XL-HeadTags: Leveraging Multimodal Retrieval Augmentation for the Multilingual Generation of News Headlines and Tags

Faisal Tareque Shohan[♣*] **Mir Tafseer Nayeem**[♠*†]
Samsul Islam[♣] **Abu Ubaida Akash**[◇] **Shafiq Joty**[♥♦]
[♣]Ahsanullah University of Science & Technology [♠]University of Alberta
[◇]Université de Sherbrooke [♥]Salesforce Research [♦]Nanyang Technological University
faisaltareque@hotmail.com mnayeem@ualberta.ca samsulratul98@gmail.com
abu.ubaida.akash@usherbrooke.ca sjoty@salesforce.com

Abstract

Millions of news articles published online daily can overwhelm readers. Headlines and entity (topic) tags are essential for guiding readers to decide if the content is worth their time. While headline generation has been extensively studied, tag generation remains largely unexplored, yet it offers readers better access to topics of interest. The need for conciseness in capturing readers' attention necessitates improved content selection strategies for identifying salient and relevant segments within lengthy articles, thereby guiding language models effectively. To address this, we propose to leverage auxiliary information such as images and captions embedded in the articles to retrieve relevant sentences and utilize instruction tuning with variations to generate both headlines and tags for news articles in a multilingual context. To make use of the auxiliary information, we have compiled a dataset named **XL-HeadTags**, which includes 20 languages across 6 diverse language families. Through extensive evaluation, we demonstrate the effectiveness of our *plug-and-play* multimodal-multilingual retrievers for both tasks. Additionally, we have developed a suite of tools for processing and evaluating multilingual texts, significantly contributing to the research community by enabling more accurate and efficient analysis across languages.[1]

1 Introduction

The headline serves as a concise and attention-grabbing summary of a news article. Articles with compelling headlines are more likely to attract increased views or shares (Gu et al., 2020; Song et al., 2020). Unlike summaries, which provide a broad overview (Nayeem et al., 2018), headlines aim to produce brief and engaging statements that draw readers into the full article (Xu et al., 2019; Zhang and Yang, 2023; Akash et al., 2023). They are often the most-read part of the article, guiding readers in deciding whether the content merits their time (Bukhtiyarov and Gusev, 2020).

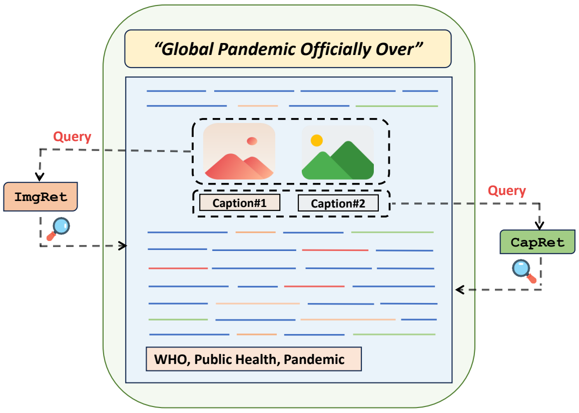

"Global Pandemic Officially Over"
Query
ImgRet
Caption#1
Caption#2
Query
CapRet
WHO, Public Health, Pandemic

Figure 1: Our content selection approach. Auxiliary information, including images and captions, is used as *queries* to extract salient and relevant sentences from documents via two modules: **ImgRet** for images (`visual modality`) and **CapRet** for image captions (`textual modality`). Our modules are designed as `plug-and-play` components that can be integrated with language models of any size and type.

Another important feature of a news article are topic tags or "`semantic markers`", which serve as dynamic connectors and navigational aids, enhancing the coherence and accessibility of information across articles. The task of tag generation is related to keyphrase generation (Meng et al., 2017). While keyphrases summarize the main themes succinctly of an article, tags provide a broader overview, guiding readers to related articles and facilitating navigation through connected themes. Despite their significant role, the generation of news article tags has remained unexplored in existing literature.

In this work, we introduce a unified framework for generating headlines and tags for news articles in a multimodal-multilingual context, covering

[*] **Equal contribution.**
[†] Corresponding author.
[1] Our code, dataset, model checkpoints, developed tools are available at XL-HeadTags.

Findings of the Association for Computational Linguistics: ACL 2024, pages 12991–13024
August 11-16, 2024 ©2024 Association for Computational Linguistics

20 languages across 6 diverse language families. Despite the emergence of open-source Large Language Models (LLMs) like Llama (Touvron et al., 2023) and Mistral (Jiang et al., 2023), multilingual models such as mT5 (Xue et al., 2021) and mT0 (Muennighoff et al., 2022), fine-tuned with task-specific data, remain the preferred solution for multilingual tasks, especially for low-resource languages (Ahuja et al., 2023; Zhao et al., 2024; Aggarwal et al., 2024). News articles often include extensive content, incorporating additional details, author quotes, historical context, and advertisements, among others. This poses a challenge in identifying article segments that are both salient and relevant for the creation of highly concise outputs like headlines and tags. This complication leads to what is known as the "Lost-in-the-Context" problem (Liu et al., 2023), wherein vital information embedded within lengthy documents is frequently overlooked by the models (Ravaut et al., 2024).

Fortunately, it is now very common for both online and printed news media to use multimedia content to enhance visibility, support, and context for articles (Oostdijk et al., 2020). Digital assets, like images, often serve as thumbnails across social media, blogs, and various platforms. Equally important are the captions accompanying these images, which not only clarify and enrich the image but also optimize articles for search engines and make news more accessible to those with vision impairments (Liu et al., 2021). This has motivated us to utilize auxiliary information (images and captions as illustrated in Figure 1) to distill salient and localized information from news articles in a multimodal-multilingual context using the **CLIP-ViT-B32** model (Radford et al., 2021). This model aligns text and images within a unified dense vector space. Our method is inspired by the **Cognitive Load Theory** (Sweller, 2011), reflecting how humans employ visual cues and summaries to understand the essence of lengthy texts without being overwhelmed by information. Building on this, with the contents selected using our multimodal retrievers, we utilize instruction tuning to generate both headlines and tags for news articles in a multilingual context.

Our contributions are summarized as follows:

- We compile the **XL-HeadTags** dataset for headline and tags generation tasks, expanding it to include 20 languages across 6 diverse language families (§2.1).

- We present a new approach to content selection that utilizes auxiliary information from both textual and visual modalities to identify the most salient content within news articles in a multilingual setting. Our modules are crafted as plug-and-play components, allowing for seamless integration with language models of any size and type (§3).

- We utilize instruction tuning to generate both headline and tag words. Our model is capable of producing tag words in both controlled and unrestricted manners through instructions (§3.2). Furthermore, we introduce novel tag words evaluation metrics designed to evaluate scenarios of both controlled and unrestricted generation effectively (§5.1).

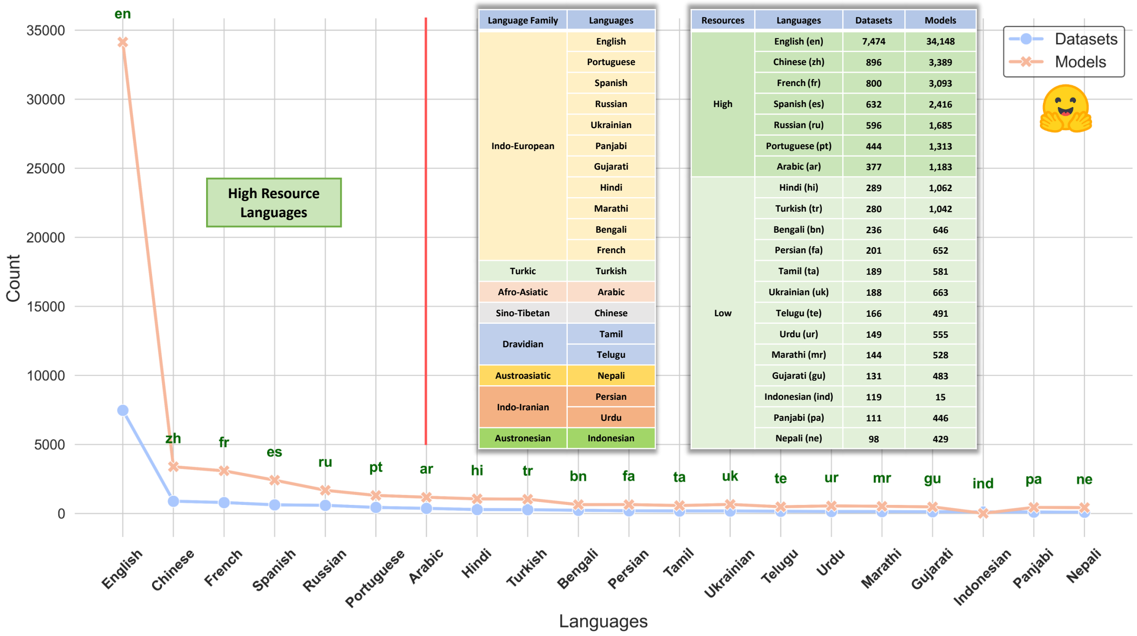
- We also develop tools by accumulating open-source resources for processing and evaluating multilingual texts, making it an easy-to-use one-stop destination. These tools include **(1)** Multilingual ROUGE Scorer, **(2)** Multilingual Sentence Tokenizer, and **(3)** Multilingual Stemmer. These resources are invaluable to the research community focusing on multilingual NLP (§4).

2 XL-HeadTags: Dataset & Tasks

2.1 Dataset

Our study focuses on exploring techniques for multilingual and multimodal retrieval, aiming to extract salient and localized information from documents. This is intended to support the generation of succinct headlines and tags. Our goal is to utilize auxiliary information, such as images and image captions, as queries to retrieve salient information from the document. To achieve our goals, we explore existing large-scale multilingual and multimodal datasets to identify auxiliary information, such as images and their captions. Table 1 highlights several well-known datasets that serve summarization purposes. We choose M3LS (Verma et al., 2023) and XL-Sum (Hasan et al., 2021) as our primary data sources. Since both datasets share the BBC as their source, we anticipate minimal distributional and structural shifts, which could enhance the coherence and efficiency of our retrieval process.

M3LS is a large-scale dataset designed for Multimodal Multilingual Summarization, featuring headlines, articles, summaries, images, captions and tag words. In contrast, XL-Sum focuses on Multilingual



Language Family	Languages
Indo-European	English
	Portuguese
	Spanish
	Russian
	Ukrainian
	Panjabi
	Gujarati
	Hindi
	Marathi
	Bengali
	French
Turkic	Turkish
Afro-Asiatic	Arabic
Sino-Tibetan	Chinese
Dravidian	Tamil
	Telugu
Austroasiatic	Nepali
Indo-Iranian	Persian
	Urdu
Austronesian	Indonesian

Resources	Languages	Datasets	Models
High	English (en)	7,474	34,148
	Chinese (zh)	896	3,389
	French (fr)	800	3,093
	Spanish (es)	632	2,416
	Russian (ru)	596	1,685
	Portuguese (pt)	444	1,313
	Arabic (ar)	377	1,183
Low	Hindi (hi)	289	1,062
	Turkish (tr)	280	1,042
	Bengali (bn)	236	646
	Persian (fa)	201	652
	Tamil (ta)	189	581
	Ukrainian (uk)	188	663
	Telugu (te)	166	491
	Urdu (ur)	149	555
	Marathi (mr)	144	528
	Gujarati (gu)	131	483
	Indonesian (ind)	119	15
	Panjabi (pa)	111	446
	Nepali (ne)	98	429

Figure 2: Distribution of Datasets and Models Across Languages. Data were sourced from the **Huggingface** resource ranking (`https://huggingface.co/languages`) as of February 5, 2024.

Abstractive Summarization, comprising headlines, articles, summaries, and news links. It is noteworthy that although the `M3LS` dataset includes images, captions, and tag words, these were not leveraged for the development of their models. Instead of recrawling the data points, we specifically utilized the auxiliary information from `M3LS`, for our retrieval augmentation framework. Furthermore, to expand our dataset to additional language families, we incorporated Arabic, Turkish, and Persian from the `XL-Sum` dataset. A notable limitation of `XL-Sum` is its absence of images, image captions, and tag words. To address this gap, we utilized the news URLs provided in the `XL-Sum` dataset to gather the necessary auxiliary information for our framework. We employ a `scrapy`[2] framework-based web crawler to systematically collect detailed information, including images, captions, and tag words from news articles. Our criteria require the presence of key elements (headlines, articles, images, captions, and tags) in each data point.

Our **`XL-HeadTags`** dataset consists of 20 languages across six diverse language families. Detailed statistics of our dataset are presented in Table 2 and described in the Appendix A. Additionally, we have developed a mechanism to classify these languages into high-resource and low-resource categories. Conneau et al. (2020) utilized Common-Crawl (CC) pretraining data, quantified in gigabytes (GB), to highlight the disparities between high- and low-resource languages. This differentiation is determined by examining the volume of pretraining data used in developing pretrained language models. In contrast, we define high-resource and low-resource languages based on the availability of task-specific resources. This includes both datasets and pretrained and/or task-specific fine-tuned models, as per the Huggingface resource ranking[3], which offers real-time and up-to-date information. Languages ranked in the top 10 of this list are classified as high-resource, while the others are low-resource. Figure 2 illustrates the distribution of datasets and models across languages.

Datasets	Multi-lingual	Multi-modal	Task	Auxiliary Information					
				Img		Cap		Tag	
				A	U	A	U	A	U
`MSMO` (2018)	✗	✓	Summ	✓	✓	✓	✗	✗	✗
`E-DM` (2018)	✗	✓	Summ	✓	✓	✓	✓	✗	✗
`XSum` (2018a)	✗	✗	Summ	✗	✗	✗	✗	✗	✗
`CNN/DM` (2017)	✗	✗	Summ	✗	✗	✗	✗	✗	✗
`MMSS` (2018)	✗	✓	Summ	✓	✓	✗	✗	✗	✗
`MLASK` (2023)	✗	✓	Summ	✓	✓	✗	✗	✗	✗
`MLSUM` (2020)	✓	✗	Summ	✗	✗	✗	✗	✓	✗
`XL-SUM` (2021)	✓	✗	Summ	✗	✗	✗	✗	✗	✗
`M3LS` (2023)	✓	✓	Summ	✓	✗	✓	✗	✓	✗
`XL-HeadTags`	✓	✓	Headline & Tags	✓	✓	✓	✓	✓	✓

Table 1: Comparison of the **`XL-HeadTags`** (*ours*) dataset with existing multi-lingual and multi-modal datasets. **'A'** means whether the auxiliary information is **"Available"** and **'U'** indicates whether the information is **"Utilized"** by the models. **"Summ"** denotes Summarization Task.

[2] `https://scrapy.org`

[3] `https://huggingface.co/languages`

Language Family	Languages	#Samples	Avg. Word	Avg. Sent	Avg. Tok	Avg. H Word	% of novel n-grams				CR	Avg. I / C	Avg. Tag W	% Pre Tag W
							uni	bi	tri	quad				
Indo-European	English	200,813	569	15.34	909	8.43	33.27	84.22	96.54	98.98	98.51	2.85	2.93	50.49
	Portuguese	4,112	4,769	39.00	8,778	14.19	22.74	71.1	90.70	96.58	99.70	4.40	4.14	31.35
	Spanish	28,406	3,745	34.94	6,084	17.82	22.34	66.03	86.71	93.87	99.52	5.10	4.00	33.04
	Russian	28,272	709	30.79	1,516	9.48	42.32	85.46	96.33	98.83	98.66	3.38	3.54	19.24
	Ukrainian	16,997	611	34.21	1,440	8.56	41.31	86.82	96.64	98.93	98.60	3.35	3.28	20.84
	Panjabi	8,195	798	41.09	2,140	13.48	31.14	77.72	32.12	96.75	98.31	3.46	5.24	47.41
	Gujarati	7,218	832	51.83	2,284	10.80	41.61	83.25	94.99	98.06	98.70	3.35	5.33	42.74
	Hindi	7,191	1,251	65.17	2,579	13.10	23.48	68.09	88.58	95.82	98.95	3.36	4.35	65.31
	Marathi	9,396	803	56.55	2,137	9.82	44.75	83.87	95.81	98.84	98.77	3.39	5.28	43.95
	Bengali	12,954	530	36.51	1,388	9.49	35.17	81.49	94.66	98.20	98.20	3.41	3.43	60.32
	French	6,344	575	20.23	1,115	9.99	29.72	74.50	91.23	96.32	98.26	3.15	3.42	28.98
Turkic	Turkish	5,031	556	26.72	1,225	10.29	39.70	80.53	93.55	97.53	98.14	2.28	3.46	44.67
Afro-Asiatic	Arabic	6,922	653	65.80	1,499	11.43	36.35	81.04	94.16	97.88	98.25	2.60	4.47	46.65
Sino-Tibetan	Mandarin	12,279	1,266	42.66	1,429	13.87	21.13	70.48	88.66	95.37	98.90	4.77	4.69	44.94
Dravidian	Telugu	9,579	536	51.50	1,568	9.02	50.95	87.51	96.80	98.99	98.31	2.69	5.07	38.42
	Tamil	9,973	440	34.75	1,110	8.84	47.94	86.67	96.65	99.06	97.99	2.29	4.14	39.56
Austroasiatic	Nepali	6,185	440	18.77	1,178	9.47	44.17	86.63	96.47	99.12	97.85	2.30	3.40	54.70
Indo-Iranian	Persian	8,830	716	27.19	1,352	11.42	24.82	70.32	89.24	95.67	98.41	2.39	3.08	52.37
	Urdu	13,469	964	36.84	1,712	14.02	22.55	63.25	82.91	91.29	98.55	2.94	3.88	57.48
Austronesian	Indonesian	12,951	779	36.74	1,432	11.38	29.34	76.65	93.18	97.78	98.53	4.82	3.04	34.55
Summary		**415,117**	**902**	**27.17**	**1,632**	**10.13**	**33.60**	**80.83**	**94.37**	**98.89**	**98.88**	**3.21**	**3.47**	**44.64**

Table 2: Statistics about our **XL-HeadTags**. '#Samples' denotes the total number of samples. 'Avg. Word,' 'Avg. Sent,' and 'Avg. Tok' represent the average number of words, sentences, and tokens per document (computed using the BERT-multilingual model (Devlin et al., 2019)), respectively. '% of Novel N-grams' (for n = 1, 2, 3, 4) indicates the proportion of novel n-grams in headlines. 'CR' refers to the compression ratio of headlines. 'Avg. I/C' shows the average number of image-caption pairs per sample. 'Avg. Tag W' signifies the average number of tag words per document. '% Pre Tag W' denotes the average percentage of present tag words in the documents.

2.2 Tasks

Our goal is to simultaneously generate headline and tags for news articles. Given that headlines provide a condensed summary and tag words serve as semantic markers, we propose that these two tasks can be effectively learned in a unified learning framework. Formally, Given a news Article $\mathcal{A}$ consisting of sequence of words $\{a_1, a_2, \cdots, a_n\}$, our objective is to generate an abstractive Headline, $\mathcal{H}$ consisting of sequence of words $\{h_1, h_2, \cdots, h_m\}$ and a set of $\mathcal{T}$ Tag Words $\{\{T_1\}, \{T_2\}, \cdots, \{T_o\}\}$, where each T can be a word $T = \{t_1\}$ or a sequence of word $T = \{t_1, t_2, \cdots, t_p\}$. Thus the task is **XL-HeadTags**$(\mathcal{A}) \implies \mathcal{H}, \mathcal{T}$.

Controlled Generation During the tag word generation phase, we introduce an additional setting to control the number of tag word generations. This task is formally defined as **Con-XL-HeadTags**$(\mathcal{A}, \mathcal{N}) \implies \mathcal{H}, \mathcal{T}^{con}_{1:n}$. Here, $\mathcal{N}$ acts as the control factor determining the number of tag words to be generated, and $\mathcal{T}^{con}_{1:n}$ represents the resulting set of n controlled tag words.

