

School of Engineering

Department of Electronic & Computer Engineering

# Test the testers (come up with a fancy name)

Author: Hlib Fedchuk

Supervisor: Stephen Farrell

A thesis submitted in partial fulfilment of the degree of BAI in

Computer & Electronic Engineering

May 2018

***Table of Contents (needs fixing):***

Abstract……………………………………………………………………………………………………………………………………3

Background………………………………………………………………………………….…………………………………….xx  
 -xcrap web sites  
-getting less crap

-measuring via n/w visible indicators  
-assumption is that if you tell people "here's HOWTO get an A"  
-research question: what's the value of the different measures/scores?  
ANS:if they had commonality then site owners might  
                  have less work to do to test their shit and hence  
                  more time to improve their shit

Methods………………………………………………………………………………………………………………………………….xx  
-x internet measurement, now practical wasn't before  
        - measuring the measurers  
        - commensurate/incommensurate scoring, 105/100 good example   
          is A == A ? etc. github.com mention that sites start with initial 100 score and get +/- points for features  
        - not all measures make sense (e.g. CSP depends on  
          content being served)  
        -x loadsa text about specific tests done (heartbleed, hsts,...)  
        - xinitial list of possible testers  
                =x reduce list based on scriptability  
                =x try include most commonly used (sslabs, moz,..)  
        - xreproducible: what hlib does needs to be re-doable by all

Design/implementation………………………………………………………………………………………………xx

-x set of test sites  
        - set of testers  
        - xscripting/screen-scraping  
                = try find CLI/json-output  
                = worst case: hack screen-scraping  
                = even worst case: headless browser & har files  
                        - ok to say thought about but didn't need  
                          (if true)

Results…………………………………………………………………………………………………………………………..xx

Conclusion and future work……………………………………………………………………………………………xx

- it's great but could be better, here's how  
        - ask testers to do X,Y,Z to improve commonality

Acknowledgements……………………………………………………………………………………………………….xx

References…………………………………………………………………………………………………………………….xx

Appendices…………………………………………………………………………………………………………………….xx

***Abstract:***

This project is about comparing and contrasting different web site testers. During the period of this project’s progress, a list of tester and test sites will be picked in order to come up with results. The aim of this project is to differentiate between the tester’s results output in order to see and pick out the key differences between them. A set of web testers will be broken into three different categories: DNS testers, TLS (mail) testers and overall website testers. This separation was introduced in order to have a set of results that would relate to each other and get rid of confusion, where it would be very impractical to compare a TLS tester with a DNS tester, where there may be some similarities between them, but we want to obtain results of the same lookup area. During this project, the results would depend on the:

1. Overall score:   
   Typically, a tester would provide the overall score to the user that would indicate how well or bad their website/mail server is implemented in terms of security aspects. This result value cannot be discarded and should be taken very seriously, as it is the final score that the tester provides on how well the security features are implemented in this website, whether they are coming from a DNS/TLS or overall web tester.
2. Security and privacy issues:  
   A majority of the web testers would provide a detailed description on how they came up with the end score (privacyscore doesn’t). This description will provide the information on the points (that this specific tester thinks are important), whether they are implemented or not. For example, 2 different test may consider the same feature differently, some might penalize for it, giving it a negative mark, and some may just leave it and give a 0 mark for it.
3. Scope of implemented features:   
   This aspect will consider the current amount of implemented features that the web site has.   
   We can consider a small web site and a much bigger one, but the smaller one can get more points depending on the implementation, if more measures are implemented – it wouldn’t be a surprise that the score will be higher.

***Background:***

***Bad Websites:***

There was a need to do some research and find some bad websites that aren’t good regarding the results that the testers would provide. This had to be done in order to see the vast difference that the tester would provide between the results of a tester. It would be rather interesting to see what results will be present in a good test site compared to what will be shown for a bad site.

A way to choose if a web site is badly implemented was to:

1. Surf the internet with a query for “bad sites” or something similar. There are many blog pages that would provide us with a list where these sites are discussed and some pictures are also provided. Mostly these sites are ranked by aesthetics and the way they look, but there are some outliers that actually talk about a website not being secure and these guys are the ones to definitely look into.
2. When the research was done and a list of “bad” websites was created, the next step would be to actually check if this website is bad security wise or it’s just bad aesthetically. To check it, there is a simple way – creating a batch file with that list and running it through Privacyscore tester. It’s rather easy and simple, because there will only be around 20 lines of a script and this script will be copy-paste, but the name of the test sites will be changed.
3. In order to pick a bad web site, a certain threshold had to be created to be able to call this web site bad, when the result comes back. For myself, I didn’t think that long and decided to go for a score that is less than 40 (taking 40 as a pass mark). This would mean that all the test sites that undergo Privacyscore ‘ s test and fail it are considered bad and will be kept in order to have the comparison that would help to understand how the testers work and what aspects they consider important.

***Methods:***

***Internet measurement:***

Internet has become a huge aspect of human lives for the past ten years. It includes different branches like: international commerce, communication and technological development. It is important that various groups that affect and use internet, like researches, entrepreneurs, service providers and other members of internet community understand the growth characteristics and its limitations.

At early stages internet measurement was neglected giving it a low priority to increasing network’s speed, capacity and coverage. The increase in interest towards the internet only come pretty recently and it is a logical outcome as internet has expanded, which led to an increase in the number of users and malicious behavior. In order to counter these malicious elements, there needs to be a huge investment into Internet measurement and data analytics. Merit Network operated the NSFnet backbone in its various forms, measured the backbone’s traffic volumes and produced summary statistics through April 1995. But these were primarily oriented toward short-term operational requirements or periodic simplistic traffic reports for funding agencies. As such, they weren’t conducive to workload or performance characterization, much less network-dynamics modeling. As the NSFnet and attached regional infrastructures exploded in popularity among academic and commercial sectors, operators acutely focused on increasing link speeds and router/switch-traffic capacities. Developers worked on improving protocols and inventing new ones to support emerging services. The evolutionary context of the infrastructure left little room for more than mild interest in network measurement. For ordinary people, the Internet has become an integral part of everyday life; it is now used continually to find information, buy products, meet people, do our jobs, and play. As if these circumstances weren’t sufficiently revolutionary, the pervasive adoption of mobile computing expectations and requirements is now prompting service providers to take a strong interest in more strategic measurement and charging schemes.

With its ever-growing user community, the Internet has gradually been forced over the past decade to deal with the “real world.” Like chemical pollutants from industrial production processes, infrastructural pollution — such as viruses, worms, and spam traffic — has become significant in volume and impact on user productivity. Protective technologies such as firewalls and NAT gateways have changed the Internet’s simple end to-end connectivity model. Although these devices can effectively block some malignant packets, they do so by filtering packets according to access control lists (ACLs), which can prevent many applications from working properly.

The Challenges Collection, interpretation, and modeling of empirical Internet data remains challenging. The technologies and protocols involved in generating and delivering Internet traffic were designed for technical expediency, architectural clarity, and functionality, rather than for measurement and analysis. New developments often introduce specifications that are independent of their predecessors; technology developers often deploy them as rapidly as possible, without concerted systematic testing on the vast set of heterogeneous components encountered on the Internet. Indeed, it would be impossible to test certain behaviors against all possible combinations of equipment, software, and configuration. Furthermore, many who develop technologies and protocols contend that the Internet has evolved splendidly thus far without extensive measurement and modeling. Others believe that we should not begin measurement and modeling efforts until doing so proves cheaper than simply expanding the currently available bandwidth. To make matters harder, a variety of legal and privacy issues serve as active disincentives to measurement research and development activity. Nonetheless, every constituency of the Internet (providers, vendors, policymakers, and users) realizes that we need a better understanding of Internet structure and behavior, including the influence of various components and functionalities on macroscopic dynamics. Floyd and Paxson’s landmark paper provided several insights into why the Internet is hard to measure, and thus hard to simulate, making it resistant to modeling and predictive insight.2 the first big challenge is that everything keeps changing. For example, HTTP traffic grew from zero in 1995 to more than 80 percent of the network traffic at many sites by the early 2000s. Yet, HTTP’s proportion of total traffic is now dropping on most links, and peer-to-peer traffic is steadily rising as developers find more ways to use P2P technology. The Internet’s global scale also complicates measurement efforts, as does the fact that many aspects of traffic and behavior change from location to location. Thus, statistics gathered at one location often prove unrepresentative of the global Internet. Instead, we need to make measurements at many sites and correlate the results to derive a comprehensive view. Finally, few Internet protocols and applications were designed to inherently support fine-grained measurement. Instead, researchers have had to find indirect ways to measure network phenomena. For example, traffic-flow measurements rely on data collected from packet headers as they pass across links; counting packets and bytes and classifying them into flows on the basis of values taken from the headers is easy but yields limited insight into higher-layer behavior. Measuring application performance generally remains a challenge, since applications differ as to how they transport application-specific data. For example, while effective tools exist for measuring Web server performance, such tools are often not effective at measuring performance of other applications.

***Reproducibility:***

This section is important, as the following text will be about recreating my work up to date in order to generate the results.

Installation instructions for SSLLabs, Mozilla-observatory and privacyscore.

### Windows:

**SSLLabs:**

Before installing SSLLabs, Go has to be of version, which is at least 1.3 or higher should be installed.

<https://golang.org/doc/install?download=go1.10.windows-amd64.msi> -> use this link to download go.

To ensure that Go was installed properly: go build command should be typed when in C://go/src directory from cmd.

When that has worked out, next step would be to clone the repo of SSLLabs by using gitbash with a simple command: git clone <https://github.com/ssllabs/ssllabs-scan.git>

Next step is to change directory to C://users/go/src/ssl-labs and use the command to build it again: “go build”

When that was done, we go to cmd from the directory above, we use a command: SSLLabs-scan --json-flat jell.ie (or other test site)

There is at least 3-4 delay for me until a json would appear.

On the next run of the same tester, there result output is almost instant.

A command for SSLLabs-scan --json-flat --host file (that would allow to test multiple test sites from a file where they are separated by a line)

**Mozilla Observatory:**

Before installing Mozilla-observatory, Npm or node.js that has Npm in it has to be installed by following: <https://www.npmjs.com/get-npm>

Next step: in command line use the command Npm install -g observatory-cli. Before doing the installation, a directory should be created for example C://user/"..."/Mozilla-obs.

We clone the repository by using git clone <https://github.com/mozilla/observatory-cli.git> into the above directory.

The command "npm install -g observatory-cli" is used to install Mozilla-obs.

When that command went through, observatory jell.ie --zero --format=report command is used to generate a report that includes 0 scores also. The output here is json format.

A command observatory some.site.name --rescan is used to force a rescan. This could be used as fetching results from cache.

**Privacyscore:**

Privacyscore also uses go, so the same procedure is followed. The only difference is just creating a new directory for privacyscore in C://users/"..."/go/

Then we clone the repo <https://github.com/asciimoo/privacyscore> and then use the command in this directory: go build.

When build succeeds, privacyscore.exe appears in the directory.

To run it in cmd, a command privacyscore.exe some.site.name is used to run the scan.

### Virtual Machine:

That scan only provides a score for the test site, thus a command created by my project supervisor is used.

This command is used for Ubuntu, thus VM is installed.

This command is: “ cat index.html | sed -e 's/<span/\n<span/g' | sed -e 's/<span .\*">//' | sed -e 's/</span>//g' | sed -e 's/"/"/g' | grep -v "<" | grep -v ">" | head -1652 | tail -1629 | json\_pp “

To get this file, a filename.html was saved when results were generated and this script gets rid of "spans" and "quots" and json\_pp is used to get the json from this.

For some reason this doesn't work form me and i get an error. If json\_pp (pretty print) is not used, i get the required information for the results, but not in json format.

***List of testers:***

The list of testers was created after the research was done. This list is a good help later on when the installation process will occur. This list will then be the end list that would show which tester will actually be used to generate the results from the scripts run from the command line.

***DNS testers:***

| **Tester name** | **JSON export available from tester** |
| --- | --- |
| <http://dnsviz.net/> | No |
| <https://manytools.org/network/query-dns-records-online/go> | No |
| <https://www.ultratools.com/domainHealthReport> | No |
| [https://www.zonemaster.fr](https://www.zonemaster.fr/) | Possible to get results from HTML, because there is iFrame present |
| [https://mxtoolbox.com](https://mxtoolbox.com/) Domain Health --- possibility of getting rid of it just because mxbox is used as a mail tester. | No, has commands to use but only on their web 👎 |
| [http://www.dnsstuff.com](http://www.dnsstuff.com/) | Allows export of a .pdf file |
| [https://dnschecker.org](https://dnschecker.org/) | Yes <https://dnschecker.org/dns-tool.php> |
| [http://dnscheck.pingdom.com](http://dnscheck.pingdom.com/) | No |
| [https://observatory.mozilla.org](https://observatory.mozilla.org/) -> from your list | Yes |
| [https://pingability.com](https://pingability.com/) IP stuff | Yes, Report of results in email though: / |
| [https://mxtoolbox.com](https://mxtoolbox.com/) SMTP | No |
| <http://www.dnsstuff.com/mstc> | Needs code to get it from HTML |
| [https://www.ultratools.com](https://www.ultratools.com/) | Code to get JSON format, not a lot of results there |
| <https://www.wormly.com/test-smtp-server> SMTP | No |
| [http://emailsecuritygrader.com](http://emailsecuritygrader.com/) Web and mail, a pretty good one, stay for sure. | No |

| **Tester name** | **JSON export available for export** |
| --- | --- |
| [https://website.grader.com](https://website.grader.com/) | Looks like it requires a subscription |
| [https://www.dareboost.com](https://www.dareboost.com/) | Yes allows export of CSV |
| [https://www.uptrends.com](https://www.uptrends.com/) | Yes <https://api.uptrends.com/v3/checkpointservers?format=json> |
| [https://testmysite.withgoogle.com](https://testmysite.withgoogle.com/) | Yes F-12 audits |
| [https://privacyscore.org](https://privacyscore.org/) --- a pretty good one for results | <https://github.com/PrivacyScore/PrivacyScore> Data goes on the their disk and could be pulled after |
| [https://www.ssllabs.com](https://www.ssllabs.com/) | Yes |

From this list, we can see that the list is reduced to only those testers that can only be scripted in order to automate the result fetching process.

***Set of test sites:***

Initially, the set of test sites was composed of mainly 4 sites. 2 of these sites were run by my supervisor and the other 2 are trinity sites, tcd.ie and scss.tcd.ie. A simple text file was created and pushed to GitHub in order to keep track, either adding or getting rid of some sites, which didn’t happen throughout the project. The results could be written in a table at the end, which can be seen below.

| **Supervisor Sites** | **TCD sites** | **My site** | **Other Irish test sites** |
| --- | --- | --- | --- |
| Jell.ie | Tcd.ie | Herdwatch.ie | Heanet.ie |
| Responsible.ie | scss.tcd.ie | Jamilin.com | iedr.ie |
|  |  | Pennyjuice.com | inex.ie |
|  |  | 007museum.com |  |
|  |  | Arngren.net |  |
|  |  | georgermartin.com |  |
|  |  | theroommovie.com |  |
|  |  | electrifyingtimes.com |  |

**Some background on different tests:**

**Heartbleed:**

The Heartbleed Bug is a serious vulnerability in the popular OpenSSL cryptographic software library. This weakness allows stealing the information protected, under normal conditions, by the SSL/TLS encryption used to secure the Internet. SSL/TLS provides communication security and privacy over the Internet for applications such as web, email, instant messaging (IM) and some virtual private networks (VPNs).

The Heartbleed bug allows anyone on the Internet to read the memory of the systems protected by the vulnerable versions of the OpenSSL software. This compromises the secret keys used to identify the service providers and to encrypt the traffic, the names and passwords of the users and the actual content. This allows attackers to eavesdrop on communications, steal data directly from the services and users and to impersonate services and users.

#### **What is being leaked?**

Encryption is used to protect secrets that may harm your privacy or security if they leak. In order to coordinate recovery from this bug we have classified the compromised secrets to four categories: 1) primary key material, 2) secondary key material and 3) protected content and 4) collateral.

#### **What is leaked primary key material and how to recover?**

These are the crown jewels, the encryption keys themselves. Leaked secret keys allow the attacker to decrypt any past and future traffic to the protected services and to impersonate the service at will. Any protection given by the encryption and the signatures in the X.509 certificates can be bypassed. Recovery from this leak requires patching the vulnerability, revocation of the compromised keys and reissuing and redistributing new keys. Even doing all this will still leave any traffic intercepted by the attacker in the past still vulnerable to decryption. All this has to be done by the owners of the services.

