Graphing Trees: The Nodes and Edges of Nabokov's Worlds

Shakeeb Arzoo[1,*], Aidan Hogan[2] and Stephen H. Blackwell[3]

[1]*Development Editor, International Vladimir Nabokov Society*
[2]*IMFD; DCC, University of Chile, Chile*
[3]*Department of World Languages and Cultures, University of Tennessee, Knoxville*

Abstract

This paper motivates and demonstrates the use of knowledge graphs (KGs) in the literary domain by annotating a specific set of entities, namely trees, in the works of Vladimir Nabokov. We identify and collect 130+ distinct tree species with 2000+ mentions across 20 books and 5300+ print pages in the Nabokovian corpus. These allusive mentions are linked to external databases, putting forward a novel, contextually enriched knowledge base. We also propose a semantic data model, the Tree Annotation Vocabulary (TAV), to capture these annotations alongside their broader literary associations. Lastly, we present a sample evaluation of existing tools to highlight the complexities of entity recognition in a literary context.

Keywords

Knowledge Graphs, Ontology, Semantic Annotation, Entity Recognition, Entity Linking, Literary Criticism, Nabokov, Trees

1. Introduction

Philosopher Richard Eldridge [1] defined art as something which "presents a subject matter as a focus for thought and emotional attitude, distinctively fused to the imaginative exploration of material." To explore art—an artwork—as an object of "knowledge" is a deceptive, tricky, amorphous, ephemeral project. This is in part because any artwork—visual, literary, performance—is not a "thing"; rather, it is a nexus of experiences, of stimuli, and of patterns organised (intentionally and unintentionally) by a creator, and consciously perceived by an audience. As the once-famous Russian literary critic Yuli Aikhenvald put it: "Art is not a fact; art is an act" [2]. An artwork exists in the fluid dynamics of its reception by an intensely engaged audience; the formal study of artworks offers not definitive knowledge, but various snapshots of that evolving reception process, and the materials that provoke it.

Any artwork (including literary works) is a collection of significant units; these units are arranged by an artist with varying degrees of complexity and intentionality. They exist on several interlinking planes: lexical, rhetorical, imagistic, figurative, symbolic, rhythmic, self-reflective, intertextual, and so on. Some artists strive not only to create good stories, but also to create exceedingly complex systems of "significance", based on subsets of any number of the planes mentioned above. These systems are designed as patterns, and the patterns on different planes may interconnect, or "weave," into a multiplanar tapestry. In order to experience the richness— the artistic essence—of such systems, patterns, and tapestries, one must be willing to develop extensive sensitivity to, or "knowledge" of, significant units in a selection of planes. This constitutes the "intense engagement" on the part of its audience.

Vladimir Nabokov is among the artists most committed to creating these multidimensional networks within their works. He was explicitly interested in the creation and discovery of patterns, in both Art and Nature: he also worked as an entomologist for several years. Some of the most important scholarship on Nabokov's work has been grounded on discovering and deciphering various of these interlinked patterns [3, 4, 5, 6, 7, 8, 9] in order to elucidate otherwise hidden meanings contained by

*Corresponding author.

✉ arzoo.ebbs@gmail.com (S. Arzoo); ahogan@dcc.uchile.cl (A. Hogan); sblackwe@utk.edu (S. H. Blackwell)

the texts. The present project builds on Blackwell's analysis [10], which identifies and explores several significant networks built around patterns of trees within Nabokov's literary art. While that study presented only a handful of deep excursions into the world of trees within the art-space, the current project aims to make Nabokov's world accessible specifically through the lens provided by the totality of its 2,000+ trees. Though much remains to be discovered, we hope that the tools resulting from this project will enhance both knowledge and artistic experience surrounding Nabokov's creations.

Nabokov was an intense observer and writer of Nature. This is brought out by his "tree-networks", i.e., the patterns through which trees occur in his works that connect their natural, tangible, and specific reality to their symbolic and functional aspects. We find a plethora of references to named and unnamed trees alongside their ecological habitats, their anatomical aspects (e.g. twigs, tree trunks), functional aspects (shadows, climbability), and craft uses (furniture, canes, paper, etc.). Due to their density and their natural camouflage, it is not until we pay close attention to the way these are structured that we notice the broader thematic concerns, the "web of sense" [11] underlying the pervasive tree mentions.

Knowledge Graphs offer a way to model domain data with a seamless interconnectivity across different datasets that is flexible enough to accommodate incomplete data. Furthermore, they are intuitive to understand (both for humans and machines) and expressive enough to represent complex statements. Finally, the matching of basic graph patterns in the model allows for complex querying, whereby a faceted view of data can be obtained (with different filters) in a tabular format [12]. For our purposes, KGs help to capture the finer details of literary knowledge and link them to encyclopaedic knowledge which would capture its broader context. This conveys the essence of literary annotation which mostly introduces relevant explanatory information to the main text through the use of inline notes, footnotes, endnotes, sidenotes, references, among others [13].