3 MultiRAGen

Retrieval-Augmented Generation (**RAG**) (Lewis et al., 2020a) involves two key phases: **first**, retrieving contextually relevant information, and **second**, using this information to guide the generation process (Zhao et al., 2023). RAG has been applied to various tasks such as *machine translation* (He et al., 2021), *dialogue generation* (Cai et al., 2019), and *abstractive summarization* (Peng et al., 2019). Inspired by these applications, we leverage multimodal information like images and captions to select salient content from news articles, introducing **MultiRAGen** (**Multi**modal **R**etrieval **A**ugmented **Gen**eration). MultiRAGen comprises two main components: **(1)** Multimodal Retrievers (§3.1) and **(2)** Instruction Tuning (§3.2).

3.1 Multimodal Retrievers (MultiRet)

Our approach utilizes auxiliary information, such as images and captions, to extract salient and relevant sentences from documents through two modules: **ImgRet** for images (visual modality) and **CapRet** for image captions (textual modality). Formally, these modules are described as follows:

$$\texttt{MultiRet} = \begin{cases} \textbf{ImgRet}(\mathcal{A}_{1:n'}, \mathcal{I}_{1:m'}, \mathcal{K}, \mathcal{L}) \implies \mathcal{A}^{I}_{1:k}, \\ \\ \textbf{CapRet}(\mathcal{A}_{1:n'}, \mathcal{C}_{1:m'}, \mathcal{K}, \mathcal{L}) \implies \mathcal{A}^{C}_{1:k} \end{cases}$$

where $\mathcal{A}_{1:n'}$ represents the article consisting of $n^{'}$ sentences, $\mathcal{I}_{1:m'}$ and $\mathcal{C}_{1:m'}$ denote the set of images and image captions within the document, respectively, with $m^{'} \geqslant 1$. $\mathcal{K}$ is the number of sentences

Algorithm 1: ImgRet

```
function ImgRet(Article, L, Images, K)
    sentences ← SenSeg(Article, L)
    for sen in sentences do
        for img in Images do
            scr ← SimScr(sen, img)
            sim ← sim + scr
        end
    end
    // Sort the sentences on sim
    article ← sentences[1:K]
    // Sort the sentences on order
    return article
```

to be retrieved, $\mathcal{A}^{I}_{1:k}$ and $\mathcal{A}^{C}_{1:k}$ are the resulting subsets containing k sentences selected based on visual and textual modalities, respectively. Here, $k = \mathcal{K}$ if $\mathcal{K} \leqslant n^{'}$, otherwise $k = n^{'}$. $\mathcal{L}$ specifies the language of the article.

We tokenize the documents into sentences using our `Multilingual Sentence Tokenizer` (introduced later in Section 4) and use images and captions as queries to compute semantic similarity with the sentences. This process employs a multilingual version of the **`OpenAI CLIP-ViT-B32`** model[4] (Radford et al., 2021), which maps text (*in 50+ languages*) and images to a shared dense vector space (Reimers and Gurevych, 2019), aligning images closely with their corresponding texts. We then retrieve the top $\mathcal{K}$ sentences from the document based on the similarity scores.

Handling Multiple Images and Captions We observed that a single document often contains multiple images and captions without a proper one-to-one mapping between them. Consequently, we treat each image and caption as distinct entities and propose a greedy algorithm for aggregating multiple retrievals. The detailed procedural depiction of one this simple yet effective algorithmic processes is presented in Algorithm 1 and detailed in Appendix B.

3.2 Instruction Tuning

Language models can be fine-tuned with supervised datasets containing natural language prompts and their corresponding target completions (Wei et al., 2022; Sanh et al., 2022; Ouyang et al., 2022; Min et al., 2022). This process, known as *"instruction tuning,"* enhances the models' ability to follow instructions accurately. Typically, task-specific prefixes are used to guide the model towards the desired output format. Inspired by the adaptability and success of these approaches in managing diverse tasks via a unified text-to-text framework, we apply this methodology to generate headlines and tags for news articles within a multilingual setting. We introduce two instructional variations: one for **unrestricted** and another for **controlled** tag word generation along with headline.

Unrestricted Generation allows the model to independently determine the optimal number of tag words to generate. In the following input instruction, **bold** indicates the input prefix, while *underline* signifies the selected content (described below). Conversely, in the output instruction, **bold** marks the output prefix, and *underline* denotes the generated output, encompassing both the headline and a variable number of tag words (T_1, T_2, ...) corresponding to each article. This approach aims to simultaneously address two text generation tasks, utilizing output prefixes to distinguish between the outcomes of each task. The complete instruction format is as follows:

Instruction for Unrestricted Generation

`Input` → **Generate Headline and Tag Words:** *Selected Content*.

`Output` → **Headline is:** *Headline*. **Tag words are:** $T_1, T_2, \cdots, T_o$.

Controlled Generation To tackle the challenge of determining the correct number of tag words, we have adjusted our unrestricted prefix. This modification allows us to directly specify the desired number of tag words in the prefix. Our revised input format for controlled tag word generation is:

Instruction for Controlled Generation

`Input` → **Generate Headline and $\mathcal{N}$ Tag Words:** *Selected Content*.

Here, $\mathcal{N}$ refers to the number of tag words to generate. During model training, this number is represented as the count of tag words associated with the original article from the training dataset, verbalized in natural language (such as One, Two, Three, ...). The output format remains unchanged.

Selected Content For the input format of the instruction, we utilize three settings: **(1)** Article,

[4] clip-ViT-B-32-multilingual-v1

where the original article is placed; **(2)** Top $\mathcal{K}$ sentences retrieved by our **MultiRet**; and **(3)** Top $\mathcal{K}$ sentences from our **MultiRet** concatenated with the Article. **MultiRet** comprises two modules: **ImgRet**, which retrieves relevant sentences using images as queries, and **CapRet**, which uses image captions as queries to retrieve pertinent sentences from the input article as detailed in Section §3.1.

4 Multilingual Tools

We also develop tools by accumulating open-source resources for processing and evaluating multilingual texts, making it an easy-to-use one-stop destination. These tools include **(1)** Multilingual ROUGE Scorer, **(2)** Multilingual Sentence Tokenizer, and **(3)** Multilingual Stemmer. These resources are invaluable to the research community focusing on multilingual NLP, providing essential support for accurate processing and evaluation[5] (also detailed in Appendix C).

Multilingual ROUGE Scorer Hasan et al. (2021); Chronopoulou et al. (2023) identified a significant issue in evaluating multilingual summarization performance: the absence of stemmers for certain low-resource languages hindered the processing of generated summaries, resulting in lower ROUGE scores (Lin, 2004). Facing a similar challenge, Aharoni et al. (2023) calculated ROUGE scores using a multilingual tokenizer. However, the lack of word tokenizers for some languages still presents a challenge for fair ROUGE score assessment across different language families. To address this issue, we developed a Multilingual ROUGE Scorer that leverages Byte-Pair Encoding (BPE) tokenization from BERT-multilingual (Devlin et al., 2019), ensuring more accurate evaluation across 104 languages.

Multilingual Sentence Tokenizer Sentence tokenization aims to divide a given document into individual sentences. To achieve this, we integrated various open-source resources for multiple languages into a unified library. This tool can perform tokenization in 41 different languages (details of the sources in Appendix (Table 6)).

Multilingual Stemmer For evaluating the performance of keyphrase generation, both generated and reference keyphrases are normalized before assessing exact matches (Meng et al., 2017; Chen et al., 2020). Given the need to evaluate generated tag words across various languages, we developed a Multilingual Stemmer that integrates open-source stemmers for 18 distinct languages (as detailed in Table 7 in the Appendix). However, open-source stemmers for **Chinese** and **Telugu** are unavailable. Therefore, when evaluating tag words in these languages, we report the scores without normalizing the tag words.

5 Evaluating Tags Generation

The terms *"Tag Words"* and *"Keyphrases,"* while sharing a common objective, differ in their application. In the keyphrase generation process, a model predicts a set of distinct keyphrases $\hat{\mathcal{Y}} = (\hat{y}_1, \ldots, \hat{y}_m)$ from a given source text, with these predictions $\hat{y}_i$ being ordered based on their relevance (Yuan et al., 2020). The ground truth keyphrases for the source text are denoted as $\mathcal{Y}$. It's important to note that in the context of tag words, the order of predictions does not necessarily reflect the quality of each prediction. To measure predictive performance, three standard evaluation metrics—namely macro-averaged precision, recall, and F-measure (F_1)—are commonly used (Meng et al., 2017). Formally, the metrics of precision, recall, and F_1 score are defined as follows:

$$P = \frac{|\hat{\mathcal{Y}} \cap \mathcal{Y}|}{|\hat{\mathcal{Y}}|}, \qquad R = \frac{|\hat{\mathcal{Y}} \cap \mathcal{Y}|}{|\mathcal{Y}|}, \qquad F_1 = \frac{2 * P * R}{P + R}. \tag{1}$$

5.1 Proposed Tag Words Evaluation Metrics

As detailed in Section 3.2, our work spans both controlled and unrestricted tag word generation. Inspired by Yuan et al. (2020), we introduce three metrics for assessing performance.

Unrestricted Generation In unrestricted generation, a varying number of tag words are generated. The evaluation employs the metric $F_1@\mathcal{M}$, where $|\hat{\mathcal{Y}}| = \mathcal{M}$. Here, $\mathcal{M}$ varies with each article, reflecting the model's autonomous decision on the number of tag words.

Controlled Generation For controlled generation, the goal is to generate a fixed number of tag words. We evaluate this using $F_1@\mathcal{K}$ and $F_1@\mathcal{O}$. $\mathcal{K}$ is predefined as 3 and 5, and $\mathcal{O}$ corresponds to $|\mathcal{Y}|$, the actual number of tag words. This means we assess controlled generation for producing ex-

[5] These tools are available here XL-HeadTags, and can be easily installed via 'pip'.

		R1	R2	RL	BLEU	METEOR	LR (↓)	BERT Score
Models		**Baselines**						
mT5		37.86	17.20	33.53	12.95	25.55	0.84	75.79
mT0		38.33	17.66	33.90	14.64	26.44	0.94	75.83
FLAN-T5		31.46	12.73	28.15	8.75	24.61	0.71	70.87
LEAD-1		14.86	5.48	11.36	2.30	11.84	3.99	65.59
EXT-ORACLE		25.90	15.29	21.57	6.13	20.11	2.96	69.33
Gemini-Pro		20.02	10.10	17.99	5.68	11.15	0.62	68.36
Mixtral		11.15	3.63	10.26	1.58	7.14	0.86	64.70
Modality	**Models**	**MultiRAGen (ours)**						
Text (*Caption*)	**mT5**							
	└ w/C (**K=5**)	39.06 (+1.20)	**18.35** (+1.15)	**34.64** (+1.11)	**14.19** (+1.24)	26.94 (+1.39)	0.87 (+0.03)	**76.23** (+0.44)
	└ w/C (**K=10**)	39.04 (+1.18)	18.20 (+1.00)	34.51 (+0.98)	14.03 (+1.08)	26.86 (+1.31)	0.87 (+0.03)	76.20 (+0.41)
	└ w/C (**K=15**)	**39.13** (+1.27)	18.30 (+1.10)	34.63 (+1.10)	14.16 (+1.21)	**26.99** (+1.44)	0.87 (+0.03)	76.22 (+0.43)
	mT0							
	└ w/C (**K=5**)	39.07 (+0.74)	18.27 (+0.61)	34.62 (+0.72)	14.29 (-0.35)	27.06 (+0.62)	0.88 (-0.06)	76.17 (+0.34)
	└ w/C (**K=10**)	**39.13** (+0.80)	**18.35** (+0.69)	34.61 (+0.71)	14.29 (-0.35)	**27.24** (+0.80)	0.88 (-0.06)	76.21 (+0.38)
	└ w/C (**K=15**)	**39.13** (+0.80)	18.30 (+0.64)	**34.63** (+0.73)	14.16 (-0.48)	26.99 (+0.55)	**0.87** (-0.07)	**76.22** (+0.39)
	FLAN-T5							
	└ w/C (**K=5**)	31.41 (-0.05)	12.66 (-0.07)	28.30 (+0.15)	8.52 (-0.23)	24.40 (-0.21)	**0.69** (-0.02)	70.88 (+0.01)
	└ w/C (**K=10**)	31.65 (+0.19)	12.80 (+0.07)	28.44 (+0.29)	8.64 (-0.11)	24.59 (-0.02)	0.70 (-0.01)	70.89 (+0.02)
	└ w/C (**K=15**)	**31.75** (+0.29)	**12.93** (+0.20)	**28.51** (+0.36)	8.74 (-0.01)	**24.75** (+0.14)	0.70 (-0.01)	**70.92** (+0.05)
Visual (*Image*)	**mT5**							
	└ w/I (**K=5**)	**39.17** (+1.31)	**18.41** (+1.21)	**34.77** (+1.24)	**14.36** (+1.41)	**27.03** (+1.48)	0.87 (+0.03)	**76.22** (+0.43)
	└ w/I (**K=10**)	38.94 (+1.08)	18.17 (+0.97)	34.44 (+0.91)	14.08 (+1.13)	26.87 (+1.32)	0.87 (+0.03)	76.18 (+0.39)
	└ w/I (**K=15**)	39.03 (+1.17)	18.19 (+0.99)	34.56 (+1.03)	14.01 (+1.06)	26.90 (+1.35)	0.87 (+0.03)	76.19 (+0.40)
	mT0							
	└ w/I (**K=5**)	**39.21** (+0.88)	**18.49** (+0.83)	**34.74** (+0.84)	14.51 (-0.13)	**27.26** (+0.82)	**0.88** (-0.06)	**76.26** (+0.43)
	└ w/I (**K=10**)	39.16 (+0.83)	18.33 (+0.67)	34.61 (+0.71)	14.27 (-0.37)	27.11 (+0.67)	**0.88** (-0.06)	76.22 (+0.39)
	└ w/I (**K=15**)	39.14 (+0.81)	18.38 (+0.72)	34.62 (+0.72)	14.27 (-0.37)	27.16 (+0.72)	0.89 (-0.05)	76.18 (+0.35)
	FLAN-T5							
	└ w/I (**K=5**)	**31.56** (+0.10)	12.76 (+0.03)	**28.38** (+0.23)	8.54 (-0.21)	24.49 (-0.12)	0.70 (-0.01)	70.89 (+0.02)
	└ w/I (**K=10**)	31.55 (+0.09)	**12.82** (+0.09)	**28.38** (+0.23)	8.65 (-0.10)	24.58 (-0.03)	**0.69** (-0.02)	**70.90** (+0.03)
	└ w/I (**K=15**)	31.46 (0.00)	12.70 (-0.03)	28.23 (+0.08)	8.55 (-0.20)	24.40 (-0.21)	0.70 (-0.01)	70.82 (-0.05)

Table 3: Headline Generation Evaluation. **Selected Content** (Important Sentences + Article). **K** denotes the number of sentences retrieved for both text and visual modalities. (↓) indicates lower values for better performance. The best results compared to their respective baseline models are marked in **bold**, and Δ gains are shown in round brackets and highlighted with green and red colors.

actly 3, 5, and $|\mathcal{Y}|$ tag words, as specified by a natural language prefix.

6 Experiments

6.1 Data & Evaluation Metrics

Data To ensure a balanced distribution, our dataset division for training, validation, and test sets across all languages consists of 95% (394,353 samples), 1% (5,187 samples), and 4% (15,577 samples) respectively from our **XL-HeadTags** dataset (detailed in Appendix (Table 2)). Our goal is to enhance the task's generalizability by developing a unified model capable of performing both **controlled** (*a fixed number of tag words*) and **unrestricted** (*the model decides the number of tag words*) tag word generation (also elaborated in Section 3.2). To achieve this, we introduce a **prefix mixture strategy** during training, using a 70:30 allocation ratio. Here, 70% of the data is formatted for controlled tag word generation, while the remaining 30% is free from such constraints. This approach, applied consistently across training, validation, and test datasets for both baseline and our models, aims to facilitate a model adept at navigating both task variations. Details on data processing are presented in Appendix A.

Evaluation Metrics For the evaluation of generated headlines, we utilize F1 score of our `Multilingual ROUGE Scorer` (§4) and the BLEU score (Papineni et al., 2002), F1 BERT score (Zhang* et al., 2020), METEOR score (Banerjee and Lavie, 2005), and Length Ratio (LR). For assessing tag words, we apply the metrics we have proposed in Section 5.1 and normalize the tag words using our `Multilingual Stemmer` (§4).

6.2 Models

Baselines We use `mT5-base` (Xue et al., 2021), `mT0-base` (Muennighoff et al., 2022) and `Flan-T5-large` (Chung et al., 2022) checkpoints available in the Hugging Face (Wolf et al., 2020). Notably, these models are multilingual and pretrained on the mC4[6] multilingual corpus. It is essential to highlight that the mT0 model underwent additional fine-tuning within a multitask frame-

[6] https://huggingface.co/datasets/mc4

work, utilizing the crosslingual task mixture xP3 (Muennighoff et al., 2022) dataset to enhance crosslingual generalization. We conducted fine-tuning of the mT5-base, mT0-base and Flan-T5-large models using the original article in the instruction. Additional baselines like LEAD-1 and EXT-ORACLE are detailed in Appendix D.

Gemini Pro and Mixtral We employed the Gemini Pro (Team, 2023) and Mixtral (Jiang et al., 2023) models for evaluating their efficacy in **XL-HeadTags** multilingual tasks. This assessment occurred in zero-shot prompting conditions, with instructions specifying input (*i.e., article*) and output formats (§3.2). This encompassed sampling 50 instances from each language.

MultiRAGen (*ours*) For visual modality, we use images, and for textual modality, we utilize image captions. The number of sentences to retrieve is determined by the parameter $\mathcal{K}$, with explored values of 5, 10, and 15, corresponding to the retrieval of the top 5, 10, and 15 sentences, respectively. After retrieval, we reorder the top-$\mathcal{K}$ sentences to their original sequence in the article, ensuring the narrative flow remains coherent. Our experimentation includes two approaches: **(1)** inputting only the top-$\mathcal{K}$ retrieved sentences and **(2)** combining these sentences with the original article. We apply the same set of hyperparameters for all the baselines and our models, as detailed in Appendix E.

7 Results and Discussion

The evaluation results for headline generation are detailed in Table 3, and the outcomes for tag word evaluation are shown in Table 5. We also compare the performance of our baseline models with three extractive methods—**TF-IDF** (Sparck Jones, 1988) and **TextRank** (Mihalcea and Tarau, 2004) as unsupervised, and **KEA** (Witten et al., 1999) as supervised—on the English test set. These extractive methods are implemented using the **PKE** module (Boudin, 2016), with a detailed comparison presented in Table 4.

7.1 Headline

Baselines Table 3 shows LEAD-1's poor performance, indicating its inability to capture the abstractive essence of headlines. EXT-ORACLE, however, provides a robust baseline with improved ROUGE and BLEU scores, though it lacks headline conciseness. mT5, mT0, and Flan-T5 outperform extractive methods, with mT0 closely matching the conciseness of reference headlines. Flan-T5 excels in high-resource languages but falls short in low-resource setting (also in Table 12, and 18).

Models	F_1@3	F_1@5	F_1@$\mathcal{M}$	F_1@$\mathcal{O}$
Extractive Unsupervised				
TF-IDF	10.9	11.4	-	10.85
TextRank	0.71	1.05	-	0.74
Extractive Supervised				
KEA	10.55	11.04	-	10.87
Abstractive				
Gemini-Pro	18.01	21.12	19.22	18.37
Mixtral	7.20	10.69	7.53	8.54
Abstractive Baselines-Ours				
mT5	44.51	37.91	45.34	47.80
mT0	51.52	43.57	54.36	57.00
Flan-T5	50.2	43.17	53.0	55.37

Table 4: Tag Words Evaluation on English.

