AUDIOBOOK OF PRECISE SUMMARY USING DATA SCIENCE

[1] Abhya Singh
[2] Prof. Manju More
[3] ShahriarSaheen Ahmed Choudhur
[4]AbhishekRjendran,

[2]Assistant Professor,[1][3][4] Student,
School Of Computer Science And Engineering,
Reva University

Abstract—Due to advancement in technologies, such as AI ,ML,deep learning and NLPfields in computer science, the demand for automation has increased drastically.Gradually, AI is making high growthin the field of Audiobook industry. BookSnap, a web-application platform aimsatknowingthe gist of any book in a short span of time in the form of short audiobooks starting with Hindi language. Based on Machine Translation(MT), a given piece of text , is translated from source language to the target language. Text-to-Speech Synthesiser technology is used to convert these short snaps into audiobooks.Having made audiobooks available in Hindi will fill a gap for those who are not very fluent in English or are more comfortable with Hindi language. This paper will not only save people's time but will also help the differently abled people (Blind). Providing the audiobooks of precise summary in Hindi language will also help a vast group of people gain maximum insights of the book. Aim and Goal is to educate more and more citizens and help them make their lives productive.
BookSnap algorithms such MLIR, MLDR, MDC, the Multi-8 two-years-on retrieval challenge, and the Multi-8 results merging task are discussed in this research.

INDEX TERMS:MDC,MLIR,MLDR,differently abled people(Blind),NLP,MT,Text-to-speech,Multi-8 two-years-on retrieval task, Multi-8 results Merging task

I.**1. INTRODUCTION**

Over the last few years, the Audiobook Industry has immensely expanded significantly. Audible and Spotify have changed and are changing millions of lives.With the advancement of technologies people can now access the knowledge of books in the form of audiobooks in hours or minutes on their smart-phone, tablet/ipad, loudspeakers and other electronic devices.The journey of AI is a vast one ,especially when it comes to match the quality of human performance. TTS technologies plays a vital role to convert text into speech and these days due to the advancement in AI field TTS technologies are sounding more natural. This is a long struggle of many leading companies to improve the voices of their AI personal assistants. The features of these systems lacks the nuances of human speech and therefore sounds robotic. Thus, at this point, DL enters the scenario.The audiobook industry is gradually gaining advantages from the audio system technology, submitted by Amazon as a patent. Lyrebird, clones human sounds using AI technology that gives us the feature of having the option to alter the narrator's accent according to the listener's choice.[1]
The Text-to-Speech synthesis are rapidly growing in the AI Industry. Utilizing the right AI voice generator, voice chatbot can avoid inaccuracies and sound more realistic. TTS technology is gaining popularity among the different ventures in the recent years, as it saves a lot of time and cost. [2]
Natural Language Processing (NLP) is one of the disciplines of Artificial Intelligence (AI) technology that is gaining popularity. The design of computational models that process and understand natural language is the focus of this discipline. NKP models effectively teach the computer how to grasp semantic grouping of objects (for example, the terms "cat and dog" and "cat and bat" are semantically quite similar), text to speech, language translation, and so on.It is possible for a computer to understand, interpret, and utilise human languages and vocal speech like English, German, or another "natural language." Natural Language Processing (NLP)." Today, a wide range of NLP applications may be found in use.
This article contains information like history, architecture, working, mathematical representation, about the different types language translation algorithms and models like MLIR,MLDR,document clustering.

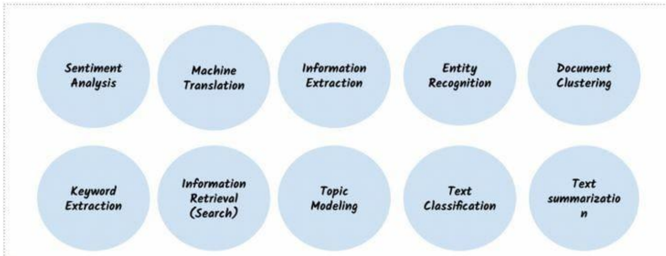

Fig.1 : Different Applications of NLP
Reference :https://in.pinterest.com/pin/536632111861377473/
(Opinosis Analytics)

In the given figure above we can see the variety of domain and applications supported by NLP.Speech recognition, dialogue systems, information retrieval, question answering, and machine translation, for example, have begun to transform the way individuals identify, retrieve, and use information resources.Artificial Intelligence is a type of Natural Language Processing.Statistical learning, in which you train your computer to learn patterns in English, is one amongst the most complex techniques. You could even create your programme once and train it to work in a variety of human languages if you do it this way.
The goal of NLP is to make human languages understandable so that a computer can interpret and comprehend the writings. The manuscript is the language script provided to the programme, and the machine is the programmed mechanism. As a result, the computerised algorithm extracts linguistic data as digital knowledge.Rather than using statistical learning models, the computer converts the language features into a rule-based, statistical technique that may be used to solve specific problems and execute the task of language processing.The components of analysis, transfer, and synthesis were not always clearly separated in many older systems, notably those of the 'direct translation' kind. They also combined data (dictionary and grammar) with processing rules and routines in some cases.New systems have shown varying degrees of modularity, allowing system components, data, and programmes to be altered and updated without compromising overall system efficiency. The reversibility of analysis and synthesis components, which means that the data and transformations employed in the analysis of a particular language are applied in reverse when synthesising texts in that language, is another level in several modern systems.[3]