In this undertaking, we construct a Tree-KG that explicitly maps out the kind of role(s), meaning(s), and importance attached to trees in Nabokov's works. There is substantial evidence to show that he used them as a crucial design choice in his stories [14, 15] and we wish to use them as a critical and interpretive device that brings out their importance to the conceptualization and design of a Nabokov story. This KG actualizes the implicit knowledge through which one can focus on a particular tree, see all its appearances across Nabokov's work, look at the various meanings it embodies, i.e., the thematic sense, geographical occurrence (whether endemic or exotic to a particular location), botanical classification, quantitative aspects, as well as the exact location of the phrases. KGs are not always well-suited to complex utterances (for that we have books and articles), but it can connect a myriad different things, from trees in Shakespeare (weeping willow, mulberry), Goethe (ginkgo trees), Chekhov's Cherry Orchard (cherry trees), the cypress trees in a painting by Böcklin, or the Bosch painting of The Garden of Earthly Delights, all that are absorbed and recreated by Nabokov in his own work.

The main contributions of this paper are as follows:

1. Identification and collection of tree-like entities from the Nabokovian corpus, presenting them as an annotated dataset.

2. An ontological data model adhering to the best practices of Linked Data Principles that captures the annotated data.

3. An illustrative benchmarking of existing technology that helps in entity recognition.

The rest of the paper is organized as follows: Section 2 provides background to the work; Section 3 discusses the ontology that captures the literary annotations; Section 4 presents our approach where we identify and link the tree-like entities in the Nabokov corpus; Section 5 presents some of our experiments with entity recognition; Section 6 proposes some directions for future work; and Section 7 concludes.

2. Background

Nabokov's stories have often been compared to a hypertext-like artifact [6, 16] with surprising linkages and associations joining disparate elements. Such links of meaning have captivated readers and garnered

scholarly attention ever since he gained popularity as a major literary giant of the 20th century. These covert patterns and the hidden layers of significance have been observed to emerge gradually, as we begin to pay attention to the verbal texture and the density of allusions behind them.

A character in Nabokov's Pale Fire [11] observes that "Our poet shared with the English masters the noble knack of transplanting trees into verse with their sap and shade". This holds true for Nabokov as well, as trees in his prose are invoked through sensory details (visual, aural, olfactory, etc.), functional use (as wooden objects) in human culture, and as objects with a figurative dimension (conceptual objects). The following passages illustrate those meanings with the relevant entities marked in bold.

SCENARIO 1: Specific Reality to the Setting of a Story

*A **pea-tree hedge** (the "**yellow acacia**" of northern Russia), with a midway opening, corresponding to the court's screen door, ran parallel to the enclosure and to a path dubbed **tropinka Sfinksov** ("path of the **Sphingids**") because of the **hawkmoths** visiting at dusk the **fluffy lilacs** along the border that faced **the hedge** and likewise broke in the middle.*

Speak, Memory [17], Ch. 2, p. 41

In the above passage from his autobiography, Nabokov is quite specific about the botanical identity of the Siberian pea-tree [18] which evokes the Russia of his boyhood. He advocates following the path of exactness and fidelity (to the best of a memoirist's ability) that would evoke a particular locale with its own set of flora and fauna, conveying a sense of "reality" of a particular time and space to his reader. No doubt he was aware that many readers would not care whether the tree is a locust tree, or an acacia, or a pea-tree, just as most readers fail to distinguish between an oak and an elm. However, he hoped that through curiosity, through repetition, or with the help of scholarly annotation, we would be more attentive and sensitive to the world surrounding us.

SCENARIO 2: Objects of Human Use

*It was not a hexagonal beauty of **Virginia juniper** or **African cedar**, with the maker's name imprinted in silver foil, but a very plain, round, technically faceless old **pencil of cheap pine**, dyed a dingy lilac.*

Transparent Things [19], Ch. 3, p. 6

Trees are often transformed into new objects through a combination of craftmanship and violence (such objects can also help in creation of art, e.g., violins, penholders), which are so ubiquitous in human culture that one ceases to think how they are produced. Such objects are often used by Nabokov at crucial points which would not only advance the plot (e.g., Lips to Lips [20], where a "wooden cane" plays a significant role in the plot) but also in a symbolic sense (e.g., in Invitation to a Beheading [21], where he describes "a beautifully sharpened pencil [...] as long as the life of any man").

SCENARIO 3: Figurative Way

*The novel was the famous **Quercus**, and Cincinnatus had already read a good third of it, or about a thousand pages. Its protagonist was an **oak**. The novel was a biography of that **oak**…Employing the gradual development of **the tree** (growing lone and mighty at the edge of a canyon at whose bottom the waters never ceased to din), the author unfolded all the historic events—or shadows of events—of which the **oak** could have been a witness…It seemed as though the author were sitting with his camera somewhere among the **topmost branches of the Quercus**, spying out and catching his prey. Various images of life would come and go, pausing among the green macules of light. The normal periods of inaction were filled with scientific descriptions of the **oak** itself, from the viewpoints of dendrology, ornithology, coleopterology, mythology—or popular descriptions, with touches of folk humour. Among other things there was a detailed list of all the **initials carved in the bark with their interpretations**.*

The above passage is a tour de force of how literature works with its elaborate artifice and the mechanisms of perspective. Nabokov proposes a meta-novel about an oak tree that serves as a focal point in time, space, and memory through which a narrative could be anchored. Undoubtedly, Nabokov carries documentary realism to its absurdist pitch with this example (by imagining the protagonist as an oak, a passive observer with limited agency) but there is little doubt that its symbolic purpose is to underline the role of an artist who is at the boundary of everyday, mundane details and the creative, extra-mundane existence.

