Live Blog Corpus for Summarization

Avinesh P.V.S., Maxime Peyrard, Christian M. Meyer

Research Training Group AIPHES and UKP Lab
Computer Science Department, Technische Universität Darmstadt
www.aiphes.tu-darmstadt.de, www.ukp.tu-darmstadt.de

Abstract

Live blogs are an increasingly popular news format to cover breaking news and live events in online journalism. Online news websites around the world are using this medium to give their readers a minute by minute update on an event. Good summaries enhance the value of the live blogs for a reader but are often not available. In this paper, we study a way of collecting corpora for automatic live blog summarization. In an empirical evaluation using well-known state-of-the-art summarization systems, we show that live blogs corpus poses new challenges in the field of summarization. We make our tools publicly available to reconstruct the corpus to encourage the research community and replicate our results. `https://github.com/UKPLab/lrec2018-live-blog-corpus`

Keywords: Live blogs, Summarization Corpus, Corpus Construction, Focused Crawling, Online Journalism

1. Introduction

A live blog is a dynamic news article providing a rolling textual coverage of an ongoing event. It is a single article continuously updated by one or many journalists with timestamped micro-updates typically displayed in chronological order. Live blogs can contain a wide variety of media, including text, video, audio, images, social media snippets and links. At the end of the broadcasting, a journalist usually summarizes the main information about the event. For more extended events, journalists may also write intermediate summaries. Figure 1 and 2 show an example live blog provided by the *BBC* on "Last day of Supreme Court Brexit Case" and *The Guardian* on "US elections 2016 campaign". The timestamped information snippets are on the right, the human-written bullet-point summary is at the top left.
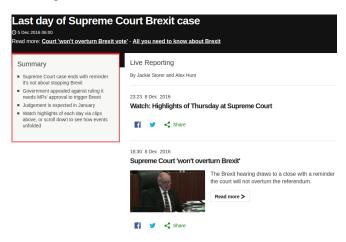

Figure 1: BBC.com live blog on "Last day of Supreme Court Brexit Case"

In the last decade, live-blogging has become very popular. It is commonly used by major news organizations, such as the *BBC*, *The Guardian* or *The New York Times*. Several different kinds of events are regularly covered by live blogs, including sport games, elections, ceremonies, protests, conflicts and natural disasters. Thurman and Schapals (2017, p.1) report a journalist's view that "live blogs have transformed the way we think about news, our sourcing, and ev-
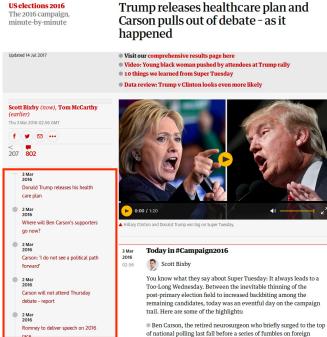

Figure 2: TheGuardian.com live blog on "US elections 2016 campaign"

erything". Thanks to this new journalistic trend, many live blogs – and their human-written summaries – are available online and new ones are generated every day.

In this work, we propose to leverage this data and investigate the task of automatic live blog summarization by crawling a new dataset. Live blog summarization has more direct applications in Journalism than the traditional but rather artificial tasks of a single document and multi-document summarization. Systems capable of summarizing live streams of heterogeneous content can be directly beneficial to users and even assist journalists during their daily work.

However, this new task also comes with new challenges. Live blogs are a list of short snippets of heterogeneous information and they do not form one coherent piece of text. The non-cohesive snippets make the task different from sin-

gle document summarization. Furthermore, most single documents are easily summarized by the baseline extracting the first few sentences. Such an approach is not effective for live blogs due to their heterogeneity and chronological order. The snippets are typically small, focused, numerous and rarely redundant which contrasts with the well-studied task of multi-document summarization. The topic is continually shifting and many sub-topics may arise and become central at some point. This even differs from the single topic shift found in classical update summarization tasks. For example, the live blog in Figure 1 on "Last day of Supreme Court Brexit case" consists of topic shift across "Supreme court judgment", "Government appeal", "Opinions of MP's on Brexit" and others. Moreover, when summarizing a live blog, one has to account for the whole past and all sub-topics previously discussed, which differs from real-time summarization setups like TREC.

