Chilling Effects

[Proposal for CSE 534 Project]

Zhi Chai
Department of Computer Science
Stony Brook University
SBU ID: 109953687

Xiaowei Liu
Department of Computer Science
Stony Brook University
SBU ID: 109972354

ABSTRACT
In our project, we are trying to find useful information from the chilling effects online database. First, we analyze the data and brainstorm what we can get. Then we perform the data mining and analysis based on the chilling effects online database to find patterns and answer our questions. Finally, we will visualize and present our results.

Keywords
Chilling effects, patterns

1. INTRODUCTION
A chilling effect is the inhibition or discouragement of the legitimate exercise of natural and legal rights by the threat of legal sanction.

In the online world, the potentially infringing activities of individuals are stored and transmitted through the networks of third parties. Web site hosting services, Internet service providers, and search engines that link to materials on the Web are just some of the service providers that transmit materials created by others. Section 512 of the Digital Millennium Copyright Act (DMCA) protects online service providers (OSPs) from liability for information posted or transmitted by subscribers if they quickly remove or disable access to material identified in a copyright holder’s complaint.

The Chilling Effects database in "chillingeffects.org" collects and analyzes legal complaints and requests for removal of online materials, helping Internet users to know their rights and understand the law. These data enable us to study the prevalence of legal threats and let Internet users see the source of content removals.

The current data in "chillingeffects.org" is mainly about the complaining of infringement towards certain DMCA, Trademark, Defamation, Court Order, Law enforcement Request, Private information and so on.

2. OVERVIEW OF THE RELATED WORK
The data in chillingeffect.org is organized in Json. JSON, or JavaScript Object Notation, is an open standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs. Some example of the chilling effects data is like the following: "https://www.chillingeffects.org/notices/2000000.json"

There are some current API that can help us get the data and searching for the database in "chillingeffect.org" and some online Json validator which can help us view Json data: "http://pro.jsonlint.com/

3. PROJECT PLAN
We are going to use the API to get the data from Chillingeffect.org first. Then we will choose the right method to analyze the data, to look for patterns in DMCA and other reports. And try to answer the questions such as the following:

- Who is most active in submitting reports?
- Are there trends/decreases in reports issued?
- Are there trends that seem to indicate abuse of the reporting mechanism?
- The comparison of the hot or sensitive of the topic, what is the most sensitive area? Or which area has the most complaints?
- Are there any trend among the take down letters sent by the government, are these kind of notices targeted?
- Is there exists a model which can describe and predict the number of the complaints a certain company received by time?
- Does the percentage of actions actually taken has any trend associated with certain kind of notices towards certain organizations?

We are going to follow these steps:

- Getting the data from the chillingeffect.org using the current API from Github
- Analyze the data and brainstorm what we can get from the data itself, think of a few questions need to be solved
• Pick up the appropriate methods and model to analyze the data

• Build the website to present what we get and use proper visualization tools

We decide to use Java and python for parsing the data, current related library like org.json, json-lib and json in python.

As for the visualization tool, we plan to use Gnuplot or some other similar tools.

Some of the important data mining skills and methods we want to apply such as data clustering, Affinity Grouping and association analysis etc. For example, firstly, to answer the question "who is most active in submitting reports?", we need to find all the sender name with their numbers of reports and then make a comparison. To answer the question "Are there trends/decreases in reports issued?", we are going to analyze the association between the report and their report time, topic field, language and so on. We hope to discover more interesting fact through brainstorm and creative thinking. Currently, we had got about 2000000 reports in JSON.

4. TIME LINE

Before Feb 16 Discuss the project ideas and plans

Feb 16 Submitting project proposal

Feb 17–Feb 26 Preprocessing

• Build the website which is used to keep update our project process and also for our final presentation

• Learn how to use the current API and get the data from the chillingeffect.org

• Get the data and format the data

Feb 26–April 5 Research and coding

• Analyze the data and try to find the patterns of the DMCA and other report

• Brainstorm what we can get from the data and according to the questions, think of appropriate analysis methods.

• Perform the analysis and the data mining of the data

April 6 Midterm report

• The midterm report will focus on the data mining methods and analysis methods that we applied

April 7–May 9 Present our result using proper visualization tools

May 10 Final report

5. FURTHER READING

DMCA
http://www.dmca.com/

Chilling effects blog
https://www.chillingeffects.org/blog_entries

Google Transparency Report
http://www.google.com/transparencyreport/removals/copyright/