BLN600: A Parallel Corpus of Machine/Human Transcribed Nineteenth Century Newspaper Texts

Callum W Booth, Alan Thomas, Robert Gaizauskas
University of Sheffield, Sheffield, United Kingdom
{cwbooth1, alan.thomas, r.gaizauskas}@sheffield.ac.uk

Abstract

We present a publicly available corpus of nineteenth-century newspaper text focused on crime in London, derived from the Gale British Library Newspapers corpus parts 1 and 2. The corpus comprises 600 newspaper excerpts and for each excerpt contains the original source image, the machine transcription of that image as found in the BLN and a gold standard manual transcription that we have created. We envisage the corpus will be helpful for the training and development of OCR and post-OCR correction methodologies for historical newspaper machine transcription—for which there is currently a dearth of publicly available resources. In this paper, we discuss the rationale behind gathering such a corpus, the methodology used to select, process, and align the data, and the corpus' potential utility for historians and digital humanities researchers—particularly within the realms of neural machine translation-based post-OCR correction approaches, and other natural language processing tasks that are critically affected by erroneous OCR.

Keywords: language resource, parallel corpus, ocr, transcription, newspapers, historical texts

1. Introduction

Historical documents present a number of unique challenges to automated digital transcription technologies, such as optical character recognition (OCR). In previous research into historical OCR, Holley (2009) found that the physical effects of the original media such as print and preservation quality, paper thickness, scan quality and contrast, and human-inflicted imperfections (such as ink transfer and fading in areas of frequent hand contact in the case of newspapers) act as confounding factors for OCR. Coupled with OCR systems that may lack the training needed to accommodate older typefaces (Springmann and Lüdeling, 2017) and newspaper layouts, it is clear that a means to correct historical newspaper OCR is still sought if historical document research is to be carried out where (1) high quality transcription is necessary for large scale automatic text analysis, such as text mining, to be carried out accurately and (2) full human re-key from source images is infeasible and even mass re-OCRing of source images processed in earlier digitisation projects using newer OCR technologies is not economically feasible. Work in this area is of particular importance for downstream natural language processing (NLP) tasks, the impact on which has been assessed by van Strien et al. (2020), who find that OCR quality degradation has a significant negative effect on a number of common NLP tasks. We present a language resource to facilitate the improvement of OCR and post-OCR correction systems, that also serves as a source of high quality historical text for NLP tasks—*BLN600*, a parallel corpus of source images, OCR transcriptions and manual transcriptions for 600 excerpts from the *British Library Newspapers (BLN) Corpus Parts 1 and 2* (Gale, 2024). This corpus was created as part of wider research into entity linkage between the *BLN* corpus and the *Digital Panopticon*[1], and consists of excerpts mostly pertaining to crime and reports of criminal justice processes that took place in nineteenth-century London.

We believe that this corpus will be a welcome addition to what we believe is a dearth of such resources. As we will show in section 2, the availability of parallel corpora within the realms of historical text resources, particularly with a focus on the nineteenth century, is quite limited.

In this paper, we discuss the compilation and alignment methodology of the *BLN600* corpus, and outline possible use cases both within the post-OCR correction context which prompted us to create this corpus, and within other contexts where a high quality source of gold-standard historical prose is required. We begin by reviewing relevant recent literature within the field.

2. Related Work

The task of OCR post-correction of historical texts and the study of the effects of poor quality OCR in historical research has seen consistent coverage within the literature (Kantner et al., 2011; Strange et al., 2014; Hu et al., 2020; Kettunen et al., 2022). Relatedly, previous research carried out as part of this work's wider project attempted to find methods of evaluating historical OCR where no parallelised gold standard existed (Booth et al., 2022). We may consider both the ICDAR2017 and 2019 *Competitions on Post-OCR Text Correction* as not only

[1]The Digital Panopticon is a structured dataset of the lives of historical UK criminals, available at https://digitalpanopticon.org

LREC-COLING 2024, pages 2440–2446
20-25 May, 2024.

