5-9-2018

Extending Safe Search Functionality for Identifying Child-Safe and Educational Web Resources

Oghenemaro Anuyah  
*Boise State University*

Ion Madrazo Azpiazu  
*Boise State University*

Maria Soledad Pera  
*Boise State University*
Motivation

- Children prefer to utilize popular search engines for their educational and leisure searches. However, they may have access to information deemed unsafe for them.
- Even if popular and children-oriented search engines make available Safe Search, a strategy meant to prevent children from accessing inappropriate content such as pornography or hate-speech, Safe Search may be strict in filtering educational resources.

Goal

1. Investigate the functionality of Safe Search filters available on a number of search engines designed for diverse users, as well as for children.
2. Introduce KiseRF, a filtering strategy that addresses some of the limitations identified with traditional Safe Search filtering strategies.

Research Questions:
1. Are existing Safe Search filters too restrictive when it comes to retrieving resources that are valid in an educational context?
2. Do traditional Safe Search filters effectively identify web resources with sexually explicit content?
3. Are traditional Safe Search filters effective in disregarding web resources that potentially promote violence?

Related Work

- DMOZ open directory project - Kids Safe Web Resources.
- Alexa - Adult Content.
- Idaho Digital Learning Academy - Educational Web Resources.
- Google’s Bad word list - Collection of Bad Keywords.
- Hatebase.org: Hate Speech Lexicons.

References


Limitations of Traditional Safe Search Filters

- DMOZ open directory project - Kids Safe Web Resources.
- Alexa - Adult Content.
- Idaho Digital Learning Academy - Educational Web Resources.
- Google’s Bad word list - Collection of Bad Keywords.
- Hatebase.org: Hate Speech Lexicons.

Data Sources

- Baseline Safe Search strategies:
  - Popular search engines: Google and Bing.
  - Children-oriented search engines: Kidsearch and Kidrex.
- Gold Standard:
  - Cyren Web URL categorizer.
- Classification Algorithm:
  - Random Forests - Achieved high performance than logistic regression, GBC and MLP classifiers.
- Dataset:
  - 80:20 Train and Test Split ratio.
- Queries:
  - Hate speech, Educational and Sexually-explicit keywords.

Initial Assessment

- Some educational resources were disregarded by the Safe Search filters.
- Children oriented search engines were particularly strict in disregarding sexually explicit content.
- Results show that there is a need to improve existing Safe Search filtering strategies.

In the Future:

- Conduct more exhaustive evaluation.
- Propose novel features for retaining educational web resources.