

Towards a robust framework for the semantic representation of temporal expressions in cultural legacy data

Daniel Isemann	Gerard Lynch	Raffaella Lanino
Natural Language Processing Group	Centre for Applied	Documentation and
Department of Computer Science	Data Analytics Research	Digitisation
Leipzig University	University College Dublin	National Gallery of Ireland
lastname@informatik.	Clonskeagh, Dublin 4	Dublin
uni-leipzig.de	Ireland	rlanino@ngi.ie
	firstname.lastname@ucd.ie	

Abstract

Date and time descriptors play an important role in cultural record keeping. As part of digital access and information retrieval on heritage databases it is becoming increasingly important that date descriptors are not matched as strings but that their semantics are properly understood and interpreted by man and machine alike. This paper describes a prototype system designed to resolve temporal expressions from English language cultural heritage records to ISO 8601 compatible date expressions. The architecture we advocate calls for a two stage resolution with a “semantic layer” between the input and ISO 8601 output. The system is inspired by a similar system for German language records and was tested on real world data from the National Gallery of Ireland in Dublin. Results from an evaluation with two senior art and metadata experts from the gallery are reported.

1 Introduction

Preserving a memory of past events has been central to human culture for millennia and may even be seen as a defining element of cultural life in general. The practice of specifying locations in time for this purpose transcends cultural boundaries. The earliest precursors of the Chinese lunisolar calendar can be traced back to the second millennium before Christ. In ancient Attica the “eponymous archon” lent his name to the year he ruled in and a similar system was employed by republican Romans. The introduction of the Julian calendar and its Georgian reform, although haphazardly adopted, has eventually led to a widely accepted standard for locating events in time (although alternative calendars exist and thrive to this day). The advent of the computer age has brought with it stricter requirements for such standards, for instance that of unambiguous machine readability. A number of such standards for encoding the meaning or extension of temporal expressions have emerged in recent years (ISO 8601, TimeML, VRA core). However, legacy records in the field of cultural heritage still abound with natural language descriptions of dates and date ranges that are not expressed in a standardised form, such as “around 1660”, “late 15th century”, “1720-30 (?)”. The non-standard nature of such expressions is compounded by inherent uncertainty about the dates which is expressed through uncertainty markers such as “around”, “(?)” or similar. While human experts have little difficulty interpreting such expressions these are not amenable to machine-based processing and thus are not directly useful for querying databases based on dates, for instance. The latter purpose is much better served by date ranges with a clear beginning and end.

We argue for conceptually splitting the process of “translating”, “converting” or otherwise associating informal descriptions of dates with concrete date ranges with an unambiguous beginning and end. A first step should capture the semantics of the original expression including possible uncertainty markers with as little loss in meaning as possible.¹ The target of this step should be a language independent ontology or semantic standard, such as the VRA core 4.0 date element. A second step should then map from a

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹What we mean by minimising “loss in meaning” is essentially that in a first step uncertainty markers should be kept (albeit in a standardized form) rather than resolved to a date range (i.e. “c.1888” → “c(1888)” instead of “c.1888” → “1878-1898”).

representation in the semantic standard to a date range with concrete beginning and end according to user, institution or context specific preferences, using intelligent defaults in the absence of preferences.

In this paper we describe a prototype system designed to resolve temporal expressions from English language cultural heritage records to ISO 8601 compatible date expressions. The system is inspired by a similar system for German language records and was tested on real life data from the National Gallery of Ireland in Dublin and evaluated by two senior art and metadata experts from the gallery. The default rules for converting the “meaning” of date expressions to date ranges were found to be superior to the heuristics currently configured by the National Gallery in their collection management system.

