







Reading Beyond the Center. Modeling Book Encounters in the Danish Periphery (1800-1850)

Alie Lassche¹ , Rie Eriksen¹ , Pascale Feldkamp¹ , Johan Heinsen² ,
Katrine Baunvig³ , and Kristoffer Nielbo¹ 

¹ Center for Humanities Computing, Aarhus University, Aarhus, Denmark

² MASSHINE, Aalborg University, Aalborg, Denmark

³ Center for Grundtvig Studies, Aarhus University, Aarhus, Denmark

Abstract

This paper investigates the reading culture of peripheral regions in nineteenth-century Denmark by analyzing book advertisements in six local Danish newspapers, a source that provides valuable insights into the characteristics of literary culture and its mediation through the local press. Drawing on previous studies and theories of reading culture, book circuit, and cultural peripherality, this study challenges center-oriented narratives in studies of Danish reading culture. Using a pipeline that combines article classification, book title extraction with generative language models, and manual genre clustering, this paper presents a dataset comprising 900,000 digitized news articles, along with over 6,000 book titles extracted from newspaper advertisements. These titles are categorized into *knowledge-* and *leisure-oriented* reading, revealing a more diverse literary landscape than previously assumed. The data indicate coexisting practices of *intensive* and *extensive* reading and show significant variation across towns in both genre distribution and circulation patterns. The analysis reveals that each town in the periphery had its own distinctive reading profile, demonstrating that literary engagement in these regions was varied and locally rooted, rather than uniform or marginal.¹

Keywords: historical newspapers, reading culture, periphery, book history, information extraction, GenAI

1 Introduction

The nineteenth century was a formative period in the development of Danish reading culture, as well as that of Europe more broadly [34]. As the publishing industry expanded and literacy became more widespread, the diversity of texts accessible to readers also increased. Traditionally, studies of nineteenth-century reading culture have primarily focused on the *production* of literary works, and only in a limited sense paid attention to the *spread* and *reception* of books [18]. Moreover, when nineteenth-century Danish book culture has been addressed, it has often centered on urban, educated elites – especially in relation to the literary movement of the Modern Breakthrough, which dominated cultural life in the late nineteenth century and was largely shaped by and for the cultural elite [5; 6; 7].

That said, nineteenth-century reading culture was not limited to literary works alone, nor was the circulation of books confined to the upper segments of society. Although Copenhagen was already the largest city in Denmark, it had only approximately 130,000 citizens around 1850. The

Alie Lassche, Rie Eriksen, Pascale Feldkamp, Johan Heinsen, Katrine Baunvig, and Kristoffer Nielbo. “Reading Beyond the Center. Modeling Book Encounters in the Danish Periphery (1800-1850).” In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 95–118. <https://doi.org/10.63744/YxKHrB4ZCThp>.

© 2025 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

¹ To ensure reproducibility, code, embeddings, and raw data are available via <https://github.com/centre-for-humanities-computing/chr-book-ads>.

vast majority of Danes (around 80% in 1840) lived in the countryside [10; 27]. Most of them were farmers and agricultural servants. By studying their literary horizon, we expand our knowledge about the Danish book landscape, its diversity, nineteenth-century reading culture, as well as center/periphery dynamics in a historical information society.

In this paper, we explore the dissemination of books in peripheral regions of nineteenth-century Denmark, focusing on the literary horizon of communities geographically distant from major cultural centers (see Figure 1). These geographical areas have long and are still being viewed as marginal in relation to the urban and institutional core.² Because this segment of the population left behind few traceable records, we face significant gaps in traditional sources such as library holdings, lending registers, and other indicators of reading culture. To gain insight into reading culture in the periphery, we must turn to alternative sources. One such source are newspaper book advertisements – often placed by booksellers [3] – which have been shown to reflect local demand and circulation patterns [12].

Book advertisements in newspapers only constitute one component of the commercial interface between booksellers and the readers, but their relative prominence must be understood in light of the specific conditions of early nineteenth-century Denmark, namely censorship and the Danish state bankruptcy of 1813. Censorship imposed substantial restrictions on the circulation of printed material, thereby limiting the extent to which alternative media such as pamphlets or posters could be freely employed [17]. Newspapers publishers, by contrast, often held licenses permitting them to advertise and were more resilient in financial terms [33]. This contributed to making newspapers not only the most stable medium for commercial communication but also the most consistently expanding medium throughout the century [18].

While book advertisements can thus not be assumed to reflect exactly what was read, they offer a valuable proxy for what was available, promoted, and likely to have circulated locally.³ As commercial texts responding to local demand and aimed at a broad regional public, they provide a more immediate and inclusive snapshot of reading culture – particularly in areas underrepresented in formal archives, where lending libraries were scarce and state-library infrastructure limited [18].⁴ Moreover, advertisements may capture demand across categories more evenly. Whereas lending library catalogues – the most common source for book circulation – tend to reflect a selective image of book culture skewed toward literary fiction, books for personal or devotional use, such as psalmbooks and religious texts, were more often purchased than borrowed [18, p. 345].

We therefore study book advertisements in local newspapers to gain insights into the types of texts available and read, how reading culture was mediated through the local press, and how these peripheral regions were connected and related to one another in terms of reading culture. We seek to answer three simple questions about reading culture in the Danish periphery in this paper: *what did people read, how did people read, and where did people read what?*

This paper is structured as follows: section 2 reviews related work on nineteenth-century reading culture and information extraction from historical texts. In section 3 we provide a description of the formation and content of the corpus used in this study. Our methodological pipeline is outlined in section 4, which includes the training of classifiers for article categories and book announcements, the extraction of book titles using generative language models, and the clustering procedure applied to the extracted book titles. In section 5, we will answer the questions outlined above. In the final section, we present the conclusions and offer suggestions for future research.

² The terms *Udkantsdanmark* (Peripheral Denmark) and *den rådne banan* (the rotten banana) are still frequently used in political and public discourse in Denmark. The ‘rotten banana’ refers to the outer regions of Denmark, which curve around the country in a banana-like shape. It reflects a perception that these rural areas are underdeveloped and place an excessive financial burden on society [9; 37].

³ Provincial booksellers would either acquire books from bigger booksellers in Copenhagen and pay them back later or they would avoid debt by only buying books that they knew were on demand [3].

⁴ Note that lending libraries often were owned by the booksellers [18].



Figure 1: Map of Denmark with in dark green the towns whose newspapers are included in our corpus.

2 Related Work

2.1 Nineteenth-century reading culture

The nineteenth century saw transformative changes in European book and reading culture, driven by technological advances, lower production costs, and socio-cultural shifts. In Denmark, improved printing technologies and the introduction of compulsory education in 1814 significantly widened access to printed materials, extending literacy beyond urban elites and into broader segments of the population, including rural communities [16; 18]. Scholarly debates about reading practices have traditionally highlighted a shift from *intensive* reading (deep engagement with a limited selection of books) towards *extensive* reading (covering a wider range of texts). This shift was described by Engelsing [13] as the ‘reader revolution’, driven by a rising bourgeois reading public in the late eighteenth and early nineteenth centuries. However, recent research suggests that this development was neither uniform nor immediate, particularly in rural or peripheral regions. Studies from the Netherlands, for instance, indicate that middle-class readers lacked literary socialization, often struggling to navigate a rapidly expanding literary market [25]. Similar dynamics can be assumed for rural Denmark, where existing research highlights a strong dominance of religious literature among rural readers [19].

Yet, despite these insights, detailed empirical evidence on reading practices in nineteenth-century Danish rural communities remains sparse. Existing studies largely employ qualitative methods and have predominantly focused on urban centers or earlier periods [2; 3; 18; 28]. This creates significant gaps, especially regarding what rural populations actually read, how they accessed literature, and how reading habits evolved over time. Newspapers, alongside the novel, played a key role in addressing these gaps. As Benedict Anderson [1] influentially argued, newspapers were instrumental in shaping national public spheres by simultaneously circulating information and cultural narratives to geographically dispersed readers, thereby fostering shared identities. By providing regular updates on books available to readers, newspaper book advertisements functioned as crucial mechanisms of literary dissemination, selection, and reader formation in peripheral regions.

In Denmark, newspapers significantly influenced the literary culture of peripheral areas by serving as a primary source of information about available literature. This influence was partic-

ularly important given the scarcity of public libraries or established literary institutions outside major urban centers. Consequently, examining newspaper book advertisements provides direct insight into how literary materials were marketed to, and perceived by, peripheral readers who may not yet have developed clear strategies for selecting literature – thus offering a clearer picture of literary socialization processes outside elite urban circles. Moreover, in *What is the History of Books?* [12], Robert Darnton highlights the key role of booksellers in the circulation of literature. His example of Isaac-Pierre Rigaud, who ordered Voltaire’s *Questions sur l’Encyclopédie* based on customer demand, illustrates how a bookseller’s stock can reflect a community’s literary interests.⁵

In summary, we identify three main gaps in the existing literature. First, peripheral reading cultures remain underexamined in quantitative research. While recent book history has adopted computational methods, these have primarily focused on urban or institutional contexts, often overlooking rural regions and reading as a daily, non-prestigious activity. Second, empirical evidence on nineteenth-century Danish reading practices is scarce. Aside from case studies, there is little systematic knowledge of the types of books that were read or made available outside Denmark’s urban centers – or of how reading habits evolved, for instance, from intensive to more extensive modes of reading. Third, the mechanisms of book dissemination remain understudied concerning reader formation. Despite Darnton’s call to attend to intermediaries like booksellers, little attention has been paid to how book advertisements functioned as selective interfaces between supply and demand, and in curating literary socialisation – particularly in regions with weak literary infrastructure. By directing our attention to peripheral communities – where the middle and lower classes made up 97% of the population in 1801 and 1834 [20] – and using newspaper book advertisements as a proxy for literary availability and interest, we are better positioned to examine reading culture beyond the cultural elite.

2.2 Information extraction from historical texts

Extracting information from historical texts is challenging due to factors like spelling variation, OCR errors, and the complexity of low-resource languages such as Danish. While Named Entity Recognition (NER) is commonly used for structured data extraction in historical sources [23; 24; 30], it is not well-suited to the task of identifying book titles, which are highly variable, context-dependent, and often embedded in noisy advertisement layouts. Previous work has shown that even combining manual and automated methods for detecting book-related content remains a complex, interpretive task [31].

