Overview of the Third Shared Task on Indian Language Summarization (ILSUM 2024)

Shrey Satapara[1], Parth Mehta[2], Sandip Modha[3], Asha Hegde[4], H. L. Shashirekha[4] and Debasis Ganguly[5]

[1]*Indian Institute of Technology Hyderabad, India*
[2]*Parmonic, USA*
[3]*LDRP-ITR, Gandhinagar, India*
[4]*Mangalore University, India*
[5]*University of Glasgow, Scotland, UK*

Abstract

This overview paper presents a synopsis of the third edition of the shared task on Indian Language Summarization (ILSUM 2024) [1] organized as a part of the 16th edition of Forum for Information Retrieval Evaluation (FIRE) 2024 [2]. In this edition of ILSUM, we continue the text summarization task from ILSUM 2022 [3, 4] and Task 1 of ILSUM 2023 [5, 6]. In this edition, we introduce three Dravidian languages Kannada, Tamil and Telugu in addition to all the languages from previous editions. We also expanded both train and test datasets for Hindi and English, and the test datasets for Gujarati and Bengali. Further, we build upon the misinformation detection subtask from ILSUM 2023 and offer a cross-lingual misinformation detection task in Hindi and Gujarati languages. We used the same evaluation metrics as ILSUM 2023. Standard ROUGE metrics and BertScore were used for the summarization subtask, while the macro-F1 score was used for the cross-lingual misinformation detection subtask. ILSUM 2024 received registrations from over 50 teams. A total of 20 teams submitted runs across both subtasks and 12 teams submitted working notes.

Keywords

Automatic Text Summarization, Headline Generation, Misinformation Detection, Indian Languages, Indo-Aryan Languages, Dravidian Languages

1. Introduction

In the past decade, platforms such as the FIRE [2] and AI4BHARAT[1] have played a critical role in creating large-scale reusable corpora for advancing language technology research in low resource languages. While AI4Bharat exclusively works with Indian Languages, FIRE has traditionally included languages from other linguistic families as well, like Persian, Arabic, etc. With such initiatives several open benchmark datasets have been made available for problems like hate speech detection [7, 8, 9, 10, 11], sentiment analysis [12, 13, 14], machine translation [15], mixed script IR [16, 17], Indian legal document retrieval and summarization [18, 19, 20, 21, 22, 23], fake news detection [24, 25, 26], authorship attribution [27, 28], IR from microblogs [29], IR for software engineering [30, 31] among others.

The current shared task on ILSUM, which has been organized as a part of FIRE since 2022, is aimed at replicating this for the summarization and misinformation detection tasks. We build upon the past successful editions and introduce new languages (Kannda, Tamil and Telugu) as well as expanded datasets for other languages (Gujarati, Hindi, Bengali and English). We also offered a new sub-task on misinformation detection in machine-generated cross-lingual summaries. The misinformation detection subtask was offered in Gujarati and Hindi. In this overview paper, we present a detailed analysis of the dataset, evaluation metrics, approaches used by the participating teams and the results.

Forum for Information Retrieval Evaluation, December 12-15, 2024, India
shreysatapara@gmail.com (S. Satapara); parth.mehta126@gmail.com (P. Mehta); sjmodha@gmail.com (S. Modha); hegdekasha@gmail.com (A. Hegde); hlsrekha@mangaloreuniversity.ac.in (H. L. Shashirekha); debforit@gmail.com (D. Ganguly)
0000-0001-6222-1288 (S. Satapara); 0000-0002-4509-1298 (P. Mehta); 0000-0003-2427-2433 (S. Modha); 0000-0003-0050-7138 (D. Ganguly)

[1]https://ai4bharat.iitm.ac.in

Table 1
Dataset Statistics of Task 1

Language	Training Set	Validation Set	Test Set
Hindi	24225* + 10427#	1500	3000
Gujarati	36629*	-	1457
Bengali	15307*	-	2206
Kannada	10694#	1188	5093
Tamil	4104#	456	1955
Telugu	9583#	1065	4564
English	31237* + 9376#	1500	2500

# Indicates new data added this year
* Indicates data from previous editions included in this year's Training set

We introduced the misinformation task in ILSUM 2023 with the aim of countering the possible misuse of Large Language Models (LLMs), such as GPT, Llama etc., for generating fake news and spreading misinformation. This year, we extended the task to a cross-lingual setup, attempting to make misinformation detection language independent. The dataset consisted of a source article in English and a summary, with possible misinformation, in Hindi and Gujarati. In Section 2 we describe the two tasks and talk about the dataset and evaluation, we cover the approaches used by the participation teams in section 3 followed by results in section 4 and finally the conclusion and acknowledgement in sections 5 and 6.

