

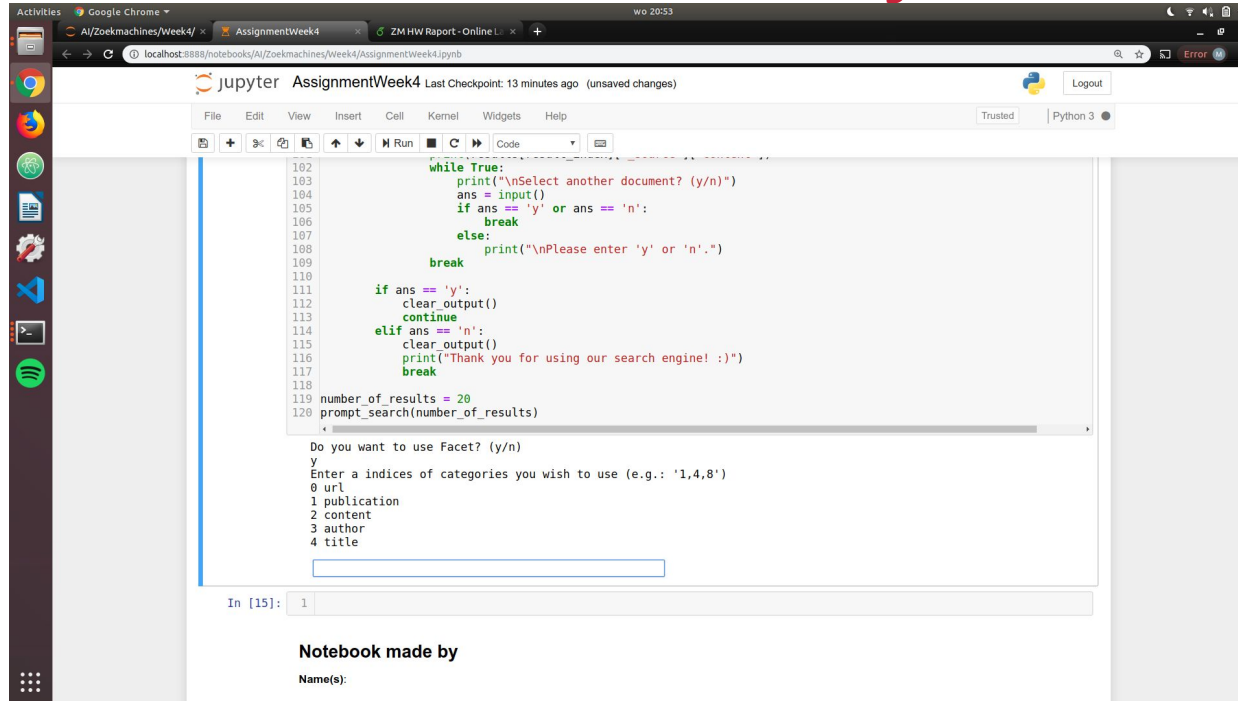
# Search-engine with Elasticsearch API

Mats Gonggrijp  
Midas Amersfoort  
Rutger van Woerkom

# Considerations (all implemented)

- Usability: intuitive and easy to understand
- Speed: largely determined by the API
- Facet search: choice to use categories
- Use Wordcloud to represent documents