2. LITERATURE REVIEW:

In the comparative study: Wei Gao, Cheng Niu, Ming Zhou, and Kam-Fai Wong used the Learning-to-Rank (L2R) framework to approach Web MLIR Ranking.Existing methods focus on collecting relevance scores from multiple retrieval settings, rather than learning the ranking function directly. As a result, the authors used the Web MLIR Ranking technique. Joint ranking models can be built by taking advantage of document correlations to estimate the chance of all documents being relevant together. This technique can be used to boost the relevance estimate of papers in a wide range of languages by utilising a relevant document in one language. Mean average precision and other information retrieval assessment metrics are directly enhanced by training model parameters to detect relevant documents more correctly. The study seeks to find a way to integrate the incomparable scores associated with each group of outcomes.To obtain the final ranking score, the scores are normalised using methods such as Min-Max, Z-score, CORI, and so on, and then merged using CombSUM or logistic regression. When it comes to MLIR relevance, they still focus on changing the scores of documents from various monolingual result lists, rather than directly modelling many aspects.The authors concluded that the new models, which are based on a generic ranking mechanism, first identify important subjects among recovered documents and then cooperatively identify relevant documents and topics based on content similarities. As a result, there is a huge improvement in ranking.[4]
Santosh GSK, Kiran Kumar N and VasudevaVerma have mentioned about the MLDR approach. By using Bilingual dictionary as their primary language resource, they were able to extract a variety of monolingual and multilingual similarity traits. They used the FIRE (Forum of Information Retrieval Evaluation) to do tests on different ranking algorithms and compare the results. The results suggest that the elements addressed in improving Multilingual Document Ranking are successful (MLDR).To summarise, the authors improved the MLDR performance from a QA standpoint, which outperformed the BM25 baseline by a significant margin. They're currently using the FIRE – 2010 datasets to extend the method to different Indian languages.[5]
Luo Si and Jamie Callan introduced 2 tasks namely :There have been two instances of this merging: the retrieval of Multi-8 data two years after it was first collected and the combining of Multi-8 findings. The Multi-8 two-years-on retrieval and the Multi-8 results merging tasks are two of the CLEF 2005 assessment tasks described here. The major goal of the Multi-8 job is to produce and aggregate multilingual search results based on basic bilingual or monolingual ranked lists.. Their efforts are mostly focused on multi-8 merging. The Multi-8 results merging job tries to create a multilingual ranked list by combining two lists of eight bilingual (or monolingual for English) ranks.Indexing and translation of the highest-rated texts in each ranking list is used to get scores that are comparable

across lists. As a result of this research, specialised and language-specific logistic models have been developed for these articles. All articles in ranked lists in different languages were sorted according to these logistic models, which were designed to estimate similar document scores for all publications. Research using the recommended methodology surpass previous studies and only require as little as 10 documents each pair (e.g., 10 per pair) to deliver correct conclusions.[9]

In the paper : Multilingual Wordnet sense Ranking using nearest context, authors have introduced OMW(Open Multilingual Wordnet) that has over 150 different languages with word-nets built automatically. Multiword expressions from wordnets are used to train the pre-trained models Word2Vec and Polygot2. This allows the multiword expressions to be rated as well. As a consequence, this model has been trained to produce embeddings for single and multi words. Five languages are represented in the lexicon's WSD. Semcor sense corpora in five languages are used to test the results using the Word2Vec and Glove model. Compared to Word2Vec, the Glove model has an average accuracy of 0.47 for languages including English, Italian, Indonesian, Chinese, and Japanese. Ranking correlation is mostly dependent on human ranks, as according studies using OMW sense ranking A distributional semantics method to Wordnet Sense Ranking may be helpful.[8]

On similar corpora, Kiran Kumar N, Santosh GSK, and VasudevaVerma discussed Multilingual Document Clustering (MDC). Wikipedia is a great example of a multilingual information repository, with 257 language versions now available. The authors of this research have conducted a thorough examination of several methods for maximising MDC's effectiveness by utilising its vast multilingual knowledge base.Bilingual dictionaries are used in the paper's proposed effort to translate Japanese and Russian papers into English (anchor language). Because of the time commitment, it is not always preferable to translate a complete document into an anchor language.[7]

3. MLIR (Multilingual Information

Retrieval)

[4]Searching for documents in many languages is possible using MLIR. A bilingual dictionary, machine translation software, or a parallel corpus can be used to translate requests prior to their monolingual reply.In order to properly combine numerous ranked lists from various languages, re-ranking is then used. The necessity to compare and integrate material in several languages makes it challenging to do multilingual information retrieval (MLIR) for websites. As a result of the information that is lost when queries are interpreted, it is difficult to determine cross-lingual relevance.[4]

i.Learning for MLIR Ranking

The aim of the MLIR ranking learning system is to develop a unique ranking function that can estimate the same scores for articles published in numerous languages with different accents and terminology. It is essential to build a multilingual feature area for the materials. For MLIR ranking, current monolingual L2R algorithms can be employed with these characteristics. In this case, we assume that each query q ∈ Q (Q is the provided query set) is connected with the list of retrieved documents Dq = {di} where di is the rank label of di and may take one of m ranks in the R set R = "r1, r2,...,rm" (r1 r2 ... rm, where denotes the order relation). As a result, the training corpus may be described as {q ∈Q|Dq, Lq}. we designate each query-document pair by denoting Φ [fk(q, di)]: " f(q, di)" If we choose K k=1, then fk is one of the relevant feature functions for our analysis (q, di). The ranking function F: Φ → will be used to assign a relevance score to each recovered document's feature vector (which represents the real value space). When Dq is ordered according to F, the order of the documents is indicated by an integer permutation (q,Dq,F), and the position of di in the results list is reflected by an integer permutation (di). As a consequence, the ranking aim refers to the search for an optimum function: argminF = F For all queries, q E(π(q,Dq, F), Lq) minimises an error function E representing the discrepancy between π(q,Dq, F) and the appealing rank order supplied by Lq. Different algorithms for ranking employ the ranking function and the error function in different ways. Probabilistic classification (e.g., Support Vector Classifier) and metric regression may be used to rate documents (e.g. Support Vector Regression). SVM (large-margin ordinal regression), RankBoost, RankNet, and other ranking algorithms all try to maximise the pair-wise loss and estimate the order of relevance between two articles based on order preference. [5–6].SVM-MAP [20] has recently been presented to directly improve the IR assessment metric – Mean Average Precision – directly (MAP).[4]

ii.Joint Ranking Models for MLIR

This task might be made more difficult by the query translation problem when using monolingual ranking algorithms. A collaborative ranking approach that uses the relationship between papers/documents authored in multiple languages is offered in addition to query-document relevancy. To estimate the joint relevance probability distribution, the Boltzmann machine (BM) simulates any relationship between items.

