

**TRABAJO FIN DE MÁSTER**

**2018**

**Generación automática de itinerarios  
basados en LOD**

**ALICIA LARA CLARES**

**MÁSTER UNIVERSITARIO EN LENGUAJES Y  
SISTEMAS INFORMÁTICOS**

**UNED**

**Directora: Dra. ANA GARCÍA SERRANO**



# Resumen

El trabajo de los guías turísticos no es sólo describir lo que quieren mostrar en cada momento de la visita guiada. Las relaciones, las historias, los cuentos y las anécdotas que envuelven aquello en que se centran se convierten, a menudo, en el motivo por el que las personas prefieren que un experto les cuente lo que van a ver desde su punto de vista. Hoy en día, es posible encontrar en Internet la lista de qué visitar al llegar a un lugar, e incluso información detallada del mismo. Pero no es tan fácil recrear el valor que aporta una persona experta, que sea capaz ya no sólo de encontrar el modo de crear interés en las personas, sino también de contestar sus dudas, relacionar las historias y saber cómo contarlas.

El objetivo de este trabajo consiste en acercar a los usuarios una forma de realizar visitas turísticas basada en temáticas con descripciones variadas de los elementos concretos de la visita. Para ello, se presenta un nuevo modelo narrativo para la generación automática de itinerarios, que organiza de manera coherente información recuperada de diferentes fuentes de datos en abierto (*Linked Open Data*, LOD), y la modela de forma coherente en itinerarios. Además, se ha creado un prototipo que permite tanto la supervisión de los itinerarios por parte de los expertos, como la navegación por parte de los usuarios.

Para validar el modelo propuesto se ha seleccionado como caso de estudio el Museo del Prado. Se han generado automáticamente 581 itinerarios que, a su vez, están formados por 4823 obras, 1504 personajes, 14166 referencias y 25569 eventos. Los usuarios podrán visualizar estos itinerarios, para profundizar, o no, en la información detallada que contienen.

Palabras clave: Humanidades Digitales, modelos narrativos, datos enlazados en abierto, extracción de información.



# Abstract

The work of tour guides is not only to describe what they want to show at each moment of the guided tour. The relationships, stories and anecdotes that surround what guides focus on often become the reason why people prefer an expert to tell them what they see from their point of view. Nowadays, it is possible to find on the Internet the list of what to visit when you arrive at a place, and even detailed information about it. However, it is not so easy to recreate the value that an expert person brings, someone who is able not only to find the way to create interest in people, but also to answer their questions, connect stories and know how to tell them.

The aim of this work is to bring users closer to a way of making tourist visits based on topics with varied descriptions of the specific elements of the visit. To this end, a new narrative model is presented for the automatic generation of itineraries. It coherently organizes information retrieved from different open data sources (Linked Open Data, LOD), and models it in itineraries. In addition, a prototype has been created to allow both expert supervision of the routes and user navigation.

The Prado Museum has been selected as a case study to validate the proposed model: 581 itineraries have been automatically generated, which in turn consist of 4823 works, 1504 characters, 14166 references and 25569 events. Users will be able to view these itineraries, and to deepen, or not, in the detailed information they contain.

Keywords: Digital Humanities, narrative models, Linked Open Data, Information Extraction.



# ÍNDICE

---

1	Introducción.....	12
1.1	Objetivos y propuesta.....	12
1.2	Caso de estudio: Museo del Prado .....	14
1.3	Estructura del documento .....	15
2	Estado del arte .....	16
2.1	Aproximación teórica.....	16
2.2	Aproximación computacional .....	20
2.2.1	Tecnologías y estándares relacionados con la web. ....	22
2.2.2	Datos enlazados en abierto en la Web. ....	24
2.2.3	Estándares de metadatos en el ámbito de las Humanidades Digitales.....	24
2.2.4	Modelos de organización del conocimiento asociados a las narrativas. ....	29
2.3	Herramientas y recursos .....	41
3	Creación del Modelo Narrativo.....	44
3.1	Elementos narrativos .....	45
3.2	Trama .....	45
3.3	Narrativa .....	45
3.4	Discurso narrativo .....	46
3.5	Sistema de diálogo .....	46
3.6	Itinerario .....	46
4	Extracción y procesamiento de la información.....	48
4.1	Extracción de información desde distintas fuentes.....	48
4.2	Procesamiento e integración de metadatos y textos descriptivos desde distintas fuentes	54
4.2.1	Enriquecimiento automático de la red conceptual utilizando las herramientas LOD presentes en Wikipedia y Wikidata .....	57
4.2.2	Enriquecimiento automático de la red conceptual a partir de las obras del Museo del Prado	60
4.2.3	Enriquecimiento automático de la red conceptual integrando los autores del Museo del Prado con sus datos en Wikipedia y Wikidata .....	62
4.2.4	Descubrimiento de información automática de las descripciones de las obras del Museo del Prado	63
4.2.5	Resultados obtenidos de los puntos anteriores .....	67
4.3	Segmentación de los textos obtenidos e identificación de elementos narrativos.....	69
5	Creación de la red conceptual y selección de tramas.....	72
5.1	Creación de la red. Selección de nodos y aristas. ....	72
5.2	Selección de tramas asociadas a un conjunto de obras .....	74

