Ciobănița Andreea Sorana

Filtrarea conținutului la nivel de client pe baza interesului utilizatorilor

Cuprins

[1. Introducere 3](#_Toc69326113)

[1.1. Prezentare generală 3](#_Toc69326114)

[1.2. Motivație 3](#_Toc69326115)

[2. Fundamente Teoretice 5](#_Toc69326116)

[2.1. Filtrarea conținutului web la nivel de client 5](#_Toc69326117)

[2.2. Soluții existente 5](#_Toc69326118)

[2.3. Rezultate teoretice 5](#_Toc69326119)

[3. Soluția propusă 6](#_Toc69326120)

[4. Evaluarea rezultatelor 7](#_Toc69326121)

[5. Concluzii 8](#_Toc69326122)

[Bibliografie 9](#_Toc69326123)

# Introducere

## Prezentare generală

Filtrarea conținutului web este reprezentată de tehnicile de blocare a afișării diferitor elemente din paginile web, în funcție de diferite criterii.

În primul rând, necesitatea unei filtrări la nivelul conținutului paginilor web a apărut odată cu facilitarea accesului la internet, deoarece persoane de toate vârstele au învățat să îl folosească în viața de zi cu zi. Din această cauză, a apărut nevoia de implementare a unor metode ce să poată oferi un control parental, ajutând astfel părinții să își protejeze copiii împotriva conținutului nepotrivit ce poate fi foarte ușor accesibil pe internet.

O altă utilitate a filtrării este folosirea acesteia în instituții și companii, pentru a limita site-urile la care un utilizator să aibă acces în cadrul rețelei interne de internet. Acest lucru ajută, de exemplu, în rețelele folosite în școli pentru a împiedica elevii să acceseze site-uri cu conținut nepotrivit sau diferite platforme care ar putea afecta calculatoarele instituției prin instalarea unor viruși. De asemenea, companiile pot folosi un sistem de filtrare pentru a împiedica angajații să petreacă prea mult timp navigând pe diferite pagini non-productive, astfel fiind posibilă o creștere a eficienței și a productivității în timpul programului de lucru.

De asemenea, o altă utilitate importantă este posibilitatea de a filtra conținutul pe baza intereselor utilizatorilor. Datorită creșterii foarte mari din ultimii ani a internetului, au apărut foarte multe informații ușor accesibile, însă uneori când căutăm anumite lucruri de care suntem interesați primim și foarte multe reclame care de cele mai multe ori acoperă conținutul paginilor. Uneori ceea ce primim ca și reclamă ar putea fi ceva de interes pentru noi, mai ales că în ultima perioadă și sistemele de recomandare și afișare a reclamelor s-au dezvoltat, fiind capabile să prezinte cât mai mult lucruri din aria de interes a utilizatorilor, însă de cele mai multe ori când navigăm pe internet am dori să ne putem concentra atenția pe lucrurile pe care le căutăm în momentul respectiv. Din acest motiv s-au dezvoltat diferite aplicații software de tip ad-blocker pentru a ascunde cât mai eficient reclamele nedorite ce apar în timpul navigării pe internet.

## Motivație

Deși ad-blocker-ele au adus un mare avantaj în navigarea pe internet, acestea nu reușesc întotdeauna să ascundă toate reclamele care apar. De asemenea, filtrarea conținutului pe baza intereselor utilizatorilor poate fi extinsă să însemne mai mult decât ascunderea unor reclame, există situații în care ne putem dori să ascundem dintr-o anumită pagină și altfel de elemente, care nu au cum să fie incluse în cele ce sunt filtrate de un ad-blocker.

De exemplu, daca navigăm pe un site de știri și suntem interesați de cele dintr-un anumit domeniu, ar fi mai ușor dacă am avea posibilitatea ca în momentul în care găsim o știre din afara domeniului nostru de interes să avem opțiunea de a o ascunde din pagină. Astfel, în final putem vizualiza mult mai ușor ceea ce ne interesează de fapt, fără să ne fie distrasă atenția cu alte lucruri. Un alt exemplu din această categorie ar putea fi diferite animații, gif-uri care se mișcă în continuu undeva pe pagina pe care o vizualizăm și am dori să putem să le ascundem.