a) Boltzmann Machine (BM) Learning

Nodes in a BM's state vector can travel in any direction, and as a result, it can make random predictions about the values they will take as they do so. s = [s1s2...sn], where si = ±1 is the state of node I and n is the total number of network nodes. In layman's terms, it is described as follows: Wij is the edge weight of node I in relation to node j and sisj is si's threshold in E(s) = 1 2 ijwijsisj = 1 θisi..The probability of finding the network in a global state after a significant period of time in the dynamics process is dependent only on the states of the nodes and their neighbours, and follows the Boltzmann distribution, which is P(s) = 1 Z exp(-E(s)), where Z is the normalisation function for all possible states.As a result of machine training, the Boltzmann distribution will begin to look more and more like P˜(s).Kullback-Leibler measures the difference between the two distributionsK(P˜||P) = s P˜(s) log P˜(s) P (s) .Using gradient descent, the divergence should be reduced. This sort of weight-updating rule may be obtained.:

Δwij = α(<sisj>clamped − <sisj>f ree) (1)

Δθi = α(<si>clamped − <si>f ree) (2)

b) Joint Relevance Estimation Based on BM

For each q, we denote $\mathbf{sd}_q = [sd_i]$ and $\mathbf{st}_q = [st_j]$ as the state vectors of the document and topic nodes respectively, then the energy of the machine becomes:

$$E(\mathbf{s}, q) = E(\mathbf{sd}_q, \mathbf{st}_q, q) = -\sum_i \Theta \cdot \mathbf{f}(q, d_i) sd_i - \frac{1}{2} \sum_{i,j} \mathcal{W} \cdot \mathbf{g}(d_i, t_j) sd_i st_j \quad (3)$$

where $\mathbf{f} = [f_x(q, d_i)]_{x=1}^X$ and $\mathbf{g} = [g_y(d_i, t_j)]_{y=1}^Y$ are the X-dimension feature vector of query-document relevancy on document nodes and the Y-dimension document-topic relevancy on edges respectively, and Θ and $\mathcal{W}$ are their corresponding weight vectors. Then the probability of the global state $P(\mathbf{s}, q) =$

$P(\mathbf{sd}_q, \mathbf{st}_q, q)$

c) Multilingual Clustering for Identifying Salient Topics

The measure of cross-lingual document similarity is widely used because of its simplicity and efficiency. A cosine-like function with an expansion of TF-IDF weights is utilised for cross-lingual keyword translation. The following is an explanation of the metric at issue:

$$sim(d_1, d_2) = \frac{\sum_{(t_1,t_2) \in T(d_1,d_2)} tf(t_1, d_1) idf(t_1, t_2) tf(t_2, d_2) idf(t_1, t_2)}{\sqrt{Z'}} \quad (4)$$

where Z' is given as

$$Z' = \left[\sum_{(t_1,t_2) \in T(d_1,d_2)} (tf(t_1, d_1) idf(t_1, t_2))^2 + \sum_{t_1 \in \overline{T}(d_1,d_2)} (tf(t_1, d_1) idf(t_1))^2 \right] \times \left[\sum_{(t_1,t_2) \in T(d_1,d_2)} (tf(t_2, d_2) idf(t_1, t_2))^2 + \sum_{t_2 \in \overline{T}(d_2,d_1)} (tf(t_2, d_2) idf(t_2))^2 \right]$$

$T(d_1, d_2)$ denotes the sets of word pairs where t_2 is the translation of t_1, and t_1 (t_2) occurs in document d_1 (d_2). $\overline{T}(d_1, d_2)$ denotes the set of terms in d_1 that have no translation in d_2 ($\overline{T}(d_1, d_2)$ is defined similarly). $idf(t_1, t_2)$ is defined as the extension of the standard IDF for a translation pair (t_1, t_2): $idf(t_1, t_2) = \log\left(\frac{n}{df(t_1)+df(t_2)}\right)$, where n denotes the total number of documents in two languages and df is the word's document frequency.

d) BM trainer as a classifier

Gripper and free phases are alternated in order to avoid maxima, which must be done multiple times with different start masses. State values of hidden units are determined during the clamping phase in contrast to output units whose status is established by human labels. model is used for both stages.

The training is to adjust the weights and thresholds in such a way that for each query the predicted probability of document relevancy, i.e., $P(\mathbf{sd}_q, q) = \sum_{\mathbf{st}_q} P(\mathbf{sd}_q, \mathbf{st}_q, q)$, approximates to the target distribution $\tilde{P}(\mathbf{sd}_q, q)$ as closely as possible, where $\tilde{P}(\mathbf{sd}_q, q) = \begin{cases} 1, & \text{if } \mathbf{sd}_q = L_q; \\ 0, & \text{otherwise} \end{cases}$ is obtained from the training data. By minimizing the *K-L Divergence*, we obtain the updating rules

$$\Delta\theta_x = \alpha \sum_{q,i} f_x(q, d_i) (< sd_i >_{clamped} - < sd_i >_{free}) \quad (5)$$

$$\Delta w_y = \alpha \sum_{q,i,j} g_y(d_i, t_j) (< sd_i st_j >_{clamped} - < sd_i st_j >_{free}) \quad (6)$$

e) BM interface for MLIR Ranking

A node's state distribution in the mean field approximation is exclusively dependent on the states of its adjacent nodes, each of which is set to its average state value. As a result of this, here's what we already have::