5.3	Descomposición del grafo.....	76
5.3.1	Primer grafo. Segmentos y elementos narrativos. ....	76
5.3.2	Segundo grafo. Elementos narrativos y entidades de Wikidata.....	77
5.4	Filtrado de información.....	78
5.4.1	Aproximación 1. Cálculo de umbral no supervisado.....	80
5.4.2	Segunda aproximación. Cálculo del umbral con dos iteraciones. ....	80
5.4.3	Tercera aproximación. Cálculo del umbral basado en el número de obras. ....	81
5.5	Selección de tramas. ....	82
5.5.1	Selección de tramas a partir de elementos narrativos. ....	83
5.5.2	Resultados obtenidos.....	86
6	Evaluación .....	88
6.1	Diseño e implementación del prototipo para la visualización de itinerarios .....	88
6.2	Evaluación con usuarios.....	88
6.2.1	Preguntas cerradas. Evaluación de la experiencia de usuario .....	89
6.2.2	Preguntas abiertas. Obtención de valoraciones y sugerencias. ....	91
6.2.3	Estudio de comportamiento .....	91
6.3	Resultados obtenidos.....	91
6.3.1	Análisis de las encuestas .....	91
6.3.2	Análisis de comportamiento en logs.....	100
7	Conclusiones y trabajo futuro.....	102
	Publicaciones .....	104
	Referencias.....	106
	Anexos.....	112
	Anexo I .....	114
	Anexo II .....	137
	Anexo III .....	139



