London Journals Press

IMAGE: OBSERVATORY WITH STAR TRAILS ON MOUNTAINS FOR CLEAR SKY

COMPILED IN UNITED KINGDOM

ENGLISH

They were leaders in building the early foundation of modern programming and unveiled the structure of DNA Their work inspired environmental movements and led to the discovery of new genes They've gone to space and back taught us about the natural world dug up the earth and discovered the origins of our species They broke the sound barrier and gender barriers along the way The world of research wouldn't be the same without the pioneering efforts of famous research works made by these women Be inspired by these explorers and early adopters- the women in research who helped to shape our society We invite you to sit with their stories and enter new areas of understanding This list is by no means a complete record of women to whom we are indebted for their research work but here are of history's greatest research contributions made by...

Read complete here:
https://goo.gl/1vQ3lS

Women In Research

Computing in the cloud!

Cloud Computing is computing as a Service and not just as a Product Under Cloud Computing...

Read complete here:
https://goo.gl/VvHC72

Writing great research...

Prepare yourself before you start Before you start writing your paper or you start reading other...

Read complete here:
https://goo.gl/np73jP

Journal Content
In this Issue

London
Journals Press

Editorial Board

Curated board members

Dr. Saad Subair

College of Computer and Information Sciences,
Alazaeim Alazhari University, Khartoum North,
Sudan, Associate Professor of Computer Science
and Information Ph.D., Computer Science
Bioinformatics, University of Technology
Malasiya

Gerhard X Ritter

Emeritus Professor, Department of Mathematics,
Dept. of Computer & Information,
Science & Engineering Ph.D.,
University of Wisconsin-Madison, USA

Dr. Ikvinderpal Singh

Assistant Professor, P.G. Deptt. of Computer
Science & Applications, Trai Shatabdi GGS
Khalsa College, India

Prof. Sergey A. Lupin

National Research,
University of Electronic Technology Ph.D.,
National Research University of Electronic
Technology, Russia

Dr. Sharif H. Zein

School of Engineering,
Faculty of Science and Engineering,
University of Hull, UK Ph.D.,
Chemical Engineering Universiti Sains Malaysia,
Malaysia

Prof. Hamdaoui Oualid

University of Annaba, Algeria Ph.D.,
Environmental Engineering,
University of Annaba,
University of Savoie, France

Prof. Wen Qin

Department of Mechanical Engineering,
Research Associate, University of Saskatchewan,
Canada Ph.D., Materials Science,
Central South University, China

Luisa Molari

Professor of Structural Mechanics Architecture,
University of Bologna,
Department of Civil Engineering, Chemical,
Environmental and Materials, PhD in Structural
Mechanics, University of Bologna.

Prof. Chi-Min Shu

National Yunlin University of Science
and Technology, Chinese Taipei Ph.D.,
Department of Chemical Engineering University of
Missouri-Rolla (UMR) USA

Prof. Te-Hua Fang

Department of Mechanical Engineering,
National Kaohsiung University of Applied Sciences,
Chinese Taipei Ph.D., Department of Mechanical
Engineering, National Cheng Kung University,
Chinese Taipei

Dr. Fawad Inam

Faculty of Engineering and Environment,
Director of Mechanical Engineering,
Northumbria University, Newcastle upon Tyne,
UK, Ph.D., Queen Mary, University of London,
London, UK

Dr. Rocío Maceiras

Associate Professor for Integrated Science,
Defense University Center, Spain Ph.D., Chemical
Engineering, University of Vigo, SPAIN

Muhammad Hassan Raza

Postdoctoral Fellow, Department of Engineering
Mathematics and Internetworking,
Ph.D. in Internetworking Engineering,
Dalhousie University, Halifax Nova Scotia,
Canada

Rolando Salgado Estrada

Assistant Professor,
Faculty of Engineering, Campus of Veracruz,
Civil Engineering Department, Ph D.,
Degree, University of Minho, Portugal

Abbas Moustafa

Department of Civil Engineering,
Associate Professor, Minia University, Egypt, Ph.D
Earthquake Engineering and Structural Safety,
Indian Institute of Science

Dr. Babar shah

Ph.D., Wireless and Mobile Networks,
Department of Informatics,
Gyeongsang National University,
South Korea

Dr. Wael Salah

Faculty of Engineering,
Multimedia University Jalan Multimedia,
Cyberjaya, Selangor, Malaysia, Ph.D, Electrical and
Electronic Engineering, Power Electronics
and Devices, University Sians Malaysia

Prof. Baoping Cai

Associate Professor,
China University of Petroleum,
Ph.D Mechanical and Electronic Engineering,
China

Prof. Zengchang Qin

Beijing University of Aeronautics
and Astronautics Ph.D.,
University of Bristol,
United Kingdom

Dr. Manolis Vavalis

University of Thessaly,
Associate Professor, Ph.D.,
Numerical Analysis,
University of Thessaloniki, Greece

Dr. Mohammad Reza Shadnam

Canadian Scientific Research and Experimental
Development Manager-Science,
KPMG LLP, Canada, Ph.D., Nanotechnology,
University of Alberta, Canada

Dr. Gang Wang

HeFei University of Technology,
HeFei, China, Ph.D.,
FuDan University, China

Kao-Shing Hwang

Electrical Engineering Dept.,
Nationalsun-Yat-sen University Ph.D.,
Electrical Engineering and Computer Science,
Taiwan

Mu-Chun Su

Electronics Engineering,
National Chiao Tung University, Taiwan,
Ph.D. Degrees in Electrical Engineering,
University of Maryland, College Park

Zoran Gajic

Department of Electrical Engineering,
Rutgers University, New Jersey, USA
Ph.D. Degrees Control Systems,
Rutgers University, United States

Dr. Homero Toral Cruz

Telecommunications,
University of Quintana Roo, Ph.D.,
Telecommunications Center for Research
and Advanced Studies National Polytechnic
Institute, Mexico

Nagy I. Elkalashy

Electrical Engineering Department,
Faculty of Engineering,
Minoufiya University, Egypt

Vitoantonio Bevilacqua

Department of Electrical and Information
Engineering Ph.D., Electrical Engineering
Polytechnic of Bari, Italy

Dr. Sudarshan R. Nelatury

Pennsylvania State University USA Ph.D., Signal
Processing Department of Electronics and
Communications Engineering,
Osmania University, India

Prof. Qingjun Liu

Professor, Zhejiang University, Ph.D.,
Biomedical Engineering,
Zhejiang University, China

Assessment of the Sewage Coverage Index in the World from 2000 to 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

Vinícius Chaves Campos & Joel Carlos Zukowski Junior

ABSTRACT

There are millions of people who do not benefit from sanitary sewage systems in Brazil and in the world. The absence of sanitary sewage systems increases the number of people affected by waterborne diseases and consequently their quality of life. The present research used the bibliographic research as a technical procedure, obtaining secondary data on sewage coverage (%) extracted from the database of The International Benchmarking Network (IBNET), for the years 2000, 2010, 2015 and the data more actions made available by the concessionaires of the countries to the platform. This article will allow people and companies that work with sanitation to know the reality of sewage coverage in their countries and with that they can seek to work with a process of continuous improvement. In relation to the most current data provided by 101 countries, 30.69% of the countries have sewage coverage with sewage coverage equal to or less than 40%.This article is a mechanism that allows sanitation companies to direct investments to countries that really need it, in order to contribute to improvements in global public health, with millions of people benefiting from a better quality of life.

Keywords: sewage coverage. indicator. countries. health. quality of life.

Classification: DDC Code: 363.728 LCC Code: TD730

Language: English

London Journal of Engineering Research

Assessment of the Sewage Coverage Index in the World from 2000 to 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

Avaliação Do Índice De Cobertura De Esgoto No Mundo No Período De 2000 a 2020: Situação Do Indicador Nos Países E Import Ncia Do Monitoramento Desta Informação

Vinícius Chaves Campos[α] & Joel Carlos Zukowski Junior[σ]

RESUMO

Existem milhões de pessoas que não são beneficiadas com sistemas de esgotamento sanitário no Brasil e no mundo. A ausência de sistemas de esgotamento sanitário, aumenta o número de pessoas afetadas por doenças de veiculação hídrica e consequentemente a qualidade de vida dessas. A presente pesquisa utilizou como procedimento técnico a pesquisa bibliográfica, com a obtenção de dados secundários de cobertura de esgoto (%) extraídos do banco de dados do The Internacional Benchmarking Network (IBNET), para os anos de 2000, 2010, 2015 e os dados mais atuas disponibilizados pelas concessionárias dos países à plataforma. O presente artigo permitirá que pessoas e empresas que trabalham com saneamento conheçam a realidade de cobertura de esgoto de seus países e com isso possam buscar trabalhar com um processo de melhoria contínua. Em relação aos dados mais atuais disponibilizados por 101 países, 30,69% dos países apresentam cobertura de esgoto com coberturas de esgoto iguais ou menores que 40%. Este artigo é um mecanismo que permite que empresas de saneamento direcionem os investimentos aos países que de fato necessitem, de modo a contribuir para que ocorram melhorias na saúde pública mundial, com milhões de pessoas sendo beneficiadas com uma melhor qualidade de vida.

Palavras-chave: cobertura de esgoto. indicador. países. saúde. qualidade de vida.

Author α: Instituição: universidade federal do tocantins- uft país: brasil, cep 77019-448.

σ: Instituição: universidade federal do tocantins- uft país: brasil, cep 77015-522.

ABSTRACT

There are millions of people who do not benefit from sanitary sewage systems in Brazil and in the world. The absence of sanitary sewage systems increases the number of people affected by waterborne diseases and consequently their quality of life. The present research used the bibliographic research as a technical procedure, obtaining secondary data on sewage coverage (%) extracted from the database of The International Benchmarking Network (IBNET), for the years 2000, 2010, 2015 and the data more actions made available by the concessionaires of the countries to the platform. This article will allow people and companies that work with sanitation to know the reality of sewage coverage in their countries and with that they can seek to work with a process of continuous improvement. In relation to the most current data provided by 101 countries, 30.69% of the countries have sewage coverage with sewage coverage equal to or less than 40%.This article is a mechanism that allows sanitation companies to direct investments to countries that really need it, in order to contribute to improvements in global public health, with millions of people benefiting from a better quality of life.

Keywords: sewage coverage. indicator. countries. health. quality of life.

London Journal of Engineering Research

I. INTRODUÇÃO

O saneamento básico é um setor importante da sociedade e quando implantado e operado corretamente geram benefícios as populações e ao meio ambiente. Com isso, a ausência de saneamento é um dos principais fatores que contribuem para o aumento das doenças, que por sua vez é responsável por grande utilização dos recursos financeiros do setor da saúde. (BILIBIO *et al* 2021)

O direito ao saneamento básico é um direito de todos os cidadãos assegurados por lei, porém no Brasil a universalização desse sistema é uma realidade ainda distante, para milhões de pessoas. (BILIBIO *et al* 2021)

A preocupação com a qualidade da água consumida e a destinação do esgoto foi aumentando no decorrer da história da humanidade, com a evolução dos conhecimentos em relação aos efeitos danosos que a ausência de cuidados com estas estruturas, poderiam causar ao homem e ao meio ambiente. (DE PAULA FERREIRA, GARCIA; 2017)

Segundo a lei nº 11.445 de 2007, um sistema de esgotamento sanitário é composto por atividades, infraestruturas e instalações operacionais de coleta, transporte, tratamento e disposição final adequados dos esgotos sanitários, desde as ligações prediais até a sua destinação final no meio ambiente. (BRASIL, 2007)

Em muitas cidades, uma parte da população é beneficiada com coleta de esgoto onde realizam afastamento dos efluentes, por meio de rede coletoras a uma disposição final em um corpo receptor, sem a realização de qualquer tipo de tratamento. Essa situação proporciona, contaminação do solo, rios, lagos, oceanos e águas subterrâneas. No país ainda existem muitas cidades que que não possuem sistemas de esgotamento sanitário completos, com redes coletoras e tratamento. (DOS SANTOS LISBOA et al. 2019). Destaca-se que um sistema de esgotamento sanitário só é considerado completo se possuir a etapa de tratamento. (DE FUNASA, 2007)

O sistema público de esgoto convencional é constituído por ramal predial, coletor de esgoto, coletor tronco, poços de visita, interceptor, emissário, elevatória e estação de tratamento de esgoto (ETE). Após o efluente ser tratado, os esgotos podem ser destinados ao corpo d'água receptor ou serem aplicados no solo. Em ambos os casos, há que se levar em conta os poluentes eventualmente ainda presentes nos esgotos tratados, com destaque, para organismos patogênicos e metais pesados (DE FUNASA, 2007). A definição dos componentes a serem implantados em cada sistema de esgotamento sanitário são determinados de acordo com as características da região, como por exemplo, características do efluente do local (vazão, carga, etc.), clima, disponibilidade de áreas, etc.

A operação e a manutenção de um sistema de esgotamento sanitário englobam as atividades necessárias para o funcionamento de todas as suas operações unitárias, com objetivo de garantir a sua eficiência, eficácia e sustentabilidade. (DE FUNASA, 2017)

A viabilidade e a sustentabilidade do sistema de esgotamento sanitário dependerão dos recursos financeiros disponíveis, seja estes oriundos de tarifas ou taxas, seja por outras receitas do titular ou do operador do serviço, desde que suficientes para cobrirem as despesas de operação e manutenção. Estes custos devem ser estimados considerando mão de obra, energia elétrica, produtos químicos, combustíveis, materiais e equipamentos para manutenção, entre outros. (DE FUNASA, 2017)

No Brasil, a baixa cobertura dos serviços de saneamento implica desdobramentos nefastos para a saúde da população, principalmente as de baixa renda, além de proporcionarem consequências negativas ao meio ambiente, considerando as inúmeras possibilidades de agravo ao equilíbrio dos ecossistemas. (BOVOLATO; 2010)

A ausência na coleta e tratamento do esgoto doméstico contribui para a contaminação do meio e aumento da proliferação de doenças, de modo a

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

2 Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

interferir na qualidade de vida das pessoas. VAZ (2017).

A presente pesquisa utilizou como procedimento técnico a pesquisa bibliográfica, com a obtenção de dados secundários de cobertura de esgoto (%) extraídos do banco de dados do The Internacional Benchmarking Network (IBNET), para os anos de 2000, 2010, 2015 e os dados mais atuas disponibilizados no perfil dos países na plataforma.

A pesquisa é importante, para mostrar a situação que se encontra a cobertura de esgoto (%) no mundo. O primeiro passo para resolver um problema é compreender este e suas condições. Com isso este artigo, permitirá que vários países conheçam um problema mundial que são a baixa presença da cobertura de esgoto em diversos países, situação está que ocorre no final do século XX até o presente momento.

Uma grande parte da população mundial continua a não ter acesso adequado à água potável e ao esgotamento sanitário, o que é o resultado de injustiça social e desigualdades estruturais. As decisões relacionadas as políticas dos serviços de água e esgotamento sanitário em todos os continentes, incluindo a Europa, têm quase sempre sido implementadas com ausência de debates públicos. (CASTRO; 2016)

A (IBNET) é uma iniciativa de pesquisa, que possui mais de 20 anos de existência e que realiza a coleta de dados sobre o desempenho de concessionárias. Esta iniciativa também fornece informações sobre o desempenho de concessionárias e promove o uso de informações comparativas sobre estas, de modo a proporcionar novas perspectivas à profissionais de muitas organizações.

Os principais interessados nestas informações comparativas de indicadores do saneamento básico como oportunidades de melhorarias em seus trabalhos são: funcionários de concessionárias, agências reguladoras, grupos de clientes e ONGs, agências de ajuda Internacional, consultores e investidores. Com isso, a IBNET é uma iniciativa para incentivar as concessionárias que prestam serviços públicos de saneamento a

reunir e compartilhar um conjunto de indicadores básicos de custo e desempenho, de modo a atender diversas partes interessadas.

O conceito de cobertura de esgoto, que será o indicador avaliado nessa pesquisa, segundo a plataforma é a população com serviços de esgoto, conexão direta, em relação a população total sob responsabilidade da concessionária. Este valor é expresso em porcentagem.

II. METODOLOGIA

A presente pesquisa utilizou dados do IBNET referente a cobertura de esgoto dos países e territórios existentes na plataforma de iniciativa de pesquisa. As obtenções dos resultados dos indicadores foram obtidas da seguinte forma. Primeiramente, clicou-se Banco de dados de Benchmarking, depois em perfil do país. Este procedimento foi realizado, para todos os países e territórios que estavam nesta base, para os anos de 2000, 2010, 2015 e dados mais atuais.

Após a visualização dos valores, na plataforma, foi realizado um quadro com o software Microsoft Excel (2019) para registrar as informações dos países que possuem informações na base de dados.

Ao selecionar o indicado cobertura de esgoto dentro do perfil do país, aparece os períodos em que este possui dados na plataforma.

Após o registro e análise do dados, foi realizado gráficos com as médias dos indicadores de cobertura de esgoto por continentes para cada período, sendo que nesses estará presente uma média mundial desse indicador, para os anos analisados. Os desvios padrões dos indicadores para cada continente também foram calculados.

O indicador de cobertura de esgoto será analisado, para os seis continentes, sendo que a América terá seus territórios analisados de forma independente, visto que as Américas do Norte, Central e do Sul apresentam características distintas.

No período dos dados mais atuais de cada país será realizado um terceiro gráfico, que foi realizado com base em classificação em grupos

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0

3

dos percentuais de cobertura de esgoto, sendo que o agrupamento será determinado com base nos seguintes parâmetros. Os países que apresentavam de 0 a 20% de cobertura de esgoto, encontravam-se no grupo A, de 20,01% a 40% no grupo B, 40,01% a 60% no grupo C e acima de 60 % no grupo D.

Os gráficos com as coberturas de esgoto por países e continentes, podem auxiliar profissionais e empresas de saneamento a terem acesso a informações do panorama médio deste indicador em diversos países do mundo e isso permitirá avaliarem os seus desempenho, além de servirem como incentivo a um processo de melhoria contínua tanto de empresas quanto de profissionais.

III. RESULTADOS E DISCUSSOES

3.1 Análise Da Cobertura De Esgoto Em 2000

No que se refere ao indicador Cobertura de Esgoto, 23 países disponibilizaram dados na plataforma, sendo que na Europa estavam 39,13% destes. Na sequência, tem-se África e Ásia com 21,74% cada e América com 17,39%. Os países da américa que forneceram dados deste indicador estavam localizados na América do Sul. A figura 01 a seguir, demonstra a cobertura de esgoto no ano de 2000.
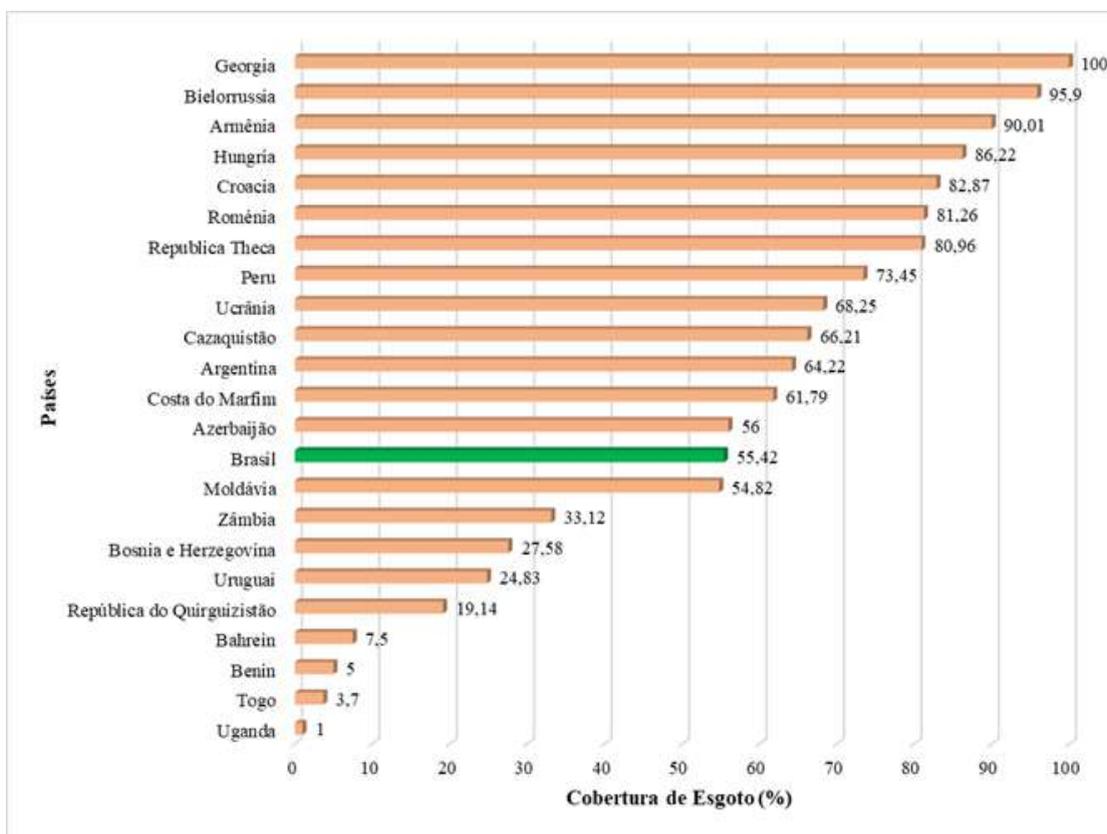
Figura 01: Cobertura De Esgoto (%) No Ano De 2000 Por Países

O menor índice de cobertura de esgoto foi registrado em três países africanos, sendo estes, Uganda com 1%, Togo com 3,7% e Benin com 5%. Os três maiores percentuais de cobertura de esgoto estão na Geórgia (Europa) com 100%, Bielorrússia (Europa) com 95,9% e Armênia (Ásia) com 90,01%. Em relação ao quantitativo de 23 países que disponibilizaram informações, o

Brasil ocupa a 14º posição em relação aos países com maiores coberturas de esgoto, com 55,42%.

Em relação aos países sul-americanos, o Brasil em 2000, apresentou indicador de cobertura de esgoto maior que o do Uruguai que era de 24,83% e menor que o da Argentina e Peru, que eram respectivamente de 64, 22% e 73, 45%. A figura

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

4   Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

02 a seguir demonstra o panorama da cobertura de esgoto no ano de 2000 por continentes.
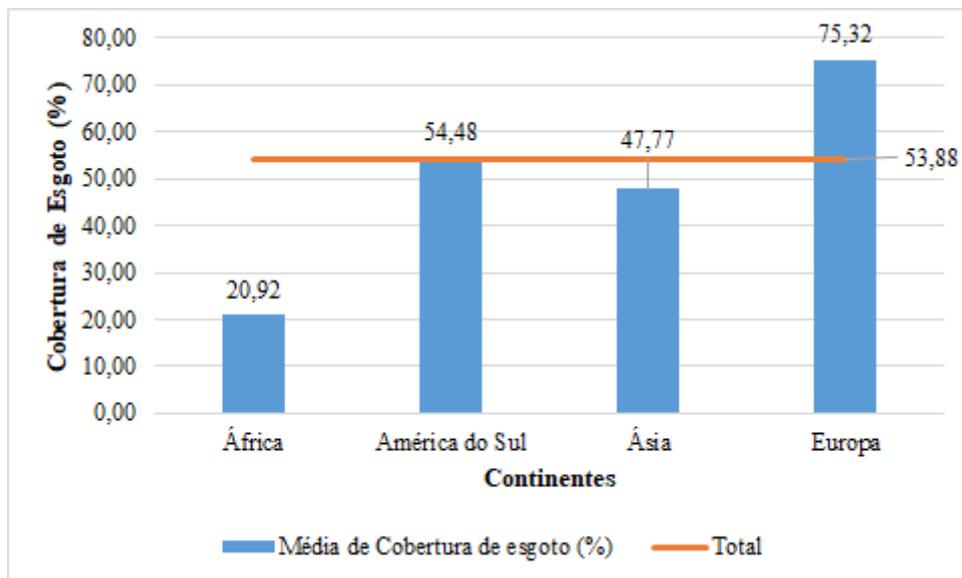
Figura 02: Cobertura de Esgoto (%) nos Continentes: Ano de 2000

O Continente com menor média de cobertura de esgoto foi a África com 20, 92% (desvio padrão de 26,30), seguido da Ásia com 47,77% (desvio padrão de 34,03). Já o continente com maior cobertura deste indicador foi a Europa com 75, 32% (desvio padrão de 22,43) e América do Sul com 54,48% (desvio padrão de 21,09).

A média mundial de cobertura de esgoto em 2000 foi de 53, 88% e os continentes da África e Ásia estão com suas médias, para este indicador abaixo da média mundial para o período.

3.2 Análise da Cobertura de Esgoto em 2000

No ano de 2010, foi disponibilizado dados do indicador cobertura de esgoto no IBNET por 43 países, sendo que no continente Europeu estavam 34,88% destes. Na sequência, tem-se a Ásia com 25,58%, América com 20,93%, África com 16,28% e Oceania com 2,33%. No continente Americano, a América do Sul registrou informações de 6 países e a América do Norte de 3 países. A figura 03 abaixo apresenta a cobertura de esgoto (%) no ano 2010.

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0   5
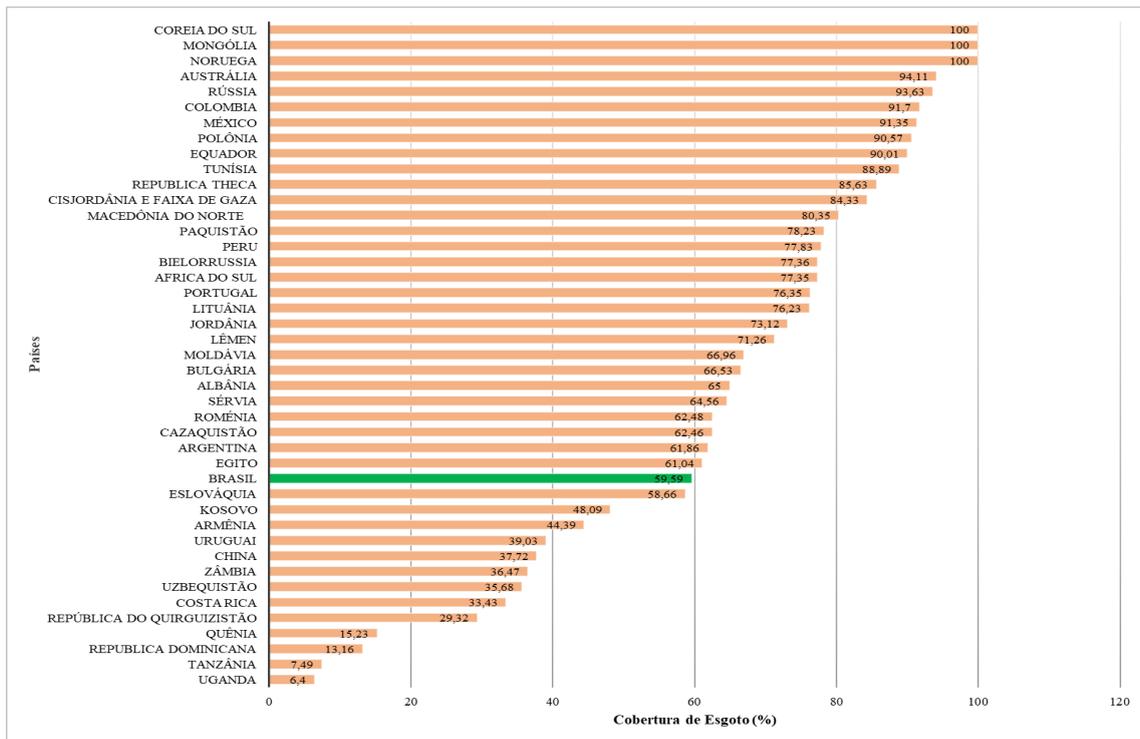
Figura 03: Cobertura De Esgoto (%) No Ano De 2010 Por Países

No ano de 2010, os três países com menores percentuais de cobertura de esgoto (%) foram Uganda (África) com 6,4%; Tanzânia (África) com 7,49% e Republica Dominicana (América do Norte) com 13,16%. Três países apresentaram este indicador com valores de 100%, sendo estes, Noruega no continente Europeu e Mongólia e Coreia do Sul na Ásia.