$$P(sd_i = r) = \frac{\exp\left[\sum_j \mathcal{W} \cdot \mathbf{g}(d_i, t_j) < st_j > r + \Theta \cdot \mathbf{f}(q, d_i) r\right]}{\sum_r \exp\left[\sum_j \mathcal{W} \cdot \mathbf{g}(d_i, t_j) < st_j > r + \Theta \cdot \mathbf{f}(q, d_i) r\right]} \quad (7)$$

$$P(st_j = r) = \frac{\exp\left[\sum_i \mathcal{W} \cdot \mathbf{g}(d_i, t_j) r < sd_i >\right]}{\sum_r \exp\left[\sum_i \mathcal{W} \cdot \mathbf{g}(d_i, t_j) r < sd_i >\right]} \quad (8)$$

$$< sd_i > = \sum P(sd_i = r) r \quad (9) \qquad < st_j > = \sum P(st_j = r) r \quad (10)$$

The relevance probability of a document is calculated using the average rank labels for all topics in Equation (7). Eq. (8), on the other hand, uses the average rank labels of all articles to estimate the topic's relevance probability. Eqs. (9) and (10) use the probability distributions produced in Eqs. (7) and (9) to estimate the average rank labels (8).
There is a fixed-point solution method for the iterative method of solving the mean field equations (7)–(10).
1. Assume that each node has an average state value.
2. Eqs (7) and (8) could be used to estimate the probability of each node's state value based on the average values of its neighbours.
3. Update each node's average state values using Eqs. (9 and 10).
4. Step 2 must be repeated until the average values of the states are in agreement..

f) BM training with MAP Optimization

The MAP represents the average of all the queries' average precision.Rather than simply maximising MAP, we aim to achieve the following:

$$MAP - C \sum_y ||w_y||^2 - C \sum_x ||\theta_x||^2 \quad (11)$$

The L-2 regularisation terms in the model's last two terms describe its level of complexity. C. handles the trade-off between model correctness and complexity. L-2 norm and MAP loss hinge relaxation were minimised by using the same strategy as in the previous paper. Because MAP is not a continuous function, the Powell's Direction Set Method is used instead, which eliminates the need for any derivation calculations.To acquire the best results using Powell's method, the BM's weights are altered several times.In order to improve classification accuracy, we train the BM with a certain set of starting weights (d). In model inference, the mean field approximation (e) is also used.

iii) Results

The suggested MLIR ranking algorithms were tested in the field. Both Chinese and English multilingual Web search data and TREC5&6 English-Chinese CLIR data were used in the research. The ScoreComb ranking score combination method serves as a basis. When learning how to rank Chinese and English texts separately, several methods such as Ranking SVM and SVM-MAP are used. It is then used to aggregate the scores using a log regression model.
In order to examine the MLIR ranking's performance, three common L2R methods were utilised: SVM classifiers include SVC (SVM classifier with probability estimate), RSVM (Ranking SVM), and SVM-MAP.There are three basic kinds of ranking strategies for these algorithms: First, there's the widely used SVC algorithm; second, there's the cutting-edge RSVM technique, which uses paired-wise preference order classification; and third, there's the SVM-MAP algorithm, which ranks by optimising the IR relevance directly.
Comparisons are made between the BM classifier (BM) and the BMC-MAP classifier (BM classifier with MAP optimizer).We removed the hidden units and edges from the BMC and BMCMAP models and used the resulting log linear models to directly examine the role that relevance plays in the data. As a consequence, LOG and LOG-MAP are the two extra systems to be evaluated.

a) Experiments on TREC CLIR data

Cross-lingual document similarity is a focus of our work at CLIR. The CLIR job for TREC5 and TREC6 is defined as the retrieval of Chinese materials using English inquiry.

Combining translations from three free machine translation engines, Okapi-BM25 (BM25) retrieves Chinese content from English queries. There are 25 regularly used query-document relevance models implemented to train the ranking models using translations of the queries and scores from TFIDF and BM25 and language modelling IR, etc. The original query is used to acquire English documents from TIPSTER, and the BM25 scores are used to create BM for the joint relevance rating. Twenty documents are chosen and given one of two labels due to the lack of a relevancy annotation in English: 0 for the last 10 documents in the result; 1 for the first 10 documents.

recall	BM25	SVC	RSVM	SVM-MAP	LOG	BMC	LOG-MAP	BMC-MAP
0	0.658	0.736	0.788	0.798	0.715	0.796	0.797	0.815
0.1	0.495	0.476	0.531	0.598	0.475	0.583	0.592	0.591
0.2	0.411	0.393	0.427	0.486	0.391	0.469	0.480	0.502
0.3	0.345	0.354	0.385	0.414	0.349	0.412	0.411	0.423
0.4	0.289	0.324	0.346	0.368	0.324	0.367	0.366	0.376
0.5	0.251	0.282	0.299	0.316	0.281	0.312	0.315	0.323
0.6	0.203	0.222	0.241	0.245	0.214	0.247	0.241	0.269
0.7	0.164	0.174	0.200	0.185	0.175	0.183	0.182	0.220
0.8	0.074	0.099	0.101	0.086	0.099	0.088	0.084	0.107
0.9	0.010	0.020	0.027	0.016	0.018	0.017	0.016	0.030
1.0	0.002	0.007	0.012	0.006	0.004	0.007	0.006	0.008
AP	0.249	0.253	0.280	0.301	0.250	0.299	0.299	0.314

Table1 : TREC6 CLIR performance by 11-point precision-recall and AP measure

As indicated in Table 1, the CLIR results are presented using AP and an 11-point precision-recall measurement. BM25 is used to evaluate the translated query's quality to Chinese material because there is no multilingual result merging.