These three examples give a hint of the 2000+ mentions comprising more than 130 types of tree in Nabokov's corpus. It can be readily seen that a tree in Nabokov, apart from its tree-like qualities[1], is embedded in a wealth of additional details. Far from being excessive, these details form patterns of significance that engage the general reader as well as the specialist in varying ways. In other words, by examining the different contexts in which the tree-like entities occur, we find a range of meanings associated with them (including those mentioned in the above examples). These meanings not only link the botanical notion of trees to their implied geographical distribution, they also encompass certain literary devices (themes, allusions, etc.) that connect the tree-units across a set/s of works. In effect, Nabokov intends us to break free of the linear flow of the narrative and adopt a multidimensional perspective by engaging with the text in a heightened state of attention [23, 24].

3. Ontological Representation

Our aim is to semantically annotate Nabokov's works by identifying entity mentions and describing them according to an ontological model. Central to the idea of capturing entity mentions within the confines of an ontological model lie three aspects. First, the *information object* (a website, a book, a line of verse) that serves as the primary "source" of information; second, the *fragment* within the source object that contains the exact information ("mentions" containing the "target entities"), and finally, the *targets* themselves, that have to be highlighted and linked to an external knowledge base.

The ontology development followed the approach laid down by the W3C recommended Open Annotation (OA) model [25], which describes an annotation as "a set of connected resources, typically including a body and target, and conveys that the body is related to the target." The exact nature of this relationship changes according to the intention of the annotation, but the body is most frequently somehow about the target. Although the OA model is primarily recommended to link web resources (as reflected by their vocabulary) it can be tweaked to encompass physical documents. This is consistent with the best practices of ontology development and conceptual modelling [26].

Let us take an example that demonstrates our modeling. The information object is Nabokov's first novel, Mashen'ka (1926), which was translated by the author himself (1970). The electronic editions used in our graph model are the Vintage International editions (standard reference for scholars of Nabokov). The selected text reads as:
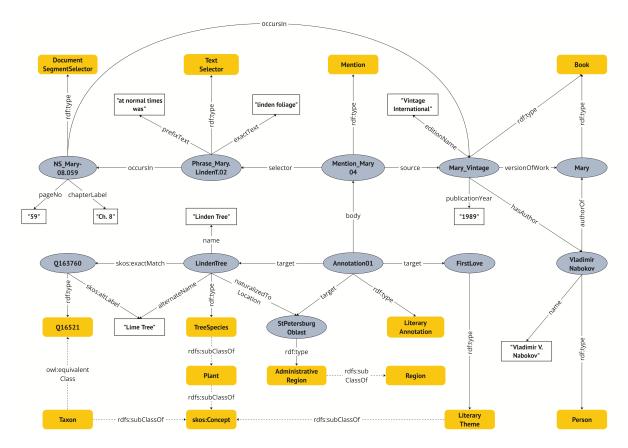
And when the whole surface had turned a ridiculous lilac color and Mary's fingers looked as if she had just been picking bilberries, Ganin, turning away and staring hard through narrowed eyes at a yellowy-green, warm, flowing something which at normal times was linden foliage, announced to Mary that he had been in love with her for a long time.

The RDF graph annotating the tree mentions in this text per our model is shown in Figure 1; for these purposes, we define the Tree Annotation Vocabulary (TAV).

Our model encompasses the detailed bibliographic information that includes the different manifestations of the textual object, Mary. It is based on the Schema.org vocabulary that was extended by

---

[1]An engaging discussion in this regard is provided by Colin Tudge [22].

Figure 1: An example annotation in RDF illustrating the TAV ontology. External vocabularies appear with a prefix. The TAV ontology is available at https://w3id.org/tav. Q163760 is the Wikidata ID for the Tilia x europaea (common lime) species of the Linden (Tilia) genus. Q16521 is the Wikidata Taxon class with which TAV is linked.

the W3C community[2] to share information about bibliographic objects. The entity mentions form an instance under the `tav:Mention` (`rdfs:subClassOf oa:SpecificResource`) class. This forms the body of the annotation (`tav:LiteraryAnnotation`) while the target "LindenTree" forms an instance under `tav:TreeSpecies`. The place mentioned in the text is "StPetersburgOblast", an instance under `tav:AdministrativeRegion` (`owl:equivalentClass` of `schema:AdministrativeArea`), while the associated literary theme is "FirstLove", both of which serve as targets of the annotation. The purpose in re-creating equivalents of external terms under our TAV vocabulary was to keep our model self-contained, i.e., avoid external dependencies (a.k.a. changing versions of external vocabularies).