#### **What is leaked secondary key material and how to recover?**

These are for example the user credentials (user names and passwords) used in the vulnerable services. Recovery from this leak requires owners of the service first to restore trust to the service according to steps described above. After this users can start changing their passwords and possible encryption keys according to the instructions from the owners of the services that have been compromised. All session keys and session cookies should be invalidated and considered compromised.

#### **What is leaked protected content and how to recover?**

This is the actual content handled by the vulnerable services. It may be personal or financial details, private communication such as emails or instant messages, documents or anything seen worth protecting by encryption. Only owners of the services will be able to estimate the likelihood what has been leaked and they should notify their users accordingly. Most important thing is to restore trust to the primary and secondary key material as described above. Only this enables safe use of the compromised services in the future.

#### **What is leaked collateral and how to recover?**

Leaked collateral are other details that have been exposed to the attacker in the leaked memory content. These may contain technical details such as memory addresses and security measures such as canaries used to protect against overflow attacks. These have only contemporary value and will lose their value to the attacker when OpenSSL has been upgraded to a fixed version.

**HSTS:**

HTTP Strict Transport Security (HSTS) is a web server directive that informs user agents and web browsers how to handle its connection through a response header sent at the very beginning and back to the browser.

This sets the Strict-Transport-Security policy field parameter. It forces those connections over HTTPS encryption, disregarding any script's call to load any resource in that domain over HTTP. HSTS is but one arrow in a bundled sheaf of security settings for your web server or your web hosting service.

You never close your physical store or home without locking the doors, right? You may even have metal detectors at the door to control shrinkage. Data can be just as valuable as physical items in your shop or house, so it’s just as important to keep them locked up and secure. Padlocking your website is sometimes not enough as people will still find a way to reach your website over http://. HSTS forces browsers and app connections to use HTTPS if that is available. Even if someone just types in the www or http://.

HTTPS is a small ranking factor in Google and is categorized as a ‘site quality’ score along with many other factors such as page speed and mobile responsiveness.

Setting up 301 redirects from http:// to https:// is not enough to completely secure your domain name. The window of opportunity still exists in the insecure redirection of HTTP.

The hacker(s) are still able to capture site cookies, session ID (usually sent as a URL parameter) or force a redirection to their phishing site that looks exactly like your website.

By having a Strict-Transport-Security header installed, it will be nearly impossible for the bad guys to glean any information at all!

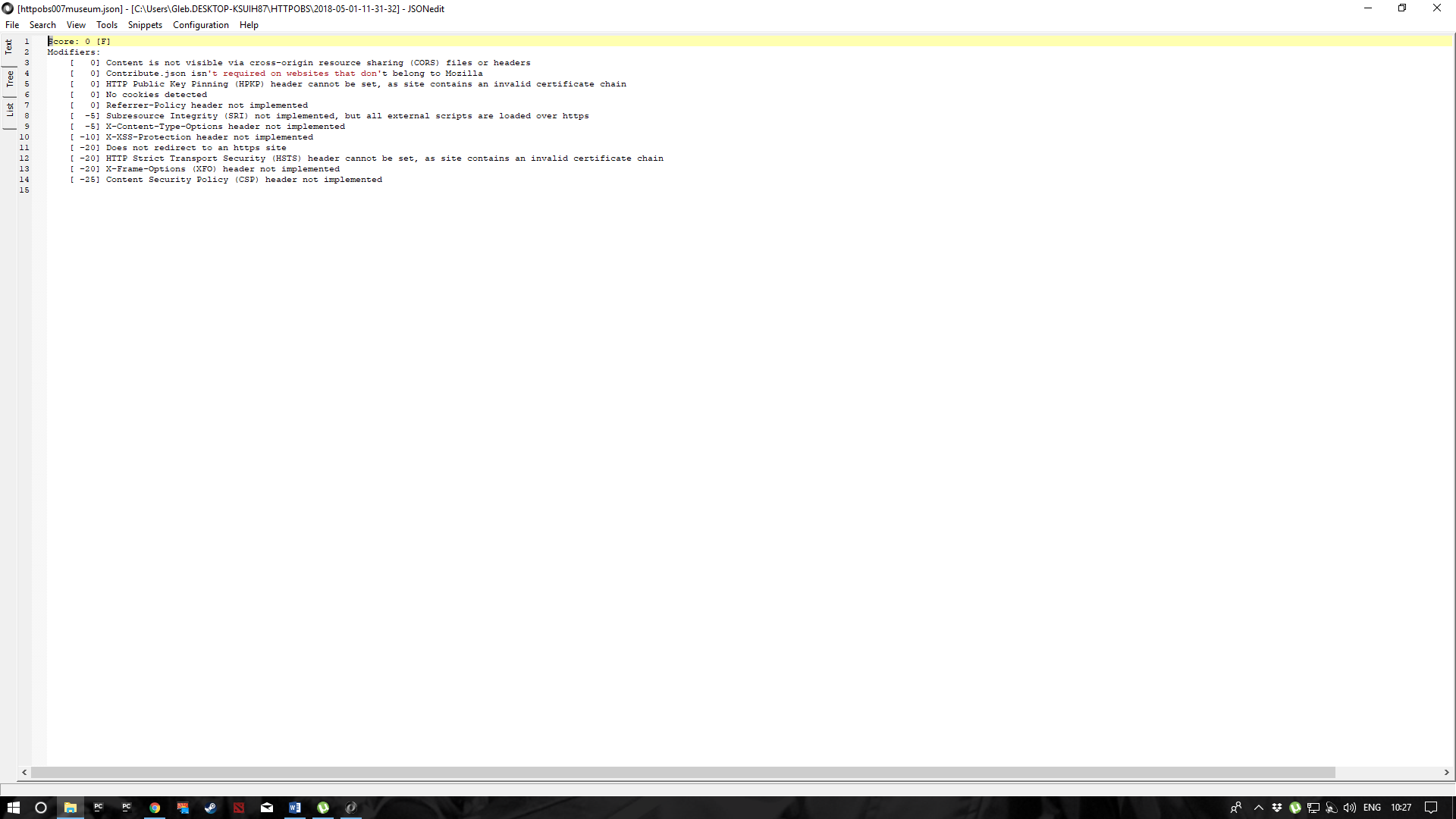
***Results:***

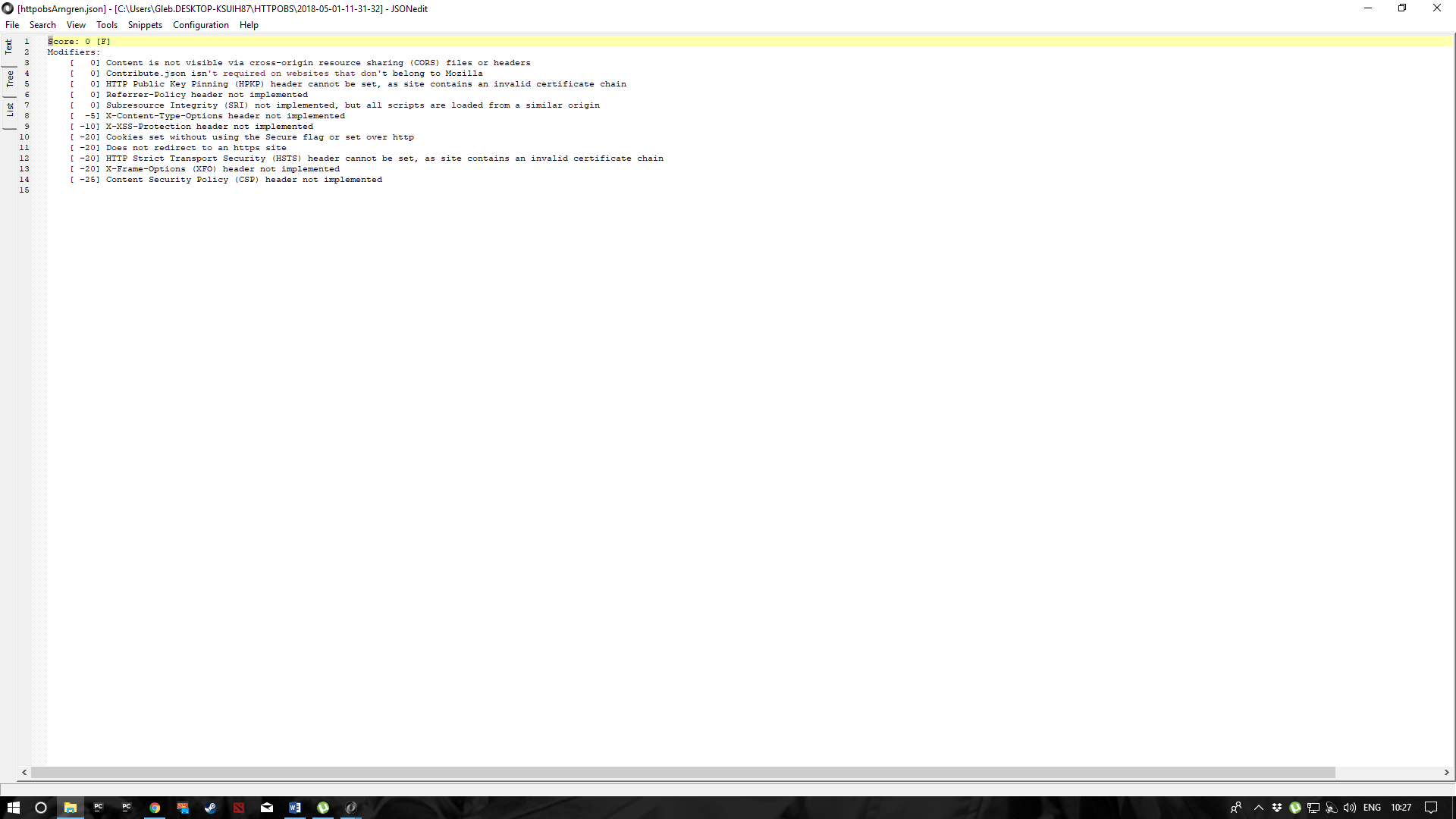
***Privacyscore:***

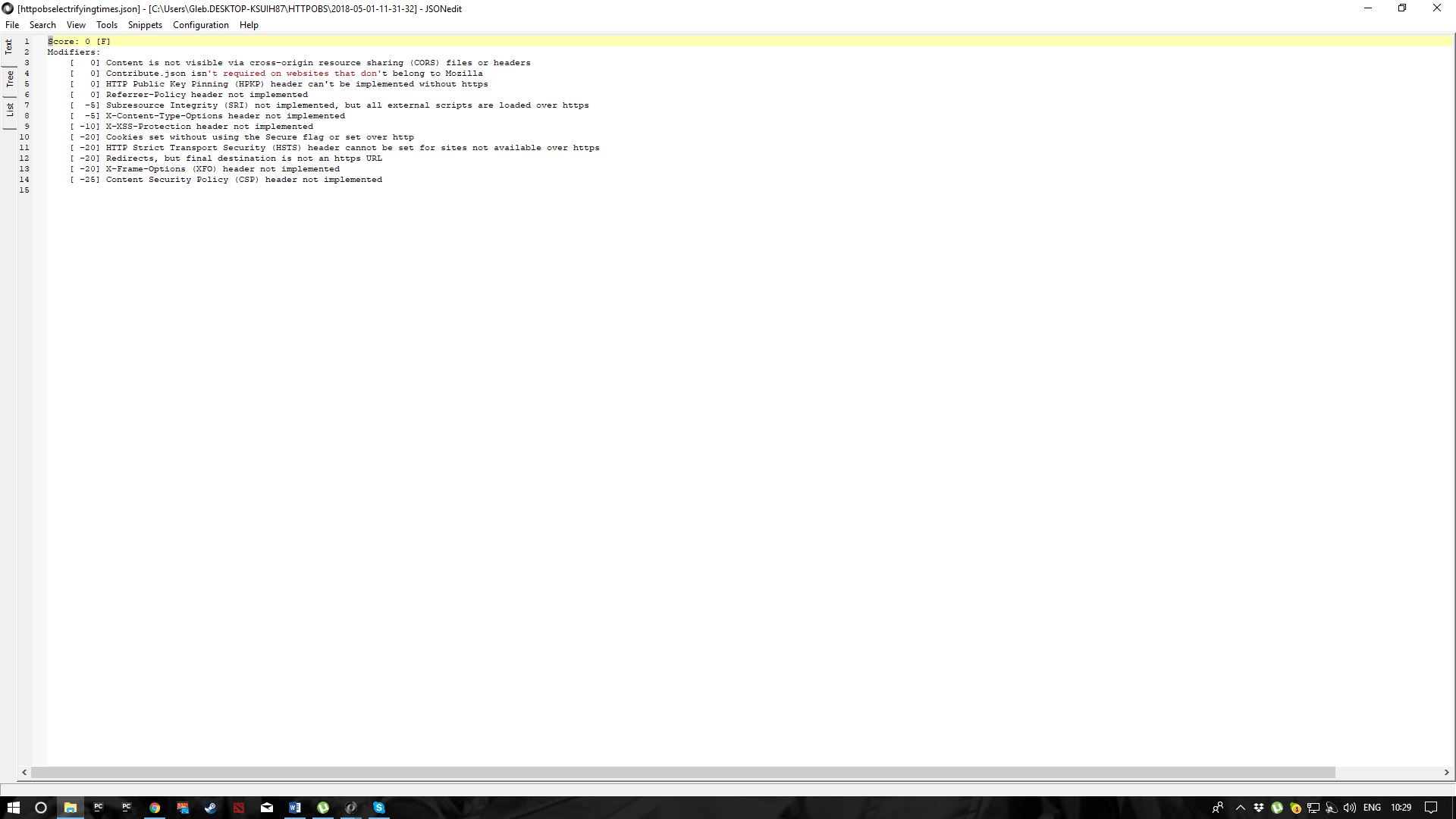
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Jell.ie | Respobnsble.ie | Scss.tcd.ie | Tcd.ie | Herdwatch.ie | Heanet.ie | Iedr.ie | Inex.ie |
| 100 | 91 | 65 | 53 | 50 | 85 | 51 | 58 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Jamilin.com | Pennyjuice.com | 007museum.com | Argngren.net | Georgematin.com | Theroommovie.com | Electrifyingtimes. com |
| 11 | 35 | -33 | -22 | 63 | 28 | 8 |

