REFERENCES

[1]   J. Srivastava and R. Cooley, "Web usage mining: Discovery and applications of usage patterns from web data," *ACM SIGKDD*, pp. 12–23, 2000.

[2]   P. Chwalinski, "Detection of Unsolicited Web Browsing with Clustering and Statistical Analysis," University of Middlesex, London, 2013.

[3]   B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and evaluation of aggregate usage profiles for Web personalization," *Data Min. Knowl. Discov.*, vol. 6, no. 1, pp. 61–82, 2002.

[4]   Sunena and Kamaljit kaur, "Web Usage Mining-Current Trends and Future Challenges," in *International Conference on Electrical, Electronics, and Optimization Techniques 2016*, 2016.

[5]   F. M. Facca and P. L. Lanzi, "Mining interesting knowledge from weblogs: A survey," *Data Knowl. Eng.*, vol. 53, no. 3, pp. 225–241, 2005.

[6]   R. Kosala and H. Blockeel, "Web Mining Research: A Survey," *ACM Sigkdd Explor. Newsl. 2.1*, vol. 2, no. 1, 2000.

[7]   C. Ghezzi, M. Pezzè, M. Sama, and G. Tamburrelli, "Mining behavior models from user-intensive web applications," *Proc. 36th Int. Conf. Softw. Eng. - ICSE 2014*, pp. 277–287, 2014.

[8]   N. M. Khairudin, A. Mustapha, and M. H. Ahmad, "Effect of Temporal Relationships in Associative Rule Mining for Web Log Data," *Sci. World J.*, vol. 2014, 2014.

[9]   J. Pei, J. Han, B. Mortazaviasl, and H. Zhu, "Mining Access Patterns Effciently from Web," *Knowl. Discov. Data Mining. Curr. Issues New Appl.*, 1999.

[10]   K. S. Y. Fu, M.Y. Shih, "Clustering of Web Users Based on Access Patterns," *KDD Work. Web Mining. San Diego, CA. Springer-Verlag*, 1999.

[11]   Lukas Cenovsky, "Web Usage Mining on is.muni.cz," Masaryk University, 2003.

[12]   V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, p. 74, 2009.

[13]   B. Aaron and D. E. Tamir, "Dynamic Incremental K-means Clustering," in *Computational Science and Computational Intelligence (CSCI)*, 2014, pp. 308–313.

[14]   W. Wang and O. R. Zaiane, "Clustering Web sessions by sequence alignment," *13th Int. Work. Database Expert Syst. Appl. 2002. Proc.*, pp. 394–398, 2002.

[15]   R. Cooley, B. Mobasher, and J. Srivastava, "Web mining: Information and pattern discovery on the world wide web," *Tools with Artif. Intell. 1997. Proceedings., Ninth IEEE Int. Conf.*, no. DECEMBER 1997, pp. 558–567, 1997.

[16]   M. Etzioni and M. Perkowitz, "Adaptive Web Sites: Automatically Synthesizing Web Pages," *Proc. AAAI*, 1998.

[17] D. Ngu and X. Wu, "Sitehelper: A localized agent that helps incremental exploration of the world wide web," *Comput. Networks ISDN Syst.*, 1997.

[18] T. Joachims, D. Freitag, and T. Mitchell, "Webwatcher: A tour guide for the world wide web," *IJCAI (1)*, 1997.

[19] P. Kumar, R. Bapi, and P. Krishna, "SeqPAM: a sequence clustering algorithm for Web personalization," *nternational J. Data Warehous. Min.*, 2007.

[20] M. Eirinaki and M. Vazirgiannis, "Web mining for web personalization," *ACM Trans.Inter.Tech.*, vol. 3, no. 1, pp. 101–127, 2003.

[21] Y. Xie and S. Tang, "Online Anomaly Detection Based on Web Usage Mining," in *2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum*, 2012, pp. 1177–1182.

[22] H. Liu, Z. Huang, H. Zhang, and J. Lv, "A Security Analysis of Web Users' Behavior Based on Web Log Mining," pp. 89–92, Nov. 2013.