2 Background

The National Gallery of Ireland (NGI) has developed a set of in-house standards for cataloguing date expressions related to the creation of artworks (Appendix A and B). These standards complement the editorial guidelines outlined in the NGI house style guide for works of art in the collection and they must be applied when entering the data into the relevant field on TMS² (The Museum System), the collection management system used by the Gallery. These guidelines have been created based on best practice standards for cataloguing date expressions. There are several authoritative resources that institutions can consult to draft their own in house cataloguing standards, including date format and epoch descriptors: AAT³ (Art & Architecture Thesaurus), CDWA⁴ (Categories for the Description of Works of Art) or the AAE Style Guide⁵ (Association of Art Editors Style Guide), to mention just a few.

As shown in Appendix A and B, the NGI standards cover a diverse set of date expressions, from specific dates to more generic ones, giving the opportunity to enter into the system a range of years, decades or centuries. The date values are expressed as four digit years. More specific dates related to other events connected to the creation of the art work (for example for published volumes or different print editions), are recorded in the ‘Historical Dates field’ where the required date can be selected from a pop up calendar and the type of date can be selected from a drop down list (for example ‘Published’).

The Date label on TMS consists of three main fields: *Date*, which displays the actual date or range of dates related to the creation of the art work and which appears on the main object record screen as part of the basic object tombstone information; *Begin Date* and *End Date*, which represent the earliest and the latest possible years from a range of dates during which the artwork was created (Fig. 1). The Begin Date and End Date are not displayed in the Date label on the data entry screen of a record, as they are used for indexing and searching purposes only. Through the Simple Search and the Advanced Search functionality in the system it is possible to retrieve records with a range of dates, by either searching for earliest date, latest date or a certain time between these dates, the resulting records being drawn from the values recorded in the Begin and End Date.

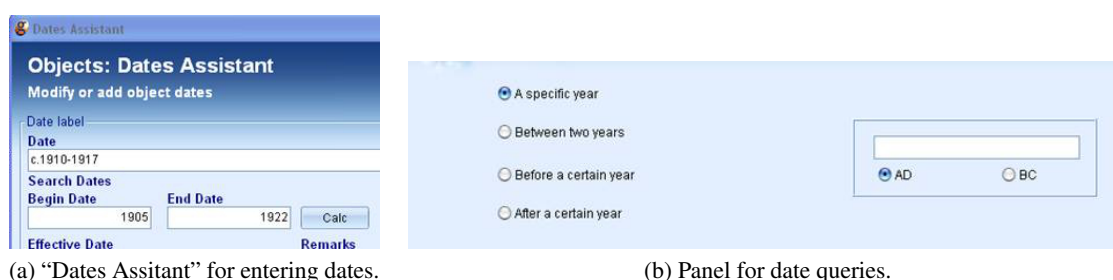


Figure 1: The “Dates Assistant” for entering dates into the database and a selection panel for date queries currently in use in the collection management system of the National Gallery of Ireland.

²<http://www.gallerysystems.com/tms> (last accessed 13/07/2014).

³http://www.getty.edu/research/tools/vocabularies/guidelines/aat_4_2_appendix_b_dates.html (last accessed 13/07/2014).

⁴http://www.getty.edu/research/publications/electronic_publications/cdwa/14creation.html#CREATION-DATE (last accessed 13/07/2014).

⁵<http://www.artedit.org/styleguide.htm> (last accessed 13/07/2014).

The Begin and End Dates can be inserted automatically by the system either by pressing the 'Calc' button or by accepting a suggestion for both Begin Date and End Date which is updated automatically every time a new value is inserted in the Date field. The suggestions can be accepted or modified manually and then saved. A date expression can also be suggested by the system when entering the relevant years directly into Begin and End Date. In this case the 'Calc' button prompts a pop up window with different date expressions based on the years inserted as beginning and end. For example, by entering '1575' and '1578' in the Begin and End Date, the suggestion box for the Date field will list the following options: '1575-1578'; 'c.1576'; 'late 16th century'. When date ranges include two specific years (for example in the case of 'YYYY/YYYY' or 'YYYY-YYYY') the two year values are automatically suggested in the Begin and End Date fields. When a single year is inserted in the Date field, the Begin and End Date are automatically filled with that same year value.