2. Task Description

This edition of the ILSUM shared task included two subtasks. The first subtask continued the text summarization task from previous years. We expanded datasets for existing languages and added datasets for three new languages. The second subtask built upon the misinformation detection subtask from ILSUM 2023 and included a cross-lingual setup with English as the source language for the article and Gujarati and Hindi as the target languages for the summaries.

2.1. Task 1: Text Summarization for Indian Languages

This is a standard text summarization task and a continuation from ILSUM 2022 and the Subtask 1 of ILSUM 2023. Given an article, participants are expected to generate a fixed-length summary. The summary could be extractive or abstractive. The ground truth consists of human-written headlines extracted from the beginning of the new articles. In the current edition, we added additional articles for Hindi, Bengali, Gujarati and English articles. We also added three Dravidian languages Kannada, Tamil and Telugu. Table 1 contains the dataset statistics for Training, Validation and Test split for each language. This year's Training data included all article-summary pairs from previous editions. Table 1 shows both the newly added documents in the current edition and those from the past editions.

As with the previous editions, the current dataset poses a unique challenge of code-mixing and script-mixing. It is common for news articles to borrow phrases from English, even if the article is written in an Indian language. Examples such as the ones shown below commonly occur both in the headlines and article bodies:

- Gujarati: "IND vs SA, 5મી T20 તસવીરોમાં: વરસાદે વિલન બની મજા બગાડી" (India vs SA, 5th T20 in pictures: rain spoils the match)
- Hindi: "LIC के IPO में पैसा लगाने वालों का टूटा दिल, आई एक और नुकसानदेह खबर" (Investors of LIC IPO left broken hearted, yet another bad news)

Table 2
Dataset Statistics of Task 2

Class	Training Set	Test Set	Total
Misrepresentation	294	25	319
Inaccurate Quantities	195	10	205
False Attribution	250	13	263
Fabrication	250	32	282
Correct	5000	143	5143

2.2. Task 2: Misinformation Detection in Machine Generated Cross-lingual Summaries

This task builds upon subtask 2 from ILSUM 2023. In this task, participants were provided with a source article in English and a corresponding summary in Hindi and Gujarati. The aim was to identify factually incorrect summaries and further classify them into one of the four known categories of incorrectness listed below:

- **Misrepresentation**: This involves presenting information in a way that is misleading or that gives a false impression. This may be achieved by exaggerating certain aspects, understating others, or twisting facts to fit a particular narrative.
- **Inaccurate Quantities or Measurements**: Factual incorrectness can occur when precise quantities, measurements, or statistics are misrepresented, whether through error or intent.
- **False Attribution**: Incorrectly attributing a statement, idea, or action to a person or group is another form of factual incorrectness.
- **Fabrication**: Making up data, sources, or events is a severe form of factual incorrectness. This involves creating "facts" that have no basis in reality.

For creating the dataset, we used the OpenAI GPT models. GPT-4 was used to generate incorrect summaries corresponding to each class, as well as to translate English summaries to Hindi and Gujarati. The GPT-3.5 model was used to generate the correct summaries. For a given news article, we first used carefully crafted prompts to generate automatic summaries corresponding to each type of misinformation without any manual intervention. Detailed procedure used for creating the english summaries with misinformation is described in our dataset paper [32]. Next we generated factually correct summaries. This was done by simply prompting GPT to generate a summary and then manually validating a subset.

Contrary to the general belief, we did not find any factual inaccuracy or hallucination in the manually evaluated subset for the correct summaries. This is the same dataset that was used in ILSUM 2023. For the cross-lingual dataset, we translated the correct summaries and summaries with misinformation in English to Hindi and Gujarati using GPT. Our experiments showed this to be more accurate and reliable than generating the incorrect summaries directly in Hindi or Gujarati. Participants were then provided the original source article as well as the translated summaries, but not the English summaries.

Each text article and generated summary is categorized into one of the four predefined types of misinformation in Training data. Participants were asked to predict all possible labels associated with text summaries in Test data, as one summary can have multiple types of incorrectness. Example articles for each category of misinformation is available at https://ilsum.github.io/ilsum/2023/index.html. Table 2 contains dataset statistics for Task 2 dataset. The class predictions on test data are evaluated using Macro F1 score.

