

# Ranking of Query Results using Structure Mining Based on Automatic Speech Recognition

Meena H. Gadge<sup>1</sup>, Prof. P. B. Sambhare<sup>2</sup>

*P. G. Student<sup>1</sup>, Assistant Professor<sup>2</sup>*

*<sup>1,2</sup>Department of Computer Science & Engineering, P. R. Pote (Patil) Welfare & Education Trust's college of Engineering & Management, Amravati, India*

**Abstract** — Speech Recognition is the process of converting an acoustic waveform into text containing the similar information conveyed by speaker. A model is learned from a set of audio recordings whose corresponding transcripts are created by taking recordings of speech as audio and their text transcriptions. In case of speech recognition, research followers are mostly using three different approaches namely Acoustic phonetic approach, Pattern recognition approach and Artificial intelligence approach. And for ranking using structure mining a new algorithm Weighted Page Content Rank user can get relevant and important pages easily as it employs web structure mining and web content mining. A webpage ranking analysis can be applied on the scenario where the searching and interaction with the numerous web data is required, so in order to provide effective result. Spoken content retrieval refers to directly indexing and retrieving spoken content based on the audio rather than text descriptions. When the Internet era increases, sharing of resource also increases and this leads to develop an automated technique to rank each web content resource. Different search engine uses different techniques to rank search results for the user query. This leads to business motivation of bringing up their web resource into top ranking position.

**Keywords** — ASR, NLP.

## 1. INTRODUCTION

As automatic speech recognition (ASR) technologies continue to advance, it is reasonable to believe that speech and text offerings will eventually be symmetric, since they are alternative representations of human language, in the spoken and written form, respectively, and the transformation between the two should be direct and straightforward. With this perspective, spoken content retrieval, or indexing and retrieving multimedia content from its spoken part, is an important key to easier browsing and retrieving of multimedia content in the future. In cases where the essence of the multimedia content is captured by its audio, especially for broadcast programs, lecture, meetings, etc. Indexing and retrieving the content based on the spoken part not only eliminates the extra requirements of producing the text description for indexing purposes, but can precisely locate the exact time when the desired information appears in the multimedia. In recent years, spoken content retrieval has achieved significant advances by primarily cascading ASR output with text information retrieval techniques.[13]

The spoken content is first converted into word sequences or lattices via ASR. The cascade approach was subsequently found to work well mainly for relatively high ASR accuracies. Speech recognition technique is the process of mapping an acoustic waveform into a text (or the set of words) which should be equivalent to the information

to be conveyed by the spoken word. Speech recognition in computer system domain is defined as the ability of computer systems to accept input in the form of spoken words(audio format) and map into a text format in order to proceed with further computations. For this purpose, Natural Language Processing (NLP) is a very active area of research and development in Computer Science. Important applications of NLP are machine translation and automatic speech recognition. For NLP, a basic unit of speech recognition is the intermediate form of speech information around which many of the recognition processing is organized for human beings or for machines. ASR for Indian languages is still at its infancy where as western languages like English and Asian languages like Chinese are comparatively well matured. Hence ongoing decade shows growing interest in this field and also huge scope for research work.[5]

Speech recognition involves different functions:

- Speech analysis
- Feature extraction
- Acoustic modeling
- Language and lexical modeling
- Recognition

## 2. EXISTING SYSTEM

Data mining is defined as computer based method of finding and analyzing large amount of data and the process of discovering interesting useful format and its relationships. The key role is to extract the unfamiliar, useful and logical patterns from large database. Data mining concentrate on automatic discovery of patterns, creation of action statements and expected output. The development of internet data in website, make the user difficult to browse webpage effectively.[3]

Every result of the keywords and content words are compared against dictionary by full word matching. If a match is found then a point is awarded to each words based on their position (keyword / content) using weighted technique. Finally all matched keywords and content words are summarized and normalized so that the cumulative total must be less than or equal to one. At last, the normalized value of each result is sorted in descending order to get the most relevant content for the user query. Re-ordered results are sent back to the user so that the top most page is more relevant for the user query.[2]

