From Objects to Data

Helge Moes

11348801

Work Group 1

Dr. ir. J. Kamps

Assignment 1: #1: BIY Search Engine

Word count: 1774

Introduction of the Assignment

For this assignment, an own search engine had to be built and to be compared to another search engine, such as Google. The goal of this assignment was to make a search engine that would also might over throne Google. To perform such a task, a vertical search engine is considered. This can be characterized as a specialized search engine that considers a specific topic. The performance of general search engines, such as Google and Bing, has been criticized to give a limited number of queries which are also regarded highly ambiguous. A vertical search engine is expected to provide more focused search results compared to a search engine that allows for a general source of knowledge.

The specific topic that will be used for the search engine is Vincent van Gogh. I will take all historical information into account as to portray a non-biased image of him as possible. To practically carry this out, Google Co-op is used. This program, that is provided by Google, enables the user to add certain sources to create a vertical search engine. Moreover, to make a representative comparison, duckduckgo.com is used to find these sources as to compare them to Google.

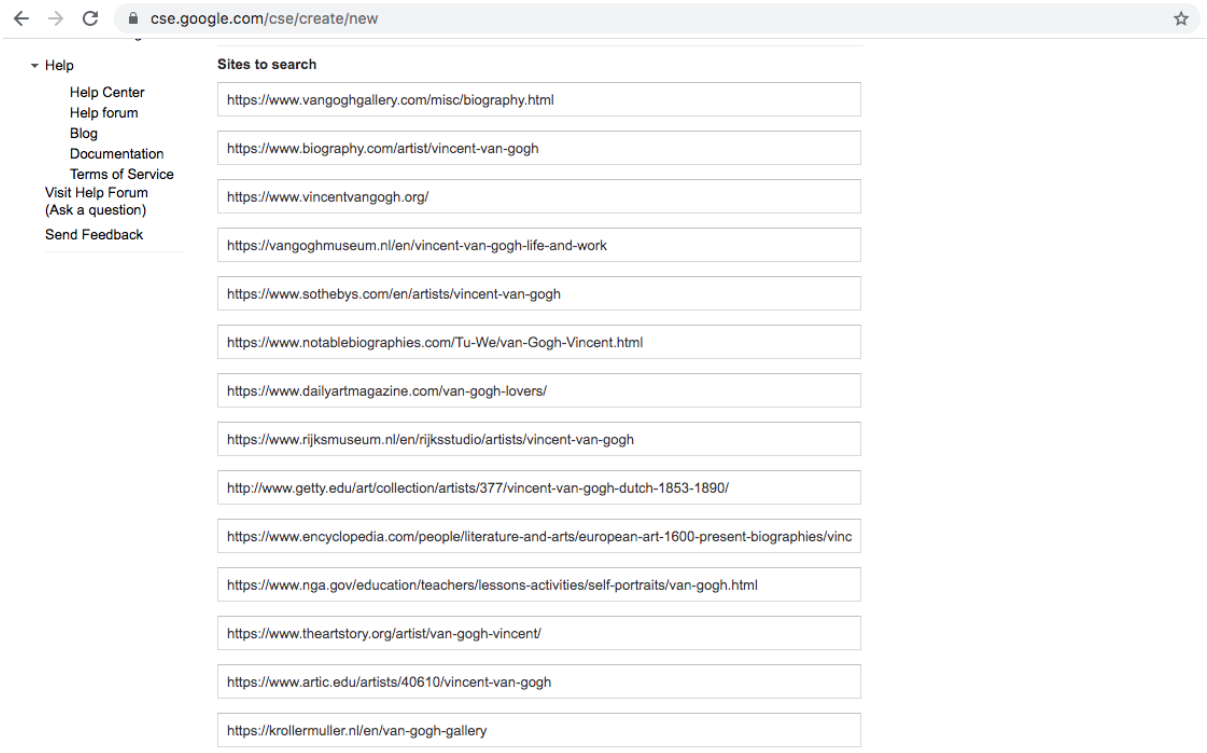
Subject's purpose, findings and issues

The purpose of the vertical search engine is to provide a clear image on Vincent van Gogh. The image of the painter is clouded in mystery, for instance how he presumably killed himself or how he was murdered when working on one of his paintings and was seen walking back to his attic room with a bullet in the chest. Furthermore, artists create work in his art as to sell paintings worth thousands of euros. I want to allow accessibility for art connoisseurs to a realistic collection of sources that show all his work and history to prevent