We focus on two online news websites for acquiring live blogs, the *BBC*[1] and *The Guardian*[2], because they contain a lot of easily accessible live blogs that we automatically crawl and process.

In summary, our contributions are:

- We introduce a new task: live blog summarization.

- We suggest a pipeline to collect and extract live blogs with human-written summaries from two major online newspapers and release it for the community[3].

- We benchmark the dataset with commonly used summarization methods to stimulate further research into this challenging task.

The rest of the paper is structured as follows: Section 2. details existing summarization corpora and related works. Section 3. discusses our approach to collect live blogs from *BBC* and *The Guardian*, followed by a discussion on the statistics and properties of our live blog corpus in section 4.. The performance of well-established summarization baselines on this new dataset is discussed in section 5., followed by the conclusion and future work.

2. Related Work

In this section, we describe previous works related to summarization corpora. They were focused on single and multi-document summarization, update summarization, and real-time summarization. We are not aware of any previous work on live blog summarization.

Single and multi-document summarization. The most widely used summarization datasets have been published in the Document Understanding Conference[4] (DUC) series. In total, there are 139 document clusters with 376 human-written reference summaries across DUC '01, '02, and '04. Although the research community has often used these corpora, creating the manual summaries is time-consuming and labor-intensive.

Large datasets typically exist for single document summarization tasks, for example, the ACL Anthology Reference Corpus (Bird et al., 2008) and the CNN/Daily Mail dataset (Hermann et al., 2015). The latter contains large pairs of 312k online news articles and multi-sentence summaries used for neural summarization approaches (Nallapati et al., 2016; See et al., 2017). However, their dataset contains only one source document, whereas live blogs have a larger number of information snippets, typically more than 100.

Another recent work uses social media's reactions on Twitter to create large-scale multi-document summaries for news (Lloret and Palomar, 2013; Cao et al., 2016). Cao et al. (2016) use hashtags to cluster the documents into the same topic and use tweets with hyperlinks to generate optimal reference summaries. Their corpus consists of 204 document clusters with 1,114 documents and 4,658 reference tweets. Although this approach uses social media information to create a summarization corpus, they produce synthetic summaries, which are not written by a human. Moreover, they only use the corpus for training supervised learning approaches and not for evaluating summarization systems.

Other multi-document summarization datasets focus on heterogeneous sources (Zopf et al., 2016; Benikova et al., 2016; Nakano et al., 2010), multiple languages (Giannakopoulos et al., 2015), and reader-aware multi-document summaries (Li et al., 2017), which jointly aggregate news documents and reader comments.

Update summarization. After the DUC series, the Text Analysis Conference[5] (TAC) series ('08, '09) introduced the update summarization task (Dang and Owczarzak, 2008). In this task, two summaries are provided for two sets of documents and the summary of the second set of documents is an update of the first set. Although the importance of text to be included in the summary solely depends on the novelty of the information, the task usually observes only a single topic shift. In live blogs, however, there are multiple sub-topics and the importance of the sub-topics changes over time.

Real-time summarization. Real-time summarization began at the Text REtrieval Conference[6] (TREC) 2016 and represents an amalgam of the microblog track and the temporal summarization track (Lin et al., 2016). In real-time summarization, the goal is to automatically monitor the stream of documents to keep a user up to date on topics of interest and create email digests that summarize the events of that day for their interest profile. The drawback of this task is that they have a predefined time frame for evaluation due to the real-time constraint, which makes the development of systems and replicating results arduous. Note that live blog summarization is very similar to real-time summarization, as the real-time constraint also holds true for live blogs if the summarization system is applied to the stream of snippets. Moreover, the Guardian live blogs do consist of updated and real-time summaries, but this requires different real-time crawling strategies which are out of the scope of this work.

---

[1] http://www.bbc.com
[2] https://www.theguardian.com
[3] https://github.com/UKPLab/lrec2018-live-blog-corpus
[4] http://duc.nist.gov/

[5] http://www.nist.gov/tac/
[6] http://trec.nist.gov/

3. Corpus Construction

In this section, we describe the three steps to construct our live blogs summarization corpus: (1) live blog crawling yielding a list of URLs, (2) content parsing and processing, where the documents and corresponding summaries with the metadata are extracted from the URLs and stored in a JSON format, and (3) live blog pruning as a final step for creating a high-quality gold standard live blog summarization corpus.