[23] P. Wang, X. Ma, and J. Yu, "An Effective Network Security Log Mining Algorithm based on Fuzzy Clustering," *Appl. Math. Inf. Sci.*, vol. 1, no. 1, pp. 307–315, 2016.

[24] M. Rai and V. Mishra, "Detection of UDP and HTTP Anomalies on Real Time Traffic Based on NIDS using OURMON Tool," *Int. J. Sci. Res. Sci. Eng. Technol.*, 2015.

[25] A. Tang, S. Sethumadhavan, and S. Stolfo, "Unsupervised anomaly-based malware detection using hardware features," *Res. Attacks, Intrusions Defenses. Springer Int. Publ.*, pp. 109–129, 2014.

[26] A. Juan-Verdejo and H. Baars, "Decision support for partially moving applications to the cloud: the example of business intelligence," in *Proceedings of the 2013 international workshop on Hot topics in cloud services. ACM*, 2013.

[27] A. Abraham, "i-miner: A web usage mining framework using hierarchical intelligent systems," in *Fuzzy Systems, 2003. FUZZ'03. The 12th IEEE International Conference on*, 2003.

[28] D. Martin, "Abstract for the Invited Talk: The IBM SurfAid Project: Transactive Analysis and Prediction," 1999.

[29] R. W. Cooley, "Web Usage Mining: Discovery and Application of Interesting Patterns from web Data," University of Minnesota, 2000.

[30] V. Losarwar and M. Joshi, "Data Preprocessing in Web Usage Mining," *Int. Conf. Artif. Intell. Embed. Syst.*, 2012.

[31] T. Aye, "Web log cleaning for mining of web usage patterns," *Comput. Res. Dev. (ICCRD), 2011 3rd Int. Conf. on. IEEE*, vol. 2, 2011.

[32] S. Alspaugh, A. Ganapathi, M. Hearst, and R. Katz, "Better Logging to Improve Interactive Data Analysis Tools," *ACM SIGKDD Work. Interact. Data Explor. Anal.*, 2014.

[33] A. Nanopoulos, Y. Manolopoulos, M. Zakrzewicz, and T. Morzy, "Indexing web access-logs for pattern queries," *Proc. fourth Int. Work. Web Inf. data Manag. - WIDM '02*, p. 63, 2002.

[34] R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," *Knowl. Inf. Syst.*, vol. 1, no. 1, pp. 5–32, Jul. 1999.

[35] P. Sharma, S. Yadav, and B. Bohra, "A review study of server log formats for efficient web mining," in *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on. IEEE*, 2015, pp. 1373–1377.

[36] P. Nithya and P. Sumathi, "An enhanced pre-processing technique for web log mining by removing web robots," in *2012 IEEE International Conference on Computational Intelligence and Computing Research*, 2012, pp. 1–4.

[37] N. Algiriyage, S. Jayasena, G. Dias, A. Perera, and K. Dayananda, "Identification and characterization of crawlers through analysis of web logs," in *2013 IEEE 8th International Conference on Industrial and Information Systems*, 2013, pp. 150–155.

[38] L. Zheng, H. Gui, and F. Li, "Optimized data preprocessing technology for web log mining," *Comput. Des. Appl. (ICCDA), 2010 Int. Conf. on. IEEE*, vol. 1, 2010.

[39] J. Chang-bin and C. Li, "Web Log Data Preprocessing Based On Collaborative Filtering," *Educ. Technol. Comput. Sci.*, vol. 2, 2010.

[40] Y. Li, B. Feng, and Q. Mao, "Research on path completion technique in web usage mining," *Comput. Sci. Comput. Technol. 2008. ISCSCT 08. Int. Symp. on. IEEE*, vol. 1, 2008.

[41] T. Hussain, S. Asghar, and N. Masood, "Hierarchical sessionization at preprocessing level of WUM based on swarm intelligence," *Emerg. Technol. (ICET), 6th Int. Conf. on. IEEE*, 2010.

[42] D. Tanasa and B. Trousse, "Advanced data preprocessing for intersites web usage mining," *Intell. Syst. IEEE*, 2004.