Through the configuration menu it is possible to specify the range of years to be 'suggested' in the Begin and End Date when entering a particular date expression in the Date field. By default this applies for the circa label (in the NGI case the range is 5 years before and after the specified date) and decades.

Although the automatic suggestions for Begin and End Date are configurable through the back end of the system, manual input is still necessary for accuracy when entering certain date expressions. Centuries for example (in all their formats, from 'xxth century' to 'early/mid/late xxth century') are not recognized by Begin and End Date, which in these cases need to be filled in manually. However the process works in reverse: when inserting the correct earliest and latest year that indicate a century span, the suggestion box for the Date field displays different options, including the correct 'xxth century' format.

On the other hand, in the case of decades, the relevant Begin Date and End Date are correctly suggested when inserting the 'YYYYs' format in the Date field, while, when entering the relevant years indicating the time span of a decade in the Begin and End date, the options listed as suggestions for the Date field do not include the correct format, giving instead the option of selecting 'YYYY-YYYY' as an alternative.

Similarly the Date field does not distinguish between years separated by an 'or', a dash or a hyphen when displaying the suggestions based on years inserted in Begin and End Date: when two different years are inserted in the Begin and End Date, the only relevant option listed by the system is the range of years separated by a hyphen. However, when entering the same date expressions in the Date field whether separated by 'or', dash or hyphen, the correct values are inserted in the Begin and End Date.

As the Date field is a free-text field on TMS, the process of manually entering date values, especially the ones that indicate uncertainty and include a prefix and non-numerical values, gives more room for error. In addition to this, not every date expression inserted in the Date field is recognised by the Begin and End Dates, in which case these also have to be entered manually.

At the same time the automatic suggestions given for the Date field when entering Begin Date and End seem to be more comprehensive and work better and they are helpful in giving the opportunity to select the correct option without having to manually enter the data, thus reducing the possibility of error. In the case of the NGI some configuration is further needed to make the most of the automated system already in place. In particular it would be useful to include in the provided suggestions for the Begin and End Date, those date expressions that are not currently recognised by the system.

3 Methodology and data set

The development of our system is inspired by an earlier system of temporal expression resolution for German language date expressions, an auxiliary part of a research project concerned with information retrieval on digital repositories of works of art, (Isemann and Ahmad, 2014). The approach was an iterative development cycle of successively resolving ever more complex date and time descriptors and mapping them to unambiguous time spans in ISO 8601 format. The data used were German date entries in a commercially available digital collection of 40,000 works of art.⁶ Example expressions from this data set are: "1707-1712", "1734/39", "1790-3", "12./13. Jh.", "1. Drittel 16. Jh.", "1420-1375 v. Chr."

These examples represent date ranges that have a fairly well defined beginning and end. One may perhaps argue whether the 13th century should include the year 1300 or not, but in general the intended

⁶The DVD collection "40000 Meisterwerke", published by Directmedia Publishing.

boundaries are reasonably clear. The following examples, however, are compounded by the fact that they contain uncertainty markers which leave the precise date range that should be assigned to them up to context and interpretation: “um 1568”, “1642 (?)”, “Vor 1650”, “ab 1486”, “nach 1776-77”.

For the experiments presented here, we obtained a similar although much smaller English language data set from the National Gallery of Ireland. The data consisted of 939 records from the NGI database, comprising date expressions such as “1791 to 1794”, “1870/72”, “1740s”, “18th century”, “1st February 1751”, “?c.1893”, “after 1752”, “late 16th century”, “mid-1930s”. Unlike in the German data set, most date expressions in the NGI data are already associated with a ‘Begin Date’ and ‘End Date’ either calculated by the NGI collection management system or manually entered by NGI staff (compare Section 2). These date ranges sanctioned by art experts present a valuable additional resource which may serve as training data for statistical learning or as a benchmark to compare against.