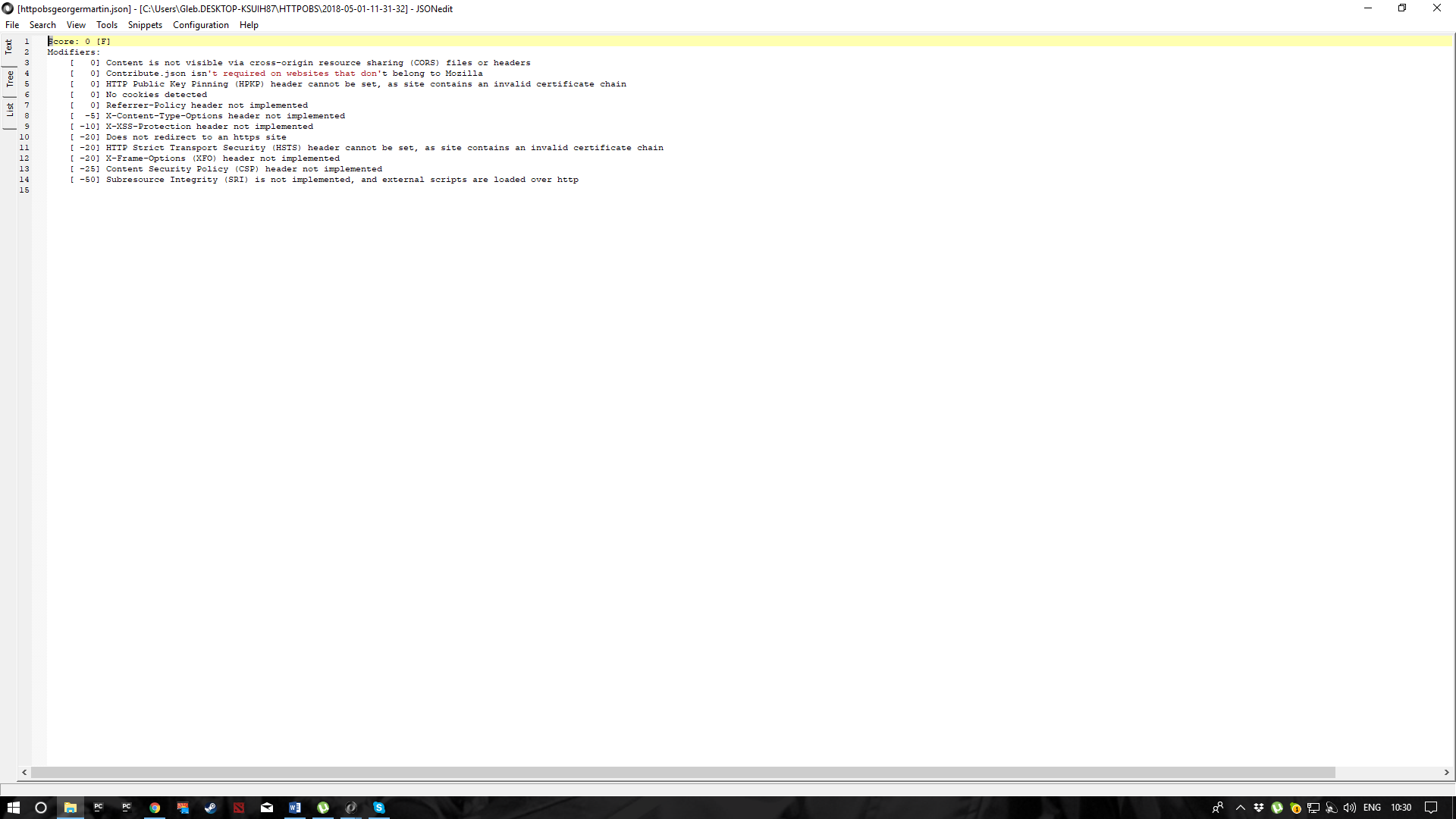
***HTTPobservatory:***

**007museum:**

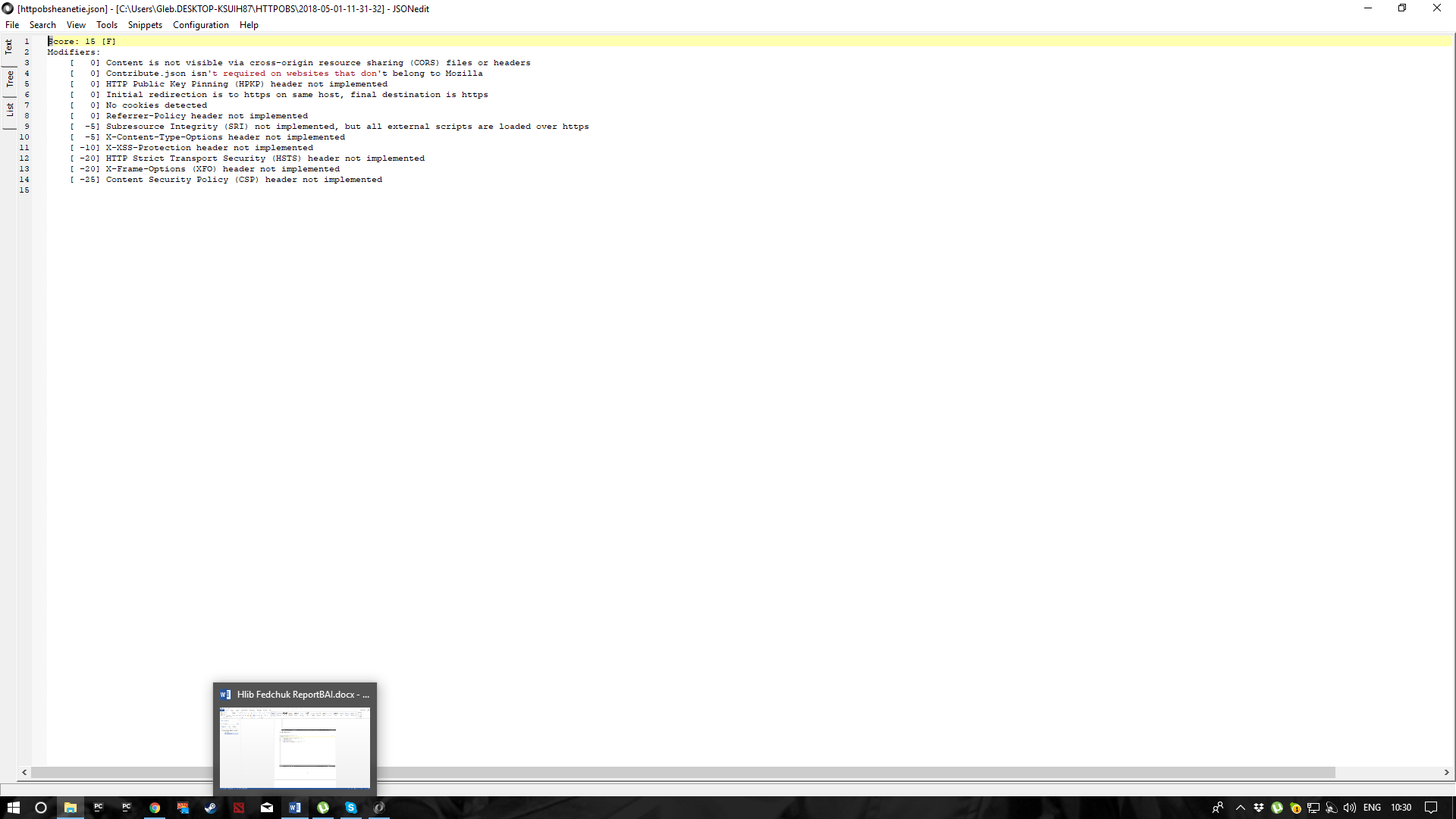
**Arngren.net:**

**Elecrifyingtimes.com:**

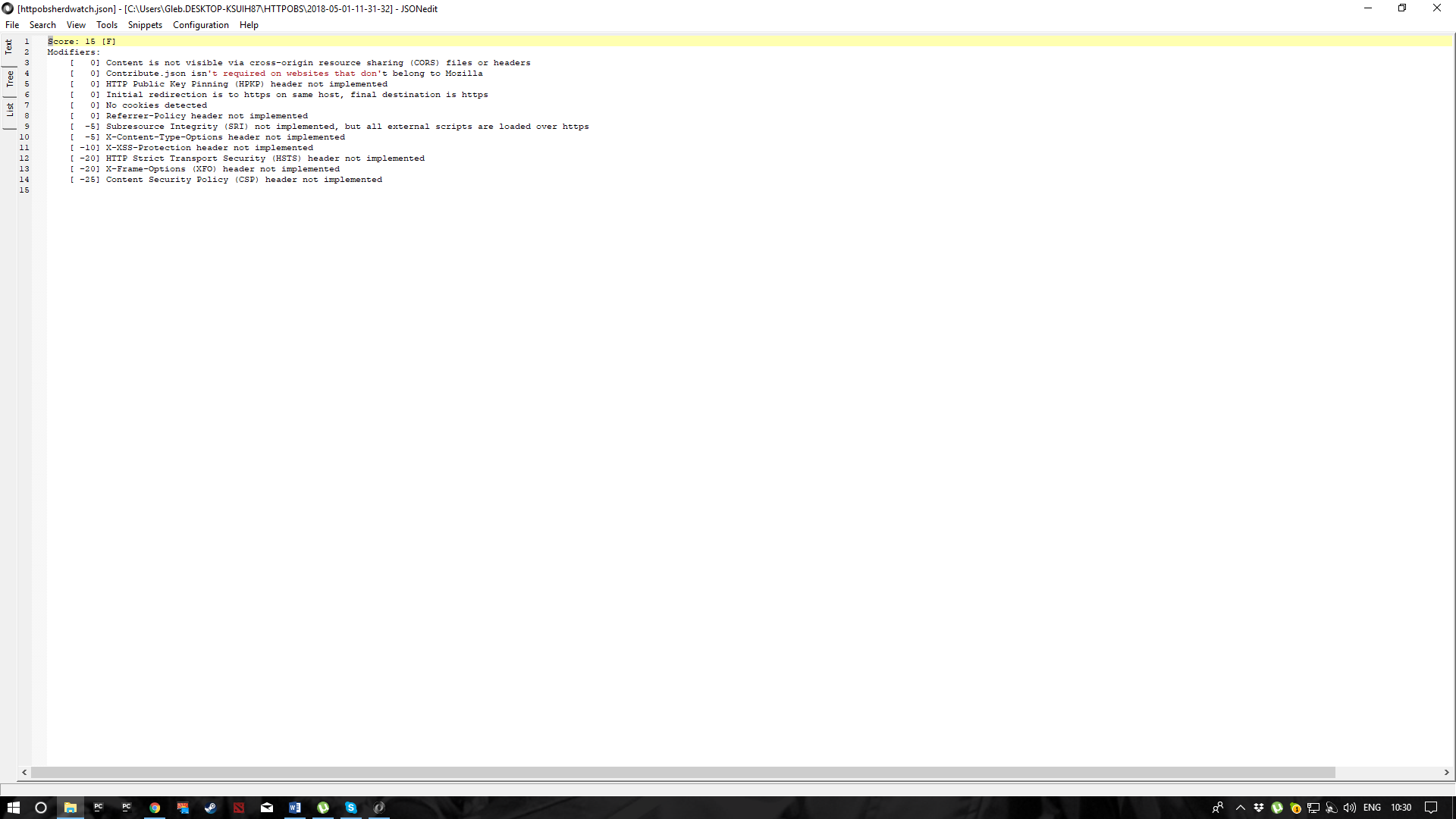
**Georgermartin.com:**

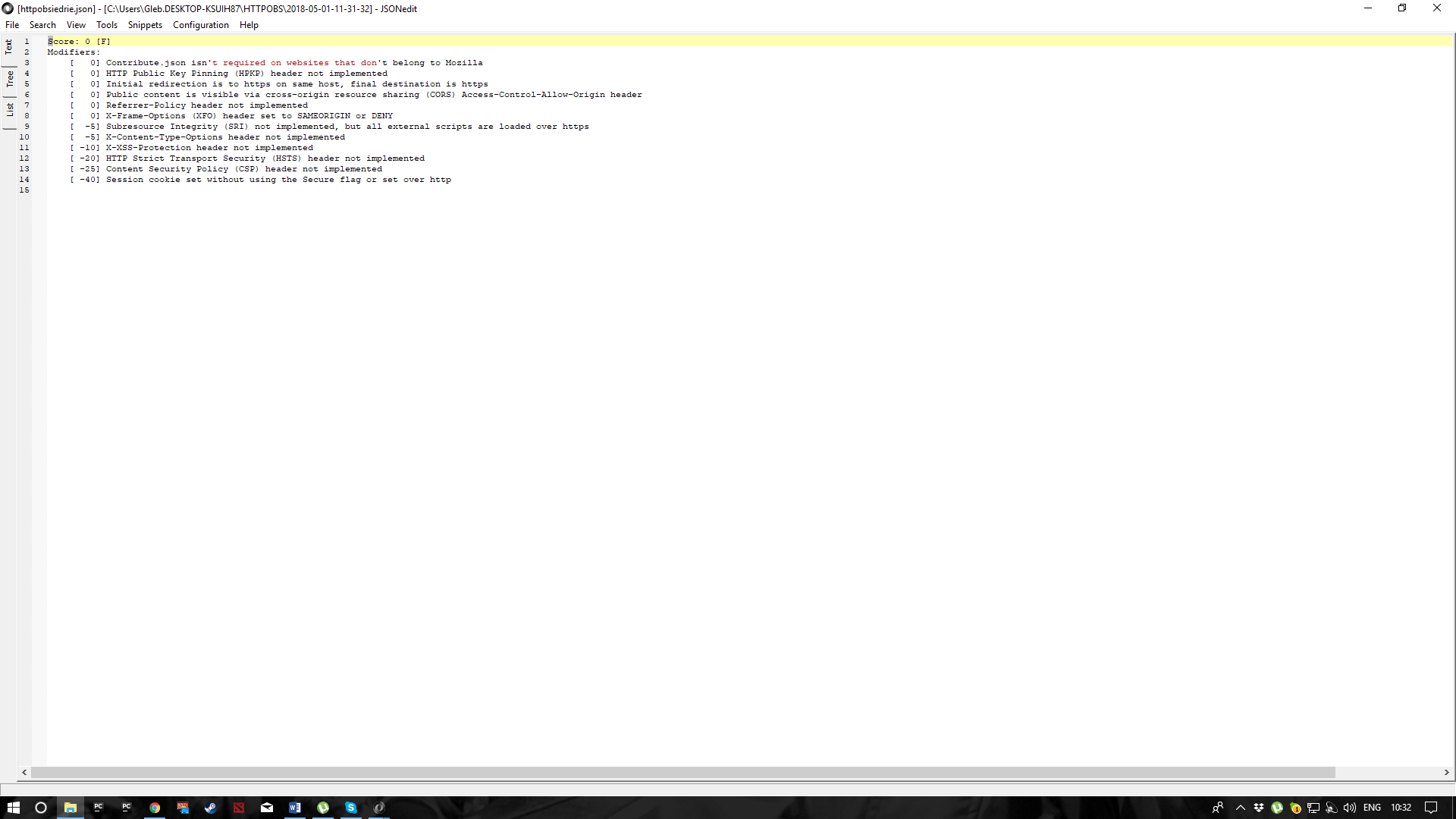
****

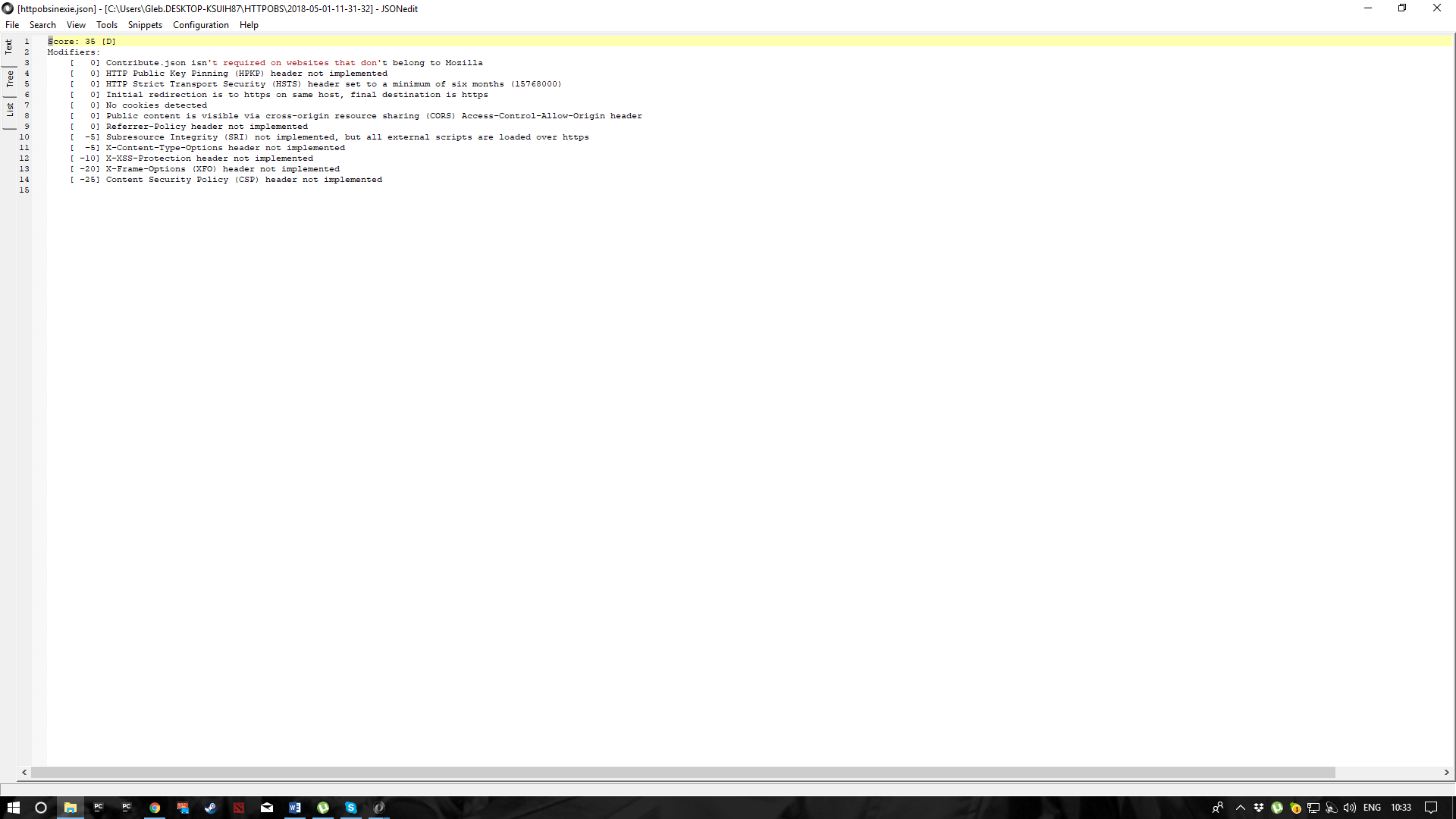
**Heanet.ie**

****

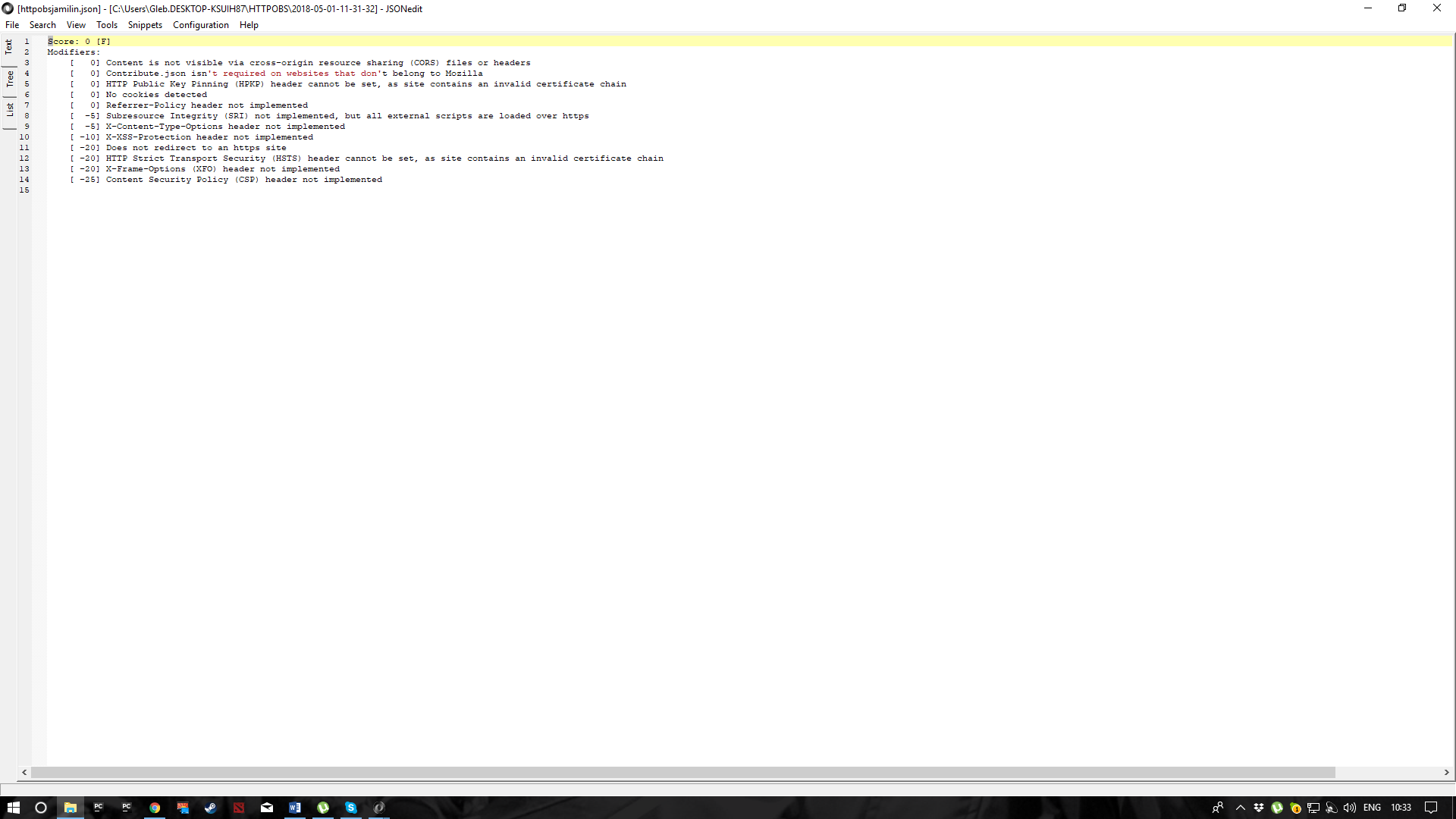
**Herdwatch.ie:**

****

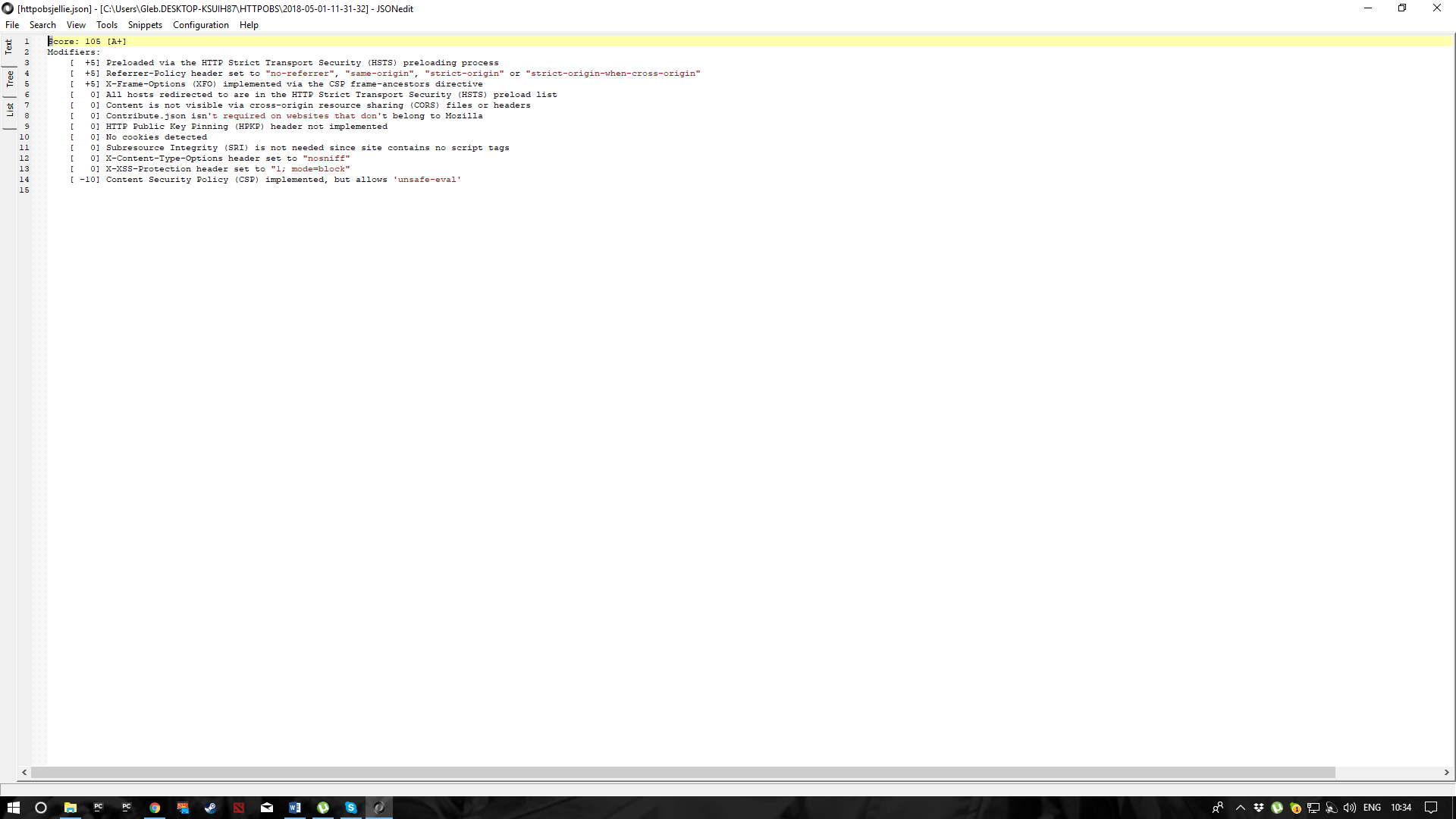