[43] Y. Fang and Z. Huang, "A session identification algorithm based on frame page and pagethreshold," *Proc. - 2010 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol. ICCSIT 2010*, vol. 6, pp. 645–647, 2010.

[44] H. Sha, T. Liu, P. Qin, Y. Sun, and Q. Liu, "EPLogCleaner : Improving Data Quality of Enterprise Proxy Logs for E ffi cient Web Usage Mining," vol. 17, pp. 812–818, 2013.

[45] J. Vellingiri and S. C. Pandian, "A Novel Technique for Web Log mining with Better Data Cleaning and Transaction Identification," vol. 7, no. 5, pp. 683–689, 2011.

[46] B. Thakur, S. Abbas, M. Beg, and S. Rizvi, "Preprocessing of web usage data for log analysis," *Int. J. Sci. Eng. Res.*, pp. 1773–1779, 2015.

[47] M. Spiliopoulou, "A framework for the evaluation of session reconstruction heuristics in web-usage analysis," *Informs J. Comput. 15.2*, pp. 171–190, 2003.

[48] M. Baglioni, U. Ferrara, A. Romei, S. Ruggieri, and F. Turini, "Preprocessing and Mining Web Log Data for Web Personalization," *AI* IA 2003 Adv. Artif. Intell. Springer Berlin Heidelb.*, pp. 237–249, 2003.

[49] P. Langley, "User modeling in adaptive interfaces," *Proc. Seventh Int. Conf. User Model.*, 1999.

[50] N. Singh, A. Jain, and R. Raw, "Comparison analysis of web usage mining using pattern recognition techniques," *Int. J. Data Min. Knowl. Manag. Process 3.4*, pp. 137–158, 2013.

[51] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Rec.*, vol. 22, no. May, pp. 207–216, 1993.

[52] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *Data Min. Knowl. Discov.*, vol. 15, no. 1, pp. 55–86, Jan. 2007.

[53] X. Huang, A. An, and N. Cercone, "Comparison of interestingness functions for learning web usage patterns," *Int. Conf. Inf. Knowl. Manag. Proc.*, no. March, pp. 617–620, 2002.

[54] B. Mortazavi-Asl, "Discovering and mining user web-page traversal patterns," Simon Fraser University, 2001.

[55] R. Srikant and R. Agrawal, "Mining sequential patterns: generalizations and performance improvements," *Proc. 5th Int. Conf. Extending Database Technol.*, pp. 3–17, 1996.

[56] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on web usage mining," *Commun. ACM*, 2000.

[57] D. Stevanovic, N. Vlajic, and A. An, "Unsupervised clustering of web sessions to detect malicious and non-malicious website users," *Procedia Comput. Sci.*, vol. 5, pp. 123–131, 2011.

[58] M. Ester, H. Kriegel, X. Xu, and D.- Miinchen, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Kdd. Vol. 96*, 1996.

[59] S. Kisilevich, F. Mansmann, and D. Keim, "P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos," *Proc. 1st Int. Conf. Exhib. Comput. Geospatial Res. Appl. - COM.Geo '10*, p. 1, 2010.

[60] A. Denton, "Kernel-density-based clustering of time series subsequences using a continuous random-walk noise model," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 122–129, 2005.

[61] M. Parimala, D. Lopez, and N. Senthilkumar, "A survey on density based clustering algorithms for mining large spatial databases," *Int. J. Adv. Sci. Technol. 31.1*, vol. 31, pp. 59–66, 2011.

[62] C. Fraley and A. E. Raftery, "How Many Clusters? Which Clustering Method?

Answers Via Model-Based Cluster Analysis," *Comput. J.*, vol. 41, no. 8, pp. 578–588, 1998.

[63] N. Mustapha, M. Jalali, A. Bozorgniya, and M. Jalali, "Navigation Patterns Mining Approach based on Expectation Maximization Algorithm," *Eur. J. Sci. Res.*, vol. 3, no. 2, pp. 855–859, 2009.