The emergence of generative language models (GenAI) has opened new possibilities for extracting information from historical corpora [22], though performance still varies by domain and task [8; 11; 36]. For this study, we developed a tailored approach that combines LLM-based extraction with human validation, specifically designed for the structure and ambiguity of nineteenth-century Danish book advertisements.

3 Data

Our dataset comprises digitized articles from six nineteenth-century Danish local newspapers, which were selected based on their distribution in peripheral regions of Denmark. This includes

⁵ How we study reading culture really depends, of course, on how we choose to define it – and that remains an open, and surprisingly underdiscussed, question, both conceptually and methodologically. In peripheral and historical settings like this one, the challenge is sharper: sources are sparse, indirect, and often shaped by the institutions we are trying to look past. Book advertisements offer one possible proxy for reader interest and literary availability, while naturally capturing only one moment in the circulation process: the commercial interface between supply and anticipated demand. Other routes are possible: marginalia, lending records, letters, even oral accounts. But each method makes a different kind of reading visible – and leaves others out.

newspapers from the towns of Aalborg, Aarhus, Maribo, Odense, Thisted, and Viborg (see Figure 1). All newspapers are digitally available at the Danish Royal Library’s online platform.⁶ Although OCR-processed, the digitized texts exhibit a high word-level error rate, generally around 50%. This is due to deterioration, small print, and cheap paper, but also cost-cutting decisions made by the Danish Royal Library to digitize from existing microfilm copies. Faulty column identification is another issue.

A re-digitization initiative has been led by the ENO project [15],⁷ which involves re-OCRing the newspapers using models developed in Transkribus [21], a tool originally designed for Handwritten Text Recognition (HTR). The word-level error rates are typically less than 5% for the re-OCRred data, and the column segmentation is near faultless. ENO is an ongoing project, beginning with the oldest newspapers and progressing forward.⁸ The current study utilizes reprocessed newspapers from the first half of the nineteenth century, when newspapers had become widely available across Denmark. While they do not all cover the same period, the overlapping time span across all titles is from 1824 to 1838.⁹ The size of each newspaper dataset differs, reflecting the frequency with which the newspapers appeared at the time. Refer to Table 1 for statistics on the corpus.

Newspaper	Population	Period	Frequency/week	Editions	Articles	Words
Aalborg	6,500	1818-1842	3-5	5,075	156,011	19,872,861
Aarhus	6,000	1794-1845	2-4	8,288	258,414	28,130,531
Maribo	1,000	1809-1838	2-3	2,990	85,092	7,845,803
Odense	8,000	1809-1848	4-6	8,240	202,178	31,558,778
Thisted	1,600	1824-1848	1-3	3,238	72,102	9,468,784
Viborg	3,100	1809-1842	2-4	5,161	169,638	17,421,450
<i>Total</i>				32,992	943,435	114,298,207

Table 1: Descriptive statistics on the corpus and subcorpora. The *population* is expressed in the number of inhabitants of that town, taken from the average of inhabitants in 1801 and 1845. These years were selected because they correspond with national census data collection periods in Denmark. The *frequency* column describes the number of times a newspaper appeared per week. The number of editions refers to the number of unique newspapers from that town in our dataset.

As part of the ENO-initiative, the raw output from Transkribus has been post-processed through a pipeline that combines rule-based components with modeling based on structural features at the line level. Features include elements such as line length, punctuation, and hyphenation. The aim is for each newspaper edition to be segmented into its individual articles. The complexity of this task varies across newspapers and is heavily dependent on their layout. The segmentation process is ongoing and still imperfect, with common errors including the merging of unrelated text segments or the splitting of single articles into smaller fragments – both of which complicate downstream categorization using LLMs.

4 Methods

Figure 2 shows the four-step pipeline used to extract and cluster book titles. The subsequent sections provide a detailed description of each step.

⁶ <https://www.mediastream.dk>.

⁷ Hosted by *Masshine* and *The Historical Data Lab* at Aalborg University: <https://hislab.quarto.pub/en/>.

⁸ The most recent version of the ENO-corpus is available on <https://huggingface.co/datasets/JohanHeinsen/ENO>.

⁹ The version of the corpus used in this paper, is available on <https://huggingface.co/datasets/chcaa/chr-book-ads-articles>.

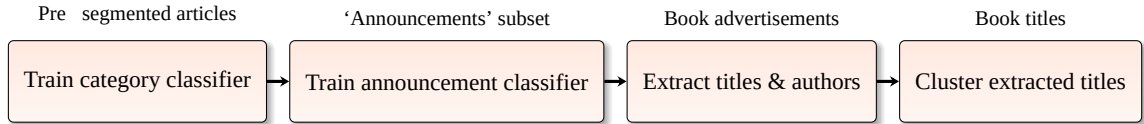


Figure 2: Four-stage pipeline for title extraction and clustering.

4.1 Training classifier for article categories

In the steps taken to post-process the articles, a separate set of rules was applied to assign a category to each article: National News (*Indenlandsk/Fædrelandet*), International News (*Udenlandsk/Blandet*), Announcements (*Bekjendtgørelser*), and a small section for Reader Mail (*Indsendt*). The accuracy of this categorization step varies by newspaper and depends largely on layout structure. Manual inspection confirmed that the Maribo newspaper achieved accurate segmentation and categorization, owing to its particularly clear and consistent layout. To evaluate the performance of different text representations for article classification, we used a balanced sample (24,000 articles for each category) of pre-categorized articles from the Maribo newspaper as a testing ground. We experimented with two types of representations: TF-IDF and semantic embeddings generated using the `multilingual-e5-large` model.¹⁰ Because the *Indsendt* category was frequently misclassified as *Indenlandsk*, we chose to merge it into the latter category, seeing that our main goal was to label the ‘Announcements’ correctly. Proceeding with a three-class classification task, we trained and evaluated a Logistic Regression model using a 70/30 train-test split.¹¹ Both representations yielded similar results, with F1-scores for individual classes ranging between 0.84 and 0.95.

The trained model was then used to predict the category of an unseen set of 1,000 articles, taken from the remaining newspapers. We manually evaluated the predicted category for each article – creating a balanced gold standard – and calculated the classification accuracy for both feature types. Since the semantic embeddings slightly outperformed the TF-IDF features (with accuracies of 0.91 vs. 0.89), we used semantic embeddings in the subsequent classification. For the final classification model, we expanded the gold standard with a random sample of 200 additional articles from the correctly-segmented Maribo newspaper, ending up with a gold standard in which all newspapers were equally represented. The model trained on this benchmark dataset, whose performance metrics are included in Table 2, was then used to predict the category of the remaining unlabeled articles, resulting in the distributions shown in Figure 3.

Class	Precision	Recall	F1-score
Announcements	0.88	0.82	0.85
National news	0.72	0.82	0.77
International news	0.89	0.82	0.85

Table 2: Performances of the classification task to label newspaper articles by category.

4.2 Training classifier for book announcements

We created a gold standard sample from the Announcements category, consisting of articles that mention book titles. This involved manually evaluating articles that contained search terms we

¹⁰ <https://huggingface.co/intfloat/multilingual-e5-large>. Note that the articles were divided into chunks of the maximum length supported by the model (514). Each segment was individually embedded, and the mean of these embeddings was used to represent the article as a whole. The embeddings of the articles used in this study are available on <https://huggingface.co/datasets/chcaa/periphery-aviser-e5>.

¹¹ We used the `scikit-learn` package in Python [29].

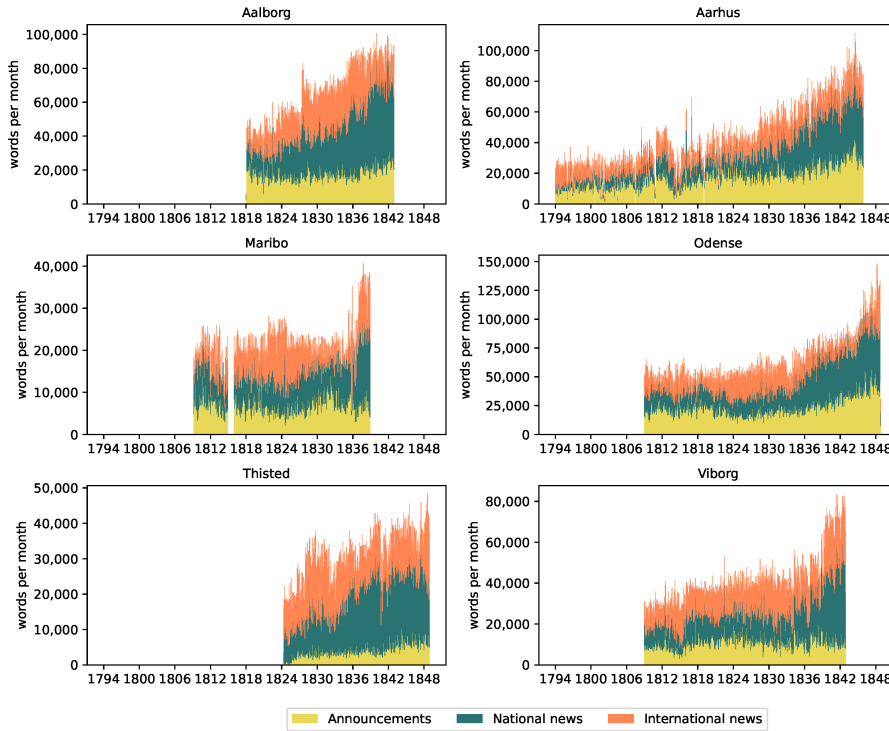


Figure 3: Number of words per month for the six newspapers in our corpus. The colors correspond with the type of article.

had previously identified as indicators of book announcements.¹² From the manually assessed articles, we selected a balanced sample of 540 – half book announcements and half non-book announcements – to train a Logistic Regression model. The model was evaluated using fivefold cross-validation, with semantic embeddings serving as input features. Performance metrics are presented in Table 3. We then applied the model trained on the fifth fold to the remaining articles in the Announcements category, identifying 7,531 likely book announcements.¹³

Class	Precision	Recall	F1-score
Book announcement	0.91	0.92	0.91
Non-book announcement	0.92	0.90	0.91

Table 3: Performances of the classification task to label book announcements.