O Brasil ocupou a 30º posição em relação aos maiores indicadores de cobertura de água em 2000, com 59,59%. Posição superior em relação a países como Uruguai (América do Sul) que ocupa a 34º posição com 39,03%, Kosovo (Europa) na 32º com 48,09% e Eslováquia (Europa) na 31º com 58,66% e posição inferior em relação a países como Argentina na 28º com 61,86%, Peru na 15º com 77,83%, Equador na 9º com 90,01% e Colômbia na 6º com 91,7%. A figura 04 a seguir apresenta o panorama da cobertura de esgoto no ano de 2010 por continentes.

Figura 04: Cobertura de esgoto (%) no ano de 2010 por Continentes

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

6 Volume 22 | Issue 6 | Compilation 1.0

© 2022 London Journals Press

O Continente com menor percentual médio de cobertura de esgoto foi a África com 41,84% (desvio padrão de 32,18). Na sequência, tem-se a América do Norte com 45,98% (desvio padrão de 32,18), Ásia com 65,14% (desvio padrão de 25,32), América do Sul com 70,00% (desvio padrão de 20,33%) ,Europa com 74,16% (desvio padrão de 14,26), Já o continente com maior média de coberturas de esgoto foi a Oceania com 94,11%.

O continente da África e América do Norte, apresentaram médias de cobertura de esgoto com valores abaixo em relação ao da média mundial no ano de 2010 que foi de 64,51% .

3.3 Análise da Cobertura de Esgoto em 2015

Em 2015, os dados de cobertura de esgoto foram disponibilizados na plataforma por 45 países, sendo que na Europa estavam 37,78% destes. Na sequência, tem-se a Oceania com 24,44%, África com 17,78%, Ásia com13,33% e América com 6,67%. Neste último, a distribuição foi equivalente entre as Américas, com 1 país disponibilizando os dados para a América do Sul, Central e do Norte. A figura 05 a seguir apresenta a cobertura de esgoto no ano de 2015.
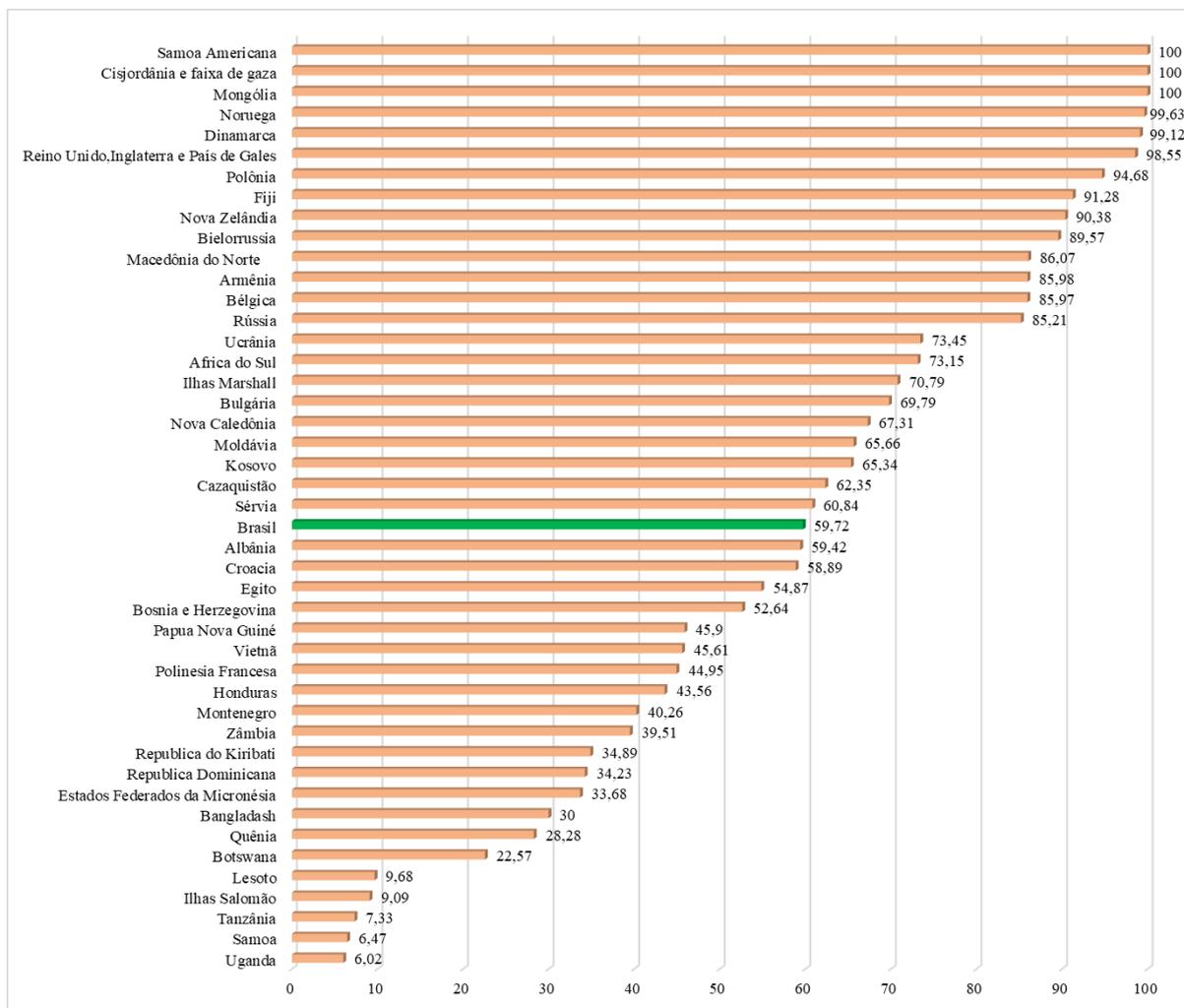
Figura 05: Cobertura de esgoto (%) no ano de 2015 por Países

Os três países com menores percentuais de cobertura de esgoto foram Uganda (África) com

6,02%, Samoa (Oceania) com 6,47% e Tanzânia (África) com 7,33%. Três países registraram

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0

7

cobertura de esgoto de 100% sendo estes, Mongólia (Ásia), Cisjordânia e Faixa de Gaza (Ásia) e Samoa Americana (Oceania).

O Brasil ocupa a 24º dos países com maiores coberturas (%) com 59,72% à frente de países europeus como a Albânia que ocupa a 25º posição com 59,42%, Croácia na 26º com 58,89%, Bósnia e Herzegovina na 28º com 52,64% e Montenegro na 33º com 40,26%.A figura 06 a seguir apresenta panorama da cobertura de esgoto no ano de 2015 por continentes.
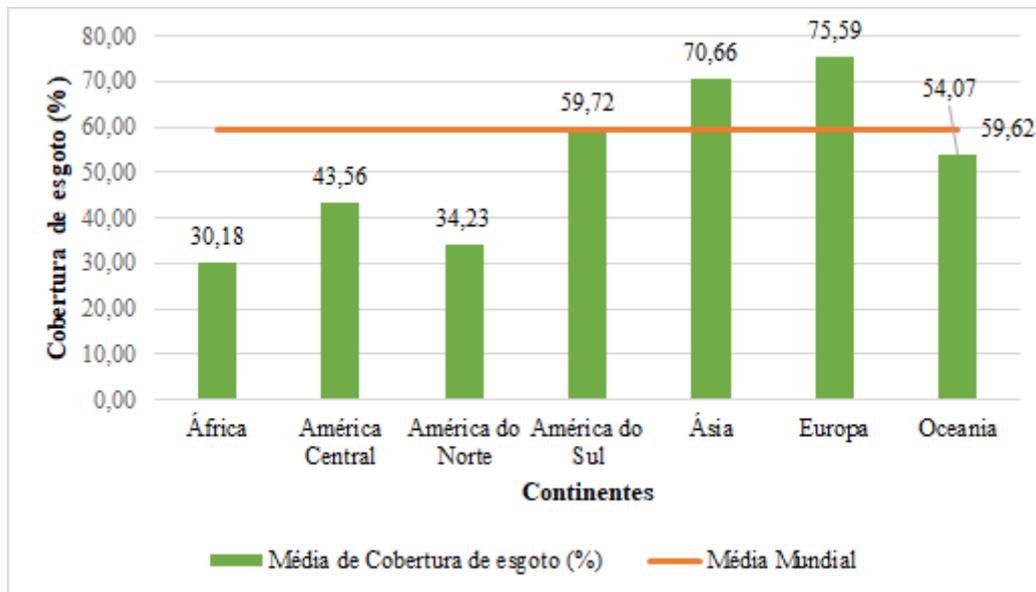
Figura 06: Cobertura de esgoto (%) no ano de 2015 por Continentes

O Continente com menor percentual médio de Cobertura de Esgoto foi a África com 30,18% (desvio Padrão de 24,29). Na sequência, tem-se a América do Norte com 34,23%, América Central com 43,56%, Oceania com 54,07% (desvio Padrão de 32,39), América do Sul com 59,72%, Ásia com 70,66% (desvio Padrão de 29,35). O continente com maior cobertura de esgoto foi a Europa com 75,59% (desvio Padrão de 18,21).

Os continentes da África, América Central, América do Norte e Oceania apresentaram média de cobertura de esgoto inferior à média mundial em 2015.

3.4 Cobertura De Esgoto Com Dados Mais Atuais Dos Países

A informação dos dados mais atuais, foi obtido entrando no perfil de cada país e selecionando esse indicador . Após isso, buscou-se a informação mais atual que cada país disponibilizou em relação ao indicador, com intuito de analisar o panorama da situação da cobertura de esgoto (%) de uma maior quantidade de países. Um quantitativo de 101 países apresentaram essas informações, com a Europa apresentando dados de 25,74% destes. Na sequência, tem- se a África com 21,78%, Ásia com 19,80%, América com 17,82% e Oceania com 14,85%. No continente americano, a América do Sul apresentou informação de 10 países, a América Central e do Norte de 4 países cada. A figura 07 a seguir apresenta a cobertura de esgoto (%) mais atual registrada por países de diversos países.
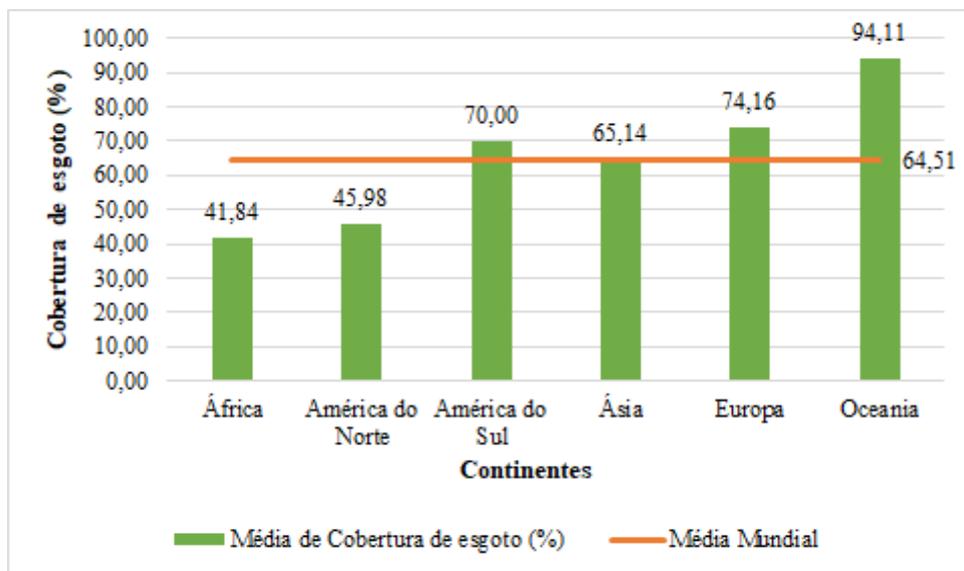
Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

8 Volume 22 | Issue 6 | Compilation 1.0 © 2022 London Journals Press

Figura 07: Cobertura de Esgoto (%) no mundo

Os países com menores percentuais de cobertura de esgoto são o Sri Lanka (Ásia) com 2,61%, Ilhas Salomão (Oceania) com 4,72% e Gambia (África) com 5,36%. Já dez países apresentaram 100% de cobertura de esgoto, sendo estes: Guam (Oceania), Polinésia Francesa (Oceania), Samoa Americana (Oceania), Mongólia (Ásia), Reino Unido, Inglaterra e País de Gales (Europa), Cisjordânia e Faixa de Gaza (Ásia), Coreia do Sul (Ásia), Estados Unidos da América (América do Norte), Namíbia (África) e Singapura (Ásia).

O Brasil ocupa a 55º posição em relação as maiores coberturas de esgoto com 59,72%, possuindo melhor índice em relação a países sul americanos como Paraguai 80º com 31,83% e Uruguai 72º com 38,67% e pior índice do que nações como Argentina que ocupa a 49º posição com 62,65%, Bolívia 45º com 67,49%, Venezuela 37º com 73,62%, Peru 31º com 83,79%, Equador 22º com 90,01%, Colômbia 19º com 91,7% e Chile 16º com 95,45%. A figura 08 a seguir apresenta o percentual dos grupos de cobertura de esgoto, para os dados mais atuais.
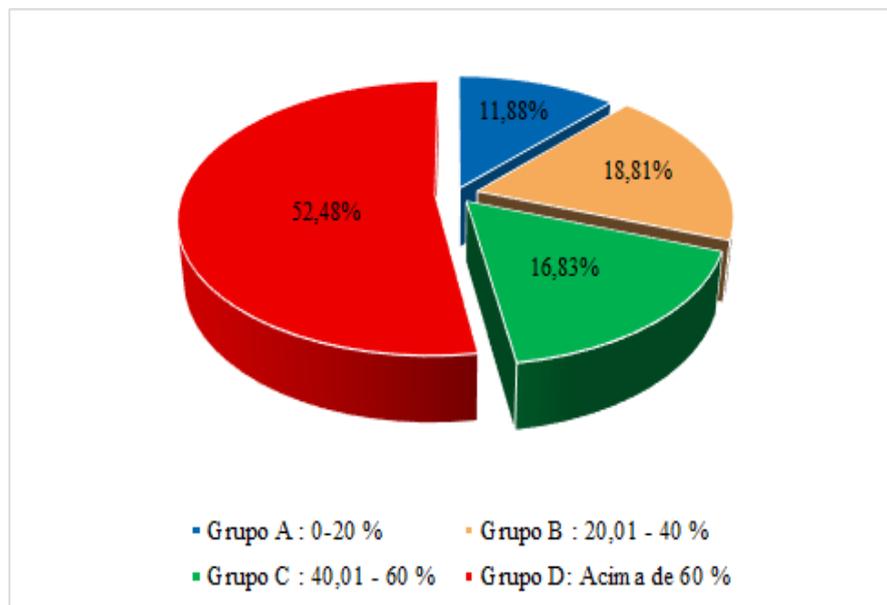

Figura 08: Grupos de Cobertura de Esgoto

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

O grupo A apresenta 11,88% dos países que disponibilizaram dados neste período, o grupo B abrange 18,81%, o grupo C 16,83% e o grupo D é o que apresenta a maior quantidade de países, 52,48%. Apesar da maior parte dos países comporem o grupo D, 47,52% de nações ainda possuem cobertura de esgoto menor do que 60%.

Foi analisado também a cobertura de esgoto por continentes. A figura 09 a seguir apresenta panorama da cobertura de esgoto mais atual, para cada continente sendo observado dados de 2004 a 2020.
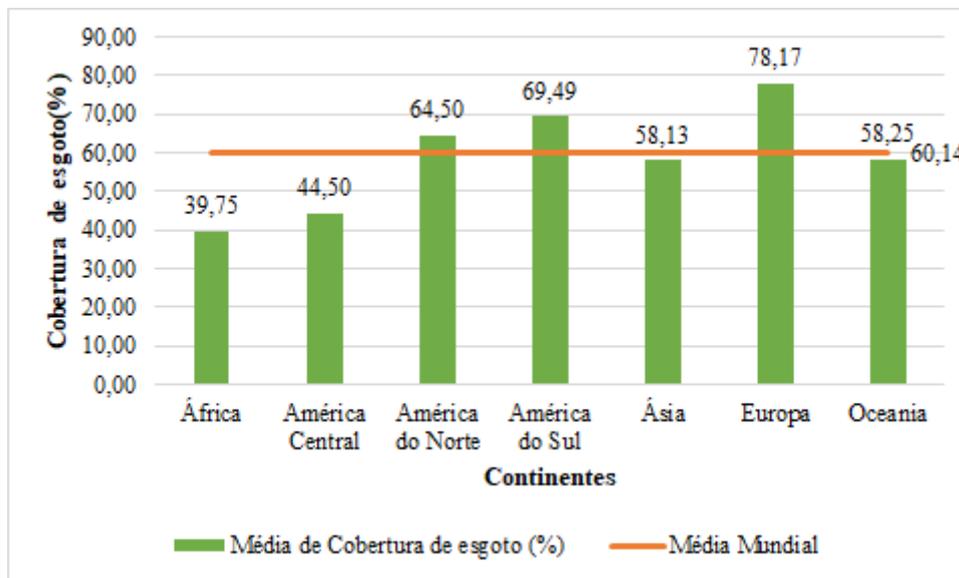
Figura 09: Cobertura de Esgoto (%) no Mundo por Continentes

O Continente com menor percentual médio de Cobertura de esgoto foi a África com 39,75% (desvio Padrão de 28,11). Na sequência, tem-se a América Central com 44,50% (desvio padrão de 7,20), Ásia com 58,13% (desvio padrão de 31,67), Oceania com 58,25% (desvio padrão de 33,02), América do Norte com 64,50% (desvio padrão de 35,67), América do Sul com 69,49% (desvio padrão 21,92). O continente com maior cobertura de esgoto foi Europa com 78,17% (desvio padrão 17,27).

Os continentes da África, América Central, Ásia e Oceania apresentaram média mundial de cobertura de esgoto menor do que a média mundial no período que foi de 60,14%.

IV. CONSIDERAÇÕES FINAIS

A Europa apresentou a maior cobertura de Esgoto em 2000, 2015, atualmente com respectivos percentuais de 75,32%, 75,59% e 78,17% e a segunda maior cobertura em 2010 com 74,16% e o continente africano apresentou a menor cobertura de esgoto em 2000, 2010, 2015 e no último período com respectivos percentuais de 20,92%, 41,84%, 30,18% e 39,75% de corbertura. Com isso, a Europa é o continente que aparece com maior cobertura de esgoto atualmente e a África com menor percentual de esgoto.

A pesquisa analisou em relação aos dados mais atuais disponibilizados, pelas concessionárias dos países no IBNET que 47,52% dos países possuem cobertura de esgoto menor que 60% e que 30,69% dos países apresentam cobertura de esgoto em até 40%.

A ausência na coleta e tratamento do esgoto doméstico é um dos fatores que aumentam o número de diversas doenças. Portanto a presente pesquisa é um mecanismo que permite que as empresas direcionem os investimento aos países que de fato necessitam, de modo a contribuir em melhorias na saúde pública de diversos países, com reduções de doenças de veiculação hídrica, após implantações de estruturas de esgotamento

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

10 Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

sanitário e avanço do indicador cobertura de esgoto.

Qualidade Sócio-Ambiental. Acesso em: out, 2017.

REFERÊNCIAS BIBLIOGRÁFICAS

1. BILIBIO, ELAINE CLAUDIA FARINELI et al. SANEAMENTO BÁSICO UM DIREITO FUNDAMENTAL DA POPULAÇÃO. Percurso, v. 2, n. 39, p. 274-282, 2021.

2. BOVOLATO, Luís Eduardo. Saneamento básico e saúde. Escritas: Revista do Curso de História de Araguaína, v. 2, 2010.

3. BRASIL, CONSTITUIÇÃO FEDERAL. Lei nº 11.445, de 5 de janeiro de 2007. Estabelece diretrizes nacionais para o saneamento básico. Diário Oficial da União, 2007.

4. CASTRO, José Esteban. O acesso universal à água é uma questão de democracia. Boletim Regional, Urbano e Ambiental, v.15, p. 59-65, 2016.

5. DE FUNASA, MANUAL DE SANEAMENTO; ORIENTAÇÕES TÉCNICAS,2007

6. DE FUNASA, Manual de Orientações técnicas, para elaboração e apresentação de propostas e projetos para sistemas de esgotamento sanitário, 2017; disponível em: http://www.funasa.gov.br/biblioteca-eletronica/publicacoes/engenharia-de-saude-publica/-/asset_publisher/ZM23z1KP6s6q/content/manual-de-orientacoes-tecnicas-para-elaboracao-e-apresentacao-de-propostas-e-projetos-para-sistemas-de-esgotamento-sanitario?inheritRedirect=false.

7. DE PAULA FERREIRA, Mateus; GARCIA, Mariana Silva Duarte. Saneamento básico: meio ambiente e dignidade humana. Dignidade Re-Vista, v. 2, n. 3, p. 12, 2017.

8. DOS SANTOS LISBOA, Danielle Cristina et al. Sistema de esgotamento sanitário estação elevatória de esgoto e linha de recalque em um residencial. Engineering Sciences, v. 7, n. 3, p. 9-26, 2019

9. IBNET, The International Benchmarkig Network. Disponível em: https://www.ib-net.org/ acessado: 21/07/2021

10. VAZ, Alexssandra Juliane. A Importância da Rede Coletora de Esgoto na Promoção da

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0

11

This page is intentionally left blank

London Journal of Engineering Research

Assessment of the Sewage Coverage Index in the World from 2000 To 2020: Situation of the Indicator in the Countries and the Importance of Monitoring this Information

12 | Volume 22 | Issue 6 | Compilation 1.0

© 2022 London Journals Press

An Automated Web Structure-Based Method for Predicting the Importance of a Webpage

Syed Tauhid Zuhori & James Miller

University of Alberta

ABSTRACT

The aim of this article is to develop a method to find the importance of web pages without using web browser data or invading the privacy of users. Rather, it works on the structure of a website. To achieve this goal, we propose a novel method that can take webpage content as input and produce a score for each page automatically. Initially, we extract content from a web page in real-time. Subsequently, we consider two important factors based on the website structure: (1) "What is the minimum number of clicks needed to access web pages in a website?" and (2) "How a web page is linked with other web pages in a website?" We use a learning method to train our model by using the "web page views" results generated by "Google Analytics" and "SimilarWeb". Experiments and Case studies on the world's most popular websites show that our method can produce very effective results in real-time.

London Journals Press

An Automated Web Structure-Based Method for Predicting the Importance of a Webpage

Syed Tauhid Zuhori[α] & James Miller[σ]

ABSTRACT

The aim of this article is to develop a method to find the importance of web pages without using web browser data or invading the privacy of users. Rather, it works on the structure of a website. To achieve this goal, we propose a novel method that can take webpage content as input and produce a score for each page automatically. Initially, we extract content from a web page in real-time. Subsequently, we consider two important factors based on the website structure: (1) "What is the minimum number of clicks needed to access web pages in a website?" and (2) "How a web page is linked with other web pages in a website?" We use a learning method to train our model by using the "web page views" results generated by "Google Analytics" and "SimilarWeb". Experiments and Case studies on the world's most popular websites show that our method can produce very effective results in real-time.

I. INTRODUCTION

The most noticeable developments in the Twenty-First century are the innovations that led to the Information Age. The Twenty-First century has all the characteristics of an Information Age as e-commerce takes center stage in our modern life. This is evident in the different enterprises that heavily depend on websites such as banking, shopping, education, hotelier services, and transport. Online shopping is probably one of the most successful innovations in e-commerce.

Through online shopping, many startups have developed different franchises which depend on users' past buying history. This includes advertising, accessing additional customers through social media and marketing in general. Therefore, the primary target is to make a website more intuitive.

One of the most popular applications in this category is "Google Analytics" (www.analytics.google.com) In this case, the developer will review the most viewed web pages, a user's interest in a specific web page and the time spent on that specific web page. This process has been automated by "Google Analytics", perhaps the premier website analyzer in the marketplace. For "Google Analytics" to be practical, code has to be facilitated for the webserver, which the admin uses to manage the analytics. Figure 1 shows the "Google Analytics" code segment that an admin has to set on their server to retrieve results.

"Similarweb" (https://similarweb.com), a web mining application on website traffic, also analyzes the audience behavior of a website. However, it also uses the user's personal information.