Live blog Crawling. On the Guardian, a frequently updated index webpage[7] references all archived live blogs. We took a snapshot of this page that provided us with 16,246 unique live blogs.

In contrast, the BBC website has no such live blog archive. Thus, we use an iterative approach similar to BootCaT (Baroni and Bernardini, 2004) as described in Algorithm 1 to bootstrap a corpus utilizing a set of seed terms extracted from ten BBC live blog links from the web. The iterative procedure starts with a small set of seed terms (K_0) and gathers new live blog links using automated Bing queries[8] by exploiting patterns (P) in live blog URLs (i.e. "site:http://www.bbc.com/news/live/[key term]" as in line 5). We collect all the valid links returned by the Bing queries (line 6) and look for new key terms in the recently retrieved live blogs (line 10). In our implementation, key terms are terms with high TF*IDF scores. The new key terms are then used in the Bing queries of the subsequent iterations. The process is repeated until no new live blogs are discovered anymore (line 7). With this process, we ran 4,000 search queries returning each around 1,000 results on average and we collected 9,931 unique URLs.

Although our method collected a majority of the live blogs in the 4,000 search queries, a more sophisticated key terms selection could minimize the search queries and maximize the unique URLs. Additionally, this methodology can be applied to other news websites featuring live blogs like *The New York Times*, *Washington Post* or *Der Spiegel*.

An important point to note is that we find the collected BBC live blog URLs predominantly cover more recent years. This usage could be due to the Bing Search API preferring recent articles for the first 100 results. To collect a broad range of news articles the queries need to be precise.

Content Parsing and Processing. Once the URLs are retrieved, we fetch the HTML content, remove the boilerplate and store the cleaned data in a JSON file.

During this step, unreachable URLs were filtered out. We discard live blogs for which we could not retrieve the summary or correctly parse the information snippets. Indeed, live blogs can have changing patterns over time rendering the automatic extraction difficult.

Parsing of BBC live blogs can be automated easily because both bullet-point summaries and information snippets follow a consistent pattern. For the Guardian, we identify several recurring patterns which cover most of the live blogs.

---

[7] http://www.theguardian.com/tone/minutebyminute

[8] https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api

---

Algorithm 1 Iterative Live Blog Retrieval

1: **procedure** LIVEBLOGRETRIEVAL()
2:    **input:** Seed terms K_0, Live blog Pattern P
3:    $L_0 \leftarrow \emptyset$
4:    **for** $t = 1...T$ **do**
5:       $Q_t \leftarrow \text{makeQueries}(K_{t-1}, P)$
6:       $L_t \leftarrow \text{getLinks}(Q_t)$
7:       **if** $\cup_{i=0}^{t-1} L_i = \cup_{i=0}^{t} L_i$ **then**
8:          **return** $\cup_{i=0}^{t} L_i$
9:       **else**
10:         $K_t \leftarrow \text{extractKeyTerms}(L_t) - \cup_{i=0}^{t-1} K_i$
11:       **end if**
12:    **end for**
13: **end procedure**

Dataset	Crawling	Processing	Pruning
BBC	9,931	7,307	974
Guardian	16,246	6,405	1,681

Table 1: Number of topics for BBC and the Guardian

The Guardian live blogs were in use since 2001 but were in experimental phase till 2008. Due to the lack of a specific structure or a summary during this experimental phase, we remove 10k of the crawled live blogs. However, after 2008, live blogs have had a prominent place in the editorial with a consistent structure.

We parse metadata like URL, author, date, genre, summaries and documents for each live blog using site-specific regular expressions on the HTML source files.

After this step, 7,307 live blogs remain for BBC and 6,450 for Guardian.

Live blog Pruning. To further clean the data, we decided to remove live blogs exhibiting several topics as they can be quite noisy. For example, BBC provides some live blogs covering all events happening in a given region within a given time frame (e.g., *Essex: Latest updates*). We also prune live blogs about sport games and live chats, because the summaries are based on simple templates.

We further prune live blogs based on their summaries. We first remove a sentence of a summary if it has less than three words. Then, we discarded live blogs whose summaries have less than three sentences. This is to ensure the quality of the corpus, as overly short summaries would yield a different summarization goal similar to headline generation and they are typically an indicator for a non-standard live blog layout.