As mentioned before, the various mentions of "trees" serve several purposes, ranging from intertextual to intratextual and the extratextual [28]. These allusive mentions are modelled into our ontology with the help of classes such as `tav:LiteraryTheme`, `tav:WoodenObject`, object properties like `tav:exoticToLocation`, `tav:naturalizedToLocation` with respect to a particular location (`tav:Region rdfs:subClassOf schema:Place`) mentioned in the text. Some competency questions (CQs) posited to help with ontology development are:

1. Where does (X) tree appear in a work by Nabokov?

2. What botanical species is appropriate for the (X) tree appearing in Nabokov?

3. What places are associated with (X) tree in a work by Nabokov?

4. What literary themes are associated with (X) tree in a work by Nabokov?

5. What wooden objects are thematically significant in a work by Nabokov?

---

6. What other notable authors or creative works mention (X) tree?

7. What other tree-like entities appear in metonymic relation to (X) tree?

More CQs can be accessed from ontology documentation online (see Appendix A). Finally, the annotated entities are linked to external knowledge bases, in our case, Wikidata and DBPedia. From our purview of botanical databases (e.g., WFO Plant List[3], Kew Plants of the World Online[4], Global Tree Portal[5]), Wikidata was found to be the most intuitive to understand, browse, and link. This linking is expressed through the various sub-properties under `skos:mappingRelation` [29]. The associative relationships offered by the SKOS vocabulary allow us to link and match entities, even when an exact match is not available in the target knowledge base. We consistently used `skos:broadMatch` when linking a particular tree mention to its genus level and `skos:closeMatch` when there were different varieties of a particular species (e.g., *Quercus robur*, or the English Oak has several sub-species and hybrids depending on the region where they are found). `skos:exactMatch` was used when there was substantial evidence to show that a particular species is found in the region mentioned in the text or is mentioned as such by Nabokov. The next section will provide some examples of the same.

At present, the TAV ontology comprises 48 classes, 38 object properties, and 19 data properties. It was developed on Protégé 5.5.0 [30]. It can be accessed from https://w3id.org/tav which contains some sample individuals. When mapped to external vocabularies, TAV comprises 78 classes, 55 object properties, and 29 data properties. The various versions of TAV can be accessed from Appendix A mentioned at the end of the paper.

The next section discusses how entities were spotted in the text and linked to knowledge bases.

4. Entity Recognition and Linking

To spot and identify the different species of trees mentioned in the Nabokov corpus, we adopted a laborious but precise approach. By the Nabokov corpus, we refer to a total of 20 books published in English (which includes his autobiography, a collection of 68 short stories, novella, etc.) with a count of over 2.6 million words (2,665,000 words) spanning 5300+ pages in print. We were aware of the different kinds of surface mentions that the tree-like entities [10] take in the text. Our priority, however, was to demarcate the boundaries of the entity mentions in the Nabokov corpus, i.e., the exact phrase, section, page number, and book where they occur, in a structured way (according to TAV vocabulary). To do so, we gathered a list of the common names of trees (e.g., sallow, willow, baobab, pine, maidenhair, ginkgo, oak, juniper, Judas tree, etc.) and performed iterative searches (and filtered out false matches) over the corpus. This approach corresponds to the dictionary or gazetteer approach [31] in NLP entailing a semi-automated workflow where the matches are verified manually.

It should be noted here that in the botanical sense, "trees" are not stable, classifiable units within the plant kingdom [32]. Although the concept of "woodiness" is closely linked with the idea of "true trees" [33], monocots (such as the bamboo, palm, banana) are also considered among the arborescent members [32]. They are also sub-divided in the Plant kingdom on the basis of height (min. of 2 mts.) with emphasis laid on a single or multi-stem plant. In the world of Nabokov, these have their own thematic relevance to the design and unfolding of the plot, hence are worthy of inclusion. In developing a list of commonly occurring tree species in Europe and America (as the majority of tree species found in Nabokov are found in these locales), the following three resources were consulted:

- Wikipedia's Lists of Trees[6]

- Appendix from Stephen Blackwell's *Nabokov's Secret Trees* [10]

---

[3] https://list.worldfloraonline.org
[4] https://powo.science.kew.org
[5] https://www.bgci.org/resources/bgci-databases/globaltree-portal
[6] https://wikipedia.org/wiki/List_of_trees_and_shrubs_by_taxonomic_family

- Index from Colin Tudge's *The Secret Life of Trees* [22]

The entity matching across the Nabokovian corpus was done through a simple Python program with a reference list at hand. The list mostly contains root words of different tree species ("larch", "pine", "oak", etc.) which encourages the maximum number of matches (see Appendix A). Two expert annotators refined, verified and validated the data (i.e., the authors of this paper) over the course of several months. The metadata details as well as the textual integrity were corroborated through Calibre[7]. The primary rationale for this brute-force approach is that the tree entities appear with a textual undercurrent, where meanings are translucent rather than transparent: Nabokov's strategy aims for overall thematic coherence through a "network of subtextual, cross-linguistic, and transcultural echoes" [34]. Let us illustrate such subtext with some examples.