This was confirmed by t-testing, which showed that BMC beat the LOG ($p = 0.009$) and the RSVM ($p = 0.011$). CLIR performance can be improved by using insights from monolingual IR studies. SVM-MAP and LOG-MAP to BMC-MAP AP improvements aren't as significant as the leap from LOG to BMC.The optimization of Eq(11) may have resulted in less benefit than initially expected. BMC-MAP training, unlike SVM-MAP training, does not provide a global optimum. BMC-MAP surpasses SVM-MAP by 4.15 percent despite the fact that it has received less learning.

b) MLIR experiments on web-search data

Search results that are available in a variety of languages. Our Web search data comes from commercial search engine query records. Separate records are kept for queries in English and Chinese.. Retrieved web sites are rated on a scale of 0 to 5, with 0 being the most irrelevant and 5 being the most relevant (excellent). In order to extract query-dependent properties, we use the query and four separate sources for each web page of a given query: the anchor text, URL, document title, and content. PageRank, for example, is a query-independent property that is also extracted. Each of the two languages has a total of 352 of these characteristics.

The Chinese log includes both the original English log and translations thereof. A multilingual ranking corpus is built using the results of these searches and the labelled results The appropriate Chinese and English websites are put together for an English-language search. There are 32,049 pages of English and 17,791 pages of Chinese in the entire collection.

The MAP, precision@1, 5,10, and NDCG@1,5,10 (NDCG—Normalized Discounted Cumulative Gain) were all successful, as illustrated in Figure 1. In general, multilingual feature space models outperform the simple ScoreComb model in terms of results. All improvements are statistically significant, as shown by the t-test (p 0.05). In this case, L2R approaches show their value in training appointed head straight from features..

RSVM is considered to perform better than SVC since it optimises the ranking order of document pairings. The MLIR results confirms on this. BMC gets comparable results with RSVM, as does TREC, showing that classification-based ranking algorithms can perform as well as state-of-the-art ranking models by utilising the relevancy among individual documents. Surprisingly, SVM-MAP performs worse than RSVM. That the RSVM can use fine-grained 6-level relevance whereas the SVM cannot is one plausible cause.

The BMC-MAP model is the best. To put it more simply, it outperforms all of the other models by at least 30.22 % ($p = 0.003$) and by at least 15.12 % ($p = 0.006$), the SVC, the BMC, the RSVM, and the SVM-MAP by at least 7.40 % ($p = 0.009$) in terms of the measure of mean overall performance.

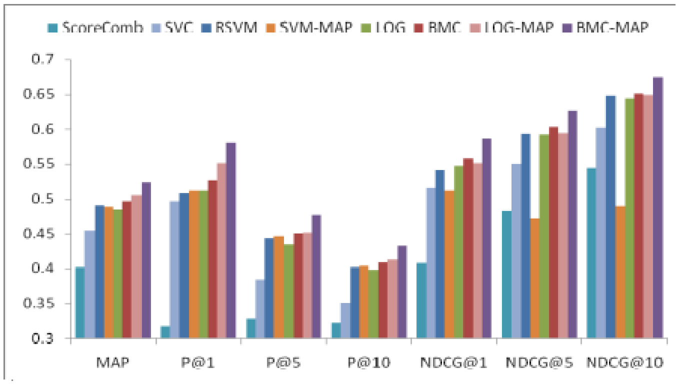

Fig.2: Comparison of ranking results using Multilingual web search data

	MAP	P@1	P@5	P@10	NDCG@1	NDCG@5	NDCG@10
LOG	0.484	0.511	0.435	0.397	0.546	0.591	0.641
BMC	0.497	0.527	0.451	0.409	0.557	0.604	0.651
LOG-MAP	0.504	0.552	0.452	0.413	0.551	0.594	0.649
BMC-MAP	0.523	0.580	0.478	0.432	0.587	0.626	0.674

Table 2 : The comparison results of using and without using clusters in BM models

Compare BMC with LOG, as shown in Table 2, to see how well the combined ranking model performs (LOG-MAP). When comparing the BMC-MAP difference to the BMC-LOG difference, the impact of inter-document relevance can be readily seen.

IV) MLDR(Multilingual Document Ranking)

By extracting simple and efficient characteristics from multilingual texts and themes, the performance of Multilingual Document Ranking (MLDR) is enhanced. In order to enhance MLDR, it is possible to use commonalities between candidate documents for comparison. Similar papers often score similarly, making crosslingual relevant documents a useful tool for estimating crosslingual relevance. Various similarity metrics between documents in the same language and documents in other languages may be obtained from result lists in two separate languages and their queries.. Using the same set of similarity measures, both monolingual and multilingual texts may be assessed and compared to one other.
Only bilingual dictionaries are utilised to calculate the multilingual document similarity. External knowledge sources, such as Wikipedia, are also used to improve the effectiveness of similarity measurement. If the fundamental language resource (bilingual dictionary) is available, this technique may be expanded to additional language pairs.
The test was performed on the FIRE 2010 corpus. Experiments are carried out using the derived features to model multiple ranking systems.The NDCG is used as the evaluation metric to compare their results. There may be a considerable real improvement compared to utilising the BM25 baseline ranking method.
To meet the cross-lingual document retrieval challenge, the FIRE 2010 dataset includes newspaper articles from regional news sources in the supported languages: Bengali, English, Hindi, and Marathi..In each of these languages, there are 50 query topics to choose from. For the experiments, they considered English and Hindi articles. For each topic-document pair, only binary relevance judgments are supplied. All of the relevant papers for each subject were collected into a single group of files that included some irrelevant information. The quantity of noise examined is twice as much as the number of documents that are relevant.
SVC (SVM Classification), RSVM (Ranking SVM), SVM Regression, and Logistic Regression are used to learn ranking functions by modelling the characteristics that are obtained.Logistic regression is carried out using the source codes of LibSVC2, SVMLight3, SVM-Rank4, and Logistic Regression5.. These learning algorithms' predicted probabilities are utilised to rank the documents.Human labellers use a scale of 0 (irrelevant) to 5 (very relevant) to verify the publications' position in the rankings (excellent).It is necessary to use the results of the NDCG@5,10,15,20 test to compare the two systems..

