Contenido

[2. Personalize your own filter 2](#_Toc19002633)

[1. Get sintaxStructure from consolidate.csv file 3](#_Toc19002634)

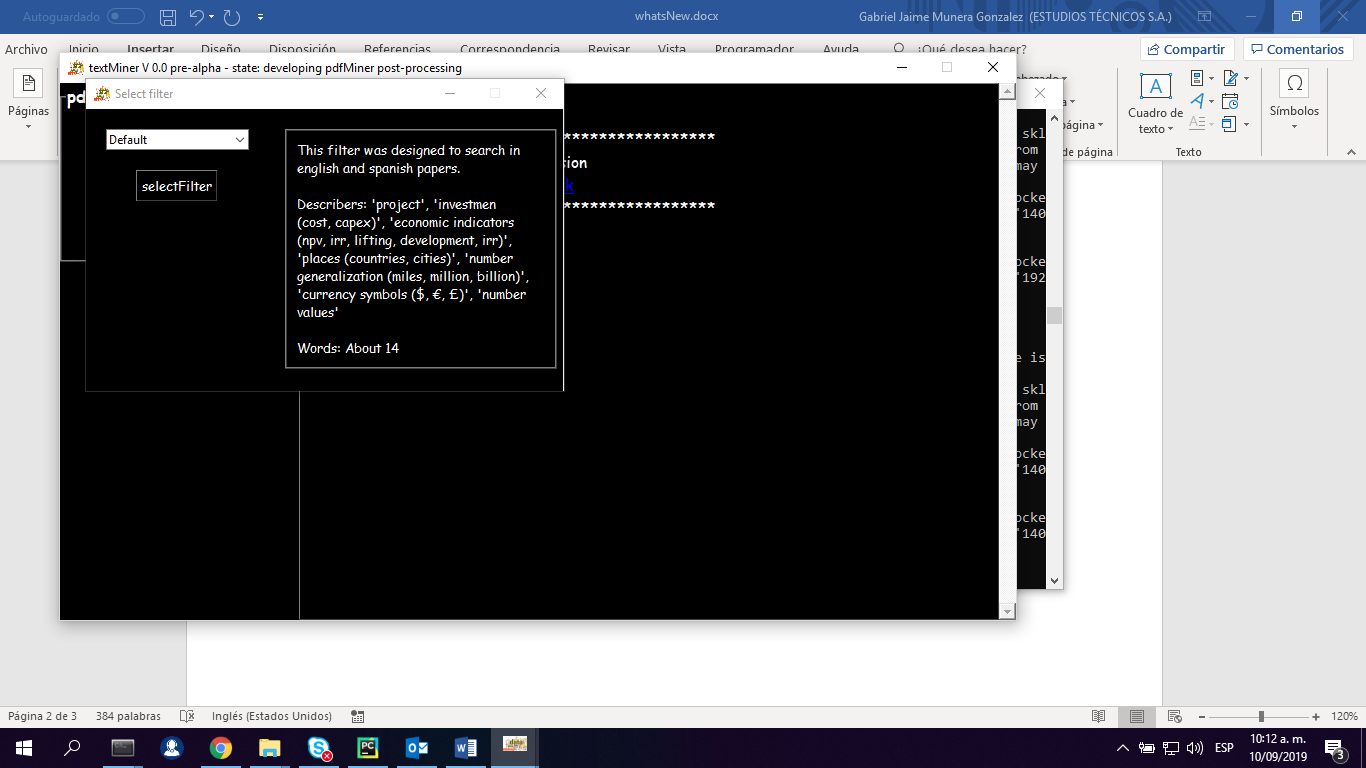
**DATE: September 09, 2019**

* We clean a little beat the code preventing trash accumulation.
* There are three more available filters in the application. To implement this new feature we include a new pop up window to select and edit filters:
  + Complex filter: Searching for project complexities in text.
  + Proj Issues: Search for issues references in text.
  + Personalized: This filter type let you select structure to find in the text, I’m sure is a great and versatile tool to find what you want in papers.