3. Methodologies of the Participating Teams

This section provides a brief description of the methodologies submitted by the participating teams.

- **GJCoders** - participated in Task 1 for English language only. They used a mix of LDA and TextRank for generating an extractive summary [33]. Further, they propose a dynamic damping factor for TextRank. The damping factor is a weighted combination of several variables like node connectivity, entropy, variance in topics across sentences, document length and number of topics. The weights were tuned using grid search.

- **TextTitans** - participated only in Task 2. They used zero shot prompting for misinformation detection [34]. Specifically, they prompted GPT 3.5 with various temperature settings, explaining what each category of misinformation is and then asking whether a given summary has misinformation. They also proposed an ensemble that combines results from different temperature settings.
- **Iem Inturns** - participated in both the tasks [35]. For subtask 1, they submitted runs only for English language and used T5-small model with a beam size 4. For subtask 2, they submitted runs for both Hindi and Gujarati. They trained several classifiers like Logistic Regression, SVC, Decision Tree, Random Forest, and Naive Bayes.
- **Sangita_NIT_Patna** - participated only in Task 1 and submitted runs for all the languages [36]. They used SVD based summarization, where a N-gram TF-IDF matrix is first decomposed using SVD and then top-k sentences are selected as the summary.
- **Data Lovers** - participated only in Task 1 and submitted runs for all languages except Kannada [37]. They used pre-trained models to generate summaries. Specifically they used BART for English, Indic Bart for Hindi and Tamil and mT5 for Gujarati, Telugu and Bengali.
- **Curious Coders** - participated in both tasks for all languages [38]. For Task 1, the team used mT5 pre-trained model for all languages. For Task 2, they fine-tuned Llama3.
- **SynopSizers** - participated only in Task 1 for Bengali, Gujarati, English and Tamil languages [39]. They employed several techniques including frequency based summarization, tf-idf based summarization and pre-trained models. For pre-trained models, they report results on mT5, XLSum, mT5-Tamil, MultiIndic, Tamil-Bert and Indic-Bert.
- **CSSG** - participated only in Task 1 for Hindi [40] and used pre-trained Indic BART for the same.
- **Squad** - participated in both the tasks [41]. For Task 1, they submitted runs for English, Tamil, Telugu and Gujarati. They used several pre-trained LLMs like BART, T5, mT5, and mT5_m2m_CrossSum for this. For Task 2, they used Support Vector Machine, Logistic Regression and Random Forest classifiers, for both Hindi and Gujarati.
- **IdliVadaSambar** - participated only in Task 1 for English and used pre-trained models like Gemini and T5[42].
- **INITIATORS** - participated only in Task 1 for English, Tamil, Telugu and Kannada [43]. They used pre-trained IndiBARTSS model for the Dravidian languages and T5-Base for English. Instead of using pre-trained models directly like many other teams, they used post-processing in the form of beam search, sentence length control and Heading Integration. This allows them to avoid repetition, control summary size and align the focus of the summary with the heading.
- **iVSum** - participated only in Tsk-2 for both the languages [44]. The team explored several classifiers for this including Logistic Regression, Logistic Regression with Class Weights and bert-base-uncased BERT with Focal Loss.

From the above description, it is evident that most of the teams have explored pre-trained models for both the tasks.

4. Results

In this section, we present the best runs for each team for each task-language pair.

4.1. Task 1

We include results separately for all languages, as well as for both ROUGE and BertScore. Tables 3 and 4 show ranking of the participating teams for Bengali language based on ROUGE and BertScore respectively. Similarly, we include separate tables for Gujarati (Tables 7 and 8), Hindi (Tables 5 and 6), Tamil (Tables 11 and 12), Telugu (Tables 9 and 10), Kannada (Tables 13 and 14) and English (Tables 15 and 16). Overall, we observed a strong positive correlation between the rankings generated by the ROUGE scores and those by the BertScores.

The best-performing approaches in Task 1 were dominated by two teams, Initiators [43] and DataLovers [37]. Both teams used pre-trained models T5 and BART or their variants to get the best performance.

