

Censorship of Search Engines

Simple Case Study: Exclusion of an Offensive Game

N.L.M. de Jonge; B.Eng. (mail@norbertdejonge.nl)
November 27, 2014

Summary

All major, English-language search engines are aware of the existence of a web page about a simple GNU/Linux game called *Pedophile's Quest*. Yet, with only one exception, they do not include it in their search results. While the game is offensive to some people, all of its elements are legal. This paper documents a possible instance of Internet censorship, and is a first step towards finding out who imposed the censorship.

Table of Contents

Introduction.....	1
Facts and Findings.....	3
Analysis and Interpretation.....	5
Annex A.....	7
Annex B.....	10
Annex C.....	12
Lycos Error.....	12
SaveSearch.....	12
Contents of .htaccess.....	13

Introduction

In July 2007, I published a simple GNU/Linux game called *Pedophile's Quest*. It was first published at <http://www.clogo.org/PedoQuest/> and later at http://www.norbertdejonge.nl/Pedophiles_Quest/. The game's provocative nature and its satirical style are offensive to some - possibly, most - people; in particular some of its texts. However, all of the game's elements are legal in my country, the Netherlands, and in America, and probably also in the rest of the Western world. The game did *not* cause great commotion. It was mentioned by *la Repubblica*¹ and Dutch columnist Bert Brussen.²

For more than two years now, since 25 June 2012, the game has been available on my personal website, at http://www.norbertdejonge.nl/Pedophiles_Quest/. Ever since that day, its web page has included just *one* level 1 heading element, the highest-level heading, namely `<h1>Pedophile's Quest</h1>`, and *one* title element, namely `<title>Pedophile's Quest</title>`. Ever since aforementioned date, it's being linked to from <http://www.norbertdejonge.nl/software/> with `P. Q.`, that in turn is being linked to from <http://www.norbertdejonge.nl/> with `Software`.

Various other software packages have their own web pages on my personal website, including *apoplexy*, at <http://www.norbertdejonge.nl/apoplexy/>, and *Wappo Level Creator*, at http://www.norbertdejonge.nl/Wappo_Level_Creator/. All are at the same directory level, use *one* level 1 heading element, *one* title element, and are linked to in the same fashion as *Pedophile's Quest*. The only difference for the links to their web pages is that the link *text* itself is not abbreviated.

I have noticed for years that major search engines, such as Google Search and Bing, do not mention, nor include links to, the web page about *Pedophile's Quest* on my personal website when searching for either `Pedophile's Quest` or `"Pedophile's Quest"`. I have no reason to believe these search engines are unaware of the existence of aforementioned web page. In fact, my experience

1 <http://www.repubblica.it/2007/06/sezioni/cronaca/pedofilia-tre/videogioco-pedofilo/videogioco-pedofilo.html>

2 <http://www.bbrussen.nl/2007/07/17/norbert-de-jonge-lanceert-pedogame/>

as a web developer who's aware of SEO fundamentals, my Informatics education, and my experience with system administration, cause me to believe my web page about *Pedophile's Quest* is deliberately, and against my wishes, being excluded from the search engines' results.

The purpose of this document is twofold:

1. to document a possible instance of (global) Internet censorship, and
2. to take a first step towards finding out who imposed/authorized the censorship.

While the exclusion of this simple game is insignificant by itself, it may point to, and be proof for, censorship of search results on a larger scale. Do search engines know the web page about *Pedophile's Quest* exists, and, if they do, are they deliberately excluding it from their search results?

I have no reason to believe that my content provider, the host of my personal website, is using an IP list to prevent search engine bots (web crawlers) from reaching the web page in question.

The rest of this document consists of two additional chapters and three annexes. I will first give you all facts and findings that I believe to be relevant for this case study. This will be followed by an analysis and interpretation of said data. This document is not an academic essay, it includes no methodology chapter, does not list any references; it's merely an overview of facts and findings, followed by my opinion.

Facts and Findings

What I've looked into first is: do search engines know about the web page <http://www.norbertdejonge.nl/software/> that links to the web pages about *Pedophile's Quest*, *apoplexy* and *Wappo Level Creator*, and do search engines' results list http://www.norbertdejonge.nl/Pedophiles_Quest/, <http://www.norbertdejonge.nl/apoplexy/> and http://www.norbertdejonge.nl/Wappo_Level_Creator/? There is a reason I picked those three software packages out of the twenty-three available packages. *Pedophile's Quest* is an obvious choice. I picked *apoplexy* and *Wappo Level Creator* because their web pages are respectively linked to from *several dozen* external web pages (other websites), and *no* external web pages.

As you can see in **Annex A**, practically all (non-Chinese) search engines know about the software overview, about *apoplexy*, and about *Wappo Level Creator*. Note that searching for "Wappo Level Creator" shows there are no known web pages on the Internet that include that particular phrase other than the software overview and the web page about *Wappo Level Creator* on my personal website.