**Iedr.ie:**

**Inex.ie:**

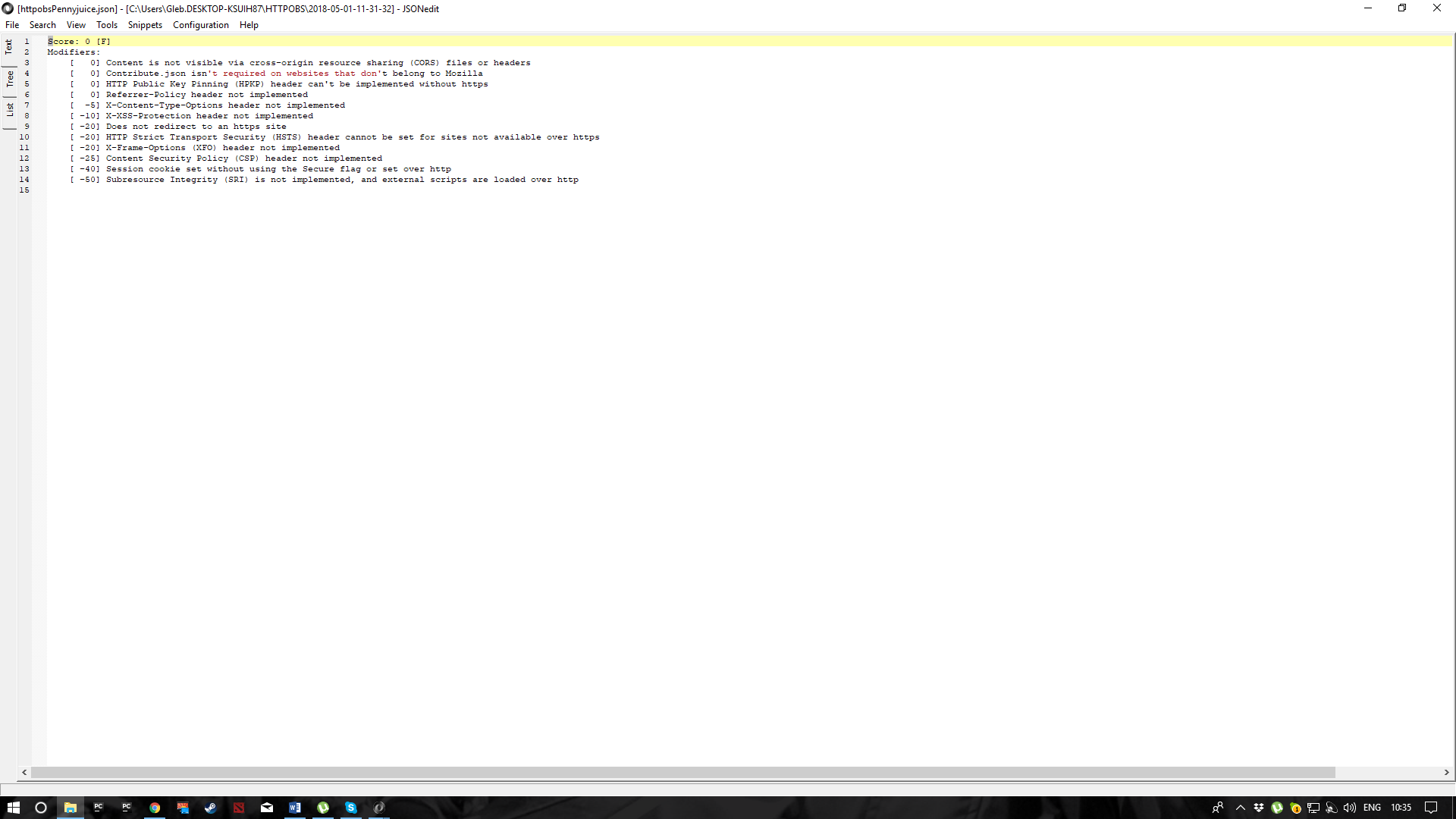
**Jamilin.com:**

****

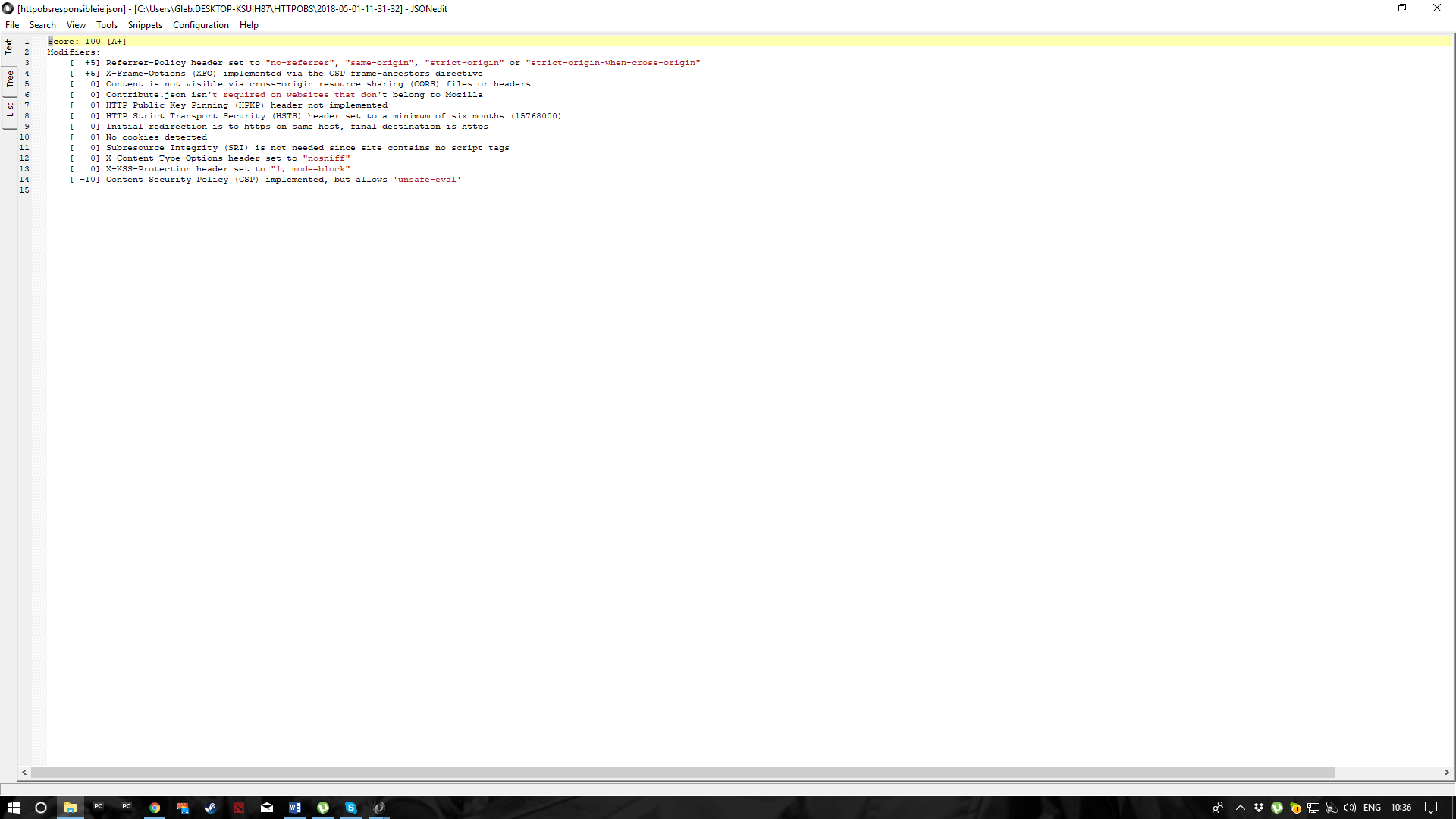
**Jell.ie:**

****

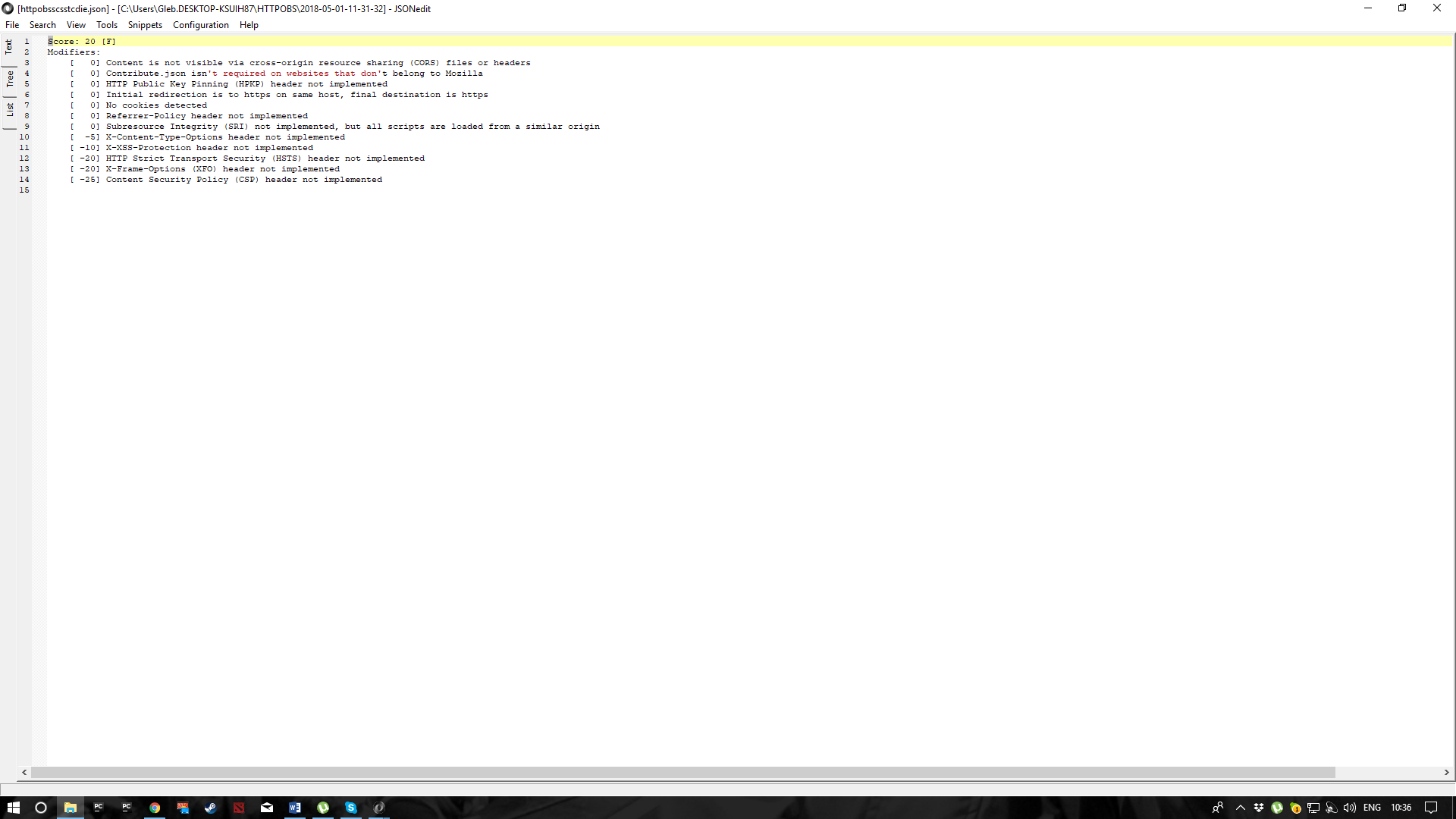
**PennyJuice.com:**

****

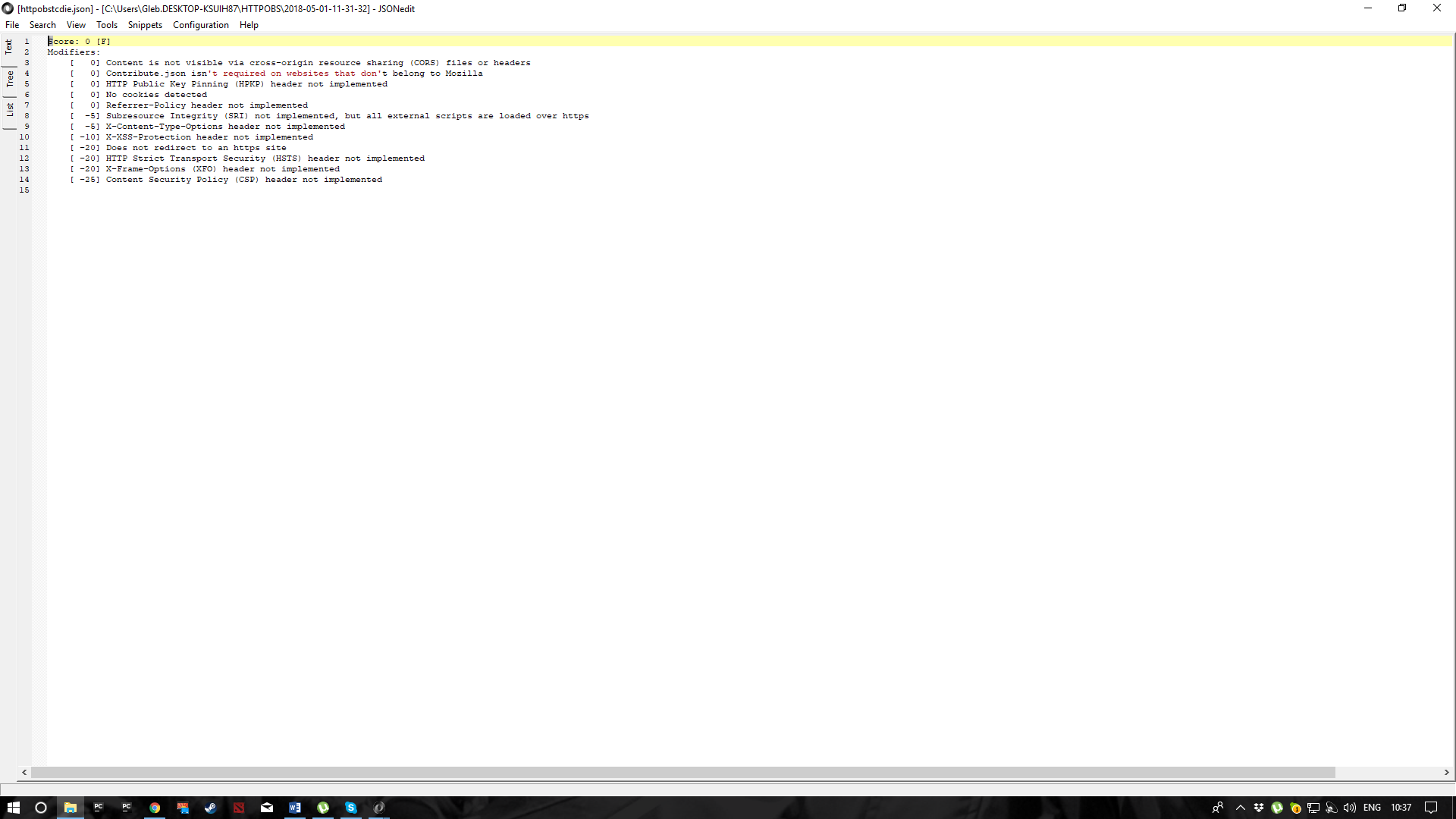
**Responsible.ie:**

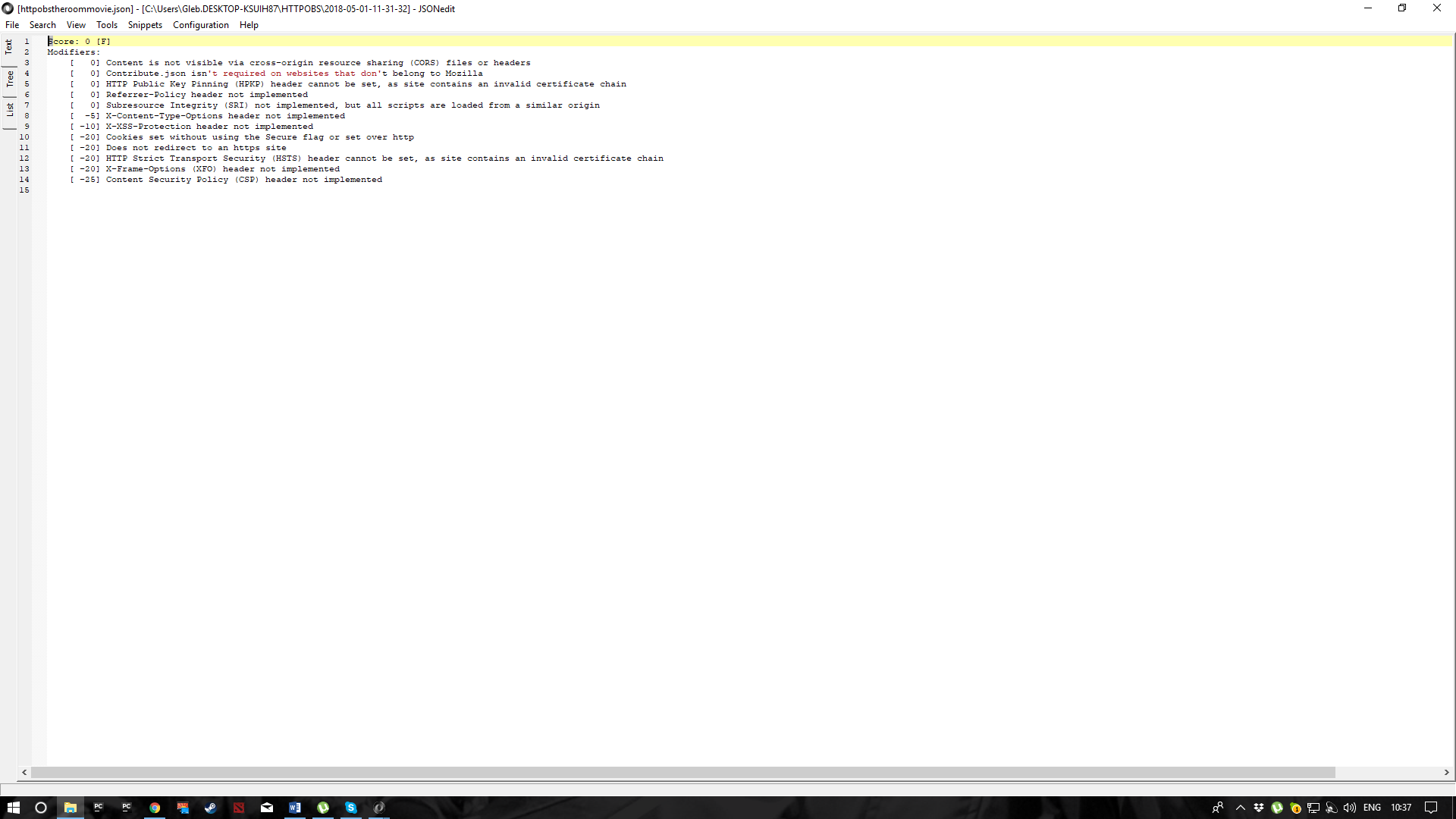
****

**Scss.tcd.ie:**

****

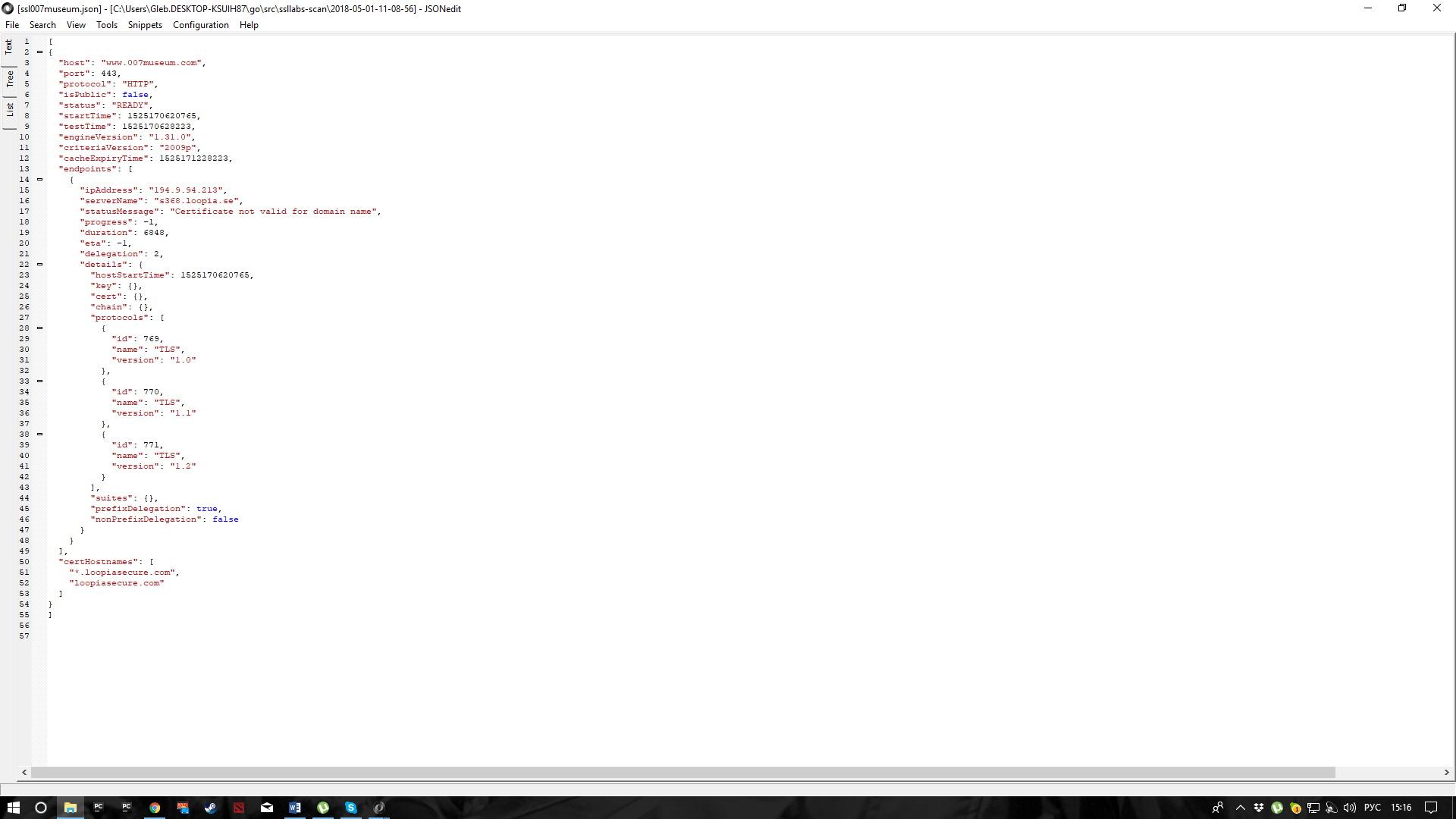