MultiRAGen results show that using captions for retrieval with mT5 outperforms the standard mT5 baseline, a pattern also seen with mT0 and Flan-T5. This indicates that including the most relevant information in the context window enhances headline generation. Similarly, using images for retrieval with mT5, mT0, and Flan-T5 also surpasses their respective baselines, performing nearly as well as caption-based retrieval. This suggests potential benefits from integrating both textual and visual modalities in future research to improve retrieval effectiveness (See Limitations (§8)).

7.2 Tag Words

Baselines Table 4 compares tag word generation across baseline models and extractive methods (**TF-IDF**, **TextRank**, and **KEA**) on an English test corpus. Extractive methods, limited to identifying tag words already present in the text, demonstrate varied performance, with TextRank lagging due to its tendency to extract longer phrases, misaligned with the brevity of reference tags. In contrast, TF-IDF and KEA perform moderately better but are outshined by abstractive models like mT5, mT0, and Flan-T5, which excel in generating concise and relevant tag words, surpassing extractive approaches that cannot synthesize new tag words. This distinction highlights the superior capability of abstractive methods in handling the tag word generation task.

MultiRAGen Table 5 extends the tag word evaluation results to include baselines mT5, mT0, Flan-T5, Gemini Pro, Mixtral and MultiRAGen across a multilingual text corpus. The abstractive baselines exhibit commendable performance, with mT0

		F_1@3	F_1@5	F_1@$\mathcal{M}$	F_1@$\mathcal{O}$
Models		**Baselines**			
mT5		45.01	39.82	44.67	46.79
mT0		51.58	44.94	52.50	54.39
Flan-T5		30.76	26.3	31.86	33.4
Gemini-Pro		5.90	7.75	6.57	6.65
Mixtral		1.93	2.90	2.49	2.28
Modality	**Models**	**MultiRAGen (ours)**			
	mT5				
	├ w/C (**K=5**)	51.16 (+6.15)	45.18 (+5.36)	52.20 (+7.53)	54.36 (+7.57)
	├ w/C (**K=10**)	**53.08** (+8.07)	**47.00** (+7.18)	**54.00** (+9.33)	**56.24** (+9.45)
	├ w/C (**K=15**)	47.66 (+2.65)	42.56 (+2.74)	47.37 (+2.7)	49.96 (+3.17)
	mT0				
	├ w/C (**K=5**)	52.10 (+0.52)	46.41 (+1.47)	53.50 (+1.00)	55.62 (+1.23)
Text (*Caption*)	├ w/C (**K=10**)	53.88 (+2.30)	47.95 (+3.01)	**55.29** (+2.79)	**57.49** (+3.10)
	├ w/C (**K=15**)	**54.05** (+2.47)	**48.19** (+3.25)	55.18 (+2.68)	57.36 (+2.97)
	Flan-T5				
	├ w/C (**K=5**)	30.58 (-0.18)	26.12 (-0.18)	31.48 (-0.38)	32.96 (-0.44)
	├ w/C (**K=10**)	31.18 (+0.42)	26.65 (+0.35)	32.16 (+0.3)	33.77 (+0.37)
	├ w/C (**K=15**)	**31.48** (+0.72)	**26.90** (+0.60)	**32.4** (+0.54)	**34.00** (+0.60)
	mT5				
	├ w/I (**K=5**)	50.72 (+5.71)	44.60 (+4.78)	51.70 (+7.03)	53.52 (+6.73)
	├ w/I (**K=10**)	**53.62** (+8.61)	**47.57** (+7.75)	**54.76** (+10.09)	**56.95** (+10.16)
	├ w/I (**K=15**)	47.67 (+2.66)	42.39 (+2.57)	47.21 (+2.54)	49.80 (+3.01)
	mT0				
	├ w/I (**K=5**)	50.85 (-0.73)	44.84 (-0.10)	52.15 (-0.35)	53.81 (-0.58)
Visual (*Image*)	├ w/I (**K=10**)	**53.79** (+2.21)	**47.69** (+2.75)	**55.00** (+2.50)	**57.12** (+2.73)
	├ w/I (**K=15**)	53.45 (+1.87)	47.46 (+2.52)	54.43 (+1.93)	56.65 (+2.26)
	Flan-T5				
	├ w/I (**K=5**)	29.11 (-1.65)	24.63 (-1.67)	29.86 (-2.0)	31.14 (-2.26)
	├ w/I (**K=10**)	30.74 (-0.02)	26.25 (-0.05)	31.40 (-0.46)	33.21 (-0.19)
	├ w/I (**K=15**)	**31.54** (+0.78)	**26.80** (+0.5)	**32.49** (+0.63)	**34.24** (+0.84)

Table 5: Tags Words Evaluation. **Selected Content** (Only Important Sentences). **K** denotes the number of sentences retrieved for both text and visual modalities. The best results compared to their respective baseline models are marked in **bold**, and Δ gains are shown in round brackets and highlighted with green and red colors.

demonstrating a noticeably higher gain over mT5 and Flan-T5. This improvement is attributed to the further fine-tuning of mT0 on a multitask framework, enhancing its crosslingual generalization. Although Flan-T5 produces results similar to mT0 and mT5 in English (see Table 4), it falls short of generating any tag words when it comes to certain languages with limited resources. This disparity is particularly noticeable in Table 21.

7.3 Discussion

Gemini Pro and Mixtral exhibit notably low performance, indicating a tendency to generate headlines and tag words in English or romanized rather than the native language of the article. More details with *"Prompt for LLMs"* and example outputs are presented in Appendix F.

Better content selection approach Our experiments explored two methods: **(1)** using only top-$\mathcal{K}$ retrieved sentences and **(2)** top-$\mathcal{K}$ merging these with the original article. We found combining retrieved sentences with the full article improved headline generation (see Table 3 vs. 8), while using solely retrieved sentences was more effective for tag generation (see Table 5 vs. 9). This difference arises because tags, being more concise, benefit from focused inputs, whereas headlines may require broader context for optimal generation.

High-Resource vs. Low-Resource Languages Our model, **MultiRAGen**, achieves balanced performance across high-resource and low-resource languages. Detailed results, grouped by resource availability, are presented in the Appendix (from Table 10 to Table 21).

8 Conclusion

In this paper, we compile the **XL-HeadTags** dataset for headline and tag generation tasks, including 20 languages across 6 diverse language families. We introduce a novel content selection approach that leverages auxiliary information from both textual and visual modalities to pinpoint the most salient content within news articles. We employ instruction tuning to generate both headlines and tags in controlled and unrestricted manners. Furthermore, we have developed a suite of tools by accumulating open-source resources for processing and evaluating multilingual texts.

Limitations

Potential Bias in Dataset Source Our dataset exclusively comprises articles sourced from the BBC, which may introduce a bias toward specific narratives or ideologies, potentially impacting the dataset's representativeness and diversity.

Handling Multiple Images and Captions In our analysis, we noted that documents frequently contain several images and captions without a direct one-to-one correspondence. Ideally, a precise mapping between each image and its caption would enhance our retrieval process. However, due to the absence of such mappings, we treat each image and caption as separate entities for independent retrieval processes. Integrating both images and captions for simultaneous retrieval presents an opportunity for future research, potentially refining the information extraction process.

Computational Constraints Given the computational constraints and resource limitations, we opted for the base versions of models for our experimentation. Despite these constraints, we introduce a novel content selection strategy that leverages auxiliary information from both textual and visual modalities. This approach aims to pinpoint the most pertinent content in news articles across multiple languages. Our designed modules are `plug-and-play`, ensuring they can be effortlessly integrated with language models of varying sizes. This flexibility facilitates broader applicability and enhances the adaptability of our approach in diverse computational environments.

Ethics Statement

Data Crawling We took ethical consideration into account when scraping data. The data we have collected is intended exclusively for non-commercial research purposes. We conducted our web scraping activities at a reasonable rate, with no intention of causing a Distributed Denial of Service (**DDoS**) attack. Additionaly, we read the instructions listed in robots.txt[7] of each website to ensure we were able to crawl the desired content as per the Robots Exclusion Protocol (REP) standards[8].

Protection of Privacy We intentionally chose not to collect certain information, including the author name, the time when the article was written, and any personal contact details such as email addresses, phone numbers, etc, for the purposes of our experiments. Consequently, our dataset does not contain any Personal Identifying Information (**PII**). This decision underscores our commitment to placing user privacy as a top priority.

Acknowledgements

We thank all the anonymous reviewers for their valuable feedback and constructive suggestions for improving this work. Mir Tafseer Nayeem is supported by a Huawei PhD Fellowship.

References

Divyanshu Aggarwal, Ashutosh Sathe, and Sunayana Sitaram. 2024. Maple: Multilingual evaluation of parameter efficient finetuning of large language models. *Preprint*, arXiv:2401.07598.

Roee Aharoni, Shashi Narayan, Joshua Maynez, Jonathan Herzig, Elizabeth Clark, and Mirella Lapata. 2023. Multilingual summarization with factual consistency evaluation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 3562–3591, Toronto, Canada. Association for Computational Linguistics.

Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Mohamed Ahmed, Kalika Bali, and Sunayana Sitaram. 2023. MEGA: Multilingual evaluation of generative AI. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4232–4267, Singapore. Association for Computational Linguistics.

Abu Ubaida Akash, Mir Tafseer Nayeem, Faisal Tareque Shohan, and Tanvir Islam. 2023. Shironaam: Bengali news headline generation using auxiliary information. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 52–67, Dubrovnik, Croatia. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *Preprint*, arXiv:1409.0473.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor,

[7] `https://moz.com/learn/seo/robotstxt`

[8] The robots.txt file is part of the robots exclusion protocol (REP), a group of web standards regulating how robots crawl the web.

Michigan. Association for Computational Linguistics.

Rishi Bommasani and Claire Cardie. 2020. Intrinsic evaluation of summarization datasets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8075–8096, Online. Association for Computational Linguistics.

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George van den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, et al. 2021. Improving language models by retrieving from trillions of tokens. *arXiv preprint arXiv:2112.04426*.

Florian Boudin. 2016. pke: an open source python-based keyphrase extraction toolkit. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 69–73, Osaka, Japan.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Alexey Bukhtiyarov and Ilya Gusev. 2020. Advances of transformer-based models for news headline generation. In *Artificial Intelligence and Natural Language*, pages 54–61, Cham. Springer International Publishing.

Deng Cai, Yan Wang, Wei Bi, Zhaopeng Tu, Xiaojiang Liu, Wai Lam, and Shuming Shi. 2019. Skeleton-to-response: Dialogue generation guided by retrieval memory. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1219–1228, Minneapolis, Minnesota. Association for Computational Linguistics.

Ziqiang Cao, Wenjie Li, Sujian Li, and Furu Wei. 2017. Improving multi-document summarization via text classification. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 3053–3059. AAAI Press.

Yllias Chali, Moin Tanvee, and Mir Tafseer Nayeem. 2017. Towards abstractive multi-document summarization using submodular function-based framework, sentence compression and merging. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 418–424, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Hou Pong Chan, Wang Chen, Lu Wang, and Irwin King. 2019. Neural keyphrase generation via reinforcement learning with adaptive rewards. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2163–2174, Florence, Italy. Association for Computational Linguistics.

Jingqiang Chen and Hai Zhuge. 2018. Abstractive text-image summarization using multi-modal attentional hierarchical RNN. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4046–4056, Brussels, Belgium. Association for Computational Linguistics.

Wang Chen, Hou Pong Chan, Piji Li, and Irwin King. 2020. Exclusive hierarchical decoding for deep keyphrase generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1095–1105, Online. Association for Computational Linguistics.

Wenhu Chen, Hexiang Hu, Xi Chen, Pat Verga, and William Cohen. 2022. MuRAG: Multimodal retrieval-augmented generator for open question answering over images and text. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5558–5570, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Radia Rayan Chowdhury, Mir Tafseer Nayeem, Tahsin Tasnim Mim, Md. Saifur Rahman Chowdhury, and Taufiqul Jannat. 2021. Unsupervised abstractive summarization of Bengali text documents. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2612–2619, Online. Association for Computational Linguistics.

Alexandra Chronopoulou, Jonas Pfeiffer, Joshua Maynez, Xinyi Wang, Sebastian Ruder, and Priyanka Agrawal. 2023. Language and task arithmetic with parameter-efficient layers for zero-shot summarization. *Preprint*, arXiv:2311.09344.

Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling instruction-finetuned language models. *Preprint*, arXiv:2210.11416.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the*

North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186.

Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, and Haifeng Wang. 2015. Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1723–1732, Beijing, China. Association for Computational Linguistics.

Tanvir Ahmed Fuad, Mir Tafseer Nayeem, Asif Mahmud, and Yllias Chali. 2019. Neural sentence fusion for diversity driven abstractive multi-document summarization. *Computer Speech & Language*, 58:216–230.

Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1631–1640, Berlin, Germany. Association for Computational Linguistics.

Xiaotao Gu, Yuning Mao, Jiawei Han, Jialu Liu, You Wu, Cong Yu, Daniel Finnie, Hongkun Yu, Jiaqi Zhai, and Nicholas Zukoski. 2020. Generating representative headlines for news stories. In *Proceedings of The Web Conference 2020*, WWW '20, page 1773–1784, New York, NY, USA. Association for Computing Machinery.

Han Guo, Ramakanth Pasunuru, and Mohit Bansal. 2018. Soft layer-specific multi-task summarization with entailment and question generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 687–697, Melbourne, Australia. Association for Computational Linguistics.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. 2020. Retrieval augmented language model pre-training. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3929–3938. PMLR.

Tahmid Hasan, Abhik Bhattacharjee, Md. Saiful Islam, Kazi Mubasshir, Yuan-Fang Li, Yong-Bin Kang, M. Sohel Rahman, and Rifat Shahriyar. 2021. XL-sum: Large-scale multilingual abstractive summarization for 44 languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4693–4703, Online. Association for Computational Linguistics.

Qiuxiang He, Guoping Huang, Qu Cui, Li Li, and Lemao Liu. 2021. Fast and accurate neural machine translation with translation memory. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3170–3180, Online. Association for Computational Linguistics.

Tatsuru Higurashi, Hayato Kobayashi, Takeshi Masuyama, and Kazuma Murao. 2018. Extractive headline generation based on learning to rank for community question answering. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1742–1753, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Kango Iwama and Yoshinobu Kano. 2019. Multiple news headlines generation using page metadata. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 101–105, Tokyo, Japan. Association for Computational Linguistics.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. 2020. Generalization through memorization: Nearest neighbor language models. *Preprint*, arXiv:1911.00172.

Mateusz Krubiński and Pavel Pecina. 2023. MLASK: Multimodal summarization of video-based news articles. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 910–924, Dubrovnik, Croatia. Association for Computational Linguistics.

Aman Kumar, Himani Shrotriya, Prachi Sahu, Raj Dabre, Ratish Puduppully, Anoop Kunchukuttan, Amogh Mishra, Mitesh M Khapra, and Pratyush Kumar. 2022. Indicnlg suite: Multilingual datasets for diverse nlg tasks in indic languages. *arXiv preprint arXiv:2203.05437*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020a. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA. Curran Associates Inc.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020b. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Haoran Li, Junnan Zhu, Tianshang Liu, Jiajun Zhang, and Chengqing Zong. 2018. Multi-modal sentence summarization with modality attention and image filtering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4152–4158. International Joint Conferences on Artificial Intelligence Organization.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Fuxiao Liu, Yinghan Wang, Tianlu Wang, and Vicente Ordonez. 2021. Visual news: Benchmark and challenges in news image captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6761–6771, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. Lost in the middle: How language models use long contexts. *Preprint*, arXiv:2307.03172.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. *Preprint*, arXiv:1711.05101.

Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi. 2017. Deep keyphrase generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 582–592, Vancouver, Canada. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Sewon Min, Mike Lewis, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2022. MetaICL: Learning to learn in context. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2791–2809, Seattle, United States. Association for Computational Linguistics.

Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng-Xin Yong, Hailey Schoelkopf, et al. 2022. Crosslingual generalization through multitask finetuning. *arXiv preprint arXiv:2211.01786*.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018a. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018b. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807, Brussels, Belgium. Association for Computational Linguistics.

Mir Tafseer Nayeem and Yllias Chali. 2017a. Extract with order for coherent multi-document summarization. In *Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing*, pages 51–56, Vancouver, Canada. Association for Computational Linguistics.

Mir Tafseer Nayeem and Yllias Chali. 2017b. Paraphrastic fusion for abstractive multi-sentence compression generation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 2223–2226, New York, NY, USA. Association for Computing Machinery.

Mir Tafseer Nayeem, Tanvir Ahmed Fuad, and Yllias Chali. 2018. Abstractive unsupervised multi-document summarization using paraphrastic sentence fusion. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1191–1204, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Mir Tafseer Nayeem, Tanvir Ahmed Fuad, and Yllias Chali. 2019. Neural diverse abstractive sentence compression generation. In *Advances in Information Retrieval*, pages 109–116, Cham. Springer International Publishing.

Nelleke Oostdijk, Hans van Halteren, Erkan Başar, and Martha Larson. 2020. The connection between the text and images of news articles: New insights for multimedia analysis. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4343–4351, Marseille, France. European Language Resources Association.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Hao Peng, Ankur Parikh, Manaal Faruqui, Bhuwan Dhingra, and Dipanjan Das. 2019. Text generation with exemplar-based adaptive decoding. In *Proceedings of the 2019 Conference of the North American*

Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 2555–2565, Minneapolis, Minnesota. Association for Computational Linguistics.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. In *Journal of Machine Learning Research*.

Mathieu Ravaut, Aixin Sun, Nancy F. Chen, and Shafiq Joty. 2024. On context utilization in summarization with large language models. *Preprint*, arXiv:2310.10570.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.

Victor Sanh, Albert Webson, Colin Raffel, Stephen Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M Rush. 2022. Multitask prompted training enables zero-shot task generalization. In *International Conference on Learning Representations*.

Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. 2020. MLSUM: The multilingual summarization corpus. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8051–8067, Online. Association for Computational Linguistics.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Yun-Zhu Song, Hong-Han Shuai, Sung-Lin Yeh, Yi-Lun Wu, Lun-Wei Ku, and Wen-Chih Peng. 2020. Attractive or faithful? popularity-reinforced learning for inspired headline generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8910–8917.

Karen Sparck Jones. 1988. *A Statistical Interpretation of Term Specificity and Its Application in Retrieval*, page 132–142. Taylor Graham Publishing, GBR.

John Sweller. 2011. Chapter two - cognitive load theory. volume 55 of *Psychology of Learning and Motivation*, pages 37–76. Academic Press.

Sho Takase, Jun Suzuki, Naoaki Okazaki, Tsutomu Hirao, and Masaaki Nagata. 2016. Neural headline generation on Abstract Meaning Representation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1054–1059, Austin, Texas. Association for Computational Linguistics.

Jiwei Tan, Xiaojun Wan, and Jianguo Xiao. 2017. From neural sentence summarization to headline generation: A coarse-to-fine approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, page 4109–4115. AAAI Press.

Gemini Team. 2023. Gemini: A family of highly capable multimodal models. *Preprint*, arXiv:2312.11805.

Ottokar Tilk and Tanel Alumäe. 2017. Low-resource neural headline generation. In *Proceedings of the Workshop on New Frontiers in Summarization*, pages 20–26, Copenhagen, Denmark. Association for Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Yash Verma, Anubhav Jangra, Raghvendra Verma, and Sriparna Saha. 2023. Large scale multi-lingual multi-modal summarization dataset. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3620–3632, Dubrovnik, Croatia. Association for Computational Linguistics.

Jason Wei, Maarten Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V Le. 2022. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*.

Ian H. Witten, Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig G. Nevill-Manning. 1999. Kea: Practical automatic keyphrase extraction. *Preprint*, arXiv:cs/9902007.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Peng Xu, Chien-Sheng Wu, Andrea Madotto, and Pascale Fung. 2019. Clickbait? sensational headline generation with auto-tuned reinforcement learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3065–3075, Hong Kong, China. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Wenhao Yu, Chenguang Zhu, Zaitang Li, Zhiting Hu, Qingyun Wang, Heng Ji, and Meng Jiang. 2022. A survey of knowledge-enhanced text generation. *ACM Comput. Surv.*, 54(11s).