London Journal of Engineering Research
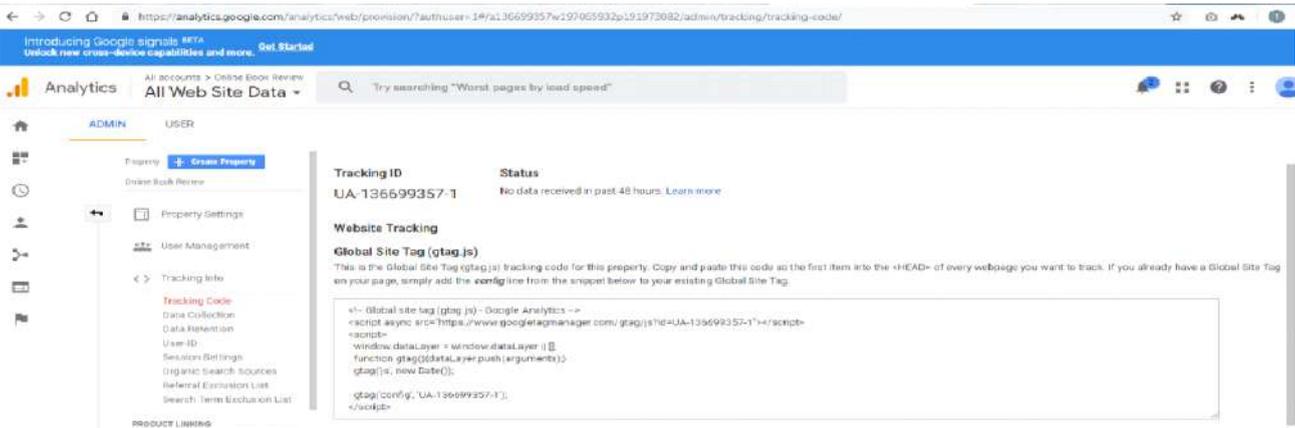
Figure 1: Google Analytics Tracking code

Figure 2 shows that there is a message about using cookies displayed on their application. Therefore, the majority of the tools are using the user's personal information to determine the user's browsing behaviors. However, it is difficult to find the user's personal information or personal choices on websites. So a plausible, less intrusive, solution to this challenge is the use of a website's structure. Hence, we propose a system that tracks web pages in real-time and determines their importance by analyzing the structure of their website.
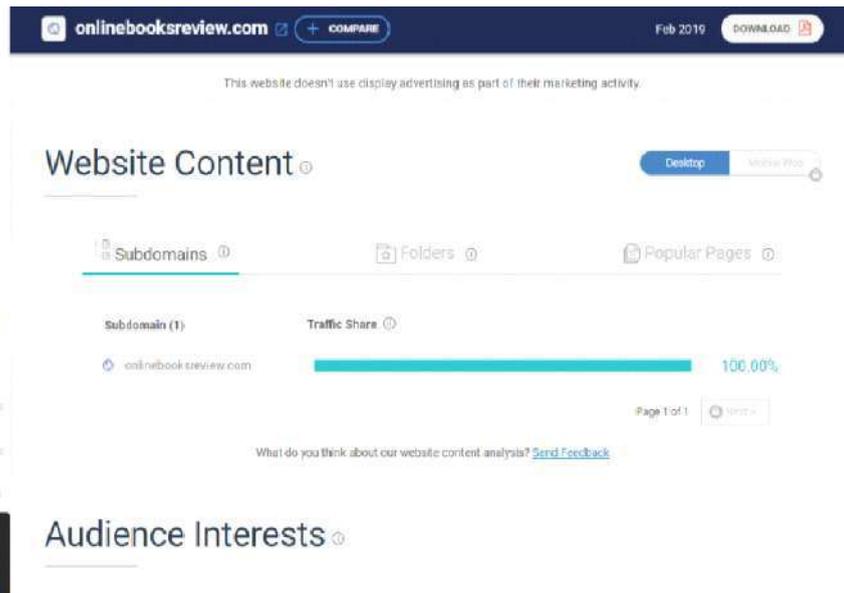
Figure 2: Web Application of "SimilarWeb"

To analyze the structure of web pages, we reviewed one hundred web pages. We selected these pages from the top twenty websites ranked by Alexa. By analyzing the structure, we find five important factors; i) The accessibility of the web pages, ii) The influence of a web page on a website, iii) The content of a web page, iv) Interacting web pages and v) Sharable web pages. The case study on the factors that influence the web pages' importance is represented in section 2. This work aims to provide a solution for online advertisements agencies, by providing an insight into the most viewed pages and providing suggestions to the web developer. This paper makes the following contributions;

- We develop an automated system that suggests areas that require improvement to make a particular web page more important. This is based on the structure of the website, and therefore no user data is required.

- By considering five different factors from the results of a Google Analytics case study on

London Journal of Engineering Research

different websites, we propose a numeric measurement of the importance of web pages on a specific website and also represent the rank (Best, Good, Average and Poor) of the web page.

- We successfully conduct two case studies, by observing and analyzing the web pages of an "Online Book Review" website for twelve weeks, and conducting analysis on five hundred different websites from Alexa using the "Similar Web" tool, since we don't have server access to these sites.
- We conduct an additional case study on the web page "Contact Us" of the website "Online Book Review". We make four versions of it and show how this page can achieve more views by adopting our proposed system's suggestions.
- To validate our work, we use two types of validity – internal and external.
  - For internal validity, we represent the results in a confusion matrix. We automatically generate the features for the web pages using our extension that can extract the number of images, videos, links etc. Then we check the page manually and analyze the content of a web page. After that, we compare the results with manual results and produce a further confusion matrix.
  - For external validity, we also use both cases. We generate the features for both cases using our extension. Then we apply "CatBoost" to produce the importance value and rank. In the case of the "Online Book Review" website, we use our generated features as input and the important value produced by "Google Analytics" as output. On the other hand, for websites ranked by Alexa, we also use the automatically generated features as input and the importance value generated by the "SimilarWeb" as output. Finally, in both cases, we use the "Pearson Correlation Coefficient" and "Spearman Correlation Coefficient" results to show the effectiveness of our work.
- Finally, we show four case studies on four types of rankings generated by our system with automatic suggestions. We manually check the effectiveness of our suggestions.

The rest of this chapter is organized as follows: in section 2.2 we represent our case studies for finding the important factors. Then in section 2.3, we review recent research on the topics of web mining. Section 2.4 describes the architecture of the proposed system for finding the importance of web pages. The experimental results are presented; and a discussion about the evaluation of these results, case studies, and validations are presented in Section 2.5. Finally, Section 2.6 summarizes the chapter and forms some conclusions.

II. FACTORS BEHIND THE WEB PAGES IMPORTANCE: CASE STUDIES

A case study is conducted to ascertain the factors behind the importance of a web page. The most popular websites from "Alexa" are selected for this study. Alexa describes each of the websites on their list based on the user's interest. Twenty websites are selected among the top list of websites for the case study. Some criteria are taken into consideration before choosing the websites for the study. Below we discuss which websites are excluded:

- "Google.com" is excluded because it is comprised of a search page where users need to type in their keywords. So, other website web pages depend on the user's keyword search. This made us exclude "Google.com" from our case study as we choose websites that are not dependent on any specific web page.
- Websites such as "Yahoo", "Facebook", and "Twitter" which require user accounts to access them are also excluded. The reason for its exclusion in our case study is that these sites can't be accessed as a guest.
- We ensure that all websites we work with have all the essential features such as images, texts, videos, and user interactions. So, YouTube is also excluded as most of its features include videos and hence this concentration is a single media type is considered problematic.
- We also ensure that a website written in English is selected for the case study. This is important as we feel that the website text is an important feature. So, we exclude websites

An Automated Web Structure-based Method for Predicting the Importance of a Webpage

that don't use the English language such as "baidu.com", "sohu.com", "Qq.com", and "Tmail.com". In reality, this rationale is simply to accommodate the limitations of the researchers.

- Pornographic websites are avoided because of their adult content.
- We also excluded one-page websites such as "thestartmagazine". The rationale behind this is that we feel that it's important to take into account the minimum number of clicks which won't be possible with a single-page website. Hence, the work presented in this paper only considers multiple-page websites as its domain of research.
- Websites like Wikipedia are also avoided because it is essentially a one-page website whereby any information clicked on appears on another Wikipedia web page which makes it difficult to measure web page hierarchy.

Hence, it is important to understand that our domain of application is limited to those websites which are not examples of our exclusion rules. We believe that the included sites are still the majority of websites (we use 500 websites from the first 656 websites from Alexa in this research). After developing the selection criteria, we spent a period of three months September 2021 to November 2021 monitoring suitable sites.

Table 1 shows the name and rank of the website that is selected from the "included list" for the case studies. The rank is recorded on a monthly basis (September 2021, October 2021 and November 2021) and changes over time; however, the rank is selected for the maximum number of days within the month. For example, if "amazon.com" was ranked 10 for 25 days in September, we choose that; a significant number of the websites are related to "e-commerce" in our case study.

Table 1: Twenty Websites Selected From the Top List of "Alexa"

Name of the Website	Ranking of the website according to "Alexa"		
	September 2021	October 2021	November 2021
Amazon.com	10	10	8
Blogspost.com	23	21	27
Microsoftonline.com	33	28	28
Ebay.com	41	45	37
Github.com	47	47	47
Imdb.com	48	48	48
office.com	50	52	55
stackoverlfow.com	51	49	49
Fandom.com	55	57	59
wordpress.com	57	56	52
imgur.com	58	60	60
Apple.com	61	61	61
Adobe.com	62	67	67
Amazon.in	65	65	69
Quora.com	79	81	78
Bbc.com	85	82	85
Roblox.com	90	95	96
Popads.com	91	93	93
Cnn.com	102	99	100
Spotify	107	120	120

Therefore, twenty websites are chosen from the "Alexa" top list for our case study. The selected websites were observed for three months. We chose 10 web pages all from a website for the case study. 5 out of these 10 pages are most visited while the other 5 represent the less visited pages. These data are collated from the "Similar Web" web application. Below are the steps we used for this study:

- The top 5 most visited and 5 less visited webpages names along with their number of views were extracted from the "Similar Web" app using an automated API (https://github.com/druidoff/similar-web-api/blob/master/SimilarWeb.php). The name of these web pages was collected for 3 months, between September and November 2021. There were variations to the most visited and less-visited pages daily. Based on our case study on 20 websites and 10 web pages each from selected websites, it implies that we collate data from 20 "Alexa" websites daily. This means within three months, we collected data from 18,200 web pages. After data collection from "Alexa", we proceeded to collect data for our case study from web pages we have earlier identified. We focus on the following critical features such as i) Web page contents, ii) Web page influence on a website, iii) Web page accessibility iv) web page interactions and v) sharable web pages.

- To collect web page accessibility data, a site map is created automatically. The technique used will be discussed in detail in section 4. An extension is created to generate the site map. We use this extension to all the 20 Websites manually and an XML sitemap (XML sitemap is a simple list of all the website pages) was produced. Since we cannot define the hierarchical structure in an XML sitemap, therefore, after we have generated the sitemap we will then use it to find the names of all websites' webpages. Also, several duplicate entries were observed in the sitemap, so after we have generated it automatically we then presented it manually in a hierarchical tree structure. We took the names of the web pages from the sitemap and searched for them manually on the websites. After we have found out the webpage name through our search, we then put down the current web page name as "child" and "parent" for where the webpages were found. For example, if we found the "profile" page on the "Home" page that implies that we note the "profile" page as a child while the "home" page will be denoted as a parent. After we completed this pairing process, we were able to easily generate the website's tree structure. We were then able to discover the web page accessibility for all 18,200 web pages automatically for our case study. Table 2 shows the 3 monthly results of the accessibility of the web pages.

- After we found the tree structure of the websites, we then used the tree structure to generate a similarity graph of the websites. However, the graph is not sufficient enough to represent a website because of the high amount of edges appearing on the actual graph representation. For instance, let us assume we can access "Profile" web page from 3 separate web pages of the websites. On the tree structure, we set the "Profile" as a child of the "Home" page. So when a similarity graph is generated, only one edge will be shown while on the actual graph, 3 more edges are shown. To solve this issue, after we generated the similarity graph, we then automatically extracted the web pages' name that can be visited through the current web page. After that, we then deleted the links that are not presented on the same websites (Suppose a link for sharing Facebook, a different website is found). We find these links manually also. Then, edges were set for all the web pages from the current in the similarity graph (Nodes represent web pages' names in the similarity graph). We then find the web page's influence from that graph. We gave in-depth details in section 4. Table 3 shows the 3 months results of the accessibility of the web pages.

- We then collected the web page contents data (number of images, words, videos, weblinks), web page interactions (login, signup, checkout etc.), and shareable web pages (web pages capable of being shared to other social media websites) automatically. All these data were collected within 3 months.

Table 2: Accessibility Value of the Web Pages for Both Categories

Name of the Website	Accessibility of web pages					
	Category 1			Category 2		
	Max	Min	Median	Max	Min	Median
Amazon.com	3	2	2	1	0	1
Blogspost.com	2	1	2	1	0	1
Microsoftonline.com	3	2	2	2	0	1
Ebay.com	3	2	2	1	0	1
Github.com	2	1	1	1	0	1
Imdb.com	4	2	2	2	0	0
office.com	3	2	2	1	0	0
stackoverlfow.com	4	2	2	2	0	0
Fandom.com	2	1	1	1	1	0
wordpress.com	3	2	2	2	1	0
imgur.com	2	1	2	1	0	0
Apple.com	3	2	1	1	0	0
Adobe.com	2	1	2	1	1	0
Amazon.in	3	1	1	2	0	1
Quora.com	3	2	2	1	0	1
Bbc.com	4	2	3	1	0	1
Roblox.com	3	1	2	2	1	0
Popads.com	4	2	1	1	0	0
Cnn.com	3	1	2	1	1	0

Table 3: Influence of a Web Page on Both Categories

Name of the Website	Influence of a web page on a website					
	Category 1			Category 2		
	Max	Min	Median	Max	Min	Median
Amazon.com	1.818	0.672	1.221	1.818	0.672	1.221
Blogspost.com	1.234	0.427	0.872	1.234	0.427	0.872
Microsoftonline.com	1.126	0.482	0.756	1.126	0.482	0.756
Ebay.com	1.112	0.426	0.728	1.112	0.426	0.728
Github.com	1.781	0.657	1.025	1.781	0.657	1.025
Imdb.com	1.289	0.429	0.821	1.289	0.429	0.821
office.com	2.114	0.782	1.412	2.114	0.782	1.412
stackoverlfow.com	1.782	0.678	1.129	1.782	0.678	1.129
Fandom.com	1.987	0.698	1.231	1.987	0.698	1.231
wordpress.com	2.112	0.772	1.467	2.112	0.772	1.467
imgur.com	1.123	0.419	0.758	1.123	0.419	0.758
Apple.com	1.256	0.425	0.857	1.256	0.425	0.857
Adobe.com	1.984	0.678	1.241	1.984	0.678	1.241
Amazon.in	2.123	0.782	1.435	2.123	0.782	1.435
Quora.com	1.678	0.578	1.098	1.678	0.578	1.098
Bbc.com	1.876	0.612	1.287	1.876	0.612	1.287
Roblox.com	1.987	0.682	1.257	1.987	0.682	1.257
Popads.com	1.876	0.662	1.298	1.876	0.662	1.298
Cnn.com	1.276	0.452	0.872	1.276	0.452	0.872

Table 4: Number of images of web pages for both categories

Name of the Website	Images on a webpage					
	Category 1			Category 2		
	Max	Min	Median	Max	Min	Median
Amazon.com	39	7	17	13	2	6
Blogspost.com	10	2	7	5	0	4
Microsoftonline.com	11	3	6	4	1	2
Ebay.com	32	6	11	16	2	6
Github.com	4	0	2	2	0	1
Imdb.com	12	3	6	6	2	2
office.com	15	3	5	5	3	1
stackoverlfow.com	6	1	3	3	2	1
Fandom.com	11	2	5	4	2	2
wordpress.com	10	2	4	5	2	1
imgur.com	13	1	9	5	0	3
Apple.com	21	3	17	11	1	9
Adobe.com	12	2	10	4	1	4
Amazon.in	32	5	24	16	2	12
Quora.com	8	1	6	3	1	2
Bbc.com	12	2	5	6	1	2
Roblox.com	17	2	9	6	1	3
Popads.com	12	1	7	6	1	4
Cnn.com	31	3	17	11	2	1

Table 5: Number of Words of Web Pages for Both Categories

Name of the Website	Number of words in a web page					
	Category 1			Category 2		
	Max	Min	Median	Max	Min	Median
Amazon.com	265	102	221	67	17	51
Blogspost.com	900	321	567	78	21	55
Microsoftonline.com	121	81	109	189	101	156
Ebay.com	123	89	98	23	19	20
Github.com	289	218	265	28	12	22
Imdb.com	247	127	187	34	14	29
office.com	265	123	210	93	29	56
stackoverlfow.com	1081	657	891	244	128	182
Fandom.com	230	124	189	332	159	218
wordpress.com	429	128	321	134	29	98
imgur.com	129	98	121	321	129	228
Apple.com	287	129	189	453	239	329
Adobe.com	127	80	102	32	12	21
Amazon.in	328	213	289	24	15	18
Quora.com	821	578	682	87	12	56
Bbc.com	928	456	781	33	18	25
Roblox.com	278	189	221	29	21	27
Popads.com	210	178	192	21	11	18
Cnn.com	316	135	219	87	28	67

An Automated Web Structure-based Method for Predicting the Importance of a Webpage

Number of Interactions of Web Pages for Both Categories

| Name of the Website | Number of words in a web page | | | | | |
| | Category 1 | | | Category 2 | | |
	Max	Min	Median	Max	Min	Median
Amazon.com	265	102	221	67	17	51
Blogspost.com	900	321	567	78	21	55
Microsoftonline.com	121	81	109	189	101	156
Ebay.com	123	89	98	23	19	20
Github.com	289	218	265	28	12	22
Imdb.com	247	127	187	34	14	29
office.com	265	123	210	93	29	56
stackoverlfow.com	1081	657	891	244	128	182
Fandom.com	230	124	189	332	159	218
wordpress.com	429	128	321	134	29	98
imgur.com	129	98	121	321	129	228
Apple.com	287	129	189	453	239	329
Adobe.com	127	80	102	32	12	21
Amazon.in	328	213	289	24	15	18
Quora.com	821	578	682	87	12	56
Bbc.com	928	456	781	33	18	25
Roblox.com	278	189	221	29	21	27
Popads.com	210	178	192	21	11	18
Cnn.com	316	135	219	87	28	67

Table 7: Number of Sharable Web Pages the Web Pages for Both Categories

| Name of the Website | Sharable web pages | | | | | |
| | Category 1 | | | Category 2 | | |
	Max	Min	Median	Max	Min	Median
Amazon.com	6	4	4	2	0	2
Blogspost.com	4	2	4	2	0	2
Microsoftonline.com	6	4	4	4	0	2
Ebay.com	6	4	4	2	0	2
Github.com	4	2	2	2	0	2
Imdb.com	8	4	4	4	0	0
office.com	6	4	4	2	0	0
stackoverlfow.com	8	4	4	4	0	0
Fandom.com	4	2	2	2	2	2
wordpress.com	6	4	4	4	2	2
imgur.com	4	2	4	2	1	1
Apple.com	6	4	2	2	0	0
Adobe.com	4	2	4	2	2	2
Amazon.in	6	2	2	4	0	2
Quora.com	6	4	4	2	0	2
Bbc.com	8	4	6	2	0	2
Roblox.com	6	2	4	4	1	2
Popads.com	8	4	2	2	1	1
Cnn.com	6	2	4	2	1	1

The overall result is represented in Tables 8 – Table 10. We show a total of the six features here while we further show the number of times category 1 (5 most visited pages) exceeds category 2 (5 less visited pages) for the twenty websites and six features. The results are shown for three months. We can see that category 1 leads over category 2 in all features.

Table 8: Number of Times Category 1 Exceeds Category 2 or Vice Versa (In Case of Maximum Value)

| Feature | The maximum value of both categories in three months | | | | | |
| | Sep 2021 | | Oct 2021 | | Nov 2021 | |
	Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
Accessibility of the web pages	20	0	20	0	20	0
Influence of a web page in a web site	19	1	20	0	20	0
Images of the webpage	15	5	16	4	18	2
Texts of the webpage	16	4	16	4	17	3
User interactions of the web pages	18	2	17	3	17	3
Sharable web pages	19	1	17	3	19	1

Table 9: Number of Times Category 1 Exceeds Category 2 or Vice Versa (In Case of Minimum Value)

| Feature | The minimum value of both categories in three months period | | | | | |
| | Sep 2021 | | Oct 2021 | | Nov 2021 | |
	Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
Accessibility of the web pages	20	0	20	0	20	0
Influence of a web page in a web site	19	1	20	0	20	0
Images of the webpage	15	5	16	4	18	2
Texts of the webpage	16	4	16	4	17	3
User interactions of the web pages	18	2	17	3	17	3
Sharable web pages	19	1	17	3	19	1

Table 10: Number of times category 1 exceeds category 2 or vice versa (in case of median value)

| Feature | The median value of both categories in three months period | | | | | |
| | Sep 2021 | | Oct 2021 | | Nov 2021 | |
	Category 1	Category 2	Category 1	Category 2	Category 1	Category 2
Accessibility of the web pages	20	0	20	0	20	0
Influence of a web page in a web site	19	1	20	0	20	0
Images of the webpage	15	5	16	4	18	2
Texts of the webpage	16	4	16	4	17	3
User interactions of the web pages	18	2	17	3	17	3
Sharable web pages	19	1	17	3	19	1

After the case study and analysis of the results, we find the proposed factors can influence the web pages' importance value:

- *Accessibility of the web pages:* This is based on the landing page of the website and the tabs available on the site. The landing page is technically the home page. Accessibility in this case, therefore, means that all the web pages that are accessible by one click from the landing page or the home page are more accessible than the pages that are accessible by two or more clicks from the home page.

- *Influence of a web page on a website:* Besides the accessibility of the web pages, we also observe that the web page that has more links and that can be accessed from more web pages has more influence on web users. Because the user comes to that page to visit the related web pages. In that case, the page views will be increased.

- *Content of the Web pages:* Text or information, images, and videos are referred to as content of the web pages. The page that contains more of this is considered more important.

- *Interacting Web pages:* Interactive web pages which require user input and show different outputs for different users by utilizing their inputs are as well considered to be more important.
- *Shareable Web pages:* This is in association with social media platforms, where some web pages contain links to social media platforms and their usability is clearly stated.

III. RELATED WORK

We examine the research work related to our study in this section; our work is related to web mining, which can be categorized into three active research areas depending on what components of web data are mined. The first one is Content Mining which is the process of extracting relevant information from the content of websites. The next one is Structure Mining which uses links and references within web pages. After analyzing that, It can obtain the underlying topology of the interconnections between web objects. The final one is usage mining which studies user access information from log server data. Our paper is based on website structure mining. However, our research uses structure mining to predict user behaviors.   Hence, we include research work related to both mining topics.

Multazim et al. [2015] analyze whether classified ads can increase search engine rankings and increase the number of visitors to a website. They note that "Firefox" and "Google Chrome" are the most popular search engines. Hence, their study is based on ad's data generated by those search engines. They point out that posting and advertising are carried out by various auto-submit programs. They concluded that the installation of classified ad with 'Auto submit' increases the number of visitors.

Verma et al. [2015] make it clear that it is prudent for every organization to have a good website. Nonetheless, e-commerce is still in the developing stages in some countries such as India. They postulate that the challenging and dynamic needs of consumers are not satisfied in such countries where e-commerce is not well established. They make arguments based on research work that

focuses on the design of a page ranking algorithm (SNEC). They explain that SNEC aids customers to search and compare products before purchasing. Finally, they recommend that business organizations need to structure their e-commerce websites to be more effective and usable.

Gleich et al. [2015] describe Google's PageRank method which evaluates the importance of web pages through their link structure. . They explain the process involved in determining the importance of web pages through various illustrations and mathematical formulae.

Khan et al. [2017] propose a new model, the popularity and productivity model (PPM). The model is based on a modular approach to finding the most influential bloggers. They describe in-depth the roles of the model's existing features and evaluate the proposed model by using data from real-world blogs. , they validate that PMM identifies influential bloggers. They make use of performance evaluation measures for the comparative analysis.

Tamimi et al. [2015] present the results of an experiment in which participants view fictitious web pages. They postulate various conceptual methods that are involved. Their study indicates that star reviews and familiarity with e-tailor (e-Bay or Amazon) are the main attributes that influence an individual's likelihood of purchasing products online. They further claim that their results are consistent with findings of previous research (Kim et al. [2010], Stocks et al. [2011]).

They point out that while they encountered various limitations, their research can help to provide a more realistic task for a better comprehension of the attributes that have implications on consumers' decisions concerning the purchase of products online.

Zhen et al. [2016] combine the h-index and the PageRank algorithm. Their main aim is to find out the impact value of a publication. They construct the resulting PR-index for any publication by evaluating the popularity of the source as well as the source publication authority. Therefore they

propose their method should be added to technical indices.

Fatehu et al. [2016] propose a two-stage supervised approach to suggest news articles to users for a given state of Wikipedia. Initially, they suggest news articles to Wikipedia entities (article-entity suggestions) relying on a rich set of features. Then they determine the exact section in the entity page for the input article (article-section placement) guided by class-based section templates. They perform an evaluation of their approach based on ground-truth data that is extracted from external references in Wikipedia.

Zhen et al. [2017] observe that while there are many hypertext links on the web, only a few are clicked regularly. Based on this observation, they make use of mixed-effects hurdle models supplemented with descriptive insights and find out user preferences involved in clicking links on the web. They adopt the PageRank algorithm in their study. They utilize o large-scale data sets from Wikipedia (English version only) for their experiment. They conclude that Wikipedia users have a preference for navigating to articles that are in the periphery of the Wikipedia link network, compared to semantically similar articles, and to articles that are linked at the top of the left-hand side of the source article.

Thomas et al. [2019] research work is highly related to our research work. A research model was created with the use of a stimulus-organism-response (S-O-R) model (S. W. Khun et al. [2018]) to explicate how the social commerce features affect the website attention (stickiness) through ideas about cognitive and emotional factors. The meaning of the word "Website stickiness" entails the amount of attention received by a website from its users.

E-commerce websites will find this very useful to their operations. Originating from environmental psychology, the S-O-R model postulates that certain stimulus affects the cognitive and emotional states of an individual; this then informs the individual's response or behavior.

Based on the S-O-R model, the cognitive and emotional states of the individuals facilitate a stimulus and response relationship. In the field of e-commerce, the S-O-R model has been widely tested by several studies to note how particular web features like stimuli (e.g., pictures, product descriptions, navigation aids) can influence consumers' responses like buying behavior. Their research model is assessed in a controlled online experiment with 164 participants using e-commerce website variants with different social commerce feature richness levels. It was indicated in their results that cognitive and affective dynamics affect feature richness positively, thereby increasing a website's stickiness. The result further concludes that e-commerce websites can be enhanced with a combination of functionally varied social commerce features.

Unfortunately, they only work with high-level (abstract) issues of the website such as; user satisfaction, the usefulness of the websites, how users trust to share their data on websites etc. For validating this they design four versions of a website and take user responses on these issues.

The high-level issues are varied from user to user. They take responses from a total of 212 participants and use the responses of 164 participants in their work (as 164 participants give an acceptable response) but it is still a very low number of users. Ultimately, this work fails to produce any guidance which would be meaningful to web designers or programmers. This is a key objective of our work.

To review the above-related work, we observe that, for finding the importance of web pages the previous research works on user navigation patterns, cookie information or other private data. Therefore, there are two significant problems with the previous research in this field. The first one is to collect user data. The second one is the use of previous data to find the solutions based on past user behaviors. So, most of the previous research work considers an old dataset of the website. For instance, suppose a new web page called "Donate Now" is included in a website for any incident. At that time that web page may attract more visitors but as the model learns from the previous dataset where the "Donate Now" link is not available, it would not be shown as the most visited page. So

An Automated Web Structure-based Method for Predicting the Importance of a Webpage

the previous works fail to decide in real-time and algorithms cannot quickly adapt to maintenance changes on an ongoing basis. In our work, we design a model according to the website structure. So, our work can give real-time predictions without using any private data of users and automatically adapts to maintenance changes. To our best knowledge, this is the first-ever work that can analyze the effectiveness of a web page by only analyzing the structure of the web pages.

