International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 532-535 Jan 2019-June 2019 www.csjournals.com

A Voice Based Effective Content Mining and Indexing for Multimedia Data

[1]Ms. Chetana M. Tatte, [2]Prof. Dr. G.D. Dalvi, [3]Prof. Dr. S. D. Wakade
[1]M.E. Scholar, [2]Associate Professor, [3]Principal
EXTC, P. R. Pote College of Engg. & Management, Amravati

Abstract: Research in peer-to-peer file sharing systems has targeted on coping with the design constraints encountered in distributed systems, whereas little or no attention has been dedicated to the user experience: these systems constantly assume the user is conscious of the regarding the file they are wanting. But average users rarely even apprehend that file exists. File sharing systems that do ponder the user experience and allow users to travel searching for files by their name, usually gift centralized management which they show several severe vulnerabilities that make the system unreliable and insecure. The aim of this methodology is to vogue a further complete distributed file sharing system that is not entirely trustful, climbable and secure, but in addition leverages the user's psychological feature employment. We have got an inclination to gift a method that by mining a file's information designates relevant keywords for the file automatically.

Subject headings: Voice Recognition, Data Mining, Text clustering, Pattern Matching, Convolution, Cross Co-relation

INTRODUCTION

Basically analysis of mining tool is that the strategy of creating judgment regarding the worth, importance and quality of mining tool, once considering Mining tools fastidiously. The analysis of Mining tools has not been maintaining with the advancement of their development. Mining tools work otherwise supported altogether completely different mode of interface, features, coverage of the ranking ways that during which, delivery of advertising and far of any such factors. it's strong to gauge them on one basis. There are some ways for evaluating Mining tools like automatic analysis; human connection judgment primarily based completely analysis.

The aim of this paper is to review the Mining tool analysis ways that propose Associate in Nursing exaggerated methodology for evaluating Mining tools. Peer-to-peer systems are a specific variety of distributed systems, wherever all computers, additionally spoken as nodes, gift identical responsibilities and capabilities. Peer-to- peer systems have several edges over ancient centralized systems: they gift higher accessibility, quality, and fault tolerance, lower maintenance prices additionally as lower operation and preparation prices. The balk of those systems is that they encounter many vogue challenges. as associate degree example the system got to keep wise, despite the variable vary of uncontrolled collaborating nodes. in addition, despite the systems size, data search on peer-to-peer systems got to be quick and sturdy (scalable).

OBJECTIVES

[1] Mrs. Chetana M. Tatte is a M.E Scholar doing her study from P.R Pote College of Engineering and Technology, Amravati. She is interested in audio signal processing, data mining and doing here research in text data mining for physically handicap users.

[2] Prof. Dr. Gopal D. Dalvi is a doctorate in Electronics and telecommunication. His Main Research area is to secure wireless data transmission using visual cryptography.

The aim of this paper is information mining is to get structure within unstructured data, extract that means from crying knowledge, discover patterns in apparently random knowledge, and use all this info to raised perceive trends, patterns, correlations, and ultimately predict client behavior, market and competition trends, in order that the corporate uses its own knowledge additional meaningfully to raised position itself on the new waves.

LITERATURE REVIEW

George et al. (2007) urged a model to gauge Min-ing tools on the press through data of past users. The model used two variables i.e. A(attractively) and C(consideration) to ascertain the probability of choosing a chunk out of the list of relevant pages through that he successes to a distance d ; once considering up to distance d- 1 elements. The conclusion of study shows that the area model represents the data over quality model. The complete analysis illustrates that the purpose biasing of connectedness is also resolved by click through data. Here it's attending to seem counter-intuitive to use this model to measure performance. This toy model is unable to represent clearly the user behavior but it's going to be any improved to implement click through data methods.

Ya-Lan et al.(2007) projected two major problems hygiene issue and motivation issue. Hygiene factors area unit those plenty of elementary wants for a Mining tool and build users willing to use a Mining tool, and motivation factors area unit those plenty of more services of a Mining tool and build users willing to remain exploitation an analogous Mining tool.

Rashid et al.(2009) devised AN automatic search analysis system supported rough set based totally rank aggregation technique. Basically, all fully completely different ranking results obtained from different techniques unit combined. 2 phases unit used, ranking rules learning half and rank aggregation half. Author used fifteen queries in rank learning half. The output of this half might be a collection of ranking rules.

