

Zoekmachines Homework 4 Report

Mats Gonggrijp, Midas Amersfoort, Rutger Van Woerkom

October 2019

1 Introduction

During the building proces we had to make some considerations of how we ask for queries and present the data. In the following section we explain our considerations and why we made our choices. The main focus was simplicity for the user.

2 Considerations

The search engine has to be easy to use and helps users to make efficient queries. We have chosen for a simple prompt based system that first asks users for wether or not they want to use faceted search, their input will be 'y'=yes, or 'n'=no. If they choose 'n', then they are simply asked to input a query, that will be forwarded to the ElasticSearch API. If they choose 'y', they will be presented with another prompt and a set of possible features to choose from. They are asked for an input, which will be a string of numbers seperated by commas, where each number corresponds to a feature. This facet search will narrow down and make the process nmore efficient and accurate when the user already knows in what sort of category they are looking for.

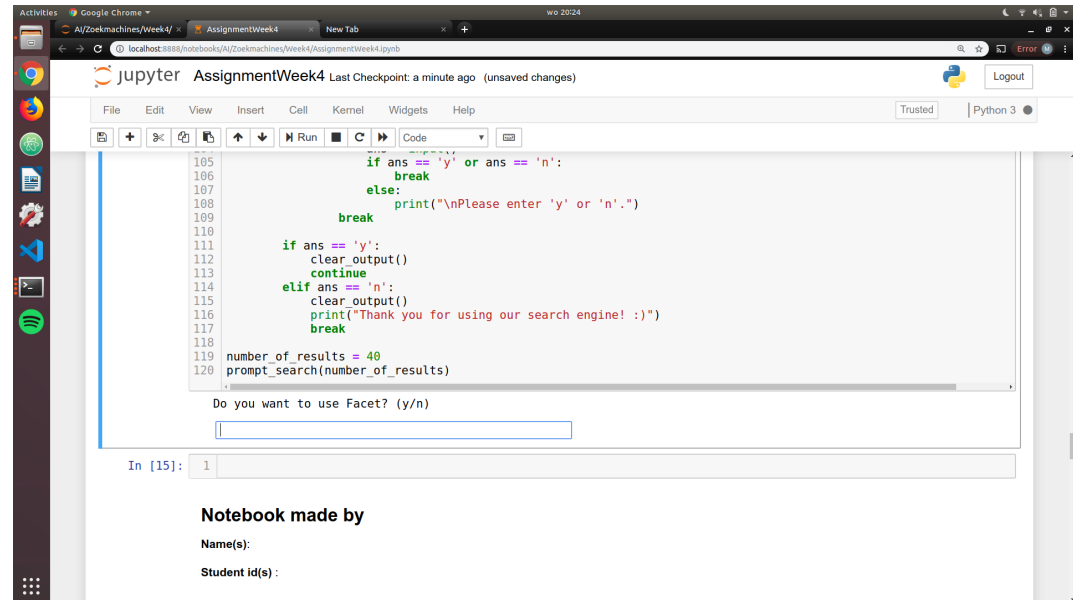
3 Output

The engine returns a ranked list of documents with some extra information to make it clear what is going on. For the text filled datasets that we use, we return the first 200 characters from the text content and the name of the author. This will help the user in deciding document relevancy. We return a list of the 10 most relevant documents.

