Story Settings: A Dataset

Kaley Rittichier
University of Connecticut
kaley.rittichier@uconn.edu

Abstract

Understanding the settings of a given story has long been viewed as an essential component of understanding the story at large. This significance is not only underscored in academic literary analysis but also in kindergarten education. However, despite this significance, it has received relatively little attention regarding computational analyses of stories. This paper presents a dataset of 2,302 time period setting labeled works and 6,991 location setting labeled works. This dataset aims to help with Cultural Analytics of literary works but may also aid in time-period-related questions within literary Q&A systems.

1 Introduction

The setting of a story is the time and place in which the events in the story are purported to occur. Understanding the setting of a story is important to understanding the story's composite pieces, such as characters, events, and plot. This significance is even unscored in children's early education with the United States Common Core standards having setting detection as a key area of English education for kindergartners (Pearson, 2013). Part of the reason for the significance is that settings can enable us to make inferences as varied as customs/practices, technology, and character limitations. The inferences we can make have various levels of granularity depending on our knowledge of the time period or location.

Apart from such inferences, story settings are advantageous when conducting Cultural Analytics using literature. One of the reasons is what is called in philosophy the epistemic role of fiction (Green, 2022; García-Carpintero, 2016). Stories have a remarkable impact on people's understanding of the world. Empirical studies have shown this is the case even when people know the story is fictional (Murphy, 1998; Strange and Leung, 1999; Strange, 1993, 2002). This carries particular weight with historical fiction. These studies seem to suggest that it is hard to read or watch "War and Peace" without it shaping our view of the actual transpiring of the War of 1812.

Such epistemic uses of fiction seem to have had large social effects. For instance, the carefully researched 1852 novel "Uncle Tom's Cabin" was said to have a profound effect on the public's negative perception and consequential response to slavery (Reynolds, 2011). However, some works have been said to misrepresent racial relations, such as "Gone With The Wind"'s portrayal of the Civil War (Coates, 2018).

Having a dataset that distinguishes the time period the work was written in and the time period of its setting enables analysis of how truthful the work is in comparison to historical records. It also enables additional literary analysis. For instance, when doing cultural analysis of fictional character presentations, we may analyze not only how Victorian authors presented women in their "modern-day" novels but also how they presented women of the past.

However, despite the value of identifying a story's time period and location, there are currently no large or diverse datasets for this purpose. This paper presents such a dataset. The dataset is available for download under a Creative Commons Attribution 4.0 license at https://github.com/krittichier/StorySettings.

This paper is organized as follows: Section 2, describes related work that focused on determining time period and location from texts. In Section 3, the time period dataset is outlined, including the retrieval process as well as the cleaning, labeling, and baseline classification from the data. Section 4, outlines the location dataset construction and classification using simple metrics. Section 5, concludes the paper.

Proceedings of the The 5th Workshop on Narrative Understanding, pages 65–72

2 Related Work

There has been some work focused on the time period setting or other temporal aspects of stories. The first Narrative Q & A dataset (Kočiský et al., 2018) was offered in 2018 to evaluate reading comprehension. Of the over 1,500 textual works, only 6% had a time period setting Q & A combo and 7.5% had a location setting Q & A combo. Although there were some that dealt with the setting for a particular event, these also had low representation.

There was a publication on the passage of time within fictional works (Kim et al., 2020), where segments of text are labeled by the time of day they take place (i.e., morning, daytime, evening, and night.) Similarly, an annotation guideline for temporal aspects was published as part of the SANTA (Systematic Analysis of Narrative levels Through Annotation) project. This dataset addresses the issue of how to deal with jumps in the story timeline, such as Analepsis (a flashback) or Prolepsis (a flashforward).

A few projects have focused on time period classification for non-story texts. Most of the literature focuses on news data (Ng et al., 2020). The problem with this, as with many other cases of news data's use in natural language processing, is that news data is often written more explicitly than other text of text (Bamman et al., 2020); in this case particularly, temporal aspects are very clear, sometimes down to the hour (Ng et al., 2020). Additionally, news text is often much shorter, and therefore its time spans are smaller. Another example of non-story detection is EVALITY's 2020 task (Basile et al., 2020; Brivio, 2020). In this task, works were collected about the former prime minister of Italy, Alcide De Gasperi (Menini et al., 2020; Massidda, 2020). In this task, there were 2,759 works that were then split into five different categories for coarse-grained analysis and 11 different ranges for fine-grained analysis.

