UDC 004.942(045)

BUILDING A SCALABLE DATASET FOR FRIDAY SERMONS OF AUDIO AND TEXT (SAT)

Samah A. A. – Postgraduate student of Department of Information Systems, Faculty of Computing and Information Technology, and Lecturer of Department of Management Information Systems, Faculty of Economics and Administration, King Abdul Aziz University, Jeddah, Mecca, Saudi Arabia.

Dimah H. A. – PhD, Associate Professor, Associate Professor of Department of Information Systems, Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Mecca, Saudi Arabia.

Hassanin M. A. – Dr. Sc., Professor, Professor of Department of Information Technology, Faculty of Computing and Information Technology, King Abdul Aziz University, Jeddah, Mecca, Saudi Arabia.

ABSTRACT

Context. Today, collecting and creating datasets in various sectors has become increasingly prevalent. Despite this widespread data production, a gap still exists in specialized domains, particularly in the Islamic Friday Sermons (IFS) domain. It is rich with theological, cultural, and linguistic studies that are relevant to Arab and Muslim countries, not just religious discourses.

Objective. The goal of this research is to bridge this lack by introducing a comprehensive Sermon Audio and Text (SAT) dataset with its metadata. It seeks to provide an extensive resource for religion, linguistics, and sociology studies. Moreover, it aims to support advancements in Artificial Intelligence (AI), such as Natural Language Processing and Speech Recognition technologies.

Method. The development of the SAT dataset was conducted through four distinct phases: planning, creation and processing, measurement, and deployment. The SAT dataset contains a collection of 21,253 audio and corresponding transcript files that were successfully created. Advanced audio processing techniques were used to enhance speech recognition and provide a dataset that is suitable for wide-range use.

Results. The fine-tuned SAT dataset achieved a 5.13% Word Error Rate (WER), indicating a significant improvement in accuracy compared to the baseline model of Microsoft Azure Speech. This achievement indicates the dataset's quality and the employed processing techniques' effectiveness. In light of this, a novel Closest Matching Phrase (CMP) algorithm was developed to enhance the high confidence of equivalent speech-to-text by adjusting lower ratio phrases.

Conclusions. This research contributes significant impact and insight into different studies, such as religion, linguistics, and sociology, providing invaluable insights and resources. In addition, it is demonstrating its potential in Artificial Intelligence (AI) and supporting its applications. In future research, we will focus on enriching this dataset expansion by adding a sign language video corpus, using advanced alignment techniques. It will support ongoing Machine Translation (MT) developments for a broader understanding of Islamic Friday Sermons across different linguistics and cultures.

KEYWORDS: Friday Sermons, Khutbah, Arabic speech recognition, Audio and text dataset, Machine translation.

ABBREVIATIONS

AI is an Artificial Intelligent;
ArSL is Arabic Sign Language;
ASR is Automatic Speech Recognition;
CMP is a Closest Matching Phrase;
P1 is a Name of Preacher;
P2 is an Age of Preacher;
P3 is an Original Country of Preacher;
P4 is an Academic Qualification of Preacher;
P5 is a Years of Experience of Preacher;
DL is a Deep Learning;
IFS is an Islamic Friday Sermons;
ML is a Machine Learning;
MT is a Machine Translation;
NLP is a Natural Language Processing;
PCM is a Pulse Code Modulation;
SAT is a Friday Sermon Audio and Text
S6 is a Title of Sermon;
S7 is a Type of Sermon (topic);
S8 is a Duration of Sermon;
S9 is a Date of Sermon;
S10 is a Place of Sermon;
S11: Language of Sermon,
S12 is other languages of Sermon translated into;
S13 is other sign languages used for translating Sermon;
S14 is a Language complexity of Sermon;
S15 is a Reliability of Manarat Al-Haramain Website;
S16 is a Reliability of AL-Khutaba Forum Website.
SL is a Sign Language;
WER is Word Error Rate.

NOMENCLATURE

A_i is an Audio recording to the ith item;
FT is a full text of one Friday Sermon;
k is representing the number of raters or judges;
N is total number of pairs in the dataset;
n is a number of observation (items) or cases being assessed;
R is a similarity between transcript and current phrase using Sequence Matcher (ratio);
r is a Pearson correlation coefficient;
SS is a sum of squares for total ranks;
T_i is a corresponding transcript to the ith item;
W is Kendall's Coefficient of Concordance;
X and Y are indicating the variables;
$\bar{x}$ and $\bar{y}$ are indicating the means of the two variables;

DOI 10.15588/1607-3274-2024-2-10

x_i is a rank or score given to the *i*th item by raters;

$\bar{x}$ is a mean (average) rank of all items assessed.

INTRODUCTION

In the past few years, large-scale datasets have become an essential step in applying artificial intelligence (AI) technologies, such as machine learning (ML) and deep learning (DL), in various sectors. These datasets can be created or collected from different types of data, which could be text, audio, video, or pictures. Each of these types of datasets has different ways of annotating, processing, and analyzing it, in order to develop or enhance the system. Overall, it supports decision-making in a specific sector. Thus, the task of collecting and creating a dataset, usually, requires a huge extensive effort from researchers in order to reach the expanded dataset in a certain domain [1,2].

In Arab countries, many researchers conducted their efforts to create and collect a huge Arabic dataset that serves many fields, such as education and healthcare, where they used AI technologies [3–5]. However, some fields like religion did not receive more attention from researchers, especially, in creating and collecting a dataset of sign language (SL) which is considered as a main unified communication language used by deaf communities [6].

Generally, in the religious domain, Islamic Friday Sermons (IFS), which are a key aspect of religious practice delivered during congregational prayers on Fridays, remain understudied. A few researches have been introduced in analyzing and understanding religious texts that are related to Sermons. Their focus was on the linguistic perspective (rhetorical structure of the Sermon), specifically, from pragmatics and discourse analysis aspects [7–9]. This ISF is a rich source of theological, cultural, and linguistic knowledge. Due to that, the aim of this research is to create a comprehensive Sermon dataset, a beneficial resource for researchers working with Islamic Sermons.

Despite increasing interest in this type of scientific research, leading to the development of various Natural Language Processing (NLP) and ML applications [10], these works still have limitations in scope and are not suited for large-scale computational analysis. The reason behind that is a lack of a Sermon dataset that has volume, value, variety, and metadata availability, a gap that this study aims to address.

The Object of Study is the process of collection, creation, and analysis of a comprehensive large-scale Friday Sermon dataset including audio, and text. This process includes creating a dataset and an algorithm implemented to enhance the recognition.

The subject of study is a methodology for creating a dataset of Friday sermons and identifying the type of dataset (audio, text, Sign Language (SL) videos). Another subject is identifying significant parameters that need to be considered from the Friday sermons presenter (Preacher) and the Friday sermons content in the collected dataset. In addition, the way of evaluating this created Friday sermons dataset.

The purpose of the work is to create, collect, and evaluate that aimed at enhancing language processing and recognition technologies. Moreover, this work was conducted to fill the existing gap in large-scale computational analysis of Friday Sermons by providing a rich dataset that has volume, value, variety, and accessible metadata. In addition, it supports the fields of Islamic Studies, Social Sciences, Linguistics, and AI with a useful resource.

The Islamic Friday Sermon (IFS) which is called in Arabic (Khutba AL-Jummah or Friday Khutbah) is a formal religious speech introduced on each Friday of the week. In Islam, Friday is considered the greatest day for Muslims to prepare themselves by praying in the mosque and listening to the Sermon [11].

Some researchers indicate that the IFS have a significant influence on humans' beliefs, attitudes, and behaviors. Also, it can influence the religious and cultural identity of Muslim communities. It can solve some issues in communities, like social and political issues such as inequality, injustice, and discrimination. Moreover, the Friday Sermon may play an important role in shaping national identity. From this standpoint, we can consider the Friday sermon data as valuable data that deserves study to under-stand the nature of its impact on different societies. In addition to the possibility of benefiting from the impact of Friday sermons on strengthening national identity, consolidating beliefs, controlling the behavior of community members, and directing them in the right direction [12].

These Islamic Friday Sermons will be composed and re-viewed based on the selected topic by the Preacher (Presenter of Friday Sermon), which is a person who delivers a Sermon to the congregation. The topic that was selected can be related to religion, community issues, morality, con-temporary challenges...etc. [13]. One of the researchers mentioned, generally, religious sermons are divided into four main types. First is religious education for the public. Second is proving faith in the souls. Third is correction of faults and prohibition of evils. Fourth is invitation to Islam or its defending [14].

In general, the speech of the Sermon on this greatest day should be introduced by the Preacher in a clear and interesting manner using understandable vocabulary. Mainly, the IFS duration without the Azan and prayer is around 30 to 40 minutes and consists of two Sermons where there is a short silence around 1 to 3 minutes between them. Usually, the first Sermon is longer than the second. There is Azan before the first Sermon and at the end of the second Sermon, there is prayer. These two parts are called the beginning and closing parts of the sermon (Sermon Prayer) included regularly in the structure of a Sermon. Whereas, the two Sermons that are in the middle are the body of the Sermon [9, 15], as shown in Fig. 1.

DOI 10.15588/1607-3274-2024-2-10

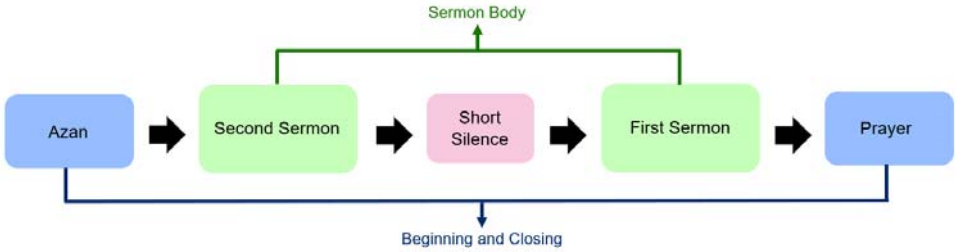


Figure 1 – Block diagram of Islamic Friday Sermon structure

Typically, the first Sermon contains the following: 1) Hamd Allah or Praise (thanks Allah). 2) Salawat, and Salam for the prophet Mohammed (Peace Be Upon Him). 3) Discussion of the main content or theme (topic) and reading some Ayat from AL-Qur'an (Recitation of Quranic Verses). 4) Advice and reminders with supplication which is called in Arabic (Dua'a). It is calling upon Allah (God) with respect and faithful for different reasons such as forgiveness, helping, and safeguarding the country and its leaders. However, the second Sermon started once again by Hamd Allah or Praising and Salawat, and Salam for the prophet Mohammed (Peace Be Upon Him). Then, the Preacher will continue the same topic that began in the first Sermon with more highlighting of more important points in the topic. Moreover, concentrates on reminding the congregation of certain Islamic obligations, virtues, or main issues related to the community. Finally, it will be ending also by supplication to all Muslim communities around the world and reminding them by doing what Allah said in order to achieve good deeds [16].

