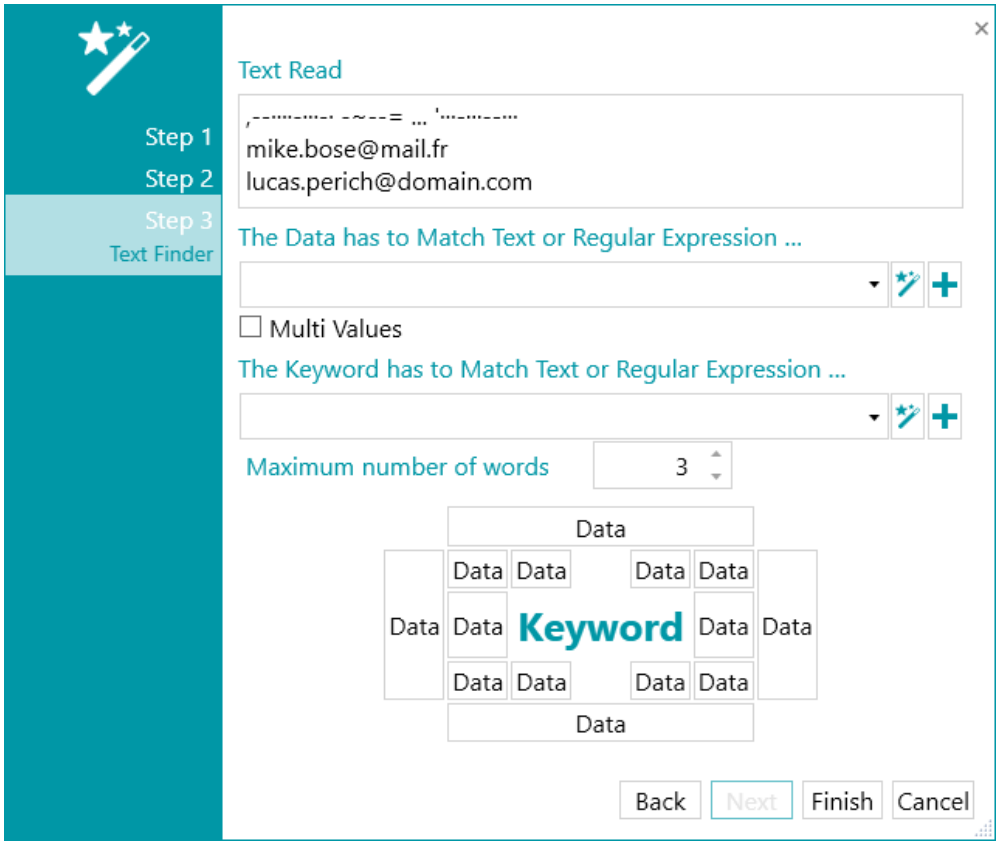
With this parameter, you will define the maximum number of words for the Data and Keyword.

A word is considered as a range of characters without space.

“E-mail” is one word, “E mail :” is three words, and so on.

This parameter allows to highly improve the speed of the extraction as it will reduce the scope of the analysis.

The default value of this parameter is 3.



Document to Knowledge™
A CANON COMPANY

4. Legal Notice

4.1 *Copyrights*

©2002-2019 I.R.I.S. All Rights Reserved.

I.R.I.S. SA/NV
Rue du Bosquet 10
B-1348 Louvain-la-Neuve

P: +32 (0) 10 45 13 64
F: +49 (0) 10 45 34 43

www.irislink.com

CANON GROUP