[64] R. C. Barros, M. P. Basgalupp, A. C. P. L. F. de Carvalho, and M. G. Quiles, "Clus-DTI: Improving decision-tree classification with a clustering-based decision-tree induction algorithm," *J. Brazilian Comput. Soc.*, vol. 18, no. 4, pp. 351–362, 2012.

[65] L. Bolelli, S. Ertekin, D. Zhou, and C. L. Giles, "K-SVMeans: A Hybrid Clustering Algorithm for Multi-Type Interrelated Datasets," *IEEE/WIC/ACM Int. Conf. Web Intell.*, pp. 198–204, Nov. 2007.

[66] H. Frigui, "SyMP : An Efficient Clustering Approach to Identify Clusters of Arbitrary Shapes in Large Data Sets," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM*, 2002.

[67] Y. Xie and V. V. Phoha, "Web user clustering from access log using belief function," *Proc. Int. Conf. Knowl. capture - K-CAP 2001*, no. JANUARY 2001, p. 202, 2001.

[68] A. Banerjee and J. Ghosh, "Clickstream clustering using weighted longest common subsequences," *Proc Work. Web Min. SIAM Conf. Data Min.*, pp. 33–40, 2001.

[69] Y. Chen, "Improving user profiles for e-commerce by genetic algorithms," *Stud. Fuzziness Soft*, vol. 105, pp. 3–6, 2002.

[70] S. Oyanagi, K. Kubota, and A. Nakase, "Application of Matrix Clustering to Web Log Analysis and Access Prediction," *Proc. WEBKDD.*, vol. Vol. 1, 2001.

[71] S. Samangooei, J. Hare, D. Dupplaw, M. Niranjan, N. Gibbins, and P. Lewis, "Social event detection via sparse multi-modal feature selection and incremental density based clustering," *CEUR Workshop Proc.*, vol. 1043, pp. 18–19, 2013.

[72] L. Chaofeng, "Research on Web Session Clustering," *J. Softw. 4.5*, vol. 4, no. 5, pp. 460–468, 2009.

[73] J. S. Cooley, Robert, Pang-Ning Tan, "Discovery of interesting usage patterns from web data," *Web Usage Anal. User Profiling. Springer Berlin Heidelb.*, pp. 163–182, 1999.

[74] S. Chakraborty and N. K. Nagwani, "Analysis and Study of Incremental DBSCAN Clustering Algorithm," *Int. J. Enterp. Comput. Bus. Syst.*, vol. 1, no. 2, 2011.

[75] K. Mumtaz, "A Novel Density based improved k-means Clustering Algorithm–Dbkmeans," *Int. J.*, vol. 02, no. 02, pp. 213–218, 2010.

[76] M. Eirinaki, C. Lampos, S. Paulakis, and M. Vazirgiannis, "Web personalization integrating content semantics and navigational patterns," *Proc. 6th Annu. ACM Int. Work. Web Inf. data Manag. WIDM 04*, no. JANUARY, p. 72, 2004.

[77] Z. Ansari, W. Ahmed, M. F. Azeem, and A. V. Babu, "Discovery of Web Usage Profiles Using Various Clustering Techniques," vol. 1, no. 3, pp. 18–27, 2011.

[78] M. Eirinaki, M. Vazirgiannis, and I. Varlamis, "SEWeP," *Proc. ninth ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '03*, no. JANUARY, p. 99, 2003.

[79] P. Fournier-viger, C. Wu, V. S. Tseng, L. Cao, and R. Nkambou, "Mining Partially-Ordered Sequential Rules Common to Multiple Sequences," vol. 27, no. 8, pp. 2203–2216, 2015.

[80] R. Agrawal and R. Srikant, "Mining sequential patterns," in *Proceedings of the Eleventh International Conference on Data Engineering*, 1995.

[81] A. I. V. Mannila, Heikki, Hannu Toivonen, "Discovering frequent episodes in sequences Extended abstract," in *The first Conference on Knowledge Discovery and Data Mining*, 1995.