1 PROBLEM STATEMENT

The problem at hand is the lack of a scalable dataset for Friday sermons in both audio and text formats, which hinders the development and evaluation of automated systems for analyzing and understanding sermon content. The absence of such a dataset limits the advancements in NLP and speech recognition research specifically tailored for sermon analysis and related applications.

The current state of available datasets for Friday sermons is either limited in size, restricted to specific languages or regions, or lacks the necessary annotations for comprehensive analysis. This scarcity prevents researchers and developers from effectively training and evaluating machine learning models and algorithms for tasks such as sentiment analysis, topic extraction, speaker identification, or content summarization within the context of sermon texts and audio recordings.

Furthermore, the complexity of these Friday Sermons, besides the intense need for accurate dataset representation in digital form (audio, text), requires a dataset that is not only extensive in volume and variety but also rich in metadata to support computational analyses. So, we can find that were created dataset through this study includes a scalable dataset for Friday sermons in both audio and text formats, encompassing diverse languages, regions, and religious denominations. This dataset has been annotated with relevant metadata such as speaker information, sermon topic, date, and location, enabling researchers and developers to explore various aspects of sermon content using both NLP and speech recognition techniques.

Thus, the created dataset presented as:

$$\mathrm{SAT} = \{(A_i, T_i)\}_{i=1}^{N},$$

where the SAT represents our created Sermon Audio and Text dataset, A_i indicates audio recording, T_i indicates the corresponding transcript, and N represents the total number of pairs in the dataset.

Moreover, for enhancing the accuracy of recognition by utilizing our SAT dataset, a similarity matching algorithm for finding the Closest Matching Phrase (CMP) between the transcript T and the full-text FT was used. Therefore, we can say that using our created SAT dataset which includes (Audio and Text) in any customized speech recognition application (tunned) will enhance the accuracy by reducing the WER of ASR as output for any Speech recognition system, as shown in the formula:

$$\mathrm{SAT}=\{(A_i,T_i)\}_{i=1}^{N} \rightarrow \mathrm{ASR}(\mathrm{WER}_{after\_tuning} < \mathrm{WER}_{before\_tuning}).$$

To identify the Closest Matching Phrase CMP within FT that most closely matches T we defined a function using similarity measure R, which is employed as:

$$f(T, FT) \rightarrow \mathrm{CMP},$$

where the similarity measure R is obtained from the Sequence Matcher algorithm which is the ratio of similarity between T and the current phrase being evaluated within FT. Also, we defined the minimum ratio minR which means if the value R is less than the minR that means can not be accepted to be similar.

Therefore, the minR ≤ R ≤ maxR, where minR=0.50 and maxR is the maximum achievable ratio, ensures R falls within this range to be considered a valid match.

Hence, the main challenge of this work seeks to address a large-scale dataset and comprehensive metadata for the Friday Sermon Audio and Text (SAT) dataset. In addition, a novelty algorithm is implemented to enhance the recognition. Overall, this work tackles the existing gap in the analysis and processing of Islamic Friday Sermons.

DOI 10.15588/1607-3274-2024-2-10

2 REVIEW OF THE LITERATURE

In terms of the linguistic studies field, some study efforts have focused on different aspects of the IFS. One of these studies used 65 texts of the Yemeni-Arab Sermon to study the usage of deixis analysis. This deixis analysis helps people understand the meaning behind certain sentences based on their context. In general, deixis is divided into five types: they are person deixis, place deixis, time deixis, social deixis, and discourse deixis. Through this study, the researchers focused on studying deixis analysis from pragmatic and discourse perspectives. They had a limitation in using a small dataset of Sermons that needed to be translated into English for conducting their experiments [17]. Similarly, a study [9] used deixis in the English Islamic Friday Sermon using 70 texts from the English Friday Sermon dataset from multiple online sources. This study ended by acknowledging the small size of the dataset as a limitation. Another study was conducted based on the interpersonal model of metadiscourse for analyzing 30 text and speech English Friday Sermon datasets that were collected from various online sources. Also, they highlighted the limitations of the Sermon dataset [15], [18]. A study [19] focused on directive speech acts performed in the Sermon using the 56 Sermon dataset from the Islamic Religious Council of Singapore. They found that Friday Sermons use different strategies of directive speech acts.

In addition, one of the studies focused on the phoneme distribution in Malay Friday Sermon derived from 52 speech transcripts that are available on a government website. They reached the same limitation of having a small number of words collected and analyzed [7].

In terms of sociolinguistics and discourse analysis, two of the studies focused on the Sermon's duration. The first study of [11] conducted an analysis of Friday Sermon duration. They found that a shorter Sermon may be indicative of the Preacher's expertise in religious affairs. The second study used a descriptive method (questionnaire) in order to assess the congregation's understanding of the Friday Sermon discourse. The result of their study was that most congregations preferred Sermons with a duration of 15–20 minutes [20]. In studying the content and thematic analysis, [21] carried out a content analysis of Friday Sermons by the Turkish-Islamic Union for Religious Affairs in Germany, integrating sociolinguistics and discourse analysis. However, this study was limited to local text Friday Sermons that may not have received more attention from all Muslims around the world. Their dataset was 481 that were obtained from 2011 to 2019 on the DİTİB website. Another study conducted a thematic analysis of the Friday Sermon in Negeri Sembilan. They highlighted the importance of selecting topics that engage the congregations while considering their cultural background and educational level. However, their limitation was that the study was confined to Sermons from one region [22].

Other studies have employed a multidisciplinary approach to scrutinize the Sermon. One of the studies used ML techniques to evaluate the impact of Turkey's Friday Sermons on Twitter users. However, this study was focused on examining only one Sermon feature, which is the topics that are handled in Sermons [10].

Based on the illustrated previous studies, we can conclude that there are a few researches that have been introduced to analyzing and understanding religious texts that are related to Sermons. Their focus was on the rhetorical structure of the Sermon, specifically pragmatics and discourse analysis aspects, by utilizing a limited speech and text Sermon dataset. These types of scientific research have gained significant attention among researchers and opened avenues for the development of various NLP and ML applications for studying more parameters of the Sermon dataset. For example, themes (topic or domain), title, duration, date of the Sermon, location (place of the Sermon), and language of the Sermon … etc. Also, from some studies, we found that we need to be aware of the Preacher's parameters, whereas a study [7] emphasized the importance of the Preacher's expertise in religious affairs in conveying the concept of the Sermon to the congregations in a short duration. Thus, we can highlight some of Preacher's parameters, such as their years of experience, their original country, and so on.

Still, these works have limitations and are not appropriate for large-scale computational analysis. The reason behind that, from our perspective, is a lack of the Sermon dataset and its metadata availability, which is a gap that this study aims to address. In our study, we are going to create a dataset of Sermons that contain Arabic speech and text.

3 MATERIALS AND METHODS

The collection and creation of our dataset followed a structured, four main phase approach. It is designed to ensure the dataset's integrity, relevance, and utility. Each phase contained specific stages (steps) that should be successfully finished to move on to the next step in the next phase.

The nine stages are illustrated in Fig. 2. Each phase and its stages will be explained in more detail.

4.1 Planning Phase: this phase includes three main stages, which are: A) Design and implement a questionnaire. B) Analyze the questioner. C) Identify the parameters of data collection. The explanation of these stages is as follows:

A) Designing and Implement Questionnaire: we used a questionnaire in order to ask the specialists in data science about the important parameters that should be included in our data and metadata. It was designed in three main parts in accordance with the axes of the questionnaire:

1) Personal Data. 2) Data for Preacher (Presenter of Friday sermon). 3) Data for Friday sermon. Each part was written and designed to collect specific data related to this study's objectives (see in the appendix Fig. A1, Fig. A2, Fig. A3, and Fig. A4).

It was distributed electronically using a Google Form. We used expert ratings for multiple parameters of Preacher and Sermon.

DOI 10.15588/1607-3274-2024-2-10

The response of (n = 50) was obtained by 28 males and 22 females. The parameters encompassed characteristics of Preacher and Sermon, with five related to Preacher and eleven related to Sermon. Each expert has rated each parameter's im-portance based on (large, medium, and little).

B) Analyze Questionnaire: our analysis of the questionnaire was conducted based on the following: (1) Describing the collected data from expert evaluations for each parameter. (2) Finding correlations between experts' evaluations of each parameter. (3) Measuring the agreement between experts' evaluations for each parameter. These steps will support decision-making about the important parameters that should be considered in creating Fraidy sermon data and its metadata.

(1) Description Analysis of Parameters (Based on Expert Evaluation): in order to analyze the expert evaluation data for identifying which parameters are important and needed to be included in our created dataset, we convert them on a scale from 1 (little importance) to 3 (large importance). Also, we add notation for each parameter. The descriptive statistics for each parameter (Preacher and Sermon), including the mean, median, and standard deviation, are shown in Table 1.

These statistics provide insights into the perceived importance of each parameter.

In Preacher parameters, the Academic Qualification P4, "Years of Experience" P5, and "Name of Preacher" P1 had a slightly high mean rating of 2.54, 2.42, and 2.34 respectively, which means these three parameters are significantly important. Conversely, the "Age of Preacher" P2 had a lower mean rating of (x =1.46), implying less perceived importance.

In Sermon parameters, the "Title of Sermon" S6 obtained (x= 2.86) mean rating, which indicates it is significantly important. Also, it received the highest median rating of 3.0, which means this parameter is important from most experts' perspectives. By looking at the standard deviation for these ratings, we can see the level of consensus or disagreement among the experts.

The parameters with a lower standard deviation indicate a greater consensus among experts regarding their importance. For example, "Title of Sermon" S6 showed the least standard deviation, which refers to a strong agreement among experts on its significance.

Moreover, the "Reliability of Manarat Al-Haramain website" S15 and the "Other sign languages used for translating Sermon" S13 received significant importance, with mean ratings of 2.74 and 2.62, respectively.