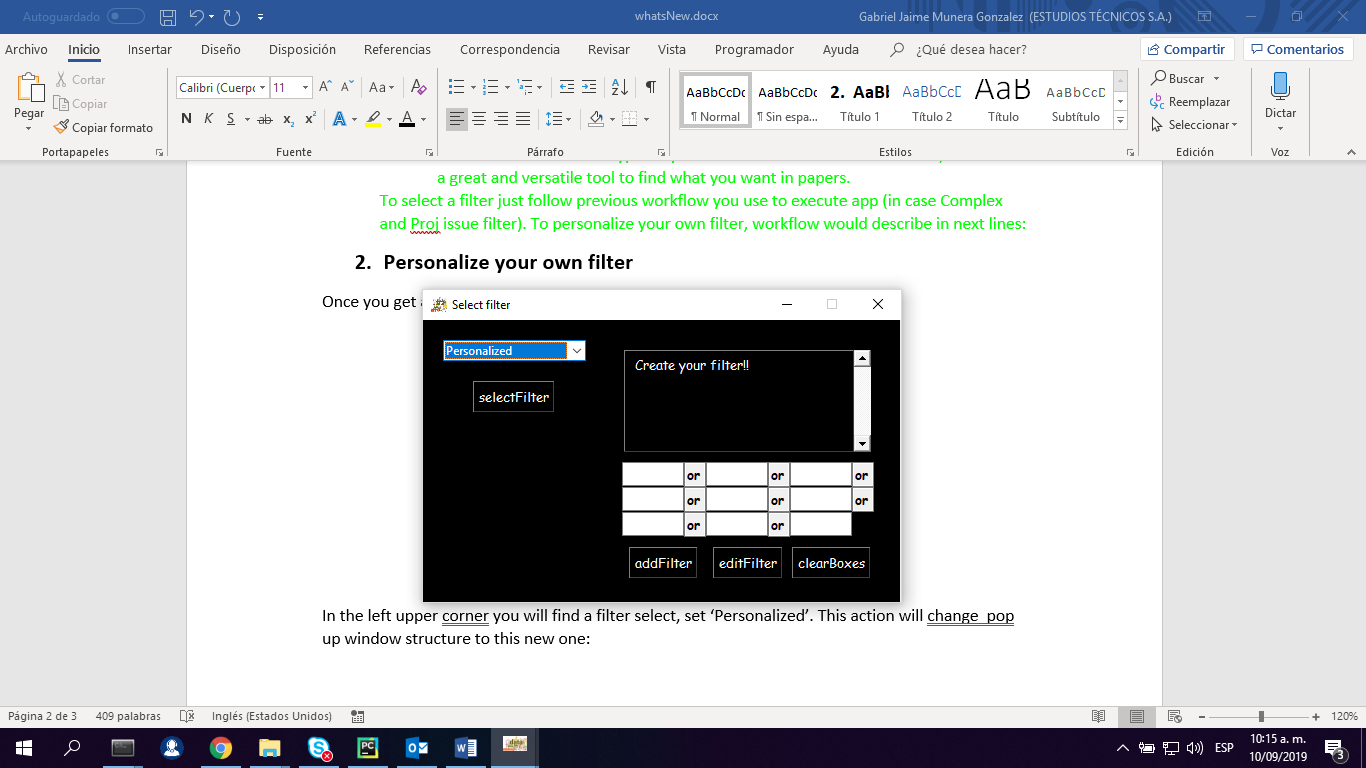
To select a filter just follow previous workflow you use to execute app (in case Complex and Proj issue filter) and click on ‘selectFilter’ button. To personalize your own filter, workflow would describe in next lines:

# Personalize your own filter

Once you get action to run a filter in text the next pop up window will be raise:



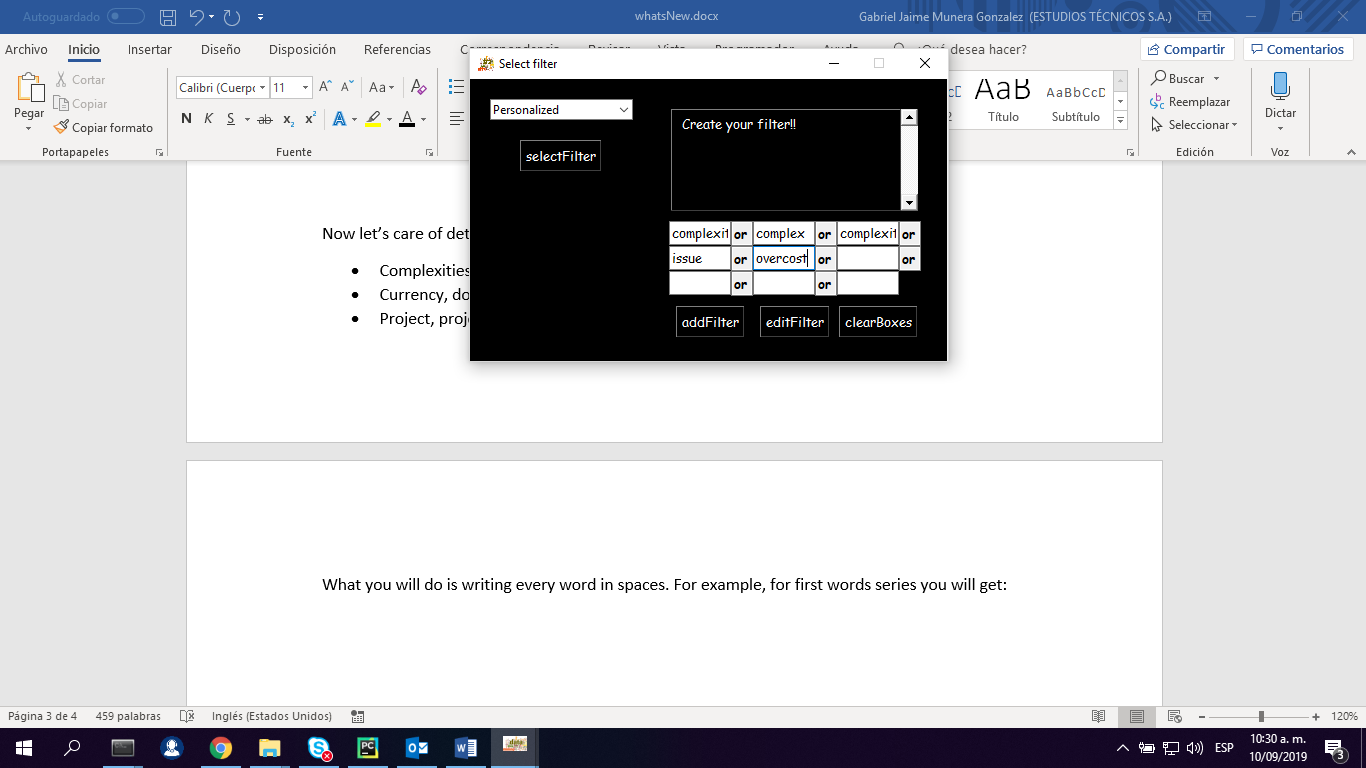
In the left upper corner you will find a filter select, set ‘Personalized’. This action will change pop up window structure to this new one:



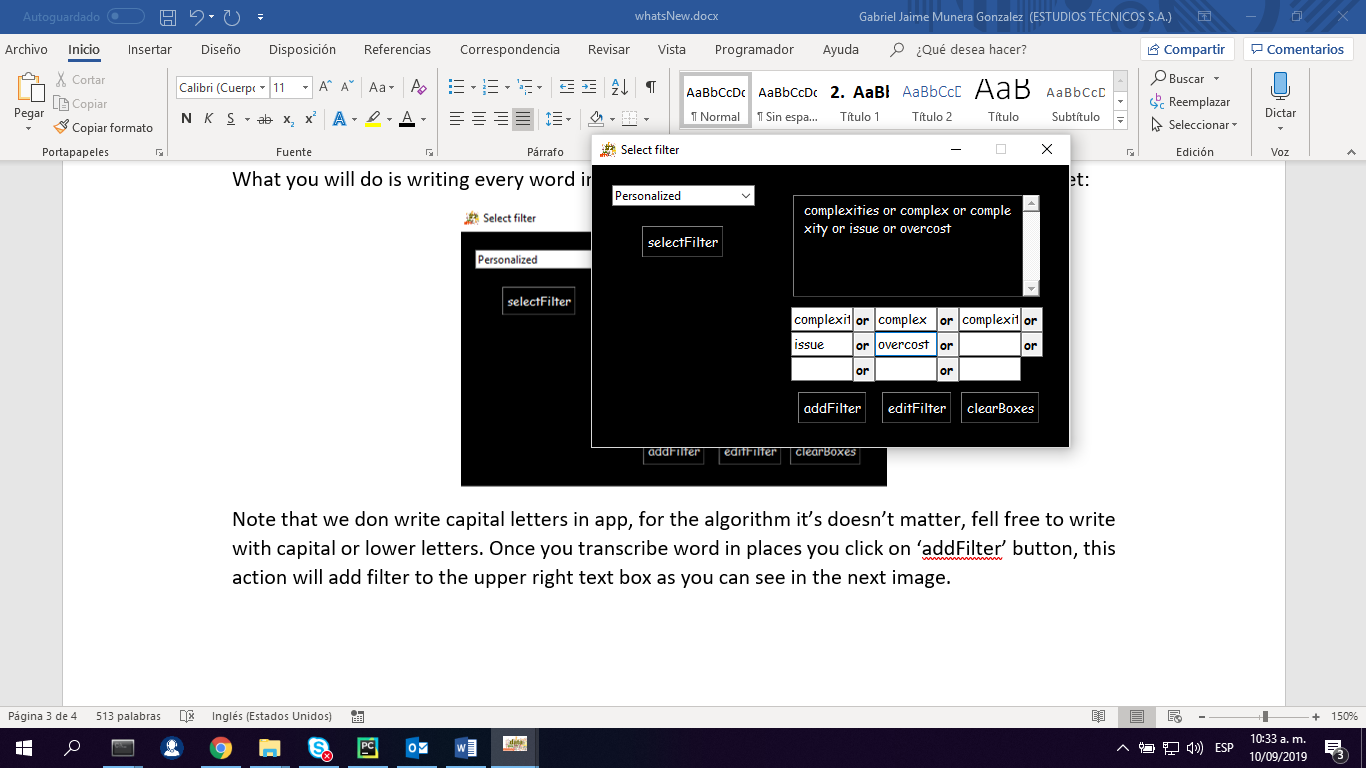
Now let’s care of details, suppose you want to find the next structures in your files:

* Complexities, complex, complexity, issue, overcost.
* Currency, dollar, peso, pound, yen, $, £, ¥.
* Project.

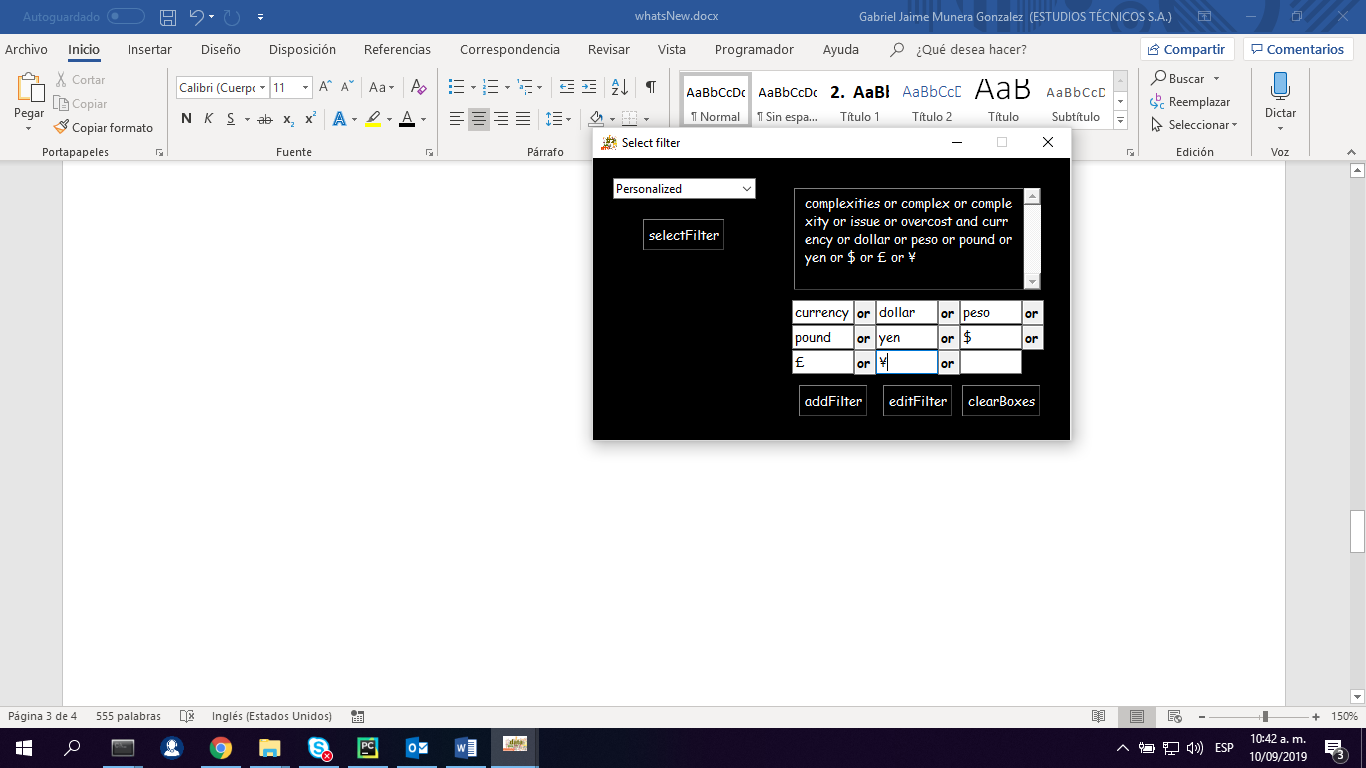
What you will do is writing every word in spaces. For example, for first words series you will get:



Note that we don’t write capital letters in app. For the algorithm it’s doesn’t matter, fell free to write with capital or lower letters. Once you transcribed words in the places, you do click on ‘addFilter’ button, this action will add filter to the upper right text box as you can see in the next image.

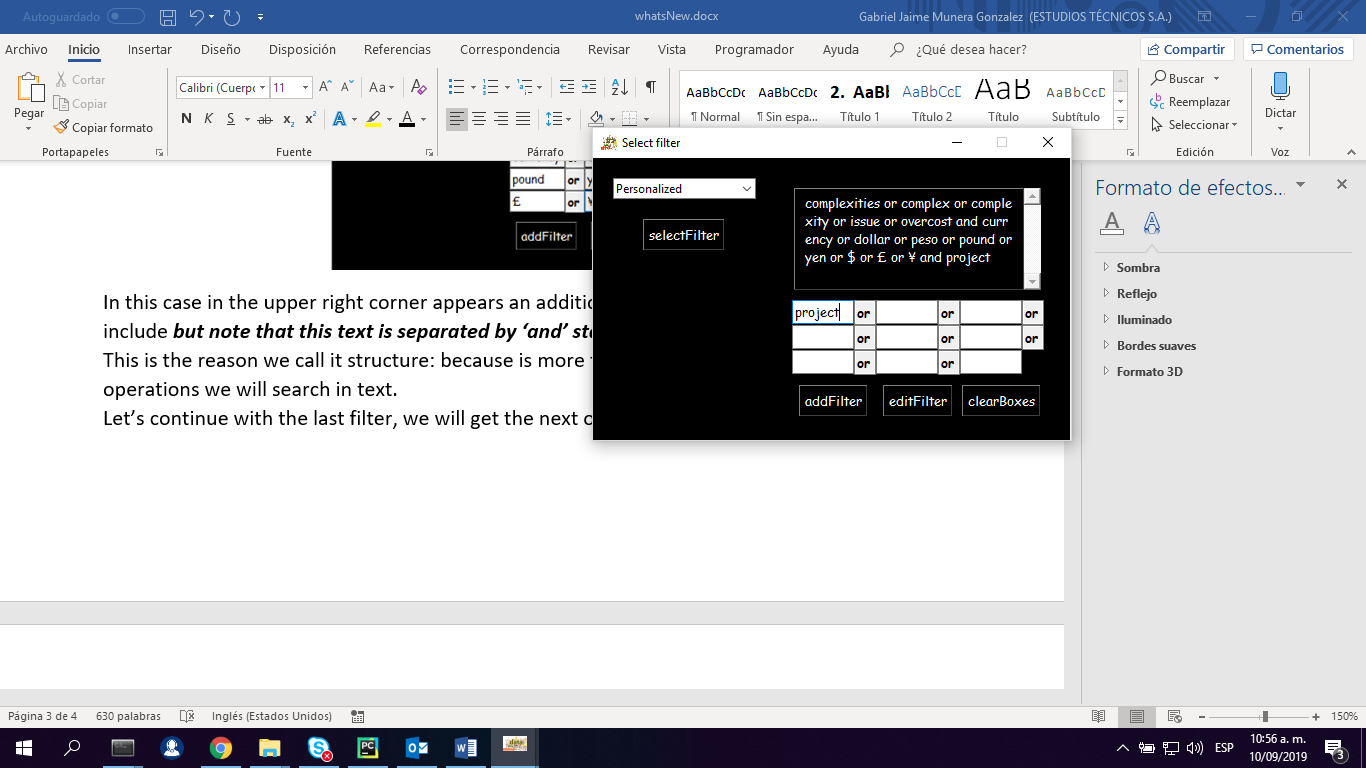


Note that app add for you a conditional structure between word (‘or’ in this case). Next to this step we do click on clearBoxes to continue editing filters. Let’s include second filter in structure, this time it will see as follow:



In this case in the upper right corner appears an additional ‘or’ separated text containing words we include ***but note that this text is* *separated by ‘and’ statement from previous filter we define.*** This is the reason we call it structure: because is more than a filter: is a sequence of logical operations we will search in text.