indications of significant interest in and efforts towards a solution, but as contributions to the necessary language resources by introduction of parallel corpora (Chiron et al., 2017; Rigaud et al., 2019).
Per the work of Chiron et al. (2017), the *ICDAR2017* parallel corpus—produced by the National Library of France and the University of La Rochelle's L3i laboratory's *AmeliéOCR* project—comprises 12M characters of OCR, equally shared between English and French. From their analysis we can see that the 6M combined English OCR and ground truth characters cover, in total, a time period between 1744 and 1911 across British Library monograph and newspaper collections, with the gold-standard characters generated jointly between the National Library of France and external projects.
The *ICDAR2019* parallel corpus (Rigaud et al., 2019) expands on this further, comprising ≈22M OCR characters (≈754K tokens) across multiple European languages. English language OCR and gold standard contribute ≈243K characters, sourced from Papadopoulos et al. (2013)'s *IMPACT* dataset—a collection of images of newspapers, books, and other text-based images and accompanying reproduced gold standard compiled from European library sources.
Significant contributions towards multi-discipline parallel corpora with a focus towards OCR engine evaluation and research are made by Jiang et al. (2021), with the *Gutenberg-HathiTrust Parallel Corpus*, a parallel corpus of crowd-proofed and OCR'd documents, primarily belonging to fiction, business, medicine, social science, world war history, and agriculture domains.
We may also consider other collections such as *Eighteenth Century Collections Online (ECCO)*, a valuable resource to historians and digital humanities researchers, however not without caveats. For example, through the *Text Creation Partnership (ECCO-TCP)*, 2000 manually re-keyed full-text sources are available, but from a glance at the documentation[2] we see that the manual transcriptions replace the machine transcriptions in this instance, therefore it is not a parallel corpus. Other hindrances include the accessibility of the corpus—*ECCO* is a commercial product and therefore requires licensing in order to access the data. Additionally, the date ranges of publications within the set are non-comparable—*ECCO* covers eighteenth-century texts, the *ICDAR2017* and *BLN* corpora cover broader ranges encompassing the nineteenth century. *ICDAR2019*'s year coverage could not be verified from the literature.
Within the restricted range of the aforementioned resources, gold standard data appears to be created either in very specific circumstances (i.e. for competitions), to be of very broad scope, or to replace OCR rather than supplement it. We believe *BLN600*'s strength in relation to these corpora lies in its focus. *BLN600* provides a set of high-quality human transcriptions of mostly crime reports (with a number of counter-examples) from English-language London-centric newspapers, across a subset of the nineteenth century, alongside source images and alternative OCR engine output, making it what we believe is a unique resource for digital humanities researchers and historians interested in the study of crime in nineteenth-century London, the linguistics of nineteenth-century journalism, or the development and improvement of historical document OCR and post-OCR correction methodologies, as we discuss next.

3. Use cases

In this section we discuss various use cases in which we believe the *BLN600* corpus will prove itself of value to researchers.

3.1. Post-OCR correction model training

The *BLN600* corpus will be a valuable resource in the training and evaluation of post-OCR correction models as demonstrated by Thomas et al. (2024). We find in the literature many previous uses of such parallel corpora for post-OCR correction methods, typically employing neural machine translation models (Amrhein and Clematide, 2018; Hämäläinen and Hengchen, 2019; Nguyen et al., 2020, 2021; Soper et al., 2021). Some approaches lean toward the use of generated synthetic erroneous OCR via character-level insertion/deletion/substitution, or via the use of autoencoders such as *BART* (Lewis et al., 2020). Real data is, in our experience, much more likely to be useful than synthetic data in this situation (Li, 2021).
A model trained from this data, if successful, could prove useful to researchers of nineteenth-century crime journalism, by facilitating the correction of OCR text generated from historical newspapers, where re-keying or training specific OCR models to cope with image defects, typeface, and layout, is infeasible. We particularly see use in cases where OCR quality has affected downstream performance in other NLP tasks. We take for example the work of Pedrazzini and McGillivray (2022), who within their dataset documentation[3] mention the effects of OCR damage on diachronic linguistic analysis—a ≈73% misspelling rate within upstream OCR which necessitated dictionary error correction by Levenshtein distance, and merging of potentially erroneous embedding vectors, which is a computation-

[2] https://historicaltexts.jisc.ac.uk/collections#ecco

[3] https://github.com/Living-with-machines/DiachronicEmb-BigHistData#pre-processing

ally expensive practice on a dataset of that scope and one whose success is hard to assess.