Xingdi Yuan, Tong Wang, Rui Meng, Khushboo Thaker, Peter Brusilovsky, Daqing He, and Adam Trischler. 2020. One size does not fit all: Generating and evaluating variable number of keyphrases. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7961–7975, Online. Association for Computational Linguistics.

Boning Zhang and Yang Yang. 2023. MediaHG: Rethinking eye-catchy features in social media headline generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5766–5777, Singapore. Association for Computational Linguistics.

Ruqing Zhang, Jiafeng Guo, Yixing Fan, Yanyan Lan, Jun Xu, Huanhuan Cao, and Xueqi Cheng. 2018. Question headline generation for news articles. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 617–626, New York, NY, USA. Association for Computing Machinery.

Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Jun Zhao, Zhihao Zhang, Luhui Gao, Qi Zhang, Tao Gui, and Xuanjing Huang. 2024. Llama beyond english: An empirical study on language capability transfer. *Preprint*, arXiv:2401.01055.

Ruochen Zhao, Hailin Chen, Weishi Wang, Fangkai Jiao, Do Long, Chengwei Qin, Bosheng Ding, Xiaobao Guo, Minzhi Li, Xingxuan Li, and Shafiq Joty. 2023. Retrieving multimodal information for augmented generation: A survey. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 4736–4756, Singapore. Association for Computational Linguistics.

Qingyu Zhou, Nan Yang, Furu Wei, and Ming Zhou. 2017. Selective encoding for abstractive sentence summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1095–1104, Vancouver, Canada. Association for Computational Linguistics.

Junnan Zhu, Haoran Li, Tianshang Liu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. 2018. MSMO: Multimodal summarization with multimodal output. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4154–4164, Brussels, Belgium. Association for Computational Linguistics.

Junnan Zhu, Qian Wang, Yining Wang, Yu Zhou, Jiajun Zhang, Shaonan Wang, and Chengqing Zong. 2019. NCLS: Neural cross-lingual summarization. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3054–3064, Hong Kong, China. Association for Computational Linguistics.

Supplementary Material: Appendices

A Data Processing & Statistics

A.1 Data Processing

The data processing involved the application of regular expressions `Regex`[9] to eliminate URLs, HTML tags, and emojis from the headline, article, and image captions. Furthermore, `Phonenumbers` package[10] was employed to specifically filter out phone numbers from the textual content.

A.2 Data Statistics

Following data processing, the dataset comprises 415,117 news samples, represented as tuples encompassing headline, article , images, image captions, and tag words. The task of headline generation is herein considered as a subproblem integral to extreme summarization. Given the dual nature of summarization extractive and abstractive, our objective is to craft abstractive headlines. In the pursuit of measuring abstractiveness, Narayan et al. (2018b) proposed evaluating the percentage of n-grams in the summary that do not find occurrence in the input article. Simultaneously, Bommasani and Cardie (2020) introduced compression as a metric for quantifying conciseness, expressed mathematically as:

$$\textbf{Compression}(H, A) = 1 - \frac{|H|}{|A|} \tag{2}$$

In this context, $|H|$ and $|A|$ represent the lengths of the headline and article, measured in words. The ideal headline is aimed to be informative, concise, and have a high ratio of novel n-grams.

The process of generating tag words is in line with Keyphrase generation. Extracting tag words seeks to capture existing words in the input document. However, it's crucial to note that extractive methods can't predict absent tag words, an important aspect of comprehensive tag word generation. In contrast, tag word generation aims to produce both present and absent tag words (Meng et al., 2017). The percentage of present tag words highlights the challenge of predicting tag words not present in the document.

In Table 2, we present a comprehensive set of quantitative statistics that includes both the previously mentioned metrics and additional ones. The average word count in articles is 902, surpassing the Transformer Encoder's contextual window capacity. Using the mt5 multilingual sentencepiece tokenizer with a vocabulary size of 250,112, the mean token count per article to 1631. This is noteworthy considering the mt5 encoder's context window is limited to 512 tokens, leading to truncation beyond this limit. This emphasizes the need for methodologies capable of capturing essential information from documents exceeding the encoder context limit.

Further analysis shows that, on average, articles require compression of 98.88% for effective headline representation, highlighting the significant challenge in abstractive headline generation. Additionally, we observe that the percentage of tag words in the document consistently falls below 50%, emphasizing the importance of generative approaches for keywords to overcome limitations posed by extraction methods. These findings enhance our understanding of the intricate dynamics of document structure and the challenges associated with information extraction and summarization.

B Handling Multiple Images and Captions

The primary distinction between our modules lies in the modalities of the auxiliary data they utilize. As discussed in Section 3.1, our goal is to retrieve salient and pertinent information from news articles. To achieve this, we developed two separate modules for retrieving the most relevant sentences, each leveraging auxiliary information from different modalities. Specifically, **`ImgRet`** uses images, and **`CapRet`** uses image captions to retrieve the most relevant sentences from a document, respectively.

We observed that a single document often contains multiple images and captions without a proper one-to-one mapping between them. Consequently, we treat each image and caption as distinct entities and propose a greedy algorithm for aggregating multiple retrievals. The following discussion predominantly focuses on the operations of the **`ImgRet`** module, with selective explication of the points where **`CapRet`** differs.

The **ImgRet** module processes input data con-

[9] https://docs.python.org/3/library/re.html

[10] https://pypi.org/project/phonenumbers/

sisting of the $\underline{Article}$, $\underline{Language}$, $\underline{Images}$, and a parameter denoted as $\underline{K}$ (indicating the top-k sentences for retrieval). Subsequently, the **SentenceSegmenter** is utilized to deconstruct the $\underline{Article}$ into individual sentences, employing both the $\underline{Article}$ and $\underline{Language}$ as inputs. Subsequent steps involve iteratively comparing each sentence with the images, accumulating similarity scores for each sentence. The sentences are then sorted based on these scores, and the top $\underline{K}$ sentences, representing the most similar ones, are selected. The final sorting is performed based on their original order in the $\underline{Article}$, yielding the most relevant sentences in their original sequence.

It is noteworthy that **CapRet** diverges from this procedure solely in the calculation of embeddings for image captions, as opposed to the image embeddings computed in the **ImgRet** module.

C Multilingual Tools

The goal of `Multilingual Sentence Tokenizer` is to split a provided document into separate sentences. The language scope and the open-source resources employed for this tool are detailed in Table 6.

`Multilingual Stemmer` is used for normalizing tag words. The languages supported and their respective implementation sources are outlined in Table 7.

D Baselines

D.1 Headline Generation Baselines

LEAD-1 In the field of generating news headlines, we often use a common benchmark called **LEAD-1** to set the minimum standard for the task (Kumar et al., 2022; Narayan et al., 2018b). To calculate LEAD-1 scores, we take the first sentence of the article as the system-generated headline and compare it with the original headline.

EXT-ORACLE On the flip side, there's another approach called **EXT-ORACLE**, which represents the best possible outcome when generating headlines using an extractive method (Kumar et al., 2022; Narayan et al., 2018b). In this case, we align a sentence from the input article with the reference headline, using the ROUGE-2 metric to assess how well they match up.

D.2 Tags Generation Baselines

TF-IDF approach (Sparck Jones, 1988) entails the ranking of extracted noun phrase candidates based on their term frequency and inverse document frequency within the provided documents.

Language	Source	Language	Source
English	NLTK	Slovak	PYSBD
Portuguese		Indonesian	Spacy
Turkish		Nepali	
Arabic	PYSBD	Ukrainian	
French		Chinese	
Spanish		Yoruba	
Persian		Vietnamese	
Russian		Thai	
Urdu		Slovenian	
Amharic		Sinhala	
Armenian		Bengali	BLTK
Bulgarian		Gujarati	IndicNLP
Polish		Hindi	
Dutch		Marathi	
Danish		Punjabi	
Burmese		Tamil	
Greek		Telugu	
Italian		Oriya	
Japanese		kannada	
German		Malayalam	
Kazakh			

Table 6: Multilingual Sentence Tokenizer supports 41 different languages.

TextRank In the case of **TextRank** (Mihalcea and Tarau, 2004), the methodology involves the conceptualization of words as web pages, followed by the application of the PageRank algorithm to identify keyphrases.

KEA On the other hand, **KEA** (Witten et al., 1999) employs lexical methods to identify candidate keyphrases, computes feature values for each candidate, and utilizes a machine learning algorithm to predict the suitability of candidates as keyphrases.

E Training & Hyperparameters

Given the computational constraints and resource limitations, we chose to work with the base versions of the models for our experiments. The fine-tuning process for the mT5-base and mT0-base models was carried out on a single RTX 3090 GPU, while the Flan-T5-large model was fine-tuned using an RTX A6000 GPU. We configured the encoder input token sequence length to 512 and the decoder output token sequence length to 64.

Language	Source
English	Snowball
Portuguese	
Spanish	
Russian	
French	
Arabic	
Tamil	
Indonesian	
Turkish	turkish-stemmer-python
Ukrainian	ukr_stemmer
Bengali	Bangla-stemmer
Persian	PersianStemmer-Python
Nepali	nepali-stemmer
Urdu	Urdu-Stemmer
Gujarati	Gujarati-NLP-Toolkit
Hindi	hindi_stemmer
Marathi	marathi_stemmer
Panjabi	Stemmer-in-punjabi

Table 7: Multilingual Stemmer Covers 18 Language.

The fine-tuning phase for the mT5-base, mT0-base, and Flan-T5-large models lasted for five epochs, employing a batch size of 8. We utilized the AdamW optimizer (Loshchilov and Hutter, 2019) throughout the training process to optimize the models. This setup was chosen to balance the trade-off between computational efficiency and model performance, ensuring effective utilization of available hardware resources.

F Performance of LLMs

Gemini (Team, 2023) and Mixtral (Jiang et al., 2023) are two of large language models used for various tasks. Google provides access to Gemini via an API[11] for free, while the Mistral AI Team has made Mixtral's weights[12] publicly available. We utilized Gemini-Pro and Mixtral models to assess their performance in a multilingual task called XL-HeadTags. This evaluation was conducted under zero-shot prompting conditions, where only input and output formats were provided as instructions. We examined 50 examples from each language.

Prompt for LLMs

Instruction

Generate Headline and some defined or undefined number for tag words from a given news article. When ask to generate defined number of tag words generate exact defined number of tag words seperated by commas (,).

A single tag word can be multiple word or single word. When asked to generate undefined number of tag words, you should generate the number of tag word you think appropriate.

While generating headline and tag words generate them in the language the article is given in.

Exemplars

For example, you will be given task in these input format "Generate Headline and Three Tag Words": "News article".
Output should be "Headline is: Generated Headline. Tag Words are: Tag1, Tag2, Tag3".

Another example "Generate Headline and Tag Words": "News article". Your output should be "Headline is: Generated Headline. Tag Words are: Tag1, Tag2, . . . , TagN".

Input: {...}

However, our analysis revealed that their performance is subpar compared to a supervised approach. Several factors may have contributed to this. Firstly, both models heavily favor English. Even when presented with non-English articles and instructed to generate in the respective languages, they often produce results in English due to the dominance of English in their pre-training corpus.

Additionally, Mistral occasionally generates output in a romanized version of the target language, indicating a lack of effort in adapting to non-English contexts. We also encountered issues where the models inconsistently generated headlines and tags in different languages or failed to follow instructions, resulting in excessive tag words or overly lengthy headlines.

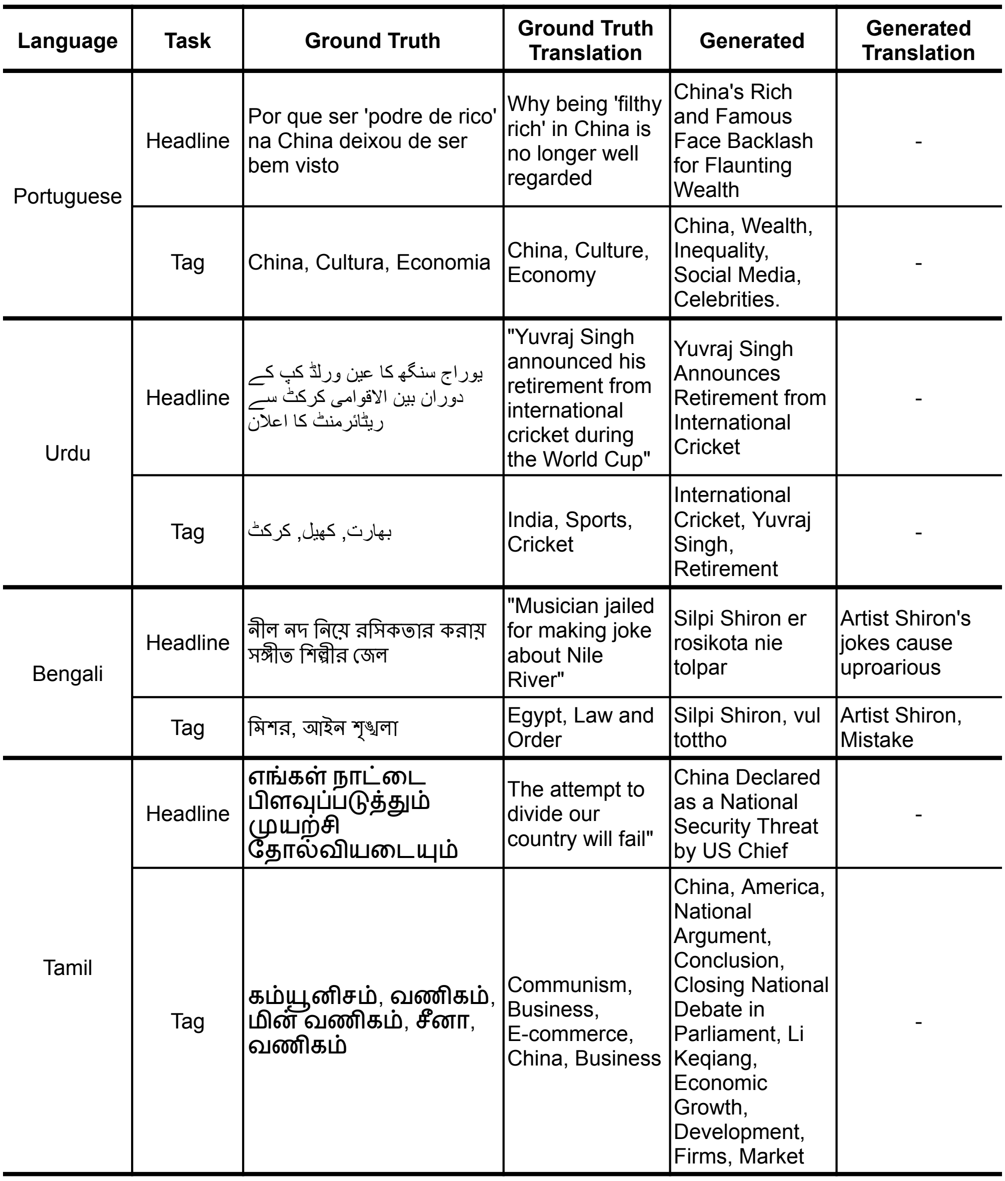
G Detailed Related Work

G.1 Headline Generation

Headline generation, a niche within abstractive summarization (Chali et al., 2017; Nayeem and Chali, 2017a; Fuad et al., 2019; Nayeem and Chali, 2017b; Nayeem et al., 2019), particularly in multilingual contexts (Chowdhury et al., 2021), has undergone notable advancements. In English, Rush et al. (2015) introduced an attention-based neural network model that employs a recurrent neural network (RNN) and attention mechanism for concise

[11] Gemini Pro API

[12] Mixtral



Language	Task	Ground Truth	Ground Truth Translation	Generated	Generated Translation
Portuguese	Headline	Por que ser 'podre de rico' na China deixou de ser bem visto	Why being 'filthy rich' in China is no longer well regarded	China's Rich and Famous Face Backlash for Flaunting Wealth	-
	Tag	China, Cultura, Economia	China, Culture, Economy	China, Wealth, Inequality, Social Media, Celebrities.	-
Urdu	Headline	یوراج سنگھ کا عین ورلڈ کپ کے دوران بین الاقوامی کرکٹ سے ریٹائرمنٹ کا اعلان	"Yuvraj Singh announced his retirement from international cricket during the World Cup"	Yuvraj Singh Announces Retirement from International Cricket	-
	Tag	بھارت, کھیل, کرکٹ	India, Sports, Cricket	International Cricket, Yuvraj Singh, Retirement	-
Bengali	Headline	নীল নদ নিয়ে রসিকতার করায় সঙ্গীত শিল্পীর জেল	"Musician jailed for making joke about Nile River"	Silpi Shiron er rosikota nie tolpar	Artist Shiron's jokes cause uproarious
	Tag	মিশর, আইন শৃঙ্খলা	Egypt, Law and Order	Silpi Shiron, vul tottho	Artist Shiron, Mistake
Tamil	Headline	எங்கள் நாட்டை பிளவுப்படுத்தும் முயற்சி தோல்வியடையும்	The attempt to divide our country will fail"	China Declared as a National Security Threat by US Chief	-
	Tag	கம்யூனிசம், வணிகம், மின் வணிகம், சீனா, வணிகம்	Communism, Business, E-commerce, China, Business	China, America, National Argument, Conclusion, Closing National Debate in Parliament, Li Keqiang, Economic Growth, Development, Firms, Market	-

Figure 3: Large Language Models (LLMs) generated samples.

summarization. Takase et al. (2016) developed an Abstract Meaning Representation (AMR) encoder within an encoder-decoder framework for headline creation, while Zhang et al. (2018) proposed a dual-attention sequence-to-sequence model tailored for question-driven headline generation. In contexts with limited resources, Tilk and Alumäe (2017) demonstrated the efficacy of pre-training neural models to enhance headline generation capabilities.

Zhou et al. (2017) approach to headline generation encompasses a tri-phase process involving sentence encoding, sentence selection via a gate network, and headline decoding. Similarly, Tan et al. (2017) introduced a method that prioritizes important sentences for context-based headline generation. On the multilingual front, Kumar et al. (2022) launched IndicNLG, a dataset aimed at headline generation, though it primarily focuses on headline-article pairs without incorporating auxiliary information.

Emerging methodologies in headline generation also explore specialized applications, such as generating headlines for community question answering (Higurashi et al., 2018) and creating multiple headlines for varied contexts (Iwama and Kano, 2019), broadening the scope of research in this field.

G.2 Keyphrases Generation

While "Tag Words" and "Keyphrases" both serve to enhance information retrieval, their application and research emphasis differ significantly. Tag word generation remains underexplored compared to the more developed field of keyphrase generation. For example, Meng et al. (2017) unveiled the CopyRNN model, which leverages an attentional encoder-decoder architecture (Bahdanau et al., 2014) integrated with a copying mechanism (Gu et al., 2016) to predict keyphrases. This method employs over-generation with an extensive beam search to then select the top N predictions, such as the top five or ten keyphrases.

Yuan et al. (2020) took an alternative approach by designing a Keyphrase Generation model that not only predicts multiple keyphrases but also determines the appropriate number of keyphrases for each document. This innovative training scheme enables the model to adaptively generate a variable number of keyphrases tailored to the specifics of each document. Further advancing the field, Chan et al. (2019) introduced the application of Reinforcement Learning to optimize the keyphrase generation process, showcasing the evolving techniques aimed at improving the precision and adaptability of keyphrase generation.

G.3 Multitask Learning and Instruction Tuning

In the realm of NLP, multitask learning (MTL) is increasingly recognized for its potential to enhance model performance on related tasks by exploiting their shared features and distinctions. For example, Cao et al. (2017) developed a model that simultaneously trains on summary generation and text classification, achieving notable improvements in text summarization. Dong et al. (2015) pioneered a multitask learning approach using a sequence-to-sequence (Seq2Seq) framework for translating a single source language into multiple target languages. Zhu et al. (2019) explored the synergies between monolingual summarization, machine translation, and cross-lingual summarization through joint training. Similarly, Guo et al. (2018) proposed a model that concurrently learns abstractive summarization and question generation. Extending the scope of multitask learning, Sanh et al. (2022) demonstrated the effectiveness of prompted multitask fine-tuning on a pre-trained T5 model (Raffel et al., 2020) for zero-shot task generalization.