After the whole pruning step, 974 live blogs remained for BBC and 1,681 for the Guardian.

Overall, 10% of the initial set of live blogs, both for BBC and Guardian remained after selective pruning. This is to ensure high-quality summaries for the live blogs. Although the pruning rejects 90% of the live blogs, the size of the live blog corpus is 20–30 times larger than the classical corpora released during DUC, TREC and TAC tasks.

Code Repository. To replicate our results and advance research in live blog summarization we publish our tools

for reconstructing the live blog corpus open-source under the Apache License 2.0. The repository consists of (a) raw and processed URLs, (b) tools for crawling live blogs, (c) tools for parsing the content of the URLs and transforming content into JSON, and (d) code for calculating baselines and corpus statistics.

4. Corpus Statistics

We compute several statistics about the corpora and report them in Table 2. The number of documents (or snippets) per topic is around 95 for BBC and 56 for the Guardian. In comparison, standard multi-document summarization datasets like DUC '04[9] and TAC '08A[10] have only 10 documents per topic.

Furthermore, we observe that snippets are quite short as there is an average of 62 words per snippet for BBC and 108 for the Guardian. Summaries are also shorter than summaries in standard datasets. Indeed, in DUC2004 and TAC2008A summaries are expected to contain 100 words. Our corpora are larger because, together, they contain 2,655 topics and 186,999 documents. With many data points, machine learning approaches become readily applicable.

Statistic	BBC	Guardian
# topics	974	1,681
# documents	92,537	94,462
# documents / topic	95.01	56.19
# words / document	61.75	107.53
# words / summary	59.48	42.23

Table 2: Corpus statistics for BBC and the Guardian

Domain Distribution. Live blogs cover a wide range of subjects from multiple domains. In Table 3, we report the distribution of different domains in our combined datasets (BBC and the Guardian). While we observe that politics, business and news are the most prominent domains, there is also a number of well-represented domains like local and international events or culture.

Heterogeneity. The resulting corpus is expected of exhibiting various levels of heterogeneity. Indeed, there contain various topics with mixed writing styles (short-to-the-point snippets vs. longer descriptive snippets). Furthermore, live blogs are subject to topic shifts which could be observed by the change in words used.

To measure this textual heterogeneity, we use information theoretic metrics on word probability distributions like it was done before in analyzing the heterogeneity of summarization corpora (Zopf et al., 2016). Based on Jensen-Shannon (JS) divergence, they defined a measure of textual heterogeneity TH for a topic T composed of documents $d_1, \cdots, d_n$ as

$$TH_{JS}(T) = \frac{1}{n} \sum_{d_i \in T} JS(P_{d_i}, P_{T \setminus d_i}) \qquad (1)$$

Domain	# topics	proportion (%)
Politics	834	31.41
Business	421	15.86
General News	369	13.90
UK local events	368	13.86
International events	337	12.69
Culture	186	7.01
Science	60	2.26
Society	27	1.02
Others	53	2.00

Table 3: Corpus distribution across multiple domains for BBC and the Guardian

	BBC	Guardian	DUC '04	TAC '08A
TH_{JS}	0.5917	0.5689	0.3019	0.3188

Table 4: Average textual heterogeneity of our corpora compared to standard datasets

Here, P_{d_i} is the frequency distribution of words in document d_i and $P_{T \setminus d_i}$ is the frequency distribution of words in all other documents of the topic except d_i. The final quantity TH_{JS} is the average divergence of documents with all the others and provides, therefore, a measure of diversity among documents of a given topic.

We report the results in Table 4. To put the numbers in perspective, we also report the textual heterogeneity of the two standard summarization datasets DUC '04 and TAC '08A. These corpora were created during shared tasks and focused on multi-document news summarization. The heterogeneity in BBC and Guardian are similar and both much higher than DUC '04 and TAC '08A, meaning that our corpora contain more lexical variation inside topics.

5. Results and Analysis

In this section, we describe the automatic summarization methods and the upper bounds we compute for our live blog summarization dataset.

5.1. Baselines

As benchmark results, we employ methods that have been successfully used for both single and multi-document summarization. Some variants of them have also been applied to update summarization tasks.