IV. METHODOLOGY

We have constructed an extension for web browsers. For this, we access the web pages' content and find the features by analyzing that content and page URL. These features are the input of our proposed system. Then for each web page, we find the importance value by using the page view results that were generated by "Google Analytics" and "SimilarWeb". We use this page view results to train our model and use the CatBoost Machine Learning (Liudmila et. al. [2010]) system to produce the final importance metric. Finally, based on the metric, we rank the web pages on the scale of "Best," "Good," "Average" and "Poor." It is hoped that this procedure is straightforward enough, to make it accessible to most web site designers without requiring them to learn additional technology.

4.1 Feature Extraction

The information extracted from web pages and used as features are: i) what is the minimum number of clicks to visit those pages from the Homepage? ii) What is the number of images, texts, videos, links and scripts of the web pages? iii) How is a web page connected with other web pages on the website? iv) Are there any interactions with the users on the web pages? The establishment of these features is based upon the case study presented in Section 2. In this section, we discuss our methodology to extract these features from the web pages.

4.1.1 What Is the Minimum Number of Clicks to Visit a Web Page

For finding the minimum number of clicks, we first extract the site map of the website using the

method used by Brawer et al. [2017]. Starting with the "home" page, we get the all links that can be accessible and save them on the site map. Then we prune all the duplicate entries and increase the value of a minimum number of clicks. After that, we repeat these steps until there is no child found in the DOM Tree. Finally, when there is no child in the DOM tree, we find the site map of the websites with the minimum number of clicks. In our proposed method we use the ease of web pages' accessibility from a website. Therefore, after finding the minimum number of clicks, we find the easiness of the accessibility value of web pages (E) using; E = D − C, Where D is the Maximum depth of a Tree and C is the minimum number of clicks.

4.1.2 What Is the Number of Images, Texts, Videos, Links and Scripts on the Web Pages

We extract the DOM structure of a page and identify a summary of the content: (1) the number of images, (2) the number of videos, (3) the number of links on the web page, (4) the amount of text (in words) and (5) the number of scripts on that page.

4.1.3 How Is a Web Page Connected With Other Web Pages on the Website

We produce an undirected graph for the web pages which represents the connectivity of all the pages on a website. Using this graph, we calculate the Eigen vector centrality for each of the nodes. (Here nodes mean the URLs of the website.) We use the adjacency matrix to find the Eigen vector centrality. For any vertex, v the relative Eigen Vector Centrality x can be defined as:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

Where, $a_{v,t}$ is the adjacency matrix ($a_{v,t} = 1$, if there is an edge between the vertex v and t), M(v) is the set of neighbors of vertex, v and λ is a constant.

4.1.4 Are there any interactions with the users on the web pages

we loop through all the "<a>", "<nav>", "<submit>", "<form>" elements of the page. Before that, we collect keywords that are used for "Login," Signup," "Share on Facebook," Share on Tweeter," "Share on Google Plus," "Checkout". We collect these keywords by analyzing the Alexa top 400 websites. For this analysis, we only consider the home page of each website (Table 11). If we find these keywords within the tag elements, we infer that there is an interaction with the users.

Table 11: Keywords Collected From the Alexa Top 400 Websites

Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
Log In	82	submit	24	Share on Facebook	15
Logon	17	Login	102	Share on Tweeter	12
Log	11	tweet	18	Share	45
Sign In	82	Facebook	28	Login Scope	18
Signin	27	googleplus	15	Share on Google	14
Signup	9	checkout	6	join	32
Sign up	31	check out	17	register	56

4.2 CatBoost Learning to Rank the Web Pages

Dealing with categorical features efficiently is one of the biggest challenges in machine learning. The most widely used technique to deal with categorical predictors is *one-hot-encoding*. The original feature is removed and a new binary variable is added for each category. Another way of dealing with categorical features is to use the so-called label-encoding technique that converts discrete categories into numerical features.

Beyond these approaches, CatBoost (Liudmila et. al. [2010]) is a specialized version of Gradient Boosting Decision Trees (GBDT), which solves problems with ordered features while also supporting categorical features. It uses a technique where the trees included in the model are not independent but sequential. In other words, each predictor learns and improves from the mistakes and errors of the previous tree. In the end, all of the trees or predictors are combined to form the model but with non-uniform weights. Each tree is constructed by the following steps: 1) splitting calculations, ii) transformation of categorical data to numerical data, iii) construction of the tree, and, iv) computation of the values in the leaf nodes.

After the first split is selected on the tree, the same step is repeated for the next split only with a condition of 'given the first split'. The same step is repeated with a similar condition until the whole

tree is constructed. The model constructed includes a tree whose leaf values provide a score which is our output — importance values. The score is further taken as input to classify itself as a rank. To categorize these values as one of the ranks, we break the range (output range from importance value) into four subranges (to decide what category a specific result falls under). (Using the ElbowMethod (Trupti et al. [2013]) estimates that k = 4 is the optimal partitioning). Further, to explain the subranges and the categories, the range in which the page is most likely to attract an audience and publicize or perform best is categorized as BEST, the one that draws a little less attention is named GOOD, the one that occasionally gets views is AVERAGE and the one that get rare or no audience at all is categorized as POOR.

V. RESULTS & EVALUATION

For evaluating our results, we generate two sorts of datasets. The first one has been generated from the "Google Analytics" results for the "Online Book Review" website (https://www.Onlinebooksreview.com/). we obtain server access to this website, for twelve weeks. The study follows best practices in maintaining users' anonymity and privacy, hence no unique information can be reported, we obtained access between July 2019 and September 2019. We used the first eight weeks' data as the training set; and the last four

An Automated Web Structure-based Method for Predicting the Importance of a Webpage

weeks' data as live data. Besides this, we also generate a second dataset using "SimilarWeb." We choose Five Hundred websites from "Alexa" and generate datasets using the "SimilarWeb." We use the first four hundred websites as the training dataset and the rest One Hundred as the live dataset. Then we show the results, and the importance values generated by our system. After that, we represent three case studies; at first one we represent how our system can extract the features from the web page, we represent four different pages for which our system generates the four different scores; "Poor", "Average", "Good" and "Best" according to their value and importance, in last one we show a case study on the web page "Contact Us" (https://www.online booksreview.com/contact) of the web site "Online Book Review". We make four versions of it and show how this page can achieve more views by adopting our proposed system's suggestions. Finally, we represent two types of validity experiments to prove the effectiveness of our system; Internal and external.

5.1 Dataset Visualization

We use a total of 8 datasets in our research. We define the datasets as "Dataset 1" to "Dataset 8". The first 3 datasets (Datasets 1 to 3) are based on the web pages of the website "Online Book Review". Dataset 1 contains the data for September 2021; Dataset 2 contains the data for October 2021, and Dataset 3 contains the data for November 2021. There are a total of 239 web pages on the "Online Book Review" website; however, we select 100 web pages from them. We discard pages that are similar (such as "Articles on Programming"). In that case, we select one web page from each group. The combination of Datasets 1 and 2 is used as the training set, and Dataset 3 is used as testing. Figure 3 shows the "Number of views" results from the "Google Analytics" for each Dataset.
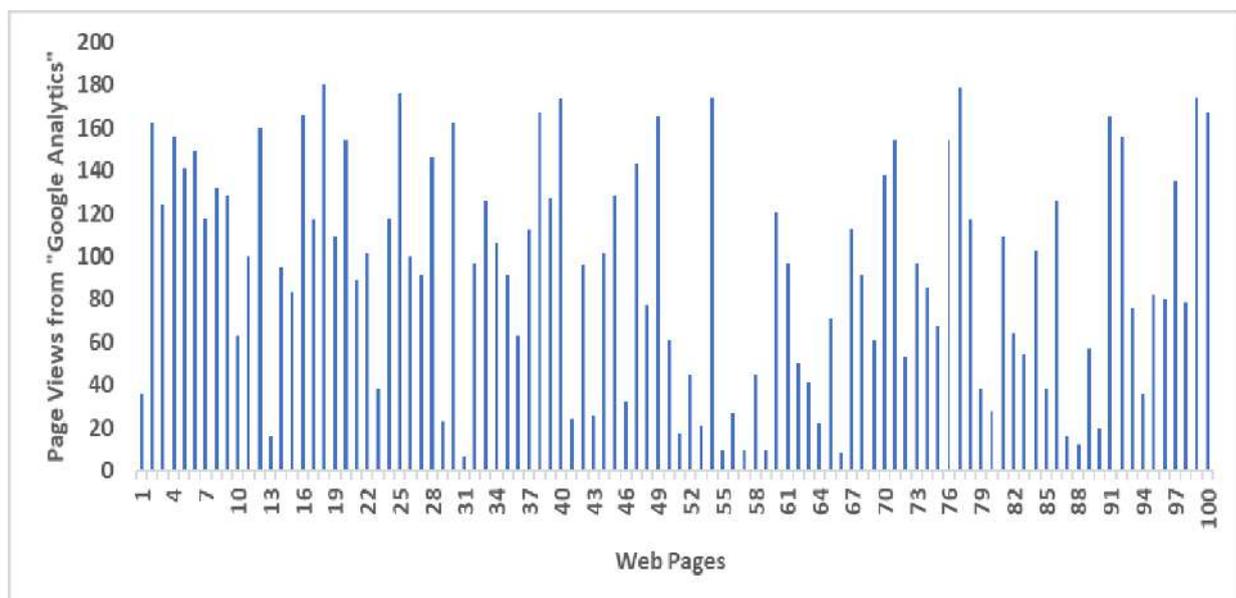
Dataset 1
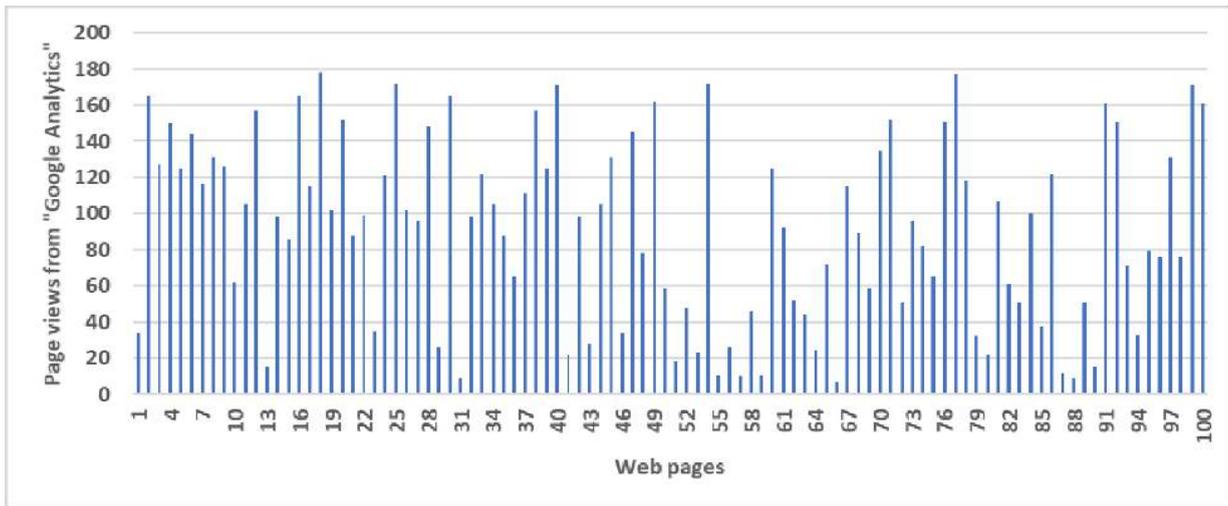
Dataset 2
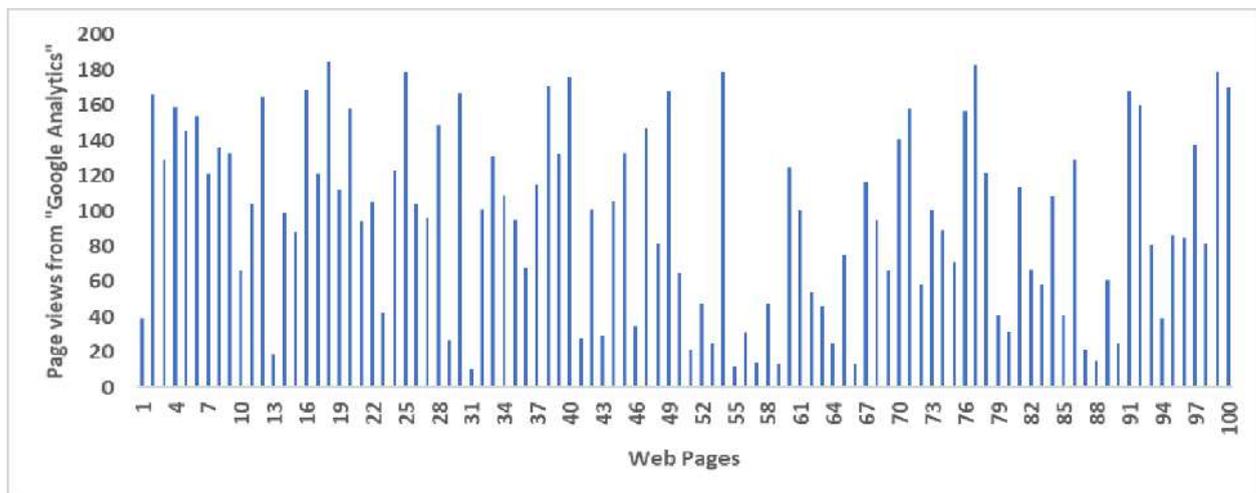
Dataset 3

*Figure 3:* Dataset Produced from "Online Books Review" Website

We also evaluate our work on the 500 popular websites ranked by "Alexa." We take the top 656 websites and remove 156 websites from the list.

There are two reasons behind that. The first one is some of the websites do not meet the criteria defined in the case study section (We delete 97 websites from the list for this reason). As an example, "Google.ca" is very different from the other websites. We consider websites with more user interaction. The second reason behind that is, that we need the number of views of any specific web pages to train our model. We use "SimilarWeb" to collect these "page views" as we don't have server access to the websites.

Therefore, we select web pages for which "SimilarWeb" can generate these results. For 59 websites from the top 656 websites of Alexa, "SimilarWeb" fails to produce results; therefore, we discard them from the list. So after cleaning the websites dataset, we include 5 datasets; naming them "Dataset 4" to "Dataset 8", where Dataset 4 to Dataset 7 are used for training and Dataset 8 is used for testing. We have 489 web pages in Dataset 4, 540 in Dataset 5, 639 on Dataset 6, 659 on Dataset 7, and 611 in Dataset 8.
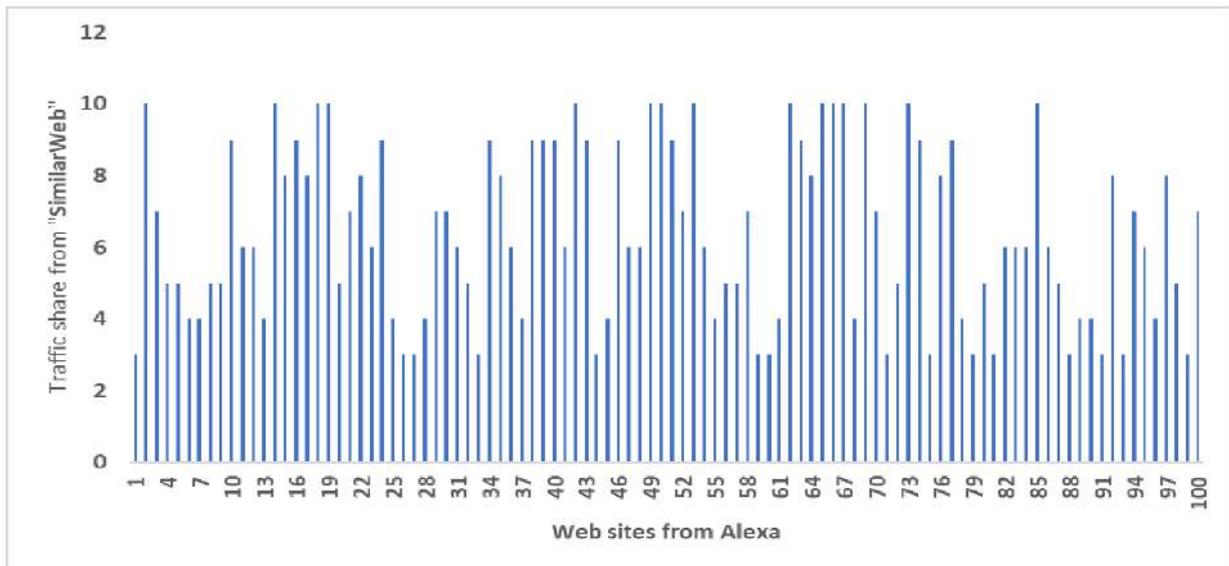
Therefore, we have a total of 2,938 web pages in the dataset where 2327 web pages are used as training for our model and 611 web pages are used as testing. In figure 4 we represent for each site the number of pages we consider in our system.
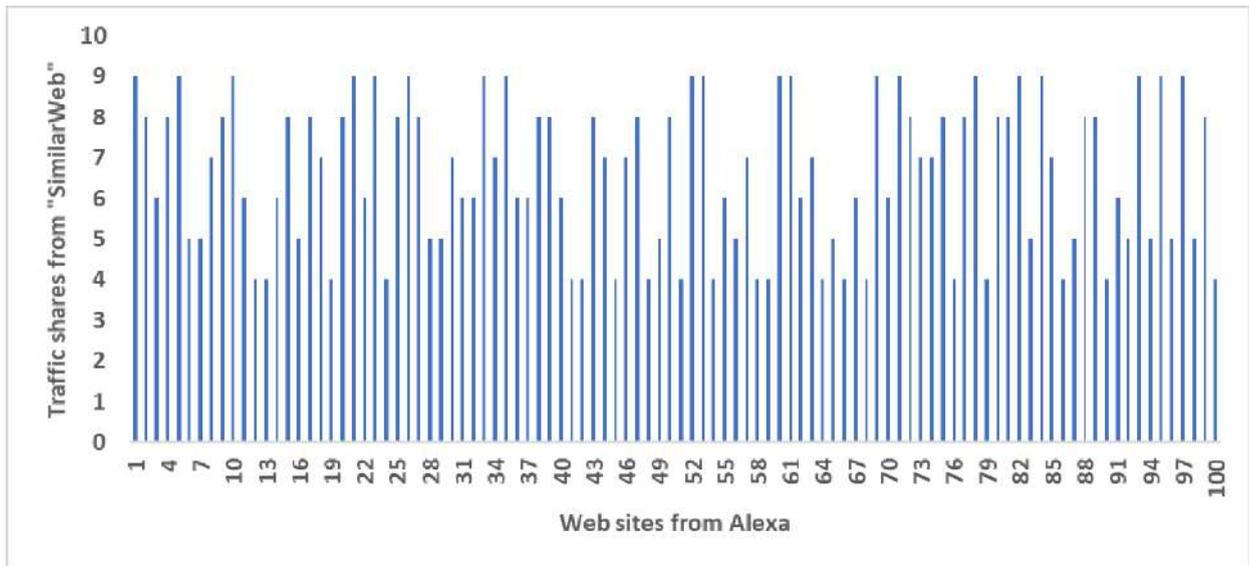
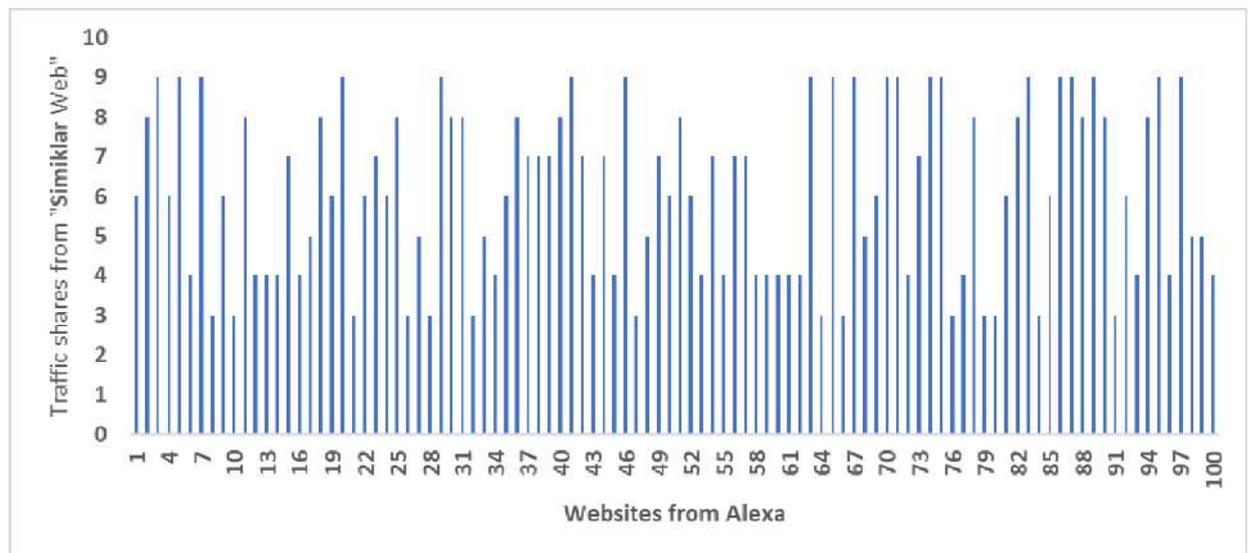An Automated Web Structure-based Method for Predicting the Importance of a Webpage
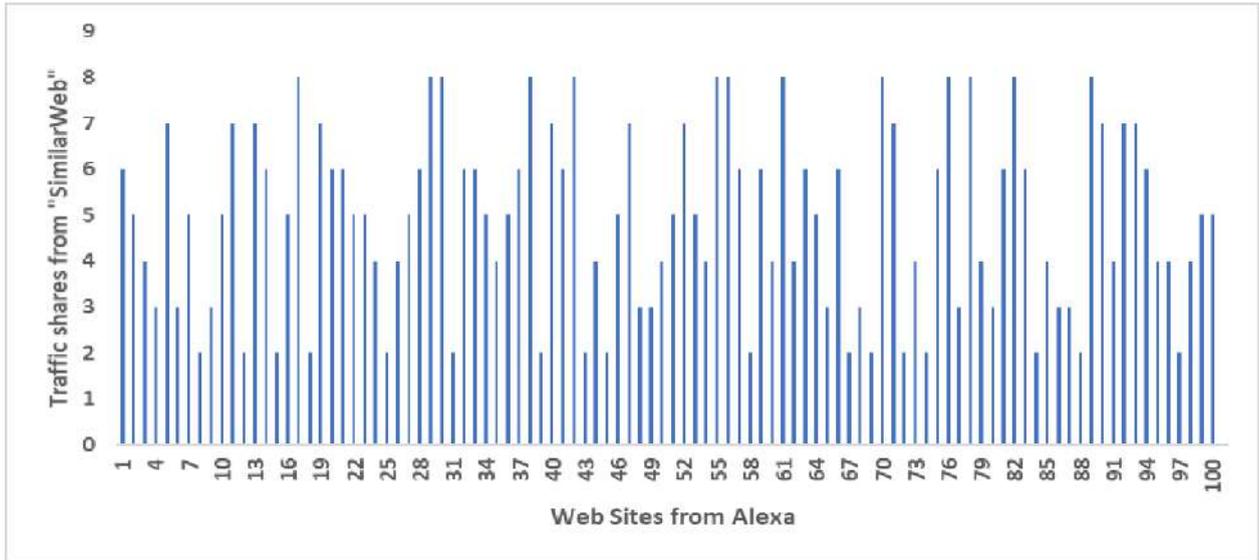
Dataset 4

Dataset 5

Dataset 6

An Automated Web Structure-based Method for Predicting the Importance of a Webpage
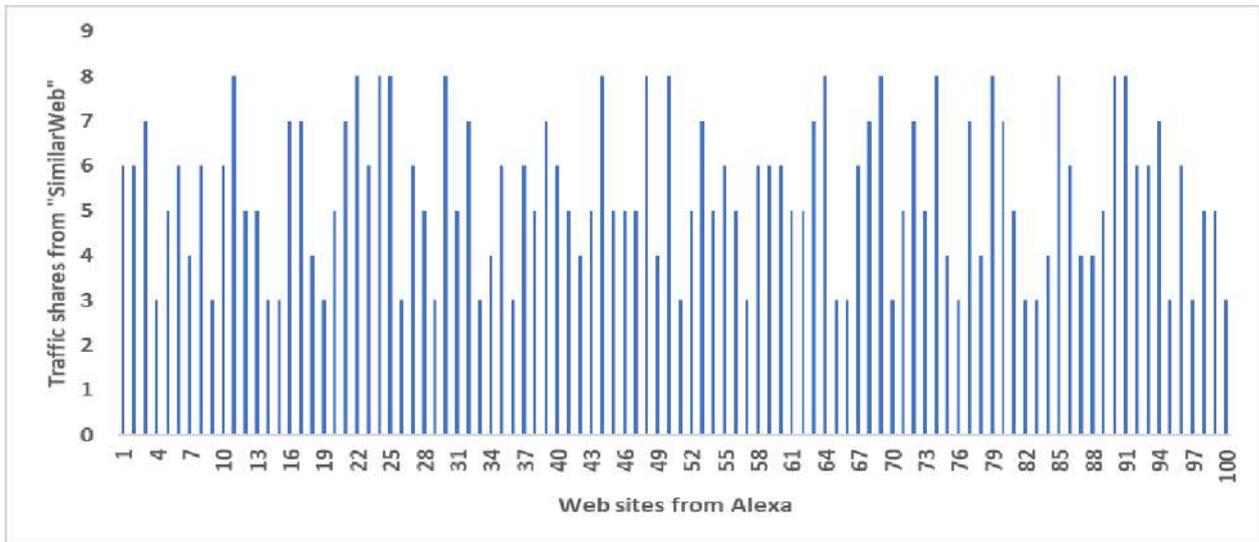
Dataset 7

Dataset 8

Figure 4: Datasets Produced from Alexa Top 500 w=Web Sites

5.2 Experimental Results

Figure 5 shows the page views from "Google Analytics" versus the Importance Value produced by our system (Note that page views from "Google Analytics" are represented as $\log_2$ values.). The data shows potential clustering of the pages, "Terms of Services" had no Importance, while, as expected, "Home" page was mostly visited, and hence the most important. It is observable that different types of Posts, Articles and Categories pages possessed higher Importance Values (I.V.).