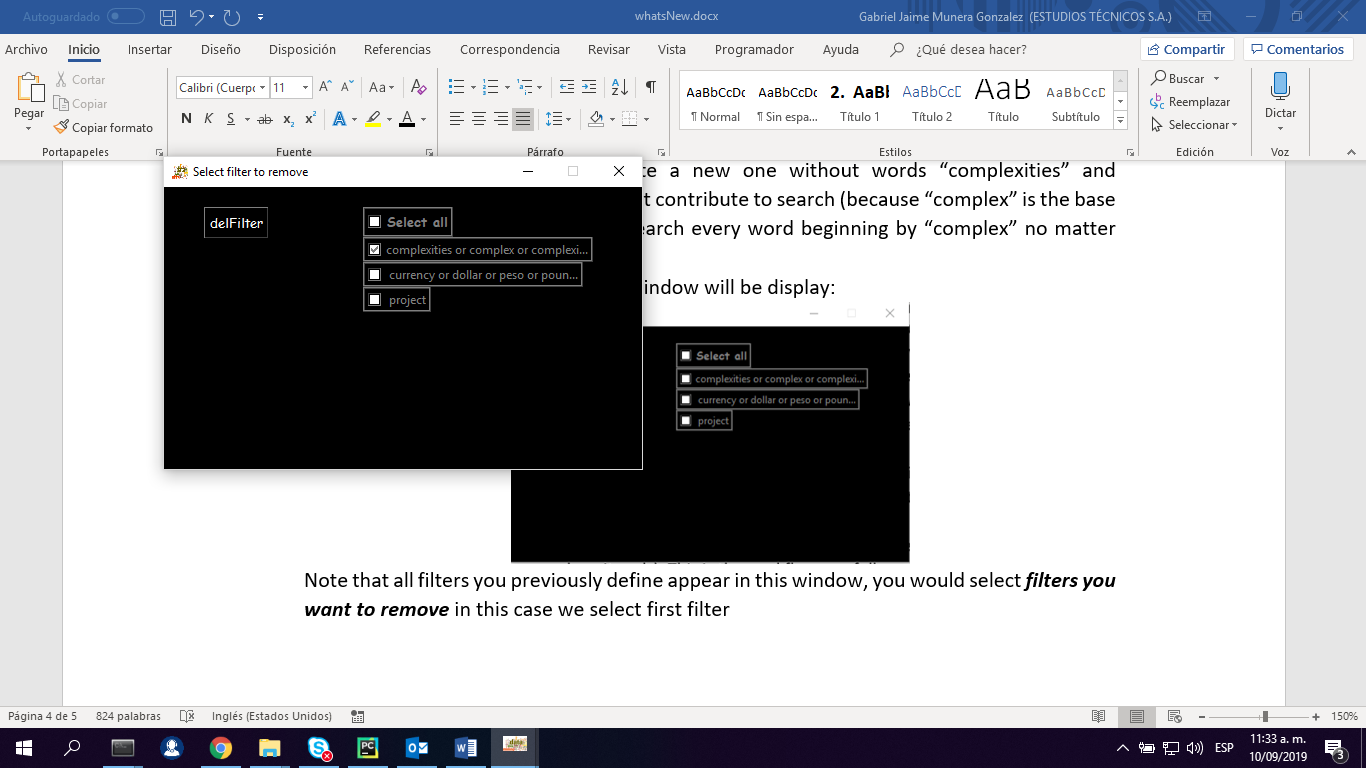
Let’s continue with the last filter, we will get the next output:



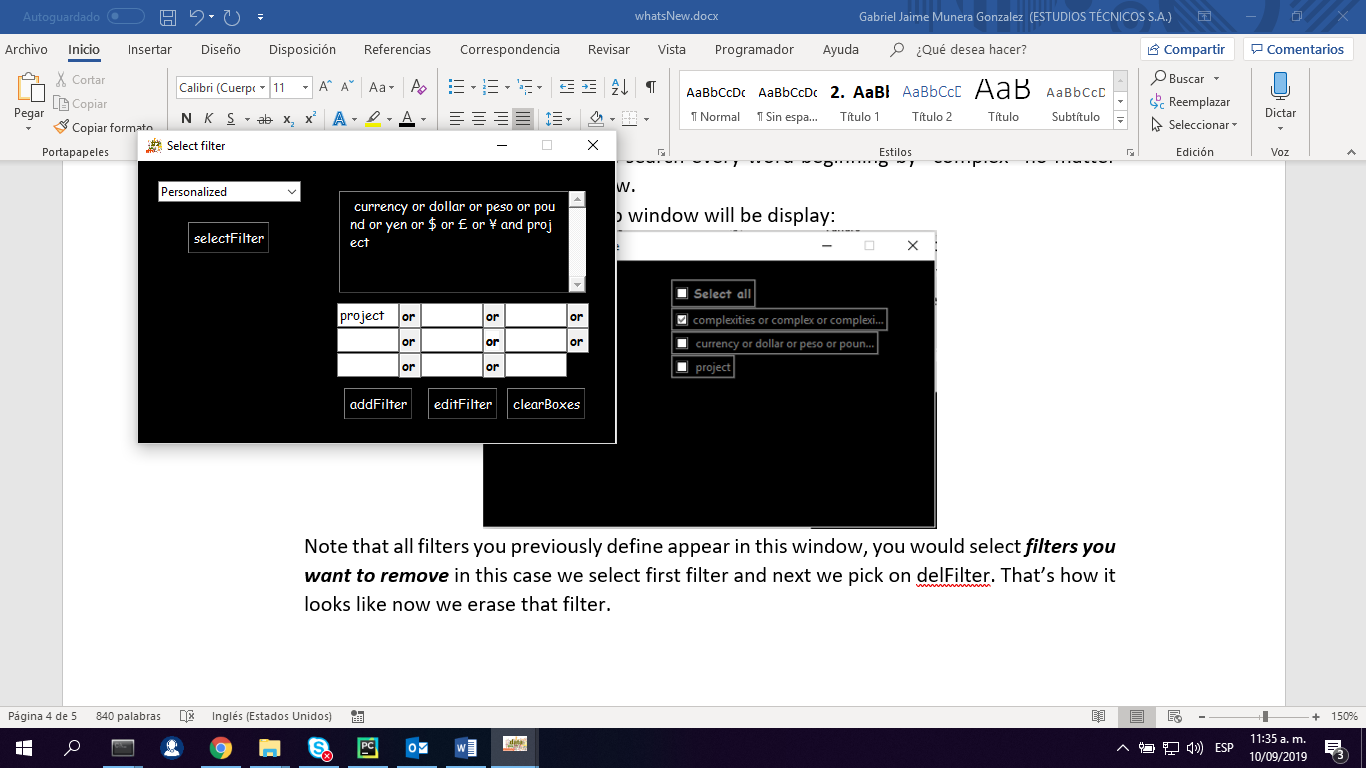
Once you finish this step you get a filtering structure to apply to the text, you can click on selectFilter button and run your filter in the text as previously you do with pre-defined filter structure. It’s important to say that ***filter will be saved while you don’t restart application, I mean, next time you run a filter and select ‘Personalized’ on upper left corner you will get the filter you define WHILE NOT RESTART APPLICATION.***There’s a last feature I want to introduce you, I mean editing your personalized filters.

For this feature suppose you want to erase filter composed by words “complexities”, “complex”, “complexity”, “issue”, “overcost“ and create a new one without words “complexities” and “complexity” since you realize that words don’t contribute to search (because “complex” is the base form of that two words and ***algorithm will search every word beginning by “complex” no matter how it ends***). This is the workflow you follow.