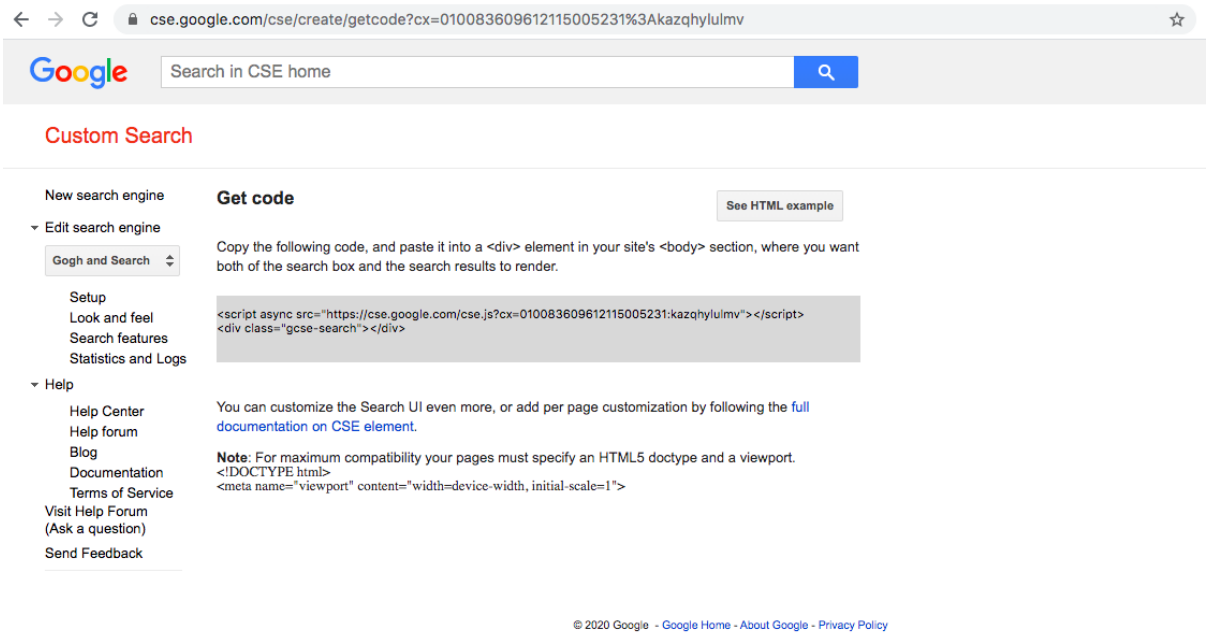
any further forgery or other malpractices in the name of the Dutch painter. Therefore, with this search engine, the main focus is to try to find clear sources on this subject matter and to compare that to the commercialized information, that is provided by Google, that may have altered art history as of today and make people aware of this.

Furthermore, the selection principles will be limited to academic sources or sources that are supported by art or history organizations, such as museums or cultural media companies, for example npofocus.nl, www.vangoghmuseum.nl or historiek.net. On the other hand, sites that will not be considered are smaller biased news outlets or regular social media related content, for instance sources as allposters.com, ebay.com, boingboing.net or ad.nl. These sites are known to promote certain topics that should be based on facts, but convey it with biased information on Vincent van Gogh that is ambiguous and challenging to find the sources of.