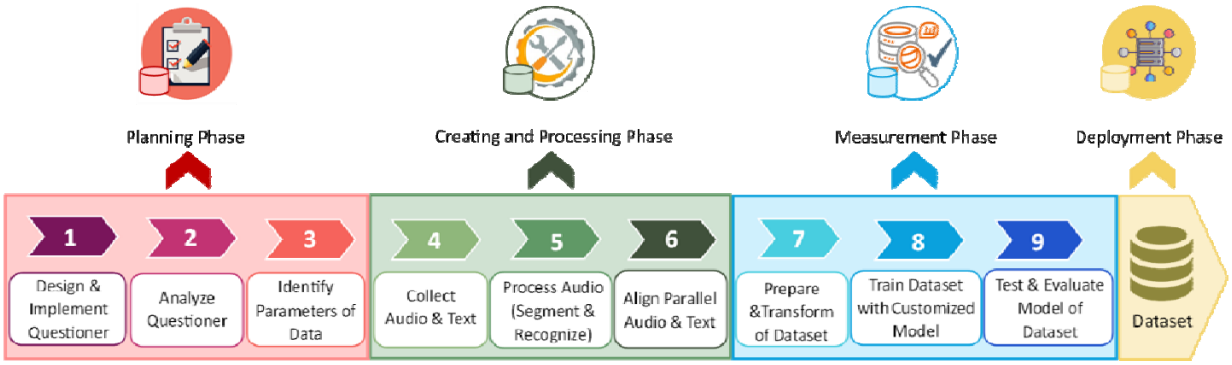


Figure 1 – Phases and stages of Sermon dataset collection

Table 1 – Descriptive statistics for the Preacher and Sermon parameters

	Notation	Parameter Name	Mean	Median	StdDev
Preacher Parameters	P1	Name of Preacher	2.34	3.0	0.772222
	P2	Age of Preacher	1.46	1.0	0.645550
	P3	Original Country of Preacher	1.84	2.0	0.817163
	P4	Academic Qualification of Preacher	2.54	3.0	0.734291
	P5	Years of Experience of Preacher	2.42	3.0	0.784805
Sermon Parameters	S6	Title of Sermon	**2.86**	**3.0**	**0.452205**
	S7	Type of Sermon (topic)	2.52	3.0	0.646498
	S8	Duration of Sermon	2.42	3.0	0.702474
	S9	Date of Sermon	2.48	3.0	0.706818
	S10	Place of Sermon	2.58	3.0	0.609114
	S11	Language of Sermon	2.48	3.0	0.706818
	S12	Other languages of Sermon translated into	2.56	3.0	0.674915
	S13	Other sign languages used for translating Sermon	**2.62**	**3.0**	**0.567486**
	S14	Language complexity of Sermon	2.26	2.0	0.694292
	S15	Reliability of Manarat Al-Haramain Website	**2.74**	**3.0**	**0.486973**
	S16	Reliability of AL-Khutaba Forum Website	2.40	3.0	0.699854

DOI 10.15588/1607-3274-2024-2-10

Also, S15 and S13 the same as "Title of Sermon" S6 received a 3.0 median rating while the standard deviation of both parameters was low, which means there is a strong agreement among experts on its significance.

(2) Correlation Analysis of Parameters (Based on Expert Evaluation): We calculated the Pearson correlation coefficient (r) between each of the two parameters as expressed in the equation (1).

$$r = \frac{\sum\left(X - \bar{x}\right)\left(Y - \bar{y}\right)}{\sqrt{\sum\left(X - \bar{x}\right)^2 \sum\left(Y - \bar{y}\right)^2}}, \quad (1)$$

where X and Y indicate the variables, $\bar{x}$ and, $\bar{y}$ indicate the means of the two variables [23].

We used the heatmap visualization using the Seaborn library in Python. Fig. 3 shows a valuable insight into the relationship between different parameters associated with the Preacher and the Sermon itself. The X-axis presents Sermons' parameters whereas the Y-axis presents Preachers' parameters Mainly, the positive correlation between Preacher and Sermon parameters suggests that the experts perceive the increasing importance of Preacher parameters the same as increasing Sermon parameters.
Interestingly, the "original country of Preacher" P3 has a strong positive correlation with the "language of the Sermon" S11, indicating that the language used in the Sermon is highly important and that its importance will increase if "Preacher's original country" is increasing.

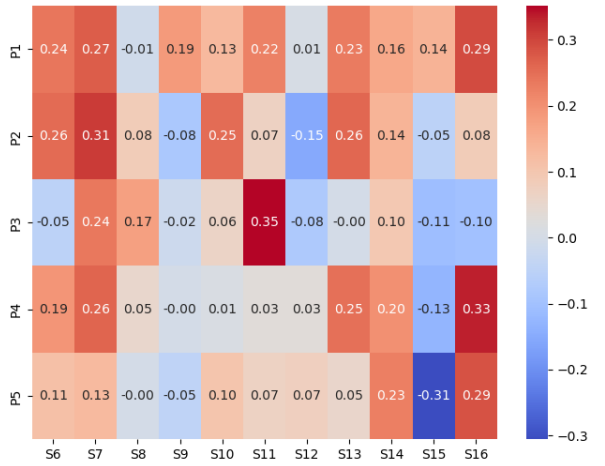


Figure 3 – Heat map correlation between Preacher and Sermon parameters. (P1) Name of Preacher; (P2) Age of Preacher; (P3) Original Country of Preacher; (P4) Academic Qualification of Preacher; (P5) Years of Experience of Preacher; (S6) Title of Sermon; (S7) Type of Sermon(topic); (S8) Duration of Sermon; (S9) Date of Sermon; (S10) Place of Sermon; (S11) Language of Sermon; (S12) Other languages of Sermon translated into; (S13) Other sign languages used for translating Sermon; (S14) Language complexity of Sermon; (S15) Reliability of Manarat Al-Haramain Website; (S16) Reliability of AL-Khutaba Forum Website

Thus, we can find that the language used in the Sermon is highly influenced by the Preacher's original country. Therefore, these two parameters should be considered in IFS metadata.

Also, we can see that "Preacher's name" P1, "academic qualifications of Preacher" P4, and "years of experience" P5 show a positive correlation with the "reliability of AL-Khutaba Forum website" S16 a website that provides a source for written texts of Friday sermon for various Sermon places. This suggests that a specific Preacher's name with higher qualifications and more experience tend to be associated with more reliable content on the AL-Khutaba Forum website.

On the contrary, the "reliability of the Manarat Al-Haramain website" S15 was not affected positively by all Preacher parameters because this website was provided by the government as a source for visual videos of Sermons. Thus, it certainly achieved high important ratings from experts without looking at other parameters' impact. Overall, this could mean that the expert's rating sees a connection between these parameters and believes they both contribute to the effectiveness or impact of IFS metadata. However, it's crucial to note that this is only an indication of how the parameters are related in terms of their perceived importance. It does not necessarily mean that they influence each other in a casual way. For studying the effectiveness and causes, a more in-depth analysis would be necessary with IFS metadata.

(3) Inter-annotator Agreement of Parameters (Based on Expert Evaluation): This study used Kendall's Coefficient of Concordance (Kendall's W), which is a measurement tool of a non-parametric test for rank correlations and for inter-reliability where its agreement is from 0 (no agreement) to 1 (complete agreement) [24]. The categories degree scale of Kendall's W is illustrated in Table 2.

Table 2 – Categories of Kendall's W interpretation

W	**Interpretation**
0	No agreement
0.10	Weak agreement
0.30	Moderate agreement
0.60	Strong agreement
1	Perfect agreement

To use Kendall's W, the rate for each item should be rearranged so that it is given by each rater as a rank starting from 1, 2, 3 … etc. If there is more than one item that has the same rate, such as item_1 =2, item_2 =2, each of the two items will have a different rank. Then, the summation of their ranks will be divided by the total number of items that are given the same rate (1 + 2 / 2). Thus, the result of 1.5 will be given to item_1 and item_2 as rank. After that, we calculate the Kendall's coefficient (W) using the following equation:

$$W = \frac{12\ SS_{Total}\ \ Ranks}{k^2\left(n^3 - n\right)}, \quad (2)$$

where SS calculated by the formula:

DOI 10.15588/1607-3274-2024-2-10

$$SS = \sum_{i=1} \left(x_i - \bar{x}\right)^2, \quad (3)$$

where x represents the total ranks for each i item that given by raters, $\bar{x}$ is the mean rank of x_i, and n is the number of items or cases being assessed, k represents the number of raters or judges [25].

In our study, we used SPSS to calculate the value of Kendall's $W = 0.210$, which indicates that there is a slight level of agreement among the raters for the ordinal or ranked data. Table 3 represents the mean ranks using Kendall's W, where this rank shows us, which parameters were rated most favorably.

Table 3 – Ranks using Kendall's W test

	Mean Rank
P1	8.15
P2	3.45
P3	5.59
P4	9.38
P5	8.75
S6	11.14
S7	9.15
S8	8.51
S9	8.75
S10	9.38
S11	9.00
S12	9.22
S13	9.63

P1: Name of Preacher, P2: Age of Preacher, P3: Original Country of Preacher, P4: Academic Qualification of Preacher, P5: Years of Experience of Preacher, S6: Title of Sermon, S7: Type of Sermon (topic), S8: Duration of Sermon, S9: Date of Sermon, S10: Place of Sermon, S11: Language of Sermon, S12: Other languages of Sermon translated into, S13: Other sign languages used for translating Sermon, S14: Language complexity of Sermon, S15: Reliability of Manarat Al-Haramain Website, S16: Reliability of AL-Khutaba Forum Website.

As we can see in Table 3, the lower parameter in rank is "Age of AL Preacher" P2 = 3.45, which indicates that it does not have a high chance of being selected as a parameter for the Sermon dataset. At the same time, in the parameter of "Original Country of Preacher" P3 has a rank of 5.59, which is a low rank and does not rate it as the most important parameter. In contrast, "Title of Sermon" S6 and "Reliability of Manarat Al-Haramain Website" S15 stand out with higher rankings of 11.14 and 10.24, respectively. Moreover, S13, P4, S10, S12, S7, and S11 have ranks around 9, which is a high level of agreement between the expert's raters regarding the importance or evaluation of these parameters (rated most favorably).

C) Identify Parameters of Data Collection: the domain of our dataset that will be created is the Islamic Friday Sermons as we can spotlight the importance of the Islamic Friday based on the expert's perspective on the evaluation questionnaire. We illustrated in the previous studies that researchers focused their studies on the collection of either text or speech Sermon data, not both [9, 21]. However, in our study, the aim is to collect a dataset with a size of 100, including 50 audio and 50 corresponding texts, with a total of 50 Islamic Friday sermons from the Grand Mosque (Masjid al-Haram) in Makkah, located in Saudi Arabia. Where experts also highlight the significance of the Grand Mosque. In addition to that, the experts' spotlight on the data type's importance in having Fraidy Sermon text and audio. Through this re-search, we will focus on collecting and creating audio and text. The intention is to utilize this comprehensive dataset in some applications of ML and DL techniques.