Table 3
Task 1 - Bengali ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**Data Lovers**	0.2471	0.1658	0.1187	0.2297
2	**Curious Coders**	0.2411	0.1610	0.1133	0.2233
3	**Sangita_NIT_Patna**	0.2096	0.1409	0.1057	0.1847
4	**SynopSizers**	0.1957	0.1224	0.0938	0.1693
5	**SCaLAR**	0.0397	0.0103	0.0010	0.0379

Table 4
Task 1 - Bengali BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**Data Lovers**	0.7372	0.7316	0.7338
2	**Curious Coders**	0.7384	0.7251	0.7310
3	**Sangita_NIT_Patna**	0.6996	0.7233	0.7099
4	**SynopSizers**	0.6755	0.7320	0.7014
5	**SCaLAR**	0.4178	0.6081	0.4948

Table 5
Task 1 - Hindi ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**Data Lovers**	0.3659	0.1975	0.1233	0.3388
2	**CSSG**	0.3421	0.1713	0.102	0.312
3	**Curious Coders**	0.331	0.163	0.0978	0.2981
4	**Sangita_NIT_Patna**	0.3015	0.1375	0.0827	0.2729

Table 6
Task 1 - Hindi BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**Data Lovers**	0.7196	0.7621	0.7396
2	**Curious Coders**	0.741	0.7343	0.7371
3	**CSSG**	0.7204	0.7449	0.7318
4	**Sangita_NIT_Patna**	0.7129	0.7308	0.7207

4.2. Task 2

We include the results for Hindi and Gujarati in Tables 17 and 18 respectively. Unlike Task 1 where the best performing approaches were dominated by pre-trained LLMS, for Task 2 the best-performing approaches mainly included classical machine learning techniques used by ivSUM [44] and Squad [41].

Table 7
Task 1 - Gujarati ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**Data Lovers**	0.2792	0.1496	0.0942	0.2669
2	**Curious Coders**	0.2723	0.1467	0.086	0.2607
3	**Sangita_NIT_Patna**	0.2526	0.1188	0.0703	0.2415
4	**SynopSizers**	0.2109	0.0835	0.0437	0.1958
5	**Squad**	0.181	0.0811	0.0347	0.1748
6	**SCaLAR**	0.0819	0.0244	0.0045	0.0802
7	**Trojan Horses**	0.0516	0.0101	0.001	0.0484

Table 8
Task 1 - Gujarati BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**Data Lovers**	0.7506	0.7301	0.7398
2	**Curious Coders**	0.7485	0.7303	0.7388
3	**Sangita_NIT_Patna**	0.7242	0.732	0.7274
4	**Squad**	0.7578	0.6844	0.7186
5	**SynopSizers**	0.6929	0.7301	0.7105
6	**Trojan Horses**	0.5796	0.6116	0.5942
7	**SCaLAR**	0.4712	0.6497	0.5458

Table 9
Task 1 - Telugu ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**INITIATORS**	0.3146	0.2318	0.1802	0.3079
2	**Data Lovers**	0.3022	0.2149	0.1606	0.2963
3	**Curious Coders**	0.1973	0.1161	0.0755	0.1916
4	**Sangita_NIT_Patna**	0.1926	0.1139	0.0769	0.1876
5	**Squad**	0.1498	0.0695	0.0226	0.1434

Table 10
Task 1 - Telugu BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**INITIATORS**	0.7488	0.7637	0.7555
2	**Data Lovers**	0.7544	0.7527	0.7532
3	**Curious Coders**	0.7175	0.7164	0.7166
4	**Sangita_NIT_Patna**	0.7006	0.7236	0.7112
5	**Squad**	0.7392	0.6765	0.7058

This could be because of the limited amount of training data available for this task, and the inability of the LLMs to perform this task out of the box.

5. Conclusion

The ILSUM 2024 track continued the efforts from the previous years in building test collections for summarization and misinformation detection tasks. A new task that focused on identifying misinformation in LLM-generated cross-lingual summaries was included. While the majority of participating teams used one or more pre-trained LLMs to approach the summarization task, some teams tried to blend additional steps and non-LLM approaches. From our experience with the three editions of ILSUM it seems using LLMs for generating summaries for Indian Languages might be reaching a saturation point and a novel more language-focused approaches might be required for further improvements.