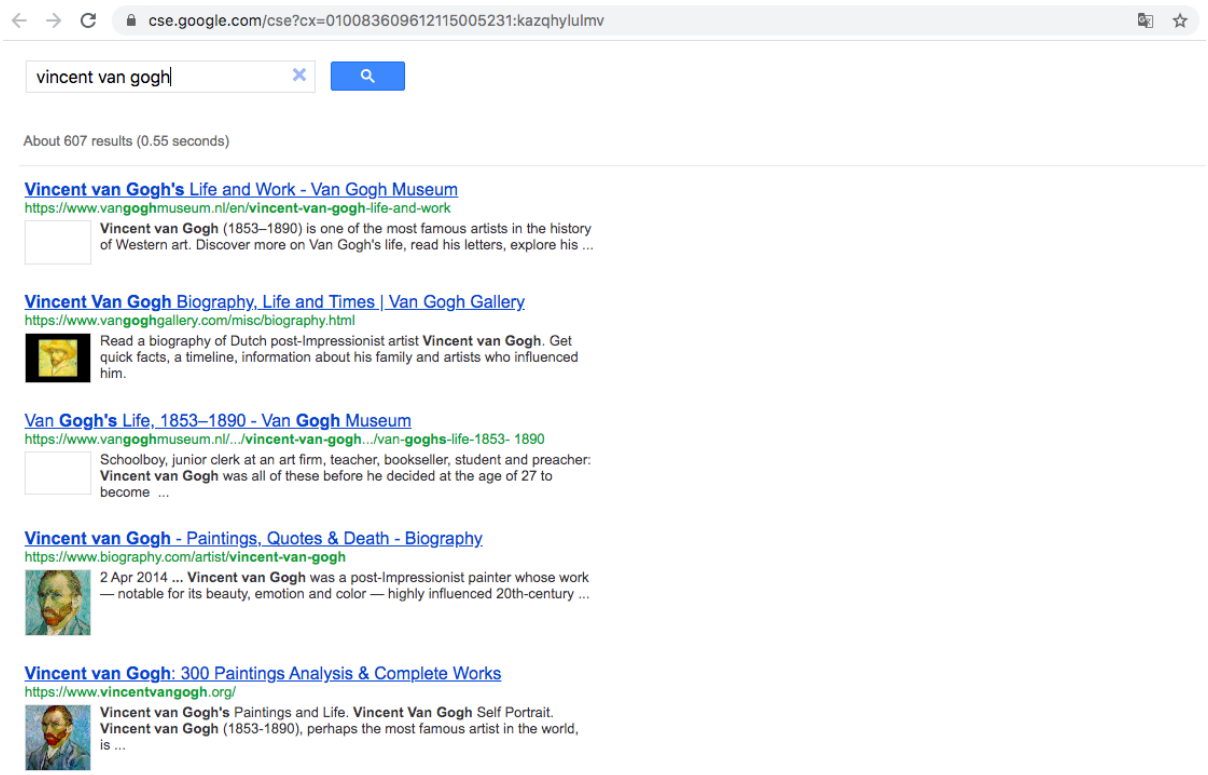
Looking at duckduckgo.com, Vincent van Gogh was still represented with different sources on how to pronounce his name and selling forgery work of him. Therefore, I assessed each and every source based on their background and information that was displayed on the interface. If the stories aligned to one another, they were considered. I found out that sources from big news outlets, such as the Guardian, proved not to be reliable, based on the biased information they presented on Vincent van Gogh. Therefore, the issue was to assess each source individually and trace where the information was based on. The search engine increased in size and eventually consisted out of 20 sources. The following image displays the emergence of the search engine; Gogh and Search and what sites were used to generate the search engine.



The Google Co-op program also allowed for a code generation. The code allowed for alterations to the search engine, such as the rendering of search results and of the search box. This information can be retrieved from the following image.



The final product of Google Co-op is displayed in the image below. Gogh and Search can be used to research mainly Vincent van Gogh and his art work.