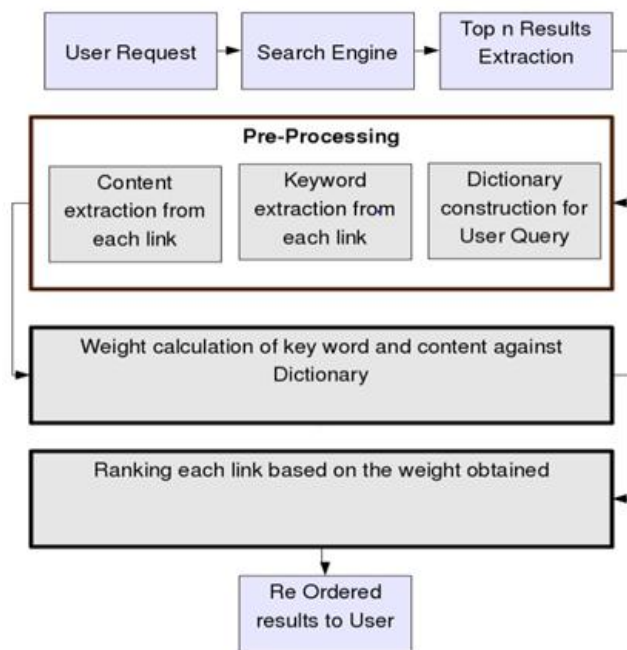


Figure 1: Architecture of Ranking

## 3. PROPOSED SYSTEM

In Proposed system by identifying the pitfall of existing system, we design proposed system, which will take voice as input which is given to ASR and retrieve the text. On the basis of that it will send then to search engine which will retrieve some link. I have applying sorting algorithm based on structure mining and rerank the links which are the outputs of search engine. In structure mining we check for in links and out links in a page. On the

retrieved results it analyzes the popularity of website and rerank according to that in result analysis. The system implements user feedback which tells about the usage of our system in comparison with existing search results.

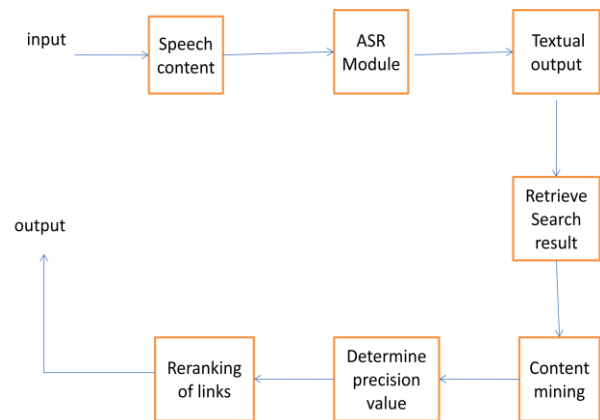


Figure 2: Data Flow Diagram of Proposed system

## 4. IMPLEMENTATION

I have done analysis on well known URLs using both proposed approach and previous approach and at last compare the results. Methodology describes the web results search as user extract the relevant information in the form of text links, from the URL and computes the number of visits using Weighted Page Rank from the URL. At the end we compute various parameters such as Execution Time, Recall, Precision and F-Measure. The whole process of implementation as described in steps:

1. Firstly, in proposed method first it takes voice as an input and converts it into text by using different technique.
2. Secondly, the entered query which is now in textual format is ready to search, and after that the results which extract from google search engine are display that means URLs.
3. Now, in third step the URLs which are given by second step on that apply content mining and after apply ranking on that result of content mining.
4. After that, apply structure mining on the ranking results of content mining and again apply ranking on that result of structure mining.
5. In fifth step we used uses mining on the result of structure mining and apply ranking on that results.
6. In this step all results are displayed which are the output of above all steps.
7. This is the last step in that apply final ranking on the results, in that it will give the results which are relevant to the entered query.

**5. RESULT ANALYSIS**

I have run our proposed approach and base approach on Various URLs. At the end we compute various parameters to conclude our results and results sets are compared using a graph and table that presents the actual research work with in comparison to previous approach. In result analysis we are calculate different types of ratios by using different formulas. I have done analysis on validation measurements as following:

**1. Content Mining:** In this we apply content mining on the search URLs and calculate the ratio for ranking by using following formula as-  
Ratio for content mining=[Term Frequency/(Total no. of results retrieved)]\*100

**2. Structure mining:** In structure mining, it calculates the ratio of Win and Wout that means WPR. By following formulas-

$$WPR = (Win/Wout)$$

**3. Usage Mining:** The usage mining is calculated by using no. of clicks, session time. According to following formula,

$$\text{Ratio of usage mining} = \frac{1}{[(\text{No. of clicks} + \text{session time})/2]} * (1/1000)$$

According to above formula following tables are calculated and on the basis of that the graph will be plotted. In that the results of previous and proposed system will be compared and try to show the difference between both of them as follows,

Table1: Results of content mining.