In addition to identifying the size and type of dataset, we identify the parameters for both Preacher and Sermon that should be collected in order to create Friday Sermons metadata. We identify the Preacher's parameters and the Sermon's parameters based on the results (Key Findings) of the evaluation questionnaire.

– Key findings in evaluating preacher and sermon parameters:

– In preacher parameters: The experts' evaluation of the importance of Preacher parameters revealed that the "Age of Preacher" has less significance, suggesting it may not contribute significantly to creating our dataset. On the other hand, the following parameters of "Name of Preacher", "Academic Qualification of Preacher", and "Years of Experience of Preacher" had more significant importance, warranting their inclusion in our dataset. Although the "Original Country of Preacher" had less importance, it has a strong correlation with the "language of the Sermon", making it relevant to the "Place of Sermon" and the "Language" used. Based on these strong correlations between the "Original Country of Preacher", "language of the Sermon", and "Place of Sermon" it is reasonable to eliminate the "Original Country of Preacher" which can be inferred from the place and language used for Sermon. Since the place and language used for the Sermon can already provide insights into the cultural and linguistic context, retaining the "Original Country" parameter may not contribute significantly to identifying the more important features for dataset creation. By eliminating this parameter, you can focus on gathering and incorporating the more essential features that have a direct impact on the Sermon.

– In sermon parameters: Several key findings appeared from the evaluation of the im-portance of Sermon parameters. The "Title of Sermon" was observed to hold significant importance, serving as a concise representation of the main theme or topic. Similarly, the "Topic of Sermon" was identified as another crucial parameter, reflecting the subject and content of the Sermon and it figures relevance to the audience's engagement. Also, the "Duration of Sermon" played a significant role in its importance, as we infer from one of the previous studies. It proved that the "Duration of Sermon" influences the audience's attention [20]. Additionally, the "Language of Sermon" was found to be significant, affecting audience accessibility and understanding. In short, both the "Date" and "Place" of the Sermon were deemed significant, as they contributed to the overall impact and resonance of the audience. The parameter of "Other languages Sermon

DOI 10.15588/1607-3274-2024-2-10

is translated into" showed slight importance. However, we believe that this parameter may gain significant importance in future work, especially when researchers who specialize in language translation focus their interest on the different languages into which Sermon is translated. Another parameter that proved to be highly important in our dataset is "Other sign languages used for translating Sermon". Including SL videos and text of Sermon in our data is crucial for making Sermon accessible to the deaf community. Moreover, for blind people, the audio of Sermons is also referred to by some experts as having significant importance to be included in our dataset.

In terms of sourcing Sermon videos and texts, we relied on two websites, Manarat Al-Haramain and AL-Khutaba Forum. Manarat Al-Haramain proved to be a more reliable source as it is backed by the government, whereas AL-Khutaba Forum, though still valuable, was considered less reliable for our dataset. In general, we will consider these two websites as significant sources for obtaining video and text for creating our dataset. On the other hand, we observed that the parameter "Complexity of Sermon" is less important compared to the other parameters. Therefore, we decided to eliminate this parameter from our dataset.

Based on Fig. 2 of the stages of 3.2 Creating and Processing Phase: in this phase, we start to create and process our Sermon dataset, considering the important parameters. Then, processing the audio (segment and analyze) was implemented. Finally, the alignment of the parallel (audio and text) was done in the last step in the creating and processing phase.

In order to go through these three stages (steps) of creating and processing phase, the two approaches (high and low level) were used; see Fig. 4 and Fig. 5.

In the high-level approach, we divide the process of creating our Friday Sermons audio and text dataset (SAT) into multiple modules. The first module is data preprocessing, which is part of the audio and text dataset collection stage. Then, the data seg-mentation and recognition, data annotation, and metadata creation modules were considered as part of the processing data (segment and recognize) stage. While the data verification, and data correction and unification modules are part of aligning parallel audio and text.

The three stages and their relevant modules in the high-level approach for collecting and creating SAT presents in Fig. 4. However, the low-level approach includes the subprocesses (steps) of each of these modules shown in Fig. 5. The deep explanation of each module shows as following:

1) Data preprocessing module: It includes the following subprocess:

– Obtaining the text and video Sermons: we collected videos of Sermons from the website of Manarat Al-Hramain [26], which is released by the Saudi Arabian government, and also had a high level of agreement between the experts' evaluations regarding the importance parameter (rated most favorably) (see Table 2).

From the Manarat Al-Haramain website, we collect around 50 mp4 videos of Fraidy Sermon with a normal resolution of 480p with 30fps that are held in the Holy Mosque of Makkah. On the other hand, we collect 50 texts of Sermons corresponding to each collected Sermon video from the website of the AL-Khutaba Forum, which encompasses Sermons delivered in multiple places in Saudi Arabian mosques, such as Riyad and Jeddah. Also, it has Sermons in different countries, such as al-Aqsa mosque, Egypt mosques, and so on [27]. Based on experts' evaluation regarding the importance parameter (rated most favorably), the AL-Khutaba Forum website recorded a slightly high mean rank (see Table 2).

Therefore, we considered this website a reliable source for gathering relevant text, where each text appeared on the website in docx format.

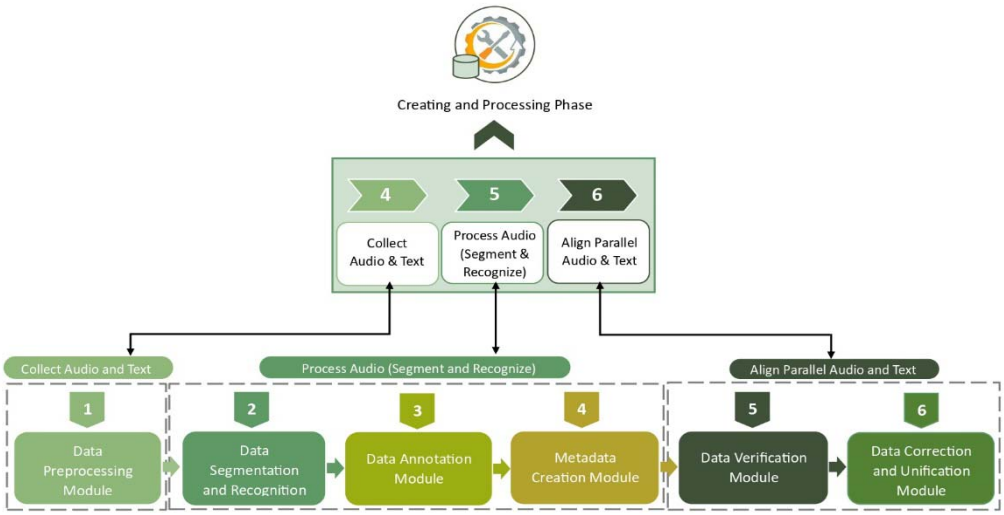


Figure 4 – High-Level approach of collecting and creating SAT

DOI 10.15588/1607-3274-2024-2-10

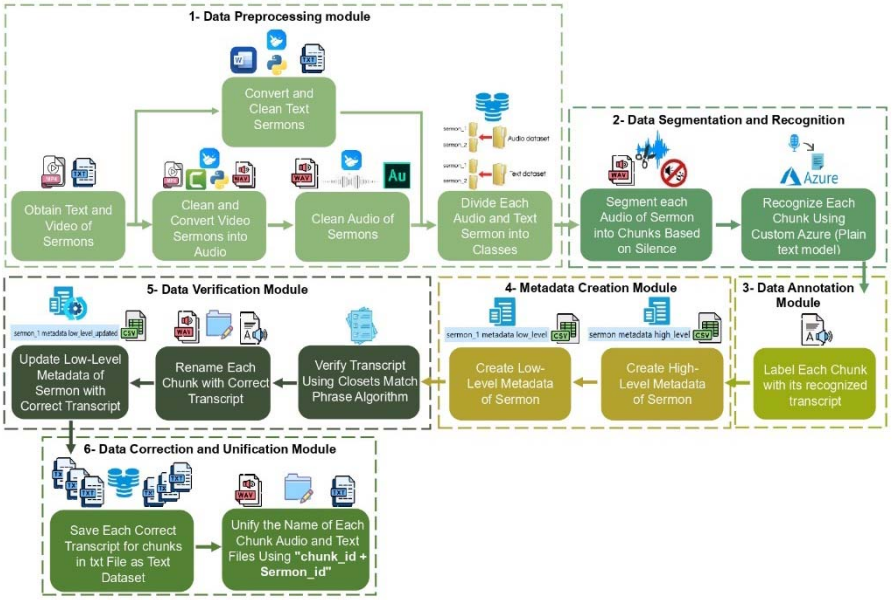


Figure 5 – Low-Level approach of collecting and creating SAT

– Cleaning and converting video into audio: as a next step, we start to clean each video using "Camtasia" (video editing software) by removing the Azan from the beginning and Prayer from the end of the Sermon and just saving the Sermon body (see the block diagram for the structure of Sermon in Fig. 1). Then, we convert the cleaned video into audio (wav format) using the Python Programming Language with the "moviepy" library. After that, we cleaned each audio from noise using "Adobe Audition" (Audio editing software). Also, we removed any stuttering, crying, and coughing…etc., that may be contained in the wav audio. As a result, the cleaned wav audio for each Sermon recorded has an average duration of 17 seconds and 29 minutes at 16 Hz sampling rate, 16-bit PCM (Pulse Code Modulation), and one number of channels.

On the other hand, the collected text Sermon converted from docx into txt format to be suitable for the next stages of annotation and validation. We normalize texts by removing diacritics from each txt file.

– Dividing each audio and text into classes: as a final step in the data preprocessing module, we divide each cleaned wav and txt file of Sermon into two mains separated folders for our dataset, where each folder obtains multiple classes of Sermons, as shown in Fig. 6.

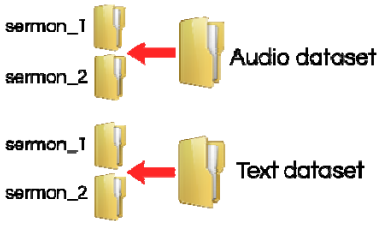


Figure 6 – SAT dataset folders

As represented in Table 4 and 5, each wav file of Fraidy Sermon was named as follows: Sermon id, Sermon date using Islamic calendar, and Preacher's name where the Preacher's name includes first, middle, and last name. For example: Sermon_23 (23-12-1443) Sheikh Usama Khayyat.

The txt file of Sermon was named as follows: Sermon title + _without_diacritics. For example: "Winning the bliss comes by following the straight path_without_diacritics".