1. Certain trees appear with a multilingual echo, such as the alder tree (Q156904) in Nabokov's Glory [35] through a village-name "Olkhovo" where the protagonist's mother is said to be born. Similarly, the place "Eichenberg" [36] etymologically contains a reference to the oak tree (Q165145) as does the more explicit "Unter den Linden" [37] to linden trees (Q163760). Other trees camouflaged under the context of place-names include but are not limited to "Larchdell Road" [38] "Pinedale" [39, 11, 40], "Ponderosa Lodge" [39], among others.

2. Nabokov has a penchant for using obscure synonyms to suit the rhythmic needs of his prose, to evoke a secondary meaning as well as delay recognition of a connection, an identity, a pattern in his verbal texture. Often, these are multilingual puns (Nabokov was trilingual). For example, the ginkgo tree (Q43284) appears under the surface mentions of "kew tree" [19], "Chinese tree", "maidenhair tree", "L'arbre aux quarantes écus d'or" [40]. Other disguised entity co-references include "balanic plum" for plum tree [19], "baguenaudier" (French for bladder senna tree, Q158890) [40], "myrrherabol brush" for the myrrh tree (Q164609), [40], etc.

3. Finally, these entity co-references are frequent, especially when some anatomical aspect acts as a metonym for the tree entity, such as "he brushed his fingertips against the wet needles", where "fir trees" are the main referent in this paragraph [37]. Other examples include "under the translucent tree" for "linden tree" [37], "powerful shower started to drum upon the liriodendron and imperialis leaves outside" for "tulip tree" and "paulownia tree" respectively [40], etc.

Our second task was entity linking, which comes with its own set of problems. It is not straightforward to correctly identify the tree species Nabokov had in his mind based on the text. However, as per his own guidelines, a close reader must make an educated guess as to what a particular word or term (denoting an entity) "really means in its contextual habitat, within the terms of a certain imagined place and in the light of a certain literary device" [18]. To do so, we check the distribution of a particular tree species with respect to the location mentioned in the text, and express our confidence level with the SKOS vocabulary [29]. Some examples are given below:

Ember was trying to recall the American name for a similar kind of fir tree in the Rocky Mountains. Two things happened together: Ember said "Douglas" and a dazzled doe plunged into the blaze of our lights.
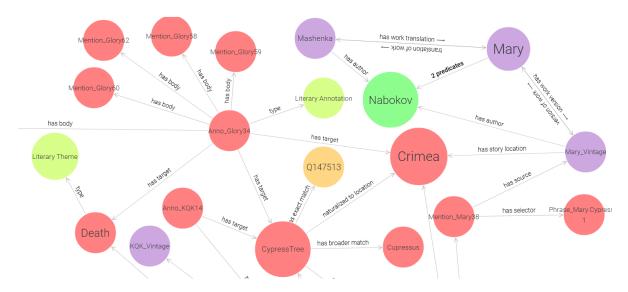
Bend Sinister [41], Ch. 17, p. 226

It is clear that the referent here is the Douglas fir, or more specifically, *Pseudotsuga menziesii var. glauca* (Q956705), which is asserted in our knowledge base as `skos:exactMatch`.

The trees seemed to be a botanist's delirium: a white rowan with alabaster berries or a birch with red bark!

The Gift [37], Ch. 2, p. 121

---

[7]https://calibre-ebook.com/download

Figure 2: A snippet from the visualization of the Tree-KG. The size of the nodes represent their RDF ranks, i.e., the number of incoming links for the most popular entities. Note that the integrated data (and thereby the RDF Rank) reflects three of Nabokov's novels.

There are two entities mentioned here, a rowan tree (`skos:broadMatch` at the genus level would be *Sorbus*) and a birch tree (`skos:broadMatch` would be *Betula*). Since the setting is in the Tibet/Western China region, we put forward the tree species (expressed as a `skos:closeMatch`) as *Sorbus oligodonta* (`Q1037196`) and *Betula albosinensis* (`Q1631520`), based on the range and habitat of the species.

> ... *thumbed a ride with a Mexican truck, found a suitable gulch in the chaparral and there, after writing a short note, began placidly eating from her cupped palm the multicolored contents of her handbag...*

Ada [40], Pt. 1, Ch. 3, p. 28

In this passage, the vegetation-related entities are the "gulch in the chaparral" which seems unremarkable in itself, except for its euphonic qualities. However, we know from our efforts, that the scrub oak appears at key places in the novel (Pt. 1, Ch. 42; Pt. 4 Ch. 1) and is thematically significant[8] [10]. Now, if we look at the etymology of the word "chaparral", we see it comes from the Spanish word "chaparro", meaning the "place of the scrub oak". Hence, the allusion is surely towards the scrub oak (`Q1017264`) and should be annotated as such. Such examples hopefully illustrate the challenges regarding entity linking and how it depends on the background information.

Data were initially collected in tabular format for convenience and later converted to RDF through Protégé's Cellfie plugin [42], which involves the writing of some rules, whereby each row in a spreadsheet is converted into the instances of a class and property values from an ontology. So far, we have integrated data from three of Nabokov novels ([27], [43], [35]) in our Tree-KG. A visual representation of the Tree-KG obtained through GraphDB[9] is shown in Figure 2.