# Output: Facet search = 'y'



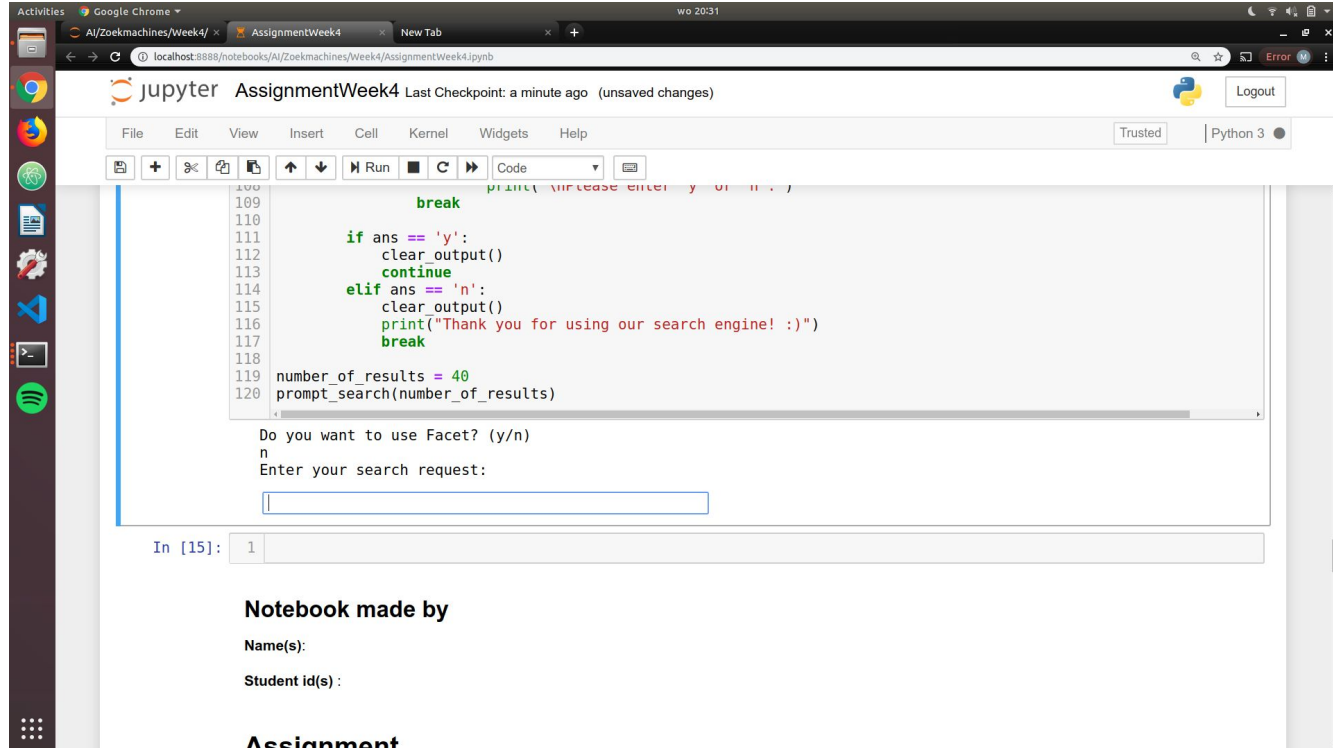
```
102     while True:
103         print("\nSelect another document? (y/n)")
104         ans = input()
105         if ans == 'y' or ans == 'n':
106             break
107         else:
108             print("\nPlease enter 'y' or 'n'.")
109     break
110
111     if ans == 'y':
112         clear_output()
113         continue
114     elif ans == 'n':
115         clear_output()
116         print("Thank you for using our search engine! :)")
117         break
118
119     number_of_results = 20
120     prompt_search(number_of_results)
```

Do you want to use Facet? (y/n)  
y  
Enter a indices of categories you wish to use (e.g.: '1,4,8')  
0 url  
1 publication  
2 content  
3 author  
4 title

In [15]: 1

Notebook made by  
Name(s):