In contrast to the German language system we are conceptually using a two stage approach in which we first attempt to represent the intended meaning of a date expression (‘intension’) and only then map it to a date range for search and retrieval (one might call this range the ‘extension’ of a date expression). For the representation of date expression semantics (intension) we have chosen the VRA core 4.0 set of metadata elements and here in particular the ‘date’ element.⁷ VRA core is a set of categories defined and maintained by the Data Standards Committee of the Visual Resources Association.⁸ The latest version 4.0 dates from 2007. The standard has been used for semantic annotation (cf. Hollink et al. (2003) which use VRA core 3.0) and defines mappings to other metadata schemata, such as Dublin Core,⁹ CDWA,¹⁰ CCO¹¹ (Cataloging Cultural Objects) and its own predecessors (VRA core 2.0 and 3.0). As value ranges the standard recommends widely used thesauri (AAT¹²) or controlled vocabularies (ULAN¹³) or in the case of dates the ISO 8601 standard. Structurally, the standard prescribes that ‘date’ elements have a ‘type’ attribute (such as ‘creation’, ‘design’, ‘alteration’) and may have an ‘earliestDate’ and ‘latestDate’ subelement, both of which should only take ISO 8601 compatible values and can be modified by a boolean ‘circa’ attribute.

The semantic representation is the point of departure for the resolution of a date expression to a concrete date range. This leaves room for interpretation, especially in cases where a ‘circa’ flag is present. Ideally this mapping should be governed by preferences at the user and/or institution level (similar to the guidelines presented in Appendices A and B).

While the interpretation of these dates may vary on a case-by-case basis and even experts may disagree, we believe that certain default rules will allow at least a rough approximation of the intended time range in many cases. Analysing the data we noticed that mentions of years are not uniformly distributed in terms of the digit they end on. Figure 2 shows the relative frequency of year end digits for the German data set (red line) in expressions involving a ‘circa’ flag (German: “um”). Assuming a uniform distribution of years the frequencies should be 801.1 throughout. It is statistically extremely unlikely that the observed deviation from a uniform distribution is due to chance variation (chi squared test, 9 degrees of freedom, $p < 0.001$). As it appears equally unlikely that artists over the centuries have had a particular propensity to be more productive in years ending in 0 and 5, we believe that the natural explanation is that art historians documenting temporality tend to gravitate to “round” numbers in cases of greater uncertainty. As an upshot we would like to suggest that all else being equal approximate dates involving years should be seen as less certain if they end in 0 or 5 than if they end in other digits. Accordingly we add ± 10 years to years ending in “0”, ± 5 to years ending in “5” and ± 1 to years ending in other digits. Table 1 shows a number of the resolutions our system can perform.

⁷<http://www.loc.gov/standards/vracore/> (last accessed 13/07/2014).

⁸<http://www.vraweb.org> (last accessed 13/07/2013)

⁹<http://dublincore.org> (last accessed 13/07/2014).

¹⁰Categories for the Description of Works of Art, cf. Section 2

¹¹<http://vraweb.org/ccoweb/cco/intro.html> (last accessed 13/07/2014).

¹²<http://www.getty.edu/research/tools/vocabularies/aat> (last accessed 13/07/2014).

¹³<http://www.getty.edu/research/tools/vocabularies/ulan> (last accessed 13/07/2014).

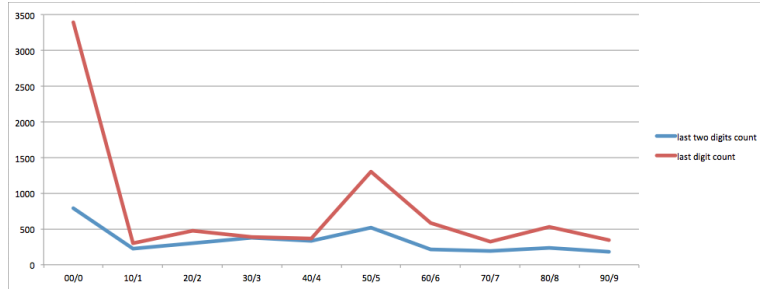


Figure 2: Distribution of year end digits in German expressions with an uncertainty marker (“um”, red line) and frequencies of “00” endings compared to other multiples of 10 (blue line).

