

Teknik Personalisasi Halaman Web dalam lingkup Metodologi Page Ranking menggunakan Pendekatan Kecerdasan Buatan

Efendi
Teknik Elektro
Politeknik Negeri Padang
Padang, Indonesia
efendi.muchtar25@gmail.com

Abstrak—Sejak diperkenalkannya Internet berkecepatan tinggi dan meningkatnya jumlah pengguna internet ponsel cerdas, jumlah data di Internet telah meroket. Tidak ada pemantauan terpusat atas data yang disimpan dan di indeks di Web menyulitkan mesin pencari untuk mengambil informasi dari Web pada tingkat minat yang tepat dan berarti yang dicari pengguna. Karena jumlah data yang tersedia di Internet semakin meningkat, mesin pencari harus mampu menemukan informasi yang diminta oleh pengguna internet. Meskipun berbagai metode untuk menentukan kepentingan relatif pengguna dalam hasil mesin pencari telah muncul seiring dengan perkembangan Internet dan pertumbuhan eksponensial informasi online, metode ini telah kehilangan keefektifannya atau menjadi ketinggalan zaman untuk memenuhi tuntutan yang terus meningkat. pengguna internet saat ini dan harapan mereka yang terus meningkat. Menggunakan metodologi peringkat halaman, artikel ini membahas berbagai skema personalisasi web untuk melihat cara kerja kecerdasan buatan. Sesuai preferensi dan minat pengguna, ini membandingkan berbagai pendekatan yang dipertimbangkan dan menyimpulkan kemajuan dan keunikannya.

Kata Kunci—*Personalized Page Ranking (PPR), User Interest Score (UIS), Term Frequency and Inverse Document Frequency (TF-IDF), User Interest Hierarchy (UIH)*

I. PENGANTAR

Saat pengguna mengetik kueri ke dalam bidang pencarian, mesin pencari berusaha untuk mengembalikan hasil web. Tetapi untuk memberikan hasil yang paling relevan dan berharga, mesin pencari harus memahami tujuan permintaan pengguna. Meskipun berbagai teknologi pencarian web, banyak pengguna mesin pencari tidak puas dengan hasil mereka. Orang yang sama mungkin memiliki berbagai prioritas dan minat dalam pencarian hasil pencarian web dari tempat yang berbeda. Misalnya, pengguna yang berbeda mungkin memiliki preferensi dan minat lain dalam hasil penelusuran web.

Konteks di mana kueri diminta dapat mempengaruhi maknanya secara signifikan, dan pengguna hanya dapat menentukan konteks ini. Membayangkan pertanyaan spesifik sebagai "perak" membantu. Ornamen perak, warna biru tertentu, atau spesies ikan tertentu dapat berada di bagian atas daftar pengguna. Kumpulan hasil yang sebanding dikembalikan oleh mesin telusur tradisional, yang tidak memahami konteks kueri. Akibatnya, hasil pencarian

yang paling relevan muncul di bagian atas hasil pencarian web yang dipersonalisasi. Efisiensi personalisasi pencarian web bervariasi tergantung pada jumlah kueri, pengguna, dan lingkungan pencarian. Cara cerdas dan efisien untuk mengambil jumlah data mentah berbasis internet yang terus meningkat menjadi semakin penting karena ukuran data ini berkembang secara eksponensial. Data digital dapat diperoleh, disesuaikan, dan digunakan dengan berbagai cara. Teknik yang jelas diperlukan untuk mengatasi penyesuaian dan modifikasi data untuk menarik informasi yang berarti, dan Data Mining siap membantu.

Banyak teknik dan tujuan yang ada untuk data mining. Algoritma evolusioner (EA) [1] termasuk di antaranya. Algoritma yang diilhami secara biologis seperti Algoritma Genetika (GA) [2] termasuk dalam kategori ini. Teknik berbasis gerombolan seperti koloni semut dan Particle Swarm Optimizations (PSO) [5] adalah contoh dari algoritma DE. Dengan algoritma Artificial Bee Colony (ABC), DM juga telah meningkatkan akurasi klasifikasi [6]. PART, DPSO, SOM, Naive Bayes, Classification Tree, dan Nearest Neighbor adalah pengklasifikasi yang paling umum digunakan untuk data mining (KNN). Pada tahun 2005, Karabog di Universitas Erciyes di Turki merancang algoritma ABC, algoritma cerdas gerombolan baru.