Evaluation of the Search Engine

To evaluate the search engine, I have browsed to a web page that was part of my initial collection. The following table shows the product of the keywords that might be used trying to find this page. After close examination, the following keywords were used: painting, painter, van Gogh, Dutch art, death and legacy to find the following page https://www.biography.com/artist/vincent-van-gogh. The method to assess both search engines can be characterized as 'know item search', as to examine the effectiveness of the search engines ("Fo2d1920-Assignment01.Pdf: From Objects to Data" n.d.).

Query	Proposed Results	Correct response	Rank (page)	Reciprocal rank
painting	6910	Paintings, painting	1st	1
legacy	10	legacy	4th	0.25
death	961	Death, died	2nd	0.5
painter	3260	Painter, painters	2nd	0.5
Van Gogh	8550	Van Gogh, Van Gogh's	3rd	0.333
Dutch art	1660	Art, Dutch	1st	1

Moreover, the same has been done for Google.com, this is depicted in the following table.

Query	Proposed Results	Correct response	Rank	Reciprocal rank
painting	32100000	Painter, painting, painter	13	0,076
legacy	3360000	legacy	14	0,071
death	7500000	Died, death	8	0,125
painter	14000000	painter	6	0,167
Van Gogh	103000000	-	-	0
Dutch art	12300000	-	-	0

Furthermore, to research the informational search of Gogh and Search, 5 general topics were used to find information. The keywords were inserted in the vertical search engine and the top ten results were examined based on relevant information for the topic. In this case, this will also be done for google.com. The precision and recall formula is used to define the precision of the search engine. Furthermore, the formula of the 'informational search' method consists out of the following ("Fo2d1920-Assignment01.Pdf: From Objects to Data" n.d.):

Number of relevant sources / the amount of results that are generated in the first page (usually 10)

The general topics that are considered randomly selected: school, labor, science, animals, history, international, rights, emotions, business, censorship. The following table is a product of what Gogh and Search generated.

General Topic	Total sources	Useful sources	Score Result
School	10 (1410)	8	0,8
Labor	5	1	0,2
Science	4	2	0,5
Animals	6	2	0,3333
History	10 (8800)	7	0,7
International	5	1	0,2
Rights	10 (800)	3	0,3

Emotions	10 (196)	4	0,4
Business	9	1	0,111
Censorship	-	-	0

When observing the general topics, it became apparent that the most useful sources were from encyclopedia and facts distributed by Van Gogh Museum or other academically approved sites. The average score of the whole table is 0,35441. In some cases, there were not even 10 results generated, since the topic did not correspond to Vincent van Gogh.

When searching these general topics in Google.com, the following results emerged.

General Topic	**Total sources**	**Useful sources**	**Score Result**
School	10 (14060000000)	6	0,6
Labor	10 (2270000000)	4	0,4
Science	10 (6760000000)	6	0,6
Animals	10 (2420000000)	5	0,5
History	10 (11650000000)	4	0,4
International	10 (12330000000)	6	0,6
Rights	10 (17940000000)	6	0,6
Emotions	10 (400000000)	5	0,5
Business	10 (19480000000)	7	0,7
Censorship	10 (40000000)	8	0,8

The sources that were generated were mainly general information sites, such as Wikipedia. For general topics, there was an overabundance of results that exceeded the 10 sites of the first page. Furthermore, the average sources that would be considered to be relevant was 0,57.

Reflection

The limitations of using a vertical search engine is that it does not take general topics into account. Therefore, to find a broad range of results, Google.com will always prevail with generating more sources. However, when considering if the content on the page is relevant and useful for writing a paper, a vertical search engine is more relevant and gives specific sources for a certain topic. In hindsight, the strict principals that were used to select certain sources for the search engine, limited the search results of the vertical search engine and effected the performance of the 'informational search' method.