¹² We used the following search terms: ‘indb.’, ‘heftet’, ‘udkomm’, ‘udgiv’, ‘udgav’, ‘indbund’, ‘oversatt’, and ‘paa adressekontoiret’. We acknowledge that this list introduces a potential bias in the gold standard, as it may exclude book advertisements that do not contain these terms. However, due to the highly formulaic and programmatic nature of advertisement phrasing in these newspapers, we expect this approach to capture the vast majority of book-related announcements.

¹³ Before making predictions for the remaining articles in the Announcement category, we filtered out announcements that appeared five times or more in the dataset, since manual inspection showed that these were never book announcements. We also filtered out all announcements that reported on lottery drawings, as well as announcements shorter than 70 characters and longer than 500 characters.

4.3 Extracting book titles

To extract book titles and authors from the classified book announcements, we employed a generative language model via the OpenAI API. Specifically, we used the `gpt-4o-mini` model, configured with `temperature = 0.3` and `max_tokens = 800`. For each article, we supplied the full text of the announcement and prompted the model to return a list of mentioned book titles and their authors, along with an English translation of each title. We experimented with several prompt formulations to determine which would yield the most accurate extractions. To evaluate the adequacy of each prompt, we used a manually annotated gold standard of 300 book announcements. This sample was randomly created and curated by two human annotators.¹⁴ For every book announcement, annotators extracted the book titles, authors, and the English translation of the title. An example illustrating the structure of this annotated sample is presented in Table 4. When a book announcement did not contain a book title, we filled in ‘NO_BOOK’ in every column.

article_id	original_title	translated_title	author
aal_007242	Jliade	Iliad	Homer
aal_028045	Bibelen forsvarer sig selv	The Bible defends itself	Biskop Balle
ode_193405	Kjerlighedens Gierninger	The Works of Love	S. Kierkegaard

Table 4: Example of the curated set of gold standard book announcements.

Extraction quality of the various prompts was assessed using fuzzy matching on the gold standard set. For each model output, we required an exact match on the `article_id`, and a partial match on either only the `original_title` or both the `original_title` and the `author`. We used the `fuzz` function in the `rapidfuzz` package and experimented with thresholds ranging from 65 to 80.

In addition to the three prompts we manually designed, we also leveraged the self-improving capabilities of generative language models by implementing a feedback loop [38]. After generating outputs for a subset of book announcements using an initial prompt, we provided the model with ten of its true positives and ten of its false positives. We then asked the model to reflect on discrepancies and suggest improvements to the prompt formulation itself. This process allowed the model to iteratively refine its own instructions based on its performance, enabling us to test whether prompt optimization through in-context feedback could yield higher extraction accuracy. We compared prompts (P) based on their precision, recall, and F1-score, which are included in Table 7 in Appendix B. P1, P2, and P3 are the prompts that were manually created, while P4, P5, P6, and P7 were created in the feedback loop of the generative model.¹⁵ While the performances of P4 and P5 were considerably lower than that of our own prompts, they were matched and slightly outperformed by P6 and P7.

The prompt that yielded the best results struck the right balance between providing enough guidelines without overcomplicating the task. This prompt (P6 in Appendix A) was used for extracting titles from all predicted book announcements, resulting in 6,391 extracted titles. Table 5 contains more detailed statistics on the extracted titles.

4.4 Clustering extracted book titles

After post-processing the extracted titles, all entries were manually assigned to semantic categories by the same two annotators, along with two student assistants with backgrounds in literary studies.

¹⁴ These were a graduate student in religious studies and a postdoctoral researcher in history, both with literary backgrounds and expert knowledge of historical newspaper material.

¹⁵ All prompts are included in Appendix A.

Newspaper	Announcements	Book announcements	Titles
Aalborg	62,615 (40%)	560 (0.9%)	965
Aarhus	113,783 (44%)	848 (0.7%)	1,313
Maribo	35,817 (42%)	629 (1.8%)	1,049
Odense	82,999 (41%)	890 (1.1%)	1,568
Thisted	15,006 (20%)	222 (1.5%)	642
Viborg	61,635 (36%)	556 (0.9%)	854
<i>Total</i>	371,855	3,705	6,391

Table 5: Number of articles, announcements, and book titles after extraction.

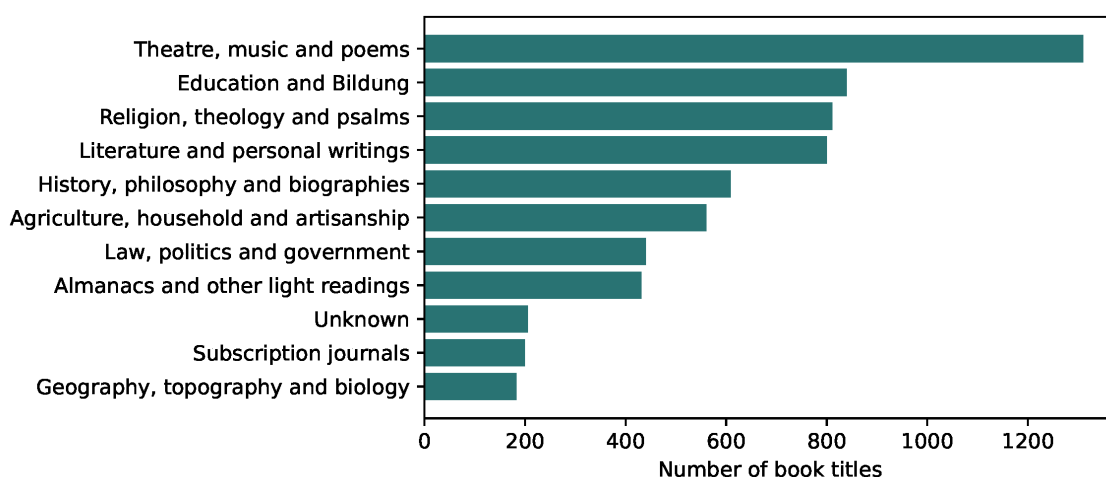


Figure 4: Number of book titles in each manually assigned category in the corpus.

The categories were derived from the corpus itself: through repeated close reading and annotation, the annotators identified recurring themes and patterns in the titles. Based on internal discussions, this resulted in the following categories: *Agriculture, household and artisanship*; *Almanacs and other light readings*; *Education and Bildung*; *Geography, topography and biology*; *History, philosophy and biographies*; *Law, politics and government*; *Literature and personal writings*; *Religion, theology and psalms*; *Subscription journals*; *Theatre, music and poems*; and *Unknown*.¹⁶

5 Results

5.1 Knowledge- and leisure-oriented reading

The horizontal bars in Figure 4 show the distribution of the 6,391 book titles across the manually assigned categories. Upon closer inspection, the ten substantive categories (excluding *Unknown*) fall into two broader thematic clusters. The first comprises texts oriented toward knowledge acquisition, instruction, or moral and civic education. This includes, first and foremost, the categories *Education and Bildung* (largely schoolbooks) and *Agriculture, household and artisanship* (practical manuals for work in and around the rural household), as well as the categories *Law, politics and government*, *Geography, topography and biology* and *Religion, theology and psalms*, the latter primarily concerned with spiritual development. We can summarize these texts as ‘knowledge-oriented reading’. The second cluster comprises texts related to leisure, cultural consumption, and

¹⁶ The *Unknown* category includes books for which no additional information could be found online, and whose titles provided no clear indication of their topic or theme. These were mostly anonymous works with very short, generic titles.

personal enjoyment. This encompasses the categories *Theatre, music and poems, Literature and personal writings, Almanacs and other light readings*, and *History, philosophy and biographies*. These fall under what we call ‘leisure-oriented reading’.

Among the 6,391 book titles we extracted, we distinguished 4,141 unique book titles.¹⁷ Of these, 3,171 appear only once in the dataset. The remaining 970 book titles are advertised between 2 and 55 times. The 20 titles that appear most frequently in advertisements are listed in Table 9 in Appendix C. Most of these fall into the cluster of ‘knowledge-oriented reading’, including several religious handbooks and psalmbooks, handbooks for reading and mental arithmetic, and a recipe book. In terms of authorship, the dataset contains 1,706 unique names – 1,049 appear only once, while the remaining 657 authors are mentioned between 2 and 144 times. The 20 most frequently advertised authors are listed in Table 8 in Appendix C. This list includes figures associated with ‘knowledge-oriented reading’ – such as A. F. Just, N. Blicher, and C. Jacobsen – as well as prominent literary authors tied to ‘leisure-oriented reading’, including Adam Oehlenschläger, August von Kotzebue, Walter Scott, Ludvig Holberg, and H. C. Andersen.

Together, these categories and clusters reveal that the reading culture in nineteenth-century peripheral Denmark was far more diverse than previously assumed. While *Religion, theology and psalms* – the category most expected based on previous studies [19] – does appear in the top three with its 811 titles, it accounts for less than 13% of the overall corpus. Notably, the category *Theatre, music and poems* is even more prominent, with 1,310 titles – 1.5 times as many. Moreover, we observe that ‘leisure-oriented reading’ encompasses both high-culture and popular texts. Next to titles of the above mentioned well-known literary authors, the corpus also contains a substantial number of almanacs. Particularly noteworthy are the frequently occurring German ‘Taschenbücher’. These small, portable pocketbooks were designed for convenient reading and typically featured novellas, essays, and (often romantic) short stories, in addition to the poetry and calendar content characteristic of almanacs more generally. Several of them specifically targeted the female reader, being titled ‘Frauentaschenbuch’ [35]. In addition to these popular publications, the dataset includes more than 300 German titles, suggesting that readers in the Danish periphery also read works in German alongside Danish literature.

5.2 Intensive and extensive reading

To get more insight into the longevity of a book title, we examine its advertising span – that is, the time between the first and last time a book title appears in an advertisement within our corpus. Across all categories, most titles appear in advertisements for only one day. Table 6 contains the percentages of books per category that have only been advertised for one day, and it shows they vary between 66% (*Religion, theology and psalms*) and 93% (*Unknown*). In the box plots in Figure 5, we show the distribution of the books that have been advertised for more than one day.