Method	NDCG@5	NDCG@10	NDCG@15	NDCG@20
SVC	67.99	73.53	76.28	79.50
SVM-Reg	71.00	77.87	77.94	**82.11**
RSVM	**75.04**	**78.84**	**78.03**	79.83
LogReg	69.89	74.53	76.69	80.37
BM25	64.19	69.38	68.82	72.21

Table 3 : Comparison of MLDR performances

A glance at Table 1 reveals that the BM25 system was surpassed by all the learning algorithms. In terms of performance, our ranking functions are on par with the baseline system. MLDR's performance has greatly improved.The best results were produced via ranking SVM and SVM regression. These precisions indicate that the characteristics under consideration have shown to be beneficial in improving MLDR performance.

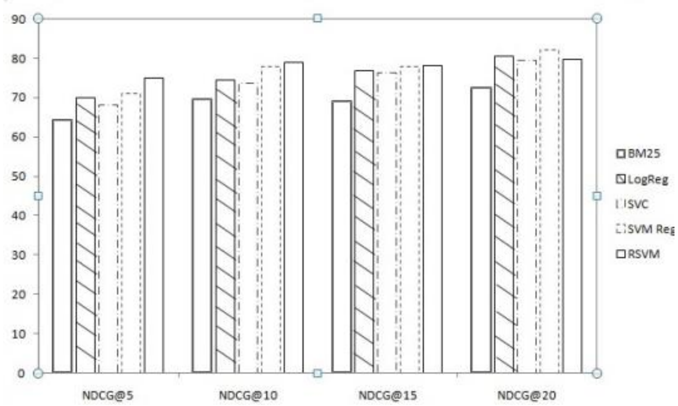

Fig.3: Graphical comparison of performances of ranking algo.

IV)MDC (Multilingual Document Clustering)

Development of applications is necessary since more and more papers are being authored in a variety of different languages. The processing and administration of multilingual online content using MDC has shown to be quite effective. As part of MDC, n documents written in several languages are divided into many clusters, with the semantically related papers belonging to each cluster. CLIR, parameter training for statistical machine translation, and aligning parallel and non-parallel corpora are just a few of the many uses for this method.

"Bag of words" clustering ignores the semantic information included in each document when using typical text clustering algorithms. Two texts that use distinct sets of keywords to describe the same subject might be incorrectly categorised as independent works. There is a lack of shared concepts, even though the terms they use are likely synonyms or semantically linked in other ways.. Adding an ontology or external knowledge to the document representation is the most common method of overcoming this problem.

Wikipedia is one such multilingual knowledge base, with 257 language editions currently active.

i) Proposed Approach

Document vectors are used to represent the collection of English and Hindi text documents at first. To obtain new vectors, these document vectors are supplemented with Wikipedia knowledge base. For measuring document similarity, the basic document vectors and enriched document vectors are linearly merged. The similarity metric is used to create clusters for English and Hindi documents. To integrate these clusters, the centroids' similarity is measured.

a) Document representation

Both English and Hindi text documents are represented using the vector space paradigm. A "bag of words" refers to a document's keyword-based properties when no sorting information is provided..The vector's values are TFIDF scores. A stopword is one that occurs in at least 50% of the texts in a given language, as opposed to maintaining a list of such terms for each language. Despite the elimination of stopwords, the document vectors still include a small number of erroneous words. Based on their TFIDF scores, only the top-k terms in each document were examined. They experimented with k values ranging from 40% to 100% with a ten percent increment. For k=50 %, the best cluster results are obtained.

b) Document clustering

Automatic document clustering can result in highly similar documents in one group, but highly different documents in different groups.In order to categorise the enriched texts, many clustering approaches (such as Hierarchical clustering, Sectioned clustering, and so on) can be utilised. In order to combine the advantages of partitional and hierarchical clustering techniques, researchers used the Bisecting k-means strategy, which divides the largest cluster into many sub-clusters. The bisecting k-means approach's final clustering result was accomplished by selecting random fifteen k values ranging from 30 to 70 on average.

Steinbach et al. evaluated various algorithms and came to the conclusion that bisecting k-means outperforms normal k-means and agglomerative hierarchical clustering.The basic vector space model is used in the basic k-means algorithm, which is a partitional clustering algorithm. The bisecting k-means method is applied to the improved document vectors and the basic keyword vector to construct different clusters for English and Hindi texts. A basic Keyword vector, a Category vector, an Outlink vector, and an Infobox vector are all included in each page.
In order to determine if two papers di and dj are similar, the following criteria must be met:

$$sim(d_i, d_j) = sim^{basic\_keyword} + \alpha * sim^{Category} + \beta * sim^{Outlink} + \gamma * sim^{Infobox}$$

The cosine similarity of the documents di,dj is given by sim(di,dj). The sim is calculated as follows:

$$sim = cos(v_i, v_j) = (v_i.v_j)/(|v_i| * |v_j|)$$

The main keyword, category, outlink, and infobox vectors from the documents di and dj are represented by these vectors.To measure the degree to which two texts are conceptually similar, we use the coefficientsα, β and γ to reflect the importance of these vectors. Hu et al. suggested a new measure of similarity in which Wikipedia Concepts and Categories were utilised to group the monolingual publications in the equation of text similarity. For the classification and grouping of monolingual texts, Hu et al., Wang, and Domeniconi all used Wikipedia Category information.

ii) Results

A total of 1563 papers, 650 of which are in English and 913 of which are in Hindi, have been compiled from 50 different themes. F-score and Purity measures are used to gauge cluster quality. An F-score is computed based on the accuracy and recall of the data. Accuracy in an assignment can be gauged by how many correct assignments there are divided by the total number of assigned documents.
Wikipedia data: Data dumps for several languages are published on a regular basis by Wikipedia. The most recent dump, which included 2 million English and 50,000 Hindi documents, was utilised. The information was in XML format. Retrieve and analyse Wikipedia data to create a vector such as Categories outlinks, Infoboxes, and Re - direct.