Maninder et al.(2011) compared and evaluated five Mining tools (Google, yahoo, being ask, AltaVista) on the premise of their search capabilities into two sections[9]. at intervals the first section, choices of five Mining tools area unit compared that area unit gettable to the user whereas looking the info.

LIMITATIONS OF EXISTING SYSTEM

Data mining remains associate art, requiring skillful analysis and selection of methodology. It needs branch of knowledge experience, expertise with giant information bases, and skills with data processing algorithms. As indicated by the list higher than, the techniques get-ting used come back from existing classical statistics or AI techniques. Issues encountered with information together with over-fitting existing information, missing and droning information, and coping with terribly giant databases and extremely high spatially have to be compelled to be resolved. Moreover, for large-scale, real- world tasks, high activity algorithms (such as neural net - works and genetic algorithms) should deal with long com- putation times and difficulties in creating interpretations. Techniques related to probabilistic learning have to be compelled to be improved.

PROPOSED METHODOLOGY

Data Flow Diagram

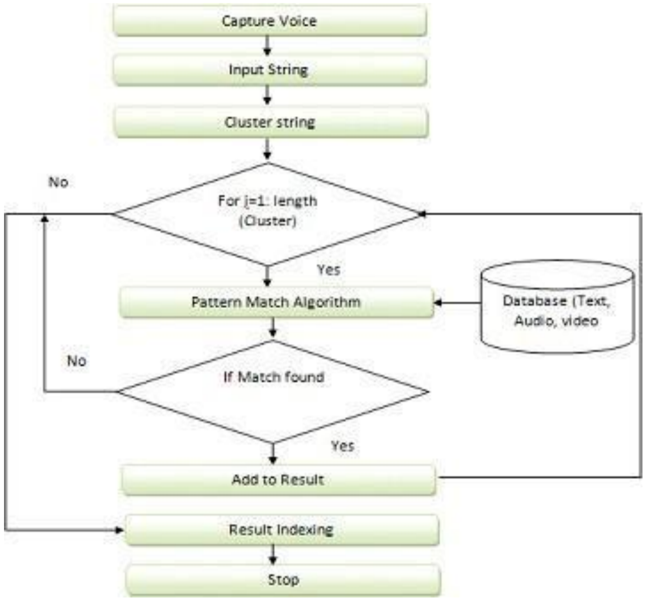


Fig. 1. Data Flow Diagram

Voice Recognition

Our aim is to facilitate data mining to those who are physically handicap and also to those who are not aware with computer system. A voice recognition step will au- tomatically recognize voice and convert it into text. to achieve this , a convolution theorem for cross correlationis used as below let samples sets of voice in database represented as

$S = [s1, s2, s3, s4.....sn]$

User input sample represented with

$Us = [u1, u2, u3, u4, ...nn]$
likewise fx is calculated for all samples values in data set. fx indicated correlation with User Samples.

$f\,x = [f\,1, f\,2, f\,3, f\,4, ...fn]$

$$BestM\,t = Sort(f\,x)$$

User Sample with Max Match is considered as recog-nize voice and then its corresponding text is loaded from database which formulate query string (Q).

QUERY CLUSTERING

Length of clusters depends on

$L\,(Q)$

Clustering is a decomposition of Query string into sev- eral sub string with its length as a Euclidean Distance.

$f \qquad C = (C1, C2, C3, C4....Cn)$

Where

$L(C) = L(Q)$

$L\,(C1) = L(Q) - 1$

$L\,(C2) = L(Q) - 2$

$L\,(Cn) = L(Q) - n$

or length wise or possibilities wise cluster string can also be represented as

$C > C1 > C2 > C3 > ..... > Cn$

DATA MINING IN NON RELATIONAL DATABASE

Non Relational Databases are called as files where no characters or string have relation with each other. Find- ing required data pattern in these kinds of files is really a challenging task. To get mining result, we decomposed file data into size equal to size of cluster query string.

$$F\,ileD_{a}^{at} = Load(F\,ileT_{ext})$$

where $F\,ileC_{luster} = f1, f2, f3, f4, ...fn$

$L(f123....n) = L(Q)$
if $(Q(C)== f(C))$
$Patternmatch = P\,atternmatch + 1$
end
User sample is cross correlated with Samples available in database as
$f\,x = Us\varepsilon S$