**Tcd.ie:**

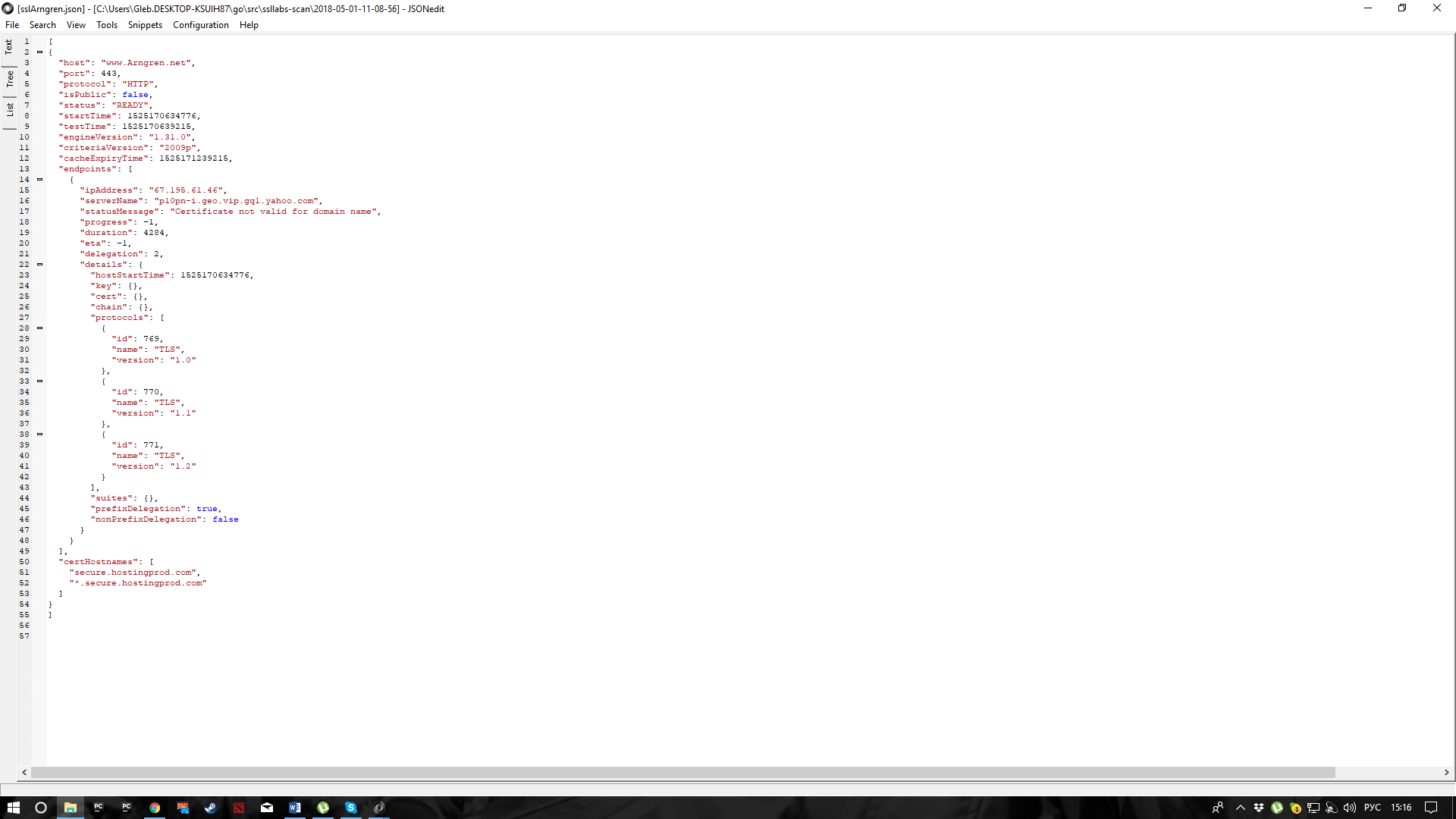
****

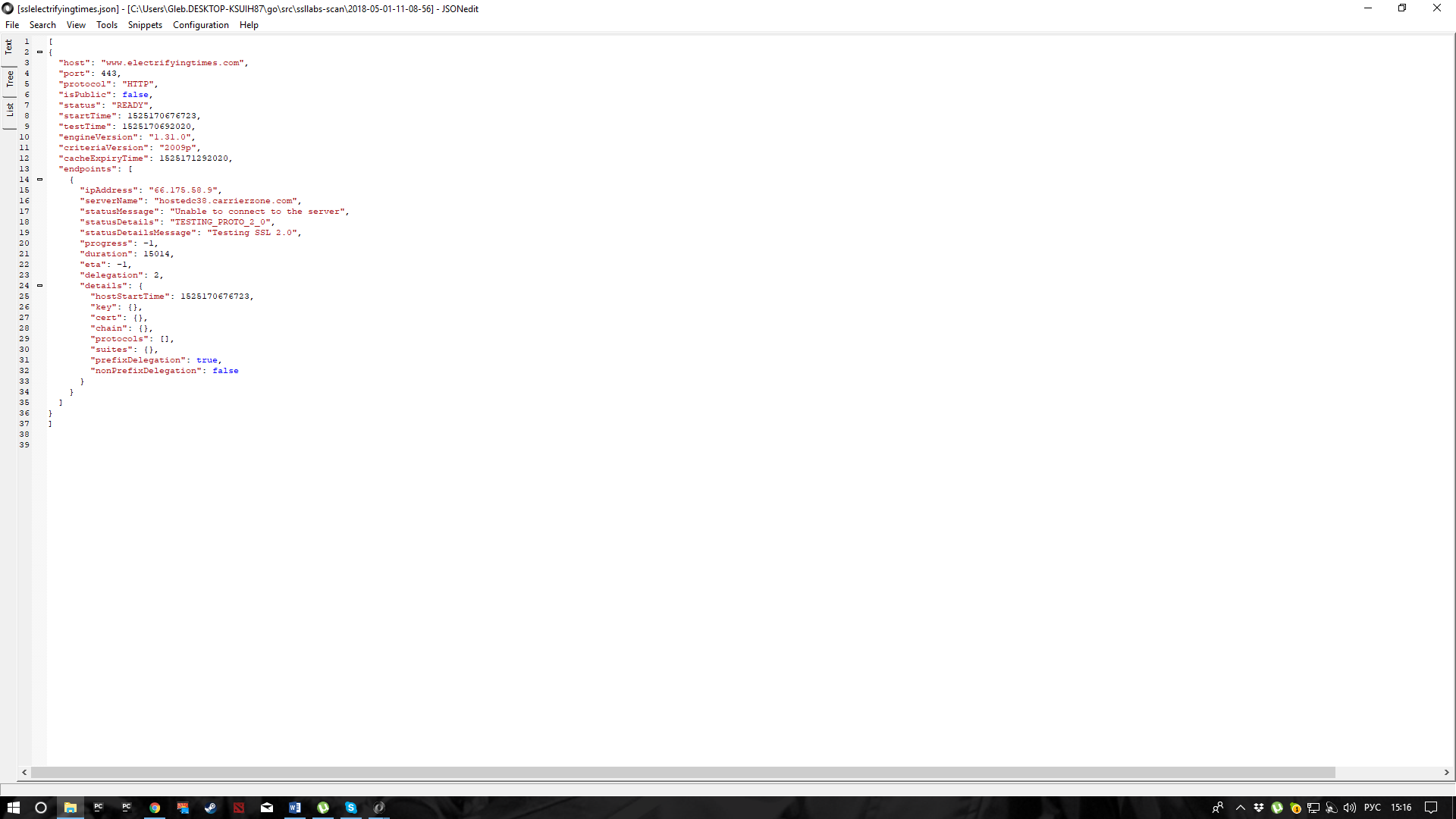
**Theroommovie.com:**

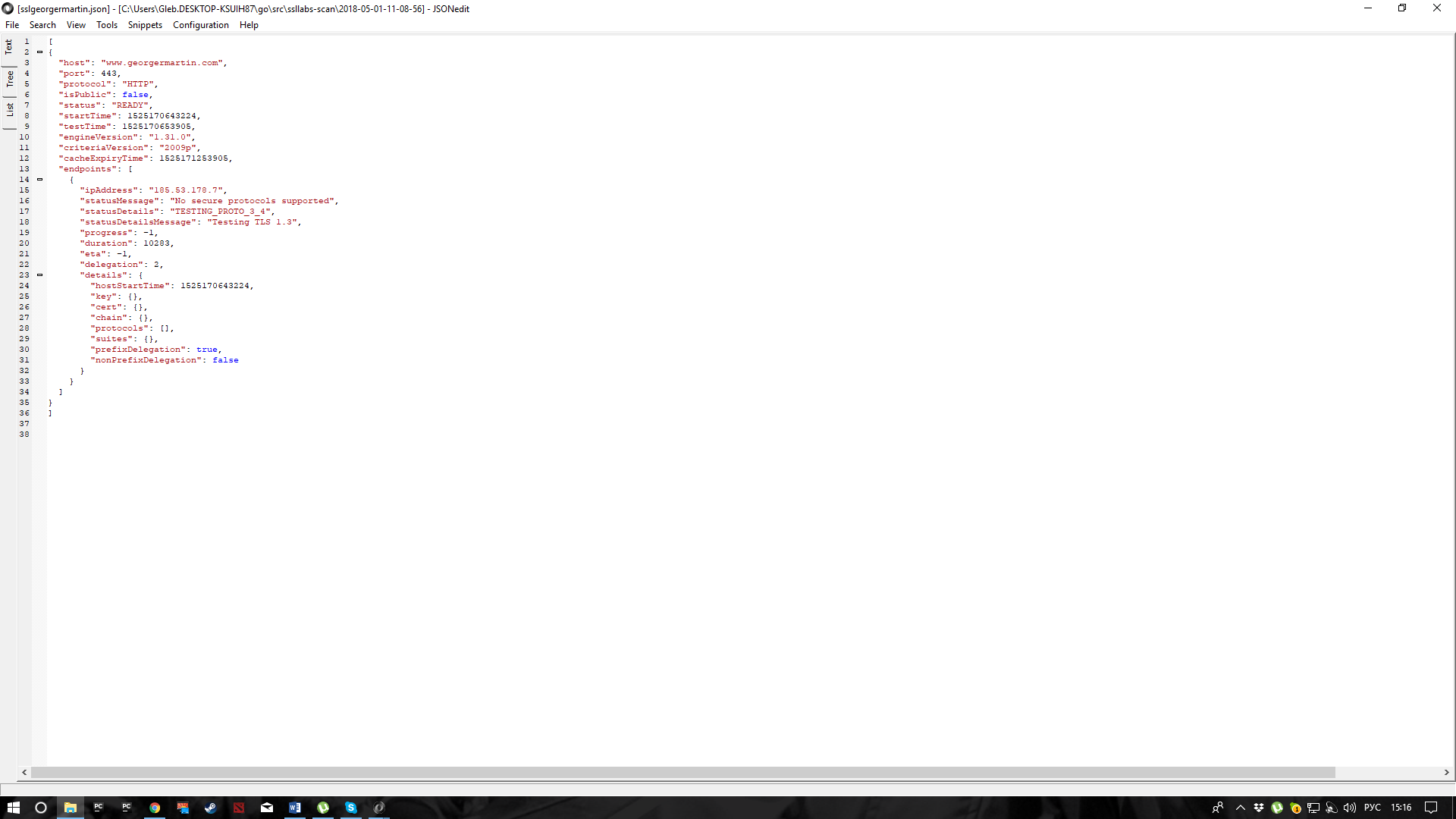
***SSLLabs:***

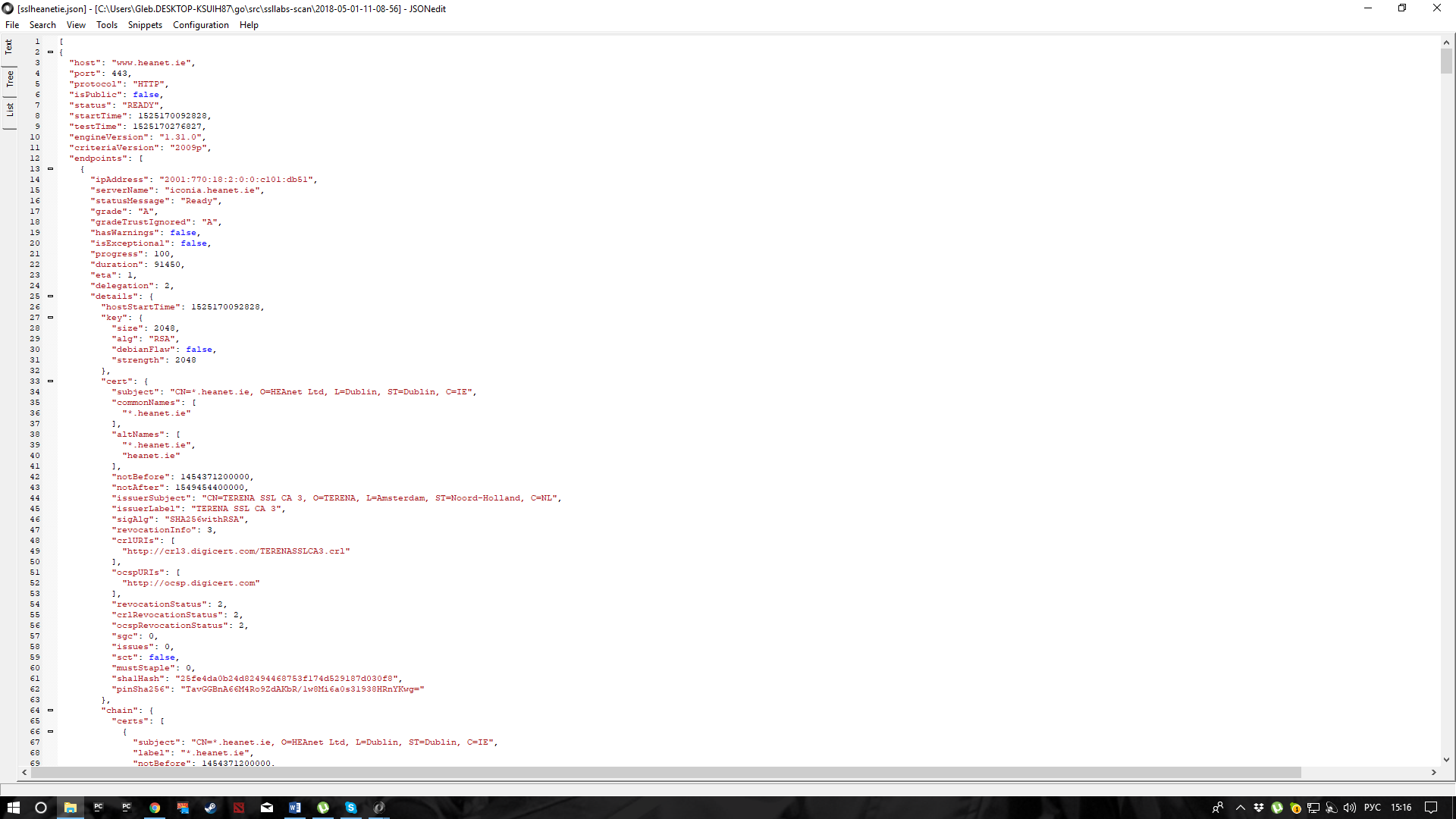
**007museum:**

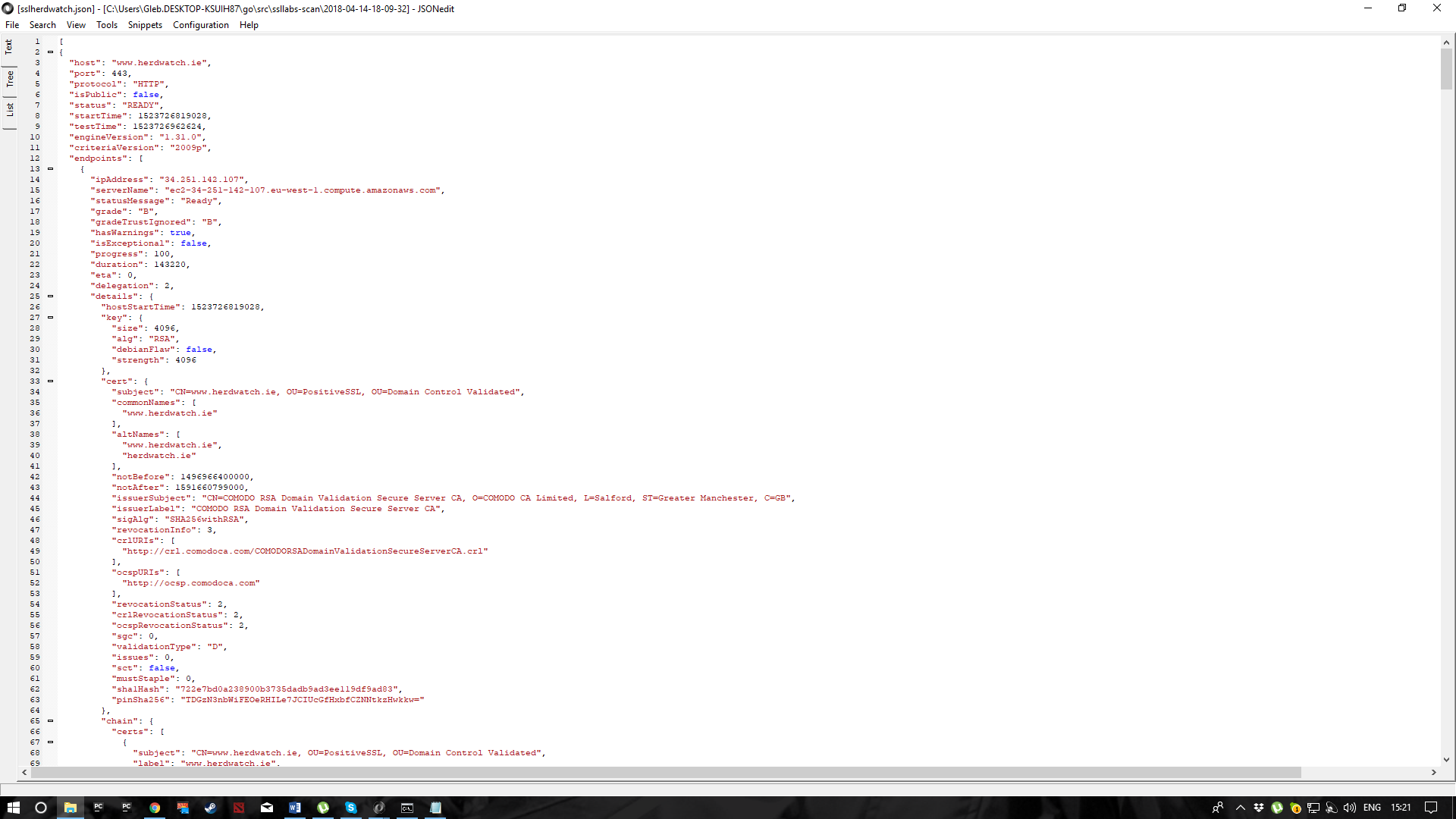


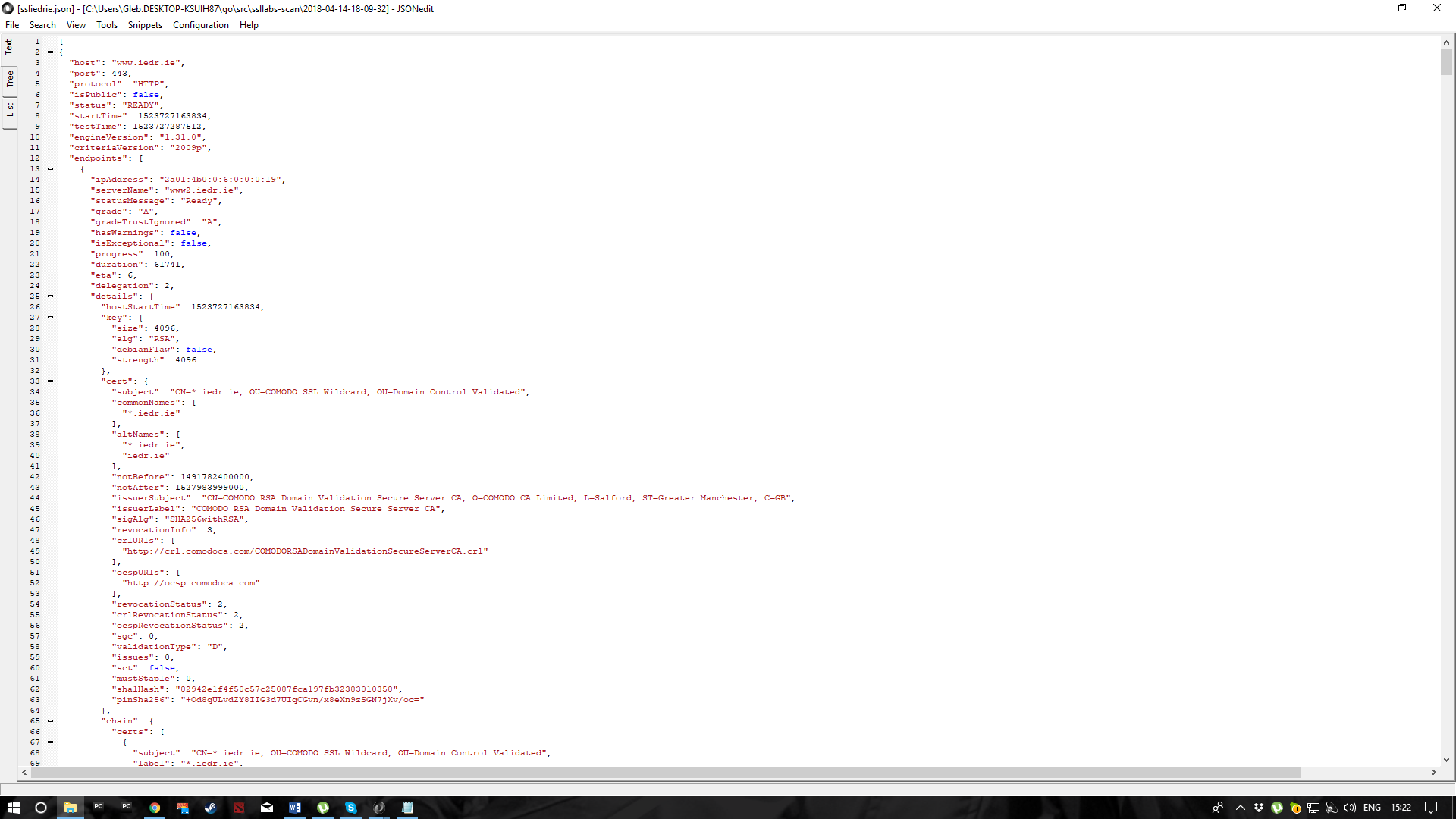
**Arngren.net:**

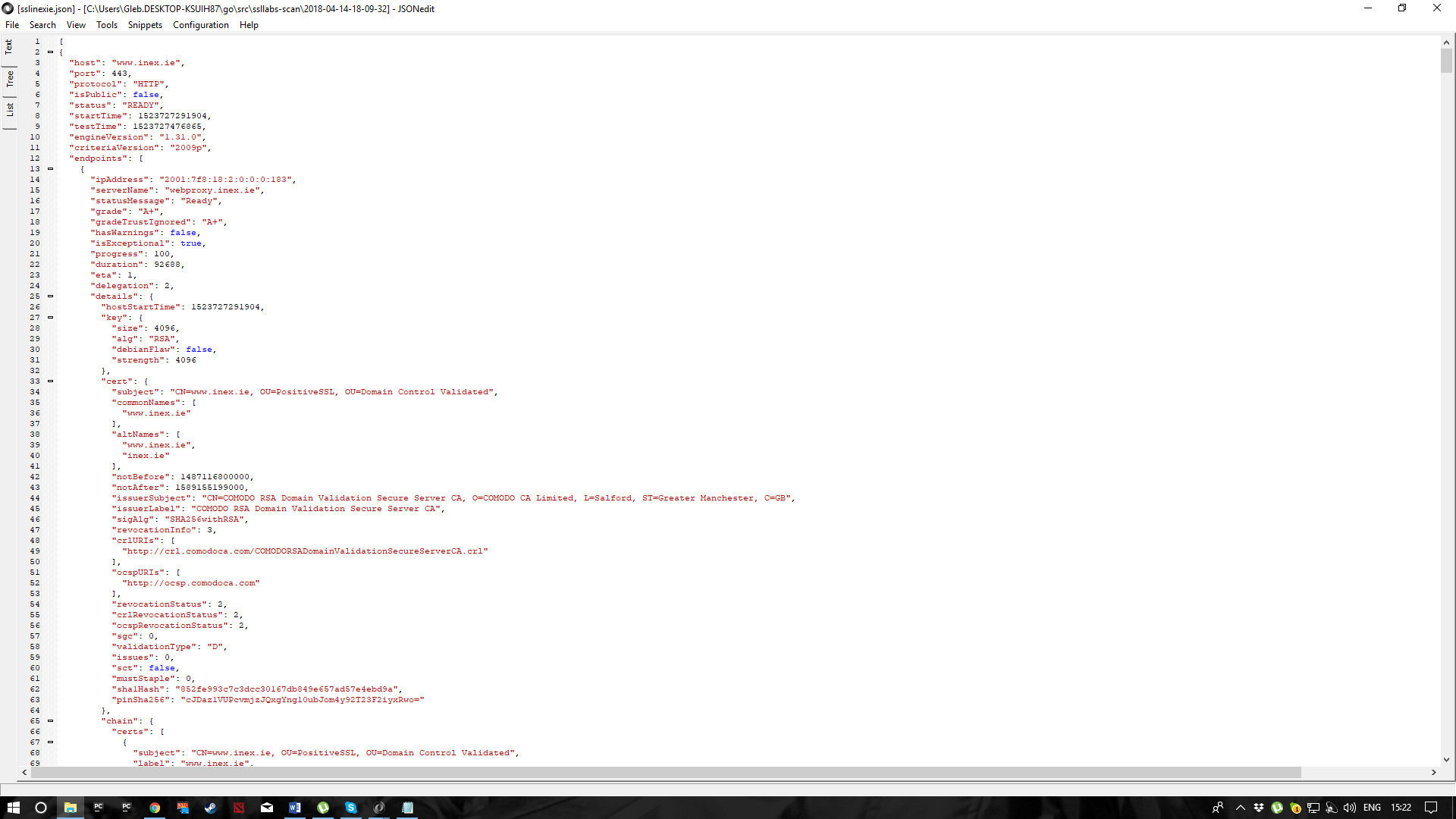
**Elecrifyingtimes.com:**

**Georgermartin.com:**

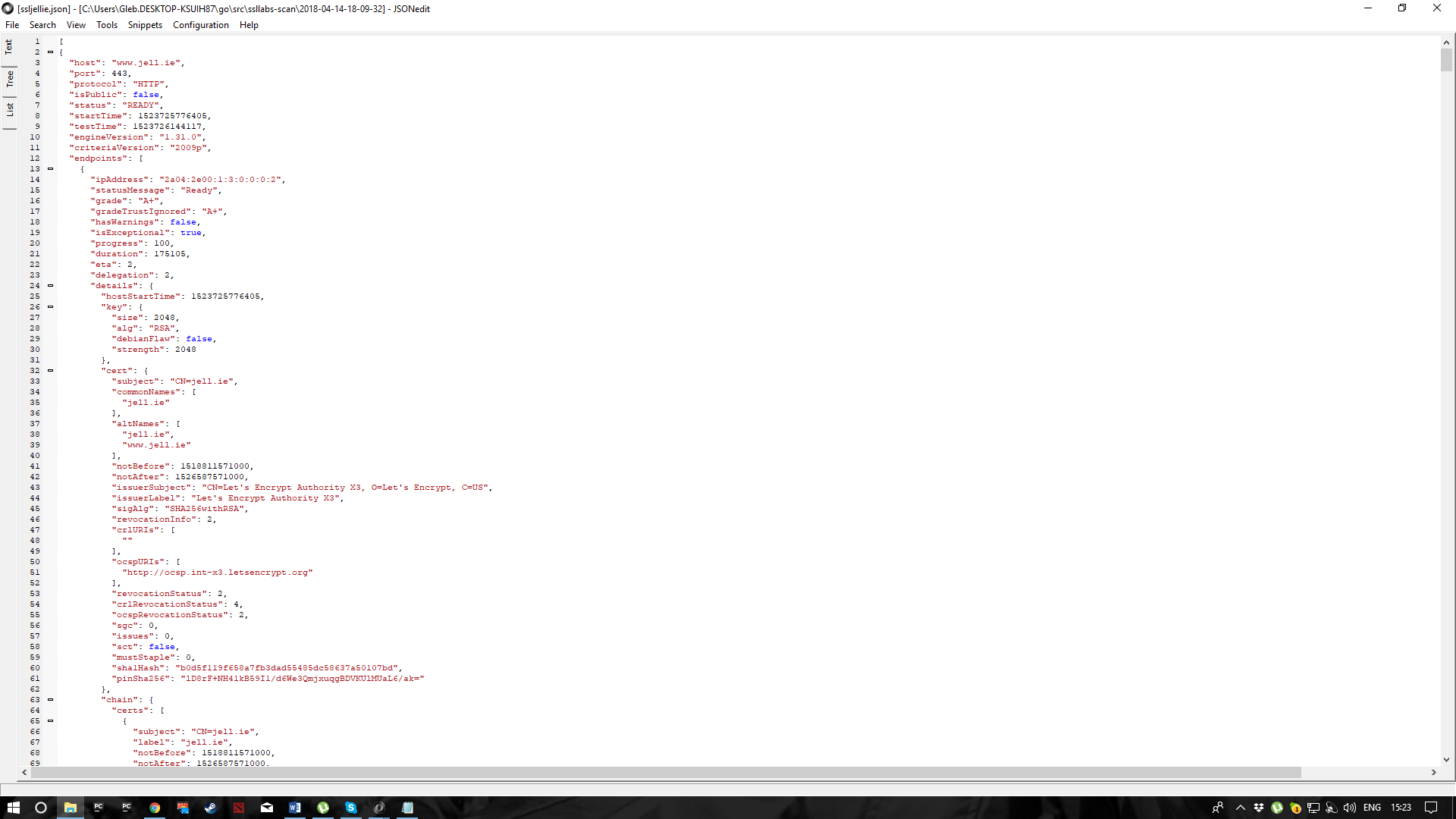