Furthermore, language models can undergo fine-tuning on supervised datasets comprising natural language prompts paired with their respective target outputs. This method, termed **"instruction tuning,"** significantly refines the models' proficiency in adhering to instructions. Employing task-specific prefixes, this technique directs the model to generate outputs in a predetermined format, showcasing the versatility and precision attainable through instruction-based training.

G.4 Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) represents a pivotal advancement in Natural Language Generation (NLG), addressing the issue of neural models' limited contextual understanding. Traditional neural models often falter when the input lacks comprehensive information for generating accurate outputs, particularly in complex real-world applications (Yu et al., 2022). To bridge this gap, KNNLM (Khandelwal et al., 2020) introduced a technique for augmenting language models with examples retrieved from a training text dataset, enhancing contextual relevance. Building on this, RETRO (Borgeaud et al., 2021) leveraged a vastly expanded text corpus, enabling models with a smaller footprint to achieve performance on par with GPT-3 (Brown et al., 2020).

Models such as REALM (Guu et al., 2020) and RAG (Lewis et al., 2020b) incorporate Wikipedia passages as external knowledge bases, significantly boosting their efficacy in tasks like Question Answering. REALM focuses on encoding information through masked language modeling, whereas RAG employs an encoder-decoder structure for generative language tasks.

Expanding on these concepts, MuRAG (Chen et al., 2022) stands out by integrating multimodal knowledge sources, encompassing both visual and textual data. This innovation extends the capabilities of knowledge-enhanced text generation, catering to the nuanced demands of intricate information landscapes.

		R1	R2	RL	BLEU	METEOR	LR	BERT Score
Models		**Baselines**						
mT5		37.86	17.20	33.53	12.95	25.55	0.84	75.79
mT0		38.33	17.66	33.90	14.64	26.44	0.94	75.83
FLAN-T5		31.46	12.73	28.15	8.75	24.61	0.71	70.87
LEAD-1		14.86	5.48	11.36	2.30	11.84	3.99	65.59
EXT-ORACLE		25.90	15.29	21.57	6.13	20.11	2.96	69.33
Modality	**Models**	**MultiRAGen (ours)**						
Text (*Caption*)	**mT5**							
	└ w/C **(K=5)**	36.93 (-0.93)	16.33 (-0.87)	32.65 (-0.88)	13.23 (+0.28)	25.28 (-0.27)	0.92 (+0.08)	75.40 (-0.39)
	└ w/C **(K=10)**	38.51 (+0.65)	17.66 (+0.46)	34.02 (+0.49)	**14.51** (+1.56)	**26.63** (+1.08)	0.93 (+0.09)	75.93 (+0.14)
	└ w/C **(K=15)**	**38.74** (+0.88)	**18.02** (+0.82)	**34.23** (+0.70)	13.99 (+1.04)	**26.63** (+1.08)	0.88 (+0.04)	**76.04** (+0.25)
	mT0							
	└ w/C **(K=5)**	37.32 (-1.01)	16.67 (-0.99)	32.93 (-0.97)	13.53 (-1.11)	25.75 (-0.69)	0.93 (-0.01)	75.52 (-0.31)
	└ w/C **(K=10)**	**39.08** (+0.75)	**18.22** (+0.56)	**34.51** (+0.61)	**15.04** (+0.40)	**27.34** (+0.90)	0.94 (0.00)	**76.12** (+0.29)
	└ w/C **(K=15)**	38.76 (+0.43)	18.03 (+0.37)	34.22 (+0.32)	13.99 (-0.65)	26.63 (+0.19)	**0.88** (-0.06)	76.04 (+0.21)
	FLAN-T5							
	└ w/C **(K=5)**	30.64 (-0.82)	12.01 (-0.72)	27.52 (-0.63)	8.00 (-0.75)	23.63 (-0.98)	0.96 (+0.25)	70.57 (-0.30)
	└ w/C **(K=10)**	31.43 (-0.03)	12.69 (-0.04)	28.26 (+0.11)	8.59 (-0.16)	24.43 (-0.18)	**0.70** (-0.01)	70.88 (+0.01)
	└ w/C **(K=15)**	**31.65** (+0.19)	**12.85** (+0.12)	**28.44** (+0.29)	**8.78** (+0.03)	**24.63** (+0.02)	**0.70** (-0.01)	**70.92** (+0.05)
Visual (*Image*)	**mT5**							
	└ w/I **(K=5)**	35.72 (-2.14)	15.37 (-1.83)	31.62 (-1.91)	12.49 (-0.46)	24.19 (-1.36)	0.91 (+0.07)	75.00 (-0.79)
	└ w/I **(K=10)**	38.24 (+0.38)	17.51 (+0.31)	33.85 (+0.32)	**14.21** (+1.26)	**26.56** (+1.01)	0.92 (+0.08)	75.82 (+0.03)
	└ w/I **(K=15)**	**38.56** (+0.70)	**17.80** (+0.60)	**34.00** (+0.47)	13.64 (+0.69)	26.55 (+1.00)	0.87 (+0.03)	**76.00** (+0.21)
	mT0							
	└ w/I **(K=5)**	35.76 (-2.57)	15.38 (-2.28)	31.56 (-2.34)	12.59 (-2.05)	24.30 (-2.14)	0.92 (-0.02)	74.98 (-0.85)
	└ w/I **(K=10)**	38.37 (+0.04)	17.59 (-0.07)	33.86 (-0.04)	14.37 (-0.27)	**26.74** (+0.30)	0.93 (-0.01)	75.84 (+0.01)
	└ w/I **(K=15)**	**38.57** (+0.24)	**17.80** (+0.14)	**34.00** (+0.10)	13.64 (-1.00)	26.55 (+0.11)	0.87 (-0.07)	**76.00** (+0.17)
	FLAN-T5							
	└ w/I **(K=5)**	29.08 (-2.38)	10.77 (-1.96)	26.14 (-2.01)	7.14 (-1.61)	21.87 (-2.74)	**0.68** (-0.03)	70.07 (-0.80)
	└ w/I **(K=10)**	30.78 (-0.68)	12.14 (-0.59)	27.67 (-0.48)	8.16 (-0.59)	23.69 (-0.92)	0.69 (-0.02)	70.61 (-0.26)
	└ w/I **(K=15)**	31.37 (-0.09)	12.63 (-0.10)	28.13 (-0.02)	8.58 (-0.17)	24.40 (-0.21)	0.71 (0.00)	70.84 (-0.03)

Table 8: Headline Generation Evaluation. **Selected Content** (Only Important Sentences). The best results compared to their respective baseline models are marked in **bold**, and Δ gains are shown in round brackets and highlighted with green and red colors.

		F$_1$@3	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
Models		**Baselines**			
mT5		45.01	39.82	44.67	46.79
mT0		51.58	44.94	52.50	54.39
Flan-T5		30.76	26.30	31.86	33.40
Gemini-Pro		5.90	7.75	6.57	6.65
Mixtral		1.93	2.90	2.49	2.28
Modality	**Models**	**MultiRAGen (ours)**			
Text (*Caption*)	**mT5**				
	├ w/C (**K=5**)	48.52 (+3.51)	**43.15** (+3.33)	48.72 (+4.05)	50.83 (+4.04)
	├ w/C (**K=10**)	48.45 (+3.44)	43.11 (+3.29)	48.83 (+4.16)	50.86 (+4.07)
	├ w/C (**K=15**)	**48.73** (+3.72)	43.13 (+3.31)	**48.88** (+4.21)	**51.05** (+4.26)
	mT0				
	├ w/C (**K=5**)	48.61 (-2.97)	42.99 (-1.95)	48.63 (-3.87)	50.77 (-3.62)
	├ w/C (**K=10**)	48.54 (-3.04)	43.17 (-1.77)	48.33 (-4.17)	50.69 (-3.70)
	├ w/C (**K=15**)	48.73 (-2.85)	43.30 (-1.64)	48.69 (-3.81)	50.95 (-3.44)
	FLAN-T5				
	├ w/C (**K=5**)	31.29 (+0.53)	26.71 (+0.41)	32.33 (+0.47)	34.02 (+0.62)
	├ w/C (**K=10**)	**31.45** (+0.69)	**26.83** (+0.53)	**32.64** (+0.78)	**34.07** (+0.67)
	├ w/C (**K=15**)	31.21 (+0.45)	26.82 (+0.52)	32.53 (+0.67)	34.06 (+0.66)
Visual (*Image*)	**mT5**				
	├ w/I (**K=5**)	48.63 (+3.62)	43.01 (+3.19)	48.66 (+3.99)	50.82 (+4.03)
	├ w/I (**K=10**)	48.76 (+3.75)	**43.32** (+3.50)	**48.80** (+4.13)	50.97 (+4.18)
	├ w/I (**K=15**)	**48.79** (+3.78)	43.24 (+3.42)	48.63 (+3.96)	**51.15** (+4.36)
	mT0				
	├ w/I (**K=5**)	48.62 (-2.96)	43.10 (-1.84)	48.49 (-4.01)	50.93 (-3.46)
	├ w/I (**K=10**)	48.85 (-2.73)	43.39 (-1.55)	48.71 (-3.79)	51.10 (-3.29)
	├ w/I (**K=15**)	48.81 (-2.77)	43.31 (-1.63)	48.73 (-3.77)	51.03 (-3.36)
	FLAN-T5				
	├ w/I (**K=5**)	31.44 (+0.68)	26.73 (+0.43)	32.51 (+0.65)	33.94 (+0.54)
	├ w/I (**K=10**)	31.40 (+0.64)	**27.00** (+0.70)	32.50 (+0.64)	34.00 (+0.60)
	├ w/I (**K=15**)	**31.49** (+0.73)	26.89 (+0.59)	**32.74** (+0.88)	**34.24** (+0.84)

Table 9: Tag Words Evaluation. **Selected Content** (Important Sentence + Article). The best results compared to their respective baseline models are marked in **bold**, and Δ gains are shown in round brackets and highlighted with green and red colors.

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	16.35	36.36	14	30.3	16.91	37.06	14.43	31.05	17.1	37.05	13.78	31.16
	Portuguese	14.21	28.37	11.98	23.61	14.95	29.14	13.69	25.12	15.88	29.53	11.91	23.91
	Spanish	15.18	30.23	13.02	27.02	17.1	32.05	14.38	28.99	16.93	31.88	13.72	28.66
	Russian	13.66	26.35	8.44	17.61	15.47	28.14	9.8	19.49	15.42	28.06	9.15	19.47
	French	15.96	32.8	12.56	26.12	16.57	33.47	13.58	27.35	16.07	32.96	12.1	27.07
	Chinese	16.92	29.77	10.48	19.71	19.07	32.18	12.33	20.09	20.35	33.54	12.17	20.44
	Arabic	15.57	27.56	12.41	21.44	18.81	30.94	14.91	24.64	19.36	31.76	14.5	24.82
	All	15.41	30.21	11.84	23.69	16.98	31.85	13.3	25.25	17.3	32.11	12.48	25.08
Low	Turkish	16.71	30.95	13.13	20.52	18.6	32.95	15.05	22.67	19.55	33.53	14.91	22.71
	Ukrainian	13.56	26.12	8.14	17.4	15.47	27.59	9.43	19.05	16.01	28.41	8.84	18.86
	Bengali	21.04	31.79	14.91	18.15	23.5	34.16	16.73	20.05	25.17	35.3	17.44	20.43
	Persian	19.39	32.38	17.75	24.51	20.03	33.31	18.12	25.28	20.85	34.52	17.84	25.59
	Nepali	21.53	33.11	16.98	20.3	24.02	34.71	19.07	23.14	24.45	34.78	18.1	21.53
	Urdu	17.86	29.8	15.34	23.05	19.78	31.96	17.46	25.85	21.07	32.95	16.93	26.16
	Gujarati	17.64	29.3	13.34	15.87	20.71	32.95	16.03	18.72	21.4	32.02	14.96	17.55
	Hindi	18.42	30.97	15.47	22.41	19.86	32.42	16.02	23.75	21.88	34.75	17.57	25
	Marathi	16.73	27.56	14.31	17.99	19.17	30.44	16.13	19.24	20.3	30.7	16.23	19.17
	Telugu	15.32	27.28	12.25	15.66	16.79	28.4	13.79	17.16	17.46	29.43	13.79	16.68
	Tamil	21.47	31.78	13.18	18.69	24.3	34.3	15.07	21.01	23.71	33.75	13.56	19.05
	Panjabi	16.23	28.73	14.3	17.52	17.8	30.18	15.83	18.42	18.83	30.83	16.18	18.65
	Indonesian	12.98	27.5	9.09	22.13	15.5	30.45	10.79	24.76	15.68	30.5	9.91	25.13
	All	17.61	29.79	13.71	19.55	19.66	31.83	15.35	21.47	20.49	32.42	15.1	21.27
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	14.7	34.45	12.59	28.06	16.71	36.75	14.26	30.74	16.9	36.9	13.56	30.97
	Portuguese	12.97	27.35	11.37	22.58	14.62	29.34	12.23	23.36	14.84	29.02	11.72	23.27
	Spanish	15.28	30.2	13.11	27.16	16.97	31.93	14.25	29.24	16.81	31.74	13.37	28.75
	Russian	13.95	26.9	8.77	17.75	16.05	28.46	10.3	20.06	15.92	28.41	9.52	19.93
	French	14.7	31.49	12.07	25.37	15.71	32.85	13.02	26.85	15.47	31.67	11.57	26.41
	Chinese	17.5	30.88	10.66	19.59	18.21	31.59	11.32	19.66	19.42	33.31	11.87	19.79
	Arabic	16.75	28.34	13.19	22.87	19.36	31.49	15.38	24.97	20.18	32.73	15.46	25.93
	All	15.12	29.94	11.68	23.34	16.8	31.77	12.97	24.98	17.08	31.97	12.25	25.01
Low	Turkish	18.18	32.61	13.91	21.91	20.56	34.87	16.42	24.67	20.91	34.84	15.45	24.39
	Ukrainian	14.72	26.8	8.88	18.08	16.73	29.23	10.37	20.25	16.3	28.72	9.22	19.38
	Bengali	18.91	29.66	13.04	17.15	22.19	33.01	15.53	19.93	23.1	34.02	15.48	19.47
	Persian	16.7	30	14.45	21.41	19.78	32.68	17.56	24.66	21.58	34.96	18.72	26.33
	Nepali	22.45	33.41	18.47	21.88	23.88	34.79	18.22	22.22	24.38	35.25	17.99	22.11
	Urdu	18.75	30.51	15.97	24.01	20.87	32.88	17.59	26.73	20.96	32.99	16.64	25.93
	Gujarati	17.55	29.67	12.5	15.63	19.93	31.65	15.11	17.52	20.47	31.66	14.67	16.72
	Hindi	18.14	31.47	14.29	22.19	21.83	34.85	16.88	24.81	21.05	33.82	16.28	24.64
	Marathi	15.95	27.05	13.13	17.75	18.7	30.26	15.38	19.76	19.42	30.51	15.38	19.47
	Telugu	12.82	24.97	10.01	14.58	15.42	27.35	12.18	16.72	16.44	28.46	12.36	16.55
	Tamil	18.61	29.36	11.58	17.5	22.31	32.14	13.42	19.81	22.84	33.39	12.97	18.8
	Panjabi	12.31	25.68	11.14	14.23	15.19	27.78	13.53	16.69	16.24	29.01	13.15	16.16
	Indonesian	13.52	28.45	9.24	23.96	15.97	30.96	11.14	25.99	16.54	31.16	10.32	26.23
	All	16.82	29.2	12.82	19.25	19.49	31.73	14.87	21.52	20.02	32.21	14.51	21.24

Table 10: Headline Generation Evaluation : **mT5**. **Selected Content** (Only Important Sentences)

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	16.78	36.77	14.33	30.9	17.55	37.56	15.03	31.82	17.11	37.05	13.78	31.16
	Portuguese	14.58	28.64	12.87	24.12	15.37	30.05	13.61	25.47	15.92	29.57	11.91	23.91
	Spanish	15.01	30.08	12.92	26.99	17.54	32.53	14.92	29.77	16.96	31.91	13.72	28.66
	Russian	13.34	25.95	8.33	17.37	15.44	28.05	9.88	19.65	15.46	28.03	9.15	19.47
	French	16.04	32.59	12.78	26.48	16.43	33.09	13.14	27.39	16.12	32.9	12.1	27.07
	Chinese	17.24	29.72	10.6	19.79	19.35	32.84	12.34	20.18	20.3	33.48	12.17	20.44
	Arabic	16.9	27.87	13.34	22.79	19.88	31.27	15.21	25.03	19.37	31.82	14.5	24.82
	All	15.7	30.23	12.17	24.06	17.37	32.2	13.45	25.62	17.32	32.11	12.48	25.08
Low	Turkish	17.02	31.16	13.71	21.79	18.93	32.98	15.16	23.55	19.68	33.61	14.91	22.71
	Ukrainian	13.87	26.17	8.36	17.51	16.27	28.86	10.15	20.13	16.01	28.4	8.84	18.86
	Bengali	21.2	32.24	15.28	18.92	24.43	34.99	18.1	22.05	25.12	35.28	17.44	20.43
	Persian	19.86	33.02	18.32	24.2	20.92	34.67	19.18	26	20.94	34.47	17.84	25.59
	Nepali	22.06	33.15	17.4	21.13	23.87	34.52	18.91	23.25	24.37	34.78	18.1	21.53
	Urdu	18.25	30.03	15.83	23.73	20.94	32.83	18.21	26.81	21.04	32.98	16.93	26.16
	Gujarati	17.51	29.84	12.64	15.58	21.7	33.14	16.63	18.76	21.44	32.01	14.96	17.55
	Hindi	19.22	32.19	15.45	22.61	21.63	33.97	17.57	25.49	21.96	34.74	17.57	25
	Marathi	17.25	28.11	14.37	18.58	20.63	31.27	17.5	20.84	20.25	30.74	16.23	19.17
	Telugu	16.33	28.39	13.13	16.92	17.32	28.93	14.44	18.28	17.41	29.43	13.79	16.68
	Tamil	21.02	31.37	13.02	18.45	24.01	34.09	15.08	20.63	23.74	33.73	13.56	19.05
	Panjabi	16.19	28.87	14.76	17.37	18.15	30.37	16.26	18.54	18.81	30.78	16.18	18.65
	Indonesian	13.97	28.35	9.55	23.21	16.31	31.32	11.49	25.76	15.72	30.44	9.91	25.13
	All	17.89	30.42	13.95	20.41	20.39	32.98	16.05	22.31	20.5	32.41	15.1	21.27
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	14.7	34.28	12.61	28.21	16.95	36.84	14.43	31.07	16.91	36.89	13.56	30.97
	Portuguese	13.48	27.68	11.71	22.43	14.53	28.9	12.82	23.59	14.88	29.05	11.72	23.27
	Spanish	15.27	30.49	12.98	27.33	16.85	31.83	14.41	29.33	16.8	31.77	13.37	28.75
	Russian	14.06	26.87	8.82	18.08	15.48	28.01	10	19.79	15.92	28.43	9.52	19.93
	French	14.83	31.59	12.09	25.69	16.93	33.99	13.28	27.5	15.39	31.73	11.57	26.41
	Chinese	17.3	30.76	10.84	19.67	19.14	32.53	12.25	19.86	19.41	33.4	11.87	19.79
	Arabic	17.23	29.17	13.89	23.24	19.62	31.41	15.91	25.88	20.12	32.76	15.46	25.93
	All	15.27	30.12	11.85	23.52	17.07	31.93	13.3	25.29	17.06	32	12.44	25.01
Low	Turkish	18.37	32.49	14.48	21.88	20.64	34.62	16.46	25.17	20.91	34.86	15.45	24.39
	Ukrainian	14.14	26.61	8.53	17.72	16.57	29.03	10.41	20.34	16.27	28.71	9.22	19.38
	Bengali	18.46	29.32	12.92	17.33	21.72	32.5	15.57	19.9	23.14	33.97	15.48	19.47
	Persian	15.96	29.27	14.43	20.76	20.2	33.27	18.38	24.99	21.61	34.94	18.72	26.33
	Nepali	21.4	32.41	17.04	20.61	23.41	34.45	18.07	22.32	24.4	35.13	17.99	22.11
	Urdu	18.61	30.78	16.05	24.07	20.6	32.59	17.88	26.61	20.91	32.97	16.64	25.93
	Gujarati	17.91	30.15	12.8	15.27	19.77	31.48	14.71	17.88	20.55	31.69	14.67	16.72
	Hindi	18.19	31.18	14.8	22.47	21.76	34.26	17.79	25.89	21.15	33.75	16.28	24.64
	Marathi	16.39	27.41	13.66	18.34	18.8	29.73	15.85	19.98	19.46	30.53	15.38	19.47
	Telugu	12.84	25	9.88	14.55	15.53	27.6	12.14	16.86	16.47	28.48	12.36	16.55
	Tamil	18.98	29.72	11.63	17.38	21.83	32.36	13.6	19.51	22.84	33.4	12.97	18.8
	Panjabi	12.97	25.95	11.5	14.65	15.02	28.06	13.3	16.64	16.16	29	13.15	16.16
	Indonesian	13.66	28.55	9.51	23.97	15.96	30.83	10.95	26.04	16.49	31.13	10.32	26.23
	All	16.76	29.14	12.86	19.15	19.37	31.6	15.01	21.7	20.03	32.2	14.51	21.24

Table 11: Headline Generation Evaluation : **mT0**. **Selected Content** (Only Important Sentences).