Rank	URLs	Google rank	Proposed rank
1	https://www.references.com/science/periodic-time-fe48db9a73b7826b	4	1
2	https://en.wikipedia.org/wiki/Frequency	6	2
3	https://answer.yahoo.com/question/index?qid....	7	3
4	www.sengpielaudio.com/calculators-period.htm	1	4
5	Dictionary.reverso.net/English../periodic%20time	8	5
6	https://www.boundless.com/.../periodic.../period-and-frequency-427-5592/	2	6

7	www.keysight.com/upload/cmc-upload/All/Exp88a.pdf	3	7
8	www.electronics-tutorials.ws/waveforms/waveforms.html	5	8

Table2. Results of structure mining.

Rank	URLs	Google rank	Proposed rank
1	https://www.references.com/science/periodic-time-fe48db9a73b7826b	6	1
2	https://en.wikipedia.org/wiki/Frequency	1	2
3	https://answer.yahoo.com/question/index?qid....	3	3
4	www.sengpielaudio.com/calculators-period.htm	7	4
5	Dictionary.reverso.net/English../periodic%20time	5	5
6	https://www.boundless.com/.../periodic.../period-and-frequency-427-5592/	4	6
7	www.keysight.com/upload/cmc-upload/All/Exp88a.pdf	8	7
8	www.electronics-tutorials.ws/waveforms/waveforms.html	2	8

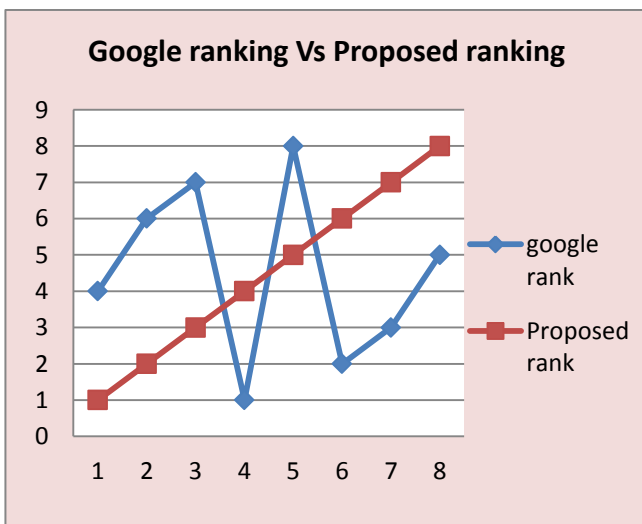
Table3. Results of usage mining.

Rank	URLs	Google rank	Proposed rank
1	https://www.references.com/science/periodic-time-fe48db9a73b7826b	7	1
2	https://en.wikipedia.org/wiki/Frequency	6	2
3	https://answer.yahoo.com/question/index?qid....	8	3
4	www.sengpielaudio.com/calculators-period.htm	3	4
5	Dictionary.reverso.net/English../periodic%20time	1	5
6	https://www.boundless.com/.../periodic.../period-and-frequency-427-5592/	4	6
7	www.keysight.com/upload/cmc-upload/All/Exp88a.pdf	5	7
8	www.electronics-tutorials.ws/waveforms/waveforms.html	2	8