The varying lengths of the boxes in Figure 5 reflect notable differences in sustained visibility across categories. Titles in the categories *Religion, theology and psalms* and *Education and Bildung* tend to have markedly longer advertising spans, ranging from 16 to 38 years. This suggests persistent availability and enduring demand, likely tied to the long-term relevance and repeated use of these works. Such patterns resonate with what Engelsing has described as the *intensive* reading practice – the repeated engagement with a limited set of texts – which he argued had largely been replaced by extensive reading by the end of the eighteenth century [13]. Our findings nuance that claim, indicating that intensive reading practices, particularly in religious contexts, remained significant in nineteenth-century Denmark.

By contrast, categories like *Literature and personal writings, History, philosophy and biographies*, and *Geography, topography and biology* display shorter advertising spans, suggesting a

¹⁷ These are unique combinations of author and book title.

Category	Percentage
Agriculture, household and artisanship	69.5%
Almanacs and other light readings	69.3%
Education and Bildung	72.5%
Geography, topography and biology	75.9%
History, philosophy and biographies	82.7%
Law, politics and government	75.9%
Literature and personal writings	76.6%
Religion, theology and psalms	66.4%
Subscription journals	66.7%
Theatre, music and poems	70.4%
Unknown	93.1%

Table 6: Percentage of books per category that have an advertising span of one day.

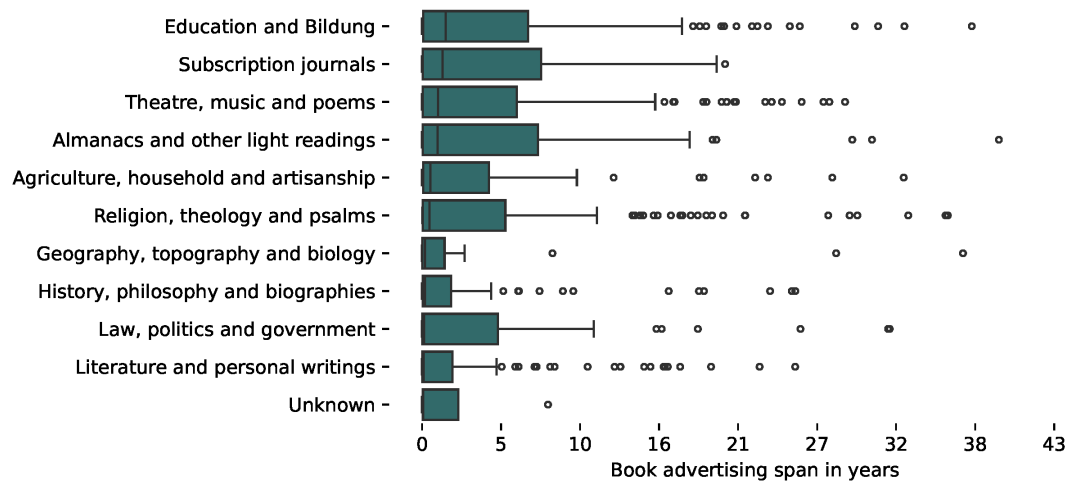


Figure 5: Advertising span of books that appear in advertisements over multiple days, grouped per category.

quicker turnover and greater influx of new titles. These categories appear more closely associated with an *extensive* reading practice, characterized by broad, fast-paced consumption. This pattern likely reflects not only the content and intended use of the books – often topical, narrative-driven, or single-read publications – but also the commercial strategies of booksellers responding to fluctuating demand. While the longer lifespan of religious and educational titles signals stability and long-term relevance, the brevity in these other categories points to a more dynamic and diversified literary market.

5.3 Unpacking the periphery

To assess whether town size is associated with the relative number of book announcements, we fitted a mixed-effects model with the *log* of population size as a fixed effect and year as a random effect.¹⁸ The results show a statistically significant negative relationship between population size and the normalized count of advertised book titles ($\beta = -0.013$, $SE = 0.001$, $z = -9.91$, $p < 0.001$).¹⁹ This means that smaller towns, such as Maribo and Thisted, tended to have proportionally

¹⁸ We used the `statsmodels` package in Python [32].

¹⁹ The number of book titles in a newspaper are normalized against the total number of editions of that newspaper.

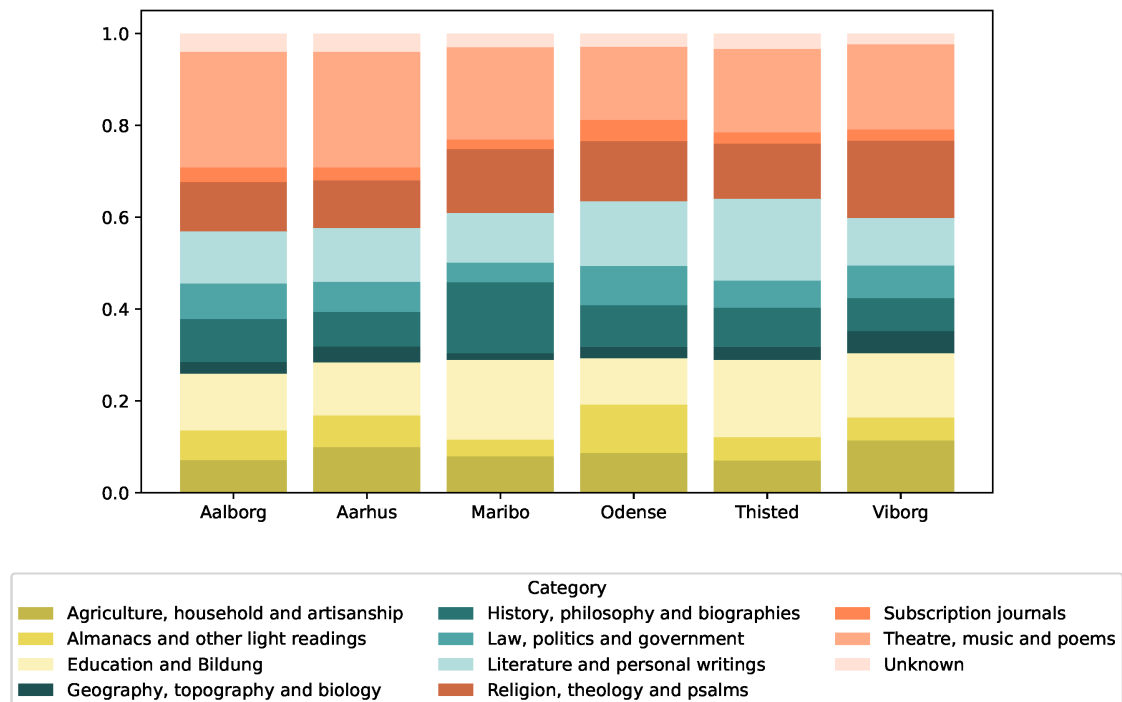


Figure 6: Normalized category distribution per newspaper.

more book advertisements in their newspapers than larger towns.

To compare how the book title offerings differ across newspapers, we visualized the normalized distribution of categories for each newspaper in Figure 6.²⁰ The stacked bar plots show that all categories are represented in each of the newspapers. Most visually dominant across all newspapers is the *Theatre, music and poems* category. This category seems to occupy roughly one-fourth to one-third of the total content of the newspapers, confirming that cultural content had a more central role in local newspapers than previously anticipated. Especially Aalborg and Aarhus – both towns with established theatres – have a higher distribution of this category, likely reflecting a large readership with access to or interest in theatrical and artistic life.

In contrast, towns such as Viborg, Maribo, and Odense show a relatively lower proportion of cultural content in this category – although Odense, the largest town in our dataset, also had a theatre. This may suggest that interest in theatre does not necessarily correlate with population size, but rather reflects the composition of a population – explaining why Aalborg and Aarhus, which are similar in geography and demographics, show comparable patterns. Instead, the newspapers in Viborg, Maribo, and Odense seem to emphasize the ‘knowledge-oriented reading’, with a stronger presence of, for example, the category *Religion, theology and psalms*. When we model the relationship between the shares of the categories *Religion, theology and psalms* and *Theatre, music and poems* in a mixed-effects model with year as random effect, we find that an increase in the *theatre* share is associated with a decrease in *religion* share ($\beta = -0.184$, $SE = 0.058$, $z = -3.19$, $p = 0.001$). This suggests that religious content in book advertisements is competing with the more cultural and artistic genres, often including profane music, satire, and critique of society. As seen in Figure 6, Viborg, in particular, stands out for this inverse relationship. This pattern aligns with the town’s long-standing status as an important religious center in Denmark, exemplified by its cathedral, which dates back to 1065.

²⁰ The number of book titles per category are normalized against the total number of book titles advertised in that newspaper.

Aalborg	0	0.0056	0.046	0.049	0.045	0.047
Aarhus	0.0056	0	0.06	0.05	0.05	0.042
Maribo	0.046	0.06	0	0.075	0.042	0.047
Odense	0.049	0.05	0.075	0	0.043	0.044
Thisted	0.045	0.05	0.042	0.043	0	0.045
Viborg	0.047	0.042	0.047	0.044	0.045	0
	Aalborg	Aarhus	Maribo	Odense	Thisted	Viborg

Figure 7: Heatmap showing the cosine distances between category distributions in newspapers. The lighter color corresponds with a lower cosine distance, in other words, a higher similarity between two distributions. Note that we would expect cosine similarity values to be naturally high when comparing vectors this short (here, 11-category distributions). Rather than measuring overall similarity, the measure helps us focus on relative differences – highlighting newspapers whose category distribution stands apart.

Overall, the newspapers of Aalborg and Aarhus show similar profiles, with balanced distributions across most categories. This is confirmed by the heatmap in Figure 7, which shows the cosine distances between the category distributions of each newspaper – where each newspaper vector expresses the normalized counts per category: Aalborg and Aarhus exhibit the highest cosine similarity. Both towns were harbor cities with comparable population sizes (approximately 6,000-6,500 inhabitants) and were located relatively close to one another. It is therefore likely that their newspapers served similar audiences, which may explain the overlap in the advertised book offerings. The heatmap furthermore shows that the newspaper of Maribo – located on the island Lolland-Falster – has the largest cosine distance to the other newspapers, with its difference from the newspaper of Odense being the greatest.