4 Demo

Below we show some screenshots of how the system works

4.1 Ask user if they want to use facet search



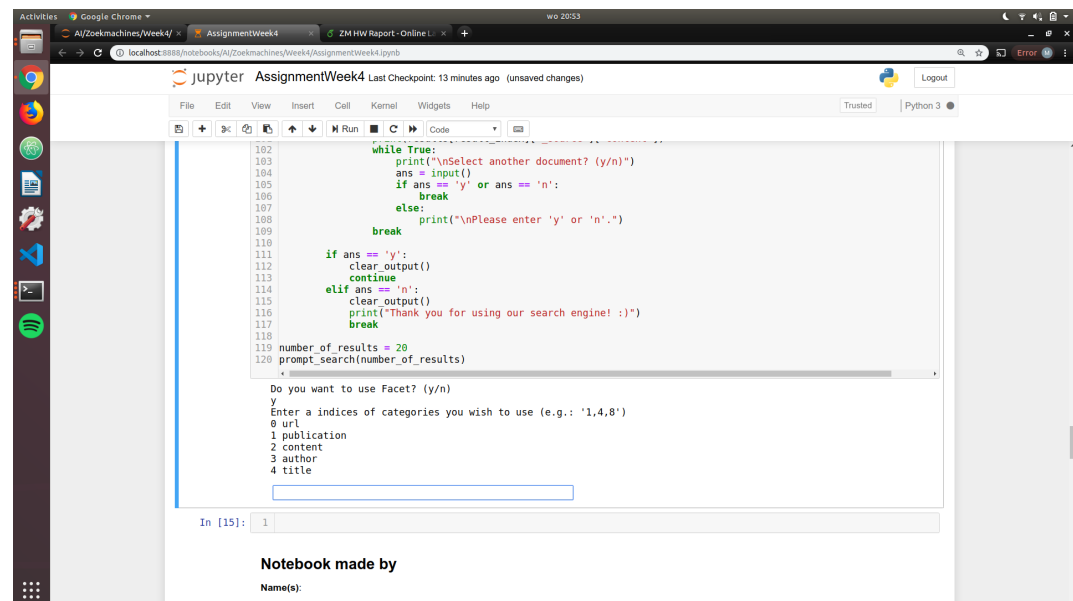
```
105         if ans == 'y' or ans == 'n':
106             break
107         else:
108             print("\nPlease enter 'y' or 'n'.")
109             break
110     if ans == 'y':
111         clear_output()
112         continue
113     elif ans == 'n':
114         clear_output()
115         print("Thank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

```

In [15]: 1

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

4.2 IF 'y', ask for facet set



```
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     if ans == 'y':
111         clear_output()
112         continue
113     elif ans == 'n':
114         clear_output()
115         print("Thank you for using our search engine! :)")
116         break
117
118 number_of_results = 20
119 prompt_search(number_of_results)
120
Do you want to use Facet? (y/n)
y
Enter a indices of categories you wish to use (e.g.: '1,4,0')
0 url
1 publication
2 content
3 author
4 title

```

In [15]: 1

Notebook made by
Name(s):

The screenshot shows a Jupyter Notebook running in a Google Chrome browser. The notebook is titled "AssignmentWeek4" and has a "Last Checkpoint: a minute ago (unsaved changes)" status. The code in the notebook is as follows:

```

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)

```

The output of the notebook shows the following text:

```

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

```

The input field at the bottom of the notebook is empty, indicating that the user has not yet entered a valid index.