Metode ini hanya memiliki beberapa kualitas dan parameter kontrol kecil, sehingga mudah dipahami dan diterapkan. Karena aplikasinya yang luas, seperti filter IIR digital [7][8], jaringan saraf tiruan [9][10][11], dan banyak lagi, teknik ini sering digunakan dalam optimasi. Sistem web menggunakan Umpan Balik Relevansi Pengguna [15] untuk memberikan informasi yang disesuaikan dengan kebutuhan dan preferensi spesifik pengguna. Misalnya, jika dua suku dapat dibandingkan, keduanya dikatakan terkait dalam model ruang vektor. Pengguna diminta untuk mengkategorikan dokumen ke dalam kelompok yang relevan dan tidak relevan untuk menerima Umpan Balik Relevansi. Untuk memperluas kueri, algoritma Rocchio digunakan. Sayangnya, umpan balik yang relevan tidak memenuhi kebutuhan pengguna karena pengguna enggan mengirimkan informasi apakah mereka tertarik dengan dokumen tertentu.

II. PENELITIAN TERKAIT

Terlepas dari berbagai macam metode pencarian, masih ada banyak area dan pengaturan di mana orang tidak puas

dengan hasil mereka. Alih-alih mempertimbangkan pilihan dan keinginan pengguna, mesin pencari tradisional mengembalikan hasil hanya berdasarkan kecocokan kata kunci. Berdasarkan struktur direktori Web, Ramadhan [16] menyarankan metode algoritmik untuk menentukan signifikansi backlink yang berbeda dengan menetapkan faktor bobot yang berbeda. Tidak ada pertimbangan untuk kepentingan pengguna dalam perhitungan peringkat ini, yang hanya didasarkan pada struktur tautan. Ada cara baru untuk menentukan peringkat halaman berdasarkan riwayat pencarian pengguna dan preferensi lokal lainnya. Halaman web pertama-tama diatur dan diberi peringkat menurut peringkatnya, dan kemudian dikembalikan ke pengguna sebagai hasil pencarian. Metode Page-Ranking digunakan untuk menghitung peringkat untuk setiap halaman web berdasarkan grafik di web. Pencarian, penelusuran, dan estimasi lalu lintas semuanya menggunakan PageRank.

Ada beberapa cara lain untuk meningkatkan peringkat mesin pencari Anda. Untuk masalah optimasi, metode Artificial Bee Colony "ABC" memiliki banyak keunggulan, seperti memori yang besar dan kemampuan untuk melakukan pencarian lokal, dan fenomena improvisasi solusi. Namun, penelitian telah menemukan bahwa ABC terhalang dalam optimalitas lokal dalam beberapa keadaan, yang merusak kinerja konvergensi dan hasil dari metode ABC yang khas. Cara baru untuk memperbaiki posisi lebah buatan telah diusulkan oleh Aderhold et al. untuk memahami bagaimana pertumbuhan penduduk mempengaruhi ABC. Menurut Stanarevic dan rekan, gagasan "lebah pintar" ditawarkan dalam ABC yang dimodifikasi yang menggabungkan kemampuan lebah untuk mengingat lokasi dan kualitas sumber makanan [22][23]. Karena ABC asli ditemukan memiliki akurasi dan kemanjuran yang rendah ketika memecahkan masalah optimasi, Lei et al. muncul dengan ide untuk menambahkan bentuk berat tertentu yang dipengaruhi oleh optimasi gerombolan partikel [24].

Algoritma Meta-Heuristik, seperti algoritma genetika (GA)[25], [26], [27], optimasi koloni semut [28],[29], optimasi gerombolan partikel (PSO)[30],[31], dan koloni lebah buatan (ABC)[32], menjadi algoritma meta-heuristik yang paling sukses selain yang telah disebutkan. Pencarian kukuk[33, 34], optimasi kepik tujuh titik (SLO)[35], dan algoritma pencarian bakteri (BFA)[36] adalah contoh meta-heuristik yang baru-baru ini dibuat. Berbagai masalah di bidang manufaktur, penjadwalan, pengiriman layanan, transportasi, geologi, astronomi, dan banyak lagi dapat diselesaikan dengan menggunakan algoritma metaheuristik ini [37]. Namun, meskipun algoritma memiliki banyak keuntungan, mereka diganggu oleh hambatan. Leher dan semuanya, Algoritme ini berkinerja cukup baik pada berbagai masalah, dan beberapa berkinerja lebih baik di domain aplikasi tertentu, sementara yang lain mungkin tidak berfungsi dalam jenis masalah yang berbeda.