Table 4 – An example of audio file name for one Sermon

Audio Dataset	
Folder Name	**File Name**
Sermon_23	Sermon_23 (23-12-1443) الشيخ أسامه خياط
	Sermon_23 (23-12-1443) Sheikh Usama Khayyat

Table 5 – An example of text file name for one Sermon

Text Dataset	
Folder Name	**File Name**
Sermon_23	الفوز بالنعيم باتباع الصراط المستقيم_without_diacritics
	Winning the bliss comes by following the straight path_without_diacritics"

2) Data segmentation and recognition module**:** this module includes two subprocesses as following:

– Segmentation based on silence: we utilized the "split_on_silence" function from Python in order to segment the audio of each Sermon into smaller chunks based on silence. By adjusting some parameters, like (silence threshold) from 40 to 50 and (minimum silence length) at least 200ms, we ensured that segments were properly identified and the segmentation was done well. In addition to that, (keep silence) 200ms of silence were added at the beginning and end of each segment to maintain the completeness of each chunk. The number of chunks for each audio Sermon is varying depending on the speech patterns and silence duration of each Preacher.

DOI 10.15588/1607-3274-2024-2-10

The generated chunks from one Sermon were named sequentially as chunk0, chunk1, chunk2, and so on.

– Recognition using Microsoft Azure: we utilized Microsoft Azure Customization in order to recognize each audio chunk of the Sermon. The plain text model, specifically the Speech Studio Custom Model, was employed because, in our case, the Sermon speech contains difficult words and is considered a special domain. Also, we imported the Speech SDK package from Azure Cognitive Services in Python to configure and build the custom model using our Azure portal's subscription key and endpoint. Training and customizing the model for Sermons' domain were used in the Speech Studio. This enabled us to obtain speech recognition for each chunk, which will be utilized in the next annotation module [28].

3) Data annotation module: the recognized speech in each audio chunk will be used for labeling each chunk with its relevant content that is recognized by Microsoft Speech Azure. All of these rec-ognized chunks of one Sermon are saved in one folder named "Sermon_id_recognizedChunks", where the Sermon_id could be Sermon_1, Sermon_2, and so on.

4) Metadata creation module: this module plays a crucial role in analyzing and organizing the Sermon dataset by creating two (CSV) files for both high-level and low-level metadata. This metadata provides descriptive information about Preacher and Sermon for better searchability, organization, classification, contextual understanding, and analysis of the dataset. In general, this module contains two subprocesses: creating high-level metadata and creating low-level metadata. This created Sermon metadata facilitates efficient management and utilization of our Sermon dataset.

The high-level metadata obtains general information about each Sermon and the person who introduced it (Preacher). Where this information based on 12 features as illustrates in following (Table 6).

In the low-level metadata of SAT, we present deeper information about each chunk's audio files of each Sermon where it captures more detailed information. This can include technical details, such as the following: (Sermon ID, Chunk ID, duration of each chunk (start and end time for each chunk), transcript of each chunk, total number of chunks, silence threshold, minimum silence length, and keep silence).

5) Data verification module**:** this module consists of three subprocess which are the following:

– Verifying transcript using Closets Matching Phrase algorithm: we create a similarity matching algorithm for finding the Closest Matching Phrase (CMP) between (Tran-script) and the (full text). Where the transcript is generated using Azure speech and saved on low-level metadata (CSV file). The way that the CPM algorithm is used to verify transcripts from any speech recognition engine is shown in Fig. 7.

Table 6 – SAT dataset's of high-level metadata features description

Features	**Variables Description**
Sermon ID	Each sermon has a unique ID, for example, sermon_1, sermon_2, sermon_3 ….
Name	Name of preacher who deliver the sermon
Academic Qualification	Educational background of preacher, such as أصول الفقه "The Principles of Jurisprudence" and so on
Experience	Number of years that being a preacher in introducing Fraidy Sermon
Title	Title of Sermon
Domain	Topic or theme that Sermon belongs
Date	Date of the Sermon that was introduced in (Arabic calendar -Hijri)
Language	Language of Sermon, it could be Arabic, English, and other language
Duration	Length of each Sermon starts from the beginning preacher's speech until the end, formatted as (hour: minutes: second)
Location	Place where the Sermon is delivered, for example, Holly Mosque in Mecca or any other mosque or religious center
Other language	Whether the Sermon is translated into other languages, formatted as (No = 0, Yes =1)
Arabic Sign Language	Whether the Sermon is interpreted or translated into Arabic Sign Language, formatted as (No = 0, Yes =1)

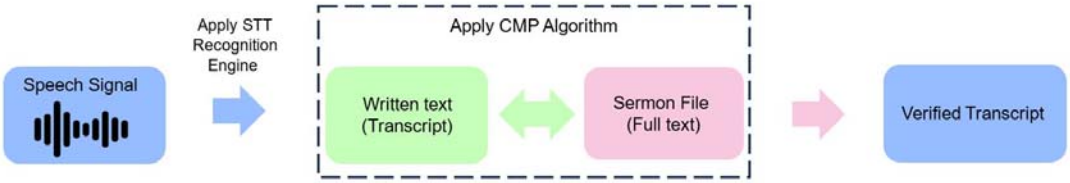


Figure 7 – Block diagram of transcript verification using CMP Algorithm

DOI 10.15588/1607-3274-2024-2-10

The algorithm allows customization through the optional parameters (window size) and (min ratio). Where window size determines the size of the sliding window used for comparison, and min ratio sets the minimum required matching ratio for a phrase to be considered a close match. After running the algorithm, the Closest Matching Phrases with their maximum ratio will be printed. This printed ratio helps us to identify phrases with low ratios that require correction.

Algorithm 1: Find Closest Matching Phrase (CMP) in Text

Require: T (string): The transcript to match
FT (string): The full text to search within
WIND (integer, optional, default=5): Number of words around the transcript to consider during matching
minR (float, optional, default=0.70): Minimum similarity ratio to consider a match
Ensure: CMP (string): Closest matching phrase within the full text
maxR (float): Similarity ratio of the closest matching phrase

```
Procedure Find_CMP (T, FT, W=5, minR=0.70)
    TW ⟵ split T into words
    FTW ⟵ split FT into words
    Initialize tracking variables
    maxR ⟵ 0
    CMP ⟵ NULL
     Iterate through words in the FT
    for I = 0 to (length of FTW – length of TW + 1) do
      Expand the window around the current position
      for j = –WIND to WIND do
        Extract words from the current window
        end_index ⟵ i + length of TW + j – 1
        current_phrase_words ⟵ subarray of FTW from index i to end_index
        current_phrase ⟵ join current_phrase_words into a string
        Calculate similarity ratio
        R ⟵ calculate similarity between T and current_phrase using SequenceMatcher (ratio)
          Update tracking variables if R is greater than maxR
          if R > maxR then
            maxR ⟵ R
            if maxR >= minR then
              CMP ⟵ current_phrase
          end if
        end if
    end for
    end for
  return the CMP and its similarity R
  return CMP, maxR
end procedure
```

– Renaming each chunk with the correct transcript: based on using the developed algorithm, the updated transcript will be automatically based on the calculated similarity between T (transcript) and current_phrase using SequenceMatcher (ratio).

If the updated max ratio is more than the minimum ratio = 0.70, a similar transcript from the full-text file will be printed as a verified transcript otherwise the current transcript will be printed where it needs to be checked manually.

If the updated max ratio is less than the minimum ratio = 0.70, we take a look at the audio chunk associated with each low-ratio phrase listened to, and a comparison is made with the transcript in the CSV file and the full text in the text file. Once the correct transcript is determined, we start to replace the transcript that needs to be updated manually based on manual records for a transcript that has a low ratio.

In some cases of the lower ratio, if the recognized transcript cannot be found in the full-text file, it means that the Preacher used a new phrase that has not been written in the full text.

– Updating low-level metadata with correct transcript: we save the result of the transcript correction based on using similarity matching in the updated CSV file for low-level metadata.

6) Data correction and unification module: in this final module, we process the following:

– Saving the correct transcript in a txt file: from the corrected CSV file that contains the transcript of each chunk we save these transcripts as separate files (txt format) for each chunk.

– Unifying the files' names in two datasets (text and audio): in order to unify the file names in the two dataset types (text and audio) to easily follow the audio file with a corresponding text file, we named each generated text file as following structure (chunk_id + Sermon_id.txt). On the other hand, we named each audio file as (as fol-

DOI 10.15588/1607-3274-2024-2-10

lowing structure (chunk_id + Sermon_id.wav). After that, each text file will be saved in the text dataset while each audio file will be saved in the audio dataset, as shown in the organized files of the SAT dataset in Fig. 8.

Figure 8 – SAT dataset with equivalent files

4 EXPERIMENTS

In this section, the outcomes of the Measurement and Deployment phases will be presented for the SAT dataset. The objectives of these last two phases were to evaluate the accuracy of our dataset, followed by deploying it strategically in order to ensure its accessibility and applicability in real-world scenarios.

3.3 Measurement Phase: to measure the accuracy of the created dataset SAT we implement three main stages, A) Prepare and transform the dataset B) Train with a customized model. C) Test and evaluate a model of the dataset.

In the preparing and transforming, in order to apply the model, we prepared SAT by adding all txt files (transcripts) in one txt file where each line contains one single audio information (audio file name with extension (.wav) followed by space then transcript). More importantly, the text should be normalized (does not include punctuations or diacritics). Also, the text encoding must be UTF-8 BOM (txt format) While the audio should be in WAV format with a sample rate of 8,000 or 16,000 Hz. The maximum length for each audio file is 2 hours for testing and 60 seconds for training. The audio files and the transcript file should be grouped in a zip file with a maximum size of 2 GB.

In the training, we used a custom model speech recognition Speech to Text (STT) provided by Microsoft Azure for training in order to improve recognition accuracy. In our case, we used 10136 audio files with a sample rate of 16,000 Hz and around 14 hours for 50 sermons.

We trained using a custom speech recognition model with 30 percent of the weight and during training; the labeled text was normalized to increase the readability.

In the testing and evaluation, we tested 8526 audio files within around 5 hours and a half. Overall, after applying our new algorithm CMP to enhance speech recognition to achieve high confidence in equivalent speech-to-text. Our customized model (fine-tuned with the SAT dataset) achieved a 5.13% Word Error Rate (WER), which indicates that our created dataset SAT with speech recognition model performed better than the base model. Our custom model performance was compared to the base model performance, the result is presented in Table 7.