[82] J. Pei, J. Han, Q. Chen, M.-C. Hsu, B. Mortazavi-Asl, H. Pinto, and U. Dayal, "PrefixSpan,: mining sequential patterns efficiently by prefix-projected pattern growth," *Int. Conf. Data Eng.*, pp. 215 – 224, 2001.

[83] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M. C. Hsu, "Mining sequential patterns by pattern-growth: The prefixspan approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1424–1440, 2004.

[84] E. Menasalvas, S. Millán, J. M. Peña, M. Hadjimichael, and O. Marbán, "Subsessions: A granular approach to click path analysis," *Int. J. Intell. Syst.*, vol. 19, no. 7, pp. 619–637, 2004.

[85] P. K. Alkan, Oznur Kirmemis, "Assisting Web Site Navigation Through Web Usage Patterns," *Recent Trends Appl. Artif. Intell. Springer Berlin Heidelberg, 2013.*, pp. 161–170, 2013.

[86] K. Bommepally, T. K. Glisa, J. J. Prakash, S. R. Singh, and H. a Murthy, "Internet activity analysis through proxy log," *2010 Natl. Conf. Commun.*, pp. 1–5, Jan. 2010.

[87] S. Chakrabarti, "Data mining for hypertext: A tutorial survey," *SIGKDD Explor.*, vol. 1, no. 2, pp. 1–11, 2000.

[88] A. Jain and R. Dubes, "Algorithms for clustering data," 1988.

[89] S. Guha, R. Rastogi, and K. Shim, "ROCK: A robust clustering algorithm for categorical attributes," *Data Eng. 1999. Proceedings., 15th Int. Conf. on. IEEE*, 1999.

[90] K. a. Smith and A. Ng, "Web page clustering using a self-organizing map of user navigation patterns," *Decis. Support Syst.*, vol. 35, no. 2, pp. 245–256, May 2003.

[91] M. A. W. Hartigan, John A., "A K-Means Clustering Algorithm," *Appl. Stat.*, pp. 100–108, 1979.

[92] Z. He, X. Xu, and S. Deng, "Discovering cluster-based local outliers," *Pattern Recognit. Lett.*, 2003.

[93]   A. Chaudhary, A. Szalay, and A. Moore, "Very Fast Outlier Detection in Large Multidimensional Data Sets.," *DMKD*, 2002.

[94]   A. Pires and C. Santos-Pereira, "Using clustering and robust estimators to detect outliers in multivariate data," *Proc. Int. Conf. Robust Stat.*, 2005.

[95]   H. Sun, Y. Bao, F. Zhao, G. Yu, and D. Wang, "CD-trees: An efficient index structure for outlier detection," *Adv. Web-Age Inf. Manag. Springer Berlin Heidelb.*, pp. 600–609, 2004.

[96]   G. Carl, G. Kesidis, R. Brooks, and S. Rai, "Denial-of-service attack-detection techniques," *Internet Comput. IEEE*, 2006.

[97]   B. Schölkopf, J. Platt, and J. Shawe-Taylor, "Estimating the support of a high-dimensional distribution," *Neural Comput. 13.7*, pp. 1443–1471, 2001.

[98]   V. Roth, "Kernel fisher discriminants for outlier detection," *Neural Comput.*, 2006.

[99]   F. J. Anscombe and I. Guttman, "Rejection of Outliers," *Technometrics*, vol. 2, no. 2, pp. pp. 123–147, 1960.

[100] H Motulsky, "Choosing a statistical test," in *Intuitive Biostatistics*, Oxford University Press Inc, 1995, pp. 1–5.

[101] A. Kind, M. Stoecklin, and X. Dimitropoulos, "Histogram-based traffic anomaly detection," *IEEE Trans. Netw. Serv. Manag.*, vol. 6, no. 2, pp. 110–121, 2009.

[102] A. Patcha and J. M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

[103] W. Fan, M. Miller, S. Stolfo, W. Lee, and P. Chan, "Using artificial anomalies to detect unknown and known network intrusions," *Knowl. Inf. Syst. 6.5*, pp. 507–527, 2004.

[104] P. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. 2006.