There have also been two attempts at identifying the time period of a text using time series(Mughaz et al., 2017; HaCohen-Kerner and Mughaz, 2010). These two papers are written by some of the same authors using different approaches. Their work differs from this one in that it deals more with the publication time period than the setting time period. The term frequency-based approach they use is not able to draw this distinction and therefore is less suited for the tasks of Digital Humanities and Cultural Analytics.

3 Time Period Dataset

`Project Gutenberg`[1] is the source of the literary works. Project Gutenberg is a resource that contains textual works in the public domain. At the time of this, The United States has the copyright set to expire 70 years after the author dies. As of 2019, all works written prior to 1924 are in the public domain. Although there are some public domain works that have been recently published, the majority of the works were published before that date. Around 80 percent of the works in Project Gutenberg are in English. Of these English texts, 40 percent are fictional texts. To determine the work's fictional status, a combination of LoC classification (namely sub-classifications of "P: Language and Literature", which were reviewed to be literature labels rather than language or literary criticism) and header terms (such as "fiction", "story", and "tale") were used.

Three primary resources were used for identifying the time period setting of the work. These resources are Library of Congress Subject Classification[2], `Wikipedia API`[3], and `SparkNotes`[4]. Library of Congress classifications of works are expert-labeled topics that include setting information. Wikipedia has categories of text related to the time period, such as "Set in the 1920s" or "Set during the Civil War." SparkNotes consist of expert reviews of works for study purposes. `BeautifulSoup`[5] was used to scrape the HTML SparkNotes webpages and retrieve the information about the works. Like many issues within machine learning, the difficulty lies in the scarcity of the data, as there were 2,302 works labeled with time periods settings after cleaning.

3.1 Resource 1: Library of Congress Data

Each work on Project Gutenberg has at least one Library of Congress subject. Most works contain multiple subjects. Each subject itself can be composed of what the Library of Congress (henceforth LoC) refers to as headers, which are separated in the subject by "– ". LoC, they are the same across

[1] https://www.gutenberg.org/

[2] https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html

[3] https://www.mediawiki.org/wiki/API:Main_page

[4] https://www.sparknotes.com/lit/

[5] https://beautiful-soup-4.readthedocs.io/en/latest/

both copies of the work. For instance, "The Scarlet Letter" has 11 subjects: "Adultery – Fiction", "Historical fiction", "Revenge – Fiction", "Psychological fiction", "Married women – Fiction", "Clergy – Fiction", "Triangles (Interpersonal relations) – Fiction", " Illegitimate children – Fiction", "Women immigrants – Fiction", "Puritans – Fiction", "Boston (Mass.) – History – Colonial period, ca. 1600-1775 – Fiction".

In order to make these headings useful, review was required. One problem was that many of the names were missing a beginning/end or listed multiple beginning dates. In these cases, the information for the person was found and filled in by hand using Wikipedia or a historical website as a resource. Sometimes different birth/death years are given than the LoC classification; in such a case, either the default or the one that offers a longer range is used.

Sometimes the reason a date is missing is more historically significant, such as in the case of the historical figure Pocahontas, where the birth is unknown. Additionally, sometimes fictional characters are the people listed with a range. This sometimes includes "(fictional character)" and other times does not but was only determined by searching. These do not have dates of death because the death of the character never took place in the book. The date for all these is approximated by using an approximate average lifespan of real people in the dataset.

Some of the ranges express uncertainty or have typos. For instance, some people have approximated deaths (and birth) times given on Wikipedia, such as Edmund Brokesbourne, whose Wikipedia page lists him as dying in either 1396 or 1397. Some cases lead to distinctly bigger ranges, but in all cases, the year that offers the longest range is selected. In order to deal with shortened versions of the names, we make sure that each piece of the names that were found has a historical name connected to it.

Lastly, there is a category of works that are "To X"; there are 47 works with only this label for the time period setting. In investigating the full subjects, the heading "To X" is embedded; there is no clear option for what to label these as. Therefore, the range used is simply (X, X) for all instances of this label. When splitting for classification, most of the works with this give a larger range than this, making it not affect the classification. However, this is made clear in the dataset and is able to be altered.