Table 7 – Comparison of WER scores for Azure Custom Model (fine-tuned with our SAT dataset) and Microsoft Azure Speech Model

Model	Dev	Test	WER	Insertion	Substitution	Deletion
Customized Speech (Our SAT dataset)	**54.29**	**45.71**	**5.13**	**0.39**	**3.49**	**1.24**
Microsoft Azure Speech	54.29	45.71	15.61	2.20	11.80	1.60

3.4 Deployment Phase: This phase is considered the last phase in the dataset collection and creation. It starts by publishing our SAT through one of the popular platforms, such as GitHub, Mendeley Data …etc, which helps other researchers to refine and deploy any other method and model depending on the needs. We will publish the documentation about SAT metadata to support users and engage them in future research for more under-standing of this dataset. As a result of this engagement, the collected and created SAT will be maintained and updated.

5 RESULTS

By following the four phases (Fig. 2) with its stages to collect and create our dataset SAT, we successfully created a dataset that included 21,253 WAV audio files and corresponding 21,253 TXT transcript files of 50 Friday sermons.

This SAT dataset can be used for the exploration and analysis of Sermon content, delivery, and various linguistic aspects by specialists and other researchers. Table 8 summarizes the information about SAT.

Table 8 – Summarizing of SAT dataset Information

	Audio	Text
Number of Sermons	50	50
Number of Files	21253	21253
Format	WAV	TXT
Total Size	1.56 GB	1.24 MB
Total Duration of All Sermons	14h 34m 31s	
Average Duration for Each Sermon	17m 29s	
Total Words	83141	
Distinct Words	25226	
Total Number of Preachers	9	

For more details about the various durations for all50 Friday sermons of Masjid al-Haram illustrates in Fig. 9.

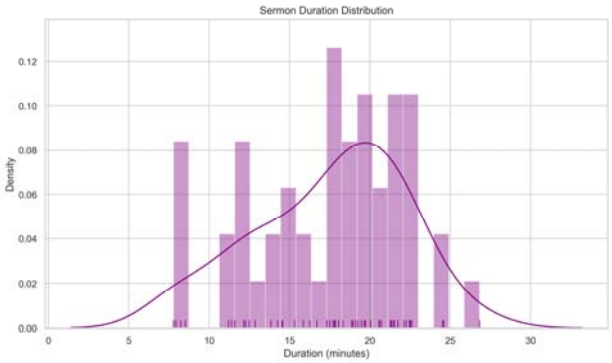


Figure 9 – Duration distribution of 50 Friday Sermons

DOI 10.15588/1607-3274-2024-2-10

Moreover, Fig.10 presents the duration of Sermon by domain where we have 4 main domains as the previous researchers specify [14]. The Invitation to Islam or Its Defending has a high duration of around 27 minutes whereas the other 3 domains of Religious Education for the Public, Proving Faith in the Souls, and Correction of Faults and Prohibition of Evils obtain around 18 minutes.

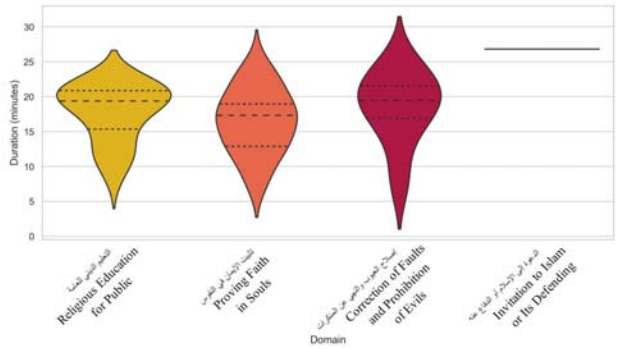


Figure 10 – Duration distribution of 50 Friday Sermons by domain

In terms of Automatic Speech Recognition (ASR), our SAT dataset can be utilized for speech recognition systems because of the potential variation in terminology, speaking styles, and SAT dataset) will be suitable for this task.

6 DISCUSSION

The successful creation of 21,253 WAV audio files and the equivalent number of TXT (transcripts) for the 50 Friday Sermons dataset marks a significant advancement in the religious discourse domain. This SAT dataset not only facilitates a comprehensive exploration of sermon content but also can be used as a valuable resource. Moreover, the diversity of our dataset and its metadata availability has the potential to support nuanced research on language use in religious settings. It can be used in discourse analysis and MT, for instance, from Arabic speech into another language, or into other Sign Languages. Despite its strengths, the dataset's scope which is limited to sermons from the Grand Mosque (Masjid al-Haram) in Makkah, suggests a direction-expanding dataset in the future direction using our methods to include a broader of sources, which seeks to be AI-driven in religious contexts.

CONCLUSIONS

In this research, we have completed four main phases, including planning, creating and processing, measuring, and deploying phases. We have curated and prepared a part of the Sermon audio and text dataset (SAT), providing valuable resources for future research and implementation in various sectors. Moreover, our created dataset SAT achieved less WER in fine-tuned using the Azure custom model compared with the Azur baseline model by 10.48 %. **Moreover, this research developed a CMP algorithm** for enhancing the customized Azure speech recognition to verify our SAT by correcting phrases that have a lower ratio, which leads to reducing the WER.

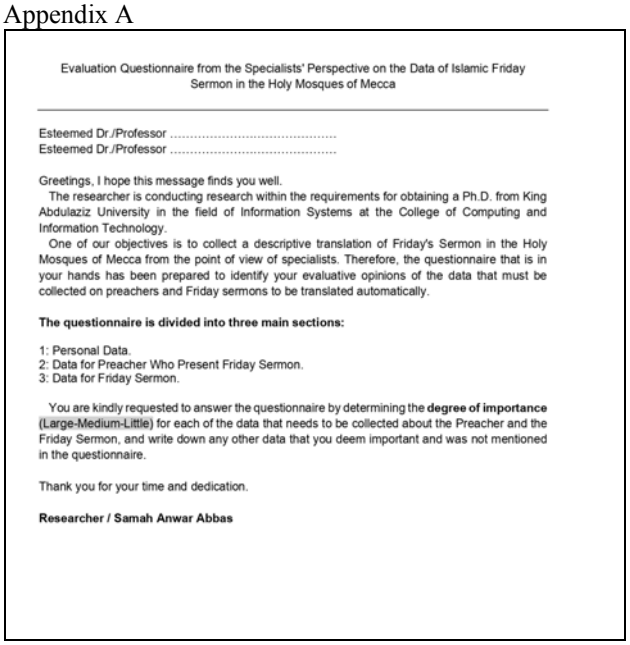
In future work, we will expand our SAT with ArSL to achieve a variety of datasets. Also, we are going to explore advanced alignment techniques and algorithms to improve the accuracy and efficiency of the ArSL video of the Sermon to support advanced machine translation techniques.

ACKNOWLEDGEMENTS

The authors extend gratitude to the Presidency of Religious Affairs at the Grand Mosque and the Prophet's Mosque for facilitating access to essential data sources, which significantly enriched this study.

APPENDICES

Appendix A


Evaluation Questionnaire from the Specialists' Perspective on the Data of Islamic Friday Sermon in the Holy Mosques of Mecca

Esteemed Dr./Professor .....
Esteemed Dr./Professor .....

Greetings, I hope this message finds you well.
The researcher is conducting research within the requirements for obtaining a Ph.D. from King Abdulaziz University in the field of Information Systems at the College of Computing and Information Technology.
One of our objectives is to collect a descriptive translation of Friday's Sermon in the Holy Mosques of Mecca from the point of view of specialists. Therefore, the questionnaire that is in your hands has been prepared to identify your evaluative opinions of the data that must be collected on preachers and Friday sermons to be translated automatically.

The questionnaire is divided into three main sections:

1: Personal Data.
2: Data for Preacher Who Present Friday Sermon.
3: Data for Friday Sermon.

You are kindly requested to answer the questionnaire by determining the **degree of importance** (Large-Medium-Little) for each of the data that needs to be collected about the Preacher and the Friday Sermon, and write down any other data that you deem important and was not mentioned in the questionnaire.

Thank you for your time and dedication.

Researcher / Samah Anwar Abbas

Figure A1: Evaluation Questionnaire from the Specialists' Perspective on the Data of Islamic Friday Sermon in the Holy Mosques of Mecca: (the purpose of investigation)

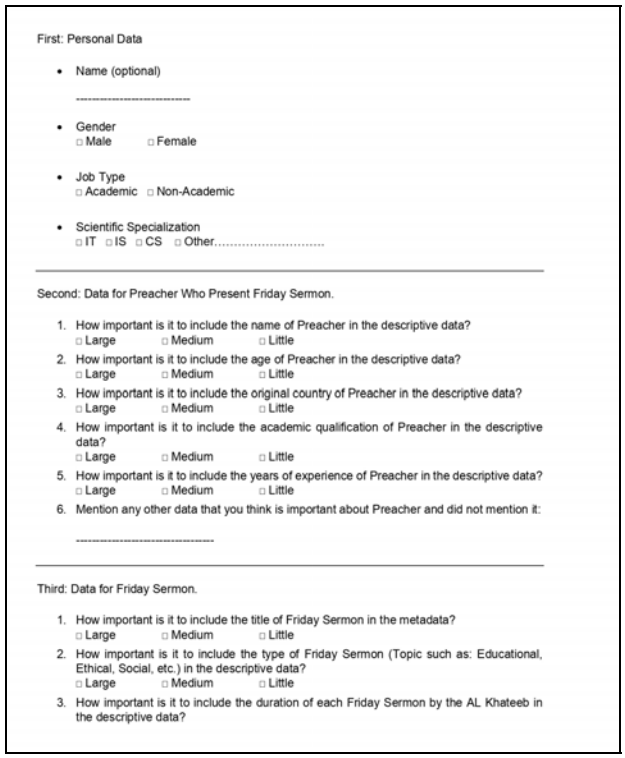

Consent for Participation in this Research:

I, the undersigned participant of this study, have been made aware of the research's objectives, as well as its potential advantages and risks. I comprehend that my involvement in this research is voluntary and I have the right to withdraw at any moment without having to justify my decision. I hereby agree to partake in this study.

□ I agree to participate in this study.
□ I do not agree to participate in this study.

Figure A2: Evaluation Questionnaire from the Specialists' Perspective on the Data of Islamic Friday Sermon in the Holy Mosques of Mecca: (the consent for participation)

DOI 10.15588/1607-3274-2024-2-10



First: Personal Data

- Name (optional)

- Gender
  □ Male □ Female
- Job Type
  □ Academic □ Non-Academic
- Scientific Specialization
  □ IT □ IS □ CS □ Other.....