The values of I.V. correlate significantly with the page views from "Google Analytics" (r=0.79;

$p<10^{-3}$). Applying a ranking of the page's procedure for both the page views from "Google Analytics" and Importance Value demonstrates a significant rank correlation $\rho=0.91$ ($p<2.2*10^{-16}$). This provides important evidence that the Importance Value is performing similar to other analytic tools.
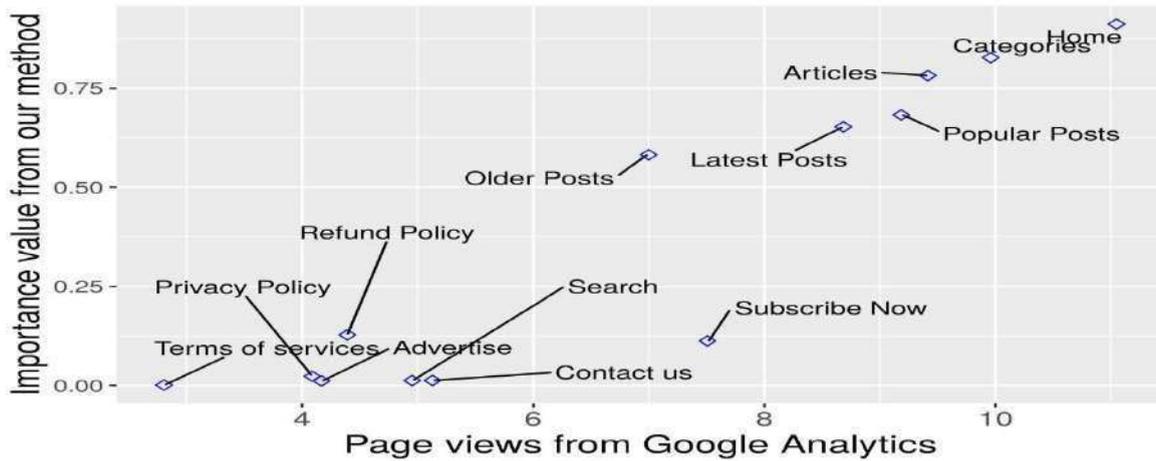
An Automated Web Structure-based Method for Predicting the Importance of a Webpage

Figure 5: Page Views from "Google Analytics" Versus the Importance Value Produced by Our System
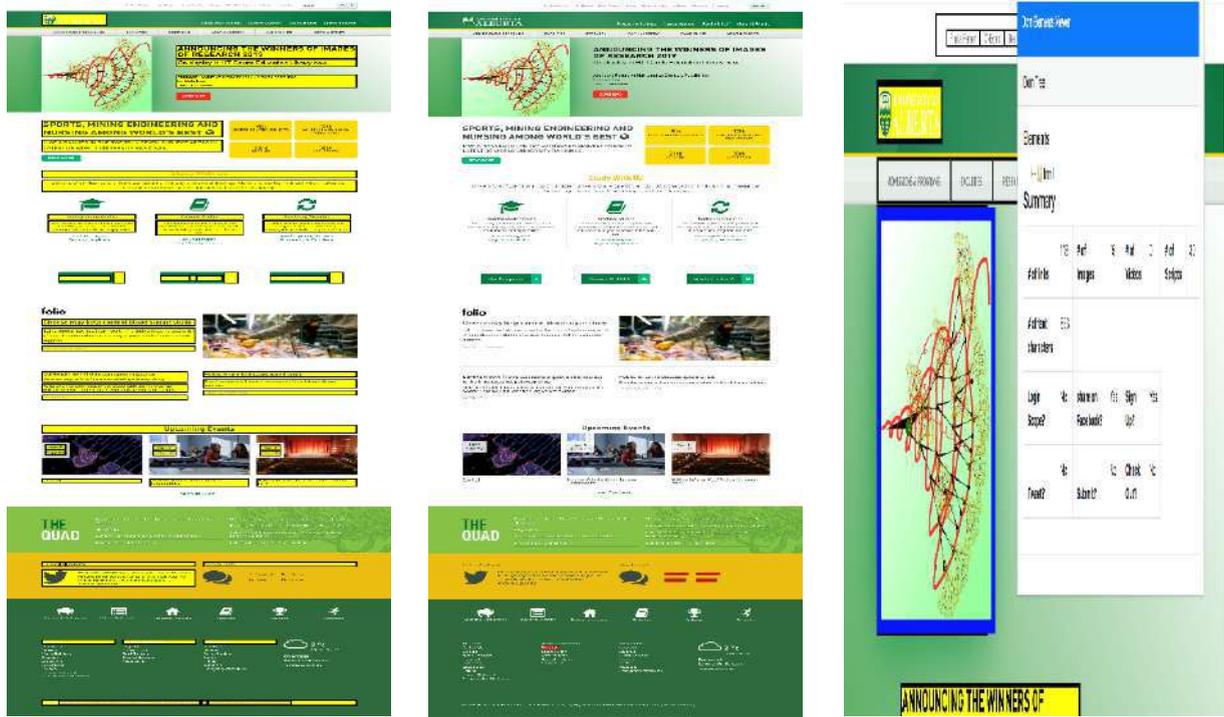
5.3 Case Study

In this section, our proposed work focuses on three case studies. We also use another different website (University of Alberta) for this purpose rather than "Online Book Review" website. The reason for selecting this website is because of the vast amount of work that can be carried out here which gives us sufficient data to analyze.

Case Study-I: In Case-I, we show how our proposed system can extract the website contents.

In figure 6(a) we represent the screenshot of the home pages of the Alberta website. In Figures 6(b) to 6(f) we represent the results. We represent images in 6(b), links in 6(c), and texts in 6(d). In figure 6(e) we show the integrations. From the figure, we see that this page has a scope of registrations, sharable on "Facebook", "LinkedIn", and "YouTube". Finally, in figure 6(f) we represent the total results in the extension window.

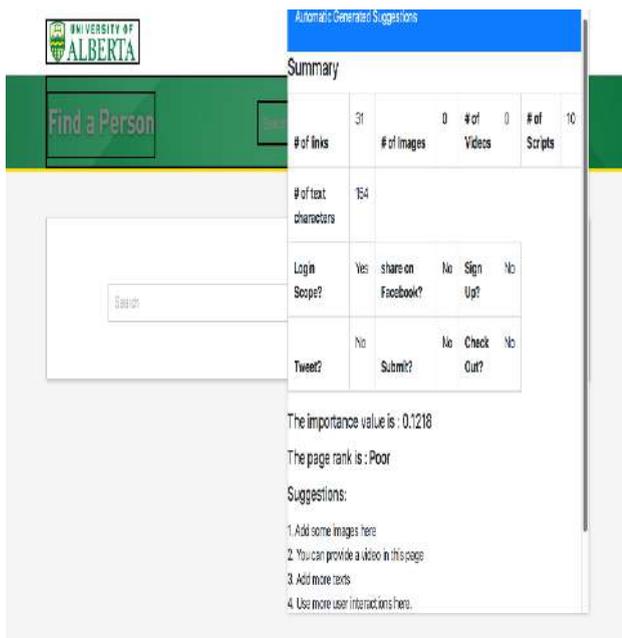Original page   Number of images   Number of links
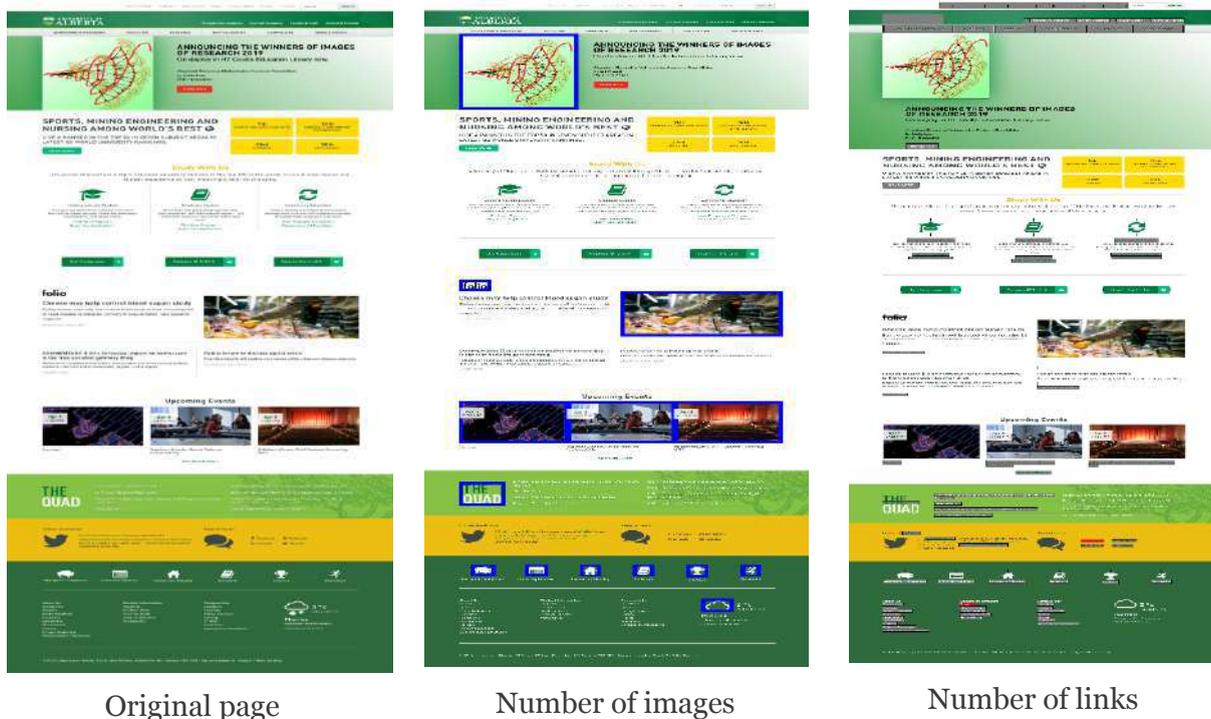
Number of Texts   Interacted with Users   With Results

Figure 6: Output of the Extension for the Homepage of the Website "University of Alberta"
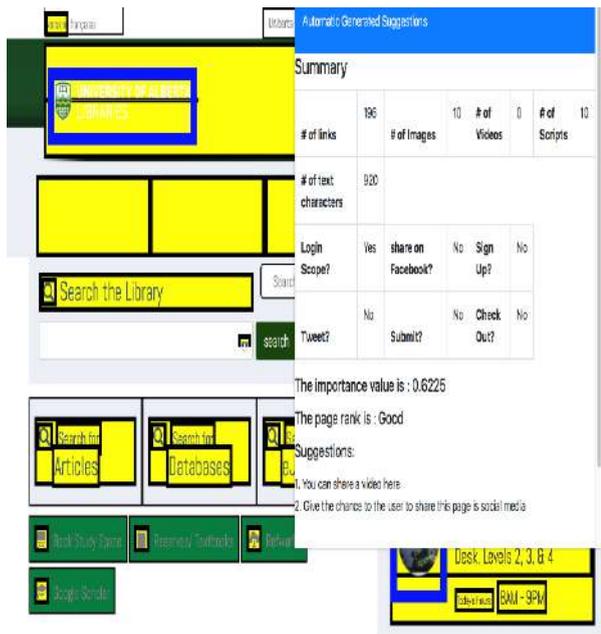
Case Study-II: In this case, we represent four different pages for which our system generates the four different scores; Poor, Average, Good and Best according to their value and importance.
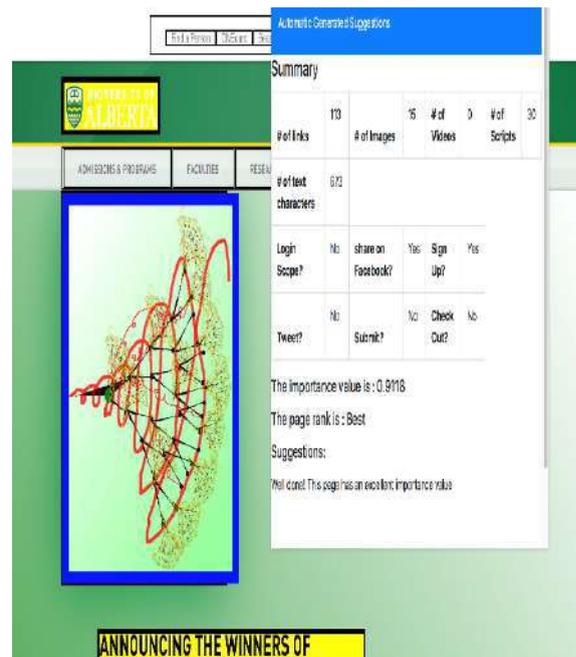

(A) Page with "Poor" Ranking   (B) Page with "Average" Ranking

An Automated Web Structure-based Method for Predicting the Importance of a Webpage

(C) Page with "Good" Ranking   (D) Page with "Best" Ranking

*Figure 7:* Automatic Suggestions Provided by Our Proposed System

In figure 7(a) we see the page "Find a person" where our system produces its score as Poor and produces a very low importance value. We see there are only some texts and one interaction with the user. Our system produces suggestions for it to increase the score. In figure 7(b) the "average" scored web page "Map" is represented. The importance value produced by our system for this page is 0.4125. We see some suggestions here. In figure 7(c) the web page "Library" is represented with the results of our system. The importance value is 0.6225. So there are fewer suggestions for that. The Best rank given by our proposed system

to the home page of the website is represented in figure 7(d). We see that the importance value is 0.9118 and there are no suggestions here. Our system can generate this suggestion automatically. These suggestions are reviewed manually and we find them very effective.

Case Study-III: A case study is also conducted on Online Book Review website. "Contact us" page(https://www.onlinebooksreview.com/contact) is chosen for this case study. 4 web pages version are made. The webpages are then updated in these four versions. Figure 8 denotes the four versions.
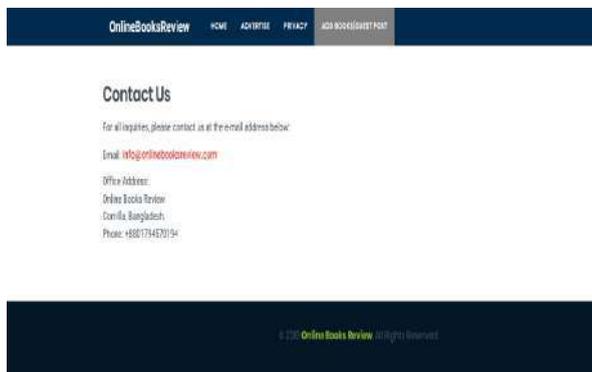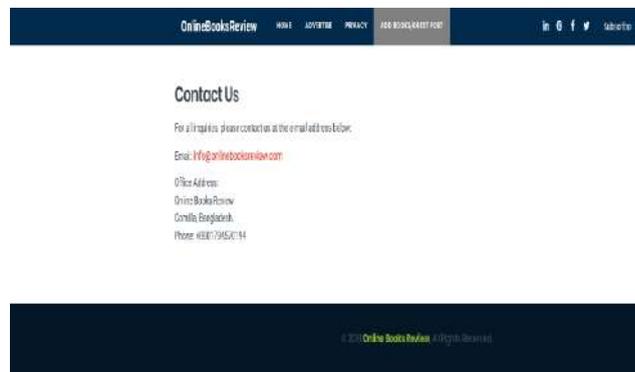
(a) Version-1   (d) Version-2
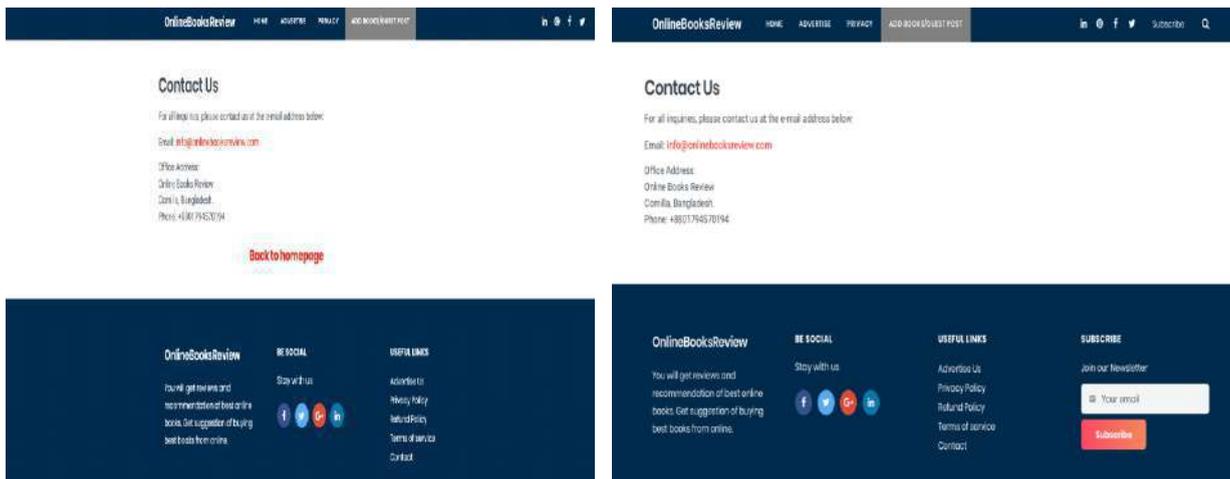
© Version-3   (d) version-4

Figure 8: Different versions of "Contact us" web page of "Online Book Review" website

The case study result is shown in Table 12. We can view the four versions of the features. Also, when we made the "contact us" page more interactive, there is an increase in the page views. So, this case study shows the effectiveness of our work.

Table 12: Page Views of "Contact Us" Web Page of "Online Book Review" Website According to Different Versions

Features	Version number			
	1	2	3	4
Header with basic information only	yes	Yes	Yes	Yes
Header with a sharable link in social media	No	Yes	Yes	Yes
Header with subscribe option	No	Yes	No	Yes
Body with basic information only	Yes	Yes	Yes	Yes
Body with a back link to home page	No	No	Yes	No
Footer with more links only	No	No	Yes	Yes
Footer with a subscribe option with a mail address	No	No	No	Yes
Number of page views (according to "Google Analytics")	5	9	22	25

5.4 Validation of Results

For validation of our work, we use two types of validity; internal and external. In internal validity, we use the confusion matrix to represent the results, and for external validity, we use the correlation matrices; Pearson and Spearman.

5.4.1 Internal Validity

For internal validity, we represent our results in a confusion matrix. To find the internal validity we checked through all 2,938 web pages manually for their features. We extract the source code of all the 2,938 web pages and then check manually all the features and compare them with the automated generated results. We discovered how

our system can find out the images, texts, videos, links, and user interactions efficiently. There are two basic measures used in evaluating the performance of these strategies. They are Precision and Recall. The recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage. On the other hand, Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. We also made use of four other parameters for more accurate analysis and to determine Accuracy, Detection and False Alarm for the extracted content. The parameters are:

An Automated Web Structure-based Method for Predicting the Importance of a Webpage

1. True Positive (TP): The number of pages in which our system discovers where login scope truly exists.
2. True Negative (TN): The number of pages in which our system does not find the login scope where login scope truly exists.
3. False Positive (FP): The number of pages in which our system finds the login scope where login scope does not exist.
4. False Negative (FN): The number of pages in which our system does not find the login scope where login scope does not exist.

Here we give the example on the basis of finding the login scopes in a web page. We follow the same parameters for finding the images, videos, links etc. Then measure the Accuracy, Precision and Recall based on this result. We represent the results for both cases in table 13 to table 17. In the evaluation process at first, we go through each page manually at their source code and see how many images, texts, links, etc. are there. Then we compare these results with our automated generated system. In Dataset 1 to Dataset 3, we use the same web pages. So in each table, we represent the results of the 3 Datasets together.

There are 100 pages in the 3 Datasets, and their design is not changed in the 3 months. So we represent the results together. For Dataset 4 to Dataset 8 we also represent the results in a confusion matrix. From the tables, we observe that our system can successfully extract the web pages' contents.

Table 13: Evaluation for Images of the Web Pages

		TP	FP	TN	FN	Accuracy	Precision	Recall
Online Book review	Dataset 1-Dataset 3	248	6	4	3	96.55%	0.9763	0.9841
Websites from Alexa	Dataset 4	1613	21	32	1	0.9682	0.9805	0.9871
	Dataset 5	2808	23	18	1	0.9933	0.9936	0.9918
	Dataset 6	1342	11	16	2	0.9868	0.9882	0.9918
	Dataset 7	3493	21	29	4	0.9906	0.9917	0.9940
	Dataset 8	1833	12	31	0	0.9834	0.9833	0.9934

Table 14: Evaluation of Videos of the Web Pages

		TP	FP	TN	FN	Accuracy	Precision	Recall
Online Book review	Dataset 1-Dataset 3	8	0	1	96	99.04%	0.8889	1
Websites from Alexa	Dataset 4	13	0	0	481	100%	1	1
	Dataset 5	16	0	1	530	99.81%	0.9411	1
	Dataset 6	18	0	0	633	100%	1	1
	Dataset 7	16	0	0	648	100%	1	1
	Dataset 8	12	0	1	589	99.04%	0.8889	1

Table 15: Evaluation of Links of the Web Pages

		TP	FP	TN	FN	Accuracy	Precision	Recall
Online Book review	Dataset 1-Dataset 3	422	11	5	9	96.42%	0.9882	0.9745
Websites from Alexa	Dataset 4	3182	54	24	44	97.63%	0.9925	0.9833
	Dataset 5	3672	402	27	49	89.66%	0.9927	0.9013
	Dataset 6	4473	490	32	56	89.66%	0.9928	0.9012
	Dataset 7	3823	417	34	62	89.59%	0.9911	0.9016
	Dataset 8	3178	348	22	55	96.42%	0.9882	0.9745

Table 16: Evaluation of Words of the Web Pages

		TP	FP	TN	FN	Accuracy	Precision	Recall
Online Book review	Dataset 1-Dataset 3	8665	256	11	9	97.01%	0.9987	0.9713
Websites from Alexa	Dataset 4	42380	1220	42	51	97.11%	0.9990	0.9720
	Dataset 5	46990	1311	49	62	97.19%	0.9989	0.9728
	Dataset 6	55602	1492	58	58	97.29%	0.9989	0.9738
	Dataset 7	57419	1598	61	65	97.19%	0.9989	0.9729
	Dataset 8	54235	1482	51	56	97.01%	0.9987	0.9713

Table 17: Evaluation of User Interactions on the Web Pages

|		TP	FP	TN	FN	Accuracy	Precision	Recall
Online Book review	Dataset 1-Dataset 3	398	5	2	0	98.27%	0.995	0.9875
Websites from Alexa	Dataset 4	1908	28	482	11	79%	0.7983	0.9855
	Dataset 5	2112	45	392	17	82.96%	0.8434	0.9791
	Dataset 6	2710	62	401	29	85.54%	0.8711	0.9776
	Dataset 7	2882	82	445	21	84.63%	0.8662	0.9723
	Dataset 8	2502	39	312	31	87.82%	0.8891	0.9846

4.2 External Validity

For representing the external validity, we use the two types of the correlation coefficient: Pearson and Spearman. We find the correlation among pairs of variables; the first one is the importance score produced automatically by our proposed system and the second variable is the "page views" results collected from the "Google Analytics" and "SimilarWeb." We use the Dataset 3 and Dataset 8 results to represent the correlation as they are used in our proposed system as testing. Pearson Correlation Coefficient is represented by r, which originally stood for regression. It is a parametric statistical measure of the strength of a linear relationship between two continuous variables. A relationship is linear when a change in one variable is associated with a proportional change in the other variable. Pearson Correlation Coefficient examines the variables concerning their deviations from the mean. On the other hand, Spearman's rank correlation coefficient is a nonparametric rank statistic proposed as a measure of the strength of the association between two variables. It is a measure of a monotone association that is used when the distribution of data makes Pearson's correlation coefficient undesirable or misleading. It assesses how well an arbitrary monotonic function can describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. Unlike the Pearson correlation coefficient, the Spearman correlation coefficient does not require the assumption that the relationship between the variables is linear, nor does it require the variables to be measured on interval scales; it can be used for variables measured at the ordinal level. However, the sign of the correlation tells something about the behavior of the two variables; the absolute value of the correlation indicates how strong the relationship is between these variables. A correlation of 1.0 is a perfect positive correlation, meaning that the two variables move upward or downward together. A correlation of -1.0 is a perfect negative correlation, meaning that the two variables move in opposite directions. The closer the correlation is to 1.0 or -1.0, the stronger the relationship between the two variables. The sign only determines the direction, positive or negative, and it does not influence the strength of the correlation. When there is no linear correlation between the variables, the value of the correlation coefficient would be 0.
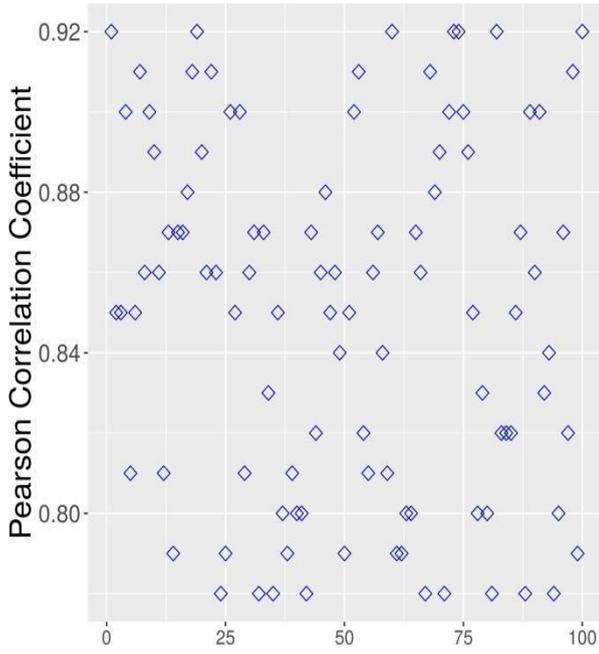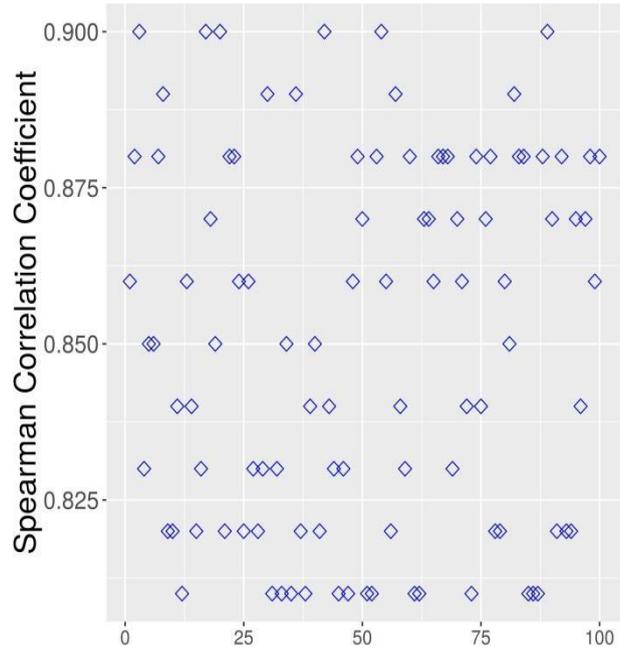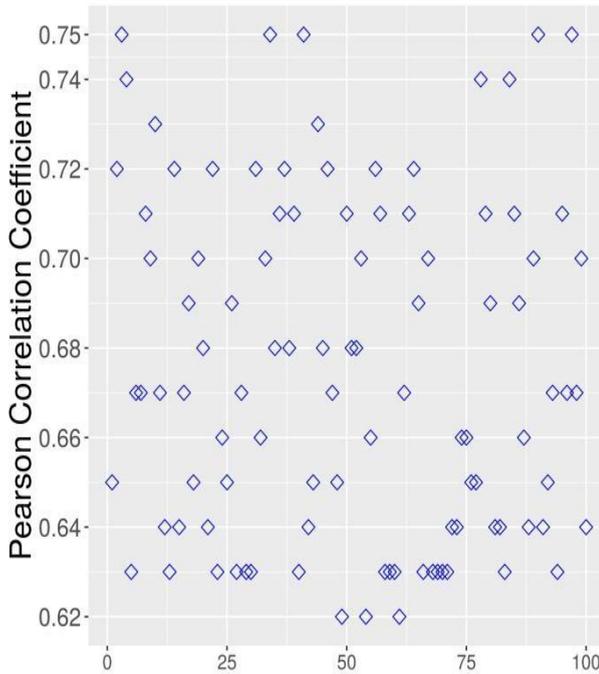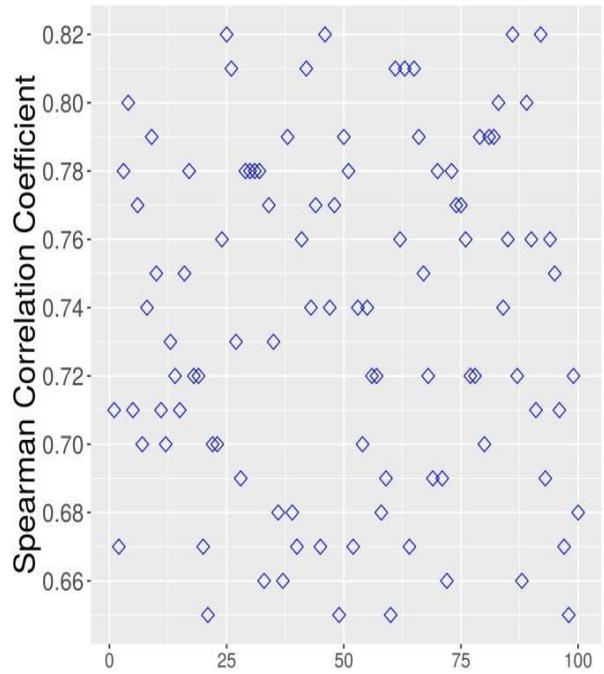
a) The Pearson Correlation   b) The Spearman Correlation