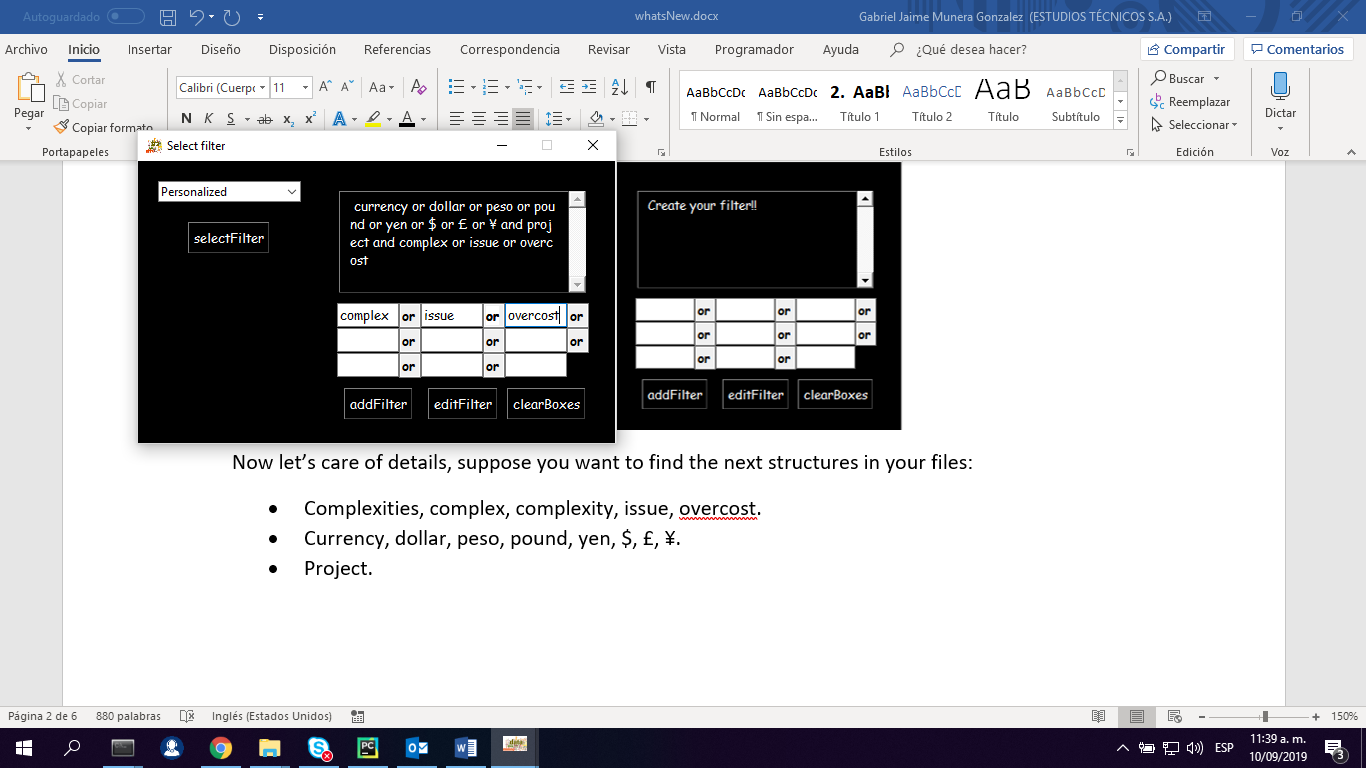
* Click on ‘editFilter’, the next pop up window will be display:



Note that all filters you previously define appear in this window, you would select ***filters you want to remove*** in this case we select first filter and next we pick on delFilter. That’s how it looks like now we erase that filter.



The first filter was remove. Now we can define a new filter that doesn’t have words “complexities” and “complexity” as we did in the previous steps (write words in white spaces and click on ‘addFilter’). This is the result:



I hope you enjoy this new feature!!!

That’s all for a while 😊

**DATE: August 15, 2019**

* We solve some technical issues in clarify function. Now function make better its role as paragraph splitter.
* Now we have available trained RandoForest classifier to make a second filter on information. To use it follow ‘sintaxSructure’ click button. This is how it works:

# Get sintaxStructure from consolidate.csv file

This functionality is based in four random forest model wich recognize patterns and classifies text in four categories:

* Scope: It relate to project scope.
* Expectations: It is basically what they think project can be in terms of monetary revenue and other benefits.
* Timeline: What project recent activity was made and if there are some issues that made project freeze.
* Cost: Monetary values associated to project.

Model learn in more than 800 examples of sentences previously categorized. Categorizing activity also return a list of engineering and petroleum companies and places mentioned in text. Of course there is too small train data and its necessary to improve the model training it again as we have more data.

**How to use it:**

* As you click on button a message advising you to previous database remove will be display. Type ‘y’ to continue.
* A window asking you for a file will raise, select ‘consolidate.csv’ from your ‘report’ folder inside PDF’s database.
* Function will run extracting text for every paragraph with features previously mentioned.