ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Optimal Page Downloading By Using Decision Making

Deepika P. Pachpute, Prof. Vina M. Lomte.

M.E., Dept. of Computer, RMD Sinhgad School of Engineering, Pune, India

Head of Department, Dept. of Computer, RMD Sinhgad School of Engineering, Pune, India

ABSTRACT: The use of ALT (Advanced Learning Technologies) creates dynamic sharing and exchanging between open source communities that diffuse e-learning systems. Crawler is designed to traverse the web to gather documents on a specific topic in E-learning. One of the most significant challenges that e-Learningface today is choosing Perfect pagethat offer them the features they want and the ease of use they need. Crawler is designed for developing new web pages on the Web as well as for refreshing the content of already downloaded pages e-learning. Page download request is frequently send to server. These requests, in turn, increase the energy consumption of the servers as hardware resources are used when serving the requested pages. In this process, the crawler issues a huge amount of HTTP requests to web servers. These requests increase the energy consumption of the web servers since computational resources are used while serving the requests. In this work, we introduce the problem of green web crawling, where the objective is to devise a page refresh policy that minimizes the total staleness of pages in the repository of a web crawler, subject to a constraint on the amount of carbon emissions due to the processing on web servers. For the case of one web server and one crawling thread, the optimal policy turns out to be a greedy one. At each page request webpage the page is refreshed with the page's staleness, its size, and the greenness of the energy consumed at the web server premises.

KEYWORDS: Advanced learning Technologies, Crawling, carbon footprint, greenness, staleness, Web Dynamics

I. INTRODUCTION

A large number of research works has investigated the techniques for reducing the query processing workload of a search engine, thus achieving reduction in energy consumption. To best of our knowledge, however, no work has investigated the issues related to the energy consumption associated with the web crawling component. problem of green web crawling, where the objective is to devise a page refresh policy that minimizes the total staleness of pages in the repository of a web crawler, subject to a constraint on the amount of carbon emissions due to the processing on web servers. Here present devise a web repository refreshing technique that takes into accounts both the greenness and staleness concepts when scheduling the download of e-learning web pages. This technique aims to reduce the total staleness of pages in the web repository while constraining web servers' total carbon footprint resulting from the activities of the crawler. We evaluate the performance of our technique using a large, real-life web server data set. Beyond creating an environment-friendly crawler, our work has implications for large-scale web search engines, which should comply with regulations about carbon footprint reduction.

II. REVIEW OF LITERATURE

1) This paper proposed a system that makes efficient use of communication resources in approximate replication environments. Advantage is when data changes rapidly, good performance. Disadvantage is that Itdoesn't ensure exact consistency [4].
2) It proposed dynamic placement of virtual machines (VMs) with deterministic and stochastic demands. In order to ensure a quick response to VM requests and improve the energy efficiency, a two-phase optimization strategy has been proposed, in which VMs are deployed in runtime and consolidated into servers periodically. Its advantage is that it gives quick response to VMrequests, but Deadlock may occur in the actual migration process [2].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