Another notable observation is that the two smallest towns in the dataset, Maribo (700-1,400 inhabitants) and Thisted (1,100-2,100 inhabitants), exhibit a comparatively higher share of the category *Education and Bildung* than the other newspapers. This pattern is visible in Figure 6, and is statistically supported by a mixed-effects model. When modeling the relationship between a town’s population size and the normalized share of the *Education and Bildung* category (with year as a random effect), we observe a significant negative association with population size ($\beta = -0.026$, $SE = 0.006$, $z = -4.62$, $p < 0.001$). This suggests that smaller towns tend to have a higher proportion of advertisements related to education compared to larger towns. One possible explanation lies in the social and institutional impact of the Danish School Acts of 1814, which made primary education compulsory across Denmark. While the legislation applied nationally, its practical implementation often had a greater visible effect in rural or less urbanized areas, where formal schooling structures were less developed beforehand [26]. In smaller towns, the new educational infrastructure – such as the construction of schools, training of teachers, and dissemination of textbooks – represented a significant cultural shift, generating increased demand for educational materials. The prominence of the *Education and Bildung* category in newspapers from small towns

like Maribo and Thisted may therefore reflect the intensity with which these reforms were being realized in those settings.

6 Discussion and conclusion

The analysis of nineteenth-century book advertisements in local newspapers from the Danish periphery has shed new light on an underexplored aspect of nineteenth-century reading culture. By applying classification models to categorize book advertisements and using a generative language model to extract book titles from them, we constructed a dataset of almost 6,400 book titles, comprising over 4,000 unique titles. With this dataset, we addressed three key questions about reading culture in peripheral Denmark: *what did people read*, *how did people read*, and *where did people read what?* The findings show that the book offers in the advertisements both include texts associated with ‘knowledge-oriented’ reading and ‘leisure-oriented’ reading. This challenges earlier claims that nineteenth-century middle-class readers primarily consumed religious books. Instead, we observe a much more diverse reading landscape – spanning genres, themes, and even languages – than previously assumed.

In terms of reading practices, an analysis of advertising spans of book titles revealed evidence for both the *intensive* and *extensive* reading practice – the former characteristic for books in the categories *Religion, theology and psalms* and *Education and Bildung*, the latter for books in categories such as *Literature and personal writings* and *History, philosophy and biographies*. This adds nuance to Engelsing’s theory that extensive reading had replaced intensive reading by the beginning of the nineteenth century.

The newspaper-level analysis yielded further insights. We identified a statistically reliable negative relationship between the population size of a town and the proportion of book titles in its newspaper, indicating that smaller towns included more book titles in their newspaper advertisements. We also observed that religious content in book advertisements competes with the more cultural and artistic genres, with some newspapers leaning more strongly towards one or the other. Finally, we found that smaller towns tend to advertise more schoolbooks than larger towns, a pattern likely connected to the institutionalization of compulsory education in Denmark in 1814, which had a pronounced impact in rural and less urbanized areas. In summary, the study demonstrates that studying book advertisements in local Danish newspapers is a fruitful way to show that reading culture in the nineteenth-century Danish periphery was more varied and dynamic than earlier accounts have suggested.

Nevertheless, while advertisements reflect commercial strategies aimed at attracting readers, they should not be interpreted as direct measures of reader demand or popularity. The decision to advertise likely varies by genre. For some genres – such as religious literature – a lower advertising frequency may reflect stable, widespread circulation that made further promotion unnecessary. In contrast, genres with high advertising intensity may signal novelty, uneven distribution, or a bookseller’s expectation of future demand, rather than confirmed popularity. As Harris et al. [14] note, advertising often operates in tension with what is already assumed to be known by the audience, rather than simply mirroring public interest [4]. Thus, continued advertising can suggest either actual sales success or an effort to create visibility for unfamiliar or aspirational content.²¹

Another limitation of the corpus and methods used in this study deserves mention. First, the segmentation of newspapers into individual articles is not always accurate, which may have affected the identification of book announcements – some of which were incorrectly merged with unrelated content. We experimented with training a classification model to detect such cases, but

²¹ Since booksellers were often indebted to bigger Copenhagen-based booksellers, it seems unlikely that repeated advertisements indicate a book’s lack of popularity. Rather than signaling unsold stock, re-advertising likely reflects ongoing expectations of demand.

these proved difficult to reliably identify. As a result, we limited the analysis to correctly segmented announcements. This likely means that the number of detected book advertisements underrepresents the actual total, though we assume this bias is consistent across newspapers and does not substantially affect comparative findings.

Interestingly, since the start of this study, significant improvements have been made in article segmentation methods for these historical newspapers. This gives us reason to be optimistic that future iterations of this study will benefit from more accurate segmentation, thereby enhancing both the completeness and reliability of the dataset. Additionally, more local newspapers are currently being digitized – both to extend the chronological range of the corpus and to include a broader selection of titles. These developments will help make future studies based on this corpus more representative of the diversity and evolution of reading culture in the Danish periphery.

Our study opens up new possibilities for further research into nineteenth-century Danish reading culture. Two findings, in particular, warrant deeper investigation: the role of German-language books in Danish reading culture and the impact of the Danish School Acts on the availability and circulation of books in the periphery. At the same time, our approach speaks to broader questions of historical information dynamics in Denmark. The corpus and methods developed here could be applied to trace the dissemination of both national and international news, and to explore how local and regional reading publics were interconnected through the press. Extending the dataset to include Copenhagen-based newspapers would further enhance comparisons between center and periphery, allowing for a more comprehensive understanding of how information flowed within a historical information society. Beyond the Danish case, the pipeline presented in this study offers a scalable, adaptable framework for studying information circulation in other Western European contexts with similar media infrastructures – contributing not only to the history of reading, but to the field of information history as a whole.

Acknowledgements

The authors affiliated with Aarhus University were supported by grants from the Carlsberg Foundation (*The Golden Array of Danish Cultural Heritage*) and the Aarhus Universitets Forskningsfond (*Golden Imprints of Danish Cultural Heritage*). We would like to thank all participants of the ENO-team, as well as Kit Morgenstjerne and Ingeborg Reinholdt Thorhauge for their annotation work. Part of the computation done for this project was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark.

References

- [1] Anderson, Benedict R. O’G. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised edition. London New York: Verso, 2016.
- [2] Appel, Charlotte. *Læsning og bogmarked i 1600-tallets Danmark*. Danish humanist texts and studies 23. Copenhagen: Museum Tusculanum, 2001.
- [3] Appel, Charlotte and Christensen, Nina. *På sporet af børn og bøger: læsekultur og medier 1750-1850*. 1. udgave. Copenhagen: Gad, 2023.
- [4] Baunvig, Katrine Frøkjær. “Fictional Realities of Modernity: The Fantastic Life of Demi-Goddess Dana in the Emerging Nation State of Denmark”. In: *Mythology and Nation Building*, ed. by Lone Kølle Martinsen, Sophie Bønding, and Pierre-Brice Stahl. Aarhus: Aarhus Universitetsforlag, June 2021, pp. 97–134.

- [5] Bjerring-Hansen, Jens, Al-Laith, Ali, Hershcovich, Daniel, Conroy, Alexander, and Rasmussen, Sebastian Ørtoft. "Literary Time Travel: Distinguishing Past and Contemporary Worlds in Danish and Norwegian Fiction". In: *Proceedings of the Computational Humanities Research Conference*, ed. by Wouter Haverals, Marijn Koolen, and Laure Thompson. Vol. 3834. CEUR Workshop Proceedings. 2024, pp. 772–787.
- [6] Bjerring-Hansen, Jens and Rasmussen, Sebastian Ørtoft. "Litteratursociologi og kvantitative litteraturstudier: Den historiske roman i det moderne gennembrud som case". In: *Passage - Tidsskrift for litteratur og kritik* 38, no. 89 (June 2023), pp. 171–189. DOI: 10.7146/pas.v38i89.137987.
- [7] Bjerring-Hansen, Jens and Wilkens, Matthew. "Deep Distant Reading: The Rise of Realism in Scandinavian Literature as a Case Study". In: *Orbis Litterarum* 78, no. 5 (2023), pp. 335–352. DOI: 10.1111/oli.12396.
- [8] Borst, Janos, Klähn, Jannis, and Burghardt, Manuel. "Death of the Dictionary? – The Rise of Zero-Shot Sentiment Classification". In: *Proceedings of the Computational Humanities Research Conference*, ed. by Artjoms Šeļa, Fotis Jannidis, and Iza Romanowska. Vol. 3558. CEUR Workshop Proceedings. 2023, pp. 303–319.
- [9] Brodersen, Lars, Staunstrup, Jan K., and Knak-Nielsen, Steffen. "Den rådne banan på kort: Visuel literacy - fra billeder til viden". In: *Dansk Biblioteksforskning* 7, no. 2/3 (2011), pp. 69–82. DOI: 10.7146/danbibfor.v7i2/3.97386.
- [10] Christensen, Søren Bitsch and Mikkelsen, Jørgen. "The Danish Urban System Pre-1800: A Survey of Recent Research Results". In: *Urban History* 33, no. 3 (Dec. 2006), pp. 484–510. DOI: 10.1017/S0963926806004081.
- [11] Dalfsen, Arjan van, Karsdorp, Folgert, Bagheri, Ayoub, Mentink, Dieuwertje, Engelen, Thirza van, and Stronks, Els. "Direct and Indirect Annotation with Generative AI: A Case Study into Finding Animals and Plants in Historical Text". In: *Proceedings of the Computational Humanities Research Conference*, ed. by Wouter Haverals, Marijn Koolen, and Laure Thompson. Vol. 3834. CEUR Workshop Proceedings. 2024, pp. 1053–1074.
- [12] Darnton, Robert. "What Is the History of Books?" In: *Daedalus* 111, no. 3 (1982), pp. 65–83.
- [13] Engelsing, Rolf. "Die Perioden der Lesergeschichte in der Neuzeit". In: *Zur Sozialgeschichte deutscher Mittel- und Unterschichten*. Vol. 4. Göttingen: Vandenhoeck & Ruprecht, 1978. DOI: 10.13109/9783666359750.112.
- [14] Harris, Richard J., Dubitsky, Tony M., Perch, Karen L., Ellerman, Cindy S., and Larson, Mark W. "Remembering Implied Advertising Claims as Facts: Extensions to the "Real World"". In: *Bulletin of the Psychonomic Society* 16, no. 4 (Oct. 1980), pp. 317–320. DOI: 10.3758/BF03329554.
- [15] Heinsen, Johan and Bøgeskov, Camilla. "A World in Print: Introducing a Danish-Norwegian Corpus of Historical Newspapers". Sept. 2025. DOI: 10.48550/arXiv.2509.02356. arXiv: 2509.02356 [cs].
- [16] Hertel, Hans. *Den daglige bog : bøger, formidlere og læsere i Danmark gennem 500 år*. Copenhagen: Forening for Boghaandværk, 1983.
- [17] Høegh-Guldberg, Ove, Jensen, Nanna Marie, and Danmarkshistorien. "Censur i Danmark fra 1750 til 1849 – danmarkshistorien: Lex". URL: https://danmarkshistorien.lex.dk/Censur_i_Danmark_fra_1750_til_1849.