[105]   and G. N. Hipp, Jochen, Ulrich Güntzer, "Algorithms for Association Rule Mining – A General Survey and Comparison," *ACM sigkdd Explor. Newsl. 2.1*, pp. 58–64, 2000.

[106] M. Qin and K. Hwang, "Frequent episode rules for internet anomaly detection," *Netw. Comput. Appl. (NCA 2004). Proceedings. Third IEEE Int. Symp. on. IEEE*, 2004.

[107] S. Cho and S. Cha, "SAD: web session anomaly detection based on parameter estimation," *Comput. Secur.*, vol. 23, no. 4, pp. 312–319, 2004.

[108] U. Manber and G. Myers, "Suffix arrays: A new method for on-line string searches," 1991.

[109] T. Kasai, G. Lee, H. Arimura, S. Arikawa, and K. Park, "Computation in Suffix Arrays and Its Applications," *Comb. pattern matching. Springer Berlin Heidelb.*, pp. 181–192, 2001.

[110] D. E. Goldberg, *Genetic Algorithms in Search , Optimization , and Machine Learning*. Pearson Education India, 2006.

[111] K. R. R. Makwana, Chintan H., "An Efficient Technique for Web Log Preprocessing using Microsoft Excel," *Int. J. Comput. Appl.*, vol. 90, no. 12, pp. 25–28, 2014.

[112] T. Jing, W.-L. Zuo, and B.-Z. Zhang, "An efficient Web traversal pattern mining algorithm based on suffix array," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*, vol. 4, pp. 1535–1539.

[113] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov, "Spamming botnets," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, p. 171, 2008.

[114] M. D. Tianyi Liang, Andrew Reynolds, Cesare Tinelli, Clark Barrett, "A DPLL (T) Theory Solver for a Theory of Strings and Regular Expressions," *Comput. Aided Verif. Springer Int. Publ.*, pp. 1–22, 2014.

[115] R. Cox, "Regular Expression Matching in the Wild," no. March, 2010.

[116] M. Becchi and P. Crowley, "Extending finite automata to efficiently match Perl-compatible regular expressions," *Proc. 2008 ACM Conex. Conf. - Conex. '08*, no. March, pp. 1–12, 2008.

[117] G. Skinner, "RegExr," 2011. [Online]. Available: http://regexr.com/. [Accessed: 12-Mar-2016].

[118] J. Goyvaerts, "RegexBuddy," *Just Great Software*, 2010. [Online]. Available: https://www.regexbuddy.com/. [Accessed: 12-Mar-2016].

[119] J. Cho and S. Rajagopalan, "A fast regular expression indexing engine," *Data Eng. 2002. Proceedings. 18th Int. Conf.*, vol. 3, pp. 419–430, 2002.

[120] D. K. Pham, Duc, *Intelligent optimisation techniques: genetic algorithms, tabu search, simulated annealing and neural networks*. Springer Science & Business Media, 2012.

[121] L. Wang, K. Huang, J. Zhang, and J. Yao, "A Complete Suffix Array-Based String Match Search Algorithm of Sliding Windows," *2012 Fifth Int. Symp. Comput. Intell. Des.*, pp. 210–213, Oct. 2012.

[122] G. Sidorov, F. Velasquez, E. Stamatatos, and A. Gelbukh, "Syntactic n-grams as machine learning features for natural language processing," *Expert Syst. with Appl. 41.3*, no. Cic, pp. 853–860, 2014.

[123] V. Arnau, R. Moreno, D. Cazorla, I. Medina, U. De Valencia, and D. De, "Acceleration of short and long DNA read mapping without loss of accuracy using suffix array," vol. 30, no. 23, pp. 3396–3398, 2014.

[124] V. K. P. Sidhu, Reetinder, "Regular expression matching can be simple and fast," *Perls Dev. Conf.*, 2007.

[125] J. H. Rosenberg, Andrew, "V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure," *EMNLP-CoNLL. Vol. 7*, 2007.

APPENDIX A: SOURCE CODE

The source code and the libraries which were used with the software have been included in the attached compact disc. A guide on how to install the software for testing has been included with the software source code.