Date expression	Date resolution	Date expression	Date semantics	Date resolution
1889	+1889/+1889	c.1824	c(+1824)/c(+1824)	+1823/+1825
1522/1523	+1522/+1523	c.1795	c(+1795)/c(+1795)	+1790/1800
1791 to 1794	+1791/+1794	c.1890	c(+1890)/c(+1890)	+1880/+1900
1870/72	+1870/+1872	?c.1893	c(+1893)/c(+1893)	+1892/+1894
1740s	+1740/+1749	after 1752	+1752/null	+1752/+1762
18th century	+1700/+1799	late 16th century	+1566/+1599	+1566/+1599
1st February 1751	+1751/+1751	mid-1930s	+1933/+1936	+1933/+1936

(a) Dates with well defined scope. (b) Dates with fuzzy scope.

Table 1: Date expressions from the National Gallery data set with default resolutions from our system. For the case of date expressions containing uncertainty or “fuzziness” we also show the semantic layer (b). Here “c(·)” represents a positive *circa* attribute in the VRA core *earliestDate* and *latestDate* subelements. Note, that not all expressions which may informally appear vague involve a *circa* attribute and that we assign a latest date by default for cases such as “after 1752”, contrary to the VRA core recommendation (which we adopt as semantic representation for such cases).

4 Experiments

We implemented a rule-based date expression resolver for the expressions in the English language National Gallery data set (achieving nearly complete coverage) with the set of heuristics outlined in the previous section (cf. Table 1). Two art history and meta data experts from the National Gallery agreed to participate in an evaluation of the output of our resolution system compared against the current date range entry in the National Gallery database. The entries in the NGI database are not a direct feature of the collection management system, but rather of how the system is currently used.

We observed that our system output agreed with the NGI entries in about half of the cases (58%). In order not to burden our volunteers’ time too much we did not evaluate on the complete data set, but on a randomly extracted subset in which we only included cases where our system output differed from the existing gallery records. We used a random number generator in Java to extract records until we reached a limit of 50 cases in which the two date interpretations were different. This limit was reached after selecting a total of 104 entries. The 50 non-trivial cases were compiled into a list comprised of the original date expression and a choice of two different date ranges each, one from the NGI records and one from our system. The order of the choices was randomized independently for each individual record.

The two evaluation participants were given this list together with a short introductory text outlining the background and purpose of the evaluation. They were then instructed to select which of the two date range alternatives they felt best captured the meaning of the date expression or indicate that they had no preference. Introductory paragraphs in the evaluation stressed that while individual context may sometimes enter into such a decision, they should think of the given date expressions as generic examples.

5 Results

Of the 100 individual decisions made by our two experts (50 each) exactly half (50) were in favour of our system’s default recommendation, less than a third were in favour of the existing database entry (29)

and just over one in five (21) had no particular preference (Table 2).

		Annotator A			Total Annotator B
		Our System	Gallery Records	No Preference	
Annotator B	Our System	8	0	3	11
	Gallery Records	19	2	6	27
	No Preference	12	0	0	12
Total Annotator A		39	2	9	

Table 2: Agreement of preferences expressed by our evaluators in 50 test cases.

While this may be seen as an encouraging result for our date range recommender system it has to be said that in their overall preference our two evaluators were leaning different ways. While one overwhelmingly agreed with our system recommendations (preferring the NGI alternative in just two cases with nine ties), the other was leaning towards the NGI records (preferring our system in just eleven cases with twelve ties). Overall the two evaluators agreed in ten of the 50 cases (Cohen’s kappa = -0.048).

We believe that the reason for the differing opinions between our two evaluators may be that one of them is working closely with the NGI database and is therefore very familiar with the status quo, including certain agreed in-house standards. The other evaluator, who was leaning towards the rules implemented in our system, is from the curatorial department and concerned with absolute and relative dating of works of art in a more theoretical way. A more thorough evaluation is needed in order to determine if the more flexible rules we are advocating would be appreciated by an expert user community.