- [18] Horstbøll, Henrik. *Menigmands medie: det folkelige bogtryk i Danmark 1500-1840 : en kulturhistorisk undersøgelse*. Danish humanist texts and studies, vol. 19. Copenhagen: Det Kongelige Bibliotek, 1999.
- [19] Jensen, F. Elle. “Vestjydsk Bondelæsning i Stavnsbaandstiden”. In: *Historie/Jyske Samlinger* (1943).
- [20] Johansen, Hans Chr. “Erhverv, arbejdskraft og boligformer på landet i begyndelsen af 1800-tallet”. dan. In: *Landbohistorisk Tidsskrift* 16 (2021), pp. 89–152.
- [21] Kahle, P., Colutto, S., Hackl, G., and Mühlberger, G. “Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents”. In: *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 04. 2017, pp. 19–24. DOI: 10.1109/ICDAR.2017.307.
- [22] Karjus, Andres. “Machine-Assisted Quantitizing Designs: Augmenting Humanities and Social Sciences with Artificial Intelligence”. Oct. 2024. DOI: 10.48550/arXiv.2309.14379. arXiv: 2309.14379 [cs].
- [23] Kettunen, Kimmo, Mäkelä, Eetu, Ruokolainen, Teemu, Kuokkala, Juha, and Löfberg, Laura. “Old Content and Modern Tools - Searching Named Entities in a Finnish OCRed Historical Newspaper Collection 1771-1910”. In: *CoRR abs/1611.02839* (2016). arXiv: 1611.02839.
- [24] Kettunen, Kimmo, Ruokolainen, Teemu, Liukkonen, Erno, Tranouez, Pierrick, Antelme, Daniel, and Paquet, Thierry. “Detecting Articles in a Digitized Finnish Historical Newspaper Collection 1771-1929: Early Results Using the PIVAJ Software”. In: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. DATECH2019. New York, NY, USA: Association for Computing Machinery, May 2019, pp. 59–64. DOI: 10.1145/3322905.3322911.
- [25] Kloek, J.J. and Mijndhardt, W.W. “Lezersrevolutie of literaire socialisatie?” In: *Nieuwe wegen in taal- en literatuurwetenschap: handelingen van het eenenveertigste Filologencongres*, ed. by J. Goedegebuure. Tilburg: University Press, 1993, pp. 211–224.
- [26] Larsen, Christian. *Fra skoleanordninger til Den Store Skolekommission: Studier af forandringer i dansk uddannelsestænkning 1814-1923 med særligt henblik på decentral uddannelse, privat uddannelse og professionsuddannelse*. Dansk. PhD thesis. 2008.
- [27] Nielsen, Sofie Krogh. “Daily Life in Denmark in the 19th Century”. In: *The Bridge* 37, no. 1 (2014).
- [28] Östling, Johan, Heidenblad, David Larsson, and Hammar, Anna Nilsson. *Forms of Knowledge: Developing the History of Knowledge*. Nordic Academic Press, 2020. URL: <http://www.jstor.org/stable/jj.919496>.
- [29] Pedregosa, Fabian et al. “Scikit-Learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [30] Pettersson, Eva, Borin, Lars, and Lenas, Erik. “SWENER-1800: A Corpus for Named Entity Recognition in 19th Century Swedish”. In: *Digital Humanities in the Nordic and Baltic Countries Publications* 6, no. 1 (Sept. 2024). DOI: 10.5617/dhnbpub.11489.
- [31] Samuelsson, Lina, Zechner, Niklas, Brodén, Daniel, Ingvarsson, Jonas, and Karimi, Aram. “Between the Arduous and the Automatic: A Comparative Approach to the Challenge of Classifying of Book Reviews in Swedish Newspapers”. In: *Digital Humanities in the Nordic and Baltic Countries Publications* 7, no. 3 (Mar. 2025). DOI: 10.5617/dhnbpub.12267.
- [32] Seabold, Skipper and Perktold, Josef. “statsmodels: Econometric and statistical modeling with python”. In: *9th Python in Science Conference*. 2010.

- [33] Søllinge, Jette Drachmann, Thomsen, Niels, and Danske, Den Store. “Avis”. URL: <https://lex.dk/avis>.
- [34] St. Clair, William. *The Reading Nation in the Romantic Period*. 1st pbk. ed. Cambridge ; New York: Cambridge University Press, 2007.
- [35] Stegemeier, Henri. “Frauentaschenbuch Für Das Jahr 1816”. In: *The Journal of English and Germanic Philology* 36, no. 3 (1937), pp. 387–390. JSTOR: 27704286.
- [36] Tudor, Crina, Megyesi, Beata, and Östling, Robert. “Prompting the Past: Exploring Zero-Shot Learning for Named Entity Recognition in Historical Texts Using Prompt-Answering LLMs”. In: *Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2025)*, ed. by Anna Kazantseva, Stan Szpakowicz, Stefania Degaetano-Ortlieb, Yuri Bizzoni, and Janis Pagel. Albuquerque, New Mexico: Association for Computational Linguistics, May 2025, pp. 216–226. DOI: 10.18653/v1/2025.latechclfl-1.19.
- [37] Winther, Malene Brandt and Svendsen, Gunnar Lind Haase. “‘The Rotten Banana’ fires back: The story of a Danish discourse of inclusive rurality in the making”. In: *Journal of Rural Studies* 28, no. 4 (2012). Growing Old in Rural Places, pp. 466–477. DOI: 10.1016/j.jrurstud.2012.05.003.
- [38] Yu, Yaoning, Yu, Ye, Wei, Kai, Luo, Haojing, and Wang, Haohan. “SIPDO: Closed-Loop Prompt Optimization via Synthetic Data Feedback”. June 2025. DOI: 10.48550/arXiv.2505.19514. arXiv: 2505.19514 [cs].

A Prompts

P1. This is an announcement in a Danish nineteenth century newspaper. Extract the book titles and authors that are mentioned. Return results in this format:

```
original_title: <title in Danish>
translated_title: <title in English>
author: <author_name>
```

If the author is missing or unclear, fill in ‘NO_AUTHOR’. For the translated_title, you can translate yourself. Be careful and look closely, sometimes the author name is a bit hidden. It can be that it is mentioned like this: ‘AUTHOR’s TITLE is now available.’ Look closely where the title begins and ends. When the announcement does not contain book titles, return exactly one row with:

```
original_title: NO_BOOK
translated_title: NO_BOOK
author: NO_BOOK
```

P2. Here is an announcement in a Danish nineteenth century newspaper. If book titles and authors are mentioned, extract those in this format:

```
original_title: <title in Danish>
translated_title: <title in English>
author: <author name>
```

If the author is missing or unclear, fill in ‘NO_AUTHOR’. The author name can appear in many ways, for example: ‘AUTHOR’s TITLE is now available.’ Look closely where the title begins and ends. For the translated title, translate it yourself in English. When the announcement does not contain book titles, return exactly one row with:

```
original_title: NO_BOOK
translated_title: NO_BOOK
author: NO_BOOK
```

Here are first a few examples of announcements containing book titles:

Example1: ”Baggesens allerældste Poesier.”

```
‘original_title’: allerældste Poesier; ‘author’: Baggesen
```

Example2: ”Kateketisk Magasin af J. C. Wegener, Forstander for det Kongelige Skolelærer-Seminarium paa Joenstrup.”

```
‘original_title’: Kateketisk Magasin; ‘author’: J.C. Wegener
```

Example3: ”Ceres. Et periodisk Skrivt for dannede Læsere. Udgiver af F. M. Lange. Femte Hefte. Det indeholder:

Juliette, eller det hemmelige Ægteskab, af Frederik Kind. - Jagtgildet, af Washington Irving. Subskription modtages hos Vogelius, Boghandler og Bogbinder."

'original_title': Juliette, eller det hemmelige Ægteskab; 'author': Frederik Kind

'original_title': Jagtgildet; 'author': Washington Irving

Here are a few examples of announcements that do NOT contain book titles:

Example1: "J. Et Parti gode hjemmegjorte Bolster og Dynevaar er i Dag arriveret og sælges billigst muligt af M. N. Samson."

Example2: "C. Andersen. Første Afdeling: "Spanierne i Odense, Vaudeville i 1 Act. Anden Afdeling: "Fem og tyve Aar derefter i Helsingør, Vaudeville i 1 Act. Billetter a 2 Mk. 8 s., (Børn det Halve) erholdes i mit Logie hos Hr. Kobbensmed Schmidt. Hvo som tager 6 Billetter erholder disse for 2 A. Werligh. Rbd." This is a theatre announcement.

Example3: "Første Binds andet Hefte, indeholdende følgende Katekisationer: 1 Den ægte kristelige Menneskekjærlighed bør være ufortrøden, virksom, uegennyttig og viis 2 Om de Glæder, den sande Menneskekjærlighed skænker os 5 Om Guds Almagt; 4 Om Guds Alvidenhed; 5 Om Guds Viisdom; 6 Til Lærebogens 6 Kap. 1. 2. 5, 7 Religion er Menneskets vigtigste Anliggende." These are chapter titles.