TF⋆IDF (Luhn, 1958) scores sentences with the TF*IDF of their terms. The best sentences are then greedily extracted.
LexRank (Erkan and Radev, 2004) is a well-known graph-based approach. A similarity graph $G(V, E)$ is constructed where V is the set of sentences and an edge e_{ij} is drawn between sentences v_i and v_j if and only if the cosine similarity between them is above a given threshold. Sentences are then scored according to their PageRank in G.
LSA (Steinberger and Jezek, 2004) is an approach involving a dimensionality reduction of the term-document matrix via singular value decomposition (SVD). The sentences extracted should cover the most important latent topics.

Systems	BBC (L)			Guardian (L)			BBC ($2*L$)			Guardian ($2*L$)		
	R1	R2	SU4	R1	R2	SU4	R1	R2	SU4	R1	R2	SU4
TF*IDF	.227	.067	.064	.153	.021	.027	.367	.115	.147	.248	.037	.065
LexRank	.276	.080	.079	.188	.029	.038	.421	.138	.176	.297	.051	.089
LSA	.212	.046	.052	.135	.013	.021	.341	.084	.123	.220	.024	.051
KL	.267	.086	.080	.178	.026	.035	.397	.132	.165	.272	.041	.076
ICSI	.302	.104	.091	.210	.046	.046	.461	.176	.201	.322	.071	.101
UB-1	**.514**	.273	.218	**.422**	.177	.145	**.754**	.388	.435	**.640**	.256	.304
UB-2	.494	**.312**	.210	.389	**.230**	.137	.709	**.453**	.419	.584	**.334**	.277

Table 5: ROUGE-1 (R1), ROUGE-2 (R2), and ROUGE-SU4 (SU4) scores of multiple systems compared to the extractive upper bounds for ROUGE-1 (UB-1) and ROUGE-2 (UB-2) extractive for summary lengths of L and $2*L$

SoA system - ICSI

o They were detained over **corruption allegations in Zurich, Switzerland**.
o The BBC's 5 live Sport will run a special programme tonight at 19:00 GMT profiling the head of Fifa.
o Fifa executives accepted bribes to help secure the 2010 World Cup in South Africa, the US Attorney General Loretta Lynch has said.
o Mr de Gregorio also stresses that **Fifa President Sepp Blatter is not involved** in the criminal cases.
o As things stand the Fifa presidential election will go ahead on Friday.
o More reaction from around the world.

Extractive Upper Bound

o In a **separate** development, **Swiss prosecutors launched a criminal case into the 2018 and 2022 World Cup bids**, won by Russia and Qatar respectively.
o **The corruption case**, filed in the US, **involves alleged bribes worth** about **$ 150m** (£ 97m; €138m) **since the early 1990s**.
o **Six of the seven Fifa officials** arrested **in Zurich** are opposing their extradition to the US.
o They were detained over **corruption allegations in Zurich, Switzerland**.
o Fifa's incumbent **president Sepp Blatter is understood not to be one of those** arrested.

Reference Bullet-point Summary

o Fourteen sports officials indicted over **corruption charges** at the sport's governing body Fifa on 27 May.
o Seven of the 14 arrested **in Zurich, Switzerland** - **president Sepp Blatter is not among them**.
o One of those held is Jeffrey Webb - Fifa's vice-president.
o **The corruption case involves alleged bribes worth** more than **$150m since the early 1990s**.
o **Six of the seven** suspects **held in Zurich are contesting their extradition to the US**.
o **Separately, Swiss prosecutors launch a criminal case into the 2018 and 2022 World Cup bids**.

Figure 3: BBC.com live blog on "FIFA corruption inquiry"

KL-Greedy (Haghighi and Vanderwende, 2009) minimizes the Kullback-Leibler (KL) divergence between the word distributions in summary and the documents.

ICSI (Gillick and Favre, 2009) is a global linear optimization that extracts a summary by solving a maximum coverage problem considering the most frequent bigrams in the source documents. ICSI has been among the state-of-the-art MDS systems when evaluated with ROUGE (Hong et al., 2014).

5.2. Upper bound

For comparison, we compute two upper bounds. The upper bound for extractive summarization is retrieved by solving the maximum coverage of n-grams from the reference summary (Takamura and Okumura, 2010; Peyrard and Eckle-Kohler, 2016; P.V.S. and Meyer, 2017). This is cast as an Integer Linear Programming (ILP) and depends on two parameters: N, the size of n-grams considered and L, the maximum length of the summaries. In our work, we set $N = 1$ and $N = 2$ and compute the upper bound for ROUGE-1 (UB-1) and ROUGE-2 (UB-2) respectively.