**Heanet.ie**

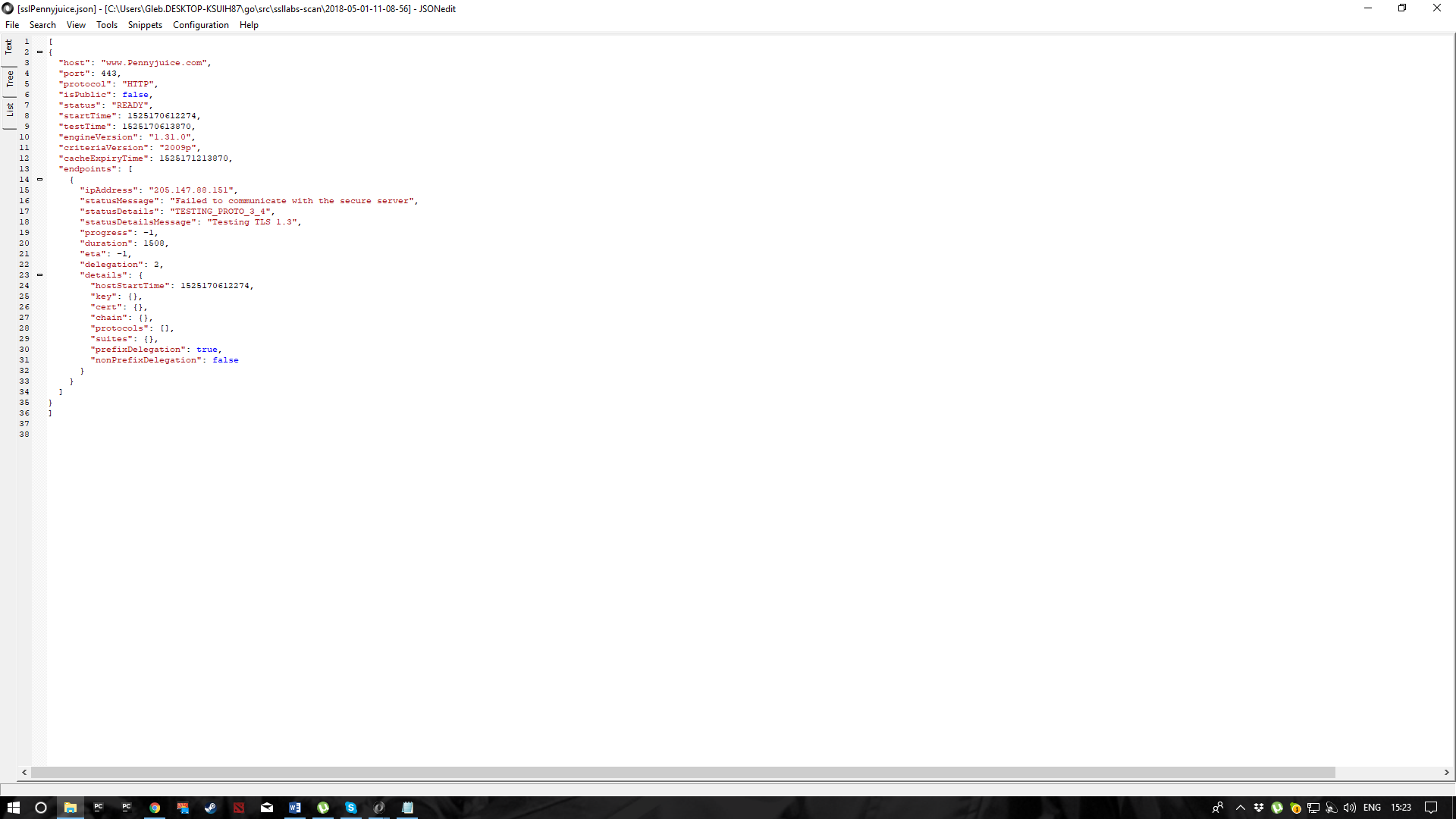
**Herdwatch.ie:**

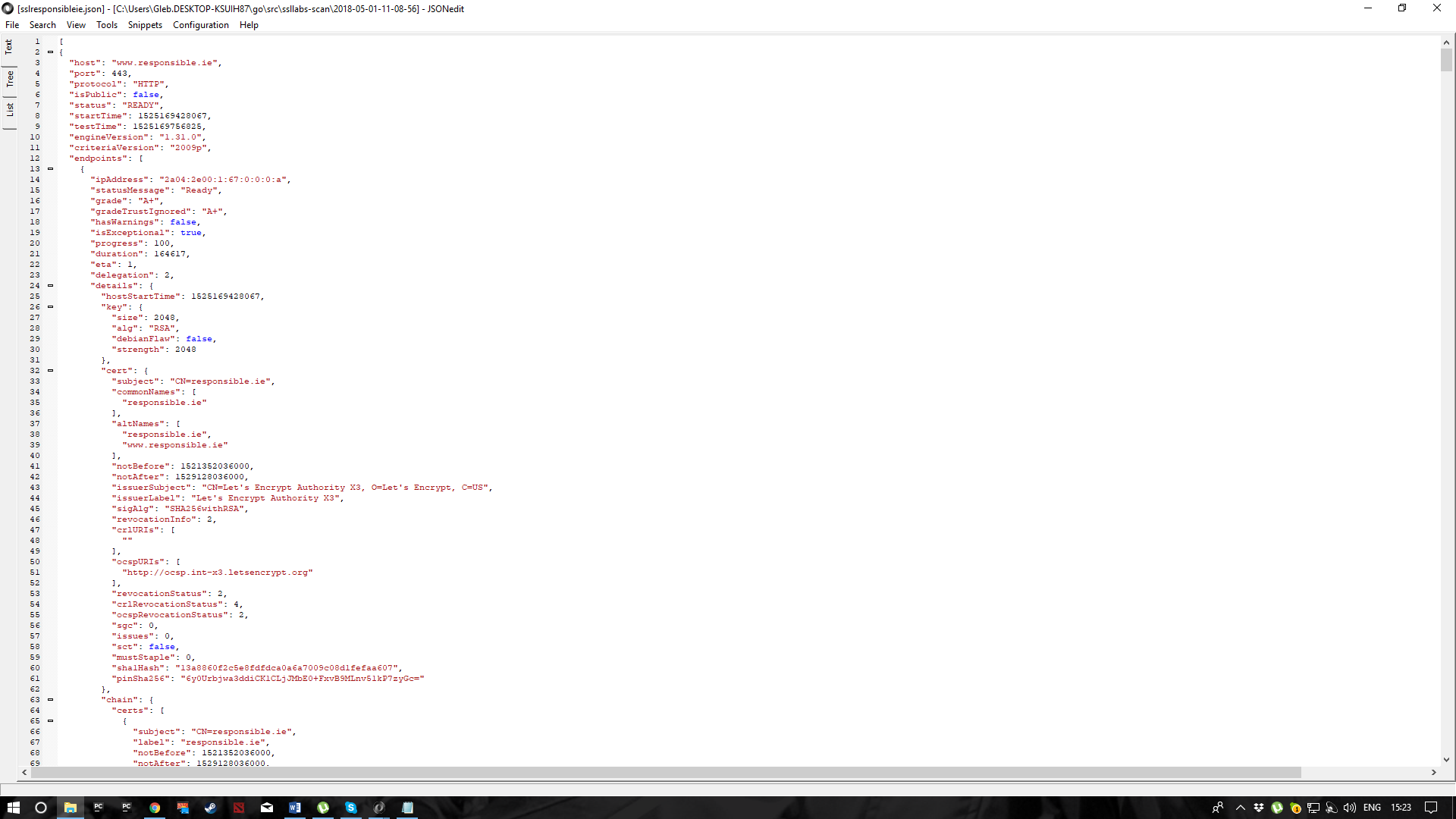
**Iedr.ie:**

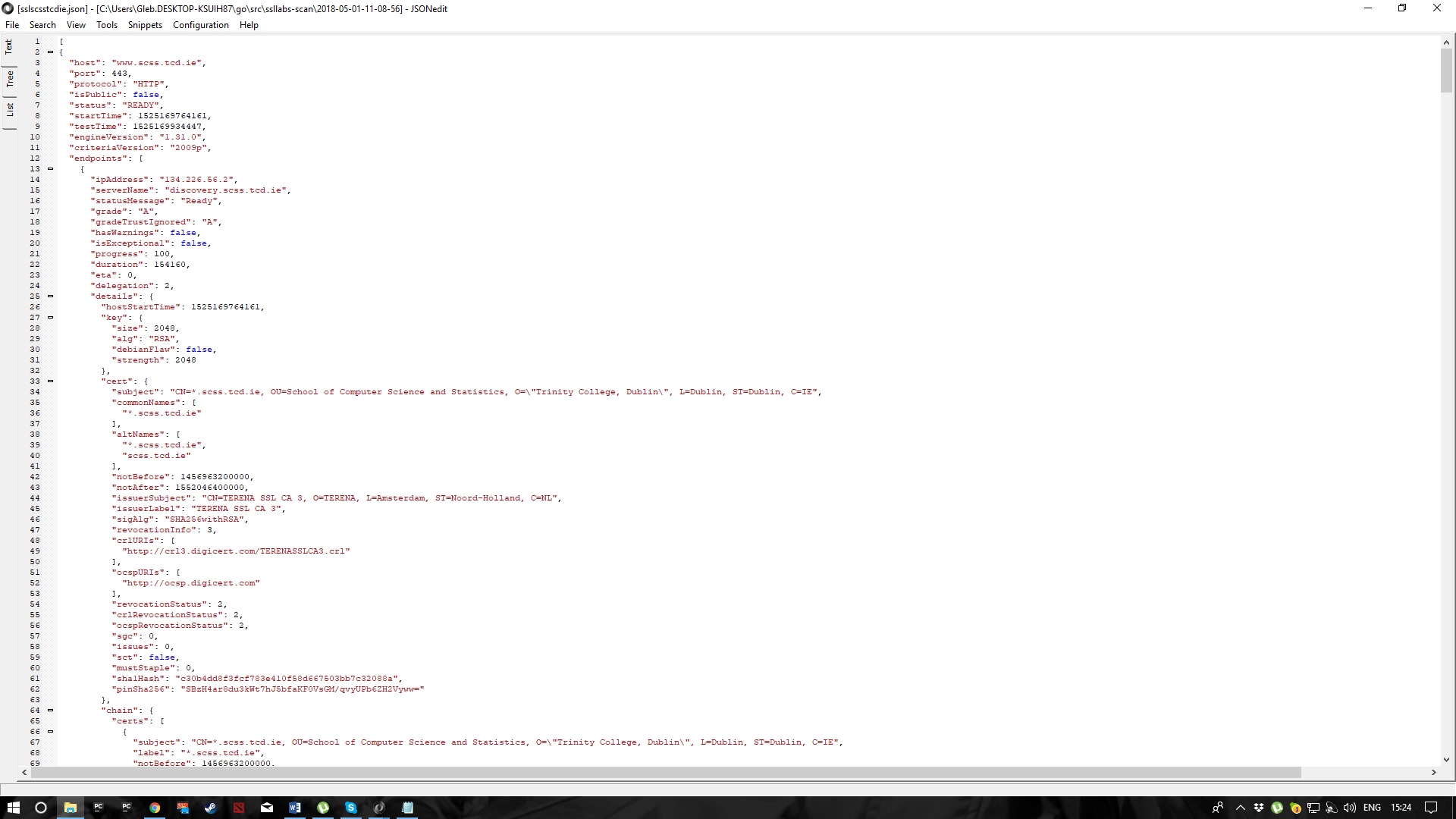
**Inex.ie:**

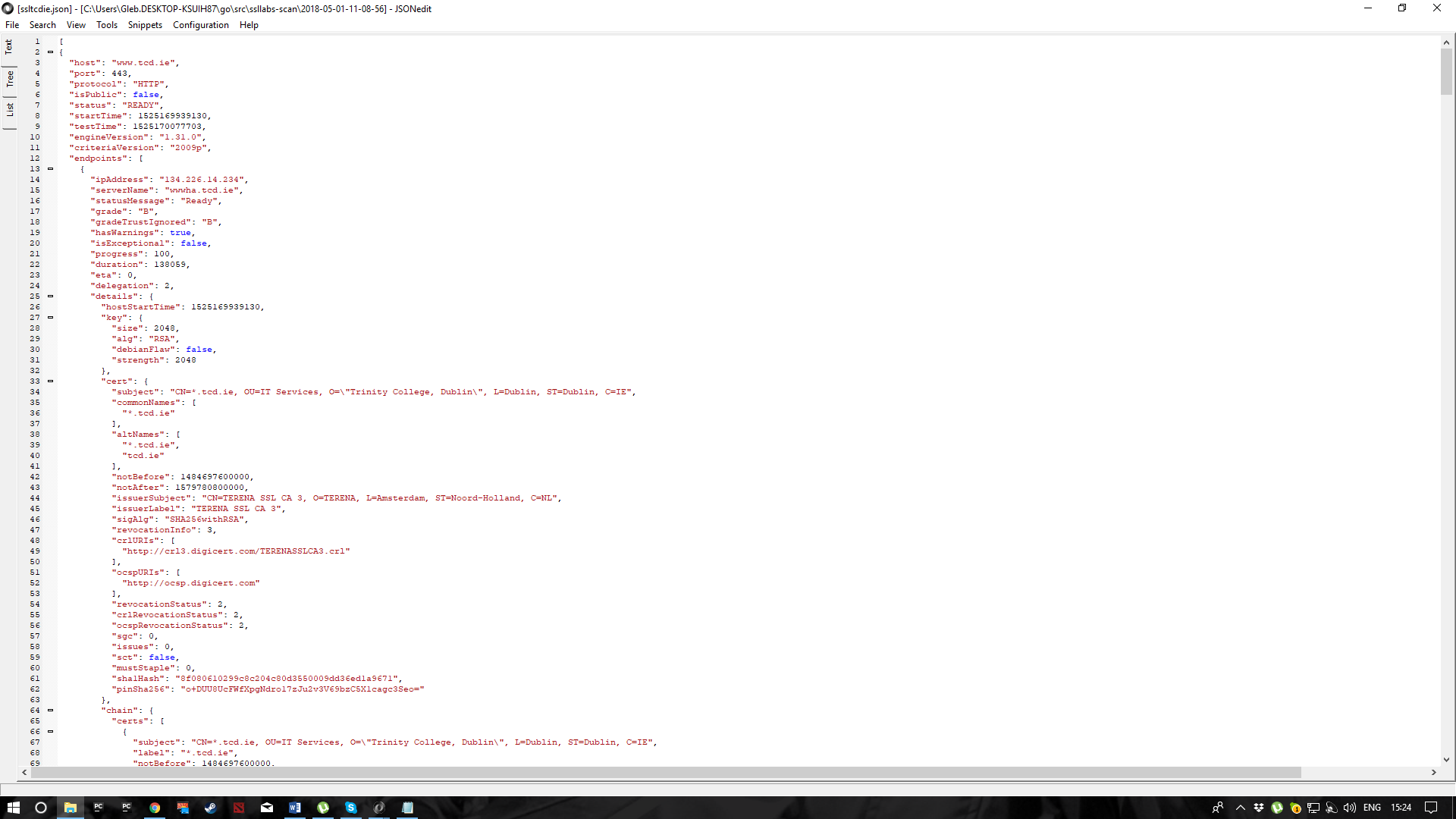
**Jamilin.com:**

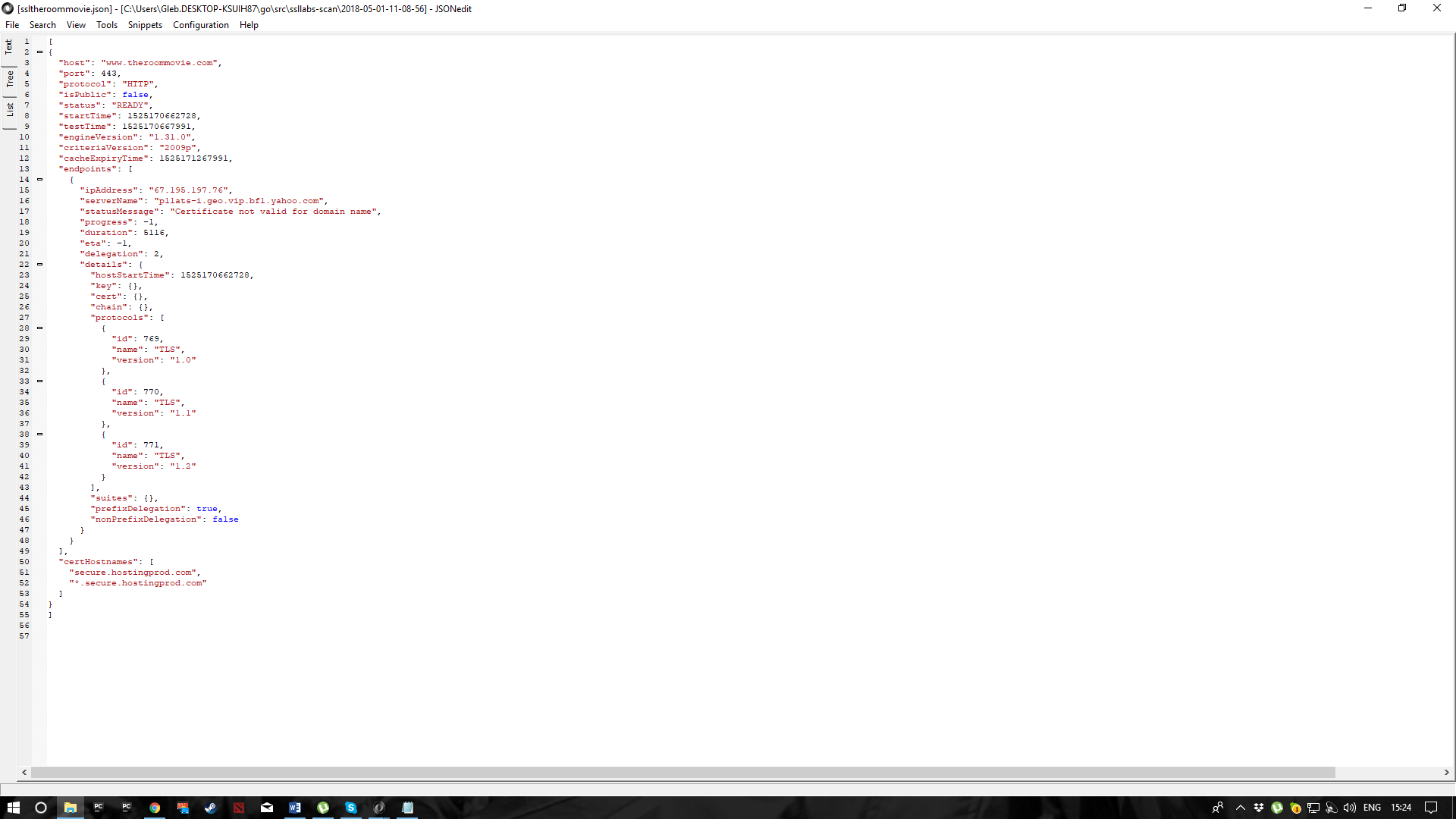
**Jell.ie:**

**PennyJuice.com:**

**Responsible.ie:**

**Scss.tcd.ie:**

**Tcd.ie:**

**Theroommovie.com:**

***Discussion:***

***Conclusion:***

***References:***

[http://www.axiompanbiog.com/Pages/HyperGraph.aspx,](http://www.axiompanbiog.com/Pages/HyperGraph.aspx)

[https://networkx.readthedocs.io/en/stable/.](https://networkx.readthedocs.io/en/stable/),

[http://courses.cs.ut.ee/2011/graphmining/Main/Demos,](http://courses.cs.ut.ee/2011/graphmining/Main/Demos)

[https://phys.org/news/2009-02-greedy-routing-enables-network.html,](https://phys.org/news/2009-02-greedy-routing-enables-network.html)

[http://math.stackexchange.com/questions/1039308/graph-theory-clique-graph,](http://math.stackexchange.com/questions/1039308/graph-theory-clique-graph)

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415>[-47572008000400001,](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572008000400001)

a lot of links for lit reviews and examples online + tester documentation