# Índice de ilustraciones

Ilustración 1. Marco narratológico planteado en Jahn [19].....	18
Ilustración 2. Imagen extraída de ficcioninteractiva.com correspondiente al juego Capitán Pirata....	22
Ilustración 3. Modelo semántico extraído de la Web del Museo del Prado .....	26
Ilustración 4. Imagen que representa los distintos nombres de artista Diego Velázquez.....	28
Ilustración 5. Estructura de clases y propiedades de la ontología Curate extraído de [70] .....	29
Ilustración 6. Imagen extraída de la Web de descarga de Storyspace [4] .....	30
Ilustración 7. Ejemplo de objeto extraído de la versión 7 de Storyscope .....	31
Ilustración 8. Ejemplo de historia extraído de la versión 7 de Storyscope .....	32
Ilustración 9. Ejemplo de evento extraído de la versión 7 de Storyscope.....	33
Ilustración 10. Ejemplo del contenido de un dossier extraído de la versión 7 de Storyscope .....	34
Ilustración 11. Ejemplo de una referencia extraída de la versión 7 de Storyscope.....	35
Ilustración 12. Ejemplo de narrativa extraída de la versión 7 de Storyscope.....	36
Ilustración 13. Ejemplo de agentes extraídos de Storyscope 7 .....	37
Ilustración 14. Ejemplo de evento con varios agentes extraído de Storyscope 7 .....	37
Ilustración 15. Ejemplo de micrositio extraído de la versión 7 de Storyscope .....	38
Ilustración 16. Representación de eventos y hechos relativos a Claude Monet[3] .....	39
Ilustración 17. Temas desde las etiquetas de historia de Claude Monet y Paul Cézanne [3] .....	40
Ilustración 18. Modelo narrativo creado para este trabajo.....	44
Ilustración 19. Consulta SPARQL para obtener las obras del museo.....	48
Ilustración 20. Consulta SPARQL para obtener la información de una obra del museo .....	49
Ilustración 21. Ilustración que muestra los datos obtenidos de la obra "El Lavatorio" desde la web del Museo del Prado .....	53
Ilustración 22. Ilustración que muestra la información obtenida desde la web del Museo del Prado referente al autor "Jacopo Robusti Tintoretto" .....	54
Ilustración 23. Captura extraída de Wikipedia sobre los eventos asociados al año 1423.....	57
Ilustración 24. Imagen extraída de Wikipedia sobre el cuadro El Lavatorio, de Tintoretto .....	58
Ilustración 25. Consulta SPARQL para obtener la información sobre guerras de Wikidata .....	58
Ilustración 26. Lista de guerras existentes en Wikidata .....	59
Ilustración 27. Lista de nacimientos y fallecimientos de personajes.....	59
Ilustración 28. Pseudocódigo que muestra el proceso de búsqueda de nuevas relaciones entre conceptos previamente desconocidos .....	60
Ilustración 29. Resultados extraídos de la consulta mostrada anteriormente.....	61
Ilustración 30. Imagen extraída de Wikidata referente al concepto "San Marcoula" .....	61
Ilustración 31. Imagen extraída de Wikipedia referente al concepto "San Marcoula" .....	62
Ilustración 32. Extracción de nombres de autores de la Web del Museo del Prado a través de Solr ..	62
Ilustración 33. Búsqueda de un concepto en Wikidata y Wikipedia .....	63
Ilustración 34. Información resaltada en los textos del Museo del Prado .....	64
Ilustración 35. Información anotada de los textos del Museo del Prado .....	64
Ilustración 36. Información anotada de los textos del Museo del Prado .....	65
Ilustración 37. Entidades nombradas existentes en una descripción.....	66
Ilustración 38. Texto anotado del Museo del Prado.....	67
Ilustración 39. Texto anotado del Museo del Prado.....	67
Ilustración 40. Texto anotado del Museo del Prado sobre "Los Borrachos" de Velázquez.....	68
Ilustración 41. Texto anotado de Wikipedia sobre "Los Borrachos" de Velázquez .....	68

Ilustración 42. Índice extraído de Wikipedia que muestra las distintas secciones textuales del Parque del Retiro.....	69
Ilustración 43. Almacenamiento en Solr de textos de Wikipedia .....	69
Ilustración 44. Segmentación de textos de distintas obras referentes a “Venecia” .....	70
Ilustración 45. Captura de la red conceptual generada en este trabajo. ....	73
Ilustración 46. Captura de la red conceptual generada en este trabajo. ....	73
Ilustración 47. Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo .....	74
Ilustración 48. Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo .....	75
Ilustración 49. Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo .....	75
Ilustración 50. Captura del grafo correspondiente con los elementos narrativos existentes en un segmento. ....	76
Ilustración 51. Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo .....	77
Ilustración 52. Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo .....	78
Ilustración 53. Gráfica que representa el DF (frecuencia de elementos narrativos / segmentos) de la información obtenida .....	79
Ilustración 54. Gráfica que representa el DF (frecuencia de elementos narrativos / segmentos) de la información obtenida aplicando los filtros. ....	81
Ilustración 55. Gráfica que representa el DF (frecuencia de elementos narrativos / obras) de la información obtenida aplicando los filtros. ....	82
Ilustración 57. Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo correspondiente a Luis Tristán.....	84
Ilustración 56. Valores de similitud correspondientes a un conjunto de segmentos.....	84
Ilustración 58. Segmentos seleccionados para un itinerario. ....	86
Ilustración 59. Creación final de un itinerario en HTML. ....	86
Ilustración 60. Definición del test Alfa de Cronbach.....	90
Ilustración 61. Criterios de calidad contemplados en el cuestionario.....	90
Ilustración 62. Captura de la página de MPOC con el resumen de resultados de las encuestas .....	92
Ilustración 63. Resultados de las encuestas por categorías de escalas .....	93
Ilustración 64. Resultados de las encuestas por escalas.....	94
Ilustración 65. Resultados finales de la encuesta .....	97
Ilustración 66. Grafo correspondiente a la primera pregunta de la encuesta en MPOC .....	98
Ilustración 67. Grafo correspondiente a la segunda pregunta de la encuesta en MPOC.....	98
Ilustración 68. Grafo correspondiente a la primera pregunta de la encuesta en MPOC .....	99
Ilustración 69. Imagen extraída de los logs correspondiente a MPOC.....	100