5.3. Experimental Setup

We report scores for the ROUGE metrics identified by Owczarzak et al. (2012) as strongly correlating with human evaluation methods: ROUGE-1 (R1) and ROUGE-2 (R2) recall with stemming and stop words not removed. For completeness, we also report the best skip-grams matching metric: ROUGE-SU4 (SU4).

5.4. Analysis

Table 5 shows the results of benchmark summarization methods widely used in the summarization community on our live blog corpus. We explore two different summary lengths: L, length of the human-written bullet-point summary, and $2*L$, twice the length of the human-written summary to give leeway for compensating the excessive compression ratio of the human live blog summaries. The results show the state-of-the-art ICSI system is .2 ROUGE-1 and .3 ROUGE-2 lower than the upper bounds for both BBC and the Guardian with length constraint L and $2*L$ respectively. ICSI is only able to reach one-third of the upper bound, which emphasizes that live blog summarization is a challenging task and we need new techniques tackling live blog summarization.

Figure 3 shows the output of the ICSI system as compared to the extractive upper bound on BBC live blog on "FIFA corruption inquiry".[11] It can be seen that the ICSI system extracts sentences with most frequent concepts (e.g., FIFA, president, world cup), but misses to identify topic shifts in these information snippets. Although the information snip-

[11] http://www.bbc.com/news/live/world-europe-32897157

pets collected by the ICSI system are related to FIFA corruption, it misses capturing relative importance of the information snippets.

Additionally, factors which determine the difficulty of the summarization task are the length of the source documents and the summary (Nenkova and Louis, 2008). The input document sizes of the BBC and the Guardian are on an average 5,890 and 6,048 words, whereas the summary sizes are around 59 and 42 words respectively. Thus, the high compression ratio makes live blog summarization even more challenging.

6. Conclusion and Future Work

We introduce a new task: live blog summarization which has direct applications for journalists and news readers. Our goal is constructing a reference corpus for this new task. In this paper, we suggest a pipeline to collect live blogs with human written bullet-point summaries from two major online newspapers, which can be extended to live blogs from other news agencies like *The New York Times*, *Washington Post* or *Der Spiegel*.

We further analyze the live blog corpus and provide benchmark results for this dataset by applying commonly used summarization methods. Our results show that off-the-shelf summarization systems cannot be used, as they are far from reaching the upper bound. This calls for new solutions that take the task characteristics into account. As future work, we plan to research novel approaches to live blog summarization and investigate algorithms to identify important information from multiple topic shifts and a large number of information snippets.

Code for constructing and reproducing the live blog corpus and the automatic summarization experiments are published under the permissive Apache License 2.0 and can be obtained from `https://github.com/UKPLab/lrec2018-live-blog-corpus`.

Acknowledgments

7. Bibliographical References

Baroni, M. and Bernardini, S. (2004). BootCaT: Bootstrapping Corpora and Terms from the Web. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, pages 1313–1316, Lisbon, Portugal.

Benikova, D., Mieskes, M., Meyer, C. M., and Gurevych, I. (2016). Bridging the gap between extractive and abstractive summaries: Creation and evaluation of coherent extracts from heterogeneous sources. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING)*, pages 1039–1050, Osaka, Japan.

Bird, S., Dale, R., J. Dorr, B., Gibson, B., Joseph, M., Kan, M.-Y., Lee, D., Powley, B., Radev, D., and Fan Tan, Y. (2008). The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco.

Cao, Z., Chen, C., Li, W., Li, S., Wei, F., and Zhou, M. (2016). Tgsum: Build tweet guided multi-document summarization dataset. In *Proceedings of the Thirtieth Conference on Artificial Intelligence (AAAI)*, pages 2906–2912, Phoenix, AZ, USA.

Dang, H. and Owczarzak, K. (2008). Overview of the TAC 2008 update summarization task. In *Proceedings of the First Text Analysis Conference (TAC)*, pages 1–16, Gaithersburg, MD, USA.

Erkan, G. and Radev, D. R. (2004). LexRank: Graph-based Lexical Centrality As Salience in Text Summarization. *Journal of Artificial Intelligence Research*, 22:457–479.