After the initial cleaning was completed, we conducted inspections to remove time period labels that indicate the time period the work was written in rather than the actual setting of the work. A notable case for this is the century label. When it is a setting, it is indicated with the heading "History" immediately preceding it. Of all the time period labels, 1,229 had only centuries as their label. Of these, 148 had "History" before it. Additionally, 47 others had those combined with other time period indicators. These history century labels were only used for the ones that did not have other time period labels, given the that they were too large of a range and the distribution: 68% of the centuries after the "History" subcategory is "19th century," while 89% span "17th century" to "19th century". Additionally, another time period indicating header that is indicative of the time written and not the setting is when author labels are included, such as "Shakespeare, William, 1564-1616". These were removed from the headings.

The total number of LoC headings for time period is 759. All of these headings were reviewed by hand to identify whether they represent a person, event, or simply a time period. Of these labels, 404 of these labels are names of people accompanied by their lifespan. 261 of the headers are events, which are broadly construed to include conspiracies and locations at particular times, such as "New Plymouth, 1620-1691". Of the remaining 93, 67 indicate year ranges, 18 of these indicate "ToX", and the remaining 8 indicate the centuries spanning from the 13th to the 20th century.

Of the 1923 works with time period labels that are not simply a century, 1,641 have only one subject (header), and 280 of them have multiple time periods indicating labels (people, events, etc.). 182 of these have at least one range that encompasses all of the other ranges. When this is not the case, a span of all of them is taken. The range of the setting years for all the works is 1000 to 2099 as "Two thousand, A.D." was used to describe two books set in the 2000s, which were written in the 1800s.

3.2 Resource 2: Wikipedia

Categories are a way that Wikipedia pages are organized and can be retrieved through the Wikipedia

API[6]. To gather the works from Wikipedia, the categories listed in the table where X stands for a number and '|' indicate different options. "BC" sometimes followed the century category:

- Novels|Fiction|Plays set in the Xth|Xst|Xnd century
- Novels|Fiction|Plays set in the Xs
- Novels|Fiction|Plays set in the X
- Novels|Fiction|Plays set in the Middle Ages

From these categories, there were 1,497 titles retrieved, and 311 (21%) were found to be unique works on Project Gutenberg. In order to avoid fiction of the same title getting mistaken for a work on Gutenberg, the Wikipedia pages were reviewed to find the author's name presented in the article. A few of the works contained the exact names of the authors. However, a by hand inspection of the remaining was needed as the authors' names come in many different variations, such as with/without accent marks, shortened/lengthened versions (e.g., Sam vs. Samuel), initials in place of names, missing middle name(s), and misspellings. The resulting number of books was 236. Part of the reason for the significant drop is that Wikipedia labels tend to focus on more recently published books, in other words, not those that are typically available on Project Gutenberg.

3.3 Resource 3: SparkNotes

In this section, we discuss the SparkNotes data. As of August 2021, the SparkNotes website has 710 works [7] of which it offers study guides that consist of descriptions and explanations. 464 of these works have a factsheet[8] associated with them containing information on specific details of the novel, such as Setting (Time Period), Setting (Location), tense, date of publication, etc. 156 of the works on the website are supplied by The Project Gutenberg, but only 101 of these works contain "factsheets" detailing aspects of the novel such as setting. By hand review of all of the 464 was done to verify the same title as SparkNotes sometimes does not use official names but rather what the work is commonly called, such as "Alice's Adventures in Wonderland" title being "Alice in Wonderland."

Of these 101 works, 10 of these are not literature/fiction. 5 of these are not the same book but simply the same title. The reference of the title to the correct work, and not another by the same title, is verified by the author's being reviewed by hand. Another 5 are removed because the time period could not be determined. The reason for this is that some have the setting marked as "unknown", and others are too vague such as in the case of "The Alchemist", where no time indications are given except for the advancements of technology. This would leave us with 81 works, yet one of the entries on SparkNotes is for both parts 1 and 2, which are separate books that cover the same time span, so these were split up. We are, therefore, left with 82 time period works from SparkNotes. In the dataset, the key is the URL for the work, and the id and filename are for Project Gutenberg retrieval.