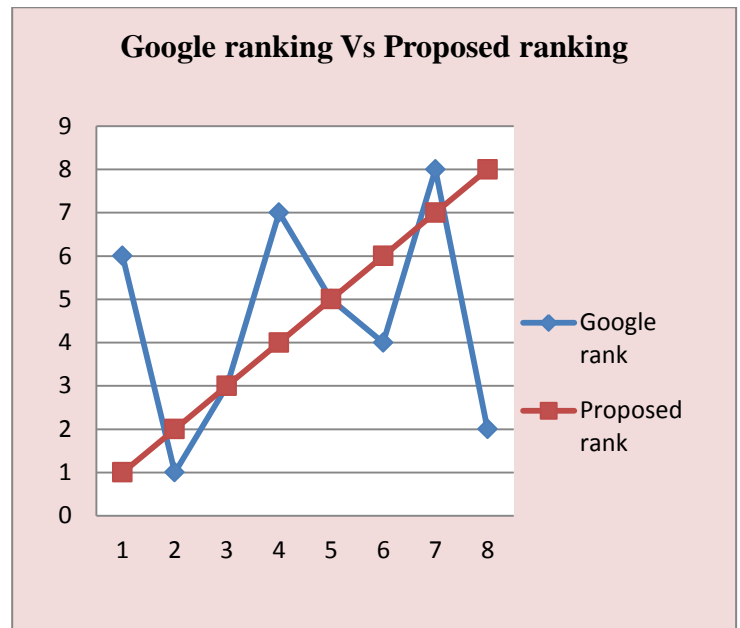
Table 4. Results of final Ranking.

Rank	URLs	Google rank	Proposed rank
1	<a href="https://www.references.com/science/periodic-time-fe48db9a73b7826b">https://www.references.com/science/periodic-time-fe48db9a73b7826b</a>	7	1
2	<a href="https://en.wikipedia.org/wiki/Frequency">https://en.wikipedia.org/wiki/Frequency</a>	6	2
3	<a href="https://answer.yahoo.com/question/index?qid....">https://answer.yahoo.com/question/index?qid....</a>	8	3
4	<a href="http://www.sengpielaudio.com/calculators-period.htm">www.sengpielaudio.com/calculators-period.htm</a>	3	4
5	<a href="http://Dictionary.reverso.net/English..../periodic%20time">Dictionary.reverso.net/English..../periodic%20time</a>	1	5
6	<a href="https://www.boundless.com/.../periodic.../period-and-frequency-427-5592/">https://www.boundless.com/.../periodic.../period-and-frequency-427-5592/</a>	4	6
7	<a href="http://www.keysight.com/upload/cmc-upload/All/Exp88a.pdf">www.keysight.com/upload/cmc-upload/All/Exp88a.pdf</a>	5	7
8	<a href="http://www.electronicstutorials.ws/waveforms/waveforms.html">www.electronicstutorials.ws/waveforms/waveforms.html</a>	2	8

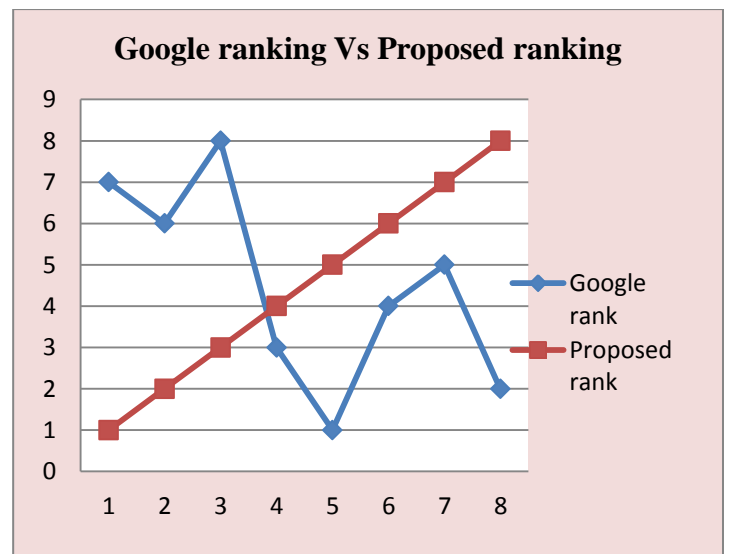
The above table shows the comparison between the results of google and proposed system. Then according to that calculation of page score of the URLs. And final output will given according to the entered query. According to above results shown in tables following graphs are calculated as-



Graph1. comparison of google and content mining results ranking.



Graph2. comparison of google and structure mining results ranking.



Graph3. comparison of google and usage mining results ranking.



Graph4: Comparison of google and final results ranking

Above graphs are the representation of the google search engine and the proposed system, it shows the positive results towards the proposed system which is very helpful in future.