User Interest Hierarchy (UIH) [38] dapat digunakan untuk mengatur profil pengguna untuk penyesuaian yang lebih baik dari pengalaman web mereka. Algoritme DHC digunakan untuk mengkategorikan hasil setelah UIH secara tidak langsung melacak minat pengguna. Banyak kualitas istilah diturunkan, dan kata-kata tersebut kemudian dinilai sesuai. Berbeda dengan hierarki waktu saat ini, tidak ada pertimbangan untuk mengintegrasikan periode ini ke dalam waktu sekarang. UIH dapat ditingkatkan dengan

menambahkan dua fitur lagi, yaitu durasi dan spesifisitas node. Hasil teratas dapat dipesan ulang menggunakan alat ini. Namun, ketika ada pertanyaan baru dari pengguna, strategi yang sama tidak berhasil. Hasil pencarian web diberi peringkat menggunakan algoritme peringkat URL berbobot yang mempertimbangkan variabel seperti kata jangkar dan domain pengguna populer. Menurut insiden token, hasil mesin pencari ditimbang dan kemudian diurutkan ulang sesuai dengan lingkungan minat pengguna, dan bobot ini dipertahankan untuk mengurutkan ulang hasil. Beberapa algoritma sisi klien telah dibuat untuk personalisasi [40] untuk menjaga perasaan pengguna, kebiasaan penggunaan, dan kecenderungan pencarian dalam pikiran. Berbagai metode analisis tautan, termasuk Page Rank (PR), Weighted Page Rank (WPR), dan Hyperlink-Induced Topic Search (HITS), diperiksa dan dikontraskan di sini.

III. PERBANDINGAN: TEKNIK PERINGKAT HALAMAN

"Metodologi peringkat halaman" adalah metode paling efektif untuk pencarian web yang dipersonalisasi. Pencarian adalah salah satu kegunaan PageRank yang paling penting. Dua mesin pencari yang signifikan menggunakan PageRank. Yang pertama adalah mesin pencari sederhana berdasarkan judul istilah pencarian. Kedua, ada mesin pencari teks lengkap, seperti Google. Google memberi peringkat hasil pencarian berdasarkan berbagai faktor, termasuk kedekatan, teks jangkar (teks tautan yang menuju ke halaman web), dan PageRank. Adalah di luar tujuan kami untuk melakukan studi pengguna yang lengkap, tetapi kami telah melakukan beberapa uji coba komparatif dan menyajikan beberapa hasil sampel. Pencarian yang tidak ditentukan mendapatkan hasil maksimal dari keuntungan PageRank. Misalnya, pencarian untuk "Universitas Stanford" menggunakan mesin pencari konvensional dapat mengembalikan halaman web apa pun yang merujuk ke Stanford (seperti daftar publikasi), tetapi halaman beranda universitas mendapat peringkat pertama menggunakan PageRank.

Hasil pencarian web diberi peringkat menggunakan algoritme peringkat URL berbobot yang mempertimbangkan variabel seperti kata jangkar dan domain pengguna populer. Menurut insiden token, hasil mesin pencari ditimbang dan kemudian diurutkan ulang sesuai dengan bidang minat pengguna, dan bobot ini dipertahankan untuk mengurutkan ulang hasil. Beberapa algoritme sisi klien dikembangkan untuk tujuan personalisasi, dengan mempertimbangkan suasana hati pengguna, kebiasaan penggunaan, dan kecenderungan pencarian. Banyak metode analisis tautan [41], seperti Page Rank (PR), Weighted Page Rank (WPR), dan algoritma Hyperlink-Induce Topic Search (HITS), telah ditinjau dan dibandingkan di bagian sebelumnya. Kepentingan pengguna diperhitungkan saat menentukan relevansi hasil. TF-IDF dan skor minat pengguna dipertimbangkan saat memilih peringkat hasil (UIS). Sistem peringkat hibrida (TF-IDF + UIS) paling mencerminkan temuan kami, seperti yang ditunjukkan pada tabel-1 di bawah, dibandingkan dengan skema peringkat lainnya.