Giannakopoulos, G., Kubina, J., Conroy, J., Steinberger, J., Favre, B., Kabadjov, M., Kruschwitz, U., and Poesio, M. (2015). MultiLing 2015: Multilingual Summarization of Single and Multi-Documents, On-line Fora, and Call-center Conversations. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 270–274, Prague, Czech Republic.

Gillick, D. and Favre, B. (2009). A scalable global model for summarization. In *Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing*, pages 10–18, Boulder, CO, USA.

Haghighi, A. and Vanderwende, L. (2009). Exploring Content Models for Multi-document Summarization. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 362–370, Boulder, CO, USA.

Hermann, K. M., Kočiský, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., and Blunsom, P. (2015). Teaching Machines to Read and Comprehend. In *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS)*, pages 1693–1701, Montreal, Canada.

Hong, K., Conroy, J., Favre, b., Kulesza, A., Lin, H., and Nenkova, A. (2014). A repository of state of the art and competitive baseline summaries for generic news summarization. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, pages 1608–1616, Reykjavik, Iceland.

Li, P., Bing, L., and Lam, W. (2017). Reader-Aware Multi-Document Summarization: An Enhanced Model and The First Dataset. In *Proceedings of the EMNLP Workshop on New Frontiers in Summarization*, pages 91–99, Copenhagen, Denmark.

Lin, J., Roegiest, A., Tan, L., McCreadie, R., Voorhees, E., and Diaz, F. (2016). Overview of the trec 2016 real-time summarization track. In *Proceedings of The Twenty-Fifth Text REtrieval Conference (TREC)*, Gaithersburg, MD, USA.

Lloret, E. and Palomar, M. (2013). Towards automatic tweet generation: A comparative study from the text

summarization perspective in the journalism genre. *Expert Systems with Applications*, 40(16):6624–6630.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research Development*, 2:159–165.

Nakano, M., Shibuki, H., Miyazaki, R., Ishioroshi, M., Kaneko, K., and Mori, T. (2010). Construction of Text Summarization Corpus for the Credibility of Information on the Web. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, pages 3125–3131, Valletta, Malta.

Nallapati, R., Zhou, B., dos Santos, C., Gulcehre, C., and Xiang, B. (2016). Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 280–290, Berlin, Germany.

Nenkova, A. and Louis, A. (2008). Can you summarize this? identifying correlates of input difficulty for multi-document summarization. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 825–833, Columbus, OH, USA.

Owczarzak, K., Conroy, J. M., Dang, H. T., and Nenkova, A. (2012). An assessment of the accuracy of automatic evaluation in summarization. In *Proceedings of Workshop on Evaluation Metrics and System Comparison for Automatic Summarization*, pages 1–9, Montreal, Canada.

Peyrard, M. and Eckle-Kohler, J. (2016). Optimizing an Approximation of ROUGE – a Problem-Reduction Approach to Extractive Multi-Document Summarization. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1825–1836, Berlin, Germany.

P.V.S., A. and Meyer, C. M. (2017). Joint Optimization of User-desired Content in Multi-document Summaries by Learning from User Feedback. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1353–1363, Vancouver, Canada.

See, A., Liu, P. J., and Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1073–1083, Vancouver, Canada.

Steinberger, J. and Jezek, K. (2004). Using latent semantic analysis in text summarization and summary evaluation. In *Proceedings of the 7th International Conference on Information Systems Implementation and Modelling (ISIM)*, pages 93–100, Rožnov pod Radhoštěm, Czech Republic.

Takamura, H. and Okumura, M. (2010). Learning to generate summary as structured output. In *Proceedings of the 19th ACM international Conference on Information and Knowledge Management*, pages 1437–1440, Toronto, Canada.

Thurman, N. and Schapals, A. K. (2017). Live blogs, sources, and objectivity: The contradictions of real-time online reporting. In *The Routledge Companion to Digital Journalism Studies*, pages 283–292. London/New York: Routledge.

Zopf, M., Peyrard, M., and Eckle-Kohler, J. (2016). The Next Step for Multi-Document Summarization: A Heterogeneous Multi-Genre Corpus Built with a Novel Construction Approach. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING)*, pages 1535–1545, Osaka, Japan.