Further, the traditional machine learning models outperformed LLMs for the misinformation detection

Table 11
Task 1 - Tamil ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**Data Lovers**	0.2376	0.1507	0.1018	0.2284
2	**INITIATORS**	0.2180	0.1336	0.0905	0.2091
3	**Curious Coders**	0.1962	0.1175	0.0801	0.1872
4	**SynopSizers**	0.1547	0.0877	0.0561	0.1468
5	**Sangita_NIT_Patna**	0.1392	0.075	0.0487	0.1334
6	**Squad**	0.0121	0.0007	0.0001	0.0120
7	**SCaLAR**	0.0097	0.0019	0.0001	0.0096

Table 12
Task 1 - Tamil BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**Data Lovers**	0.7226	0.7496	0.7354
2	**INITIATORS**	0.7301	0.7292	0.7290
3	**Curious Coders**	0.7123	0.7283	0.7197
4	**Sangita_NIT_Patna**	0.6867	0.7051	0.6948
5	**SynopSizers**	0.6606	0.7292	0.6925
6	**Squad**	0.6192	0.5747	0.5951
7	**SCaLAR**	0.3928	0.5839	0.4693

Table 13
Task 1 - Kannada ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**INITIATORS**	0.2284	0.1446	0.1032	0.2218
2	**Curious Coders**	0.0524	0.0095	0.0012	0.0491

Table 14
Task 1 - Kannada BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**INITIATORS**	0.7274	0.7438	0.7349
2	**Curious Coders**	0.6713	0.6677	0.6691

task. One reason for this could be the limited size of the dataset for these tasks. In future, we will focus on expanding these datasets and introducing new categories of misinformation. Another possibility that we previously discussed but are yet to explore is the identification of misinformation at a more granular level by extracting the factually incorrect spans within the machine-generated summaries.

Table 15
Task 1 - English ROUGE score

Rank	Team Name	Rouge-1	Rouge-2	Rouge-4	Rouge-L
1	**Data Lovers**	0.3644	0.206	0.1467	0.3133
2	**INITIATORS**	0.3435	0.1879	0.1357	0.2964
3	**IdliVadaSambar**	0.3102	0.1554	0.0856	0.2644
4	**Iem inturns**	0.3044	0.1448	0.0843	0.2646
5	**NITK_Surathkal**	0.3039	0.1293	0.0664	0.2468
6	**Curious Coders**	0.2989	0.1392	0.0927	0.2497
7	**SynopSizers**	0.2295	0.0894	0.0513	0.1895
8	**Squad**	0.2214	0.0579	0.0111	0.1761
9	**SCaLAR**	0.1391	0.0437	0.0123	0.1189

Table 16
Task 1 - English BertScore

Rank	Team Name	BertScore-Precision	BertScore-Recall	BertScore-F1
1	**Data Lovers**	0.8706	0.8862	0.8781
2	**INITIATORS**	0.8701	0.8789	0.8742
3	**IdliVadaSambar**	0.8529	0.8829	0.8675
4	**Curious Coders**	0.8684	0.8664	0.8672
5	**NITK_Surathkal**	0.8591	0.8686	0.8636
6	**Squad**	0.8764	0.8447	0.8601
7	**Iem inturns**	0.8482	0.8708	0.8591
8	**SynopSizers**	0.8195	0.8706	0.8439
9	**SCaLAR**	0.8002	0.8284	0.8136

Table 17
Task 2 - Hindi Results

Rank	Team Name	F1-Score
1	ivSUM	0.3426
2	Squad	0.3153
3	CUET_SSTM	0.2371
4	Team Bug Smashers	0.2367
5	Trojan Horses	0.2287
6	Iem inturns	0.2132
7	Curious Coders	0.181
8	ByteSpark	0.1808
9	f4ctual	0.0969
10	TextTitans	0.0677

Table 18
Task 2 - Gujarati Results

Rank	Team Name	F1-Score
1	ivSUM	0.3371
2	Squad	0.296
3	Trojan Horses	0.2456
4	Iem inturns	0.2127
5	Curious Coders	0.1892
6	Team Bug Smashers	0.1884
7	ByteSpark	0.1861
8	f4ctual	0.1179
9	TextTitans	0.0677

6. Acknowledgment

We would like to acknowledge and thank the student volunteers Dhwanit Shah (PDEU, India), Pannag Agrawal (DDIT, India) and Sushanth Kumble (Mangalore University, India) who helped us in creating the dataset.

Declaration on Generative AI

The authors confirm that no generative AI tools were used in the writing, editing, or analysis processes of this manuscript. All content was created and reviewed by the authors.

References

[1] S. Satapara, P. Mehta, S. Modha, A. Hegde, S. HL, D. Ganguly, Key insights from the third ilsum track at fire 2024, in: Proceedings of the 16th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2024, Gandhiinagar, India. December 12-15, 2024, ACM, 2024.