Tabel 1. Perhitungan perbandingan pendekatan Page Ranking

Preferred terms	TD-IDF	UIS	PPR	Hybrid
-----------------	--------	-----	-----	--------

				(US+TF-IDF)
Web	Excellent	-	Fair	Excellent
Web usage mining	Good	Adequate	Good	Fairly Good
Web structure mining	Good	Fair	-	Good
Web content mining	Good	Neutral	Fair	Good
Personalization	Excellent	Good	Fair	Excellent
Pattern analysis	Good	Good	Poor	Excellent
Usage History	-	Good	Fair	Good

Pilihan pengguna dipertimbangkan saat menghitung peringkat halaman yang dipersonalisasi. Bobot UIS dan TF-IDF terus berubah karena sifat kueri dan preferensi pengguna berubah. Teknik Hybrid menggunakan faktor yang berbeda untuk perhitungan peringkat untuk pengindeksan kata kunci, seperti yang ditunjukkan pada tabel-2 – berbeda dengan skema peringkat halaman Web biasa.

Tabel 2. Pengindeksan Kata Kunci Daftar Preferensi Istilah Kueri

	Traditional	Hybrid
1	Web	Personalization
2	Web Mining	Usage data
3	Web Structure Mining	Profile
4	Web content mining	User data
5	Internet	Access log
6	Data mining	Web usage mining
7	Web usage history	Pattern analysis

Page Re-ranking paling baik diterapkan dengan cara berikut saat menggunakan skema peringkat halaman hybrid. Pertama, mesin pencari mengambil dokumen yang cocok dengan kueri pengguna (dokumen K teratas) berdasarkan ukuran TF-IDF dan UIS untuk menemukan informasi yang relevan. Jaringan konsep yang lebih disukai pengguna untuk dibingkai kemudian dilacak untuk UIS, dan bobot fitur yang diusulkan dihitung.

Strategi PR ini dan perbandingan mendalam dari metodologi peringkat halaman terbaru juga telah dianalisis secara ekstensif (2013 hingga 2019). Menggunakan temuan kami, kami datang dengan tabel-3 berikut, yang menyoroti manfaat, keterbatasan, dan pendekatan setiap publikasi. Tentu saja, beberapa metode bekerja dengan baik dalam situasi ini, sementara yang lain tidak. Namun, mereka semua pada akhirnya memiliki kekuatan dan kebajikan unik mereka sendiri.

Tabel 3. Analisis komparatif antara pendekatan pemeringkatan halaman yang digunakan dalam penelitian dari tahun 2013 hingga 2017

The advantages and cons of each page ranking algorithm approach are laid forth in a table. Table 3 shows the merits, shortcomings, and methodology of each paper. Some methods work well in this situation, while others don't. Nevertheless, UIS outperforms the competition in tailored online search performance and yields more promising results. Thus, the Page Rank is calculated based on the user's interest score, based on the network's desired profile. Then, users' preferences are categorized and traced in an automated manner before everything else. Finally, the user is presented with the search result after mapping the matching results.

IV. KESIMPULAN

For all web pages, Page-Rank is a global ranking that considers their content. Web graph structure is all that matters here. Search results can be sorted using PageRank so that the most important and most urgent Web pages appear first in the list. A handful of regularly used documents can answer the vast majority of user inquiries using PageRank. If a remote database can't answer a search query, only a complete database on the Internet should be checked. For this reason, PageRank could be a great technique to get the most relevant and representative pages for a cluster center's search results. Using a Page-Rank method, the structure of the Web graph can help with several information retrieval tasks. There are several other page ranking algorithms out there. However, the hybrid page ranking algorithm (TF-IDF + UIS) is the most promising for consumers' pleasure. In addition, the profile convergence aspects might be thoroughly evaluated to provide even more precise and accurate, individualized web search results to improve search results ranking.