Given that literary works do not always offer specific dates for the time period setting, this ambiguity is reflected in SparkNotes labeling. For instance, the time period of "The American" by Henry James is labeled by SparkNotes as "May 1868 and the several years thereafter". Of the 82 works, only 24 contain specific ranges. To deal with the variance, each 58 with the remaining, certain rules were used. Regular expressions and named-entity recognition was used to aid the labeling of the works, but each of the works was inspected by hand. A table for numerical interpretations of terms such as "mid", "late", and "early" (and their synonyms)is given in Appendix A as well as an explanation of the rules followed for other vague terms.

3.4 Resource Overlap

Some of the resources had overlap in works attributed values LoC and Wikipedia had 51 works that overlapped: 35 of these works the range fell within one another, 12 of these works had overlap (with an overlap average of 40 years), and 4 were disjoint from one another which was, on average, only a difference of 4 years. LoC and SparkNotes had 7 works in-common and 6 of the SparkNotes within the AoC label ranges. The only one that did not was the LoC label' Revolution, 1789-1799' for "The Tale of Two Cities," which takes place "1775–1793". Between Wikipedia and SparkNotes, Wikipedia was often too large of a range. There were 3 works that were in all 3 of the datasets. Be-

[6] https://en.wikipedia.org/wiki/Wikipedia:Contents/Categories

[7] https://www.sparknotes.com/lit/

[8] For an example of a factsheet, namely A Tale of Two Cities: https://www.sparknotes.com/lit/a-tale-of-two-cities/facts/

cause SparkNotes is the most precise and expert reviewed, its value takes precedence over all other labelings. Second is Wikipedia, i.e., Wikipedia's labels are used when Wikipedia and LoC both label the same work.

3.5 Dataset Construction

For time period setting, the dataset contains a zipped folder of all 2,302 works. For labels, it contains a JSON file that can be read in as a table. In the table, the time period is in the form of a tuple indicating its range. It also includes Project Gutenberg data/metadata: file name, id, title, author and years alive (e.g., "Hawthorne, Nathaniel, 1804-1864"), the list of LoC subjects, and the list of LoC classifications (such as "PR: English literature").

For interpreting the LoC headings, a JSON dictionary is supplied. Within this, each heading has a type label (year, event, or person), a start date, and an end date. So, "William I, Prince of Orange, 1533-1584" is labeled as a person and has a start date of 1533 and an end date of 1584. Given restrictions on distributing SparkNotes data, there are no such dictionaries offered. However, A offers a breakdown of the general rules applied. Also, due to the simplicity of the Wikipedia labels, no such dictionary for it is supplied.

3.6 Results

For the classification task, we use the TF-IDF score. This score is commonly used for document classification. It works by calculating the term frequency of a document and dividing it by the inverse document frequency. By doing this, the formula captures the significance of the words to the document rather than simply prominence in the document. This can be seen in the formula 1. In this formula, $tf_{i,j}$ is the frequency of the term i in file j, df_i is the number of files that contain i, and N is the total number of files.

$$w_{i,j} = tf_{i,j} \times \log(\frac{N}{df_i}) \quad (1)$$

Before running the TF-IDF algorithm on the works, they were cleaned to remove stopwords and lemmatized.

Given the various ranges the labels offer, they must be split into categories. The difficulty of this lies in the lack of clear thresholds. For instance, some novels may cover the first few years of the Revolution, while others cover the duration and the aftermath. Given that the dataset is already fairly small, we don't want to lose many of the works. For this reason, we give some wiggle room to thresholds in comparing the works to the threshold. The formula for softening the thresholds is allowing them to be up to 10 years off as long as the difference is less than 10% of the range. This metric was used because it appeared to best represent our concept of "close", and that the majority of the work would be in that range. In future work, other metrics may be tested.

The data was tested on three different numbers of categories:

- 3-way split where the soft thresholds are 1746 and 1877
- 4-way split where the soft thresholds are 1698, 1803, and 1898
- 5-way split where the soft thresholds are 1605, 1792, 1859, and 1912

The 3-way split reduced the total works down to 1850 split 545:681:624. The 4-way split reduced the total number of works down to 1686 split 471:211:454:550. The 5-way split reduced the total works down to 1595 split 286:303:207:326:473. Given that the 3-way split offers the evenest distribution and a similar breakdown to the EVALITY task mentioned in Section 2, split-3 was used for the baseline results. Table 1 shows the results using the top 100 TF-IDF features alone. Both Random Forest and Support Vector were able to give an F1 score of 0.81.