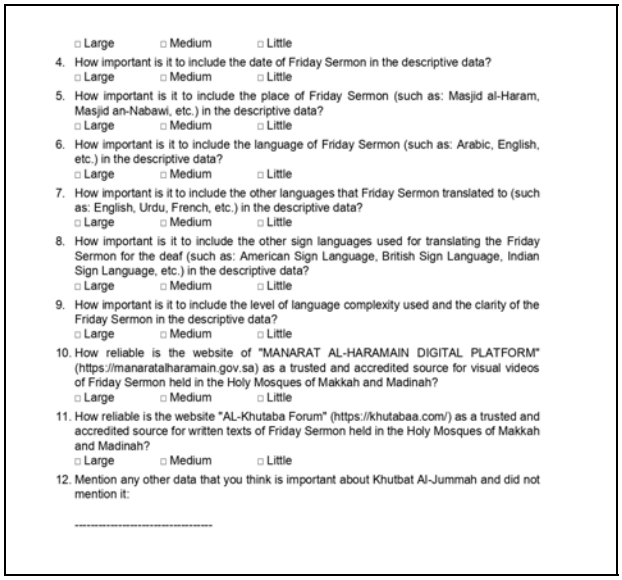
Second: Data for Preacher Who Present Friday Sermon.

1. How important is it to include the name of Preacher in the descriptive data?
   □ Large □ Medium □ Little
2. How important is it to include the age of Preacher in the descriptive data?
   □ Large □ Medium □ Little
3. How important is it to include the original country of Preacher in the descriptive data?
   □ Large □ Medium □ Little
4. How important is it to include the academic qualification of Preacher in the descriptive data?
   □ Large □ Medium □ Little
5. How important is it to include the years of experience of Preacher in the descriptive data?
   □ Large □ Medium □ Little
6. Mention any other data that you think is important about Preacher and did not mention it:

Third: Data for Friday Sermon.

1. How important is it to include the title of Friday Sermon in the metadata?
   □ Large □ Medium □ Little
2. How important is it to include the type of Friday Sermon (Topic such as: Educational, Ethical, Social, etc.) in the descriptive data?
   □ Large □ Medium □ Little
3. How important is it to include the duration of each Friday Sermon by the AL Khateeb in the descriptive data?

Figure A3: Evaluation Questionnaire from the Specialists' Perspective on the Data of Islamic Friday Sermon in the Holy Mosques of Mecca: (presenting the Three axes including questions about the level of importance of Preacher and Friday Sermon parameters)



□ Large □ Medium □ Little

4. How important is it to include the date of Friday Sermon in the descriptive data?
   □ Large □ Medium □ Little
5. How important is it to include the place of Friday Sermon (such as: Masjid al-Haram, Masjid an-Nabawi, etc.) in the descriptive data?
   □ Large □ Medium □ Little
6. How important is it to include the language of Friday Sermon (such as: Arabic, English, etc.) in the descriptive data?
   □ Large □ Medium □ Little
7. How important is it to include the other languages that Friday Sermon translated to (such as: English, Urdu, French, etc.) in the descriptive data?
   □ Large □ Medium □ Little
8. How important is it to include the other sign languages used for translating the Friday Sermon for the deaf (such as: American Sign Language, British Sign Language, Indian Sign Language, etc.) in the descriptive data?
   □ Large □ Medium □ Little
9. How important is it to include the level of language complexity used and the clarity of the Friday Sermon in the descriptive data?
   □ Large □ Medium □ Little
10. How reliable is the website of "MANARAT AL-HARAMAIN DIGITAL PLATFORM" (https://manaratalharamain.gov.sa) as a trusted and accredited source for visual videos of Friday Sermon held in the Holy Mosques of Makkah and Madinah?
    □ Large □ Medium □ Little
11. How reliable is the website "AL-Khutaba Forum" (https://khutabaa.com/) as a trusted and accredited source for written texts of Friday Sermon held in the Holy Mosques of Makkah and Madinah?
    □ Large □ Medium □ Little
12. Mention any other data that you think is important about Khutbat Al-Jummah and did not mention it:

Figure A4: Evaluation Questionnaire from the Specialists' Perspective on the Data of Islamic Friday Sermon in the Holy Mosques of Mecca: (presenting the Three axes including questions about the level of importance of Preacher and Friday Sermon parameters) cont

REFERENCES

1. Chen H., Xie W., Vedaldi A., Zisserman A. VGGSound: A Large-Scale Audio-Visual Dataset, 2020.
2. Cohn N., Cardoso B., Klomberg B., Hacımusaoğlu I. The Visual Language Research Corpus (VLRC), *An Annotated Corpus of Comics from Asia, Europe, and the United States.* Lang Resources & Evaluation, 2023, DOI:10.1007/s10579-023-09673-0.
3. Mounsef J., Hasib M., Raza A. Building an Arabic Dialectal Diagnostic Dataset for Healthcare, *IJACSA,* 2022, No.13, DOI:10.14569/IJACSA.2022.01307100.
4. Alfraidi T., Abdeen M. A. R., Yatimi A., Alluhaibi R., Al-Thubaity A. The Saudi Novel Corpus, *Design and Compilation. Applied Sciences,* 2022, No. 12, P. 6648, DOI:10.3390/app12136648.
5. Abdelhay M., Mohammed A., Hefny H.A. Deep Learning for Arabic Healthcare, *MedicalBot. Soc. Netw. Anal. Min.* 2023, No. 13, P. 71. DOI:10.1007/s13278-023-01077-w.
6. Abbas S., Al-Barhamtoshy H., Alotaibi F. Towards an Arabic Sign Language (ArSL) Corpus for Deaf Drivers. *PeerJ Comput. Sci.,* 2021, No. 7, e741, DOI:10.7717/peerj-cs.741.
7. Asyafie M. A., Harun M., Shapiai M. I., Khalid P. I. Identification of Phoneme and Its Distribution of Malay Language Derived from Friday Sermon Transcripts. In Proceedings of the 2014 IEEE Student Conference on Research and Development, 2014, December, pp. 1–6.
8. Saddhono K., Rakhmawati A. Sociolinguistic Studies of Friday Sermon Using Javanese as an Effort to Preserves Indigenous Language in Java Island, *In Proceedings of the 2nd International Conference on Sociology Education,* SCITEPRESS – Science and Technology Publications. Bandung, Indonesia, 2017, pp. 829–833.
9. Alkhawaldeh A. A. Deixis in English Islamic Friday Sermons, *A Pragma-Discourse Analysis. Studies in English Language and Education,* 2022, No. 9, pp. 418–437, DOI:10.24815/siele.v9i1.21415.
10. Aksoy O. Preaching to Social Media: Turkey's Friday Khutbas and Their Effects on Twitter, SocArXiv, 2021, May, No. 12.
11. Usman A. H., Iskandar A. Analysis of Friday Sermon Duration, *Intellectual Reflection of Classical and Contemporary Islamic Scholars. Journal of Religious & Theological Information,* 2022, No. 21, pp. 68–81, doi:10.1080/10477845.2021.1928349.
12. Gürlesin Ö. F. Understanding the Political and Religious Implications of Turkish Civil Religion in The Netherlands: A Critical Discourse Analysis of ISN Friday Sermons. *Religions,* 2023, No. 14, P. 990, DOI:10.3390/rel14080990.
13. Nor M.R.M. Multicultural Discourse from the Minbar: A Study on Khutbah Texts Prepared by Jakim Malaysia. In; Fukami N., Sato S., Eds.; *JSPS Asia and Africa Science Platform Program.* Organization of Islamic Area Studies, Waseda University. Tokyo, Japan, 2012, pp. 55–62 ISBN 978-4-904039-52-6.
14. Ismail Ali Mohammed Art of Oratory and Skills of Orator. Researches in the preparation of preacher preacher. Fifth edition, Dar Alkalema, Cairo-Egypt, 2016.
15. Mahmood I., Kasim Z. Metadiscourse Resources across Themes of Islamic Friday Sermon, 2021, No. 21, pp. 45–61, DOI: 10.17576/gema-2021-2101-01-03.
16. Sukarno S., Salikin H. The The Generic Structure Potential of Friday Sermons in Jember, *International Journal of Linguistics and Translation Studies.* Indonesia, 2022, No. 3, pp. 56–73. DOI:10.36892/ijlts.v3i1.207.
17. Mohammed Saleh Al-Hamzi A., Sumarlam, Santosa R., Jamal M. A Pragmatic and Discourse Study of Common Deixis Used by Yemeni-Arab Preachers in Friday Islamic Sermons at Yemeni Mosques, *Cogent Arts & Humanities*

DOI 10.15588/1607-3274-2024-2-10

2023, No. 10, P. 2177241, doi:10.1080/23311983.2023.2177241.
18. Mahmood I., Kasim Z. Interpersonal Metadiscursive Features in Contemporary Islamic Friday Sermon, *3L: Language, Linguistics, Literature,* 2019, No. 25, pp. 85–99. DOI:10.17576/3L-2019-2501-06.
19. Wardoyo C. Directive Speech Acts Performed in Khutbah (Islamic Friday Sermon), 2017.
20. Fahruroji F., Rakhmat M., Shodiq M. The Understanding of Friday Prayer Attendees (Mustamik) Towards Friday Sermon Discourse, 2017, P. 779.
21. Carol S., Hofheinz L. A Content Analysis of the Friday Sermons of the Turkish-Islamic Union for Religious Affairs in Germany (DİTİB), *Politics and Religion,* 2022, DOI:10.1017/S1755048321000353.
22. Jafilus M., Asha'ari M. F., Rasit R. Thematic Analysis of the Content of the Friday Sermon in Negeri Sembilan, IJARBSS, 2021, No. 11, pp. 84–98, DOI:10.6007/IJARBSS/v11-i6/10087.
23. Numeracy, Maths and Statistics – Academic Skills Kit Available online: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/strength-of-correlation.html (accessed on 19 February 2024).
24. Moslem S., Ghorbanzadeh O., Blaschke T., Duleba S. Analysing Stakeholder Consensus for a Sustainable Transport Development Decision by the Fuzzy AHP and Interval AHP, *Sustainability,* 2019, No. 11, P. 3271, DOI:10.3390/su11123271.
25. Encyclopedia of Statistics in Behavioral Science; Everitt B., Howell D. C., Eds. John Wiley & Sons. Hoboken, N. J, 2005, ISBN 978-0-470-86080-9.
26. MNARAT AL-HARAMAIN Available online: https://manaratalharamain.gov.sa/home (accessed on 25 June 2023).
27. Khutaba Forum Available online: https://khutabaa.com/en (accessed on 25 June 2023).
28. Beatman A. Improve Speech-to-Text Accuracy with Azure Custom Speech | Azure Blog | Microsoft Azure Available online: https://azure.microsoft.com/en-us/blog/improve-speechtotext-accuracy-with-azure-custom-speech/ (accessed on 23 September 2023).

Received 04.03.2024.
Accepted 26.04.2024.