REFERENSI

- [1] Tan, Teoh, Yu, & Goh, "Evolutionary algorithms", 2009; Whitley, 2001.
- [2] Sumida, Houston, McNamara, & Hamilton, "Genetic Algorithms", 1990.
- [3] Storn & Price, "Differential Evolution Algorithms", 1997.
- [4] Colomi, Dorigo, & Maniezzo, 1991.
- [5] Kennedy J, "Particle Swarm Optimizations", 2006
- [6] D. Karaboga, "Artificial Bee Colony (ABC) algorithm", 2005.
- [7] N. Karaboga, "Digital IIR filters", 2009.
- [8] Karaboga, N. "A new design method based on artificial bee colony algorithm for digital IIR filters. Journal of the Franklin Institute, 2009, 346(4), 328-348.
- [9] D. Karaboga & Akay, "Artificial Neural Networks", 2005
- [10] Karaboga, D., Akay, B., & Ozturk, C. "Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks Modeling Decisions for Artificial Intelligence, 2007 (pp. 318-329).
- [11] Karaboga, D., & Akay, B. B. "An artificial bee colony (abc) algorithm on training artificial neural networks (Technical Report TR06): Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.
- [12] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. Journal of Global Optimization, 39(3), 459-471.
- [13] Karaboga, D., & Basturk, B. (2007). Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems Foundations of Fuzzy Logic and Soft Computing (pp. 789-798).
- [14] Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. Applied Soft Computing, 8(1), 687-697.
- [15] Algarni, A., Y. Li and X. Tao, "Mining specific and general features in both positive and negative relevance feedback". Proceedings of the