6 Related Work

The resolution of temporal expressions is an important topic in the information extraction and semantic web community and employing these methods on cultural heritage texts in particular has been the focus of research spanning these fields and the emergent discipline of digital humanities.

Context-free grammars (CFG) for the resolution of temporal expressions have been employed by Angeli et al. (2012) and Kauppinen et al. (2010). Angeli et al. (2012) attempt to learn a probabilistic CFG for time and date expressions and at the same time an expectation maximization framework for the resolution of pragmatic ambiguity in time expressions (e.g. ‘Friday’ may refer to last or next Friday, ‘last Friday’ may refer to the previous Friday or the Friday two weeks ago etc.). For training their system they employ the TempEval-2 Task A dataset.¹⁴ Despite the relatively small training set (1052 time expressions) they report comparable performance of their system with leading rule-based temporal resolvers.

Kauppinen et al. (2010) employ fuzzy sets towards the representation and querying of temporally disputable periodic expressions from cultural heritage such as ‘late-Roman-era’, ‘Middle Ages’ or ‘beginning of the 1st century BC’, which can vary due to subjectivity or lack of hard records. They define a date span with a fuzzy beginning and end which encompasses the widest possible bounds for a temporal period and then a more concise beginning and end which encompasses more constrained bounds. Queries are matched against the fuzzy set using a bespoke querying model which finds the level of overlap between the query and the fuzzy set. They test their theories on a set of records from the Ancient Milan¹⁵ project, representing fuzzy date ranges as four RDF triples, one for each of the date points. They represent definite temporal expressions such as *First half of the 1st Century BC* in Backus-Naur form.

Research into frameworks for temporal expression extraction in the computational sciences, (Chang and Manning (2012), Strötgen and Gertz (2010), Sun et al. (2013)) has tended to focus on domains such as clinical texts and newswire for developing temporal expression resolution systems. We believe, however, that there is a clear and present need for systems and frameworks which can extract structured information from cultural heritage text, particularly in the domain of fine art image catalogues. These methodologies can enable the development of smarter retrieval systems for catalogues of cultural history

¹⁴Cf. <http://timeml.org/tempeval2> (last accessed 13/07/2014).

¹⁵<http://www.csai.disco.unimib.it/CSAI/space/CuRM/projects+and+research/Milano+Antica> (last accessed 13/07/2014).

data. Grandi and Mandreoli (2001), Grandi (2002) describe work on representing a geographical history resource, *il Dizionario geografico, fisico e storico della Toscana*¹⁶ created by cultural historian Emanuele Repetti in the early 19th century. They focus on the resolution of temporal expressions' indeterminacy and varying granularity in Italian temporal expressions, such as *around X*, *circa. X*, *near the end of the X century* and others. They represent such indeterminacy using a four category classification of date expressions and a *probabilistic* approach from the TSQL2 standard, (Snodgrass et al. (1994)). Lillis and others (2005) use multidimensional RDF in their representation of cultural artifacts in a museum setting.

Smith (2002) focuses on detecting events in unstructured historical text with dates forming the main focus of his study. The author investigates the co-occurrence of place names and dates in 19th century text and extracts a geo-located list of events from the text. He mentions that 98% of numerical tokens in the texts refer to dates, although in different text genres, date information may be more vaguely expressed. Furthermore, he finds that certain dates are expressed as a calendar day and others refer merely to the year an event occurred. These expressions can prove problematic for traditional date processing algorithms, and often a more complex mapping is required to convert these textual representations to a computational formalism such as the CIDOC specification. Chang and Manning (2012) focus on generic temporal expressions with their SUTIME parser, which represents date and temporal information extracted from text using the TIMEX3 tag format from the TimeML (Boguraev and Ando (2005)) standard.