3.2. Other use cases

The *BLN600* corpus provides a repository of gold-standard, manually-transcribed text covering nineteenth-century London-centric crime journalism. In context of the gold-standard side of the corpus, we see potential from the literature for other NLP-related tasks, such as named entity recognition and annotation, information extraction tasks, language modelling of nineteenth-century texts (Hosseini et al., 2021), and linguistic analysis (Pedrazzini and McGillivray, 2022). We plan to add additional layers of gold standard annotation to this corpus in subsequent stages of our research, including named entity annotation and annotation of criminal justice-related events[4]. Finally, since the BLN600 includes the original source images as well as gold standard manual transcriptions, it is also of potential use to researchers working on new approaches to improving OCR quality for historical texts.

4. Data Acquisition and Processing

The *BLN* corpus is vast, therefore a tractably re-transcribable sub-corpus was selected as follows:

1. **Querying:** over the initial *BLN* parts 1 and 2 data, a custom *Gale Digital Scholar Lab* query, shown in appendix A, was run by staff at Gale on our behalf, which returned 10K full newspaper page images with the corresponding metadata needed to locate the OCR within *Digital Scholar Lab*.

   The original intent of this corpus was to cover articles pertaining to crime within London-specific publications, hence the query used to return the images reflects this requirement. The query additionally reflects a requirement to stratify the results across decades—given the total size of the *BLN* corpus, the 10,000 image output cannot be guaranteed to be temporally homogeneous without specific intervention.

2. **Image selection:** from the resulting 10K full-page images, we selected 600 page images at random—without knowledge of the publication or year—based on whether a usable, legible excerpt pertaining to crime or criminal justice was present on the page. Some non-crime articles were permitted, to behave as counter-examples for criminal justice-specific work.

   Rejection of images was decided visually—reasons for rejection of a page image include lack of short articles, lack of relevant articles, missing OCR text, and the readability of the image. The image quality is highly variable across the *BLN* dataset, with some images being too faded or damaged to read. This is to be expected particularly with newspapers published at the beginning of the century—older newspapers suffer with quality issues in their digitised version, particularly those with faded low-contrast print, which is in turn more prominent in some publications. As the resulting image set from this phase was to be manually transcribed by a person, rejection of an image was decided simply by whether or not the authors could (1) read the image in its entirety, and (2) do so quickly, without needing to repeatedly read sections or rely too heavily on prior context to guess words, or without needing to edit the source image to increase contrast or sharpness.

   The count of 600 was chosen as it represented a compromise between time and resources available for human transcription work to be carried out. Additionally, this figure represents a compromise between a usable amount of data to be released for academic use, without negatively impacting the commercial interests of the parties that created the data.

3. **Human transcription:** from each full-page image, a single article or continuous section containing multiple articles was selected manually. The image was cropped to the region containing the article(s) of interest, and sent for re-keying to produce a gold-standard transcription.

4. **Machine transcription:** from the *BLN* corpus, the OCR text for the selected 600 pages was fetched. For each cropped section, the corresponding OCR text was gathered through a combination of manual alignment and automated search algorithms, resulting in an article-level alignment between the two transcription types.

In fig. 1, we illustrate a truncated example of the scan quality accepted in the image selection phase of data acquisition, along with the corresponding OCR text and the parallel re-keyed text, created from the original source image.

5. Analysis

The corpus in total consists of excerpts spanning a time period between 1834 and 1894, over six publications, totalling ≈1.7M characters (≈294K tokens) of manually re-keyed ground truth, averaging ≈500

[4]We are developing a set of justice-related event annotation guidelines inspired by the ACE English Annotation Guidelines for Events www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-events-guidelines-v5.4.3.pdf.