Google Chrome

AssignmentWeek4

Your Projects - Overleaf

WhatsApp

ZM HW Report - Online L

localhost:8888/notebooks/AI/Zeokmachines/Week4/AssignmentWeek4.ipynb

jupyter AssignmentWeek4 Last Checkpoint: 5 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

```

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

5

Please enter a valid index

4

Enter your search request:

christmas

In [15]: 1

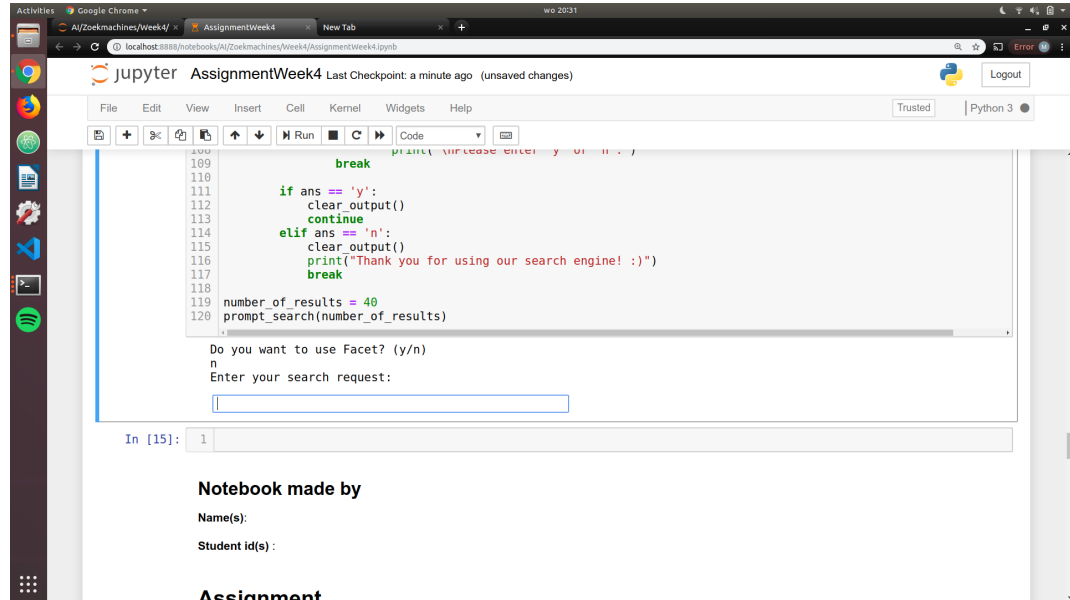
Notebook made by

Name(s):

Student id(s):

Assignment

4.3 IF 'n', ask for simple query



The screenshot shows a Jupyter Notebook titled "AssignmentWeek4" with a "Last Checkpoint: a minute ago (unsaved changes)" status. The notebook is running on a local host. The code cell contains the following Python code:

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

Below the code cell, the output shows a prompt: "Do you want to use Facet? (y/n)" with the input "n". Below that, it says "Enter your search request:" followed by an empty text input field.

In [15]: 1

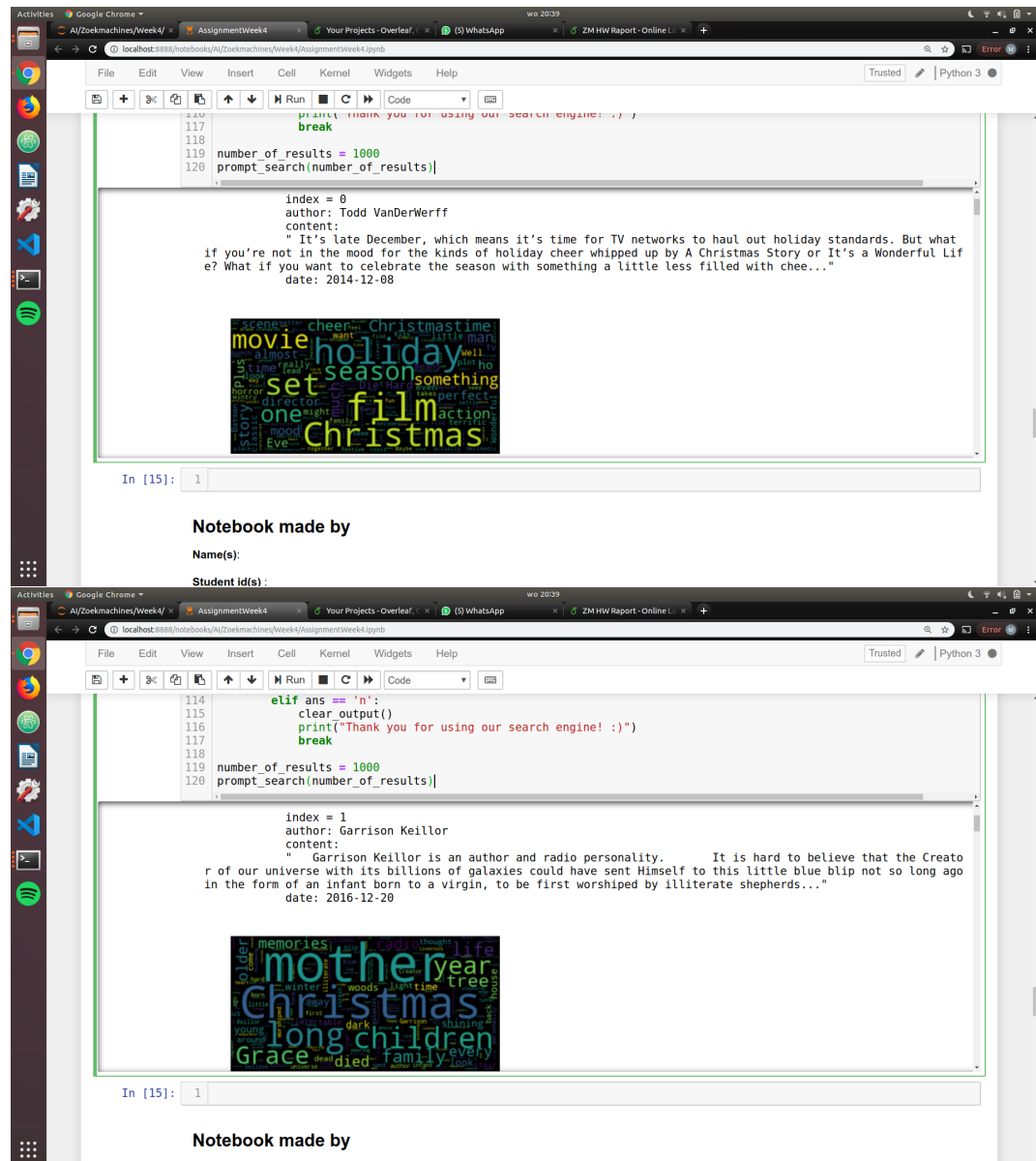
Notebook made by

Name(s):

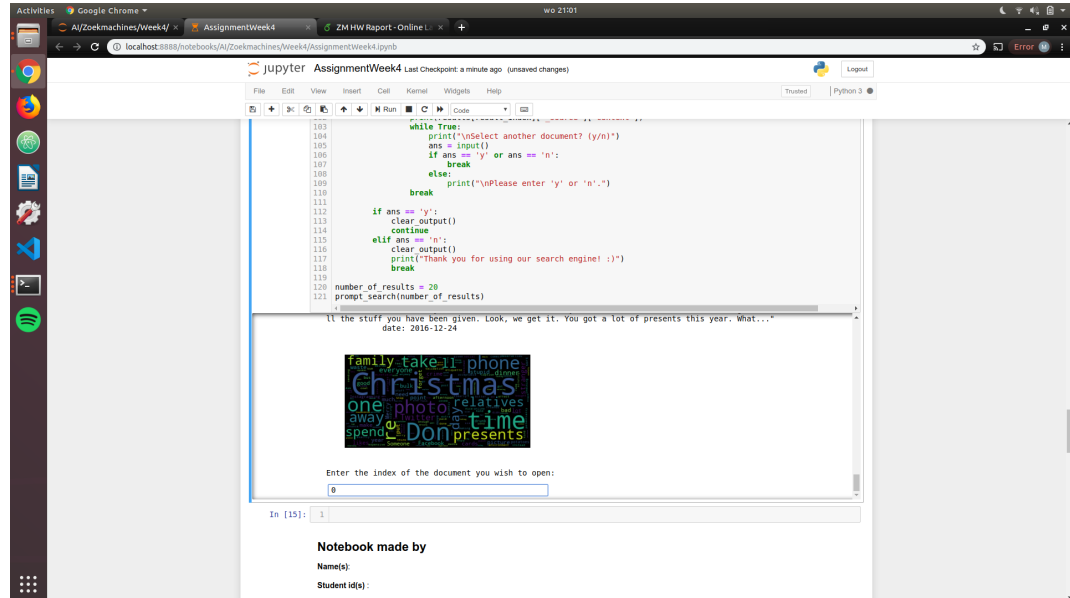
Student id(s) :

Assianment

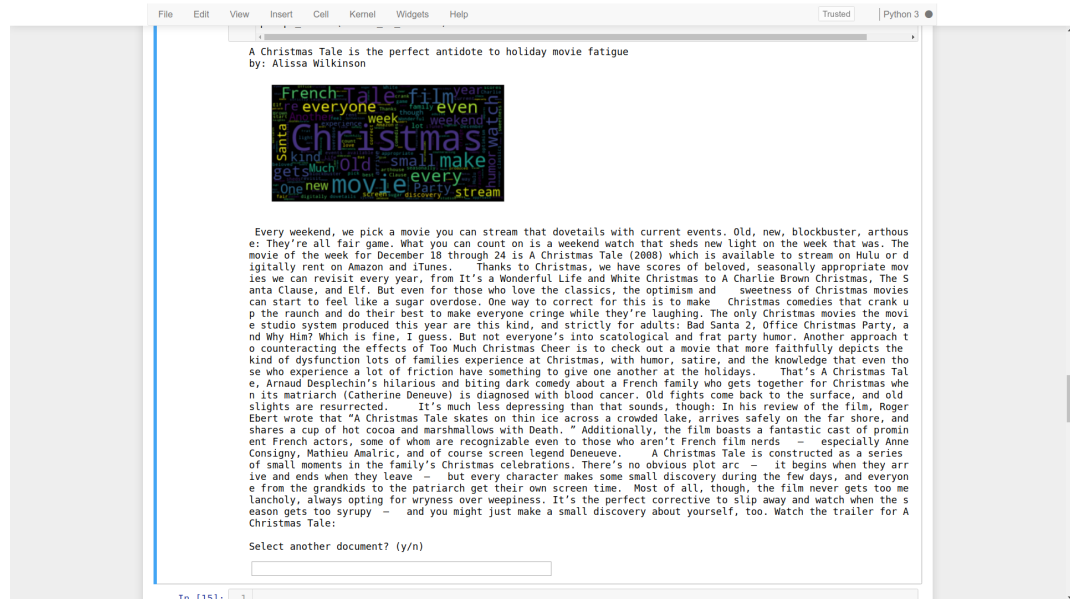
4.3.1 Return simple query results



4.4 Return results, ask document selection



4.5 Return selected document content



5 Evaluation

The search engine works well for simple queries, but quickly becomes less accurate with more complicated constructions. The faceted search clearly helps with speed and accuracy, given that the user understands the categories.

Below are three examples of queries query: "brexit demonstrations 2017" Percentage of agreement: 80%

Cohen's kappa: 0.6000000000000001

Substantial agreement

doc_id	judge1	judge2
1	1	1
2	1	1
3	1	0
4	0	0
5	1	1
6	1	0
7	0	0
8	1	1
9	1	1
10	0	0

query: "wall in mexico demonstration" Percentage of agreement: 90%

Cohen's kappa: 0.8

Substantial agreement

doc_id	judge1	judge2
1	1	1
2	1	1
3	0	0
4	0	0
5	1	1
6	1	1
7	0	0
8	1	1
9	0	1
10	0	0

query: "christmas movies about cats and dogs" Percentage of agreement: 50%

Cohen's kappa: 0

No agreement

doc_id	judge1	judge2
1	0	1
2	1	1
3	1	0
4	1	0
5	0	1
6	1	1
7	0	0
8	0	0
9	0	1
10	0	0