Figure 9: the Correlation Among Pairs of Variables in Our Proposed System's Score and "Google Analytics" Page View

a) The Pearson Correlation   b) The Spearman Correlation

Figure 10: The Correlation Among Pairs of Variables in Our Proposed System's Score and "Similarweb" Page View

We use both cases (The website of "Online Book Review" and the Top websites from "Alexa") for finding external validity. In the case of "Online Book Review" website, we consider the "page view data from the "Google Analytics" for four weeks of December 2021. Suppose for the first web page, we first produce the importance value automatically. According to our system, the importance value will not change in four weeks as the website design is not changed. So we select four different values of page views collected from "Google Analytics" and four unchanged values automatically generated from our proposed system. After that, we use this data to find the Spearman and Pearson Correlation coefficient. In this way, we go through the remaining 99 web pages and generate the correlation values. Figure 9(a) represents the Pearson Correlation coefficient, and Figure 9(b) represents the Spearman Correlation coefficients for the "Online Book Review" websites. Estimation of Pearson correlation coefficients denoted strong correspondence between variables. The values varied between 0.78 and 0.92, with a mean of 0.85. On the other hand, estimation done applying Spearman correlation evidenced a strong correlation between variables. This statistical parameter varied from 0.81 to 0.9. On average its value was 0.85. So our system can find the importance value of "Online Book Review" website successfully. In the case of top websites from "Alexa" we use Dataset 8 where we also find 100 websites. For each website, we generate the importance value of the web pages automatically and then use the results of "SimilarWeb" to compare. In this way, we generate for 100 websites and represent the Pearson Correlation in Figure 10(a) and Spearman in Figure 10(b). In the case of Person Correlation, the correspondence between variables is calculated between 0.62 and 0.75. Association between the variables, expressed by average correlation was 0.67. Spearman coefficients values are estimated as not less than 0.65. The average correspondence between analyzed variables is 0.74. The maximal correlation identified is 0.82. So we can conclude that our system can generate similar results to that of "Google Analytics" and "SimilarWeb."

5.5 State-of-the-art

The research carried out by Thomas et al. [2019] is similar to our research. An experimental study was carried out to assess social commerce's impact on website features in their research. Four versions of a website were created and they use for testing purposes. The feature of the fourth version is richer than the other three versions tested. Below are a few comparative studies between our work and that of Thomas et al. [2019].

- The website's high-level issue was worked on by Thomas et. al. [2019]. This issue varies for different users. Some of the issues considered include:
  1. Perceived satisfaction
  2. Perceived usefulness
  3. Trust
  4. Operation checks items.

  These issues are varied on the website to view users' responses to increasing or reduction. The responses are recorded with a "Yes", "No" or "Unsure". The numbers of clicks, page views per user and time spent were also recorded. However, our model focuses more on the website's low-level features to observe the responses of users to changes in features. Therefore, our research encompasses almost all website features.

- In Thomas et. Al. [2019] experiments, 4 website version was used namely "zero", "low", "medium" and "high" versions. The richness of each feature was in ascending order from zero level to high version. In our work, we chose the selected Alexa top 500 websites from the top 656 websites while the "Online Book Review" website was also considered since we have access to its server.

- They receive 212 participants' feedback with a significant number of them 164 were used and some were discarded. Likewise, we also keep track of web page users' responses through data generated from "SimilarWeb" and "Alexa". For instance, "SimilarWeb" was able to track about 1 Million Amazon website users. This gives us a robust amount of real-time participants.

London Journal of Engineering Research

VI.   SUMMARY & CONCLUSION

Web applications have infiltrated almost every aspect of our daily life. Research shows that 93% of online shopping starts from websites search. Therefore, to capture the online marketplace places the advertisement provider needs to know the right place to set up their ad so that most of the website users can see their ad. There are lots of web applications available capable of fulfilling this purpose but most of them use web users' private data. So, when users close their browser cookies information the web applications won't be able to get accurate results. The main highlight and fascinating aspect of our work are that it works on the website structure to predict the importance of the web page. It consists of two very helpful features for both web developers and advertisement providers.

- In the case of an Advertisement provider, our proposed system can show the importance score alongside the web pages' rank so that they can take a quick decision to include their advertisement in real-time. No user private data is needed.

- In the case of web developers that sometimes publish their trial version and later use feedback gotten from the users to update their web application. Our system can give them real-time suggestions with the importance score so they can design a better website in the development period.

For solving the problem, we extract the features from web pages in real-time and use CatBoost Machine learning to create the rank. We do not only use the web pages' contents (such as the number of images, number of videos, number of links, number of texts, etc.) but we also use the web page accessibility and connectivity with other web pages. To validate our work, we use two types of datasets; one is collected from the server of the "Online Book Review" website and another we prepare from the most popular 500 websites from Alexa. We represent our effectiveness in the format of case studies, confusion matrix and correlation coefficient. In all formats, our good results prove the effectiveness of our proposed system.

REFERENCES

1. B. Abrahao, S. Soundarajan, J. Hopcroft, and R. Kleinberg (2012), On the separability of structural classes of communities, in Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, ACM, New York, pp. 624–632.

2. D. Aldous and J. A. Fill (2002), Reversible Markov Chains and Random Walks on Graphs, unfinished monograph; recompiled 2014, available online from http://www.stat.berkeley.edu/ ~aldous/RWG/book.html.

3. S. Allesina and M. Pascual (2009), Googling food webs: Can an eigenvector measure species importance for coextensions? PLoS Comput. Biol., 5, e1000494.

4. R. Andersen, F. Chung, and K. Lang (2006), Local graph partitioning using PageRank vectors, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, IEEE, pp. 475–486.

5. W. N. J. Anderson and T. D. Morley (1985), Eigenvalues of the Laplacian of a graph, Linear Multilinear Algebra, 18, pp. 141–145.

6. A. Arasu, J. Novak, A. Tomkins, and J. Tomlin (2002), PageRank computation and the structure of the web: Experiments and algorithms, in Proceedings of the 11th International Conference on the World Wide Web, Poster session. www2002.org/COROM/poster.173.pdf.

7. K. Avrachenkov, N. Litvak, and K. S. Pham (2007), Distribution of PageRank mass among principle components of the web, in Proceedings of the 5th Workshop on Algorithms and Models for the Web Graph (WAW2007).

8. A. Bonato and F. C. Graham, eds., Lecture Notes in Comput. Sci. 4863, Springer, New York, pp. 16–28.

9. K. Avrachenkov, B. Ribeiro, and D. Towsley (2010), Improving random walk estimation accuracy with uniform restarts, in Algorithms and Models for the Web-Graph, R. Kumar and D. Sivakumar, eds., Lecture Notes in Comput. Sci. 6516, Springer, Berlin, Heidelberg, pp. 98–109.

10. L. Backstrom and J. Leskovec (2011), Supervised random walks: Predicting and recommending links in social networks, in Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, ACM, New York, pp. 635–644.

11. R. Baeza-Yates, P. Boldi, and C. Castillo (2006), Generalizing PageRank: Damping functions for link-based ranking algorithms, in Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2006), Seattle, WA, ACM, New York, pp. 308–315.

12. B. Bahmani, A. Chowdhury, and A. Goel (2010), Fast incremental and personalized PageRank, Proc. VLDB Endow., 4, pp. 173–184.

13. A. Balmin, V. Hristidis, and Y. Papakon-stantinou (2004), ObjectRank: Authority-based keyword search in databases, in Proceedings of the Thirtieth International Conference on Very Large Data Bases, Volume 30, VLDB '04, VLDB Endowment, pp. 564–575.

14. Z. Bar-Yossef and L.-T. Mashiach (2008), Local approximation of PageRank and reverse Page-Rank, in CIKM '08: Proceedings of the 17th ACM conference on Information and Knowledge Management, ACM, New York, pp. 279–288.

15. D. S. Bassett and E. Bullmore (2006), Small-world brain networks, The Neuroscientist, 12, pp. 512–523.

16. M. Bayati, D. F. Gleich, A. Saberi, and Y. Wang (2013), Message-passing algorithms for sparse network alignment, ACM Trans. Knowledge Discovery. Data, 7, pp. 3:1–3:31.

17. L. Becchetti, C. Castillo, D. Donato, R. Baeza-Yates, and S. Leonardi (2008), Link analysis for web spam detection, ACM Trans. Web, 2, pp. 1–42.

18. M. Benzi, E. Estrada, and C. Klymko (2013), Ranking hubs and authorities using matrix functions.

19. P. Berkhin (2005), A survey on PageRank computing, Internet Math., 2, pp. 73–120.

20. A. Berman and R. J. Plemmons (1994), Nonnegative Matrices in the Mathematical Sciences, Classics Appl. Math. 9, SIAM, Philadelphia.

21. M. Bianchini, M. Gori, and F. Scarselli (2005), Inside PageRank, ACM Trans. Internet Technologies, 5, pp. 92–128.

22. D. A. Bini, G. M. D. Corso, and F. Romani (2010), A combined approach for evaluating papers, authors and scientific journals, J. Computer. Applied Mathematics, 234, pp. 3104–3121.

23. D. M. Blei, A. Y. Ng, and M. I. Jordan (2003), Latent Dirichlet allocation, J. Mach. Learn. Res., 3, pp. 993–1022.

24. V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren (2004), A measure of similarity between graph vertices: Applications to synonym extraction and web searching, SIAM Rev., 46, pp. 647–666.

25. P. Boldi (2005), TotalRank: Ranking without damping, in Poster Proceedings of the 14th International Conference on the World Wide Web (WWW2005), ACM Press, New York, pp. 898–899.

26. P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna (2008), The query-flow graph: Model and applications, in Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM '08, ACM, New York, pp. 609–618.

27. P. Boldi, F. Bonchi, C. Castillo, and S. Vigna (2009a), Voting in social networks, in Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, ACM, New York, pp. 777–786.

28. P. Boldi, F. Bonchi, C. Castillo, and S. Vigna (2011), Viscous democracy for social networks, Commun. ACM, 54, pp. 129–137.

29. P. Boldi, R. Posenato, M. Santini, and S. Vigna (2007), Traps and pitfalls of topic-biased PageRank, in Fourth International Workshop on Algorithms and Models for the Web-Graph, WAW2006, Lecture Notes in Comput. Sci., Springer-Verlag, New York, pp. 107–116.

30. P. Boldi, M. Santini, and S. Vigna (2005), PageRank as a function of the damping factor, in Proceedings of the 14th International Conference on the World WideWeb (WWW2005), Chiba, Japan, ACM Press, New York, pp. 557–566.

31. P. Boldi, M. Santini, and S. Vigna (2009), PageRank: Functional dependencies, ACM Trans. Inf. Syst., 27, pp. 1–23.

32. J. Bollen, M. A. Rodriquez, and H. Van de Sompel (2006), Journal status, Scientometrics, 69, pp. 669–687.

33. S. B. Brawer, M. Ibel, R. M. Keller, M. Shivakumar (2016), Web Crawler Scheduler that utilizes Sitemaps from Websites, United States Patent.

34. T. Friedrich, S. Schlauderer, S. Overhage (2019), The impact of social commerce feature richness on website stickiness through cognitive and affective factors: An experimental study, Electronic Commerce Research and Applications, 36(2019).

35. 1] Lundberg S. M., Lee Su-In. A Unified Approach to Interpreting Model Predictions. 2017.http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

36. Lundberg S. M., Erion G. G., Lee Su-In. Consistent Individualized Feature Attribution for Tree Ensembles. 2017. https://arxiv.org/pdf/1802.03888.pdf.

37. Khun S. W., Petzer D. J., Fostering Purchase Intentions Toward Online Retailer Websites in an Emerging Market: An S-O-R Perspective, Journal of Internet Commerce (2018), Volume 17 Issue 3, page 255-282.

Fault Analysis of Inverter in Monopolar HVDC System

Olumoroti Ikotun & Suleiman Tijjani Shehu

Federal University Oye-Ekiti

ABSTRACT

One of the different kinds of electrical energy is high voltage direct current (HVDC). Some previous researchers work on the Monopolar HVDC systems using Matlab/Simulink environment to carry out their analyses. The analyses that are carried out include single phase to ground fault, double phase to ground fault and three phase to ground fault. It is discovered, among others, that for single phase to ground fault on AC side of the inverter the three phase voltages and three phase currents show no effect except at the rectifier side. But it is clear in our present findings that with reference to single phase to ground fault, the three phase voltages have transient and oscillation waveforms that exclude phase A that decreases to zero and the three phase currents that have a high unstable waveform. We further show that the causes of increase in DC transmission line depend on the severity of the fault that is implemented at the inverter side. There fore, transmission network is extremely important and engineers can use this result when doing their careful analyses.

Keywords: fault analysis, monopolar hvdc, rectifier, inverter, simulation.

Classification: DDC Code: 621.31 LCC Code: TK153

Language: English

Fault Analysis of Inverter in Monopolar HVDC System

Olumoroti Ikotun[α] & Suleiman Tijjani Shehu[σ]

ABSTRACT

One of the different kinds of electrical energy is high voltage direct current (HVDC). Some previous researchers work on the Monopolar HVDC systems using Matlab/Simulink environment to carry out their analyses. The analyses that are carried out include single phase to ground fault, double phase to ground fault and three phase to ground fault. It is discovered, among others, that for single phase to ground fault on AC side of the inverter the three phase voltages and three phase currents show no effect except at the rectifier side. But it is clear in our present findings that with reference to single phase to ground fault, the three phase voltages have transient and oscillation waveforms that exclude phase A that decreases to zero and the three phase currents that have a high unstable waveform. We further show that the causes of increase in DC transmission line depend on the severity of the fault that is implemented at the inverter side. There fore, transmission network is extremely important and engineers can use this result when doing their careful analyses.

Author α: Department of Electrical and Electronics Engineering, Federal University Oye-Ekiti, Oye, Nigeria.
σ: Department of Electrical and Electronics Engineering, Federal University Oye-Ekiti, Oye, Nigeria.

I.   INTRODUCTION

Electrical energy is the fundamental, the fact of being necessary for the production, distribution, the use of income wealth, and commodities for the development or progress of a country. It is not possible to make an estimate the great importance of electrical energy has done in our present day civilization. In fact or in reality, in this present world, the state of relying on electricity is high and it has become a portion of our life. Therefore, we can say that electrical power system is the backbone for the progress of any country in all areas. Similarly, another area of electrical energy is HVDC. This technology is mature and generally approved technology. HVDC is well known as an adequate power system. The HVDC can accomplish as well as being capable of sending sufficient power over several distances. The entire achievements that are used in HVDC rely on the electronic devices such as Gate Turn Off thyristors (GTOs), Insulated Gate Bipolar Transistors (IGBTs), thyristors, semiconductors devices, Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) and mercury arc valves [1] [2] [3]. In addition, HVDC is principally an overhead transmission line over a long distance. In case, there is a fault in the transmission line it will result to total blackout. Therefore, fault analysis will be of great importance to the electric power systems. Some previous researchers work on the HVDC Monopolar systems. For example, [4] Present faults on HVDC Monopolar transmission line such as a performance of the converter, DC faults (voltage & current) and line to ground fault on AC side. They use Matlab/Simulink environment to model and simulate. In their conclusion, it shows that during faulty conditions the voltages waveforms have slight decrease and currents decreases from its standard value [5] work on DC link line to line fault and line to ground fault using Matlab/Simulink environment.

The result shows that when fault occurs on the DC line it decreases and after fault it is hard to go back to its original state. [6] too carried out a proposed model with broad simulation. They carry out their research work by studying suppression of temporary overvoltage during the

London Journal of Engineering Research

various fault conditions such as DC faults, three phase AC fault, single phase AC fault and double phase AC fault and fault recovery using Matlab/ Simulink environment. The result shows that optimal PI controller is good on transient performances. Similarly, [7] present transient characteristics of DC line to line fault of HVDC Monopolar system. The model of HVDC Monopolar system was carried out using PSCAD/MATLAB. In their result the DC voltage leads to unstable which is hard to rebalance. [8] too simulate a pole to ground fault on a DC link which results to a low steady state DC fault current and further discovers to have a healthy pole it involves large overvoltage. In addition to their work, they also propose a new strategy of fault protection in HVDC grids. Authors such as [9] focus on pole rebalancing in Monopolar HVDC of DC side. In their paper, they propose post-fault restoration and fault clearing for pole rebalancing in order to deal with DC faults that result to pole-to ground fault. The test is carried out using one of the popular software which is PSCAD/ EMTDC. In this paper, we want to continue the discussion of Monopolar HVDC system using Matlab/Simulink environment with accurate detection to differentiate some fault characteristics when fault is applied at the AC side of the inverter.

II. TYPES OF HVDC SYSTEMS

2.1 Monopolar link

It has a single negative polarity connector and closes the current loop with the earth or the sea. In a Monopolar connection, there are two converters at the end of each pole. The grounding distance between the poles with the substation is fixed at 15-55 kilometers [10].

2.2 Bipolar link

It has two conductors, one of which is negative and the other is positive in relation to the earth or the sea. Each terminal station's converter midpoints are earthed through electrode wires. The voltage between the conductors is double to the voltages between any two conductors and ground [11]. In HVDC system, it is commonly used.

2.3 Homopolar link

This link always operates with the earth or metallic return and has two connectors of the same polarity of the electrodes, which is generally negative. The poles are operated in parallel in a homopolar connection at the moment homopolar system is not in use [12].

III. MATLAB/SIMULINK MODEL OF HVDC MONOPOLAR SYSTEM

The HVDC system is modeled by Matlab/ Simulink and Simpower tools in order to determine the fault of the Monopolar HVDC (high voltage direct current) system. A frequency of 60Hz of voltage source feeds a frequency of 50Hz with a load of 50kW through an AC to DC to AC converter. The voltage of 380V is from the secondary transformer of the wye-delta which is first rectified by the six pulse diode universal bridge. The insulated-gate bipolar transistor (IGBT) of the inverter side uses PWM at a 2000Hz carrier frequency (f_c).
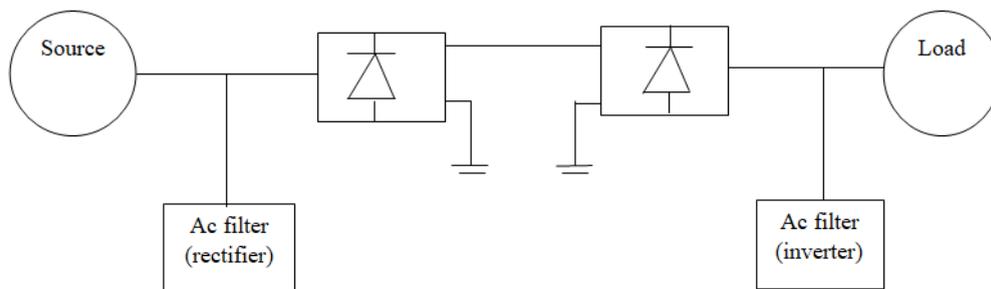

Figure 1: Model of HVDC Monopolar system

IV.   RESULTS AND DISCUSSION

4.1   *Fault Simulation to Predict Actual Behaviour*

The leading importance of this paper is to distinguish different fault conditions when they are applied at the AC side of the inverter and their behaviour on the DC line and at the rectifier side of the source. In this part, it is highly necessary for the engineers to identify where the fault is coming from HVDC power system transmission lines for quick repair so that it can go back to its continuous steady state.

4.2   Three Phase to Ground Fault (Inverter)

The three phase to ground fault (LLLG) is the most grievous fault when compared with the other different faults. At the point when the three phase to ground fault happens, the AC voltages and currents at station 2 (inverter side) which collapse to 0 are shown in figure 2a. The DC voltage will increase with oscillation during fault as shown in figure 2b and at station 1(rectifier side), the AC voltages will increase from their original values and the currents will decrease from their normal values. Both are shown in figure 2c
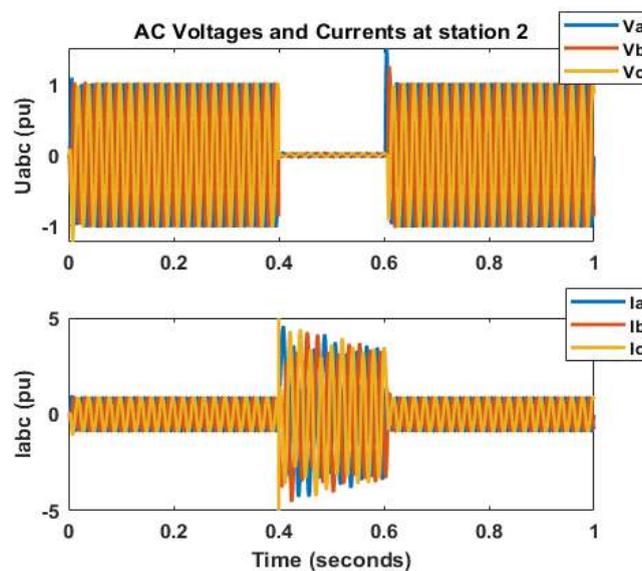

Figure 2a: Simulation Results for Three Phase Voltages and Currents During the Three Phase to Ground Fault (LLLG) at Inverter Side
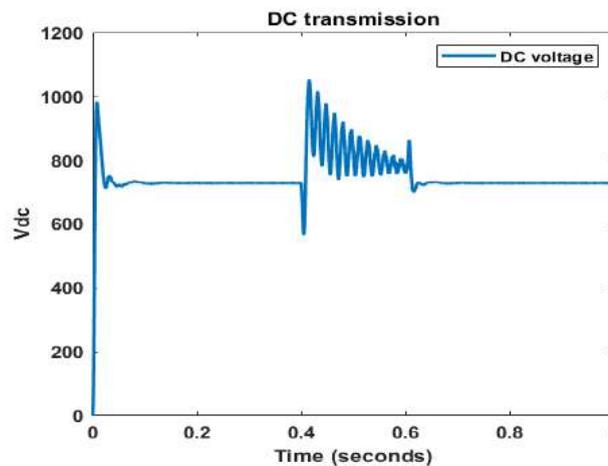

Figure 2b: Simulation Result for DC Voltage During the Three Phase to Ground Fault (LLLG) at Inverte

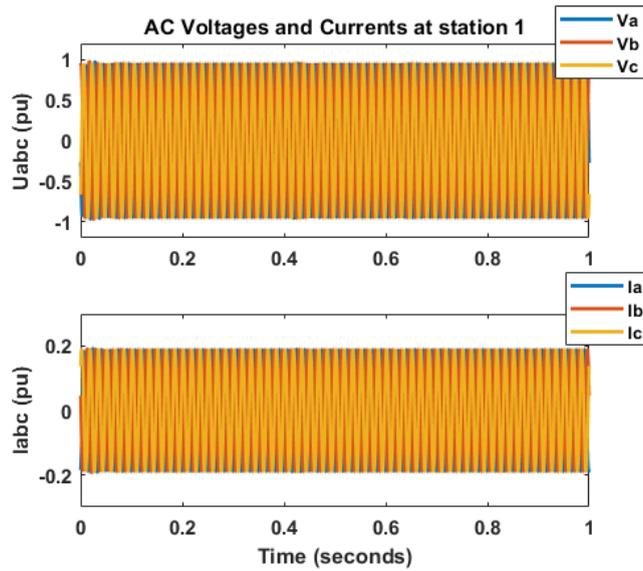Fault Analysis of Inverter in Monopolar HVDC System

Figure 2c: Simulation results for three phase voltages and currents station 1(rectifier side) during the three phase to ground fault (LLLG) at inverter side

4.3 Double Phase to Ground Fault (Inverter)

In the simulation result, the fault was applied at the station 2 (inverter) with duration of 0.2s. The AC voltages (A & B) and currents that become unstable are shown in figure 3a.The DC voltage is affected and increase in waveform during the fault is shown in figure 3b while the AC voltages and currents are also affected at station 1. The AC voltages slightly increase from the original values while the currents decrease with false waveforms. Both are shown in figure 3c.
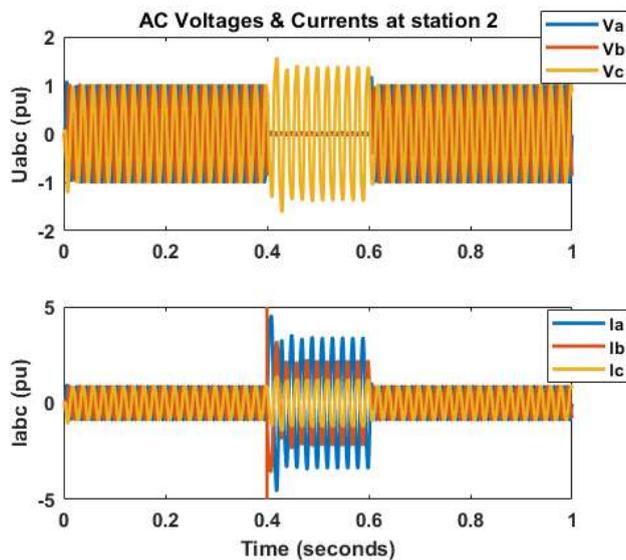

Figure 3a: Simulation Results for Three Phase Voltages and Currents During the Double Phase to Ground Fault (LLG) at Inverter Side
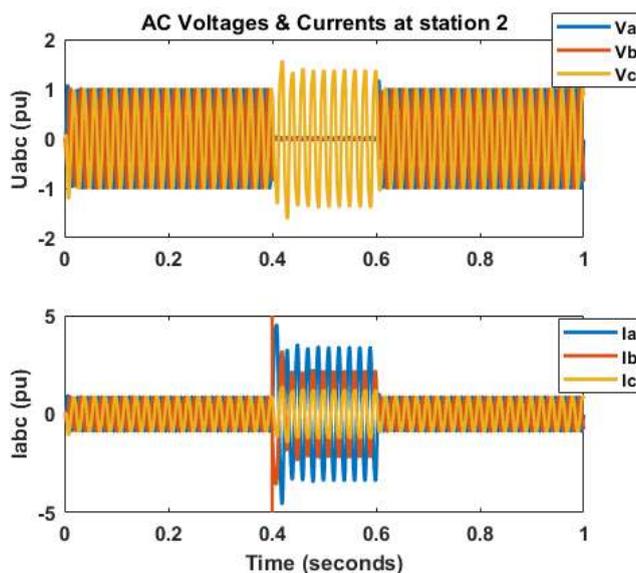
London Journal of Engineering Research

Figure 3b: Simulation Result for DC Voltage During the Double Phase to Ground Fault (LLLG) at Inverter Side
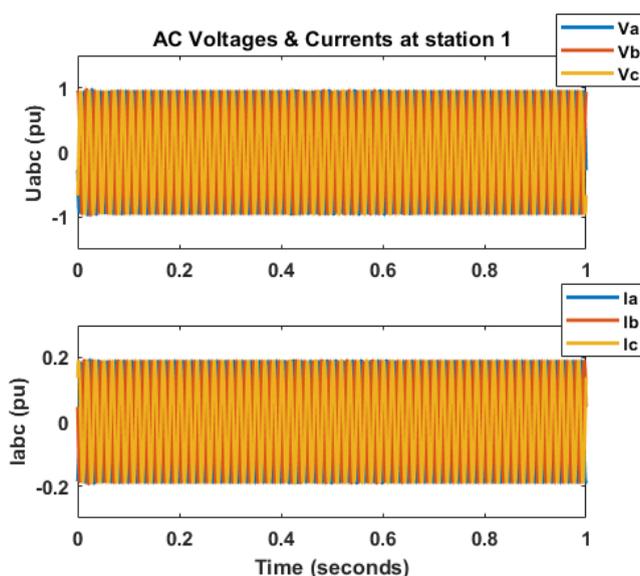

Figure 3c: Simulation Results for Three Phase Voltages and Currents at Station 1(rectifier Side) During the Double Phase to Ground Fault (LLG) at Inverter Side

4.4 Single Phase to Ground Fault (Inverter)

During the fault at 0.2s, the voltage phase A will collapse to 0 and the other phases will maintain unstable sinusoidal waveform and the currents will also show unstable waveform and this state is shown in figure 4a. Meanwhile, the DC transmission is affected having an increase in waveform as shown in figure 4b. The sending end, the AC voltages and currents that have slight increase waveforms are shown in figure 4c.