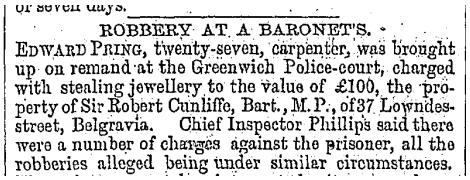
Source image crop	 *Illustrated Police News*. May 27 1882. "ROBBERY AT A BARONET'S.". In *British Library Newspapers*. Document ID: BA3200797029.
OCR	__RO BBERY AT A BARONET'S. v EDWARn PRING, twenty-seven, caipentbr, was bro1ught up on remand at the Greenwich Police-court, charged with stealing jewellery to the Value of 100, the propertyofSirRobert Cunliffe, Bart.,2,L1P., of37 Lot ndes- street, Beigravia. Chief Inspector Phillips said there were a number of cbarges against the prisoner, all the robberies alleged being tinder similar circumstances.
Gold standard	ROBBERY AT A BARONETS. EDWARD PRING, twenty-seven, carpenter, was brought up on remand at the Greenwich Police-court, charged with stealing jewellery to the value of £100, the property of Sir Robert Cunliffe, Bart., M.P., of 37 Lowndes street, Belgravia. Chief Inspector Phillips said there were a number of charges against the prisoner, all the robberies alleged being under similar circumstances.

Figure 1: Comparison between original source image crop, OCR text, and gold standard text.

tokens per document. In total, 939 individual articles are included as part of the 600 excerpts, with an 816/123 crime/non-crime split (≈87% crime)[5], for an average of ≈313 tokens per article.

Table 1 shows the distribution of transcriptions over discrete decade buckets and publication axes. Excerpts are biased towards two publications: *Lloyd's Weekly Newspaper*, and *The Illustrated Police News*. These biases are a result of the crime article bias during the image selection process carried out as documented earlier in this section—documents were initially selected based on the presence of substantial criminal justice content, and hence we would expect a bias towards publications such as *The Illustrated Police News*. Additionally, we see an increase in excerpt counts over time, starting at 1830. We reason this is a result of legibility requirements of the image selection process—images were rejected if the source image was illegible and would have presented issue to transcribers, a problem that is exacerbated by the age of the source material. It follows that scans of older documents were more likely to be rejected.

Character error rate (CER) in OCR quantifies error rates using Levenshtein distance, which compares OCR output to ground truth text by counting incorrect characters, and dividing this by the total number of ground truth characters. CER was computed between the *BLN* OCR and the manual transcriptions. As shown in Table 2, *BLN600* provides a useful middle ground in terms of both size and CER distribution with comparison to the *ICDAR* corpora. Preliminary analysis of the per-decade and per-publication CER distributions did not reveal any notable insights, however this may be explained by the unbalanced nature of the dataset. For *Lloyd's Weekly Newspaper*, the dominant publication in the dataset, CER ranges from 0.003 to 0.445, indicating substantial variability in OCR quality.

6. Conclusion and Future Work

In this paper we have presented the *BLN600* parallel corpus of machine and human transcribed nineteenth-century London-centric crime journalism. We have covered the approach taken to compile and align a selection of OCR'ed excerpts from the British Library Newspapers corpus parts 1 and 2 with a gold-standard version re-keyed from original source images. The corpus adds to the current language resource landscape for nineteenth-century journalism research, by providing a gold-standard source that may potentially be useful for post-OCR correction, natural language processing tasks, and linguistic analysis.

Our next steps for this corpus will include named entity and relation annotation of the gold standard, using a custom annotation schema tailored towards criminal justice events—based on the *ACE English Annotation Guidelines for Events*, and the Linguistic Data Consortium's *Annotation Guidelines for Individuality of Specific Entities*[6]. We believe this, coupled with the potential to expand the corpus, will add even more value to the resource in the future. We also plan to explore the application of newer OCR engines, such as Tesseract[7] to the source images to see what effect this has on recognition performance.

[5] Separate counts of crime and non-crime articles per excerpt are included in the dataset metadata, however no information on the positions of articles within the excerpts is given.