An emerging trend in date resolution literature encompasses the *big data* paradigm. Blamey et al. (2013) develop a probabilistic approach toward modelling everyday natural language date expressions¹⁷ using textual data from image descriptions and EXIF¹⁸ data from uploaded photos on the flickr website.

7 Conclusion and Future Work

We have presented and tested a system specifically designed for the resolution of date expressions in cultural heritage legacy records and we have argued for a 'semantic layer' between the literal expressions and the date range resolution. Our evaluation, although small scale, suggests that such a system may potentially be able to improve even records which already incorporate date resolutions, if slightly more complex rules than are contained in the current system or data entry guidelines are implemented.

A number of possible lines of future work suggest themselves. In order to arrive at an explicit local grammar for 'heritage dates' (cf. Kauppinen et al. (2010) and Angeli et al. (2012)), we have created a context-free grammar, that accepts roughly the same input as our current rule set. Initial examination suggests that the non-lexical part of the grammar can cover both English and German language data given appropriate lexicons. The grammar phrases can be mapped to representations in the semantic layer thereby in effect creating a system which could process multilingual input and produce consistent output.

A further extension to the system would involve the processing of semantically more complex temporal period expressions, such as "Victorian", "Edwardian", "Gründerzeit", "Gilded Age" or "Renaissance". Examples tied to the reign of a monarch tend towards a more defined scope however wider-ranging and more culturally-disputed periods such as "the Renaissance" tend to attract a less precise beginning and end-date than the former examples and may require a more complex set of semantics. Data-driven approaches could be employed to model the temporal boundaries for temporal expressions of a more vague nature. Angeli et al. (2012) demonstrate that this may be feasible even on relatively small datasets.

Examples which could benefit from an ontological augmentation involving events and periods include the practice of dating works of art with implicit reference to such periods based on a believed or previously confirmed date for a major event such as a battle or war. One example of this practice could be an artwork dated "after 1453", with the date actually representing the current dating of the fall of Constantinople. As historical information is updated or revised, the corresponding date range estimating the temporal origin of a work could be resolved based on updated information for the reference event. Similar suggestions were made in (Isemann and Ahmad, 2009). Perhaps the practically most relevant example of this kind could be cross-referencing the lifespan of an artist associated with the production

¹⁶A geographical, physical and historical dictionary of Tuscany

¹⁷Their work focuses on UK-specific cultural expressions such as Bonfire Night, first day of summer, Christmas holidays.

¹⁸Timestamps saved by digital cameras.

of a work with the temporal expression for its creation: if the expression says “after 1756” but we have concrete knowledge that the artist died in 1758, this can be used to add bounds to the creation event.

Acknowledgements

The authors would like to thank Dr Adriaan Waiboer, curator for Northern European Art at the National Gallery of Ireland, for valuable suggestions and his help in organising the expert evaluation.