# Output Facet Search = 'n'



The screenshot displays a Jupyter Notebook titled "AssignmentWeek4" running in a Google Chrome browser. The notebook contains a Python script that simulates a search engine interface. The script prompts the user to use Facet (y/n) and enter a search request. The output shows the user has entered 'n' and a search request field is displayed.

```
108         break
109     print("Increase entry of n.")
110
111     if ans == 'y':
112         clear_output()
113         continue
114     elif ans == 'n':
115         clear_output()
116         print("Thank you for using our search engine! :)")
117         break
118
119 number_of_results = 40
120 prompt_search(number_of_results)

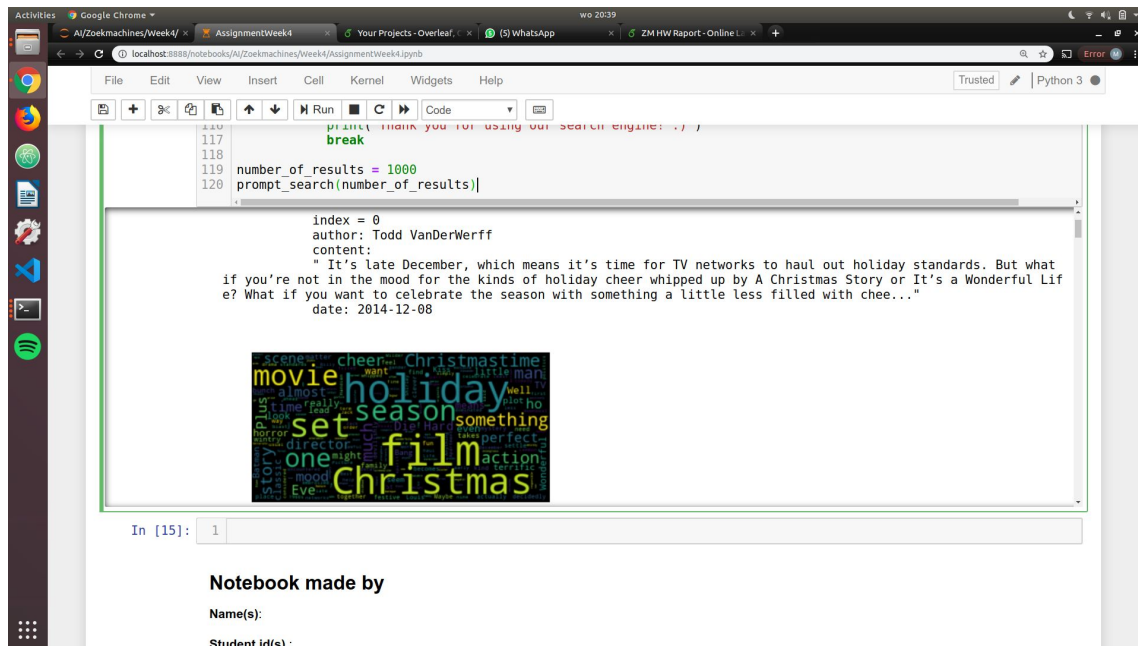
```

Do you want to use Facet? (y/n)  
n  
Enter your search request:

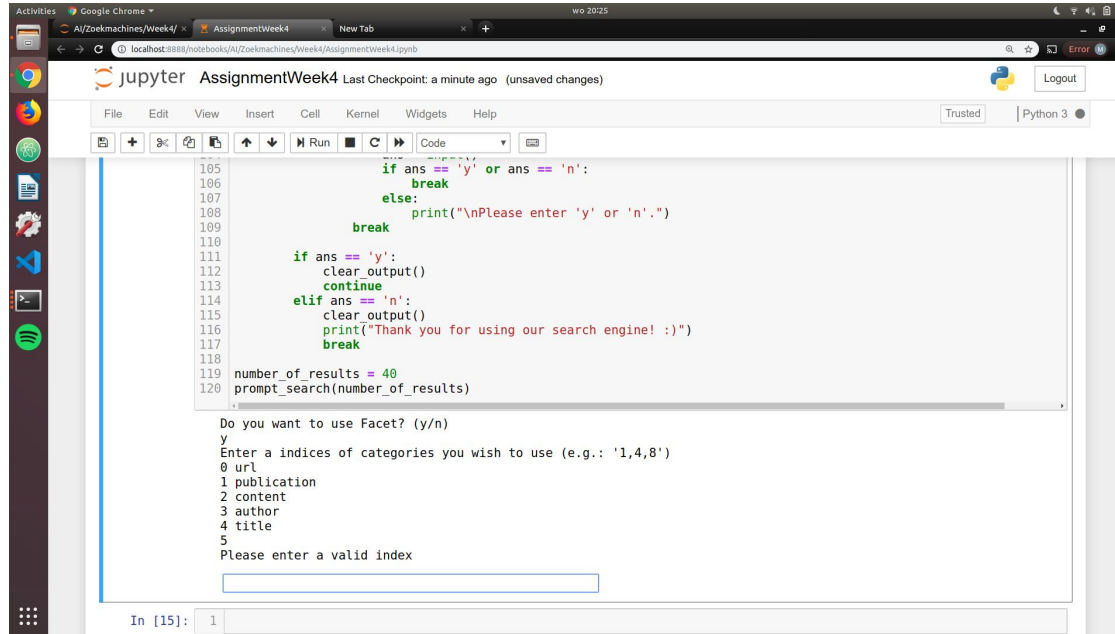
In [15]: 1

**Notebook made by**  
**Name(s):**  
**Student id(s) :**  
**Assianment**

Output: FS='y'  $\rightarrow$  correct category  
sequence = document set



# Output: FS='y' → incorrect category sequence



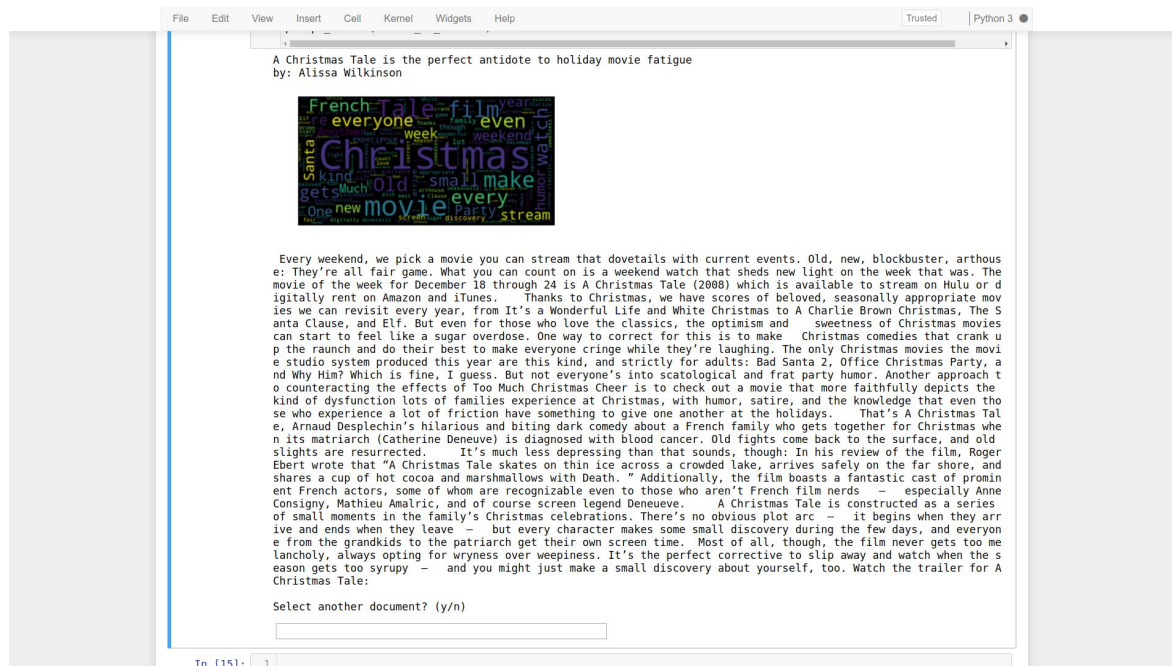
The screenshot shows a Jupyter Notebook window titled "AssignmentWeek4" with a "Last Checkpoint: a minute ago (unsaved changes)" status. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running, and code execution. The code editor displays a Python script with line numbers 105 to 120. The script contains conditional logic for handling user input 'y' or 'n', including a loop for category selection. The output area shows the execution results, including a prompt for Facet usage and a list of category indices (0 to 5) with their corresponding labels (url, publication, content, author, title). A text input field is visible below the output, and the status bar at the bottom indicates "In [15]: 1".

```
105         if ans == 'y' or ans == 'n':
106             break
107         else:
108             print("\nPlease enter 'y' or 'n'.")
109             break
110
111     if ans == 'y':
112         clear_output()
113         continue
114     elif ans == 'n':
115         clear_output()
116         print("Thank you for using our search engine! :)")
117         break
118
119     number_of_results = 40
120     prompt_search(number_of_results)
```

Do you want to use Facet? (y/n)  
y  
Enter a indices of categories you wish to use (e.g.: '1,4,8')  
0 url  
1 publication  
2 content  
3 author  
4 title  
5  
Please enter a valid index

In [15]: 1

# Output: correct facet or non-facet search → correct document ID



# Evaluation: two judge Kappa score

- Use the Kappa score as defined by:  
 $(Po - Pe) / (1 - Pe)$
- Takes into account random matching judgement
- Kappas:  
0.6, 0.8 and 0 for the queries used respectively