[6] https://tac.nist.gov/2016/KBP/guidelines/DEFT_ERE_Entities_IndividualGroup_Guidelines_V2.6.pdf

[7] https://tesseract-ocr.github.io/tessdoc

Publication	Decade (18-)							
	30	40	50	60	70	80	90	**Total**
Charter	4	2	-	-	-	-	-	**6**
Daily News	-	-	-	1	-	-	-	**1**
Illustrated Police News	-	-	-	-	-	36	176	**212**
Lloyd's Illustrated Newspaper	-	23	66	83	94	80	20	**360**
Morning Chronicle	13	-	-	-	-	-	-	**13**
The Era	-	1	6	1	-	-	-	**8**
Total	**17**	**26**	**72**	**85**	**94**	**110**	**196**	**600**

Table 1: Distribution of *BLN600* excerpts over publication decade and publication name. Columns for 1800 through to 1820 are omitted as *BLN600* contains no articles from this period, due to poor image quality judged during the image selection phase.

Dataset	Source	# character	μ CER	σ CER
BLN600	Gale BLN	1.7M	0.07	0.07
ICDAR2019	IMPACT	243K	0.21	0.20
ICDAR2017	BL Euro NP	1.8M	0.04	-
ICDAR2017	BL Monog	1.2M	0.01	-
ICDAR2017	GT BnF Eng	3.0M	0.02	-

Table 2: Character Error Rate of *BLN600* in comparison with the *ICDAR* corpora.

7. License, Access, and Permission

Express permission was sought from and granted by Gale on behalf of the company and the British Library partners—and communicated to the authors electronically—for the release of the OCR text of 600 individual excerpts from the British Library Newspapers corpus parts 1 and 2, under a non-commercial use-only license (CC BY-NC-ND 4.0) [8]. Permission was also sought from and granted by the British Library for the release of the accompanying images. *BLN600* is publicly accessible at https://doi.org/10.15131/shef.data.25439023.

8. Acknowledgements

The authors would like to thank Gale for granting us permission to release this corpus, and the British Library for allowing us to release source images. We'd also like to thank our anonymous reviewers for their helpful comments.

9. Bibliographical References

Chantal Amrhein and Simon Clematide. 2018. Supervised OCR error detection and correction using statistical and neural machine translation methods. *Journal for Language Technology and Computational Linguistics (JLCL)*, 33:49–76.

Callum Booth, Robert Shoemaker, and Robert Gaizauskas. 2022. A language modelling approach to quality assessment of OCR'ed historical text. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5859–5864, Marseille, France. European Language Resources Association.

Guillaume Chiron, Antoine Doucet, Mickaël Coustaty, and Jean-Philippe Moreux. 2017. ICDAR2017 competition on post-OCR text correction. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 01, pages 1423–1428.

Gale. 2024. British Library Newspapers. https://www.gale.com/intl/primary-sources/british-library-newspapers.

Mika Hämäläinen and Simon Hengchen. 2019. From the paft to the fiiture: a fully automatic nmt and word embeddings method for ocr post-correction. In *Proceedings of Recent Advances in Natural Language Processing*, pages 432–437, Bulgaria. INCOMA. Recent Advances in Natural Language Processing, RANLP ; Conference date: 02-09-2019 Through 04-09-2019.

Rose Holley. 2009. How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs. *D-Lib Magazine: The Magazine of the Digital Library Forum*, 15(3/4).

[8] http://creativecommons.org/licenses/by-nc-nd/4.0/

Kasra Hosseini, Kaspar Beelen, Giovanni Colavizza, and Mariona Coll Ardanuy. 2021. Neural language models for nineteenth-century english. *Journal of Open Humanities Data*.

Yuerong Hu, Ming Jiang, Ted Underwood, and J Stephen Downie. 2020. Improving Digital Libraries' Provision of Digital Humanities Datasets: A Case Study of HTRC Literature Dataset. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, pages 405–408.

Ming Jiang, Yuerong Hu, Glen Worthey, Ryan C Dubnicek, Boris Capitanu, Deren Kudeki, and J Stephen Downie. 2021. The Gutenberg-HathiTrust parallel corpus: A real-world dataset for noise investigation in uncorrected OCR texts.