- 19th Text REtrieval Conference: Relevance Feedback Track, Nov. 2010. pg. 16-19,
- [16] Ramadhan, H.A., K. Shihab and J.H. Ali., Improving the ranking capability of the hyperlink based search engines using heuristic approach. *J. Comput. Sci.*, 2006 : 638-645. DOI: 10.3844/jcssp.2006.638.645
 - [17] Zhao, H., Pei, Z., Jiang, J., Guan, R., Wang, C., Shi, X.,. A hybrid swarm intelligent method based on genetic algorithm an artificial bee colony. In: *Advances in Swarm Intelligence*. springer, 2010 pp. 558–565.
 - [18] Luo, J., Wang, Q., Xiao, X., “A modified artificial bee colony algorithm based on converge-onlookers approach for global optimization”. *Appl. Math. Computat.* 2013, 219 (20), 10253– 10262.
 - [19] Xiang, W.-L., An, M.-Q., 2013. An efficient and robust artificial bee colony algorithm for numerical optimization. *Comput. Operat. Res.* 40 (5), 1256–1265.
 - [20] Kong, X., Liu, S., Wang, Z., 2013. An improved artificial bee colony algorithm and its application. *Int. J. Signal Process. Image process. Pattern Recognit.* 2013 , 6 (6).
 - [21] Aderhold, A., Diwold, K., Scheidler, A., Middendorf, M., 2010. Artificial bee colony optimization: a new selection scheme and its performance. In: *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*. Springer, pp. 283–294.
 - [22] Stanarevic, N., Tuba, M., Bacanin, N., 2010. Enhanced artificial bee colony algorithm performance. In: *Proceedings of the 14th WSEAS International Conference on Computers: Part of the 14th WSEAS CSCC Multiconference*, vol. 2, pp. 440–445.
 - [23] Stanarevic, N., 2011. Comparison of different mutation strategies applied to artificial bee colony algorithm. In: *Proceedings of the European computing conference (ECC11)*, pp. 257–262.
 - [24] Lei, X., Huang, X., Zhang, A., 2010. Improved artificial bee colony algorithm and its application in data clustering. In: *Bio- Inspired Computing: Theories and Applications (BIC-TA)*, 2010 IEEE fifth International Conference on. IEEE, pp. 514–521.
 - [25] R. Haupt and S. Haupt, “The binary genetic algorithm,” *Pract. Genet. Algorithms*, Second, pp. 27–50, 1998.
 - [26] M. Dorigo and G. Di Caro, “The Ant Colony Optimization Meta-Heuristic,” pp. 11–32, 1999.
 - [27] M. Dorigo and T. Stützle, *Ant Colony Optimization*. 2004.
 - [28] J. Kennedy and R. Eberhart, “Particle swarm optimization,” *Neural Networks, Proceedings., IEEE Int. Conf.*, 1995.
 - [29] J. Kennedy and R. Eberhart, “Particle swarm optimization,” *Neural Networks., Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.
 - [30] R. Poli, J. Kennedy, and T. Blackwell, “Particle swarm optimization,” *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, 2007.
 - [31] D. T. Pham, D. T. Pham, a. Ghanbarzadeh, a. Ghanbarzadeh, E. Koc, E. Koc, S. Otri, S. Otri, S. Rahim, S. Rahim, M. Zaidi, and M. Zaidi, 2008.
 - [32] “The bees algorithm—a novel tool for complex optimisation problems,” *Intell. Prod. Mach. Syst. 2nd I• PROMVirtual Conf.* 3-14 July 2006, p. 454, 2006
 - [33] I. Fister, X. S. Yang, and D. Fister, “Cuckoo search: A brief literature review,” *Stud. Comput. Intell.*, vol. 516, pp. 49–62, 2014.
 - [34] X. S. Yang and S. Deb, “Cuckoo search via Lévy flights,” 2009 World Congr. Nat. Biol. Inspired Comput. NABIC 2009 - Proc., pp. 210–214, 2009.
 - [35] K. M. Passino, “Bacterial foraging optimization,” *Int. J. Swarm Intell. Res.*, vol. 1, no. 1, pp. 1–16, 2010.
 - [36] P. Wang, Z. Zhu, and S. Huang, “A New Meta- Heuristic Technique for Engineering Design Optimization : Seven- Spot Ladybird Algorithm,” in *2nd International Symposium on Computer, Communication, Control and Automation*, 2013, pp. 387–392.
 - [37] M. J. Lehane, “The Biology of Blood-Sucking in Insects,” *Liverpool Sch. Trop. Med.*, p. 321, 2005.
 - [38] Kim, H.R. and P.K. Chan., Personalized ranking of search results with learned user interest hierarchies from bookmarks. *Proceedings of the 11th SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 21-21, Chicago, Illinois, USA. 2005
 - [39] Hu, J. and P.K. Chan, 2008. Personalized web search by using learned user profiles in re-ranking. *Florida Institute of Technology, USA*.
 - [40] Teevan, J., S.T. Dumais and E. Horvitz., Personalizing search via automated analysis of interests and activities. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 15-19, ACM Press, Salvador, Brazil, pp: 449-456. DOI: 10.1145/1076034.1076111, 2005.
 - [41] Kumar, P.R. and A.K. Singh., Web Structur Mining: Exploring Hyperlinks and Algorithms for Information Retrieval. *Am. J. Applied Sci.*, 7: 840-845. DOI: 10.3844/ajassp.2010.840.845, 2010.
 - [42] Derhami V, Khodadadian E, Ghasemzadeh M, ZarehBidoki AM (2013) Applying reinforcement learning for web pages ranking algorithms. *Appl Soft Comput* 13(4):1686–1692, 2013.
 - [43] Roobam R, Vallimayli V (2014) Survey on ontology based semantic web usage mining for enhanced recommendation model. *Int J Sci Eng Res* 5(5):1164–1169, 2014.
 - [44] Kaviarasan S, Hemapriya K, Gopinath K (2015) Semantic web usage mining techniques for predicting users’ navigation requests. *Int J Innov Res Comput Commun Eng* 3(5):4261–4270, 2015.
 - [45] Moreno MN, Segrera S et al (2015) Web mining based framework for solving usual problems in recommender systems: a case study for movies recommendation. *Neurocomputing J* 176(C):73–80, 2015.
 - [46] Bairagade R, Singh N, Afre N, Bhamare D (2015) A survey on SmartCrawler: a deep-web harvesting approach. *Int J Adv Res Comput Eng Technol* 4(9):3611–3614, 2015.
 - [47] Sharma S, Lodhi SS (2016) Development of decision tree algorithm for mining web data stream. *Int J Comput Appl* 138(2):34–43, 2016.
 - [48] Aqlan HAA, Ahmed S, Danti A (2017) Death prediction and analysis using web mining techniques. In: *ICACCS, Coimbatore, India*, 978-1-5090-4559-4/17/\$31-EEE, 2017.