Fault Analysis of Inverter in Monopolar HVDC System

Figure 4a: Simulation Results for Three Phase Voltages and Currents at Inverter Side During the Single Phase to Ground Fault (LG) at Inverter Side

Figure 4b: Simulation Result for DC Voltage During the Single Phase to Ground Fault (LG) at Inverter Side
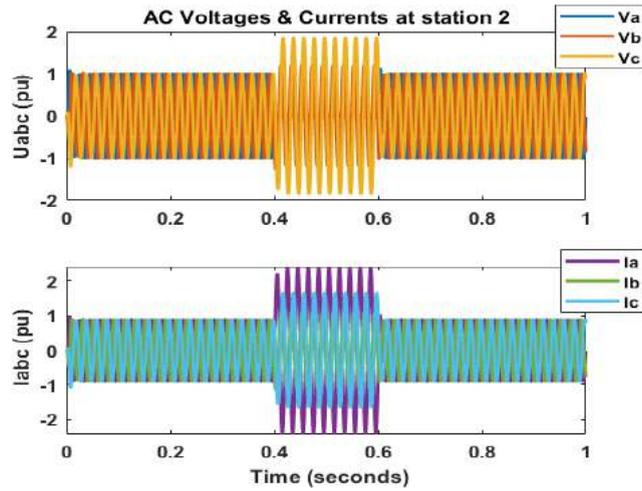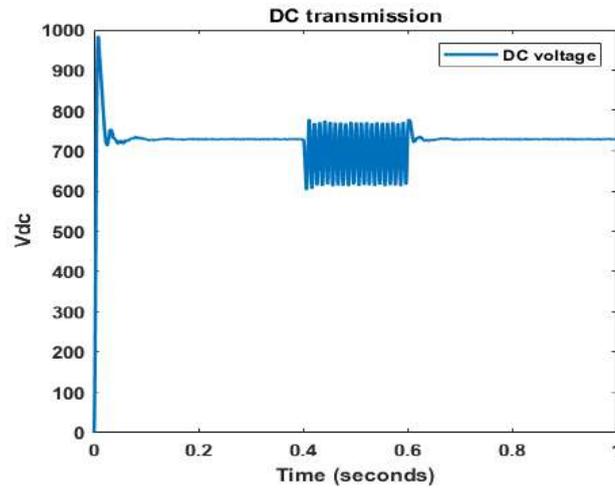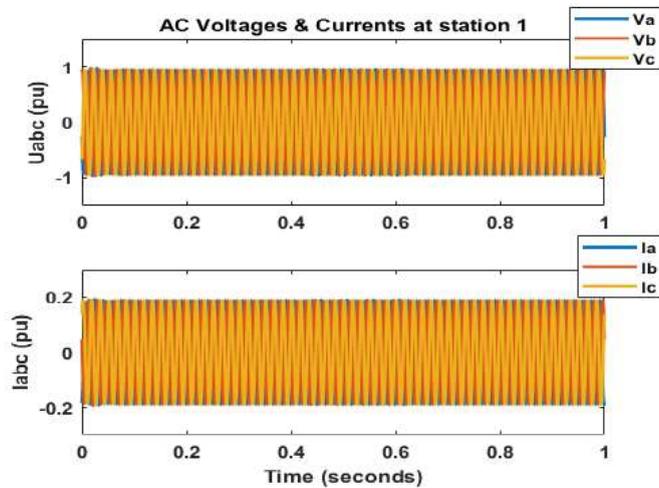

Figure 4c: Simulation Results for Three Phase Voltages and Currents at Rectifier Side During the Single Phase to Ground Fault (LG) at Inverter Side

V. CONCLUSION

In this paper, detection of fault behaviours in High voltage direct current (HVDC) system is carried out due to the alarming rate of power failure. The analyses that are done include single phase to ground fault, double phase to ground fault and three phase to ground fault. The model was carried out using Matlab/Simulink. [4] Shows that when fault was applied on single phase to ground fault at the inverter side on AC side, the voltage would decrease. But it is clear in our present findings that with reference to single phase to ground fault, the three phase voltages have transient and oscillation waveforms that exclude phase A that decreases to zero and the three phase currents that have a high unstable waveform. We further show that the causes of increase in DC transmission line depend on severity of the fault that was implemented at the inverter side. Therefore, transmission network is of extreme importance and engineers can use this result when doing their careful analyses.

REFERENCES

1. Tuan A. Le "Operation of HVDC Grids in Parallel with AC Grids" Master of Science Thesis in Electric Power Engineering, Department of Energy and Environment Division of Electric Power Engineering,2010.\

2. Power Semiconductor Devices, Available online, http://www.allaboutcircuits.com/technical-articles/areview-on-power-semiconductor-devices/

3. Yidong Hong and Tao Yu "Reliability Improvement Strategies for HVDC Transmission System" Energyand Power Engineering, 2013, 5, 52-56.

4. Naveen Gaur, Om Prakash Mahela and Ram Niwash Mahia. "Analysis of faults on Monopolar hvdc transmission line". International journal of electrical and electronics engineering (IJEEE), vol. 4, issue 1, 11-18, 2014.

5. Manish Kumar, Manjeet and Khatri, Pooja. "Study of Faults on HVDC Transmission Lines". Golden research Thoughts, volume 3, issue 8, 2014.

6. Singaravelu, S., and S. Seenivasan. "Modelling and Simulation of Monopolar HVDC Transmission System Feeding a Strong AC Network with Firefly Algorithm based Optimal PI controller." International Journal of Computer Applications 102.10 (2014).

7. Ashwini k. khairnar,p. j. shah, "DC line-to-line fault analysis for VSC based hvdc transmission system", International Journal of Advance Research in Science and Engineering (IJARSE), Vol. No. 6, Issue No. 07, July 2017.

8. Bertinato, Alberto, Pascal Torwelle, Dantas De Freitas, Manuel Colmenero, and Bertrand Raison. "Pole-to-ground fault protection strategy for HVDC grids under symmetrical monopolar configuration." In 2019 IEEE Milan PowerTech, pp. 1-6. IEEE, 2019.

9. Wang, Mian, Willem Leterme, Geraint Chaffey, Jef Beerten, and Dirk Van Hertem. "Protection and pole voltage rebalancing for pole-to-ground faults in symmetrical monopolar HVDC grids." (2019): 43-6.

10. S. Wenig, M. Goertz and J. Prieto "Effects of DC fault clearance methods on transients in a full-bridge Monopolar MMC-HVDC link" Innovative Smart Grid Technologies - Asia (ISGT-Asia), 2016 IEEE, ISBN: 978-1-5090-4302-6.

11. Dirk Van Hertem "HVDC grid layouts" Wiley-IEEE Press,2016 ISBN: 9781119115243.

12. Prashant Agnihotri, A. K. and A. Gole "A Robust Wide-Area Measurement Based Damping Controller for Networks with Embedded Multi-Terminal and Multi-Infeed HVDC Links" IEEE Transactions on Power Systems, 2017, ISSN: 0885-8950.

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

Luis Edo García Jaimes

ABSTRACT

This paper introduces the design of a Linear Quadratic Gaussian Controller (LQG) to regulate the voltage of a DC generator and compares its performance with that of a self-adjusting regulator (STR) tuned by pole placement. To implement these control strategies, the prototype of a DC generator was used and the mathematical model of the system was obtained using nonparametric identification, the theoretical basis and design of the two control strategies are presented and the performance of the two controllers is compared using integral error and temporal response metrics. Finally, the results obtained are presented from which the good performance of both control strategies is inferred.

London Journals Press

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

Diseño De Un Controlador Lineal Cuadrático Gaussiano Y Un Controlador Str Por Asignación De Polos Para El Control Del Voltaje De Un Generador Dc

Luis Edo García Jaimes

RESUMEN

En este artículo se presenta el diseño de un controlador lineal cuadrático gaussiano (LQG) para regular el voltaje de un generador DC y se compara su desempeño con el de un regulador autoajustable (STR) sintonizado por asignación de polos. Para para implementar estas estrategias de control se utilizó el prototipo de un generador DC y se obtuvo el modelo matemático del sistema utilizando identificación no paramétrica, se presentan las bases teóricas y el diseño de las dos estrategias de control y se realiza la comparación del desempeño de los dos controladores utilizando métricas de la integral del error y respuesta temporal. Finalmente, se presentan los resultados obtenidos a partir de los cuales se puede deducir el buen desempeño de ambas estrategias de control.

Palabras clave: control óptimo, filtro de kalman, controlador lqr, controlador str, función de costo.

ABSTRACT

This paper introduces the design of a Linear Quadratic Gaussian Controller (LQG) to regulate the voltage of a DC generator and compares its performance with that of a self-adjusting regulator (STR) tuned by pole placement. To implement these control strategies, the prototype of a DC generator was used and the mathematical model of the system was obtained using nonparametric identification, the theoretical basis and design of the two control strategies are presented and the performance of the two controllers is compared using integral error and temporal response metrics. Finally, the results obtained are presented from which the good performance of both control strategies is inferred.

Keywords: optimal control, kalman filter, LQR controller, STR controller, cost function.

I. INTRODUCCIÓN

La aplicación intensiva de las técnicas del control automático de procesos se basa en la evolución de las tecnologías de medición y control aplicadas al ambiente industrial. El crecimiento en la complejidad de los procesos industriales ha traído como consecuencia que el control automático de los mismos se haya desarrollado a gran velocidad, que los modos de control convencional estén siendo reemplazados por controladores óptimos como una estrategia clave para mejorar la competitividad de la empresa y alcanzar objetivos que de otro modo difícilmente podrían conseguirse.

En la literatura se encuentran diferentes publicaciones que hacen referencia al control óptimo y específicamente al control lineal cuadrático Gaussiano (LQG). En [1] se describe la implementación de un controlador LQG y de un regulador lineal cuadrático (LQR) para controlar la dinámica del vuelo lateral y longitudinal en la simulación del vuelo de una aeronave con el fin de lograr una estabilidad robusta. En trabajo realizado por [2] se presenta el diseño de dos estrategias de control, un controlador

LQG y un controlador DMPC (Discrete Time Model Predictive Control) para regular la velocidad de una turbina de vapor acoplada a un generador DC, con el fin de mantener constante la velocidad ante variaciones de presión en la tubería de vapor.

Rodríguez, Murilo y Silva [3], investigan el comportamiento de un sistema de dirección eléctrica asistida al aplicar un controlador LQG discreto, implementan este controlador debido a su estabilidad y robustez aun cuando el sistema se encuentra sometido a ruidos y perturbaciones. En [4] se implementa una plataforma experimental de un helicóptero de dos grados de libertad pivotado, se estima el modelo matemático que describe el comportamiento del sistema, y se diseña un controlador PID y un controlador LQR.

Finalmente, se realiza la comparación de los controladores para determinar su desempeño frente a perturbaciones. [5], hace referencia al diseño de un algoritmo de control óptimo para el sistema HVAC de un edificio e introduce un algoritmo de control jerárquico que se compone de controladores PID y un controlador LQR. El controlador LQR minimiza una función de coste cuadrático que tiene dos términos cuadráticos: uno tiene en cuenta el nivel de confort y el otro representa el esfuerzo de control, es decir la cantidad de energía consumida. En [6] se propone un controlador LQG de velocidad para suprimir las vibraciones torsionales en el sistema de accionamiento de un motor con una carga de dos masas giratorias. Debido a la dificultad de medir la velocidad de la carga y el par del eje, se utiliza el filtro de Kalman como un estimador de orden completo. Se realizan simulaciones para comparar el desempeño del controlador LQG con un controlador PI convencional. [7] presenta un artículo en el cual realiza el diseño de un algoritmo de control basado en la optimización de los parámetros del controlador LQG mediante el uso de algoritmos genéticos (GA) para controlar la velocidad del motor de DC. Los resultados obtenidos muestran que el enfoque seleccionado utilizado en la elección de la matriz Q y R, logra un muy buen rendimiento y reflejan la eficiencia y la accesibilidad a los requisitos de diseño del enfoque propuesto.

El objetivo de este trabajo es diseñar un controlador LQG y un controlador STR para controlar el voltaje producido por una planta prototipo constituida por un sistema de un motor y un generador DC y realizar un análisis comparativo entre el desempeño de los dos controladores, utilizando diferentes métricas de comparación.

Metodológicamente el artículo inicia con la descripción del sistema a controlar y la obtención de su modelo matemático utilizando identificación no paramétrica, a continuación, se plantean los conceptos básicos de los controladores STR y LQG, se estiman sus parámetros y, finalmente se implementan en la planta prototipo, para comparar su desempeño y analizar los resultados obtenidos.

II.   MATERIALES Y MÉTODOS

2.1   Descripción del sistema

La la figura 1 muestra la planta prototipo objeto de estudio, la cual está compuesta por: la unidad electromecánica, encargada de la generación del voltaje que se desea controlar en el sistema, está constituida por dos motores DC de 24 V, uno de ellos trabajando como motor propiamente dicho y el otro, unido al primer motor mediante un acople flexible, cumple la función de generador, la unidad de medición, encargada de adecuar el valor de la variable a controlar, formada por un divisor de voltaje que convierte la salida proveniente del generador DC en una señal de 0 a 4.5 voltios, el sistema de adquisición de datos y control basado en una tarjeta Arduino UNO®, un Módulo Driver Dual L298N para motores DC el cual maneja la potencia mediante la modulación del ancho de pulso (PWM), cuya amplitud depende de la señal de control proveniente del computador en el cual se implementan los

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

algoritmos de control en la plataforma Labview® y finalmente, una fuente de alimentación con salida de 24VDC a 10 A,
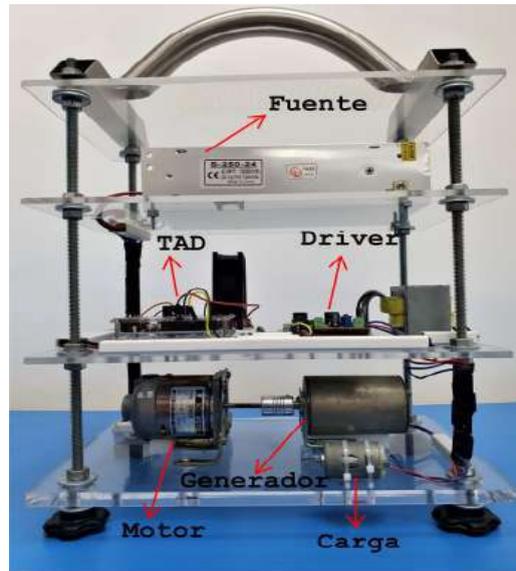
Fig. 1: Prototipo Del Generador DC

La figura 2 muestra el diagrama eléctrico que interconecta las unidades de funcionamiento que conforman el sistema.
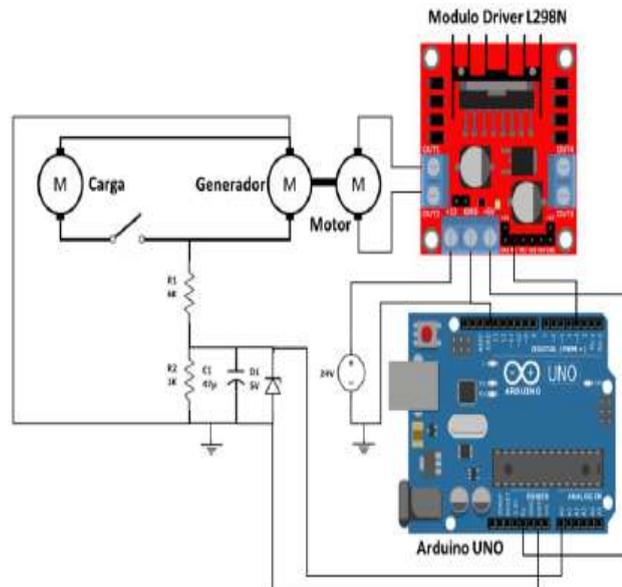
Fig. 2: Diagrama Eléctrico del Prototipo

2.2 Obtención del Modelo Matemático del Prototipo del Generador DC

La identificación de sistemas, permite obtener el modelo matemático que describe la dinámica del sistema a partir de datos experimentales [8]. El prototipo del generador DC se sometió a diferentes señales de entrada de tipo escalón, en un rango del 0 al 100%, con incrementos de 10% como se muestra en la Fig.3.

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0   51

Por las características de la respuesta y las del prototipo del generador DC, su dinámica se aproximó a un modelo de segundo orden con retardo (SOR) [9], [10] para ello se promediaron los diferentes modelos estimados a partir de la base de datos obtenida, la función de transferencia resultante está dada por:

$$G(s) = \frac{361.71e^{-0.0532}}{S^2 + 38.64S + 315.46} \tag{1}$$

La figura 4 muestra la comparación de la respuesta del modelo obtenido con respecto a la respuesta real tomada en el escalón del 50%
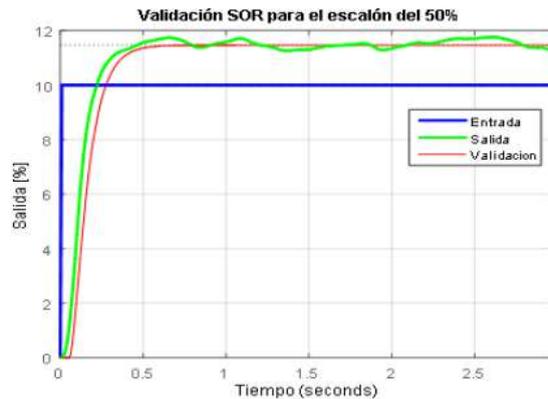
Fig. 4: Validación del Sistema Para el Escalón del 50%

2.3 Controlador STR Por Asignación De Polos

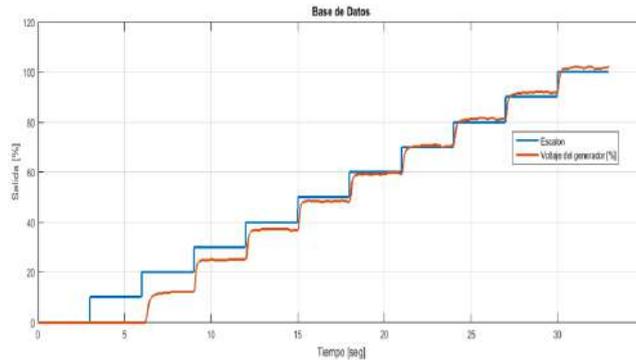Sea la planta definida por la función de transferencia de pulso:

$$G_p(z) = \frac{B(z^{-1})}{A(z^{-1})} = \frac{b_1 z^{-1} + b_2 z^{-2} + \dots b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_m z^{-m}} * z^{-d} \tag{2}$$

Para la planta dada en la ecuación 2, es posible obtener un controlador representado por:

$$D(z) = \frac{Q(z^{-1})}{P(z^{-1})} = \frac{q_o + q_1 z^{-1} + \dots q_v z^{-v}}{1 + p_1 z^{-1} + \dots + p_u z^{-u}} \tag{3}$$

De modo que sus polos de lazo cerrado queden ubicados en el lugar deseado [11]. El orden del numerador y del denominador de la función de transferencia del controlador deben ser: $v = m$ y $u = m + d$ respectivamente.

La ecuación característica deseada para el sistema en lazo cerrado está dada por:

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

52 Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

$$P(z^{-1})A(z^{-1}) + Q(z^{-1})B(z^{-1})z^{-d} = \Delta(z^{-1}) \tag{4}$$

En donde:

$$\Delta(z^{-1}) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \cdots \alpha_l z^{-l} \tag{5}$$

La ecuación característica deseada se determina a partir de las especificaciones de diseño establecidas, el orden l de la misma está dado por: $l = 2m + d$. Los parámetros del controlador se obtienen al resolver 4 mediante la ecuación:

$$
\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{m+d} \\ q_1 \\ q_2 \\ \vdots \\ q_m \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & \cdots & 0 & 0\{d & 0 & \cdots & 0 \\
a_1 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
a_1 & a_1 & \vdots & 0 & b_1 & 0 & \vdots & 0 \\
\vdots & a_2 & & 0 & b_2 & b_1 & & 0 \\
a_m & \vdots & & 0 & \vdots & \vdots & & b_1 \\
0 & 0 & & 0 & b_m & & \cdots & \\
\vdots & \vdots & \cdots & a_m & 0 & b_m & \cdots & b_m \\
1 & 1 & \cdots & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} \alpha_1 - a_1 \\ \alpha_2 - a_2 \\ \vdots \\ \alpha_m - a_m \\ \alpha_{m+1} \\ \vdots \\ \alpha_{2m+1} \\ -1 \end{bmatrix}
\tag{6}
$$

Por último, se reemplazan los parámetros hallados en 6 en la ecuación 3 que corresponde a la función de transferencia del controlador STR.

2.4 Regulador Lineal Cuadrático LQG

Para diseñar el controlador LQG se aplica el principio de separación, que permite calcular en forma independiente la matriz de realimentación de ganancia K del regulador lineal cuadrático (LQR) y un estimador de estados tipo Kalman que posteriormente se combinan para formar el controlador [12], [13]. Para calcular la matriz de ganancia K se utiliza una función de costo cuadrática que minimiza el índice de desempeño, [14].

$$J(i, N) = \frac{1}{2}x^T(N). S. x(N) + \frac{1}{2}\sum_{K=1}^{N-1} x^T(k). Q. x(j) + u^T(k). R. x(k) \tag{7}$$

Las matrices Q y R son matrices de peso positivas definidas y se seleccionan para penalizar a algunos estados o a algunas entradas más que otras. S es una matriz semidefinida positiva. Para calcular el control LQR, se parte del sistema definido por:

$$x(k + 1) = Ax(k) + Bu(k) \tag{8}$$

$$y(k) = Cx(k) \tag{9}$$

El objetivo es encontrar la señal del control $u(k)$ que lleve al sistema dado por la ecuación 8, del estado inicial $x(k) = x(0)$ al estado final $x(N) = x(f)$, minimizando la función de coste cuadrática definida por la ecuación 7. Utilizando un procedimiento iterativo [15] se puede obtener el paso del estado k al estado N lo que da como resultado:

$$u(k) = -K(k)x(k) \tag{10}$$

$$K(k) = \left[R + B^T S(k + 1)B \right]^{-1} B^T S(k + 1)A \tag{11}$$

$$S(k + 1) = [A - BK(k)]^T S(k + 1)[A - BK(k)] + Q + K^T(k)RK(k) \tag{12}$$

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0   53

Las ecuaciones 10, 11 y 12 generan el controlador LQR para el sistema discreto dado por la ecuación 8. La ecuación 12 se conoce como la ecuación de Riccati. En la Fig.5 se presenta un diagrama de bloques del sistema de control LQR.
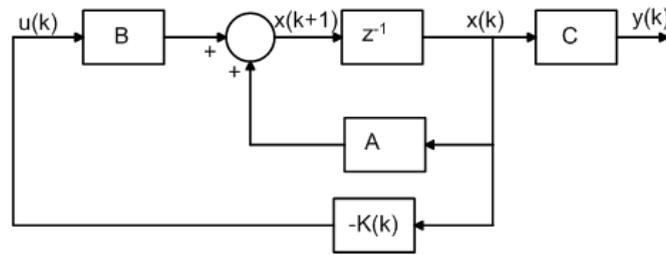
2.4 Filtro de Kalman

El filtro de Kalman es un algoritmo óptimo recursivo basado en el modelo de espacio de estados y su objetivo es estimar los estados de una manera óptima y, dada su función de filtro, eliminar el ruido, de manera que se minimice el índice del error cuadrático medio [12],
Si se considera el sistema discreto:

$$x(k + 1) = Ax(k) + Bu(k) + Gv(k) \tag{13}$$

$$y(k) = Cx(k) + w(k) \tag{14}$$

En donde: $v(k)$ y $w(k)$ son el ruido del proceso y de la medición respectivamente.