5. Experiments for Entity Recognition

Entity recognition and linking are fundamental tasks for Information Extraction. Different tools have been developed for these tasks, such as DBPedia Spotlight [44], OpenTapioca [45], ANNIE [46] etc. With the rapid advent and acceptance of large language models, systems such as ChatGPT[10], Gemini[11],

Table 1
Evaluation of Entity Recognition Capabilities

System	Precision	Recall	F1 Value
DBPedia Spotlight		75.9%	
ChatGPT	98.5%	89.6%	93.9%
Gemini	97.8%	88.3%	92.8%
Llama	98.6%	92.2%	95.3%

Llama[12] are also being put to use for information extraction. Since our dataset has been patiently and carefully annotated by human hands – and given the challenges of entity recognition and linking in this literary setting – our Tree-KG can serve as a challenging gold dataset to benchmark some of the existing tools and systems. We describe such a sample evaluation for illustrative purposes.

In the experiment, we selected DBPedia Spotlight and OpenTapioca as traditional entity linking systems and three LLM models, GPT-4, Llama 3.1, and Gemini 1.5 Pro, as their alternatives to compare their capabilities. To evaluate them, we selected 100 passages from the Nabokov corpus that contained references to "tree entities" and 100 more passages that contained no references. It should be noted that we focused our tests on entity recognition, because of the challenges mentioned in the previous section. Table 1 shows the Precision, Recall, and F1 values for the systems mentioned here.

OpenTapioca could not retrieve the relevant results from any of the passages from the novels, because of its focus on mentions of humans, organizations, and locations in their Wikidata dump [45]; thus the system was excluded from the results. We set the confidence of DBPedia Spotlight to 0.2, because we wished to maximize the recall results for the tree-related entities, as the noise can be easily filtered out (removing non-tree-related entities via DBpedia). We cannot report the Precision value for DBPedia Spotlight as it cannot be set to identify *only* tree-related entities. Even with post-filtering, the recall value for DBPedia Spotlight was not good enough for our purposes. It can be seen that the LLM models outperformed the traditional alternatives by a substantial margin, with the best performing model being Llama 3.1. The dataset comprising the passages and results are available from Zenodo (see Appendix A).

6. Future Work

Our primary goal in gathering tree-related data is to showcase the remarkable variety of trees found in Nabokov. We hope to provoke interest not only among the readers, fans, and scholars of Nabokov but also in the broader community of tree lovers. It is inspired by the vital roles of trees in our ecosystem, which include their various practical benefits, their physical aesthetics, their diversity, and their ecological complexity. In other words, although our dataset is fairly large, it remains open to a community-wide participation and discussion[13] which would be essential to generate further annotations, corrections, links and a general terminological consensus in a Wiki-style knowledge base[14].

Some other directions suggested by the present work are:

- An interactive website making the annotated data openly accessible. It would include a SPARQL endpoint alongside documentation that would be intuitive to use for the general audience. A sample SPARQL query[15] that links the different facets of Nabokov's trees is shown in Figure 3.

- Other tree-related data that were noted but not collected, e.g., trees being food plants of other organisms, such as birds [47], lepidoptera [8], etc., that have their own thematic importance in Nabokov. A fuller integration that links the different mereological and metonymical aspects of a tree remains to be achieved. References to artworks through trees [48, 10] require further

---

[12]https://www.meta.ai
[13]https://thenabokovian.org/
[14]https://wikiba.se/
[15]https://www.w3.org/TR/sparql11-query/

```
SELECT ?Novel ?Chapter ?Page ?Pref ?Exact ?Place ?Theme ?Wiki WHERE {
  ?Annotation tav:target tav:FirTree , ?Theme , ?Place ; tav:body ?Mention .
  ?Theme a tav:LiteraryTheme .          ?Place a tav:Country .
  tav:FirTree skos:closeMatch  ?Wiki . ?Phrase tav:occursIn ?Section .
  ?Mention tav:source ?Novel ; tav:selector ?Phrase  .
  ?Phrase tav:exactText   ?Exact .
  OPTIONAL { ?Phrase tav:prefixText ?Pref . }
  ?Section tav:chapterLabel ?Chapter ; tav:pageNo ?Page .
}
```

?Novel	?Chapter	?Page	?Pref	?Exact	?Place	?Theme	?Wiki
tav:KQK_Vintage	Ch. 7	139	luxuriant	fir tree	tav:Germany	tav:Magic	wd:Q146992
tav:KQK_Vintage	Ch. 7	136		Christmas tree	tav:Germany	tav:Magic	wd:Q146992
...	...	...	...	...	...	...	...

Figure 3: A SPARQL query showing the exact occurrences of Fir Tree in three of Nabokov's novels alongside their attached literary theme as well as the places mentioned in the text where they occur endemically.

expansion. Community engagement can further enrich the Tree-KG by establishing links to different art forms and further integrating with the Linked Open Data (LOD) Cloud.

- A network analysis of trees alongside other networks, i.e., characters [49], places, motifs, etc. Such force-directed graphs may require more data, in which case we could plan to collect and integrate mentions of characters, themes, places, etc., as they occur and interact across the pages of a Nabokovian text.