When considering the strengths and weaknesses of keyword search, the sources individually also should be considered. Where Google.com provides over a million results, not every source can be used. In the case of a vertical search engine, most sources are relevant, as proven by the 'know item search.' In this case the vertical search engine allowed for more relevant sources that might be used and Google.com lacked in relevant sources while providing thousands of results. Yet, to find general topics, the vertical search engine is not as suitable as Google.com. For instance, 'censorship' was not able to be found in Gogh and Search. This affected the end average and the deficit it has compared to Google.com.

It is difficult to evaluate how the outcome would be if computational language understanding was possible. In this case, artificial intelligence should be trained and be in par with Google's search engine. Since it is obvious that a search engine that has been trained over many years by millions of people that use it will perform better than a search engine that is only used by one person over a time period of two weeks.

In comparison, the average score for both search engines for the 'informational search' holds a big difference. Furthermore, the results of the informational search of Gogh and Search was more specific and contained content that was peer reviewed or academically acclaimed. After adding ten extra sources to Gogh and Search, the average of this score became an average of 0,47. Therefore, the averages do not take all aspects of the search engines into account and also do not portray a definitive conclusion.

Besides that, it was interesting to find out what can be altered in Google Co-op. For instance, the program made it possible to change the smallest details of the search engine, to make it seem original. For Gogh and Search, a self-portrait is used as the logo of Vincent

van Gogh. Furthermore, the image search is also enabled, since the paintings of Vincent van Gogh are the main reasons for his legacy. This made the search engine also seem like a personal gallery of paintings of Vincent van Gogh.

In conclusion, it is difficult to determine which search engine performs better. Google.com can be used for general topics, but for in depth relevant and academical sources a vertical search engine might allow for a fruitful result. Therefore, to assess the history and the art work of Vincent van Gogh, both search engines prove to be useful. However, it depends on the user and what they consider to be convenient for background research. It might also be interesting to take both search engines into consideration to allow for a perception that is derived from a wide range of sources instead of only taking academical sources into account.

References

"Fo2d1920-Assignment01.Pdf: From Objects to Data." n.d. Accessed March 28, 2020. https://canvas.uva.nl/courses/15702/files/2368029?module_item_id=513161.

"Custom Search - Create CSE." n.d. Accessed April 4, 2020. https://cse.google.com/cse/create/new.

Appendix

Code:

```
<script async
src="https://cse.google.com/cse.js?cx=010083609612115005231:kazqhylulmv"></script>
<div class="gcse-search"></div>
```

Custom search engine:
Gogh and Search - https://cse.google.com/cse?cx=010083609612115005231:kazqhylulmv

Results used for Gogh and Search:
https://www.vangoghgallery.com/misc/biography.html
https://www.biography.com/artist/vincent-van-gogh
https://www.vincentvangogh.org/
https://vangoghmuseum.nl/en/vincent-van-gogh-life-and-work
https://www.sothebys.com/en/artists/vincent-van-gogh
https://www.notablebiographies.com/Tu-We/van-Gogh-Vincent.html
https://www.dailyartmagazine.com/van-gogh-lovers/
https://www.rijksmuseum.nl/en/rijksstudio/artists/vincent-van-gogh
http://www.getty.edu/art/collection/artists/377/vincent-van-gogh-dutch-1853-1890/
https://www.encyclopedia.com/people/literature-and-arts/european-art-1600-present-biographies/vincent-van-gogh
https://www.nga.gov/education/teachers/lessons-activities/self-portraits/van-gogh.html
https://www.theartstory.org/artist/van-gogh-vincent/
https://www.artic.edu/artists/40610/vincent-van-gogh
https://krollermuller.nl/en/van-gogh-gallery
https://archive.artic.edu/van-gogh-bedrooms/about-paintings
https://www.moma.org/artists/2206
https://artuk.org/discover/artists/van-gogh-vincent-18531890
https://archive.org/details/Van.gogh.paintings
http://www.visual-arts-cork.com/famous-artists/van-gogh.htm
http://www.vggallery.com/painting/by_period/auvers.htm