[2] P. Mehta, T. Mandl, P. Majumder, S. Gangopadhyay, Report on the FIRE 2020 evaluation initiative, SIGIR Forum 55 (2021) 3:1–3:11. URL: https://doi.org/10.1145/3476415.3476418. doi:10.1145/3476415.3476418.

[3] S. Satapara, B. Modha, S. Modha, P. Mehta, Findings of the first shared task on indian language summarization (ILSUM): approaches challenges and the path ahead, in: K. Ghosh, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2022 - Forum for Information Retrieval Evaluation, Kolkata, India, December 9-13, 2022, volume 3395 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 369–382. URL: https://ceur-ws.org/Vol-3395/T6-1.pdf.

[4] S. Satapara, B. Modha, S. Modha, P. Mehta, FIRE 2022 ILSUM track: Indian language summarization, in: D. Ganguly, S. Gangopadhyay, M. Mitra, P. Majumder (Eds.), Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2022, Kolkata, India, December 9-13, 2022, ACM, 2022, pp. 8–11. URL: https://doi.org/10.1145/3574318.3574328. doi:10.1145/3574318.3574328.

[5] S. Satapara, P. Mehta, S. Modha, D. Ganguly, Key takeaways from the second shared task on indian language summarization (ILSUM 2023), in: K. Ghosh, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2023 - Forum for Information Retrieval Evaluation (FIRE-WN 2023), Goa, India, December 15-18, 2023, volume 3681 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 724–733. URL: https://ceur-ws.org/Vol-3681/T8-1.pdf.

[6] S. Satapara, P. Mehta, S. Modha, D. Ganguly, Indian language summarization at FIRE 2023, in: D. Ganguly, S. Majumdar, B. Mitra, P. Gupta, S. Gangopadhyay, P. Majumder (Eds.), Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE 2023, Panjim, India, December 15-18, 2023, ACM, 2023, pp. 27–29. URL: https://doi.org/10.1145/3632754.3634662. doi:10.1145/3632754.3634662.

[7] T. Mandl, S. Modha, G. K. Shahi, H. Madhu, S. Satapara, P. Majumder, J. Schäfer, T. Ranasinghe, M. Zampieri, D. Nandini, A. K. Jaiswal, Overview of the HASOC subtrack at FIRE 2021: Hatespeech and offensive content identification in english and indo-aryan languages, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 1–19. URL: http://ceur-ws.org/Vol-3159/T1-1.pdf.

[8] T. Mandl, S. Modha, G. K. Shahi, A. K. Jaiswal, D. Nandini, D. Patel, P. Majumder, J. Schäfer, Overview of the HASOC track at FIRE 2020: Hate speech and offensive content identification in indo-european languages, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2020 - Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, volume 2826 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 87–111. URL: http://ceur-ws.org/Vol-2826/T2-1.pdf.

[9] S. Modha, T. Mandl, P. Majumder, D. Patel, Overview of the HASOC track at FIRE 2019: Hate speech and offensive content identification in indo-european languages, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 167–190. URL: http://ceur-ws.org/Vol-2517/T3-1.pdf.

[10] H. Madhu, S. Satapara, S. Modha, T. Mandl, P. Majumder, Detecting offensive speech in conversational code-mixed dialogue on social media: A contextual dataset and benchmark experiments, Expert Systems with Applications (2022) 119342.

[11] S. Modha, P. Majumder, T. Mandl, C. Mandalia, Detecting and visualizing hate speech in social media: A cyber watchdog for surveillance, Expert Syst. Appl. 161 (2020) 113725. URL: https://doi.org/10.1016/j.eswa.2020.113725. doi:10.1016/j.eswa.2020.113725.

[12] M. Subramanian, R. Ponnusamy, S. Benhur, K. Shanmugavadivel, A. Ganesan, D. Ravi, G. K.

Shanmugasundaram, R. Priyadharshini, B. R. Chakravarthi, Offensive language detection in tamil youtube comments by adapters and cross-domain knowledge transfer, Comput. Speech Lang. 76 (2022) 101404. URL: https://doi.org/10.1016/j.csl.2022.101404. doi:10.1016/j.csl.2022.101404.

[13] B. R. Chakravarthi, R. Priyadharshini, V. Muralidaran, S. Suryawanshi, N. Jose, E. Sherly, J. P. McCrae, Overview of the track on sentiment analysis for dravidian languages in code-mixed text, in: P. Majumder, M. Mitra, S. Gangopadhyay, P. Mehta (Eds.), FIRE 2020: Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, ACM, 2020, pp. 21–24. URL: https://doi.org/10.1145/3441501.3441515. doi:10.1145/3441501.3441515.