References

- Gabor Angeli, Christopher Manning, and Daniel Jurafsky. 2012. Parsing time: Learning to interpret time expressions. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 446–455, Montréal, Canada, June. Association for Computational Linguistics.
- Ben Blamey, Tom Crick, and Giles Oatley. 2013. ‘The First Day of Summer’: Parsing temporal expressions with distributed semantics. In *Research and Development in Intelligent Systems XXX*, pages 389–402. Springer International Publishing.
- Branimir Boguraev and Rie Kubota Ando. 2005. Timeml-compliant text analysis for temporal reasoning. In *Proceedings of the 19th international joint conference on Artificial intelligence*, pages 997–1003. Morgan Kaufmann Publishers Inc.
- Angel X Chang and Christopher Manning. 2012. Suntime: A library for recognizing and normalizing time expressions. In *LREC*, pages 3735–3740.
- Fabio Grandi and Federica Mandreoli. 2001. The “XML/Repetti” Project: Encoding and Manipulation of Temporal Information in Historical Text Sources. In *ICHIM (2)*, pages 243–252. Citeseer.
- Fabio Grandi. 2002. Xml representation and management of temporal information for web-based cultural heritage applications. *Data Science Journal*, 1(1):68–83.
- L. Hollink, A. Th. Schreiber, B. Wielemaker, and B. Wielinga. 2003. Semantic annotation of image collections. In *Proceedings of the KCAP’03 Workshop on Knowledge Markup and Semantic Annotation*, Florida, USA, October.
- Daniel Isemann and Khurshid Ahmad. 2009. Navigating cultural heritage in style. Sketching an ontological representation of metadata: the example of artistic period and style. *Museum Ireland*, 19:149–155.
- Daniel Isemann and Khurshid Ahmad. 2014. Ontological access to images of fine art. *Journal on Computing and Cultural Heritage (JOCCH)*, 7(1):3.
- Tomi Kauppinen, Glauco Mantegari, Panu Paakkarinen, Heini Kuittinen, Eero Hyvönen, and Stefania Bandini. 2010. Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International journal of human-computer studies*, 68(9):549–560.
- Pantelis Lilis et al. 2005. A metadata model for representing time-dependent information in cultural collections. In *MTSR, First online metadata and semantics research conference, Conference Proceedings*. Citeseer.
- David A. Smith. 2002. Detecting and browsing events in unstructured text. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’02*, pages 73–80, New York, NY, USA. ACM.
- Richard Thomas Snodgrass, Ilsoo Ahn, Gad Ariav, Don S Batory, James Clifford, Curtis E Dyreson, Ramez Elmasri, Fabio Grandi, Christian S Jensen, Wolfgang Käfer, et al. 1994. Tsql2 language specification. *Sigmod Record*, 23(1):65–86.
- Jannik Strötgen and Michael Gertz. 2010. Heideltime: High quality rule-based extraction and normalization of temporal expressions. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval ’10*, pages 321–324, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Weiyi Sun, Anna Rumshisky, and Ozlem Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association*, 20(5):806–813.

A NGI definitions/explanations of date expressions

Rules for object date

Date	Definition
1513	executed in 1513
c.1513	executed sometime around 1513 (for the purposes of search dates in TMS this will be 5 years either side e.g. begin 1508, end 1518)
1513-1515	begun in 1513, finished in 1515
1513/1516	executed sometime around the years 1513 to 1516
1513 or 1515	executed in either 1513 or 1515

Do not use c. (for circa) with a / (slash).

Figure 3: Rules for interpreting object date descriptions in NGI records.

B NGI guidelines for entering dates and date ranges

	Display Date	Search Date Begin	Search Date End	Rule
Single date	1855	1855	1855	
Begin date and end date (began 1630, finished 1633)	1630-1633	1630	1633	Separate the 2 dates with a dash, no spaces between Include all 4 digits for both years
One of 2 years (work was done in either 1631 or 1633)	1631 or 1633	1631	1633	Use the word 'or'. Include all 4 digits for both years
Range of dates (work was done sometime between 1745 and 1748)	1745/1748	1745	1748	Separate the 2 dates with a slash, with no spaces between Include all 4 digits for both years
Decades	1930s early 1930s late 1930s	1930 1930 1935	1939 1934 1939	no apostrophe before 's' 'early' and 'late' all in lower case
Circa	c.1900 c.1600 or 1610 c.1510-1520 c.1510-c.1520	1895 1595 1505 1505	1905 1610 1520 1525	Use c. Don't use 'about', 'circa', 'ca' or 'c' No space between c. and date
Before	1686-before 1770 before 1686-1750	1686 no begin date	1770 1750	Use 'before' Don't use 'prior to'
After	1823-after 1941 after 1822-1900	1823 1822	no end date 1900	Use 'after' Don't use 'post'
Uncertainty	?1750	1750	1750	Place a question mark before the doubtful element without a space. Do not use 'probably'
More precise dates	November 1900 21 January 1890	1900 1890	1900 1890	Dates given in full should be entered as day month year without punctuation or ordinal abbreviation such as rd, th, nd

Table 3: Data Entry Conventions when entering Object Dates and Search Dates into TMS (excerpt).