**Ad Blog Block**

The Blocking Program against the Advertising Blogs



Team name: Lemon Grass

Team Members: SeungYun Yeom, JaeHyung Jung, JiHyuk Choi

**INDEX**

* **Introduction**
* **Product**
* **Similar programs**
* **Idea of Algorithms**
* **Roll of Each Members**
* **Rough Schedule**
* **Summary**

**Introduction:**

Blog is originally for doing memo about personal thoughts or information and sharing them with others. There are blogs which are handled by the film such as ‘Naver’. Because it is related to its searching portal, there is a clear advantage that the blogs can be disposed from web users easily. However, there are the users who abuse this advantage. With paid fee from advertisement film, the users advertise the information about certain products. The problem is that the information is not honest reviews but intended stories for personal profit from advertisers. There is no difference between reading blogs and advertisement sites.

**Product:**

The goal of our team is implementing the program which can output the certain percentage of likeliness if the content of blog is an advertisement or not.

In ideal, it will be the best if the program will determine that it is an advertisement or not. However, the unperfect classification can block useful content from user too. So, the product will show the percentage of advertisement-likeness. For practical use, the program will be implemented as the Chrome extension or something other one for easy accessibility for user.

ADV Percentage: 70%

Advertising contents

**Similar programs:**

1. messaging filtering system (Naver): writing the keywords to filter from received messages. And messages including certain words will be automatically classified as advertisement messages. This is not filtering the advertisement blog itself.
2. Blocking ‘AdPost’ (Naver): ‘AdPost’ is the advertisement on where the contents of post ended.



One can block the advertisement but can’t handle the advertisement content itself.

**Idea of Algorithms:**

There are several ways to recognize that it’s an advertisement or not.

1. The First one is catching key points from the data of blogs with web scraping. For example, there is a famous blog-advertising film ‘Weble’. And the blogs, which is influenced by Weble, remain some hints such as the fixed images brought from Weble like “This content is supported from Weble”.
2. The second one is making ‘advertisement words dictionary’ and use it on filtering the blog. The ‘advertisement words dictionary’ is a collection of words which are usually written on advertising blogs. The advantage of this algorithm from the user’s view is that it will be an integrated filtering system. Unlike the message filtering system (Naver) on ‘Similar Programs’ part, the users don’t need to search and update advertisement-like word list.
3. The third one, which is challenging and reinforced or discarded after experiments, is assuming that there are certain and unique patterns of positive/negative words. Because advertisement is appealing only good aspects of products in limited resources. So, there will be experiments if there are the patterns only found from advertisement blogs. The link below is already implemented positive/negative words dictionary. And this will be adjusted and reinforced for this project.

The link about positive/negative words. <http://web.yonsei.ac.kr/dslab/Journal/sentiment%20dictionary.pdf>

**Roll of Each Members**

|  |  |
| --- | --- |
| **SeungYun**  **Yeom** | * Implementing Naver blog parsing module * Structing Dataset for Algorithm * Implementing Algorithm |
| **JaeHyung**  **Jung** | * Data searching * Implementing Algorithm |
| **JiHyuk**  **Choi** | * Chrome extension Design and Implementing * Implementing Algorithm |

**Rough Schedule**

Yeom Jung Choi \_

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| week | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Parsing  Module |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Algorithm  Experiments |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Dataset |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Chrome  Extension |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Collecting  Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Summary:**

The program is blocking advertisement blog. For handy usage, it will be implemented to chrome extension. There are several ideas of the algorithm which calculates and shows the percentage of advertisement-likeness when the user accesses to Naver blog. The team members will search and device more ideas about algorithms and undergo experiments.