Notation	**F-Score**	**Purity**
Keyword (baseline)	0.532	0.657
Keyword_Category	0.563	0.672
Keyword_Outlinks	**0.572**	0.679
Keyword_Infobox	0.544	0.661
Category_Outlinks	0.351	0.434
Category_Infobox	0.243	0.380
Outlinks_Infobox	0.248	0.405
Keyword_Category_Outlinks	0.567	**0.683**
Keyword_Outlinks_Infobox	0.570	0.678
Keyword_Category_Infobox	0.551	0.665
Category_Outlinks_Infobox	0.312	0.443
Keyword_Category_Outlinks_Infobox	0.569	0.682

Table 4 : Clustering schemes based on different combinations of vectors

Our tests using an external knowledge resource performed better than the baseline, as shown in Table 1.Categories and Infobox information have also slipped behind Outlinks information in the popularity ratings. It is vital to include outlinks in any Wikipedia article containing references to other articles (hyperlinks). Papers are reviewed at a more abstract level when looking at the Categories, which might have resulted in lower findings when compared with Outlinks. Important statistical data may be found in the Infobox of a Wikipedia page.However, the information in all of its articles is inconsistent, which contributes to its low performance when compared to others.

IV. ALGORITHMS STUDIES FOR CLEF 2005 EVALUATION TASKS

a) Multi-8 two-years-on retrieval task

Phase one of the experiment involved using a technique known as "Multi-8 two years on" to look for content in eight different languages using only one language (in this case, English queries). Multilingual search results can be fine-tuned prior to incorporating them (or monolingual results for documents written in the same language as queries).Bilingual retrieval may be performed in several instances by fine-tuning the query translation algorithms and then creating an accurate bilingual run for each bilingual run, as shown by previous research.. In the end, a multilingual ranked list is created by combining the results of the various ways. Multi-8's two-year-old retrieval challenge aims to produce and integrate multilingual retrieval results based on fundamental bilingual (or monolingual) ranked lists. A number of multilingual retrieval results can be obtained by combining bilingual (or monolingual) retrieval results using the same retrieval algorithms and then merging the multilingual retrieval results.Research shows that merging multilingual results improves accuracy significantly over single multilingual ranked lists, and this may be done in a variety of ways.

b) Multi-8 results merging task

Using the Multi-8 results merging job, an all-language rating can be generated by combining two lists of eight bilingual (or monolingual) rankings. A federated search task results merging technique has been developed by the authors to address this issue. Indexing and translating of the highest-rated texts in each ranking list is used to get scores that are comparable across lists. Logistic models for both language- and query-specific ranking lists are built using the scores of these documents.After creating and running these models, the documents were sorted using the projected similar document scores from all of the ranking lists, in all languages. A limited amount of documents (e.g., 10 per question, language >pair) has been proved to provide accurate replies using the current proposed methods.

i) MULTILINGUAL RETRIEVAL SYSTEM

Using the findings of numerous multilingual retrieval techniques, this work delivers correct results for multilingual search. ' Authors examine both query-based and document-based techniques of retrieval in the search engine landscape. Preparing text for many languages is the first step in this section. Before providing methods for integrating results from various multilingual retrieval systems, the research delves deeply into query and document translation-based multilingual retrieval algorithmic characteristics.

a) Text Preprocessing

Stopword Lists: In order to make text searchable, stopwords must first be removed.English papers are searched using the Inquerystopword list. Snowball2 is used as a stopword list for Dutch and is utilised for the rest of the European languages save Finnish and French..
After stopwords have been eliminated, different stemming algorithms are used to stem other content words. For English words, Porter stemmer is used.
Decompounding: Languages like Dutch, Finnish, German, and Swedish have a lot of compound words. All words with a length of more than three in the CLEF corpus are considered possible base words. We only evaluate base words with greater collection frequencies than the word in question to avoid overly forceful decompounding. Dutch takes into consideration connecting elements such as –s-, –e-, and –en-, whereas no linking elements are taken into account in Finnish, and elements such as –s-, –n-, –e-, and –en- are taken into account in German. Swedish also takes into account connecting elements such as –s-, –e-, and –o-. Decompounding techniques used in previous investigations have been the same.
Multilingual information retrieval systems may benefit from the usage of online machine translation systems to translate queries and material..

b) Multilingual Retrieval via Query Translation

A multilingual search can be performed by translating English inquiries into other languages, conducting a search in that language and combining the results from several languages into one multilingual list. employing parallel corpus-based translation matrices, English query phrases are translated into various languages.. In the translation matrix of other languages, each word in the English language is translated into the top three alternatives. All three translated terms of an English phrase have normalised weights according to translation matrices (i.e., the sum of the weights is 1). Because the vocabulary of the parallel corpus is so tightly concentrated, certain English terms may not have translations in the parallel corpus at all. The online machine translation programmeSystran 4 provides word-by-word translation results as a complement. Searches in the indexes of each language are conducted using the translated queries. In the Okapi algorithm, each query phrase is weighted according to its translation representation weight..

c) Multilingual Retrieval via Document Translation

Multilingual retrieval may also be accomplished by translating all non-English materials to English and then using the same original English queries. The semantic meaning of longer texts may be better represented than the semantic meaning of short inquiries, hence this retrieval approach may have an advantage over query translation-based retrieval methods. In addition, previous study has shown that translating a phrase from another language into English and translating a word from English into this language may be effective in tandem. German words or phrases may not always have the same sense in English, but they may be translated appropriately into English.