[14] B. R. Chakravarthi, P. K. Kumaresan, R. Sakuntharaj, A. K. Madasamy, S. Thavareesan, B. Premjith, S. K, S. C. Navaneethakrishnan, J. P. McCrae, T. Mandl, Overview of the hasoc-dravidiancodemix shared task on offensive language detection in tamil and malayalam, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 589–602. URL: http://ceur-ws.org/Vol-3159/T3-1.pdf.

[15] J. Gala, P. A. Chitale, R. AK, S. Doddapaneni, V. Gumma, A. Kumar, J. Nawale, A. Sujatha, R. Puduppully, V. Raghavan, et al., Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages, arXiv preprint arXiv:2305.16307 (2023).

[16] S. Banerjee, K. Chakma, S. K. Naskar, A. Das, P. Rosso, S. Bandyopadhyay, M. Choudhury, Overview of the mixed script information retrieval (MSIR) at FIRE-2016, in: P. Majumder, M. Mitra, P. Mehta, J. Sankhavara, K. Ghosh (Eds.), Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December 7-10, 2016, volume 1737 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2016, pp. 94–99. URL: http://ceur-ws.org/Vol-1737/T3-1.pdf.

[17] R. Sequiera, M. Choudhury, P. Gupta, P. Rosso, S. Kumar, S. Banerjee, S. K. Naskar, S. Bandyopadhyay, G. Chittaranjan, A. Das, K. Chakma, Overview of FIRE-2015 shared task on mixed script information retrieval, in: P. Majumder, M. Mitra, M. Agrawal, P. Mehta (Eds.), Post Proceedings of the Workshops at the 7th Forum for Information Retrieval Evaluation, Gandhinagar, India, December 4-6, 2015, volume 1587 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015, pp. 19–25. URL: http://ceur-ws.org/Vol-1587/T2-1.pdf.

[18] P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, P. Mehta, A. Bhattacharya, P. Majumder, Overview of the FIRE 2019 AILA track: Artificial intelligence for legal assistance, in: P. Mehta, P. Rosso, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2019 - Forum for Information Retrieval Evaluation, Kolkata, India, December 12-15, 2019, volume 2517 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–12. URL: http://ceur-ws.org/Vol-2517/T1-1.pdf.

[19] P. Bhattacharya, P. Mehta, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, FIRE 2020 AILA track: Artificial intelligence for legal assistance, in: P. Majumder, M. Mitra, S. Gangopadhyay, P. Mehta (Eds.), FIRE 2020: Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, ACM, 2020, pp. 1–3. URL: https://doi.org/10.1145/3441501.3441510. doi:10.1145/3441501.3441510.

[20] V. Parikh, U. Bhattacharya, P. Mehta, A. Bandyopadhyay, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, AILA 2021: Shared task on artificial intelligence for legal assistance, in: D. Ganguly, S. Gangopadhyay, M. Mitra, P. Majumder (Eds.), FIRE 2021: Forum for Information Retrieval Evaluation, Virtual Event, India, December 13 - 17, 2021, ACM, 2021, pp. 12–15. URL: https://doi.org/10.1145/3503162.3506571. doi:10.1145/3503162.3506571.

[21] V. Parikh, V. Mathur, P. Mehta, N. Mittal, P. Majumder, Lawsum: A weakly supervised approach for indian legal document summarization, CoRR abs/2110.01188 (2021). URL: https://arxiv.org/abs/2110.01188. arXiv:2110.01188.

[22] S. Ghosh, A. Wyner, Identification of rhetorical roles of sentences in indian legal judgments, in: Legal Knowledge and Information Systems: JURIX 2019: The Thirty-second Annual Conference, volume 322, IOS Press, 2019, p. 3.

[23] S. Parashar, N. Mittal, P. Mehta, Casrank: A ranking algorithm for legal statute retrieval, Multimedia Tools and Applications (2023) 1–18.

[24] M. Amjad, G. Sidorov, A. Zhila, Data augmentation using machine translation for fake news

detection in the urdu language, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020, European Language Resources Association, 2020, pp. 2537–2542. URL: https://aclanthology.org/2020.lrec-1.309/.