3) This paper proposed a general, optimization-based framework to minimize datacenter costs in the presence of different carbon footprint reduction goals, renewable energy characteristics, policies, utility tariff, and energy storage devices (ESDs).Advantage is on-site renewable can help lower costs due to their ability to reduce the peak datacenter power draw from the utility and disadvantage is On-site and off-site these hybrid combination is the most cost-effective across the spectrum [1].
4) Proposed various refresh policies and studies their effectiveness. We first formalize the notion of "freshness" of copied data by defining two freshness metrics, and we propose a Poisson process as the change model of data sources. Based on this framework, we examine the effectiveness of the proposed refresh policies analytically an Experimentally. Advantage is proposed refresh policies improve the "freshness" of data very significantly.After drawback is it increases local data storage [3].
5]Web search engines (e.g. Google, Yahoo, Microsoft Live Search, etc.) are widely used to find certain data among a huge amount of information in a minimal amount of time. These useful tools also pose a privacy threat to the users. Web search engines profile their users on the basis of past searches submitted by them. In the proposed system, we can implement the String Similarity Match Algorithm (SSM Algorithm) for improving the better search quality results. To address this privacy threat, current solutions propose new mechanisms that introduce a high cost in terms of computation and communication. Personalized search is promising way to improve the accuracy of web search. However, effective personalized search requires collecting and aggregating user information, which often raises serious concerns of privacy infringement for many users[10].
6] The e-Learning has become matured learning paradigm with the advent of web based learning and content management tools, and shifted the focus of entire world from instructor centric learning paradigm to learner centric approach. Now for making the learning process more streamlined and standardized, the implementing agencies are emphasizing on moving towards service oriented architectural design approach to create, deploy and manage reusable e-Learning services, thus benefiting education sector. For providing the intelligence to evaluation system and other e-Learning services, various domains like data mining, web mining, semantic web etc. can be utilized intelligently. In this paper, we will describe an approach aiming to achieve personalization in e-Learning services using web mining and semantic web[5].
7] As brown energy costs grow, renewable energy becomes more widely used. Previous work focused on using immediately available green energy to supplement the non-renewable, or brown energy at the cost of canceling and rescheduling jobs whenever the green energy availability is too low. In this paper we design an adaptive data center job scheduler which utilizes short term prediction of solar and wind energy production. This enables us to scale the number of jobs to the expected energy availability, thus reducing the number of cancelled jobs by 4x and improving green energy usage efficiency by 3x over just utilizing the immediately available green energy[9].
8]Personalized web search (PWS) has demonstrated its effectiveness in improving the quality of various search services on the Internet. However, evidences show that users' reluctance to disclose their private information during search has become a major barrier for the wide proliferation of PWS. We study privacy protection in PWS applications that model user preferences as hierarchical user profiles. We propose a PWS framework called UPS that can adaptively generalize profiles by queries while respecting user specified privacy requirements. Our runtime generalization aims at striking a balance between two predictive metrics that evaluate the utility of personalization and the privacy risk of exposing the generalized profile. We present two greedy algorithms, namely GreedyDP and GreedyIL, for runtime generalization. We also provide an online prediction mechanism for deciding whether personalizing a queryis beneficial. Extensive experiments demonstrate the effectiveness of our framework [6].
9]Data mining is used for finding the useful information from the large amount of data. Data mining techniques are used to implement and solve different types of research problems. The research related areas in data mining are text mining, web mining, image mining, sequential pattern mining, spatial mining, medical mining, multimedia mining, structure mining and graph mining. This paper discussed about the text mining and its preprocessing techniques. Text mining is the process of mining the useful information from the text documents. It is also called knowledge discovery in text (KDT) or knowledge of intelligent text analysis. Text mining is a technique which extracts information from both structured and unstructured data and also finding patterns. Text mining techniques are used in various types of research domains like natural language processing, information retrieval, text classification and text clustering [8]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

10]It work on
1.Locating sources of web content.
2.Selection of relevant sources.
3.Extracting the underlying content of deep web pages. Here is the problem of retrieving unwanted pages which needs more time to crawl relevant results [1].

III. EXISTING SYSTEM

User not getting perfect solution on his query in e-learning.It is hard to find out how long it will take for crawler to notice changes you've made to your e-learning Pages. It takes too much time. Thefrequencies at which you publish new content on your website.New updated pages were not found. A web crawler can do to reduce the energy consumption it incurs to web servers without sacrificing the coverage or freshness of its web repository. This is because the amount of energy consumed on web servers depends only on factors related to the hardware and software resources that are not managed by the search engine company. Nevertheless, certain optimizations can be employed to reduce the carbon emissions that a crawler incurs to web servers.

DISADVANTAGES

1) User not getting perfect solution on his query in e-learning. New updated pages were not found.
2) Increase the energy consumption and carbon footprint of the web servers since computational resources are used while serving the requests.