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	20.34	40.42	17.08	35.45	21.31	41.5	18.08	36.56	21.57	41.73	18.34	36.94
	Portuguese	12.18	25.62	9.88	21.72	12.55	26.88	10.18	21.56	12.98	27.61	10.21	22.05
	Spanish	14.36	29.47	11.48	26.13	15.67	30.9	12.69	27.71	15.67	30.63	13.32	27.86
	Russian	1.93	12.67	0.8	8.79	2.09	12.77	0.99	8.94	2.01	12.73	0.9	8.73
	French	15.99	32.64	12.27	26.65	16.86	33.81	13.35	28.14	17.56	34.43	13.21	28.83
	Chinese	0.41	10.57	0.01	21.54	0.35	10.72	0.02	21.43	0.46	10.92	0.02	21.98
	Arabic	0.33	9.61	0	5.91	0.29	10.15	0	5.79	0.25	10.57	0	6.07
	All	9.36	23	7.36	20.88	9.87	23.82	7.9	21.45	10.07	24.09	8	21.78
Low	Turkish	11.03	24.5	8.03	13.56	12.64	25.99	8.98	15.11	12.65	25.69	8.93	14.33
	Ukrainian	1.23	11.53	0.3	8.12	1.38	11.46	0.2	8.02	1.41	11.77	0.19	7.9
	Bengali	0.48	7.99	0	7.07	0.73	8.46	0	7.39	0.74	8.68	0	6.84
	Persian	0.75	11.16	0	5.01	0.84	12.13	0	4.65	0.81	11.95	0	4.65
	Nepali	1.17	10.22	0	6.15	0.92	9.42	0.04	7.35	1.1	10.04	0.04	7.34
	Urdu	0.29	11.13	0	6.85	0.31	10.44	0	6.65	0.46	10.83	0	6.9
	Gujarati	0.95	9.15	0.08	8.05	1.28	9.73	0.04	8.96	1.14	9.8	0.04	8.56
	Hindi	0.57	10.31	0	5.89	1.28	11.39	0	6.67	0.65	10.73	0	6
	Marathi	1.41	11.94	0.05	9.28	1.31	12.23	0.1	9.07	1.4	12.36	0.08	9.14
	Telugu	2.55	12.5	0	8.86	2.83	12.62	0.05	9.37	3.04	13.22	0	9.33
	Tamil	0.4	9.15	0	8.64	0.39	9.02	0	8.62	0.34	9.12	0	8.82
	Panjabi	0.3	10.03	0.01	6.03	0.24	9.96	0.02	6.05	0.33	10.08	0.01	6.06
	Indonesian	11.25	25.73	6.89	20.68	12.52	27.09	7.9	22.11	12.95	27.5	8.11	22.48
	All	2.49	12.72	1.18	8.78	2.82	13.07	1.33	9.23	2.85	13.21	1.34	9.1
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	17.84	37.68	14.83	32.05	20.29	40.4	17.06	35.3	21.12	41.18	17.94	36.42
	Portuguese	10.49	24.98	7.72	19.46	11.02	25.31	9.19	20.38	11.95	25.5	9.88	21.29
	Spanish	13.85	28.74	11.41	25.55	15.21	30.18	12.53	27.4	15.2	30.04	12.54	27.45
	Russian	1.82	12.45	0.94	8.61	2.03	12.76	0.94	8.63	2.02	13.02	0.97	8.81
	French	14.85	31.57	11.5	25.43	16.73	32.96	13.13	28.07	17.47	33.83	13.4	28.87
	Chinese	0.43	10.58	0.02	21.45	0.45	10.92	0.02	21.88	0.41	10.94	0.02	21.68
	Arabic	0.19	10.53	0	6.26	0.25	10.48	0	5.92	0.36	10.62	0	6.16
	All	8.5	22.36	6.63	19.83	9.43	23.29	7.55	21.08	9.79	23.59	7.82	21.53
Low	Turkish	11.44	24.67	8.37	14.38	12.32	26.2	8.47	14.43	12.37	25.75	8.58	14.92
	Ukrainian	1.25	11.27	0.38	7.79	1.17	11.31	0.2	7.72	1.29	11.79	0.58	8.08
	Bengali	0.58	8.35	0	5.78	0.67	8.57	0	6.06	0.67	7.84	0	7.9
	Persian	0.83	12.46	0	4.52	0.78	11.96	0	4.6	0.79	12.23	0	4.56
	Nepali	1.12	10.14	0	5.76	1.21	10.56	0.01	6.58	1.12	9.49	0.05	7.82
	Urdu	0.52	11.09	0	6.94	0.42	11	0	7.05	0.41	10.91	0	6.99
	Gujarati	1.39	9.79	0.07	9.03	1.09	9.59	0.07	8.5	1.16	9.44	0.03	8.53
	Hindi	0.78	10.71	0	6.04	0.82	10.57	0	5.81	1.22	10.84	0.02	6.35
	Marathi	1.36	11.91	0.06	8.96	1.33	11.94	0.14	8.59	1.5	12.2	0.12	8.6
	Telugu	2.77	12.8	0.15	8.91	3.04	12.82	0.11	9.38	3.25	12.93	0.14	9.38
	Tamil	0.41	8.94	0.05	8.62	0.61	8.99	0.06	9	0.49	9.16	0.06	8.85
	Panjabi	0.35	10.02	0.01	5.67	0.31	10.06	0.02	5.85	0.41	10.05	0.03	6.09
	Indonesian	11.44	25.27	7.26	20.65	12.33	26.86	7.94	22.21	13.39	28.22	8.12	22.51
	All	2.63	12.88	1.26	8.7	2.78	13.11	1.31	8.91	2.93	13.14	1.36	9.28

Table 12: Headline Generation Evaluation : **Flan-T5**. **Selected Content** (Only Important Sentences)

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	51.85	44.4	54.8	57.81	53.31	45.6	55.93	59.05	47.46	40.78	48.24	51.25
	Portuguese	54.83	49.46	53.2	53.13	55.08	51.58	54.25	56.42	50.32	49.19	47.85	51.73
	Spanish	56.67	54.15	56.63	58.93	59.23	56.54	58.88	61.26	55.26	52.66	53.76	56.75
	Russian	56.99	51.71	56.12	58.38	59.42	53.82	58.94	61.22	57.28	52.25	55.6	58.4
	French	48.81	39.6	51.03	50.41	50.91	42.55	51.03	52.95	47.45	40.16	48.75	51.44
	Chinese	48.37	46.05	47.92	48.98	49.92	49.43	50.51	50.47	48.57	48.98	47.84	50.12
	Arabic	45.37	42.35	45.21	44.53	48.61	48.45	48.65	50.45	40.15	39.55	38.88	40.03
	All	51.84	46.82	52.13	53.17	53.78	49.71	54.03	55.97	49.5	46.22	48.7	51.39
Low	Turkish	57.45	49.51	54.4	56.64	57.36	51.48	55.05	57.75	54.72	48.19	51.22	53.06
	Ukrainian	53.32	50.8	52.78	54.66	56.37	52.42	55	56.86	53.13	51.01	50.91	54.25
	Bengali	55.79	48.41	55.37	56.33	56.6	50.46	57.81	58.51	48.51	42.59	51.12	51.12
	Persian	58.53	48.34	57.92	59.8	57.68	50.28	59.45	61.13	48.84	40.56	50.08	52.96
	Nepali	55.4	47.87	54.49	56.46	57.68	48.07	57.07	58.13	51.69	44.52	49.02	51.09
	Urdu	52.57	46.28	51.63	53.04	55.23	49.52	54.38	55.22	49.84	45.6	46.57	49.62
	Gujarati	32.91	27.4	28.88	29.57	35.51	30.05	32.21	31.79	29.66	26.9	27.86	27.59
	Hindi	44.29	42.5	43.57	44.87	50.09	45.02	48.34	49.36	34.45	31.71	32.83	34.27
	Marathi	33.73	33.05	33.2	33.77	37.61	36.58	35.81	37.03	33.17	33.41	31.51	33.52
	Telugu	38.32	37.72	36.86	37.87	42.96	40.58	41.74	42.09	34.61	34.35	33.52	34.1
	Tamil	44.97	41.16	44.78	45.29	46.6	43.39	46.28	47.53	41.29	38.76	40.85	42.77
	Panjabi	36.24	31.08	32.12	33.79	38.76	31.92	34.08	34.24	30.53	24.25	26.61	27.15
	Indonesian	53.01	45.95	53.21	53.86	55.18	47.72	54.64	55.96	52.6	46.66	51.71	54.28
	All	47.43	42.31	46.09	47.38	49.82	44.42	48.6	49.66	43.31	39.12	41.83	43.52
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	49.82	42.2	52.67	54.9	53.21	45.63	56.25	59.09	47.04	40.41	47.7	50.73
	Portuguese	55.85	50.38	53.97	56.98	55.33	53.16	56.88	58.28	50.18	50.1	49.07	51.88
	Spanish	58.84	55.49	58.26	61.33	60.55	57.4	59.91	62.72	56.54	52.07	54.78	57.75
	Russian	58.58	52.15	57.41	59.42	60.61	55.34	60.55	62.33	57.71	52.25	56.18	58.92
	French	48.93	42.99	52.08	53.53	51.42	41.8	53.36	53.42	47.39	40.85	49.3	50.15
	Chinese	48.32	47.7	49.3	49.54	50.84	49.55	50.15	52.09	48.45	49.44	47.96	50.55
	Arabic	49.24	47.33	49.11	49.99	50.35	49.2	49.7	51.32	40.41	40.43	40.74	42.25
	All	52.8	48.32	53.26	55.1	54.62	50.3	55.26	57.04	49.67	46.51	49.39	51.75
Low	Turkish	57.79	49.68	56.25	57.04	59.92	53.68	59.49	61.39	52.99	48.05	50.54	52.2
	Ukrainian	55.81	51.9	54.16	56.48	56.79	53.32	56.84	58.05	54.55	51.58	50.92	55.3
	Bengali	53.4	46.41	52.71	54.27	57.27	49.52	56.72	58.07	47.96	41.81	48.98	49.39
	Persian	58.82	50.58	58.76	61	60.56	52.41	59.94	62.35	48.87	42.13	48.87	51.47
	Nepali	56.83	48.17	55.12	56.92	58.42	50.38	58.29	60.85	52.13	44.83	49.17	52.93
	Urdu	55.14	49.2	54.23	54.94	57.56	51.59	56.11	57.85	50.69	45.29	47.45	49.93
	Gujarati	35.7	28	31.12	30.93	37.26	30.22	32.18	32.04	29	24.46	25.78	26.39
	Hindi	48.36	46.74	48.41	49.67	50.66	48.23	50.81	52.15	36.62	32.99	33.74	35.91
	Marathi	36.61	36.61	37.03	37.17	39	40.39	39.38	40.66	31.78	32.29	31.7	31.59
	Telugu	36.4	34.89	34.64	35.13	41.41	39.51	41.03	40.37	32.17	31.31	30.85	31.46
	Tamil	42.99	40.04	43.35	44.02	48.22	44.17	46.89	49.68	42.28	38.42	40.35	43.05
	Panjabi	31.27	26.2	28.17	27.62	37.83	30.22	32.91	32.99	32	27.44	28.97	28.08
	Indonesian	55.41	48.22	55.08	55.77	57.14	50.57	57.09	58.57	54.39	48.33	53.69	56.29
	All	48.04	42.82	46.85	47.77	50.93	45.71	49.82	51.16	43.49	39.15	41.62	43.38

Table 13: Tags Words Evaluation : **mT5**. **Selected Content** (Only Important Sentences).

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**F_1@3**	**F_1@5**	**F_1@$\mathcal{M}$**	**F_1@$\mathcal{O}$**	**F_1@3**	**F_1@5**	**F_1@$\mathcal{M}$**	**F_1@$\mathcal{O}$**	**F_1@3**	**F_1@5**	**F_1@$\mathcal{M}$**	**F_1@$\mathcal{O}$**
High	English	51.85	44.4	54.8	57.81	53.31	45.6	55.93	59.05	47.46	40.78	48.24	51.25
	Portuguese	54.83	49.46	53.2	53.13	55.08	51.58	54.25	56.42	50.32	49.19	47.85	51.73
	Spanish	56.67	54.15	56.63	58.93	59.23	56.54	58.88	61.26	55.26	52.66	53.76	56.75
	Russian	56.99	51.71	56.12	58.38	59.42	53.82	58.94	61.22	57.28	52.25	55.6	58.4
	French	48.81	39.6	51.03	50.41	50.91	42.55	51.03	52.95	47.45	40.16	48.75	51.44
	Chinese	48.37	46.05	47.92	48.98	49.92	49.43	50.51	50.47	48.57	48.98	47.84	50.12
	Arabic	45.37	42.35	45.21	44.53	48.61	48.45	48.65	50.45	40.15	39.55	38.88	40.03
	All	51.84	46.82	52.13	53.17	53.78	49.71	54.03	55.97	49.5	46.22	48.7	51.39
Low	Turkish	57.45	49.51	54.4	56.64	57.36	51.48	55.05	57.75	54.72	48.19	51.22	53.06
	Ukrainian	53.32	50.8	52.78	54.66	56.37	52.42	55	56.86	53.13	51.01	50.91	54.25
	Bengali	55.79	48.41	55.37	56.33	56.6	50.46	57.81	58.51	48.51	42.59	51.12	51.12
	Persian	58.53	48.34	57.92	59.8	57.68	50.28	59.45	61.13	48.84	40.56	50.08	52.96
	Nepali	55.4	47.87	54.49	56.46	57.68	48.07	57.07	58.13	51.69	44.52	49.02	51.09
	Urdu	52.57	46.28	51.63	53.04	55.23	49.52	54.38	55.22	49.84	45.6	46.57	49.62
	Gujarati	32.91	27.4	28.88	29.57	35.51	30.05	32.21	31.79	29.66	26.9	27.86	27.59
	Hindi	44.29	42.5	43.57	44.87	50.09	45.02	48.34	49.36	34.45	31.71	32.83	34.27
	Marathi	33.73	33.05	33.2	33.77	37.61	36.58	35.81	37.03	33.17	33.41	31.51	33.52
	Telugu	38.32	37.72	36.86	37.87	42.96	40.58	41.74	42.09	34.61	34.35	33.52	34.1
	Tamil	44.97	41.16	44.78	45.29	46.6	43.39	46.28	47.53	41.29	38.76	40.85	42.77
	Panjabi	36.24	31.08	32.12	33.79	38.76	31.92	34.08	34.24	30.53	24.25	26.61	27.15
	Indonesian	53.01	45.95	53.21	53.86	55.18	47.72	54.64	55.96	52.6	46.66	51.71	54.28
	All	47.43	42.31	46.09	47.38	49.82	44.42	48.6	49.66	43.31	39.12	41.83	43.52
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**F_1@3**	**F_1@5**	**F_1@$\mathcal{M}$**	**F_1@$\mathcal{O}$**	**F_1@3**	**F_1@5**	**F_1@$\mathcal{M}$**	**F_1@$\mathcal{O}$**	**F_1@3**	**F_1@5**	**F_1@$\mathcal{M}$**	**F_1@$\mathcal{O}$**
High	English	49.82	42.2	52.67	54.9	53.21	45.63	56.25	59.09	47.04	40.41	47.7	50.73
	Portuguese	55.85	50.38	53.97	56.98	55.33	53.16	56.88	58.28	50.18	50.1	49.07	51.88
	Spanish	58.84	55.49	58.26	61.33	60.55	57.4	59.91	62.72	56.54	52.07	54.78	57.75
	Russian	58.58	52.15	57.41	59.42	60.61	55.34	60.55	62.33	57.71	52.25	56.18	58.92
	French	48.93	42.99	52.08	53.53	51.42	41.8	53.36	53.42	47.39	40.85	49.3	50.15
	Chinese	48.32	47.7	49.3	49.54	50.84	49.55	50.15	52.09	48.45	49.44	47.96	50.55
	Arabic	49.24	47.33	49.11	49.99	50.35	49.2	49.7	51.32	40.41	40.43	40.74	42.25
	All	52.8	48.32	53.26	55.1	54.62	50.3	55.26	57.04	49.67	46.51	49.39	51.75
Low	Turkish	57.79	49.68	56.25	57.04	59.92	53.68	59.49	61.39	52.99	48.05	50.54	52.2
	Ukrainian	55.81	51.9	54.16	56.48	56.79	53.32	56.84	58.05	54.55	51.58	50.92	55.3
	Bengali	53.4	46.41	52.71	54.27	57.27	49.52	56.72	58.07	47.96	41.81	48.98	49.39
	Persian	58.82	50.58	58.76	61	60.56	52.41	59.94	62.35	48.87	42.13	48.87	51.47
	Nepali	56.83	48.17	55.12	56.92	58.42	50.38	58.29	60.85	52.13	44.83	49.17	52.93
	Urdu	55.14	49.2	54.23	54.94	57.56	51.59	56.11	57.85	50.69	45.29	47.45	49.93
	Gujarati	35.7	28	31.12	30.93	37.26	30.22	32.18	32.04	29	24.46	25.78	26.39
	Hindi	48.36	46.74	48.41	49.67	50.66	48.23	50.81	52.15	36.62	32.99	33.74	35.91
	Marathi	36.61	36.61	37.03	37.17	39	40.39	39.38	40.66	31.78	32.29	31.7	31.59
	Telugu	36.4	34.89	34.64	35.13	41.41	39.51	41.03	40.37	32.17	31.31	30.85	31.46
	Tamil	42.99	40.04	43.35	44.02	48.22	44.17	46.89	49.68	42.28	38.42	40.35	43.05
	Panjabi	31.27	26.2	28.17	27.62	37.83	30.22	32.91	32.99	32	27.44	28.97	28.08
	Indonesian	55.41	48.22	55.08	55.77	57.14	50.57	57.09	58.57	54.39	48.33	53.69	56.29
	All	48.04	42.82	46.85	47.77	50.93	45.71	49.82	51.16	43.49	39.15	41.62	43.38

Table 14: Tags Words Evaluation : **mT0**. **Selected Content** (Only Important Sentences).