P3. Here is an announcement in a Danish nineteenth-century newspaper. If book titles and authors are mentioned, extract those in this format:

original_title: <title in Danish>

translated_title: <title in English>

author: <author name>

If the author is missing or unclear, fill in 'NO_AUTHOR'. The author name can appear in many ways, e.g., 'AUTHOR's TITLE is now available.' Look closely where the title begins and ends. For the translated title, translate it yourself into English. If the announcement does not contain any book titles, return exactly one row with:

original_title: NO_BOOK

translated_title: NO_BOOK

author: NO_BOOK

Examples with books:

Example1: 'Baggesens allerældste Poesier'. →

'original_title': allerældste Poesier; 'author': Baggesen

Example2: 'Kateketisk Magasin af J. C. Wegener, Forstander for det Kongelige Skolelærer-Seminarium paa Joenstrup.'

→

'original_title': Kateketisk Magasin; 'author': J.C. Wegener

Example3: 'Ceres. Et periodisk Skrivt for dannede Læsere. Udgiver af F. M. Lange. Femte Hefte. Det indeholder: Juliette, eller det hemmelige Ægteskab, af Frederik Kind. - Jagtgildet, af Washington Irving. Subskription modtages hos Vogelius, Boghandler og Bogbinder.' →

'original_title': Juliette, eller det hemmelige Ægteskab; 'translated_title': Juliette, or The Secret Marriage; 'author': Frederik Kind

'original_title': Jagtgildet; 'translated_title': The Hunting Feast; 'author': Washington Irving

Examples without books:

Example1: 'J. Et Parti gode hjemmegjorte Bolster og Dynevaar er i Dag arriveret og sælges billigst muligt af M. N. Samson.'

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example2: 'C. Andersen. Første Afdeling: "Spanierne i Odense, Vaudeville i 1 Act. Anden Afdeling: "Fem og tyve Aar derefter i Helsingør, Vaudeville i 1 Act. Billetter a 2 Mk. 8 s., (Børn det Halve) erholdes i mit Logie hos Hr. Kobbensmed Schmidt. Hvo som tager 6 Billetter erholder disse for 2 A. Werligh. Rbd.' This is a theater announcement.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example3: 'Første Binds andet Hefte, indeholdende følgende Katekisationer: 1 Den ægte kristelige Menneskekjærlighed bør være ufortrøden, virksom, uegennyttig og viis 2 Om de Glæder, den sande Menneskekjærlighed skænker os 5 Om Guds Almagt; 4 Om Guds Alvidenhed; 5 Om Guds Viisdom; 6 Til Lærebogens 6 Kap. 1. 2. 5, 7 Religion er Menneskets vigtigste Anliggende.' These are chapter titles.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

P4. Here is an announcement in a Danish nineteenth-century newspaper. Your task is to extract book titles and authors using the following format:

original_title: <title in Danish>

translated_title: <title in English>

author: <author name>

Guidelines:

1. Carefully identify the beginning and end of each book title. Pay attention to punctuation and phrasing that may indicate a book title.

2. If the announcement mentions multiple book titles, extract each one separately.

3. If the author is missing or unclear, use 'NO_AUTHOR'.
4. Translate the original Danish title into English yourself for the 'translated_title'.
5. If no book titles are present, return exactly one row with:

original_title: NO_BOOK
translated_title: NO_BOOK
author: NO_BOOK

Examples with books:

Example1: 'Baggesens allerældste Poesier'.

→ original_title: allerældste Poesier; translated_title: Oldest Poems; author: Baggesen

Example2: 'Kateketisk Magasin af J. C. Wegener, Forstander for det Kongelige Skolelærer-Seminarium paa Joenstrup.'

→ original_title: Kateketisk Magasin; translated_title: Catechetical Magazine; author: J.C. Wegener

Example3: 'Ceres. Et periodisk Skrivt for dannede Læsere. Udgiver af F. M. Lange. Femte Hefte. Det indeholder: Juliette, eller det hemmelige Ægteskab, af Frederik Kind. - Jagtgildet, af Washington Irving. Subskription modtages hos Vogelius, Boghandler og Bogbinder.'

→ original_title: Juliette, eller det hemmelige Ægteskab; translated_title: Juliette, or The Secret Marriage; author:

Frederik Kind

→ original_title: Jagtgildet; translated_title: The Hunting Feast; author: Washington Irving

Examples without books:

Example1: 'J. Et Parti gode hjemmegjorte Bolster og Dynevaar er i Dag arriveret og sælges billigst muligt af M. N. Samson.'

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example2: 'C. Andersen. Første Afdeling: 'Spanierne i Odense, Vaudeville i 1 Act. Anden Afdeling: 'Fem og tyve Aar derefter i Helsingør, Vaudeville i 1 Act. Billetter a 2 Mk. 8 s., (Børn det Halve) erholdes i mit Logie hos Hr. Kobbersmed Schmidt. Hvo som tager 6 Billetter erholder disse for 2 A. Werligh. Rbd.' This is a theater announcement.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example3: 'Første Binds andet Hefte, indeholdende følgende Katekisationer: 1 Den ægte kristelige Menneskekjærlighed bør være ufortrøden, virksom, uegennyttig og viis 2 Om de Glæder, den sande Menneskekjærlighed skjænker os 5 Om Guds Almagt; 4 Om Guds Alvidenhed; 5 Om Guds Viisdom; 6 Til Lærebogens 6 Kap. 1. 2. 5, 7 Religion er Menneskets vigtigste Anliggende.' These are chapter titles.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

P5. Here is an announcement in a Danish nineteenth-century newspaper. Your task is to extract book titles and authors using the following format:

original_title: <title in Danish>
translated_title: <title in English>
author: <author name>

Guidelines:

1. Carefully identify the beginning and end of each book title. Look for capitalization, italics, or quotation marks that may indicate a book title.
2. If the announcement mentions multiple book titles, extract each one separately.
3. If the author is missing or unclear, use 'NO_AUTHOR'.
4. Translate the original Danish title into English yourself for the 'translated_title'.
5. Pay special attention to context - announcements may contain other text (e.g., product listings, theater plays) that should not be considered book titles.
6. If no book titles are present, return exactly one row with:

original_title: NO_BOOK
translated_title: NO_BOOK
author: NO_BOOK

Examples with books:

Example1: 'Baggesens allerældste Poesier'.

→ original_title: allerældste Poesier; translated_title: Oldest Poems; author: Baggesen

Example2: 'Kateketisk Magasin af J. C. Wegener, Forstander for det Kongelige Skolelærer-Seminarium paa Joenstrup.'

→ original_title: Kateketisk Magasin; translated_title: Catechetical Magazine; author: J.C. Wegener

Example3: 'Ceres. Et periodisk Skrivt for dannede Læsere. Udgiver af F. M. Lange. Femte Hefte. Det indeholder: Juliette, eller det hemmelige Ægteskab, af Frederik Kind. - Jagtgildet, af Washington Irving. Subskription modtages hos Vogelius, Boghandler og Bogbinder.'

→ original_title: Juliette, eller det hemmelige Ægteskab; translated_title: Juliette, or The Secret Marriage; author:

Frederik Kind

→ original_title: Jagtgildet; translated_title: The Hunting Feast; author: Washington Irving

Examples without books:

Example1: 'J. Et Parti gode hjemmegjorte Bolster og Dynevaar er i Dag arriveret og sælges billigst muligt af M. N. Samson.'

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example2: 'C. Andersen. Første Afdeling: 'Spanierne i Odense, Vaudeville i 1 Act. Anden Afdeling: 'Fem og tyve Aar derefter i Helsingør, Vaudeville i 1 Act. Billetter a 2 Mk. 8 s., (Børn det Halve) erholdes i mit Logie hos Hr. Kobbersmed Schmidt. Hvo som tager 6 Billetter erholder disse for 2 A. Werligh. Rbd.' This is a theater announcement.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example3: 'Første Binds andet Hefte, indeholdende følgende Katekisationer: 1 Den ægte kristelige Menneskekjærlighed bør være ufortrøden, virksom, uegennyttig og viis 2 Om de Glæder, den sande Menneskekjærlighed skjænker os 5 Om Guds Almagt; 4 Om Guds Alvidenhed; 5 Om Guds Viisdom; 6 Til Lærebogens 6 Kap. 1. 2. 5, 7 Religion er Menneskets vigtigste Anliggende.' These are chapter titles.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Use these guidelines and examples to enhance extraction accuracy and maintain the required output format.

P6. Here is an announcement in a Danish nineteenth-century newspaper. Your task is to extract book titles and authors using the following format:

original_title: <title in Danish>
translated_title: <title in English>
author: <author name>

Guidelines:

1. Carefully identify the beginning and end of each book title. Pay attention to capitalization, italics, quotation marks, or context that may indicate a book title. Titles are often followed by a description or an author's name.
2. If the announcement mentions multiple book titles, extract each one separately. Ensure each title is uniquely identified.
3. If the author is missing or unclear, use 'NO_AUTHOR'. Verify the context to ensure the correct identification of authors.
4. Translate the original Danish title into English yourself for the 'translated_title'. Ensure the translation preserves the meaning and context of the original title.
5. Pay special attention to context - announcements may contain other text (e.g., product listings, theater plays, chapter titles) that should not be considered book titles. Identify keywords that separate book titles from other content.
6. If no book titles are present, return exactly one row with:

original_title: NO_BOOK
translated_title: NO_BOOK
author: NO_BOOK

Examples with books:

Example1: 'Baggesens allerældste Poesier'.

→ original_title: allerældste Poesier; translated_title: Oldest Poems; author: Baggesen

Example2: 'Kateketisk Magasin af J. C. Wegener, Forstander for det Kongelige Skolelærer-Seminarium paa Joenstrup.'

→ original_title: Kateketisk Magasin; translated_title: Catechetical Magazine; author: J.C. Wegener

Example3: 'Ceres. Et periodisk Skrivt for dannede Læsere. Udgiver af F. M. Lange. Femte Hefte. Det indeholder: Juliette, eller det hemmelige Ægteskab, af Frederik Kind. - Jagtgildet, af Washington Irving. Subskription modtages hos Vogelius, Boghandler og Bogbinder.'