7. Conclusion

Nabokov planted trees in his works with great care and deliberate attention. Certain tree species often stand out in a particular work, acting as a kind of structuring background for that particular work, carrying a specific strand of meaning [10]. With this dedicated knowledge graph, we bring to light those rich connections that potentially offer deeper insights into his work. We do so by capturing the different mentions of trees in his corpus, annotating their thematic meaning, noting down their range and distribution across different locales mentioned in the text (e.g., Berlin, St. Petersburg, Crimea, New York, etc.) as well as their functional aspects (trees transformed into wooden objects). Often, there is an added layer of complexity, when those trees appear through "frames", as literary or mythological allusions or as tree metonyms ("olive branch", "laurel tree", "bay leaf", "apple", etc.), but knowledge graphs are well suited to map those meanings.

We developed an ontology model to describe these annotations building upon existing vocabularies and standards. Challenges were encountered in defining the tree-like entities (and their metonyms), spotting them in his corpus, and identifying the best match for the botanical unit at the genus and species level. We used these annotations to evaluate a number of tools for entity recognition in this literary setting, including traditional entity linking tools and large language models. Some of the resources stemming from this effort can be extended and reused in other contexts.

Nabokov famously said "If the mind were constructed on optical lines and if a book could be read in the same way as a painting is taken in by the eye, that is without the bother of working from left to right and without the absurdity of beginnings and ends, this would be the ideal way of appreciating a novel, for thus the author saw it at the moment of its conception" [23]. We hope to remain true to his dictum with this Tree-KG.

References

[1] R. Eldridge, An Introduction to the Philosophy of Art, Cambridge Introductions to Philosophy, 2 ed., Cambridge University Press, Cambridge, 2014. doi:10.1017/CBO9781107300538.

[2] Y. Aikhenvald, Silhouettes of Russian Writers: Literary and Philosophical Essays of Yuli Aikhenvald, Academic Studies Press, Boston, 2024.

[3] D. B. Johnson, Worlds in Regression: Some Novels of Vladimir Nabokov, Ardis, Ann Arbor, 1985.

[4] P. Meyer, Find What The Sailor Has Hidden: Vladimir Nabokov's Pale Fire, Wesleyan University Press, Connecticut, 1988.

[5] V. E. Alexandrov, Nabokov's Otherworld, Princeton University Press, Princeton, 1991.

[6] B. Boyd, Nabokov's "Pale Fire": The Magic of Artistic Discovery, Princeton University Press, Princeton, 1999. URL: https://www.jstor.org/stable/j.ctt7t9zq.

[7] B. Boyd, Nabokov's Ada: The Place of Consciousness, Cybereditions, Christchurch, 2001.

[8] D. E. Zimmer, A Guide to Nabokov's Butterflies and Moths, vollständige web-ausgabe ed., 2012. URL: http://www.d-e-zimmer.de/eGuide/PageOne.htm.

[9] D. Dragunoiu, Vladimir Nabokov and the Art of Moral Acts, Northwestern University Press, Evanston, 2021. URL: https://doi.org/10.2307/j.ctv1v090kw.

[10] S. H. Blackwell, Nabokov's Secret Trees, University of Toronto Press, Toronto, 2024. URL: https://doi.org/10.3138/9781487554446.

[11] V. V. Nabokov, Pale Fire, international ed., Vintage Books, New York, 1989.

[12] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutiérrez, S. Kirrane, J. E. Labra Gayo, R. Navigli, S. Neumaier, A.-C. Ngonga Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. F. Sequeda, S. Staab, A. Zimmermann, Knowledge Graphs, number 22 in Synthesis Lectures on Data, Semantics, and Knowledge, Springer, 2021. URL: https://kgbook.org/. doi:10.2200/S01125ED1V01Y202109DSK022.

[13] K. Lee (Ed.), What Is a Semantic Annotation?, Studies in Natural Language Processing, Cambridge University Press, Cambridge, 2023. doi:10.1017/9781108884532.004.

[14] J. B. Rowan, True to Life: A Study of Lifelikeness in Fiction through Proust, Austen, Nabokov, and Joyce, PhD Thesis, University of California, Berkeley, 2014.

[15] S. H. Blackwell, Reflections on (and of) Trees in Nabokov, in: B. Boyd, M. Bozovic (Eds.), Nabokov Upside Down, Northwestern University Press, Evanston, 2017, pp. 21–36.

[16] S. H. Blackwell, K. Johnson, Fine Lines: Vladimir Nabokov's Scientific Art, Yale University Press, New Haven, 2016. URL: https://doi.org/10.12987/9780300220582. doi:10.12987/9780300220582.

[17] V. V. Nabokov, Speak, Memory: An Autobiography Revisited, international ed., Vintage Books, New York, 1989.

[18] A. Pushkin, V. Nabokov, Eugune Onegin, revised ed., Princeton University Press, Princeton, 1975.

[19] V. Nabokov, Transparent Things, international ed., Vintage Books, New York, 1989.