4 Location Dataset and Baseline Classification

In order to detect location data, LoC headings are used, as well as some SparkNotes headings. Additionally, datasets from `Simple Maps` are used for some of the world cities[9], The USA[10] and Great Britain[11]. Additionally, given the variance in state names, in LoC classification, much of the data consists of states which are abbreviated with either standard abbreviations (e.g., "AZ") or postal abbreviations (e.g., "ARIZ"). A table containing alternative state names (full and abbreviated) and postal was used. The reason for using these resources is that it enables a more robust part-whole classification than WordNet currently offers. Having the

[9] https://simplemaps.com/data/world-cities
[10] https://simplemaps.com/data/us-cities
[11] https://simplemaps.com/data/gb-cities

	Random Forest	SVM	KNN	Naïve Bayes	Decision Tree
Accuracy	0.8090	0.8090	0.7351	0.6955	0.6649
Precision	0.8260	0.8198	0.7569	0.6962	0.6717
Recall	0.8092	0.8104	0.7362	0.7049	0.6721
F1	0.8141	0.8125	0.7399	0.6985	0.6719

Table 1: Time period classification results using the TF-IDF score

part-whole relation offered a way to detect which country/state it was falling in and whether the city location was legitimate.

The results for both LoC headings and SparkNote location labels were reviewed by hand due to non-locations with the same term. For instance, though Battle is a place in England, but many battles took place in England, which is what is most often referred to with the term Battle and England in the heading labels.

The dataset included 6,962 gathered with LoC subjects. 689 of these works are labeled as having more than one location classification. There are 556 headings with identified locations. There are also 75 SparkNotes works with location(s), with around 22 having multiple location labels. 46 works are in both the LoC headers and SparkNotes, resulting in 6,991 works.

Baseline classification results using location setting were achieved using simple term occurrence metrics. 34.5% had the setting location as the most often mentioned location. 60.5% had the setting location (or the larger location it falls within, such as the country) as mentioned. The remaining 5% did not have any terms to indicate the location, and it remains an open question what content in the stories the annotators relied on in assigning the label.

This dataset covers a more simple version of location setting, namely geolocation. Other important features for location are whether it takes place in a house or, better yet, a certain character's house. However, this more nuanced version is only reflected in a few of the SparkNotes labels we see, with most having simple geolocation (e.g., country, city, state), which indicates a need for even simple location setting labels.

The dataset for the location settings is similar to the the one for time period described in section 3.5. It has a zipped folder of the works and a table that includes all of the same Project Gutenberg information. However, instead of each work having a location label column, there is a list of location-specific headings. These headings can be used as keys in the accompanying dictionary. Each key has an associated country and may also have a city and/or state based on the granularity of the label.

5 Conclusion

This paper presents a dataset of 2,302 time period setting labeled works and 6,991 location setting labeled works. The aim is for these to help with the detection of settings within stories and interesting Cultural Analytic findings by enabling analysis of cross-time-period writing and the role settings serve for story understanding. It can also help offer refinement/investigation into literary Q&A systems.

Additionally, this project serves as a way to investigate how beneficial metadata on Project Gutenberg or from LoC can be. The aim is that this will enable the use of the LoC classifications, which, to our knowledge, have not been capitalized on in natural language processing, at least at this scale or for this aim. There is also room for tracking more carefully where different portions of the work take place as can be seen to be important in the SparkNotes' labeling.

Limitations

Some of the limitations of this dataset include that of much of the time periods and locations given are simply approximations of the time period that the work is actually set in; this is most notable in the case of Library of Congress and Wikipedia labels which make up the majority of the work. These datasets offer more coarse-grained settings of a work, such as years and geolocation, which have limitations for some purposes. An additional limitation is that the works are in English and also are more commonly set/written in the West, which should be taken into account when used for analytics.

References

David Bamman, Olivia Lewke, and Anya Mansoor. 2020. An annotated dataset of coreference in English literature. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 44–54, Marseille, France. European Language Resources Association.