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	50.07	43.08	52.37	54.61	50.74	43.79	53.32	55.83	51.05	44.22	53.51	56.28
	Portuguese	25.12	23.19	25.08	27.25	27.46	24.91	27.93	29.83	27.56	24.83	25.42	28.09
	Spanish	41.22	32.99	39.5	40.49	42.78	33.88	39.91	41.46	41.92	32.9	39.04	40.65
	Russian	0.16	0.05	0	0	0.2	0.18	0.17	0.24	0.44	0.25	0.19	0.35
	French	45.17	35.37	45.29	48.29	47.19	38.91	49.29	51.21	48.11	39.64	48.92	50.97
	Chinese	0.14	0.09	0.04	0.19	0.07	0.1	0.13	0.17	0.14	0.24	0.22	0.2
	Arabic	0	0	0	0	0	0	0	0	0	0	0	0
	All	23.13	19.25	23.18	24.4	24.06	20.25	24.39	25.53	24.17	20.3	23.9	25.22
Low	Turkish	17.54	17.86	18.6	18.14	21.04	18.79	22.06	22.98	25.68	20.68	25.87	26.42
	Ukrainian	0	0	0.06	0.06	0	0.07	0	0	0	0	0	0.11
	Bengali	0	0	0	0	0.07	0	0	0	0.07	0.07	0.08	0.08
	Persian	0	0	0	0	0	0	0	0	0	0	0	0
	Nepali	0	0	0	0	0	0	0	0	0	0	0	0
	Urdu	0	0	0	0	0	0	0	0	0	0	0	0
	Gujarati	0	0	0	0	0	0	0	0	0	0	0	0
	Hindi	0	0	0	0	0	0	0	0	0	0	0	0
	Marathi	0	0	0	0	0	0	0	0	0	0.06	0.09	0
	Telugu	0	0	0	0	0	0	0	0	0	0	0	0
	Tamil	0.15	0.07	0.09	0.13	0.15	0.13	0.09	0.07	0.15	0.06	0	0.07
	Panjabi	0	0	0	0	0	0	0	0	0	0	0	0
	Indonesian	42.23	35.18	42.08	42.9	43.6	36.72	42.86	44.13	43.8	37.63	44.59	43.91
	All	4.61	4.09	4.68	4.71	4.99	4.29	5	5.17	5.36	4.5	5.43	5.43
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	46.76	39.74	48.87	50.89	49.64	42.64	51.46	54.43	50.95	43.77	53.59	56.39
	Portuguese	28.36	24.17	28.09	28.53	29.29	26.24	27.24	27.78	29.73	24.59	27.19	28.79
	Spanish	42.69	34.11	40.36	41.13	41.81	33.7	40.74	41.1	43.12	34.89	41.39	42.57
	Russian	0.25	0	0.12	0.25	0.33	0.26	0.2	0.31	0.35	0	0.07	0.09
	French	44.96	36.82	44.91	47.3	46.38	37.77	48.44	49.54	46.49	39.93	48.92	48.58
	Chinese	0.07	0.05	0	0.16	0.14	0.12	0.06	0.06	0.07	0.22	0.09	0.18
	Arabic	0	0	0	0	0	0	0	0	0	0	0	0
	All	23.3	19.27	23.19	24.04	23.94	20.1	24.02	24.75	24.39	20.49	24.46	25.23
Low	Turkish	20.67	16.4	21.51	21.33	23.61	20.37	22.64	22.92	22.82	18.88	24.68	25.7
	Ukrainian	0	0	0	0	0.06	0.11	0	0	0	0.09	0	0
	Bengali	0	0	0	0	0	0	0	0	0	0.05	0	0
	Persian	0	0	0	0	0	0	0	0	0	0	0	0
	Nepali	0	0	0	0	0	0	0	0	0	0	0	0
	Urdu	0	0	0	0	0	0	0	0	0	0	0	0
	Gujarati	0	0	0	0	0	0	0	0	0	0	0	0
	Hindi	0	0	0	0	0	0	0	0	0	0	0	0
	Marathi	0	0	0	0	0	0	0	0	0	0	0	0
	Telugu	0	0	0	0	0	0	0	0	0	0	0	0
	Tamil	0.15	0.12	0.18	0.13	0.15	0.12	0.09	0.2	0.15	0.18	0.18	0.13
	Panjabi	0	0	0	0	0	0	0	0	0	0	0	0
	Indonesian	42.83	35.88	42.63	42.98	43.51	35.85	42.98	44.49	44.53	36.74	43.86	45.94
	All	4.9	4.03	4.95	4.96	5.18	4.34	5.05	5.2	5.19	4.3	5.29	5.52

Table 15: Tags Words Evaluation : **Flan-T5**. **Selected Content** (Only Important Sentences).

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	17.39	37.47	13.76	31.32	17.36	37.49	13.84	31.44	17.37	37.53	13.87	31.51
	Portuguese	14.46	28.39	10.87	23.48	16.31	30.17	12.25	24.35	14.97	29.19	11.51	23.42
	Spanish	16.79	31.63	13.35	28.71	16.79	31.8	13.26	29	16.6	31.51	13.33	28.69
	Russian	15.32	28.01	9	19.39	15.14	27.65	8.86	19.11	15.89	28.32	9.53	19.8
	French	16.75	33.26	12.56	26.13	15.59	32.31	11	26.09	16.12	33.09	12.06	26.32
	Chinese	20.29	33.78	12.49	20.87	19.91	33.24	12.2	20.43	20.39	34.03	12.17	20.54
	Arabic	22.62	34.19	18.6	26.98	20.67	31.56	15.89	25.38	20.68	32.65	15.22	25.93
	All	17.66	32.39	12.95	25.27	17.4	32.03	12.47	25.11	17.43	32.33	12.53	25.17
Low	Turkish	21.28	35.68	16.44	24.98	19.75	34.09	14.93	22.91	21.25	35.34	16.38	24.54
	Ukrainian	15.13	27.71	8.35	18.42	15.53	28.05	8.74	18.87	15.71	28.43	8.78	18.75
	Bengali	25.19	35.53	17.3	21.03	24.9	35.24	17.02	20.34	25.21	35.45	17.54	21
	Persian	21.93	35.31	18.64	26.8	21.07	34.25	18.04	25.52	21.46	35.43	18.71	26.28
	Nepali	22.7	33.97	17	19.83	24.64	35.39	18.46	22.02	23.94	34.56	17.81	21.47
	Urdu	21.18	32.97	16.69	25.91	20.66	32.49	16.38	25.15	22.07	33.81	17.78	26.89
	Gujarati	21.26	32.72	15.05	17.99	21.57	32.9	15.82	17.38	20.76	31.52	14.18	17.29
	Hindi	22.54	34.44	17.75	26.08	20.82	33.12	15.92	24.02	21.95	34.18	17.38	25.62
	Marathi	21.88	32.75	17.5	21.03	21.34	32.16	17.19	21.05	21.66	32.34	16.98	20.22
	Telugu	20.22	31.75	15.73	19.65	19.26	30.47	14.91	18.71	18.62	29.99	14.7	18.38
	Tamil	25.46	35.42	14.95	20.77	25.09	35.19	15.12	20.53	24.58	34.48	13.93	19.57
	Panjabi	19.46	31.63	16.52	19.34	18.51	30.56	15.65	18.4	19.57	31.33	16.5	18.97
	Indonesian	15.53	30.24	9.92	25.01	16.73	31.51	10.62	26.13	15.42	30.57	9.8	25.15
	All	21.06	33.09	15.53	22.06	20.76	32.72	15.29	21.62	20.94	32.88	15.42	21.86

		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	17.34	37.45	13.89	31.32	17.2	37.27	13.81	31.33	17.16	37.3	13.78	31.28
	Portuguese	14.97	29.5	11.65	23.84	14.44	28.46	10.93	22.8	14.48	28.97	11	23.07
	Spanish	17.47	32.37	14.03	29.49	17.1	31.78	13.61	29.2	16.96	31.84	13.59	28.98
	Russian	16.3	29.03	9.82	20.31	15.47	27.88	9.09	19.38	16.18	28.77	9.58	20.22
	French	15.56	32.46	12	26.52	15.92	32.77	11.89	26.74	16.03	33.16	12.34	26.75
	Chinese	19.55	33.52	12.18	20.67	19.89	33.36	12.63	20.19	19.12	32.87	11.1	20.47
	Arabic	22.57	33.54	16.89	27.11	22.03	34.24	16.68	27.04	20.58	32.49	15.67	25.82
	All	17.68	32.55	12.92	25.61	17.44	32.25	12.66	25.24	17.22	32.2	12.44	25.23
Low	Turkish	20.22	34.42	15.15	24.04	20.98	34.75	15.85	24.77	21.28	35.69	16.16	24.05
	Ukrainian	15.39	28.02	8.68	18.74	15.96	28.35	8.92	19.06	16.55	29.12	9.33	19.36
	Bengali	25.41	35.94	17.71	21.32	24.56	35.03	16.96	20.52	24.64	35.14	16.78	19.99
	Persian	21.47	35.41	19.04	26.02	19.99	33.25	16.79	24.85	20.79	34.1	18.44	25.97
	Nepali	22.84	33.72	17.01	20.28	23.96	34.99	17.42	21.04	23.8	34.56	17.85	21.5
	Urdu	22.31	34.07	18.05	27.1	20.85	32.62	16.46	25.84	21.63	33.69	17.3	26.69
	Gujarati	21.96	33.08	16.07	17.65	21.34	32.12	15.29	17.5	20.7	32.21	14.14	16.93
	Hindi	22.62	35.34	17.37	25.57	21.64	34.12	16.47	24.77	20.96	33.68	15.64	24.72
	Marathi	22.25	33.18	17.87	20.91	21.17	31.85	16.89	20.17	20.32	31.34	15.62	19.87
	Telugu	19.33	31.13	15.07	19.04	19.47	31.15	14.99	18.59	19.65	31.24	15.22	19.28
	Tamil	24.65	34.64	14.14	19.33	25.74	35.61	15.24	20.84	24.41	34.28	13.92	19.38
	Panjabi	17.95	30.23	15.45	17.72	18.12	30.48	15.28	17.73	17.8	30.15	14.87	17.46
	Indonesian	15.87	30.99	10.01	25.69	15.92	30.99	10.45	25.97	16.38	31.36	10.41	26.04
	All	20.94	33.09	15.51	21.8	20.75	32.72	15.15	21.67	20.69	32.81	15.05	21.63

Table 16: Headline Generation Evaluation : **mt5**. **Selected Content** (Important Sentences + Article).

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	17.22	37	37.31	31.51	17.56	37.54	14.12	31.85	17.58	37.52	14.19	31.85
	Portuguese	14.19	26.95	28.18	23.29	14.98	29.45	11.56	24.21	16.21	30.47	12.43	24.71
	Spanish	17	29.93	31.91	29.09	16.96	32.05	13.4	29.06	17.12	32.04	14	29.06
	Russian	15.27	26.2	28.04	19.21	15.42	28.03	8.99	19.51	15.47	28.33	9.37	19.84
	French	16.67	32.13	32.89	26.24	15.99	32.49	11.82	26.88	16.71	33.42	12.14	27.22
	Chinese	19.99	32.07	34.05	20.52	20	33.52	12.77	20.94	20.7	34.08	12.81	21.2
	Arabic	21.68	32.84	33.16	27.82	19.95	32.26	16.39	26.41	20.18	32.09	15.24	25.34
	All	17.43	31.02	32.22	25.38	17.27	32.19	12.72	25.55	17.71	32.56	12.88	25.6
Low	Turkish	20.94	34.33	34.95	24.28	20.36	34.68	15.77	24.07	20.79	34.66	15.88	24.19
	Ukrainian	15.28	26.36	28.08	18.89	15.99	28.69	9.09	19.34	16.08	28.67	9.19	19.48
	Bengali	25.32	35.33	36.16	21.39	24.55	35.01	17.49	21.26	24.81	35.14	17.28	20.57
	Persian	21.12	33.58	34.63	25.34	21.48	35.01	19	26.66	21.35	33.79	18.28	25.82
	Nepali	22.8	32.88	33.57	20.47	24.57	35.43	19.11	22.07	24.21	35.15	17.79	21.33
	Urdu	21.14	32.39	33.22	26.21	20.65	32.13	16.96	25.59	21.72	33.41	17.86	27.04
	Gujarati	23.09	33.49	34.22	18.22	21.93	33.26	15.63	18.46	21.6	32.25	14.94	17.55
	Hindi	21.63	32.55	33.89	24.59	21.79	34.28	16.8	24.54	22.59	34.99	17.87	25.84
	Marathi	21.76	32.13	32.63	21.29	20.84	31.14	16.94	20.73	21.22	31.48	16.88	20.54
	Telugu	19.6	29.46	31.21	19.07	19.34	30.95	14.93	19.05	18.69	30.28	14.65	17.98
	Tamil	25.56	33.52	35.44	20.83	24.79	34.48	14.86	20.19	24.86	34.6	14.87	20.38
	Panjabi	18.87	29.69	30.74	18.84	18.04	30.09	15.72	18.05	19.62	31.71	16.78	19.29
	Indonesian	16.37	28.84	31.21	26.03	16.58	31.22	10.5	26.11	16.13	31.27	9.99	25.82
	All	21.04	31.89	33.07	21.96	20.84	32.8	15.6	22.01	21.05	32.88	15.56	21.99
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	17.45	37.37	14.06	31.53	17.47	37.43	14.08	31.63	17.55	37.45	14.15	31.66
	Portuguese	14.96	29.24	11.67	24.02	15.32	29.31	12.67	24.32	15.62	30.13	12.51	24.91
	Spanish	17.76	32.69	14.4	29.84	16.84	31.75	13.57	29.12	17.27	31.91	13.99	29.55
	Russian	15.96	28.58	9.79	20.22	15.81	28.46	9.41	20.06	16.1	28.41	9.84	20.26
	French	16.42	32.8	11.9	26.86	16.78	33.09	12.6	27.4	16.99	33.63	12.35	26.83
	Chinese	19.7	33.69	12.48	20.65	19.63	33.62	12.65	20.36	19.6	33.4	11.75	20.26
	Arabic	22.37	34.21	17.45	27.98	21.66	33.69	17.09	27.73	21.13	33.48	16.55	27.06
	All	17.8	32.65	13.11	25.87	17.64	32.48	13.15	25.8	20.6	32.69	15.17	21.69
Low	Turkish	20.67	34.66	16.01	24.3	21.64	35.53	16.69	25.3	20.19	34.28	15.7	23.69
	Ukrainian	15.52	28.27	8.81	19.27	15.86	28.4	8.93	19	16.7	29.11	9.6	19.7
	Bengali	25.1	35.48	17.66	21.33	25.17	35.85	17.6	20.79	25.03	35.71	17.52	21.29
	Persian	20.42	33.96	17.92	25.5	19.54	32.7	16.58	24.42	21.29	34.13	18.74	25.78
	Nepali	23.17	34.13	17.79	21.03	23.53	34.89	17.82	21.11	24.02	35.29	18.01	21.31
	Urdu	22.33	34.29	18.13	27.49	21.22	32.79	17.13	26.03	21.59	33.45	17.41	26.81
	Gujarati	22.56	33.27	16.59	17.55	21.99	32.83	16.03	18.27	21.64	32.83	15.45	17.87
	Hindi	22.1	34.64	17.47	25.06	21.26	33.84	16.02	23.93	20.71	33.52	15.41	23.93
	Marathi	21.82	32.7	17.4	20.69	20.97	31.62	16.46	20.11	20.17	30.58	15.79	19.79
	Telugu	18.71	30.61	14.47	17.93	19.11	31.21	15.21	19.21	18.66	30.45	14.24	18.22
	Tamil	25.67	35.6	15.56	21.2	25.33	35.13	14.93	20.51	23.81	33.81	14.29	19.75
	Panjabi	18.57	30.66	15.55	18.24	17.86	30.25	15.36	17.68	16.95	29.38	14.17	16.7
	Indonesian	16.43	31.29	10.08	26.54	16.52	31.72	10.32	26.26	17.06	32.46	10.9	27.08
	All	21.01	33.04	15.65	22.01	20.77	32.83	15.31	21.74	20.6	32.69	15.17	21.69

Table 17: Headline Generation Evaluation : **mT0**. **Selected Content** (Important Sentences + Article).

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	21.3	41.56	18.05	36.55	21.44	41.65	18.21	36.85	21.69	41.85	18.43	37.09
	Portuguese	11.5	25.44	9.77	20.7	12.4	27.04	9.77	21.85	12.85	27.59	9.59	21.67
	Spanish	15.63	30.5	12.74	27.92	15.62	30.8	12.79	27.92	15.93	30.81	13.17	28.31
	Russian	1.93	12.71	1.06	8.83	2.12	12.89	1.03	9.07	2.03	12.98	0.78	9.05
	French	17.62	34.63	13.8	28.66	17.51	34.37	14.16	28.77	17.14	34.04	12.86	28.1
	Chinese	0.4	10.88	0.01	21.63	0.49	11.01	0.01	22.14	0.51	10.98	0.01	22.01
	Arabic	0.24	9.49	0	5.83	0.22	10.22	0	5.81	0.22	10.12	0	5.95
	All	9.8	23.6	7.92	21.45	9.97	24	8	21.77	10.05	24.05	7.83	21.74
Low	Turkish	11.31	24.3	8.13	13.53	12.07	25.98	8.26	14.45	11.91	24.91	8.53	14.66
	Ukrainian	1.31	11.63	0.3	8.06	1.47	11.66	0.39	8.18	1.44	11.81	0.4	8.28
	Bengali	0.86	8.72	0	7	0.78	8.73	0	6.39	0.8	8.8	0	7.14
	Persian	0.82	11.99	0	4.62	0.86	12.45	0	4.55	0.84	12.33	0	4.55
	Nepali	1.11	9.92	0.03	7.42	1.08	9.91	0.01	7.17	1.11	9.73	0.04	7.42
	Urdu	0.39	11.22	0	6.86	0.34	11.28	0	6.9	0.27	10.89	0	6.92
	Gujarati	1.15	9.49	0.11	8.61	1.29	9.76	0	8.78	1.14	9.47	0	8.31
	Hindi	0.39	10.4	0	5.59	0.64	10.55	0	5.77	0.75	10.91	0	6.14
	Marathi	1.23	12.15	0.16	8.8	1.21	11.88	0.04	8.72	1.37	12.16	0.13	9.04
	Telugu	3.05	13	0.05	9.48	3.02	12.77	0.1	9.62	2.97	13.09	0.12	9.61
	Tamil	0.37	8.89	0	8.24	0.41	9.03	0	8.54	0.42	8.85	0	8.37
	Panjabi	0.39	10.08	0.01	6.04	0.29	10.04	0.01	5.86	0.31	10.04	0	5.99
	Indonesian	12.8	27.35	7.79	22.65	13.55	28.05	8.62	22.92	13.33	27.86	8.42	22.67
	All	2.71	13.01	1.28	8.99	2.85	13.24	1.34	9.07	2.82	13.14	1.36	9.16
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**	**R2**	**RL**	**BLEU**	**METEOR**
High	English	21.39	41.6	17.94	36.57	21.51	41.67	18.26	36.79	21.31	41.31	18	36.42
	Portuguese	10.81	25.72	8.56	20.64	13.09	27.25	11.01	22.64	10.93	25.21	8.45	20.35
	Spanish	15.8	30.85	12.98	27.88	15.88	30.56	12.92	28.4	15.11	30.03	12.55	27.42
	Russian	1.91	12.78	1.23	8.97	2.18	13.16	1.25	9.12	2.16	13.18	1.17	9.07
	French	17.02	33.93	13.59	27.42	15.92	32.7	12.81	27.86	17.41	34.05	13.76	29
	Chinese	0.43	10.86	0	21.71	0.45	10.93	0.01	22.04	0.53	11.09	0.01	22.18
	Arabic	0.38	10.36	0	5.92	0.29	10.18	0	6.31	0.24	10.27	0	6.08
	All	9.68	23.73	7.76	21.3	9.9	23.78	8.04	21.88	9.67	23.59	7.71	21.5
Low	Turkish	12.2	25.94	8.32	14.45	12.39	25.8	8.75	14.81	12.52	25.86	8.79	14.86
	Ukrainian	1.36	12.01	0.25	8.19	1.24	11.85	0.42	7.97	1.39	11.66	0.38	8.03
	Bengali	0.94	8.49	0	7.73	0.67	8.6	0	6.18	0.84	8.5	0	7.5
	Persian	0.77	11.97	0	4.68	0.76	11.56	0	4.89	0.82	12.31	0	4.63
	Nepali	1.02	9.43	0.04	7.23	1.08	9.81	0.02	6.77	1.09	9.96	0.02	6.86
	Urdu	0.45	11.73	0	7.16	0.26	11.22	0	6.95	0.24	11.07	0	6.96
	Gujarati	1.47	9.86	0.1	9.4	1.02	9.46	0.03	8.31	1.19	9.59	0.03	8.66
	Hindi	0.59	10.43	0	5.63	0.67	10.57	0	5.75	0.85	10.45	0	6
	Marathi	1.84	12.58	0.18	9.54	1.47	12.05	0.08	8.76	1.56	12.25	0.3	9.05
	Telugu	3.46	13.16	0.08	9.71	3.45	13.25	0.02	9.91	3.39	13.3	0.16	9.7
	Tamil	0.49	9.06	0.04	8.52	0.41	8.86	0.05	8.39	0.54	9.18	0.05	8.67
	Panjabi	0.36	9.82	0.01	6	0.34	9.88	0.01	5.81	0.38	9.81	0.02	6
	Indonesian	12.9	27.24	8.15	22.56	13.24	27.67	8.18	22.93	13.35	27.97	8.26	22.77
	All	2.91	13.21	1.32	9.29	2.85	13.12	1.35	9.03	2.94	13.22	1.39	9.21