IV.TECHNIQUES

No.	Name	Techniques
1	Geographic trough filling for internet datacenters, in Proc. INFOCOM, 2012, pp. 2881–2885, D. Xu and X. Liu,	1) Queue-based trough filling algorithm, called QTF
2	Web Crawling	1) Batch crawling. 2) Incremental crawling.
3	Supporting Privacy Protection in Personalized Web Search, Year :2012	1)GreedyDP 2)GreedyIL , for runtime generalization.
4	Predicting Content Change on the Web	1D, 2D and 3D algorithms, Content Similarity KNN
5)	Effective Page Refresh Policies for Web Crawlers, ACM Transactions on Database Systems, Vol. 28, No. 4, December 2003	1) Poisson Process and Probabilistic Evolution of an Element
6)	Preprocessing Techniques for Text Mining - An Overview,Year-2016	Natural Language Processing
7	Personalization of e-Learning Services using Web Mining and Semantic Web, Sandesh Jain, Dhanander	*Semantic Web Languages, Web Mining*

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

	K. Jain, HariharBhojak, AnkitBhilwar, and Mamatha J. September 10, 2012.	
8	Supporting Privacy Protection in Personalized Web Search with Secured User Profile , Archana Ukande1, Nitin Shivale2	String Similarity Match Algorithm
9	Carbon-Aware Energy Capacity Planning for Datacenters ChuangangRen, Di Wang, BhuvanUrgaonkar, and AnandSivasubr amaniam	On-site Renewable Generation Off-site Renewable Generation
10	Utilizing Green Energy Prediction to Schedule Mixed Batch and Service Jobs in Data Centers BarisAksanli, JagannathanVenkatesh, Liuyi Zhang, TajanaRosingOctomber 23,2011	job scheduling methodology 1Green energy scheduling and data center modelin- 1.1 Predictive green energy schedule 1.2 Instantaneous green energy scheduler

V.CONCLUSION

This system solves the problem of sending large no. of requests to particular server. Discovering updated content is possible that increases greenness of server and decreases staleness of web pages. It must be very careful in deciding what data to check for freshness. It showed that these optimal policies can improve freshness very significantly through real Web data. These optimal policies consider how often a page changes and how important the pages are, and make an appropriate refresh decision.

VI. FUTURE WORK

Future work will focus on the implementation Of Green crawler for reducing more carbon emission by designing different policies.This proposed system developed only for e-learningin future it can be developed for another application. Also focus on storage space reduction.

REFERENCES

1) Predicting Content Change on the Web, 2013 ACM 978-1-4503-1869-3/13/02, KiraRadinsky, Paul N. Bennett
2) Carbon-Aware Energy Capacity Planning for Datacenters, 2012,ChuangangRen, Di Wang, BhuvanUrgaonkar, and AnandSivasubramaniam.
3) Effective page refresh policies for web crawlers, J. Cho and H. Garcia-Molina ACM Trans. Database Syst., vol. 28, no. 4, pp. 390–426, Dec. 2003.
4)Geographic trough filling for internet datacenters, in Proc. INFOCOM, 2012, pp. 2881–2885, D. Xu and X. Liu,
5] Personalization of e-Learning Services using Web Mining and Semantic Web Sandesh Jain, Dhanander K. Jain, HariharBhojak, AnkitBhilwar, and Mamatha J.
6]Supporting Privacy Protection in Personalized Web Search LidanShou, He Bai, Ke Chen, and Gang Chen,Year 2012

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

7] Web Crawling, Foundations and Trends in Information Retrieval, vol. 4, No. 3, pp. 175–246, 2010. Olston and M. Najork,
8] Preprocessing Techniques for Text Mining - An Overview Dr. S. Vijayarani, Ms. J. Ilamathi, Ms. Nithya Assistant Professor, M. Phil Research Scholar,Year-2016
9] Utilizing Green Energy Prediction to Schedule Mixed Batch and Service Jobs in Data Centers BarisAksanli, JagannathanVenkatesh, Liuyi Zhang, TajanaRosingOctomber 23,2011
10] Supporting Privacy Protection in Personalized Web Search with Secured User Profile ,Archana Ukande1, Nitin Shivale2 Volume 4 Issue 6, June 2015