As you can see in **Annex B**, practically none of the search engines list *Pedophile's Quest*. The only clear exception is Exalead, which not only lists it when searching for "Pedophile's Quest", but even when searching for *Pedophile's Quest* (without quotes) as the second entry of thousands of results. It also includes a cached version of August 26, 2014. Practically all other search engines claim they do not know about http://www.norbertdejonge.nl/Pedophiles_Quest/, do not know there is a (phrase) "Pedophile's Quest" on my personal website - *even* when directing the search engines to only search my personal website (with `site:www.norbertdejonge.nl`).

In addition I searched Internet Archive's Wayback Machine for http://www.norbertdejonge.nl/Wappo_Level_Creator/ and got no results. One of the Lycos searches kept returning an error. The error is partially listed in **Annex C**.

I performed all searches with Firefox 33.0, under GNU/Linux. I visited the first eight search engines in the tables on November 24, 2014; the rest on November 25, 2014. Before searching, I cleared all web browser history (Browsing & Download History, Form & Search History, Cookies, Cache, Active Logins, Offline Website Data, Site Preferences). Before searching, I also disabled any

SafeSearch options (family filters) that I could find. See **Annex C** for a listing. I've ignored advertisements in the counts - in the number of results. None of the advertisements were about my website.

Of all relevant directories...

<http://www.norbertdejonge.nl/>

<http://www.norbertdejonge.nl/software/>

<http://www.norbertdejonge.nl/apoplexy/>

http://www.norbertdejonge.nl/Wappo_Level_Creator/

http://www.norbertdejonge.nl/Pedophiles_Quest/

...only <http://www.norbertdejonge.nl/> contains a `.htaccess` file. Its contents is listed in **Annex C**. None of the `index.php` files in aforementioned directories include `NOINDEX` or `NOFOLLOW` or similar statements. None of the directories contain a `robots.txt` file.

Analysis and Interpretation

Annex A shows that there are no known web pages on the Internet that include the phrase "Wappo Level Creator" other than on my personal website. Practically all search engines know about the software overview. Of the twenty search engines tested,³ the vast majority can find *Wappo Level Creator* when searching for "Wappo Level Creator":

- 12 give it as the *first entry* of two results, where the other result is the software overview,
- 2 give it as the *first and only* result, and
- 6 cannot find it or give no results or do not list it on the first web page. Of those 6, a whopping 4 are Chinese search engines (Baidu, Sogou, Soso and Youdao).

The reason search engines know about *Wappo Level Creator*, even though no external web pages (other websites) link to its web page, is because there is a link from the software overview on my personal website to the web page about the package. Search engine bots (web crawlers) are constantly looking for links to web pages unknown to them, and if nothing prevents these software applications from parsing such web pages, they *will* do so.

http://www.norbertdejonge.nl/Pedophiles_Quest/ has been on-line for more than two years, and has been linked to from <http://www.norbertdejonge.nl/software/> for more than two years. The web page about *Pedophile's Quest* uses proper heading and title elements that include the package's name, the web page also mentions the package's name in regular text, I do not use NOINDEX or NOFOLLOW or similar statements, I do not use a robots.txt file, and the (contents of the) .htaccess file in the root directory has no impact on this case.

The Internet Archive's Wayback Machine has one capture of the software overview, and it includes the P. Q. link to http://www.norbertdejonge.nl/Pedophiles_Quest/. This (January 1, 2014) capture is available here: <https://web.archive.org/web/20140101035035/http://www.norbertdejonge.nl/software/>

Furthermore, Exalead does not only know about the web page about *Pedophile's Quest*, it returns it as the first result when searching for "Pedophile's Quest" and the second result when searching for Pedophile's Quest (without quotes), and it has

³ Internet Archive's Wayback Machine is generally seen as a digital archive and not a (true) search engine.

a cached version of the web page.

I'm aware that removing all web browser history is not enough to escape all filter bubbles,⁴ but even personalized searches (with *disabled* family filters, see **Annex C**) - if applicable - should not suddenly make web pages vanish into thin air. However, as **Annex B** shows, searching for "Pedophile's Quest" (with quotes) results in:

- 16 not found or no results responses,
- 1 successful hit (found), and
- 3 do not list it on the first web page.^(†)

^(†) This does not mean it is on any of the other web pages.

And searching my personal website (with `site:www.norbertdejonge.nl`) for Pedophile's Quest (no quotes) results in:

- 19 not found or no results responses, and
- 1 successful hit (found).