## 6. CONCLUSION

In the propose system, the fundamentals of speech recognition are discussed and its recent progress is investigated. The various approaches available for developing an ASR system are clearly explained with its merits and demerits. The performance of the ASR system based on the adopted feature extraction technique and the speech recognition approach for the particular language is compared. In recent years, the need for speech recognition research based on large vocabulary speaker independent continuous speech has highly increased. I have implementing the google query results based on the WPR. On number of clicks, usage, time spending, in-links, out-links. The ranked results are going to tested with user feedback method, which is better as compared to previous search engine results.

## 7. FUTURE SCOPE

The ranking of web pages based on ASR text retrieval. Which can be implemented best and efficient retrieval result which will decided on the basic of user feedback, which is limited in scope because it work on defined document only and it is less adequate when more challenging real-world scenarios. Encouraged by some

innovative models, developments in ASR appear to be accelerating. The outlook is optimistic that future applications of automatic speech recognition will contribute substantially to the quality of life among deaf children and adults, and others who share their lives, as well as public and private sectors of the business community who will benefit from this technology.

## ACKNOWLEDGEMENT

The author would like to present their sincere gratitude towards the, Prof.P.B.Sambhare (Co-Guide) and Prof.V.B.Bhagat (Guide) also special thanks to Prof. V. B. Gadicha (H.O.D – Computer Engineering) for their extreme support to complete this assignment.

## REFERENCES

- [1] Meena H. Gadge and Prof. P.B.Sambhare."A Survey on Ranking of query results using structure mining based on Automatic Speech Recognition."International Journal of Innovative Science, Engineering & Technology,ISSN (Online) 2348 – 7968, Vol. 3 Issue 11, November 2016, PP 360-365.
- [2] Neelam Tyagi and Simple Sharama, "Comparative study of various page ranking Algorithms in Web Structure Mining." International Journal of Innovative Technology and Exploring Engineering (IJTEE). 1(1),2012.
- [3] Syed thousif hussain B.N.Kanya "Extracting Images From The Web Using Data Mining Technique",International Journal of Advanced Technology &Engineering Research , March 2012.
- [4] Neelam Tyagi and Simple Sharma, "Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Page", International Journal of Computing and Engineering (IJSCE) July 2012.
- [5] Nidhi Desai, Prof. Kinnal Dhameliya, "Feature Extraction and Classification Techniques for Speech Recognition: A Review," International Journal of Emerging Technology and Advanced Engineering (IJETA), Vol. 3, Issue 12, December 2013.
- [6] Seifedine Kadry and Ali Kalakech, "On the Improvement of Weighted Page Content Rank", Journal of Advances in Computer Networks, Vol. 1, No. 2, June 2013
- [7] Rashmi Sharma, Kamaljit Kaur, "Review of Web Structure Mining Techniques using Clustering and Ranking Algorithms" International Journal of Research in and Communication Technology,6 June- 2014.

- [8] M. Larson and G. J. F. Jones, “Spoken content retrieval: A survey of techniques and technologies,” *Found. Trends Inf. Retr*, vol. 5, pp. 235–422, 2012.
- [9] G. Chen, O. Yilmaz, J. Trmal, D. Povey, and S. Khudanpur, “Using proxies for OOV keywords in the keyword search task,” in *Proc. ASRU*, pp. 416–421, 2013.
- [11] Ms. B. Nagarathna, Dr. M. Moorthi,, “A study on web content mining and web structure mining”, *International Journal of Modern Trends in Engineering and Research (IJMTER)*, vol. 02, issue 05, May 2015.
- [10] Karuna Katariya, Rajanikanth Aluvalu, “Agglomerative Clustering in Web Usage Mining: A Survey”, *International Journal of Computer Applications*, vol. 89, no. 08, March 2014.
- [11] Shanthi Therese S, Chelva Lingam, “Review of Feature Extraction Techniques in Automatic Speech Recognition,” *International Journal of Scientific Engineering and Technology (IJSET)*, Vol. 2, Issue 6, pp.479-484, June 2013.
- [13] Sanjivani S. Bhabad, Gajanan K. Kharate, “Overview of Technical Progress in Speech Recognition”, *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, Vol.3, Issue 3, March 2013.
- [14] Amandeep Kaur Mann, and Navneet Kaur, “Review Paper on Clustering Techniques”, *Global Journal of Computer Science and Technology (ISSN (Online): 0975-4172)*, vol. 13, Issue 5, Version 1.0, 2013.