Translators use parallel corpora to generate translation matrices to aid in the translation of documents. Each term in a language other than English is compared to the three most accurate English translations of same word in English. Untranslated words have five word spaces allocated to each of the three contenders, based on their normalised translation chances. A single database indexes and stores all translated and original English content.

d) Combine Multilingual Ranked Lists

According to the theory that various multilingual retrieval techniques prefer to return relevant publications while different case algorithms prefer to retrieve nonsensical articles, ranking lists are combined. The information retrieval field has seen the adoption of similar principles in Metasearch. The result is a simple combination strategy that favours documents recovered using several retrieval techniques as well as documents with a high rating retrieved using a single retrieval method. Assume $dr_{sk_mj}$is the resource-specific raw document score for the jth document in the mth ranked list, and that the dr_{sk_max} and minimum document scores in this list are both $dr_{sk\_m\_max}$. Following this procedure, the j^{th}document's normalised score is calculated.:

$$d_{s_{k\_mj}} = \frac{(d_{rs_{k\_mj}} - d_{rs_{k\_m\_\min}})}{(d_{rs_{k\_m\_\max}} - d_{rs_{k\_m\_\min}})}$$

4. RESULTS AND IMPLEMENTATIONS

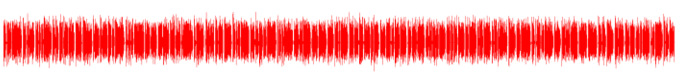

Fig.4 : Waveform of Audiobook for “5AM Club” book Summary in Hindi Language

“The 5AM Club” book written by Robin Sharma was taken as a reference for testing the results. Using googletrans module in python, translating the summarised text of the book in Hindi was successfully achieved. After successfully translating the summarized text from source language (English) to its destination language (Hindi), we converted the text into speech using gTTS module in python. The audio waveform of the obtained audio is shown in Fig.4 above.

5. CONCLUSION

Knowledge is never ending and we strongly believe that it should reach to each and every individual regardless of the location where they stay, and the background they belong to. Knowledge is the best investment and what can be a better option than investing your time in reading a nice book? Even famous authors and leaders have openly admitted how a single book has drastically changed their lives in a positive way. Practically, it is not possible for everyone to buy each and every book and carry with them where ever they go. But, this should not become a hurdle in their way, thankfully technology is advancing every day and that’s why today we have come up with our idea of providing audiobooks, which is nothing but a summarization of the whole book and translating it in Hindi using Machine Translation (MT) technology and converting it into speech with the help of TTS syntesis, without changing the essence or altering the content of the book. The accuracy of the book summary is maintained.Emerging technologies such as AI,ML,NLP,Text-to-Speech synthesizer has been a great helping hand in the audiobook industry in the recent years. Although, TTS Synthesizer lacks the nuances of human speech, making it sound more like a robot with no feelings and expressions. Majority of TTS applications are unsuccessful at producing nuances like expressive voice tones,pauses in between and so on, which ultimately leads to low-quality results. Researchers and developers are trying to train AI to make the speech sound more realistic using Speech-to-Speech(STS) voice cloning, AI powered technology that improves TTS Speech quality. It uses a person's speech and produces the speech in a different person's voice. For instance.: We can make our own voice sound like someone else. Our platform BookSnap will be accessible and free of cost to everyone. Our main motive to keep it for free of cost, was to make sure that everyone can benefit out of it, regardless of their background. Since it is an audiobook of precise summary, thus, it is very convinient for people to listen to it from where ever they are and whenever they want, which otherwise becomes a bit difficult when reading an actual book. The classic feature of any audiobook is that it helps one to keep track of different books they've read and they can resume from where they've last read for a particular book.

The world is progressing, but still majority of individuals don't have access to public library. Here, audiobooks play avital role. Today, we have a vast library of books and novels and it is not possible for everyone to buy each book of their choice. At this point, an audiobook of precise summary comes into picture, where it allows people to connect with their book of choice that too for free of cost.Back in 1824 Braille was invented for the differently abled (Blind), so that they are not left behind. Access to knowledge should not have any boundaries or limitations and thus, audiobooks are differently abled(Blind) friendly. We strongly believe that, access to knowledge should not be situational, that's why BookSnap is such a platform which is not just user-friendly, free of cost but strongly believes in growing by uplifting everyone with knowledge.

AI powered technologies have a rapid growth in the audiobook industry in the upcoming times, by introducing unique features. This will increase the demand of Audiobook industry and thus leading to high competition where competitors will indulge themselves into adding unique features in order to improve the production process. As of now, we can patiently wait for the new features AI powered technology gets in the audiobook industry.

6. REFERENCES

1) https://analyticssteps.com/blogs/how-ai-used-audiobook-industry
2) https://www.respeecher.com/blog/text-to-speech-ai-voice-generator-creating-human-like-voice
3) https://www.analyticsvidhya.com/blog/2019/01/neural-machine-translation-keras/
4) Basha, S. M., Poluru, R. K., & Ahmed, S. T. (2022, April). A Comprehensive Study on Learning Strategies of Optimization Algorithms and its Applications. In *2022 8th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE.
5) Wei Gao, Cheng Niu, Ming Zhou, Kam-Fai Wong :Joint Ranking for multilingual web search. ResearchGate Publication, April(2009)
6) Santosh GSK, Kiran Kumar N, VasudevaVarma : Ranking Multilingual Documents using Minimal Language Dependent Resources. International Institute of Information Technology , Hyderabad,India
7) Kiran Kumar N, Santosh GSK, VasudevaVarma : Multilingual Document Clustering using Wikipedia as External Knowledge. International Institute of Information Technology , Hyderabad,India
8) Authors-unknown : Multilingual Wordnet sense Ranking using nearest context
9) Luo Si, Jamie Callan -CLEF 2005: Multilingual Retrieval by Combining Multiple Multilingual Ranked Lists. Language Technology Institute, School of Computer Science Carnegie Mellon University, Pittsburgh, Pennsylvania, USA