Table 18: Headline Generation Evaluation : **Flan-T5**. **Selected Content** (Important Sentences + Article)

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	$\mathbf{F}_1@\mathbf{3}$	$\mathbf{F}_1@\mathbf{5}$	$\mathbf{F}_1@\mathcal{M}$	$\mathbf{F}_1@\mathcal{O}$	$\mathbf{F}_1@\mathbf{3}$	$\mathbf{F}_1@\mathbf{5}$	$\mathbf{F}_1@\mathcal{M}$	$\mathbf{F}_1@\mathcal{O}$	$\mathbf{F}_1@\mathbf{3}$	$\mathbf{F}_1@\mathbf{5}$	$\mathbf{F}_1@\mathcal{M}$	$\mathbf{F}_1@\mathcal{O}$
High	English	48.37	41.46	49.82	52.35	48.36	41.47	50.16	52.49	48.45	41.41	49.79	52.52
	Portuguese	51.4	48.74	48.34	51.66	50.37	49.13	48.28	51.3	51.28	49.39	49.19	51.68
	Spanish	56.6	53.43	55.22	58.63	56.38	52.86	54.93	57.98	57.03	53.05	55.26	58.55
	Russian	57.68	52.3	56.93	58.72	57.57	52.63	56.94	58.61	57.97	53.17	57.37	59.32
	French	49.04	41.56	50.85	51.36	48.68	40.7	51.26	51.48	48.62	41.6	52.55	51.65
	Chinese	48.16	48.5	47.88	50.33	47.71	48.96	48.09	49.92	48.72	48.79	47.74	49.71
	Arabic	40.65	40.69	40.87	41.95	40.46	40.74	41.06	41.73	41.51	40.34	40.98	41.26
	All	50.27	46.67	49.99	52.14	49.93	46.64	50.1	51.93	50.51	46.82	50.41	52.1
Low	Turkish	53.25	48	51.51	53.73	54.67	48.03	52.32	55.77	54.44	47.09	50.9	52.4
	Ukrainian	53.87	50.91	52.41	54.79	54.19	51.31	52.7	54.82	53.84	50.84	53.13	54.69
	Bengali	50.5	43.67	51.66	52.52	49.24	43.74	50.67	51.08	49.76	42.96	51.64	52.03
	Persian	49.97	41.37	50.23	52.01	50.09	41.77	50.51	53.3	49.78	41.41	50.78	52.01
	Nepali	51.37	45.87	51.43	51.88	50.99	45.92	50.73	51.55	51.76	46.56	50.64	53.11
	Urdu	50.1	45.15	47.24	49.65	49.86	44.82	47.68	49.63	52.06	45.3	49.5	51.14
	Gujarati	30.89	26.9	28.78	28.45	29.72	27.29	28.31	29.32	29.8	27.27	30.12	28.93
	Hindi	36.52	34.27	34.92	35.61	36.82	33.76	34.81	36.75	38.32	33.58	35.58	37.88
	Marathi	32.15	33.61	32.21	33.34	31.83	32.63	31.56	32.56	32.28	33.24	32.02	33.83
	Telugu	35.52	34.68	34.96	35.59	35.11	34.27	33.86	34.71	35.94	33.76	33.92	35.32
	Tamil	42.77	40.31	41.29	43.78	43.27	40.77	42.13	44.31	43.45	39.81	41.38	43.39
	Panjabi	31.19	24.69	26.57	25.92	32.84	25.09	26.55	26.41	32.66	26.61	28.5	28.76
	Indonesian	54.84	47.87	54.15	55.18	54.62	47.05	53.56	55.61	53.77	47.71	53.1	54.88
	All	44.07	39.79	42.87	44.03	44.1	39.73	42.72	44.29	44.45	39.7	43.17	44.49
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	$\mathbf{F}_1@\mathbf{3}$	$\mathbf{F}_1@\mathbf{5}$	$\mathbf{F}_1@\mathcal{M}$	$\mathbf{F}_1@\mathcal{O}$	$\mathbf{F}_1@\mathbf{3}$	$\mathbf{F}_1@\mathbf{5}$	$\mathbf{F}_1@\mathcal{M}$	$\mathbf{F}_1@\mathcal{O}$	$\mathbf{F}_1@\mathbf{3}$	$\mathbf{F}_1@\mathbf{5}$	$\mathbf{F}_1@\mathcal{M}$	$\mathbf{F}_1@\mathcal{O}$
High	English	48.01	41.16	49.61	51.98	48.24	41.35	49.51	52.07	48.29	41.28	49.63	52.41
	Portuguese	50.57	47.36	47.06	52.17	51.14	50.58	48.72	52.46	52.79	48.69	49.82	51.7
	Spanish	57.32	52.79	55.34	58.74	57.15	53.11	55.45	58.8	57.43	53.67	55.09	58.84
	Russian	58.12	52.93	57.36	59.38	59.3	53.59	58.28	59.89	58.62	53.31	57.56	59.79
	French	48.02	40.81	48.64	50.5	48.72	41.22	51.52	51.57	49.67	41.81	48.9	51.43
	Chinese	48.28	48.69	47.64	50.62	49.18	49.01	47.94	50.7	48.33	49.02	47.68	50.19
	Arabic	42.77	40.99	41.54	43.3	40.85	41.57	41.03	42.02	42.14	41.76	41.94	43.46
	All	50.44	46.39	49.6	52.38	50.65	47.2	50.35	52.5	51.04	47.08	50.09	52.55
Low	Turkish	55.94	49.08	52.04	53.57	53.48	47.42	50.25	53.4	55.59	48.21	50.44	53.86
	Ukrainian	54.74	51.03	52.16	55.36	54.72	51.46	52.99	55.78	54.61	51.14	52.55	55.81
	Bengali	50.52	43.2	50.63	51.36	49.52	42.83	50.66	50.53	49.53	42.78	49.89	51.25
	Persian	49.99	41.01	50.6	51.47	49.68	41.44	51.07	52.27	49.99	42.28	50.74	52.14
	Nepali	52.54	45.75	50.43	51.65	53.47	46.22	52.21	53.65	53.41	46.13	51.97	53.75
	Urdu	50.62	44.9	46.9	49.83	51.14	46.03	49.55	51.1	50.07	45.08	47.5	50.72
	Gujarati	31.36	28.11	29.95	30.3	29.54	27.89	27.4	28.89	29.99	26.17	28.49	29.19
	Hindi	37.92	33.92	37.13	37.9	38.5	34.11	35.24	37.33	37.3	33.27	36.05	37.8
	Marathi	32.74	32.24	31.78	32.86	33.57	33.74	33	33.7	32.85	33.96	32.39	32.85
	Telugu	35.87	34.31	36.36	36.01	36.08	34.54	34.72	35.93	36.53	34.18	34.73	35.57
	Tamil	44.37	41.29	41.22	44.93	42.9	41.71	41.75	43.87	43.3	40.55	40.89	43.35
	Panjabi	32.32	25.96	28.17	27.09	32.99	26.69	29.48	29.12	32.3	28.2	27.93	28.79
	Indonesian	54.05	48.46	53.92	54.93	53.82	48.27	52.95	54.94	54.88	48.39	53.39	56.14
	All	44.84	39.94	43.18	44.4	44.57	40.18	43.17	44.65	44.64	40.03	42.84	44.71

Table 19: Tags Words Evaluation : **mT5**. **Selected Content** (Important Sentences + Article)

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	48.27	41.36	49.74	52.04	48.28	41.48	49.39	51.99	48.23	41.51	49.6	52.23
	Portuguese	54.58	50.9	51.58	53.56	51.3	50.46	48.04	52.33	53.11	49.62	50.43	53.33
	Spanish	56.77	52.72	55.2	58.35	56.85	52.88	55.48	58.67	57.48	53.25	55.67	59.28
	Russian	57.97	53.11	57.34	58.92	58.06	53.41	57.19	59.24	58.69	53.17	57.44	59.4
	French	49.01	39.99	50.65	52.59	47.67	40.59	49.63	51.92	49.16	40.64	50.53	52.26
	Chinese	48.81	47.93	48.08	50.65	48.96	48.85	48.82	49.8	49.12	49.45	48.36	50.07
	Arabic	40.53	39.19	39.47	41.01	40.78	38.52	39.82	40.76	41.08	40.94	41.14	41.7
	All	50.85	46.46	50.29	52.45	50.27	46.6	49.77	52.1	50.98	46.94	50.45	52.61
Low	Turkish	55.92	47.53	52.52	54.51	54.32	48.13	51.52	54.76	53.81	48.44	50.29	52.88
	Ukrainian	54.79	51.09	52.34	55.13	54.34	50.94	52.35	55.58	54.28	51.07	52.4	55.81
	Bengali	49.78	42.97	50.91	51.2	50.72	42.69	51.67	52.04	50.16	43.01	51.47	51.63
	Persian	50.47	40.31	51.58	53.02	48.31	39.63	48.98	50.87	49.59	40.95	50.38	51.67
	Nepali	52.27	44.01	48.45	51	52.2	46.1	49.19	51.8	52.56	44.77	49.36	52.37
	Urdu	49.65	45.02	47.31	50.28	50.26	45.61	46.85	49.88	50.11	45.15	47.71	49.43
	Gujarati	30.8	29.08	28.78	30.05	29.32	28.14	27.73	27.96	31.96	29.48	29.85	30.69
	Hindi	36.25	32.39	34.89	36.32	35.68	32.63	32.69	35.89	36.53	32.69	34.76	35.91
	Marathi	32.32	32.67	30.71	32.33	33.18	33.71	32.34	33.75	32.51	33.03	31.17	32.11
	Telugu	35.79	35.58	35.61	36.61	34.31	33.94	33.42	33.55	35.17	35.6	35.19	35.4
	Tamil	43.38	40.14	40.71	43.63	43.57	41.14	40.86	44.11	43.73	40.67	41.9	44.35
	Panjabi	31.22	25.57	25.87	26.85	32.77	26.42	26.35	28.61	32.12	25.56	28.37	28.49
	Indonesian	54.15	47.51	53.75	55.37	54.34	48.51	53.18	54.54	55.04	48.69	52.71	54.95
	All	44.37	39.53	42.57	44.33	44.1	39.81	42.09	44.1	44.43	39.93	42.74	44.28
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	48.02	41.37	49.16	51.95	48.29	41.45	49.44	51.96	48.03	41.41	49.62	52.17
	Portuguese	53.11	49.66	51.26	53.51	54.19	51.3	51.72	53.41	53.28	51.92	51.38	54.13
	Spanish	57.74	52.82	56.14	59.27	57.37	53.2	56.42	59.34	57.88	54.05	56.17	58.9
	Russian	58.69	53.5	57.49	60.16	58.83	53.75	57.75	60.34	58.85	53.78	57.42	60.32
	French	47.96	40.73	50.19	50.87	48.36	41.79	50.18	50.75	47.44	39.41	48.92	49.82
	Chinese	49.11	48.72	48.34	50.99	48.46	49.18	47.42	50.8	49.14	49.25	46.95	49.67
	Arabic	41.19	39.98	39.28	41.2	40.47	40.06	40.89	41.57	43.34	40.84	42.88	42.56
	All	50.83	46.68	50.27	52.56	50.85	47.25	50.55	52.6	51.14	47.24	50.48	52.51
Low	Turkish	53.46	47.99	50.53	54.23	54.6	48.1	52.84	53.85	53.36	47.57	52.19	54.2
	Ukrainian	53.85	50.72	52.41	54.96	53.58	50.58	51.7	55.07	55.25	50.76	53.3	55.33
	Bengali	50.29	42.84	50.9	51.24	50.52	42.68	50.57	51.44	49.62	42.64	49.67	50.58
	Persian	48.91	41.11	50.83	51.85	49.56	42.52	51.07	53.22	49.44	41.14	50.14	51.72
	Nepali	51.31	44.2	49.45	51.83	53.8	46.53	50.24	52.93	51.25	44.96	49.06	51.03
	Urdu	51	44.54	47.75	50.91	51.25	46.31	48.07	52.05	50.91	45.06	47.3	49.77
	Gujarati	31.23	28.54	28.93	29.8	31.7	27.84	29.2	30.33	31.21	27.39	28.61	29.93
	Hindi	37.17	33.18	35.36	36.46	36.77	33.77	36.00	36.53	37.61	33.02	34.24	36.06
	Marathi	32.88	33.76	31.49	32.81	34.27	34.42	31.87	33.84	34.79	33.84	33.2	34.12
	Telugu	35.27	34.68	34.8	36.38	36.32	34.88	35.52	36.33	35.86	35	34.95	36.3
	Tamil	44.87	40.92	41.36	45.05	43.35	40.21	41.54	43.87	43.89	41.26	41.1	43.96
	Panjabi	32.83	26.56	27.57	27.46	33.59	27.25	27.54	28.15	32.79	26.34	29.09	28.86
	Indonesian	54.05	47.38	54.4	55.93	54.63	48.58	53.97	57.17	55.31	48.9	54.21	57.5
	All	44.39	39.72	42.75	44.53	44.92	40.28	43.09	44.98	44.71	39.84	42.85	44.57

Table 20: Tags Words Evaluation : **mT0**. **Selected Content** (Important Sentences + Article)

		Text (*Caption*)											
		-w/C (K=5)				**-w/C (K=10)**				**-w/C (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	50.91	44.03	53.64	56.42	50.92	44.08	53.9	56.23	50.86	44.28	53.93	56.47
	Portuguese	25.87	23.52	25.38	26.06	27.8	24.13	26.17	26.48	27.93	24.84	27.75	28.48
	Spanish	43.23	33.51	40.24	41.55	42.74	33.61	40.26	41.99	41.22	32.16	39.45	40.39
	Russian	0.16	0.2	0.09	0.12	0.29	0.12	0.2	0.27	0.19	0.13	0.09	0.16
	French	46.62	38.06	47.51	49.14	46.96	39.28	48.49	50.53	48.3	39.96	48.45	50.23
	Chinese	0.22	0.23	0.16	0.27	0.19	0.15	0.13	0.17	0.14	0.18	0.12	0.22
	Arabic	0	0	0	0	0	0	0	0	0	0	0	0
	All	23.86	19.94	23.86	24.79	24.13	20.2	24.16	25.1	24.09	20.22	24.26	25.14
Low	Turkish	19.36	18.05	20.01	20.22	22.41	21.68	23.37	22.49	24.1	21.09	23.68	25.23
	Ukrainian	0	0	0	0	0	0	0	0	0	0	0	0
	Bengali	0.07	0.07	0.08	0.08	0.07	0.07	0	0	0.07	0.07	0.08	0.08
	Persian	0	0	0	0	0	0	0	0	0	0	0	0
	Nepali	0	0	0	0	0	0	0	0	0	0	0	0
	Urdu	0	0	0	0	0	0	0	0	0	0	0	0
	Gujarati	0	0	0	0	0	0	0	0	0	0	0	0
	Hindi	0	0	0	0	0	0	0	0	0	0	0	0
	Marathi	0	0.07	0.09	0	0	0	0	0	0	0.07	0.08	0
	Telugu	0	0	0	0	0	0	0	0	0	0	0	0
	Tamil	0.15	0.12	0.09	0.13	0.15	0.15	0.18	0.2	0.15	0.07	0.09	0.07
	Panjabi	0	0	0	0	0	0	0	0	0	0	0	0
	Indonesian	44.28	37.35	44.65	45.82	44.22	37.94	44.9	46.4	43.02	36.75	43.61	44.33
	All	4.91	4.28	4.99	5.1	5.14	4.6	5.27	5.31	5.18	4.47	5.2	5.36
		Visual (*Image*)											
		-w/I (K=5)				**-w/I (K=10)**				**-w/I (K=15)**			
Resources	**Languages**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**	**F$_1$@3**	**F$_1$@5**	**F$_1$@$\mathcal{M}$**	**F$_1$@$\mathcal{O}$**
High	English	50.82	43.95	53.86	56.16	50.97	44.14	53.74	56.2	51.05	44.05	54.22	56.57
	Portuguese	27.92	27.84	25.8	27.39	27.61	25.63	26.39	27.77	27.2	25.09	25.43	26.12
	Spanish	44.49	34.18	41.62	42.02	43.65	35.13	41.47	42.7	42.27	33.52	40.37	41.72
	Russian	0.18	0.1	0.17	0.16	0.32	0.09	0.09	0.09	0.04	0.08	0.04	0.07
	French	47.92	37.53	48.42	47.91	46.82	39.32	48.44	49.7	47.05	38.62	48.42	49.48
	Chinese	0.16	0.17	0.07	0.3	0.07	0.11	0.09	0.12	0.18	0.2	0.21	0.24
	Arabic	0	0	0	0	0	0	0	0	0	0	0	0
	All	24.5	20.54	24.28	24.85	24.21	20.63	24.32	25.23	23.97	20.22	24.1	24.89
Low	Turkish	23.1	19.42	23.72	23.51	20.39	18.85	21.31	20.17	22.77	18.61	21.94	23.56
	Ukrainian	0	0	0	0	0	0	0	0	0	0	0	0
	Bengali	0	0.07	0	0	0.07	0.07	0.08	0.08	0.07	0.05	0.08	0.08
	Persian	0	0	0	0	0	0	0	0	0	0	0	0
	Nepali	0	0	0	0	0	0	0	0	0	0	0	0
	Urdu	0	0	0	0	0	0	0	0	0	0	0	0
	Gujarati	0	0	0	0	0	0	0	0	0	0	0	0
	Hindi	0	0	0	0	0	0	0	0	0	0	0	0
	Marathi	0	0	0	0	0	0	0	0	0	0	0	0
	Telugu	0	0	0	0	0	0	0	0	0	0	0	0
	Tamil	0.15	0.12	0.09	0.13	0.15	0.18	0.09	0.2	0.09	0.06	0.09	0.07
	Panjabi	0	0	0	0	0	0	0	0	0	0	0	0
	Indonesian	45.52	37.25	44.52	47.01	44.51	37.94	44.91	46.08	46.35	37.79	45.23	47.27
	All	5.29	4.37	5.26	5.43	5.01	4.39	5.11	5.12	5.33	4.35	5.18	5.46

Table 21: Tags Words Evaluation : **Flan-T5**. **Selected Content** (Important Sentences + Article)