Valerio Basile, Maria Di Maro, Croce Danilo, and Lucia C Passaro. 2020. Evalita 2020: Overview of the 7th evaluation campaign of natural language processing and speech tools for italian. In *Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, pages 1–7. CEUR-ws.

Matteo Brivio. 2020. matteo-brv@ dadoeval: An svm-based approach for automatic document dating. In *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Online. CEUR. org*.

Ta-Nehisi Coates. 2018. Why do so few blacks study the civil war?

Manuel García-Carpintero. 2016. Introduction: Recent debates on learning from fiction. *Teorema: Revista Internacional de Filosofía*, 35(3):5–20.

Mitchell Green. 2022. Fiction and epistemic value: State of the art. *British Journal of Aesthetics*, 62(2):273–289.

Yaakov HaCohen-Kerner and Dror Mughaz. 2010. Estimating the birth and death years of authors of undated documents using undated citations. In *International Conference on Natural Language Processing*, pages 138–149. Springer.

Allen Kim, Charuta Pethe, and Steve Skiena. 2020. What time is it? temporal analysis of novels. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9076–9086, Online. Association for Computational Linguistics.

Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, and Edward Grefenstette. 2018. The NarrativeQA reading comprehension challenge. *Transactions of the Association for Computational Linguistics*, 6:317–328.

Riccardo Massidda. 2020. rmassidda@ dadoeval: Document dating using sentence embeddings at evalita 2020. In *EVALITA*.

Stefano Menini, Giovanni Moretti, Rachele Sprugnoli, and Sara Tonelli. 2020. Dadoeval@ evalita 2020: Same-genre and cross-genre dating of historical documents. In *7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. EVALITA 2020*, pages 391–397. Accademia University Press.

Dror Mughaz, Yaakov HaCohen-Kerner, and Dov Gabbay. 2017. Mining and using key-words and key-phrases to identify the era of an anonymous text. In *Transactions on Computational Collective Intelligence XXVI*, pages 119–143. Springer.

Sheila T Murphy. 1998. The impact of factual versus fictional media portrayals on cultural stereotypes. *The Annals of the American academy of political and social science*, 560(1):165–178.

Victoria Ng, Erin E Rees, Jingcheng Niu, Abdelhamid Zaghool, Homeira Ghiasbeglou, and Adrian Verster. 2020. Application of natural language processing algorithms for extracting information from news articles in event-based surveillance. *Canada Communicable Disease Report*, 46(6):186–191.

P David Pearson. 2013. Research foundations of the common core state standards in english language arts. *Quality reading instruction in the age of Common Core State Standards*, pages 237–262.

David S Reynolds. 2011. *Mightier than the Sword: Uncle Tom's cabin and the Battle for America*. WW Norton & Company.

Jeffrey J Strange. 2002. How fictional tales wag real-world beliefs. *Narrative impact: Social and cognitive foundations*, pages 263–286.

Jeffrey J Strange and Cynthia C Leung. 1999. How anecdotal accounts in news and in fiction can influence judgments of a social problem's urgency, causes, and cures. *Personality and Social Psychology Bulletin*, 25(4):436–449.

Jeffrey John Strange. 1993. *The facts of fiction: The accommodation of real-world beliefs to fabricated accounts*. Ph.D. thesis, Columbia University.

A Appendix

Type	Term	Start	End
Century	turn-century	XX00	XX10
	early-century	XX00	XX40
	mid-century	XX35	XX75
	late-century	XX60	XX99
Decade	early-decade	0	4
	mid-decade	3	7
	late-decade	6	9

Table 2: Conversions for SparkNotes' ambiguity

Phrases like "shortly after the turn of the 20th century" is assumed to be 10 years longer than the dates given. In other smaller cases, "several years after" is assumed to mean 5 years after that time. Likewise, terms like "around" are 5 years added to both sides. Additionally, there are some eras

used, such as Renaissance, Medieval, and Victorian eras. For these, historical references were used. In the case of multiple years given for different sections of the works (chapters, acts, etc.), the highest range is used. Also, with the presence of terms like "specifically" or "especially," the more specific range is what is used.

There are also times when multiple years, centuries, decades, or eras are given. Sometimes the variance refers to different sections of the work, such as the first chapter being set in X year and the second being set in Y. In these cases, the full range is used.