None of the search engines gave any indication that any of the queries resulted in any kind of censorship or unusual filtering; that one or more results were being excluded. There is *absolutely no way* that, for example, Google Search and Bing are not aware of the web page about *Pedophile's Quest*. It has been on-line for more than two years. AWStats, a web analytics reporting tool, shows that dozens of search engine bots visit my personal website. Some almost daily, including Googlebot. They also did this around the time *Pedophile's Quest* was added to my personal website. June 2012: 20 different robots, 145 hits from Googlebot. July 2012: 23 different robots, 153 hits from Googlebot.

What other conclusion can I draw than that someone imposed/authorized the censorship of this game and its web page?

⁴ See http://en.wikipedia.org/wiki/Filter_bubble for more information.

Annex A

Search engine	www.norbertdejonge.nl/apoplexy/	www.norbertdejonge.nl/software/	Wappo Level Creator	"Wappo Level Creator"	site:www.norbertdejonge.nl Wappo Level Creator
http://www.baidu.com/	1st and only entry	1st and only entry	not on first page ^(†) (84,700 results)	no results	no results
https://www.bing.com/	1st entry (29 results)	1st entry (707 results)	1st entry (5,460 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
https://blekko.com/	1st entry (10 results)	1st entry (10 results)	1st entry (10 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
https://duckduckgo.com/	1st entry (13 results)	1st entry (many results)	1st entry (many results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
http://www.exalead.com/search/	1st entry (4 results)	1st entry (3 results)	1st result (54 results)	1st and only entry	1st entry (2 results, other is my software/ overview)
http://www.gigablast.com/	no results	not found (1 result, link to a non-related .tar.gz file on my website)	no results	no results	no results
https://www.google.com/ncr	1st entry (199 results)	1st entry (1,200 results)	1st entry (5,150 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
https://www.ixquick.com/	1st entry (6 results)	1st entry (2 results, other is a	1st entry (11,999,890 results)	1st entry (2 results, other is my	1st entry (2 results, other is my

		link to a non-related .tar.gz file on my website)		software/ overview)	software/ overview)
https://www.qwant.com/	1st entry (various pages)	1st entry (various pages)	1st entry (various pages)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
http://www.sogou.com/	unclear	unclear	not on first page ^(†) (107,515 results)	not on first page ^(†) (116,765 results)	no results
http://soso.com/	unclear	unclear	not on first page ^(†) (11,137 results)	not on first page ^(†) (90,736 results)	no results
https://search.yahoo.com/	1st entry (36,500 results)	1st entry (369,000 results)	1st entry (22,300,000 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
https://www.yandex.com/	1st entry (6 results)	1st and only entry	1st entry (2,000 results)	1st and only entry	1st entry (3 results, others also on my domain)
http://www.youdao.com/	no results	not on first page ^(†) (43 万 5000 results)	not found (20/30 results)	not found (20/30 results)	no results
http://nl.search.yahoo.com/	1st entry (863 results)	1st entry (12,600 results)	1st entry (1,620,000 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
http://www.google.nl/	1st entry (197 results)	1st entry (1,140 results)	1st entry (3,830 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
http://www.ilse.nl/	1st entry (6 results)	1st entry (26 results)	not on first page ^(†) (319 results)	not found (1 result, my software/ overview)	not found (1 result, my software/ overview)
http://www.lycos.nl/	1st entry	1st entry	1st entry	1st entry	1st entry

	(35,300 results)	(368,000 results)	(22,500,000 results)	(2 results, other is my software/ overview)	(2 results, other is my software/ overview)
http://www.vinden.nl/	1st entry (120 results)	1st entry (12,600 results)	1st entry (1,620,000 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
http://www.metaseek.nl/	1st entry (2 results)	1st entry (3 results)	2nd entry (19 results)	1st entry (2 results, other is my software/ overview)	1st entry (2 results, other is my software/ overview)
https://archive.org/	9 results (July 16, 2011 - June 20, 2014)	1 result (January 1, 2014)	not applicable	not applicable	not applicable
Search engine	www.norbertdejonge.nl/apoplexy/	www.norbertdejonge.nl/software/	Wappo Level Creator	"Wappo Level Creator"	site:www.norbertdejonge.nl Wappo Level Creator

^(†) This does not mean it is on any of the other web pages.