[25] M. Amjad, A. Zhila, G. Sidorov, A. Labunets, S. Butt, H. I. Amjad, O. Vitman, A. F. Gelbukh, Overview of abusive and threatening language detection in urdu at FIRE 2021, in: P. Mehta, T. Mandl, P. Majumder, M. Mitra (Eds.), Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021, volume 3159 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 744–762. URL: http://ceur-ws.org/Vol-3159/T4-1.pdf.

[26] M. Amjad, N. Ashraf, A. Zhila, G. Sidorov, A. Zubiaga, A. F. Gelbukh, Threatening language detection and target identification in urdu tweets, IEEE Access 9 (2021) 128302–128313. URL: https://doi.org/10.1109/ACCESS.2021.3112500. doi:10.1109/ACCESS.2021.3112500.

[27] P. Mehta, P. Majumder, Optimum parameter selection for K.L.D. based authorship attribution in gujarati, in: Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013, Asian Federation of Natural Language Processing / ACL, 2013, pp. 1102–1106. URL: https://aclanthology.org/I13-1155/.

[28] P. Mehta, P. Majumder, Large scale quantitative analysis of three indo-aryan languages, J. Quant. Linguistics 23 (2016) 109–132. URL: https://doi.org/10.1080/09296174.2015.1071151. doi:10.1080/09296174.2015.1071151.

[29] M. Basu, S. Ghosh, K. Ghosh, Overview of the fire 2018 track: Information retrieval from microblogs during disasters (irmidis), in: Proceedings of the 10th annual meeting of the Forum for Information Retrieval Evaluation, 2018, pp. 1–5.

[30] S. Majumdar, A. Bandyopadhyay, S. Chattopadhyay, P. P. Das, P. D. Clough, P. Majumder, Overview of the irse track at fire 2022: Information retrieval in software engineering, in: Forum for Information Retrieval Evaluation, ACM, 2022.

[31] S. Majumdar, S. Paul, D. Paul, A. Bandyopadhyay, S. Chattopadhyay, P. P. Das, P. D. Clough, P. Majumder, Generative ai for software metadata: Overview of the information retrieval in software engineering track at fire 2023, arXiv preprint arXiv:2311.03374 (2023).

[32] S. Satapara, P. Mehta, D. Ganguly, S. Modha, Fighting fire with fire: Adversarial prompting to generate a misinformation detection dataset, CoRR abs/2401.04481 (2024). URL: https://doi.org/10.48550/arXiv.2401.04481. doi:10.48550/ARXIV.2401.04481. arXiv:2401.04481.

[33] A. Vora, A. Shah, R. Jain, S. Sonawane, Text summarization for indian languages: Extractive summarization using graph based technique, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.

[34] A. Deroy, S. Maity, Crossfactual: A novel approach for detecting factual inaccuracies in machine-generated summaries, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.

[35] R. Jha, R. Ahuja, S. K. Mahata, M. Dey, D. Das, Enhancing accuracy in indian language summarization, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.

[36] S. Singh, J. P. Singh, A. Deepak, Supriya, Enhancing extractive summarization for low resource indian languages using tf-idf and svd, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.

[37] P. Chatwal, A. Agarwal, A. Mittal, Overcoming code-mixing and script-mixing in indian language summarization with transformer models, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.

[38] A. Balan, C. Karthik, O. Christopher, Indian language summarization and factual error detection using pretrained sequence-to-sequence models and large language models, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.
[39] T. Sriram, A. Raman, S. Anand, D. Thenmozhi, Text summarization using pre-trained models on tamil, english, gujarati and bengali, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.
[40] C. Subhedar, S. Sonawane, Abstractive summarization of large articles in hindi language using indicbart, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.
[41] D. Thenmozhi, R. R, N. A, M. Ravi, P. R, Text summarization and detection of factual incorrectness for indian languages, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.
[42] M. Saipranav, M. Sreekumar, D. Thenmozhi, S. Karthik, R. VS, Summarizing english news articles: Leveraging t5 and google gemini 1.0 pro, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.
[43] F. F. A, R. Sheik, K. A, S. J. Nirmala, Enhancing news article summarization through post-processing techniques, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, CEUR-WS.org, 2024.
[44] K. S. Potta, J. Sridharan, M. R. Ramya, S. Gopalakrishnan, D. Thenmozhi, Assessing factual accuracy in machine generated cross lingual summaries using logistic regression and bert, in: K. Ghosh, T. Mandl, P. Majumder, D. Ganguly (Eds.), Working Notes of FIRE 2024 - Forum for Information Retrieval Evaluation, Gandhinagar, India. December 12-15, 2024, 2024.