→ original_title: Juliette, eller det hemmelige Ægteskab; translated_title: Juliette, or The Secret Marriage; author: Frederik Kind

→ original_title: Jagtgildet; translated_title: The Hunting Feast; author: Washington Irving

Examples without books:

Example1: 'J. Et Parti gode hjemmegjorte Bolster og Dynevaar er i Dag arriveret og sælges billigst muligt af M. N. Samson.'

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example2: 'C. Andersen. Første Afdeling: 'Spanierne i Odense, Vaudeville i 1 Act. Anden Afdeling: 'Fem og tyve Aar derefter i Helsingør, Vaudeville i 1 Act. Billetter a 2 Mk. 8 s., (Børn det Halve) erholdes i mit Logie hos Hr. Kobbersmed Schmidt. Hvo som tager 6 Billetter erholder disse for 2 A. Werligh. Rbd.' This is a theater announcement.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example3: 'Første Binds andet Hefte, indeholdende følgende Katekisationer: 1 Den ægte kristelige Menneskekjærlighed bør være ufortrøden, virksom, uegennyttig og viis 2 Om de Glæder, den sande Menneskekjærlighed skjænker os 5 Om Guds Almagt; 4 Om Guds Alvidenhed; 5 Om Guds Viisdom; 6 Til Lærebogens 6 Kap. 1. 2. 5, 7 Religion er Menneskets vigtigste Anliggende.' These are chapter titles.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Use these guidelines and examples to enhance extraction accuracy and maintain the required output format.

P7. Here is an announcement in a Danish nineteenth-century newspaper. Your task is to extract book titles and authors using the following format:

original_title: <title in Danish>
translated_title: <title in English>
author: <author name>

Guidelines:

1. Carefully identify the beginning and end of each book title. Pay attention to capitalization, italics, quotation marks, or context that may indicate a book title. Titles are often followed by a description or an author's name.
2. If the announcement mentions multiple book titles, extract each one separately. Ensure each title is uniquely identified.
3. If the author is missing or unclear, use 'NO_AUTHOR'. Verify the context to ensure the correct identification of authors.
4. Translate the original Danish title into English yourself for the 'translated_title'. Ensure the translation preserves the meaning and context of the original title.
5. Pay special attention to context - announcements may contain other text (e.g., product listings, theater plays, chapter titles) that should not be considered book titles. Identify keywords that separate book titles from other content.
6. If no book titles are present, return exactly one row with:

original_title: NO_BOOK
translated_title: NO_BOOK
author: NO_BOOK

7. Avoid false positives by distinguishing between book titles and other textual elements like speeches, essays, or articles, which may not be standalone books.
8. If a title seems overly descriptive or contains elements like 'a speech', 'an essay', or similar, reassess if it fits the criteria for a book title.

Examples with books:

Example1: 'Baggesens allerældste Poesier'.

→ original_title: allerældste Poesier; translated_title: Oldest Poems; author: Baggesen

Example2: 'Kateketisk Magasin af J. C. Wegener, Forstander for det Kongelige Skolelærer-Seminarium paa Joenstrup.'

→ original_title: Kateketisk Magasin; translated_title: Catechetical Magazine; author: J.C. Wegener

Example3: 'Ceres. Et periodisk Skrivt for dannede Læsere. Udgiver af F. M. Lange. Femte Hefte. Det indeholder: Juliette, eller det hemmelige Ægteskab, af Frederik Kind. - Jagtgildet, af Washington Irving. Subskription modtages hos Vogelius, Boghandler og Bogbinder.'

→ original_title: Juliette, eller det hemmelige Ægteskab; translated_title: Juliette, or The Secret Marriage; author: Frederik Kind

→ original_title: Jagtgildet; translated_title: The Hunting Feast; author: Washington Irving

Examples without books:

Example1: 'J. Et Parti gode hjemmegjorte Bolster og Dynevaar er i Dag arriveret og sælges billigst muligt af M. N. Samson.'

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example2: 'C. Andersen. Første Afdeling: 'Spanierne i Odense, Vaudeville i 1 Act. Anden Afdeling: 'Fem og tyve Aar derefter i Helsingør, Vaudeville i 1 Act. Billetter a 2 Mk. 8 s., (Børn det Halve) erholdes i mit Logie hos Hr. Kobbersmed Schmidt. Hvo som tager 6 Billetter erholder disse for 2 A. Werligh. Rbd.' This is a theater announcement.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Example3: 'Første Binds andet Hefte, indeholdende følgende Katekisationer: 1 Den ægte kristelige Menneskekjærlighed bør være ufortrøden, virksom, uegennyttig og viis 2 Om de Glæder, den sande Menneskekjærlighed skjenker os 5 Om Guds Almagt; 4 Om Guds Alvidenhed; 5 Om Guds Viisdom; 6 Til Lærebogens 6 Kap. 1. 2. 5, 7 Religion er Menneskets vigtigste Anliggende.' These are chapter titles.

→ original_title: NO_BOOK; translated_title: NO_BOOK; author: NO_BOOK

Use these guidelines and examples to enhance extraction accuracy and maintain the required output format.

B Prompt performances

Prompt	Match	Threshold	Precision	Recall	F1-score
P1	original title and author	65	0.79	0.82	0.81
		70	0.76	0.80	0.78
		75	0.73	0.77	0.75
		80	0.69	0.73	0.71
	original title	65	0.77	0.82	0.79
		70	0.75	0.80	0.78
		75	0.75	0.80	0.77
		80	0.74	0.78	0.76
P2	original title and author	65	0.81	0.80	0.80
		70	0.79	0.78	0.79
		75	0.77	0.77	0.77
		80	0.75	0.75	0.75
	original title	65	0.85	0.84	0.84
		70	0.83	0.83	0.83
		75	0.83	0.82	0.82
		80	0.82	0.82	0.82
P3	original title and author	65	0.80	0.79	0.79
		70	0.77	0.77	0.77
		75	0.76	0.76	0.76
		80	0.74	0.74	0.74
	original title	65	0.84	0.84	0.84
		70	0.82	0.82	0.82
		75	0.81	0.81	0.81
		80	0.81	0.81	0.81
P4	original title and author	65	0.76	0.54	0.63
		70	0.72	0.52	0.61
		75	0.70	0.51	0.59
		80	0.69	0.50	0.58
	original title	65	0.77	0.56	0.65
		70	0.76	0.55	0.64
		75	0.75	0.55	0.63
		80	0.74	0.54	0.63
P5	original title and author	65	0.82	0.65	0.72
		70	0.79	0.63	0.70
		75	0.76	0.61	0.68
		80	0.69	0.56	0.62
	original title	65	0.81	0.66	0.73
		70	0.80	0.64	0.71
		75	0.79	0.64	0.70
		80	0.78	0.63	0.70
P6	original title and author	65	0.83	0.85	0.84
		70	0.81	0.84	0.82
		75	0.79	0.82	0.81
		80	0.68	0.70	0.69
	original title	65	0.84	0.87	0.86
		70	0.83	0.86	0.84
		75	0.82	0.85	0.83
		80	0.81	0.84	0.82
P7	original title and author	65	0.82	0.84	0.83
		70	0.79	0.82	0.80
		75	0.77	0.80	0.79
		80	0.68	0.71	0.70
	original title	65	0.82	0.86	0.84
		70	0.81	0.84	0.82
		75	0.80	0.83	0.81
		80	0.79	0.82	0.81

Table 7: Performances of book title extraction with different prompts. We look for a full match on `article_id`, and a fuzzy match with different thresholds on the author and/or `original_title`. P1, P2, and P3 are manually created, P4, P5, P6, and P7 are created by the OpenAI API in a self-evaluation loop with P3 as start input.

C Top 20 authors and books

Author	Count
A. F. Just	144
Adam Oehlenschläger	81
J. L. Heiberg	81
A. F. Elmquist	75
A. von Kotzebue	67
F. Munthe	62
N. Blicher	61
P. O. Boisen	48
N. F. S. Grundtvig	47
Walter Scott	47
St. St. Blicher	46
J. N. Høst	46
C. Jacobsen	46
Ludvig Holberg	42
P. D. Faber	41
C. F. Schulz	31
Jacob Henric Schou	31
H. C. Andersen	30
J. Kragh Høst	30
Morten Luther	30

Table 8: Most frequently mentioned authors in book advertisements.

Author	Title	Category	Count
A. F. Just	Haandbog for den læsende Ungdom	Education and Bildung	55
C. Jacobsen	Nye Kogebog	Agriculture, household and artisanship	46
F. Munthe	De vigtigste indenlandske Tildragelser og de m...	History, philosophy and biographies	34
Schou	Udtog af Forordninger	Law, politics and government	27
A. F. Just	Naturhistorie	Geography, topography and biology	25
A. F. Elmquist	Læsefrugter samlede paa Litteraturens Mark	Subscription journals	25
P. O. Boisen	Bibelske Fortællinger med Anvendelse paa Relig...	Religion, theology and psalms	21
F. Munthe	De mærkeligste Personers Levnetsbeskrivelse og...	History, philosophy and biographies	18
C. F. Schulz	Den lystige Løllænder, eller nyeste Samling af...	Almanacs and other light readings	17
H. J. Birch	Bibelhistorie	Religion, theology and psalms	17
N. Blicher	Forsøg til en kristelig Bømelærdom	Religion, theology and psalms	16
Thomas Kingo	Psalmebøger	Religion, theology and psalms	15
A. F. Elmquist	Blomstersproget, eller Blomsternes Betydning p...	Agriculture, household and artisanship	15
Th. Jensen and K. Lindholm	Veiledning til Hovedregning eller mental Regne...	Education and Bildung	15
N. Blicher	Vidoms og Dyds Tabel	Religion, theology and psalms	14
C. F. Schulz	Videnskabelig Fortegnelse over Provindsialbogs...	History, philosophy and biographies	14
NO_AUTHOR	Taschenbuch der Liebe und Freundschaft gewidmet	Almanacs and other light readings	14
Laurits Fogtman	Kongelige Reskripter, Resolutioner og Collegia...	Law, politics and government	14
N. Blicher	Lov og Rets Tabel for Danske Bønder	Law, politics and government	12
NO_AUTHOR	Taschenbuch zum geselligen Vergnügen	Almanacs and other light readings	12

Table 9: Most frequently mentioned books in book advertisements.