# Índice de tablas

Tabla 1. Número de elementos obtenidos de cada tipo en el proceso de extracción de información	52
Tabla 2. Tipos de colecciones de datos obtenidos en la etapa de extracción de información en Solr	52
Tabla 3. Fuentes desde las que se extrae información referente a " Jacopo Robusti Tintoretto" .....	54
Tabla 4. Enlaces obtenidos entre dos autores distintos entre Wikipedia y el Museo del Prado .....	63
Tabla 5. Primera aproximación del cálculo del umbral no supervisado. ....	80
Tabla 6. Segunda aproximación del cálculo del umbral no supervisado. ....	80
Tabla 7. Filtrado final de la información .....	82
Tabla 8. Resultados de la encuesta realizada para MPOC .....	94
Tabla 9. Resultados de consistencia obtenidos para cada escala.....	95
Tabla 10. Cálculo de la consistencia general de la encuesta .....	95
Tabla 11. Desviación típica e intervalos de confianza para cada escala .....	96
Tabla 12. Tamaño de la muestra estimado para obtener ciertos valores de precisión y probabilidad de error. ....	96



# 1 INTRODUCCIÓN

---

Hoy en día, gracias a las fuentes de recursos web existentes, es posible obtener información acerca de casi cualquier lugar o monumento antes de ir a visitarlo. Pese a esto, es imposible recrear los beneficios que aportan los guías turísticos a las visitas, ya que un guía turístico no sólo ofrece esta información estática, sino también su contexto, las relaciones que existen entre los objetos de la visita y las anécdotas que surgen en relación con éstos. En este trabajo se propone un modelo de representación capaz de aportar ese dinamismo a la información existente en recursos enlazados en abierto en la Web (*Linked Open Data*, LOD) que permita establecer relaciones y aporte una nueva forma de preparar una visita turística, tanto presencial como virtual.

Mulholland et al. (2016) [1] propuso un modelo narrativo como base para establecer el contexto asociado a las obras de los museos, permitiendo a los usuarios compartir su interpretación personal y estableciendo relaciones entre la obra de arte y sus propias preocupaciones, conocimientos e intereses. Este modelo teórico permite a los usuarios crear visitas personalizadas por los museos utilizando las narrativas como herramienta. El presente trabajo propone una mejora de este modelo haciendo uso de la información textual proveniente de las descripciones asociadas a los elementos que se quieren conocer. Estas descripciones contienen, a su vez, información semántica asociada (como personajes o eventos), lo que permite relacionar los distintos objetos de la visita entre sí. Además, este trabajo presenta una herramienta y un prototipo que permite al experto crear nuevos itinerarios o visitas a partir de esta información existente.

## 1.1 OBJETIVOS Y PROPUESTA

El objetivo principal de este trabajo es proponer un modelo narrativo basado en grafos y desarrollar un prototipo demostrativo para generar itinerarios a partir de sus componentes. Un itinerario se puede ver como un conjunto de narraciones acerca de los elementos que lo conforman, que comparten un tema en común.

Para llevarlo a cabo, se plantean tres objetivos:

- (1) creación de un modelo narrativo que sirva como estructura de datos para almacenar y trabajar con la información desde las distintas fuentes de recursos. Un modelo narrativo establece las relaciones entre los conceptos involucrados en todo el proceso de la creación y visualización de los itinerarios asociados a una visita turística.
- (2) extracción de la información desde las fuentes de recursos en abierto, así como la anotación automática de las narrativas y su integración en el modelo planteado anteriormente.
- (3) diseño e implementación de un sitio web para la visualización y evaluación de los itinerarios.

Además de plantear un sistema capaz de generar itinerarios de forma automática, se hace uso de herramientas y recursos en abierto, aprovechando la información que ya existe, tanto en forma de metadatos como en forma textual, para extraer, organizar, seleccionar y publicar los itinerarios clasificándolos por temáticas. Por ello, la información se obtiene de fuentes como Wikidata y Wikipedia.

El proceso de extracción de información se basa en analizar el lenguaje natural para extraer elementos de información. Parte de textos (o expresiones) y produce datos de salida no ambiguos que pueden ser directamente mostrados al usuario final, almacenados en una base de datos, analizados o usados para indexar en motores de búsqueda. La diferencia entre extracción de información y recuperación de información es que en el primer caso la aplicación analiza el texto y devuelve los elementos que necesita el usuario; mientras que en el segundo caso el sistema encuentra qué documentos son relevantes y los muestra al usuario [2].