Para obtener el filtro de Kalman se parte de la ecuación del observador de estados normal, pero teniendo en cuenta los efectos del ruido para el cálculo de la matriz de ganancia L. Esta matriz debe elegirse de forma que dé la mejor estimación del estado del sistema y que rechace los efectos de los ruidos. Esta elección define el observador óptimo conocido como filtro de Kalman dado por la ecuación: [13]

$$\hat{x}(k + 1) = A\hat{x}(k) + Bu(k) + L[y(k) - C\hat{x}(k)] \tag{15}$$

$\hat{x}(k)$ es la variable de estado estimada y L es la matriz de ganancia de Kalman.

El filtro de Kalman estima el estado del proceso en un instante de tiempo y posteriormente obtiene una realimentación de las medidas. Por eso las ecuaciones del filtro de Kalman se pueden dividir en dos partes: actualización en el tiempo (conocida como predicción) y actualización en las medidas (conocida como corrección) [14]. En la fase de predicción, se realiza el cálculo del estado y de la covarianza del error actual $P(k)$ a partir de la covarianza del error en el instante anterior, es decir:

$$\hat{x}(k + 1) = A\hat{x}(k) + Bu(k) \tag{16}$$

$$P(k + 1) = AP(k)A^T \tag{17}$$

En la fase de corrección, se actualiza la ganancia de Kalman, se corrige la estimación del estado calculado en la etapa de predicción y, finalmente se actualiza la covarianza del error o sea:

$$L(k + 1) = P(k)C^T\left[CP^T + R\right]^{-1} \tag{18}$$

$$\hat{x}(k + 1) = \hat{x}(k) + L(k)\left[y(k) - C\hat{x}(k)\right] \tag{19}$$

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

54 Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

$$P(k + 1) = [I - L(k)C](P(k) \tag{20}$$

La implementación del controlador utilizando el filtro de Kalman es similar a la utilizada con el observador tipo predictor diseñado utilizando la técnica de asignación de polos [15], es decir:

$$D(z) = -\frac{U(z)}{Y(z)} = K[zI - A + BK + LC]^{-1}B \tag{21}$$

La Fig.6 representa el diagrama en bloques del sistema de control LQG incluyendo la matriz de ganancia de realimentación K del control LQR y el estimador de estados constituido por el filtro de Kalman.
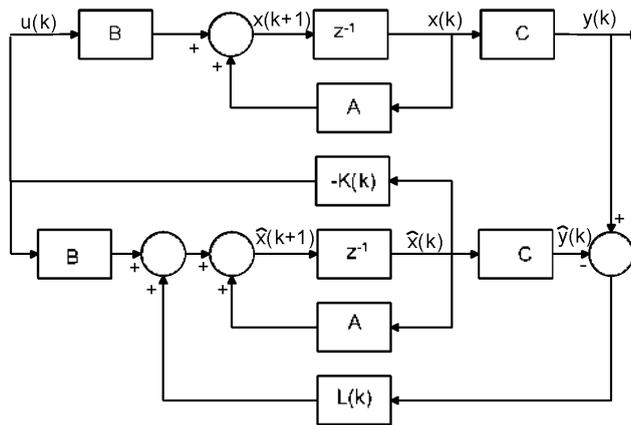

Fig. 6: Sistema de Control LQR Con Filtro de Kalman

Como el sistema de la Fig.7 no considera una entrada de referencia, se puede presentar error en estado estable, para evitar este problema, se utiliza un sistema de seguimiento con un integrador en el camino directo entre el comparador de error y la planta [14],[15] tal como se observa en la Fig.7.
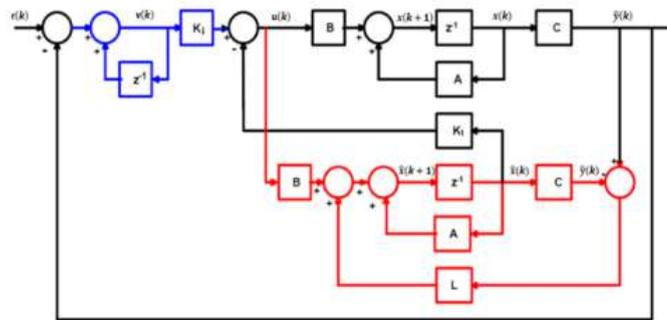
Fig. 7: Sistema LQR Con Filtro de Kalman e Integrador

La ley de control del sistema de la Fig.8 está dada por la ecuación 21, la cual combina el integrador, la matriz de ganancia de realimentación óptima y el estimador de estados del sistema (filtro de Kalman) [15]

$$D(z) = \frac{[1 + K_1[zI - A + LC]^{-1}B]^{-1}}{z - 1}$$
$$* [K_i z[R(z) - Y(z)] - (z - 1)K_1[zI - A + LC]^{-1}LY(z)] \tag{21}$$

Para el cálculo de K_1 y K_i se utiliza la ecuación:

$$[K_1 \quad K_i] = \left[\hat{K} + [0 \quad I]\right]\begin{bmatrix} A - I_m & B \\ CA & CB \end{bmatrix}^{-1}$$

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0

55

En donde $\hat{K}$ es la matriz de ganancia del estimador óptimo LQR obtenida con las matrices ampliadas:

$$A_1 = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} \qquad B_1 = \begin{bmatrix} 0 \\ I_m \end{bmatrix}$$

III.   RESULTADOS

A continuación, se presenta el diseño de las dos estrategias de control a implementar en el prototipo de generador DC.

3.1 Diseño del Controlador STR Por Asignación de Polos Para el Prototipo del Generador DC

Utilizando como periodo de muestreo $T = 0.0335$ s la función de transferencia discreta de la planta es:

$$G(z) = \frac{0.0289z^{-1} + 0.1673z^{-2} + 0.0249z^{-3}}{1 - 1.081z^{-1} + 0.274z^{-2}} * z^{-1} \tag{22}$$

De la ecuación 22, se obtiene: $m = 3$, $d = 1$, por tanto para el controlador resulta: $v = 3$ $u = 4$ y el orden de la ecuación característica deseada es $l = 7$.

Se definen como parámetros de diseño para el sistema en lazo cerrado: factor de amortiguamiento de 0.8 y tiempo de establecimiento 20% menor al correspondiente en lazo abierto. Así, los polos de lazo cerrado dominantes están ubicados en $z = 0.687 \pm j0.18$. Se deben agregar cinco polos no dominantes para completar el orden de la ecuación característica. Si se ubican en $z = 0$, se obtiene:

$$\Delta(z) = z^7 - 1.374z^6 + 0.504z^5 = 0$$

Utilizando la ecuación 6 se obtienen los parámetros del controlador:

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} -0.293 \\ -0.1657 \\ -0.4722 \\ -0.0691 \\ 2.7335 \\ -2.9053 \\ 0.7597 \\ 0 \end{bmatrix}$$

La función de transferencia del controlador es según la ecuación 3:

$$D(z) = \frac{2.7335 - 2.9053z^{-1} + 0.7597z^{-2}}{1 - 0.293z^{-1} - 0.1657z^{-2} - 0.422z^{-3} - 0.0691z^{-4}}$$

3.2 Diseño del Controlador LQG Para el Prototipo del Generador DC

La representación del sistema en el espacio de estados en su forma canónica observable está dada por:

$$x(k+1) = \begin{bmatrix} 1.081 & 1 & 0 & 0 \\ -0.297 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 0.0289 \\ 0.1673 \\ 0.0249 \end{bmatrix}$$

$$y(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

Haciendo uso del software Matlab® se realiza el cálculo de la matriz de ganancia del observador L, ingresando como parámetros de entrada las matrices A y B del sistema y como ruidos para el

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

56   Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

proceso y para la medición $v = 0.02$ y $w = 0.008$ respectivamente. La matriz de ganancia del observador está dada por:

$$L = \begin{bmatrix} 0.2624 \\ -0.092 \\ 0.0258 \\ 0.0092 \end{bmatrix}$$

La fig. 8 muestra una comparación entre los valores de las variables de estado del sistema y los valores estimados por el filtro de Kalman al aplicar un escalón unitario al sistem.
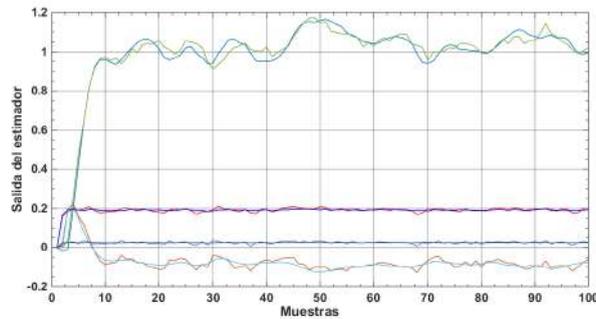

Figura 8: Respuesta del Filtro de Kalman Ante Entrada Escalón

Para el cálculo de la ganancia K del controlador LQR, se utilizaron las matrices ampliadas A_1 y B_1 debido a la inclusión del integrador. La matriz S inicial se tomó como una matriz identidad y las matrices Q y R se estimaron con el criterio de Bryson [16]:

$$Q = \begin{bmatrix} 1000 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5000 \end{bmatrix}$$

$$R = 1000$$

(4)

El resultado obtenido, utilizando recursivamente las ecuaciones 10, 11 y 12, fueron:

$$K_1 = [3.7101\ 4.7032\ 4.6264\ ] \quad K_i = [1.0453]$$

IV.   ANÁLISIS DE RESULTADOS

A continuación, se presenta el análisis de resultados de las dos estrategias de control diseñadas para el prototipo del generador DC.

4.1   Resultados Obtenidos Con el Controlador STR.

En la Fig.9 se muestra la respuesta del sistema de control STR simulada con el software Matlab®, ante escalones del 20%, 30% y 40% respectivamente. En la Fig.10 se presenta la respuesta del sistema real ante las mismas señales de entrada. Se observa coincidencia en las dos respuestas: máximo sobreimpulso cercano al 10%, buena velocidad de respuesta, error de estado estable prácticamente cero y la ley de control produce una salida bastante fuerte al inicio de la respuesta del sistema de control, llevando el sistema a estabilidad en el punto de referencia.

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

© 2022 London Journals Press

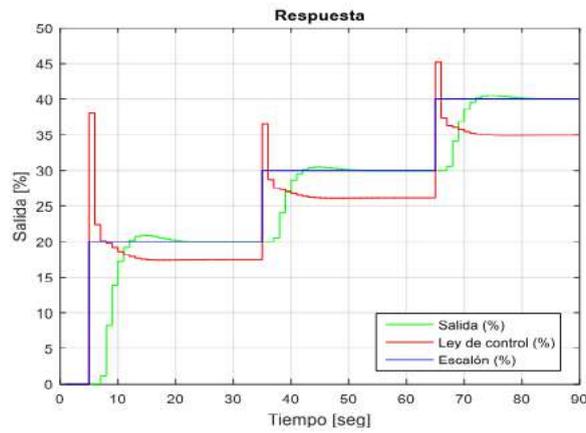Volume 22 | Issue 6 | Compilation 1.0   57

Fig. 9: Simulación del Controlador STR

Fig. 10: Control STR con el Prototipo del Generador DC

La fig. 11 muestra la respuesta del sistema real ante una perturbación: se añade una carga al generador en t=14s y se suspende en t=20s, se observa que el controlador STR responde de manera eficiente ante el cambio de carga, presentando una respuesta rápida en presencia de la misma y retornando el sistema al punto de referencia tanto al conectar la carga como al desconectarla.

Fig. 11: Respuesta del Control STR Ante una Carga

4.2 Resultados Obtenidos del Controlador LQG

La Fig.12 muestra la respuesta del sistema con el controlador LQG, simulada con el software Matlab®, para entradas en forma de escalón del 20%, 30% y 40%, y en la Fig 13 se aprecia la respuesta del sistema real ante las mismas señales de entrada. En los dos casos el sistema presenta buena velocidad de respuesta, error de estado estable igual a cero, sobreimpulso no significativo, y la ley de control no es tan agresiva como la generada por el STR en las mismas condiciones.
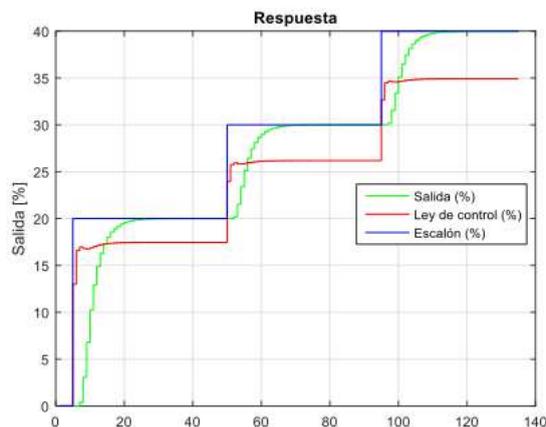
Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

58   Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press
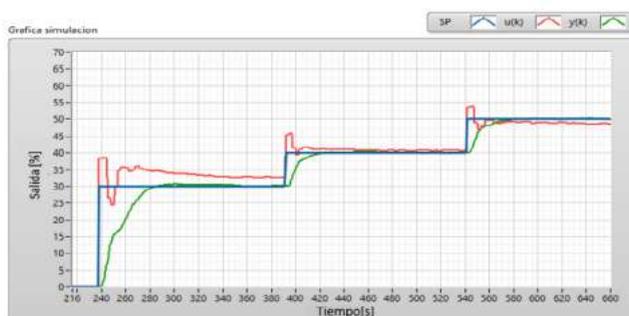
Fig. 12: Simulación del Controlador LQG

Fig. 13: Control LQG del Prototipo del Generador DC

Finalmente, en la fig. 14 se ve la respuesta del sistema ante una carga aplicada al generador, igual que la utilizada para el STR, se observa que el controlador LQG responde rápidamente ante la presencia de la carga y retorna al sistema al punto de consigna tanto al aplicar la carga como al desconectarla.
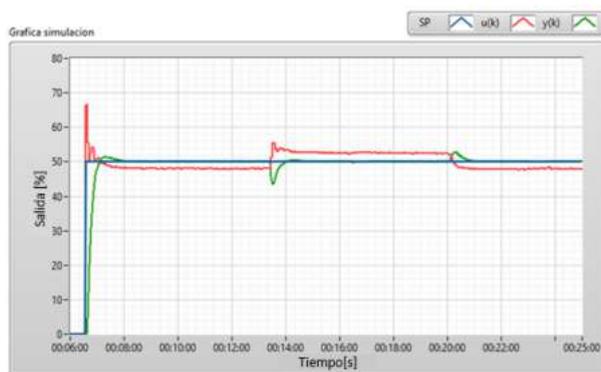
Fig. 14: Respuesta del Control LQG Ante una Carga

4.3 Comparación del Desempeño de Las Dos Estrategias de Control Para el Prototipo del Generador DC.

Para comparar las dos estrategias de control, se utilizaron métricas de respuesta temporal y de la integral evaluadas bajo las mismas condiciones de señal de entrada y de carga. Para la respuesta temporal se utilizaron: el máximo sobreimpulso (Mp), el tiempo de establecimiento (t_s) y el error de estado estable (e_{ee}). En las métricas de la integral se utilizaron entre otras, la integral del valor absoluto del error (IAE), la integral del valor del cuadrado del error (ICE) y el trabajo sobre la variable

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0   59

London Journal of Engineering Research

manipulada (TVM). Se considera una zona de operación entre el 40% y 70% con incrementos del 10% en la referencia, en la Fig.15 se observa la respuesta del sistema de control STR y en la Fig.16, se muestra la del sistema de control LQG ante el mismo estímulo.
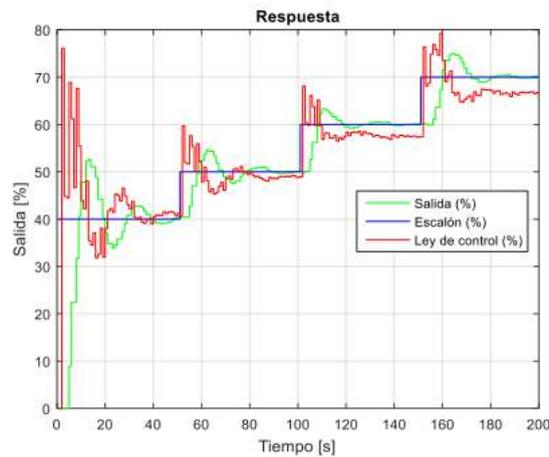

Fig. 15: Respuesta Control STR
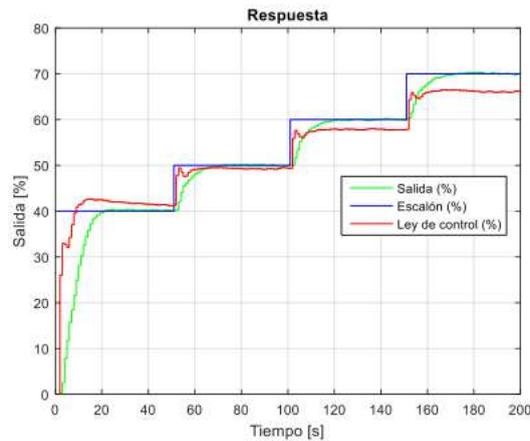
Fig. 16: Respuesta Control LQG

En las tablas 1 y 2, se presentan los resultados obtenidos con cada una de las métricas para el controlador LQG y para el STR respectivamente.

Tabla 1: Desempeño del Controlador LQG

Métrica	CONTROLADOR LQG			
	40%	50%	60%	70%
$Mp[\%]$	1.00	2.20	3.30	4.40
$e_{ee}[\%]$	0.00	0.00	0.00	0.00
t_s [s]	0.14	0.14	0.14	0.15
TVM	10609	10270	10690	10601
ECM	162.81	7.39	7.64	7.72
IAE	310.39	63.31	63.37	65.87
$IAET$	15.85	3.74	3.44	3.93
ICE	8390	369.72	382.19	386.09
$ICET$	243.09	8.11	8.21	8.74

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

60 Volume 22 | Issue 6 | Compilation 1.0   © 2022 London Journals Press

Tabla 2: Desempeño del Controlador STR

Métrica	CONTROLADOR LQG			
	40%	50%	60%	70%
$Mp[\%]$	30.18	46.07	31.18	47.31
$e_{ee}[\%]$	0.00	0.00	0.00	0.00
t_s [s]	0.09	0.09	0.09	0.10
TVM	10910	10911	10987	10945
ECM	181.10	12.06	10.689	16.578
IAE	376.86	106.90	84.06	122.38
$IAET$	33.31	11.98	6.18	10.73
ICE	9058.9	602.96	534.46	828.93
$ICET$	324.7	29.31	16.02	41.57

Los resultados de las pruebas indican que, para incrementos en el set-point del 10%, el controlador LQG presentó, en todos los casos, menor sobreimpulso que el STR. Sin embargo, el tiempo de establecimiento con el controlador LQG fue mayor al presentado por el controlador STR. En los experimentos realizados, los dos controladores presentan error de estado estable igual a cero. Analizando los resultados de las métricas de la integral del error y el TVM, se deduce que el desempeño del controlador LQG, en términos generales, fue mejor que el presentado por el controlador STR.

V. CONCLUSIONES

Se realizó el diseño de dos estrategias de control para controlar el voltaje generado por una planta prototipo formada por la combinación de dos motores de 24 V, uno trabajando como motor propiamente dicho y el otro como generador. Para el diseño de los controladores la planta se aproximó a un sistema de segundo orden con retardo. Al realizar las pruebas sobre el sistema real se observa que las dos estrategias de control muestran buen desempeño ante cambios en la señal de referencia y en presencia de perturbaciones. El tiempo de establecimiento del sistema con el controlador STR es menor que el del sistema con el LQG pero, como resultado, el controlador STR presenta mayor sobreimpulso y una ley de control más agresiva que la del sistema con el LQG haciéndose evidente la relación entre el esfuerzo de control y los parámetros de desempeño de sobreimpulso y el tiempo de establecimiento. Finalmente se observa que las variaciones en la carga no afectan la estabilidad del sistema y los dos controladores compensan los cambios en la misma, lo cual demuestra que las dos estrategias son robustas ante la presencia de perturbaciones, notándose en general, un mejor desempeño del controlador LQG.

Es importante notar que el diseño de los controladores se realizó tomando como base un modelo lineal de la planta, mientras que su implementación y análisis se hizo utilizando el modelo real, no lineal de la misma y teniendo en cuenta condiciones de carga variable.

REFERENCIAS

1. Szabolcsi, R. (2018). Robust LQG controller design for the small unmanned aerial vehicle. Review of the Air Force Academy No.1 (36). DOI: 10.19062/1842-9238. 2018.16.1.5
2. González, H y González, H. (2018). Diseño de un sistema de control avanzado para regular la velocidad de una turbina de vapor acoplada a un generador DC. ITECNE Vol. 15 Nº1. ISSN: 1692-1798 DOI: 10.15332/iteckne.v15i1.1964.
3. Da silva, R. Murilo, A. y Teixeira, E (2017). Controlador LQG aplicado a un sistema de dirección eléctrico. Revista Interdisciplinaria de Investigación de Ingeniería, 2 (31), 164-181. DOI:10.26512/ripe. v2i31.21332.

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

© 2022 London Journals Press

Volume 22 | Issue 6 | Compilation 1.0

61

London Journal of Engineering Research

4. GONZALEZ, C. (2016) Comparación de un controlador LQR vs un controlador PID implementados en un helicóptero de dos grados de libertad pivotado. Universidad Distrital Francisco José de Caldas Ingeniería en Control Bogotá D.C

5. Patel, P. (2015) Modeling and optimal control algorithm design for HVAC systems in energy efficient buildings. International Research Journal of Engineering Science, Technology and Innovation (IRJESTI). ISSN-2315-5663 Vol. 2(2) pp. 29-39.

6. Jun-Keun Ji, Dong-Choon Lee & Seung-Ki (2002). LQG based speed controller for torsional vibration suppression in 2-mass motor drive system. Conference: Industrial Electronics, Control, and Instrumentation, Proceedings of the IECON, International Conference. DOI: 10.1109/IECON.1993.339166.

7. Jamal M. Ahmed (2017). Optimal LQG Controller Based on Genetic Algorithm Technique For Speed Control of DC Motor. International Journal of Engineering and Innovative Technology (IJEIT) Volume 7, Issue 1, July 2017. ISSN: 2277-3754

8. Ljung, L. (1999). System Identification: Theory for the User. Prentice Hall. Segunda edición. ISSN 0-13-656695-2.

9. Gómez, J.C. (2009) Toolbox didáctico para el diseño y análisis de sistemas de control lineal. Revista educación en Ingeniería. Vol. 09 N°. 8. Pp 155-169• Publicada en línea por la Asociación Colombiana de Facultades de Ingeniería. ISSN 1900-8260.

10. Alfaro, V (2001). Identificación de procesos sobreamortiguados utilizando técnicas de lazo abierto. Revista Ingeniería Universidad de Costa Rica. Vol. 11 N°.

11. García, L y Arroyave, M. (2015) Controlador difuso y controlador por asignación de polos para un levitador neumático. Revista Politécnica, vol. 11, no. 21, pp. 107-116. ISSN 1900-2351 (Impreso), ISSN 2256-5353 (En línea).

12. Montoro, G (1996). Contribución al estudio y desarrollo de técnicas de control aplicadas a la linealización de sistemas. Tesis doctoral. Universidad Politécnica de Catalunya Departamento de Teoría del Señales y Comunicaciones.

13. Outanoute, M. Lachhab, A. Guerbaoui, M. Selmani, A. (2016) Synthesis of an Optimal Dynamic Regulator Based on Linear Quadratic Gaussian (LQG) for the Control of the Relative Humidity Under Experimental Greenhouse. International Journal of Electrical and Computer Engineering (IJECE) Vol. 6, No. 5, October 2016, pp. 2262~2273 ISSN: 2088-8708.

14. Ogata, K. (2006). Sistemas de control en tiempo discreto. Prentice Hall Hispanoamericana S.A. ISBN:9789688805398.

15. García J. (2013) Control Digital, Teoría y Práctica. Tercera edición. Politécnico Colombiano JIC. ISBN 978-958-9090-25-1.

16. Ariza, H y García J (2017). Diseño e implementación de un péndulo invertido doble lineal sobre base móvil como material de apoyo didáctico para el área de control 3 en la licenciatura en electrónica de la Universidad Pedagógica Nacional. Tesis de maestría Universidad Pedagógica Nacional, Bogotá.

London Journal of Engineering Research

Design of a Linear Quadratic Gaussian Controller and a STR Controller for Pole Placement to Control the Voltage of a DC Generator

62 Volume 22 | Issue 6 | Compilation 1.0 © 2022 London Journals Press

London Journal Press Membership

London Journals Press membership is an elite community of scholars, researchers, scientists, professionals and institutions associated with all the major disciplines. London Journals Press memberships are for individuals, research institutions, and universities. Authors, subscribers, Editorial Board members, Advisory Board members, and organizations are all part of member network.

Read more and apply for membership here:
https://journalspress.com/journals/membership

For Authors

For Institutions

For Subscribers

Author Membership provide access to scientific innovation, next generation tools, access to conferences/seminars /symposiums/webinars, networking opportunities, and privileged benefits.
Authors may submit research manuscript or paper without being an existing member of LJP. Once a non-member author submits a research paper he/she becomes a part of "Provisional Author Membership".

Society flourish when two institutions come together." Organizations, research institutes, and universities can join LJP Subscription membership or privileged "Fellow Membership" membership facilitating researchers to publish their work with us, become peer reviewers and join us on Advisory Board.

Subscribe to distinguished STM (scientific, technical, and medical) publisher. Subscription membership is available for individuals universities and institutions (print & online). Subscribers can access journals from our libraries, published in different formats like Printed Hardcopy, Interactive PDFs, EPUBs, eBooks, indexable documents and the author managed dynamic live web page articles, LaTeX, PDFs etc.

JOURNAL AVAILABLE IN

PRINTED VERSION, INTERACTIVE PDFS, EPUBS, EBOOKS, INDEXABLE
DOCUMENTS AND THE AUTHOR MANAGED DYNAMIC LIVE WEB PAGE
ARTICLES, LATEX, PDFS, RESTRUCTURED TEXT, TEXTILE, HTML, DOCBOOK,
MEDIAWIKI MARKUP, TWIKI MARKUP, OPML, EMACS ORG-MODE & OTHER

GO GREEN AND HELP
SAVE THE ENVIRONMENT

SCAN TO KNOW MORE

support@journalspress.com
www.journalspress.com

*THIS JOURNAL SUPPORT AUGMENTED REALITY APPS AND SOFTWARES