Annex B

Search engine	www.norbertdejonge.nl/Pedophiles_Quest/	Pedophiles Quest	Pedophile's Quest	"Pedophile's Quest"	site:www.norbertdejonge.nl Pedophile's Quest
http://www.baidu.com/	no results	not on first page ^(†) (1,610 results)	not on first page ^(†) (2,830 results)	no results	no results
https://www.bing.com/	not found (7 results)	not on first page ^(†) (32,900 results)	not on first page ^(†) (33,000 results)	not found (18 pages)	no results
https://blekko.com/	not found (7 results)	not found (10 results)	not found (10 results)	not found (3 results)	no results
https://duckduckgo.com/	not found (18 results)	not on first 'page' ^(†) (many results)	not on first 'page' ^(†) (many results)	not found (13 results)	not found (1 result, a PDF on my website)
http://www.exalead.com/search/	1st and only entry (including a cache of Aug. 26, 2014)	not on first page ^(†) (10,893 results)	2nd entry (16,976 results)	1st entry (4 results)	1st entry (2 results, other is a PDF on my website)
http://www.gigablast.com/	no results	not on first page ^(†) (248 results)	not found (2 results)	not found (1 result)	no results
https://www.google.com/ncr	no results	not on first page ^(†) (133,000 results)	not on first page ^(†) (136,000 results)	not found (8 results)	no results
https://www.ixquick.com/	no results	not on first page ^(†) (169,599 results)	not on first page ^(†) (338,802 results)	not found (23 results)	not found (1 result, a PDF on my website)
https://www.qwant.com/	not on first page ^(†) (various pages)	not on first page ^(†) (various pages)	not on first page ^(†) (various pages)	not on first page ^(†) (various pages)	no results
http://www.sogou.com/	unclear	not on first page ^(†)	not on first page ^(†)	not on first page ^(†)	no results

		(931 results)	(205,720 results)	(239,772 results)	
http://soso.com/	unclear	not on first page ^(†) (952 results)	not on first page ^(†) (66,291 results)	not on first page ^(†) (230,371 results)	no results
https://search.yahoo.com/	not on first page ^(†) (6,780 results)	not on first page ^(†) (307,000 results)	not on first page ^(†) (307,000 results)	not found (18 results)	no results
https://www.yandex.com/	no results	not on first page ^(†) (169,000 results)	not on first page ^(†) (169,000 results)	not found (4 results)	not found (1 result, a PDF on my website)
http://www.youdao.com/	no results	not found (1,540 results)	not found (3,330 results)	not found (1 result)	no results
http://nl.search.yahoo.com/	not on first page ^(†) (19,900 results)	not on first page ^(†) (32,900 results)	not on first page ^(†) (32,900 results)	not found (1,990/180 results)	no results
http://www.google.nl/	not found (1 result)	not on first page ^(†) (135,000 results)	not on first page ^(†) (3,130,000 results)	not found (8 results)	no results
http://www.ilse.nl/	no results	not on first page ^(†) (14,100 results)	not on first page ^(†) (14,300 results)	not found (2 results)	no results
http://www.lycos.nl/	not on first page ^(†) (6,780 results)	not on first page ^(†) (308,000 results)	error	not found (20 results)	no results
http://www.vinden.nl/	not on first page ^(†) (19,900 results)	not on first page ^(†) (32,900 results)	not on first page ^(†) (32,900 results)	not found (175/84 results)	no results
http://www.metaseek.nl/	no results	not found (42 results)	not found (42 results)	not found (21 results)	no results
https://archive.org/	no results	not applicable	not applicable	not applicable	not applicable
Search engine	www.norbertdejonge.nl/Pedophiles_Quest/	Pedophiles Quest	Pedophile's Quest	"Pedophile's Quest"	site: www.norbertdejonge.nl Pedophile's Quest

Annex C

Lycos Error

Zend_Locale_Exception

Missing locale file '/web/sites/common/inc/Zend/Locale/Data/en_us.xml' for 'en_us' locale.

/web/sites/common/inc/Zend/Locale/Data.php(143)

[...]

2014-11-25 15:32:30 Apache Yii Framework/1.1.14

SaveSearch

- Bing: Preferences -> SafeSearch: Off -> Save -> Agree
- DuckDuckGo: <https://duckduckgo.com/settings> -> Safe Search: Off -> Save and Exit
- Exalead: Preferences -> Safe search: Disabled -> Save my preferences
- Google Search: Already disabled. I used the gear icon to make sure it was.
- Yandex: After any search, an icon on the top right can be used to change settings. Search mode: Unsecure Also available via <https://www.yandex.com/search/customize>
- *Dutch* Yahoo!: <https://nl.search.yahoo.com/preferences/preferences> -> SafeSearch: Uit: resultaten niet filteren -> Opslaan -> Ik ga akkoord
- *Dutch* Google Search: Already disabled. Instellingen -> Zoekinstellingen -> [] Expliciete resultaten filteren.
- Vinden.nl: Already disabled. Voorkeuren instellen -> Familiefilter: Uit -> Voorkeuren opslaan

In addition, I changed Metaseek.nl from Nederland (the Netherlands) to Internationaal (International).

Contents of .htaccess

```
AddType video/webm .webm
```

```
# DO NOT USE THIS ON LOCALHOST
```

```
RewriteEngine On
```

```
RewriteCond %{HTTP_HOST} !^www\.
```

```
RewriteRule ^(.*)$ http://www.%{HTTP_HOST}/$1 [R=301,L]
```