УДК 004.942(045)

СТВОРЕННЯ МАСШТАБОВАНОГО НАБОРУ ДАНИХ ДЛЯ П'ЯТНИЧНИХ ПРОПОВІДЕЙ З АУДІО ТА ТЕКСТУ (ПАТ)

Самах А. А. – д-р філософії кафедри інформаційних систем факультету обчислювальної техніки та інформаційних технологій та викладач кафедри інформаційних систем управління факультету економіки та адміністрації Університету короля Абдула Азіза, Джидда, Мекка, Саудівська Аравія.

Дімах Х. А. – д-р філософії, доцент, доцент кафедри інформаційних систем, факультет обчислювальної техніки та інформаційних технологій, Університет короля Абдула Азіза, Джидда, Мекка, Саудівська Аравія.

Хасанін М. А. – д-р техн. наук, професор, професор кафедри інформаційних технологій, факультет обчислювальної техніки та інформаційних технологій, Університет короля Абдула Азіза, Джидда, Мекка, Саудівська Аравія.

АНОТАЦІЯ

Актуальність. Сьогодні збір і створення наборів даних у різних секторах стає все більш поширеним. Незважаючи на таке поширене створення даних, досі існує прогалина в спеціалізованих областях, зокрема в області Ісламських п'ятничних проповідей. Вона багата на теологічні, культурні та лінгвістичні дослідження, які стосуються арабських і мусульманських країн, а не лише релігійних дискурсів.

Мета. Мета цього дослідження полягає в тому, щоб усунути цю нестачу, створивши повний набір даних аудіо та тексту проповідей із його метаданими. Це спрямоване надати великий ресурс для вивчення релігії, лінгвістики та соціології. Крім того, це дозволить підтримати досягнення у сфері штучного інтелекту, таких як технології обробки природної мови та розпізнавання мовлення.

Метод. Розробка набору даних проходила у чотири окремі етапи: планування, створення та обробка, вимірювання та розгортання. Набір даних містить колекцію з 21 253 аудіо та відповідних файлів розшифровки, які були успішно створені. Удосконалені методи обробки звуку були використані для покращення розпізнавання мовлення та надання набору даних, який підходить для широкого використання.

Результати. Тонко налаштований набір даних досяг 5,13% частоти помилок у словах (Word Error Rate – WER), що вказує на значне покращення точності, порівняно з базовою моделлю Microsoft Azure Speech. Це досягнення вказує на якість набору даних і ефективність використовуваних методів обробки. У світлі цього було розроблено новий алгоритм фрази з найбільшою відповідністю, щоб підвищити високу надійність еквівалентного мовлення до тексту шляхом коригування фраз із меншим співвідношенням.

Висновки. Це дослідження створює ресурс для поєднання різних досліджень, таких як релігієзнавство, лінгвістика та соціологія. Крім того, воно демонструє потенціал у сфері штучного інтелекту і підтримує його програми. У майбутніх дослідженнях ми зосередимося на збагаченні цього розширення набору даних шляхом додавання відеокорпусу мовою жестів, використовуючи вдосконалені методи вирівнювання. Він підтримуватиме поточні розробки машинного перекладу для ширшого розуміння ісламських п'ятничних проповідей у різних мовах і культурах.

КЛЮЧОВІ СЛОВА: п'ятничні проповіді, хутба, розпізнавання арабської мови, набір звукових і текстових даних, машинний переклад.

DOI 10.15588/1607-3274-2024-2-10

ЛІТЕРАТУРА

1. VGGSound / [H. Chen, W. Xie, A. Vedaldi, A. Zisserman]. –A Large-Scale Audio-Visual Dataset, 2020.
2. The Visual Language Research Corpus (VLRC) / [N. Cohn, B. Cardoso, B. Klomberg, I. Hacımusaoğlu] // An Annotated Corpus of Comics from Asia, Europe, and the United States. – Lang Resources & Evaluation. – 2023, DOI: 10.1007/s10579-023-09673-0.
3. Mounsef J. Building an Arabic Dialectal Diagnostic Dataset for Healthcare / J. Mounsef, M. Hasib, A. Raza // IJACSA. – 2022. – No. 13. DOI:10.14569/IJACSA.2022.01307100.
4. The Saudi Novel Corpus: Design and Compilation / [T. Alfraidi, M.A.R.Abdeen, A. Yatimi et al.] // Applied Sciences. – 2022. – No. 12. – P. 6648. DOI: 10.3390/app12136648.
5. Abdelhay M. Deep Learning for Arabic Healthcare / M. Abdelhay, A. Mohammed, H. A. Hefny // MedicalBot. Soc. Netw. Anal. Min. – 2023. – No. 13. – P. 71. DOI: 10.1007/s13278-023-01077-w.
6. Abbas S. Towards an Arabic Sign Language (ArSL) Corpus for Deaf Drivers / S. Abbas, H. Al-Barhamtoshy, F. Alotaibi // PeerJ Comput. Sci. – 2021. – No. 7. – e741, DOI:10.7717/peerj-cs.741.
7. Identification of Phoneme and Its Distribution of Malay Language Derived from Friday Sermon Transcripts / [M. A. Asyafie, M. Harun, M. I. Shapiai, P. I. Khalid] // In Proceedings of the 2014 IEEE Student Conference on Research and Development. –2014. – December. – P. 1–6.
8. Saddhono K. Sociolinguistic Studies of Friday Sermon Using Javanese as an Effort to Preserves Indigenous Language in Java Island / K. Saddhono, A. Rakhmawati // In Proceedings of the 2nd International Conference on Sociology Education. – SCITEPRESS – Science and Technology Publications : Bandung, Indonesia, 2017. – P. 829–833.
9. Alkhawaldeh, A.A. Deixis in English Islamic Friday Sermons: A Pragma-Discourse Analysis / A. A. Alkhawaldeh // Studies in English Language and Education. – 2022. – No. 9. – P. 418–437. DOI:10.24815/siele.v9i1.21415.
10. Aksoy O. Preaching to Social Media: Turkey's Friday Khutbas and Their Effects on Twitter / O. Aksoy // SocArXiv. – 2021. – May. – No. 12.
11. Usman, A.H.; Iskandar, A. Analysis of Friday Sermon Duration: Intellectual Reflection of Classical and Contemporary Islamic Scholars / A. H. Usman, A. Iskandar // Journal of Religious & Theological Information. – 2022. – No. 21. – P. 68–81. DOI:10.1080/10477845.2021.1928349.
12. Gürlesin Ö. F. Understanding the Political and Religious Implications of Turkish Civil Religion in The Netherlands: A Critical Discourse Analysis of ISN Friday Sermons / Ö. F. Gürlesin // Religions. – 2023. – No. 14. – P. 990, DOI:10.3390/rel14080990.
13. Nor, M.R.M. Multicultural Discourse from the Minbar: A Study on Khutbah Texts Prepared by Jakim Malaysia / M.R.M. Nor, In N. Fukami, S. Sato, Eds. // JSPS Asia and Africa Science Platform Program. – Organization of Islamic Area Studies, Waseda University : Tokyo. – Japan, 2012. – P. 55–62. ISBN 978-4-904039-52-6.
14. Ismail Ali Mohammed Art of Oratory and Skills of Orator: Researches in the preparation of preacher preacher / Ismail Ali Mohammed. – Fifth edition, Dar Alkalema : Cairo-Egypt, 2016.
15. Mahmood I. Metadiscourse Resources across Themes of Islamic Friday Sermon / I. Mahmood, Z. Kasim. – 2021. – No. 21. – P. 45–61. DOI:10.17576/gema-2021-2101-01-03.
16. Sukarno S. The The Generic Structure Potential of Friday Sermons in Jember / S. Sukarno, H. Salikin // International Journal of Linguistics and Translation Studies. – Indonesia. – 2022. – No. 3. – P. 56–73. DOI:10.36892/ijlts.v3i1.207.
17. A Pragmatic and Discourse Study of Common Deixis Used by Yemeni-Arab Preachers in Friday Islamic Sermons at Yemeni Mosques / [A. Mohammed Saleh Al-Hamzi, Sumarlam, R. Santosa, M. Jamal] // Cogent Arts & Humanities. – 2023. – No. 10. – P. 2177241, DOI:10.1080/23311983.2023.2177241.
18. Mahmood I. Interpersonal Metadiscursive Features in Contemporary Islamic Friday Sermon / I. Mahmood, Z. Kasim // 3L: Language, Linguistics, Literature. – 2019. – No. 25. – P. 85–99. DOI:10.17576/3L-2019-2501-06.
19. Wardoyo C. Directive Speech Acts Performed in Khutbah (Islamic Friday Sermon) / C. Wardoyo. – 2017.
20. Fahruroji F. The Understanding of Friday Prayer Attendees (Mustamik) Towards Friday Sermon Discourse / F. Fahruroji, M. Rakhmat, M. Shodiq, 2017. – P. 779.
21. Carol S. A Content Analysis of the Friday Sermons of the Turkish-Islamic Union for Religious Affairs in Germany (DİTİB) / S. Carol, L. Hofheinz // Politics and Religion. – 2022. DOI:10.1017/S1755048321000353.
22. Jafilus M. Thematic Analysis of the Content of the Friday Sermon in Negeri Sembilan / M. Jafilus, M. F. Asha'ari, R. Rasit // IJARBSS. – 2021. – No. 11. – P. 84–98. DOI:10.6007/IJARBSS/v11-i6/10087.
23. Numeracy, Maths and Statistics – Academic Skills Kit Available online: https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/strength-of-correlation.html (accessed on 19 February 2024).
24. Stakeholder Consensus for a Sustainable Transport Development Decision by the Fuzzy AHP and Interval AHP / [S. Moslem, O. Ghorbanzadeh, T. Blaschke, S. Duleba] // Sustainability. – 2019. – No. 11. – P. 3271. DOI:10.3390/su11123271.
25. Encyclopedia of Statistics in Behavioral Science / B. Everitt, D. C. Howell, Eds. – John Wiley & Sons : Hoboken, N.J, 2005. ISBN 978-0-470-86080-9.
26. MNARAT AL-HARAMAIN Available online: https://manaratalharamain.gov.sa/home (accessed on 25 June 2023).
27. Khutaba Forum Available online: https://khutabaa.com/en (accessed on 25 June 2023).
28. BeatmanA. Improve Speech-to-Text Accuracy with Azure Custom Speech | Azure Blog | Microsoft Azure Available online: https://azure.microsoft.com/en-us/blog/improve-speechtotext-accuracy-with-azure-custom-speech/ (accessed on 23 September 2023).

DOI 10.15588/1607-3274-2024-2-10