Cathleen Kantner, Amelie Kutter, Andreas Hildebrandt, and Mark Püttcher. 2011. How to get rid of the noise in the corpus : cleaning large samples of digital newspaper texts.

Kimmo Kettunen, Heikki Keskustalo, Sanna Kumpulainen, Tuula Pääkkönen, and Juha Rautiainen. 2022. OCR quality affects perceived usefulness of historical newspaper clippings – a user study.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Mite Li. 2021. Using Deep Learning for Post-OCR Correction of 19th Century British Newspaper Text. University of Sheffield (MSc Dissertation).

Thi Tuyet Hai Nguyen, Adam Jatowt, Mickael Coustaty, and Antoine Doucet. 2021. Survey of post-OCR processing approaches. *ACM Comput. Surv.*, 54(6).

Thi Tuyet Hai Nguyen, Adam Jatowt, Nhu-Van Nguyen, Mickael Coustaty, and Antoine Doucet. 2020. Neural machine translation with BERT for post-OCR error detection and correction. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, JCDL '20, page 333–336, New York, NY, USA. Association for Computing Machinery.

Christos Papadopoulos, Stefan Pletschacher, Christian Clausner, and Apostolos Antonacopoulos. 2013. The IMPACT dataset of historical document images. In *Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing*, HIP '13, page 123–130, New York, NY, USA. Association for Computing Machinery.

Nilo Pedrazzini and Barbara McGillivray. 2022. Machines in the media: semantic change in the lexicon of mechanization in 19th-century British newspapers. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 85–95, Taipei, Taiwan. Association for Computational Linguistics.

Christophe Rigaud, Antoine Doucet, Mickaël Coustaty, and Jean-Philippe Moreux. 2019. ICDAR 2019 competition on post-OCR text correction. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pages 1588–1593.

Elizabeth Soper, Stanley Fujimoto, and Yen-Yun Yu. 2021. BART for post-correction of OCR newspaper text. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 284–290, Online. Association for Computational Linguistics.

U. Springmann and A. Lüdeling. 2017. OCR of historical printings with an application to building diachronic corpora: A case study using the RIDGES herbal corpus.

Carolyn Strange, Daniel McNamara, Josh Wodak, and Ian Wood. 2014. Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers. 8.

Alan Thomas, Robert Gaizauskas, and Haiping Lu. 2024. Leveraging LLMs for Post-OCR Correction of Historical Newspapers. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2024)*.

Daniel van Strien, Kaspar Beelen, Mariona Coll Ardanuy, Kasra Hosseini, Barbara Mcgillivray, and Giovanni Colavizza. 2020. Assessing the impact of OCR quality on downstream NLP tasks. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, volume 1: ARTDIGH, pages 484–496.

A. Initial DSL Query

The following represents the query used to select the initial 10,000 documents from which *BLN600* was formed. The authors cannot guarantee that the initial document search is reproducible from this query if recreated in Digital Scholar Lab. It is noted here for reference only.

```
Keyword: police court
Or Document Title: "police intelligence"
Or Document Title: "crime intelligence"
Not Document Title: "advertisements notices"

Publication Country: "England"
Publication Title: "Champion" Or "Charter" Or "Cobett's Weekly Political Register" Or "Daily News" or "The Era" Or "Examiner" Or "Graphic" Or "Illustrated Police News" Or "Lloyd's Illustrated Newspaper" Or "Morning Chronicle (1801)" Or "Morning Post" Or "The Standard"
Document Type: "Article"
Publication Section: "News"
Archive: Part I: 1800-1900 Or Part II: 1800-1900

Date: Jan 01, 1805 - Dec 31, 1814
Date: Jan 01, 1815 - Dec 31, 1824
Date: Jan 01, 1825 - Dec 31, 1834
Date: Jan 01, 1835 - Dec 31, 1844
Date: Jan 01, 1845 - Dec 31, 1854
Date: Jan 01, 1855 - Dec 31, 1864
Date: Jan 01, 1865 - Dec 31, 1874
Date: Jan 01, 1875 - Dec 31, 1884
Date: Jan 01, 1885 - Dec 31, 1894
```