[20] V. Nabokov, The Stories of Vladimir Nabokov, international ed., Vintage Books, New York, 2012.

[21] V. Nabokov, Invitation to a Beheading, international ed., Vintage Books, New York, 1989.

[22] C. Tudge, The Secret Life of Trees, Allen Lane, London, 2005.

[23] V. Nabokov, Lectures on Literature, Harcourt Brace Jovanovich/Bruccoli Clark, New York, 1980.

[24] R. Alter, The Pleasures of Reading in an Ideological Age, Simon & Schuster, New York, 1989.

[25] R. Sanderson, P. Ciccarese, H. Van de Sompel, Designing the W3C open annotation data model, in: Proceedings of the 5th Annual ACM Web Science Conference, WebSci '13, Association for Computing Machinery, New York, 2013, pp. 366–375. URL: https://doi.org/10.1145/2464464.2464474. doi:10.1145/2464464.2464474.

[26] E. F. Kendall, D. L. McGuinness, Ontology Engineering, Synthesis Lectures on Data, Semantics, and Knowledge, Springer International Publishing, Cham, 2019. URL: https://link.springer.com/10.1007/978-3-031-79486-5. doi:10.1007/978-3-031-79486-5.

[27] V. Nabokov, Mary, international ed., Vintage Books, New York, 1989.

[28] B. Boyd, "Words, Works and Worlds in Joyce and Nabokov: Or Intertextuality, Intratextuality, Supratextuality, Infratextuality, Extratextuality and Autotextuality in Modernist and Prepost-

modernist Narrative Discourse", Cycnos 12 (1995) 3–12. URL: http://epi-revel.univ-cotedazur.fr/publication/item/422.

[29] A. Miles, S. Bechhofer, SKOS Simple Knowledge Organization System Namespace Document, W3C Recommendation, 2009. https://www.w3.org/2009/08/skos-reference/skos.html.

[30] M. A. Musen, The protégé project: a look back and a look forward, AI Matters 1 (2015) 4–12. doi:10.1145/2757001.2757003.

[31] J. L. Martinez-Rodriguez, A. Hogan, I. Lopez-Arevalo, Information extraction meets the Semantic Web: A survey, Semant. Web 11 (2020) 255–335. doi:10.3233/SW-180333.

[32] R. Ehrenberg, What Makes a Tree a Tree?, Knowable Magazine (2018). doi:https://doi.org/10.1146/knowable-033018-032602.

[33] A. T. Groover, What genes make a tree a tree?, Trends in Plant Science 10 (2005) 210–214. URL: https://www.cell.com/trends/plant-science/abstract/S1360-1385(05)00053-1. doi:10.1016/j.tplants.2005.03.001.

[34] P. Tammi, Russian Subtexts in Nabokov's Fiction: Four Essays, Tampere University Press, Tampere, 1999.

[35] V. Nabokov, Glory, international ed., Vintage Books, New York, 1991.

[36] V. Nabokov, Despair, international ed., Vintage Books, New York, 1989.

[37] V. Nabokov, The Gift, international ed., Vintage Books, New York, 1991.

[38] V. Nabokov, Look at the Harlequins!, international ed., Vintage, New York, 1990.

[39] V. Nabokov, Lolita, international ed., Vintage Books, New York, 1989.

[40] V. Nabokov, Ada or Ardor: A Family Chronicle, international ed., Vintage, New York, 1990.

[41] V. Nabokov, Bend Sinister, international ed., Vintage Books, New York, 1990.

[42] M. J. O'Connor, C. Halaschek-Wiener, M. A. Musen, Mapping Master: A Flexible Approach for Mapping Spreadsheets to OWL, in: P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, I. Horrocks, B. Glimm (Eds.), The Semantic Web – ISWC 2010, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 194–208.

[43] V. Nabokov, King, Queen, Knave, international ed., Vintage Books, New York, 1989.

[44] J. Daiber, M. Jakob, C. Hokamp, P. N. Mendes, Improving efficiency and accuracy in multilingual entity extraction, in: Proceedings of the 9th International Conference on Semantic Systems, I-SEMANTICS '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 121–124. doi:10.1145/2506182.2506198.

[45] A. Delpeuch, OpenTapioca: Lightweight Entity Linking for Wikidata, CoRR abs/1904.09131 (2019). URL: http://arxiv.org/abs/1904.09131.

[46] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications, in: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), 2002.

[47] D. B. Johnson, Nabokov's Aviary in Ada, in: J. Dingley, L. Ferder (Eds.), In the Realm of Slavic Philology: To Honor the Teaching and Scholarship of Dean S. Worth from His UCLA Students, Slavica, Bloomington, 2000, pp. 163–180.

[48] G. de Vries, D. B. Johnson, L. Ashenden, Nabokov and the Art of Painting, Amsterdam University Press, 2006. URL: http://www.jstor.org/stable/j.ctt46n25r.

[49] F. Moretti, Network Theory, Plot Analysis, New Left Review (2011) 80–102.

A. Online Resources

Some of the resources stemming from this work can be accessed from:

- Annotated Dataset from Zenodo

- TAV Ontology Documentation