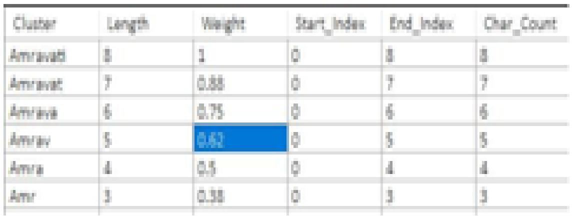
RESULT RANKING

Result Ranking is a way to display result as per user re- quirement. Generally it is found that, Google like display result with multi parameters based like pattern found, File Size, Mining Time,Waiting time etc...

FinalRank = Sort[P1, P2, P3, ...Pn]
where P1, P2...P atternM atchCount

RESULT ANALYSIS

found max- imum match in page. Means Page with max content will display at top of result page and likewise in decreasing order. We also adopt approximately same criteria to



Cluster	Length	Weight	Start_Index	End_Index	Char_Count
Amravati	8	1	0	8	8
Amravat	7	0.88	0	7	7
Amrava	6	0.75	0	6	6
Amrav	5	0.62	0	5	5
Amra	4	0.5	0	4	4
Amr	3	0.38	0	3	3

Fig. 2.— Query String Clusters

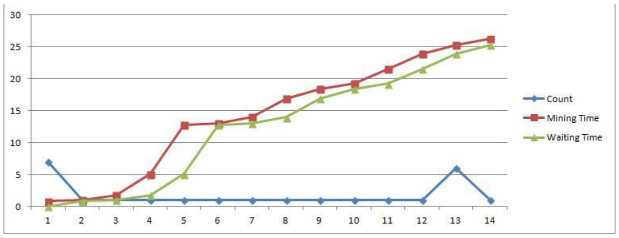


Fig.3. Data Mining

REFERENCES

[1] A.M. Cohen and W. R. Hersh, A Survey of Current Work in Biomedical Text Mining, Briefings in Bioinformatics, vol. 6, no. 1,pp. 5771, Mar. 2005.

[2] M. Huang, X. Zhu, Y. Hao, D. G. Payan, K. Qu, and M. Li, Discovering Patterns to Extract ProteinProtein Interactions from Full Texts, Bioinformatics, vol. 20, no. 18, pp. 36043612, Jul. 2004.

[3] J. Czarnecki, I. Nobeli, A. M. Smith, and A. J. Shepherd, A Text-Mining System for Extracting Metabolic Reactions from Full-Text Articles, BMC bioinformatics, vol. 13, no. 1, p. 172, Jul. 2012.

[4] A.Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic,A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering et al., STRING v9.1: Protein-Protein Interaction Networks, with Increased Coverage and Integration, Nucleic acids research, vol. 41, no. D1, pp. D808D815,Jan. 2013.

[5] S. Pletscher-Frankild, A. Pallej‘a, K. Tsafou, J. X. Binder, and L. J.Jensen, DISEASES: Text Mining and Data Integration of DiseaseGene Associations, Methods, vol. 74, pp. 8389, Mar. 2015.

[6] R. Xu and Q. Wang, A Semi-Supervised Approach to Extract Pharmacogenomics-Specific DrugGene Pairs from Biomedical Literature for Personalized Medicine, Journal of biomedical informatics,vol. 46, no. 4, pp. 585593, Aug. 2013.

[7] Lewis, P., The Characteristic Selection Problem in Recognition Systems, IRE Transactions on Information Theory, vol. 8, no. 2, pp.171178, Feb. 1962.

[8] J. D. Rennie, L. Shih, J. Teevan, D. R. Karger et al., Tackling the Poor as Sumptions of Naive Bayes Text Classifiers, in Proceedings of the 20th International Conference on Machine Learning (ICML-2003), vol. 3. Washington DC, 2003, pp. 616623.

[9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer,SMOTE: Synthetic Minority Over-Sampling Technique, Journal of Artificial Intelligence Research, vol. 16, pp. 321357, Feb. 2002.

[10] G. H. John and P. Langley, Estimating Continuous Distributions in Bayesian Classifiers, in Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence (UAI 1995). Morgan Kaufmann Publishers Inc., 1995, pp. 338345.