În acest sens, propunem o extensie Google Chrome care să ne ofere posibilitatea de a selecta și apoi a ascunde elementele pe care nu dorim să le vedem afișate în paginile pe care navigăm. Odată instalată în browser, extensia va adăuga o opțiune numită ”*Remove Element*” în meniul afișat în momentul în care dăm click dreapta pe un element din pagină, iar aceasta va ascunde elementul selectat. Mai târziu, în fereastra extensiei va fi afișată o listă cu elementele care au fost ascunse și va exista posibilitatea ca acestea să fie scoase din listă și astfel să fie afișate din nou în pagină la locul lor. De asemenea, ne propunem să găsim similarități între elementele pe care le ascunde un utilizator și după câteva utilizări să putem oferi sugestii în legătură cu posibile elemente ce nu par a fi de interes pentru acesta și ar putea fi ascunse din paginile pe care navighează.

# Fundamente Teoretice

## Filtrarea conținutului web la nivel de client

## Soluții existente

## Rezultate teoretice

# Soluția propusă

# Evaluarea rezultatelor

# Concluzii

# Bibliografie

|  |  |
| --- | --- |
| [1] | E. Bertino, E. Ferrari și A. Perego, „A general framework for web content filtering,” *World Wide Web,* 2010. |
| [2] | W. Ming, L. Rossi, Y. Li și J. Kuo, „Automatic Web Filtering Approach Based on Multimodal Content Information,” *Internet Multimedia Management System II,* 2001. |
| [3] | C. Ding, C.-H. Chi, J. Deng și C.-L. Dong, „Centralized content-based web filtering and blocking: How far can it go?”. |
| [4] | N. Churcharoenkrung, Y. S. Kim și B. H. Kang, „Dynamic web content filtering based on user's knowledge,” *Proceedings of the International Conference on Information Technology: Coding and Computing.* |
| [5] | Y. Lai, Q. Ma, Z. Yang și J. Liu, „Framework of Web Content Filtering for IPv6”. |
| [6] | P. Y. Lee, S. C. Hui și A. C. M. Fong, „Neural Networks for web content filtering”. |
| [7] | D. D. Nguyen, M. Erdmann, T. Takeyoshi și G. Hattori, „Training multiple support vector machines for personalized web content filters,” *IEICE Trans. Information & Systems,* 2013. |
| [8] | N. Narwal și S. K. Sharma, „Web informative content identification and filtering using machine learning technique,” *Data Analysis Techniques and Strategies,* 2016. |
| [9] | A. C. M. Fong, S. C. Hui și P. Y. Lee, „XFighter: An intelligent web content filtering system,” *Kybernetes,* 2016. |
| [10] | K. S. Kuppusamy și G. Aghila, „A personalized web page content filtering model based on segmentation,” *International Journal of Information Sciences and Techniques,* 2012. |
| [11] | M. Chau și H. Chen, „A machine learning approach to web page filtering using content and structure analysis,” 2007. |
| [12] | T. Sandholm și H. Ung, „Real-time, location-aware collaborative filtering of web content,” 2011. |
| [13] | M. Hammami, Y. Chahir și L. Chen, „WebGuard: a web filtering engine combining textual, structural and visual content-based analysis,” *IEEE Transactions on Knowledge and Data Engineering,* 2006. |
| [14] | P. Y. Lee, S. C. Hui și A. Fong, „A structural and content-based analysis for Web filtering,” *Internet Research: Electronic Networking Applications and Policy,* 2003. |
| [15] | E. Bertino, E. Ferrari, A. Perego și G. P. Zarri, „An Integrated Approach to Rating and Filtering Web Content,” 2005. |
| [16] | G. Sartor și A. Loreggia, „The impact of algorithms for online content filtering or moderation,” 2020. |
| [17] | E. Bertino, E. Ferrari și A. Perego, „Content-based filtering of web documents: the MaX system and the EUFORBIA project,” 2003. |
| [18] | L. Yan și Y. Shenghua, „Content filtering research based on web community structure,” *International Conference of Mechatronic Sciences, Electric Engineering and Computer (MEC),* 2013. |
| [19] | S.-J. Ko, „Predicting Typical User Preferences Using Entropy in Content Based Collaborative Filtering System,” 2004. |
| [20] | W. Paireekreng, „Mobile content recommendation system for re-visiting user using content-based filtering and client-side user profile,” *International Conference of Machine Learning and Cybernetics,* 2013. |