El proceso de descubrimiento de información a partir de la información textual da lugar a la creación de una red de conceptos que se nutre de los datos aportados a partir de cada fuente.

La metodología empleada en este trabajo es la siguiente:

1. Estudiar los distintos modelos narrativos existentes, tanto desde el punto de vista teórico (lingüístico) como computacional. Una vez identificados los conceptos asociados al marco narrativo y su organización, proponer un modelo narrativo que permita aprovechar la información existente y mejorar significativamente el alcance de las herramientas actuales de creación y edición de narrativas.
  - El resultado de esta tarea es el planteamiento, a nivel teórico, de un modelo narrativo para la creación automática de itinerarios.
2. Trasladar el modelo narrativo teórico propuesto a un punto de vista computacional. Esta tarea se subdivide, a su vez, en subtareas:
  - Estudiar cómo se presentan los datos en la web, identificando los distintos formatos y las herramientas que permiten extraer y procesar de forma automática la información. Implementar un sistema que permita extraer la información desde las fuentes seleccionadas, de manera que se obtengan los metadatos asociados a los conceptos de manera organizada. Además, proponer un sistema de anotación de textos acorde al modelo narrativo propuesto al primer objetivo.
    - El resultado de esta tarea es la construcción de una base de datos que contiene, para cada concepto, sus metadatos asociados y los textos anotados siguiendo la estructura propuesta en el modelo narrativo anterior. Además, se obtiene un grafo semántico con todos los conceptos extraídos de las distintas fuentes, organizados también según el modelo narrativo propuesto.
  - Construir el grafo que contenga toda la información obtenida en el punto anterior.
    - El resultado en esta tarea es la construcción de una base de datos que contiene, para cada segmento obtenido, información sobre a qué obra pertenece, de qué fuente de recursos proviene, el número de párrafo, sus metadatos y el texto anotado del mismo.
  - Estudiar distintos métodos para seleccionar las tramas, que conformarán los distintos itinerarios. Una trama puede ser vista a este nivel como una selección de uno o varios temas de los que tratará el itinerario. Implementar un sistema de creación de

itinerarios utilizando el método seleccionado, y exportar la información obtenida siguiendo el modelo narrativo planteado en el primer objetivo.

- El resultado de esta tarea es una colección de documentos organizada por temáticas, donde cada temática es una trama, y cada trama puede representar a uno o varios conceptos del modelo narrativo propuesto.
3. Diseñar e implementar un sistema de información basado en web que permita visualizar los itinerarios generados, además de ofrecer a los expertos la funcionalidad necesaria para crear y editar itinerarios.
  4. Evaluar el sistema creado con usuarios, utilizando un cuestionario que evalúe la experiencia de usuario.
  5. Extraer conclusiones a partir del análisis de la evaluación realizada en el punto anterior.

## 1.2 CASO DE ESTUDIO: MUSEO DEL PRADO

El caso de estudio de este trabajo está enmarcado en el proyecto MUSACCES, museología e integración social: la difusión del patrimonio artístico y cultural del Museo del Prado a colectivos con especial accesibilidad (invidentes, sordos y reclusos). El Consorcio MUSACCES propone acciones de museología e integración social en un entorno de trabajo interdisciplinar. Se busca acercar el Museo del Prado a colectivos de especial accesibilidad (invidentes, personas sordas y reclusos) con apoyo de las Tecnologías de la Información y la Comunicación, así como redefiniendo los recursos tradicionales. Los trabajos de investigación comenzaron el 1 de enero de 2016 con participantes de ámbitos multidisciplinares pertenecientes a grupos de investigación de la Universidad Complutense de Madrid, de la Universidad Nacional de Educación a Distancia (UNED) y de la Universidad Autónoma de Madrid (UAM).

Se ha aplicado el modelo narrativo sobre información de las obras del Museo del Prado para validar su viabilidad. Este dominio se ha seleccionado por dos motivos principales: (1) se ajusta a la situación planteada en la introducción sobre visitas turísticas y (2) otros autores [1,3,4] han trabajado sobre otros museos.

En este dominio, un itinerario está formado por una serie de descripciones o narraciones que corresponden a una o varias obras. La información existente proviene de tres fuentes de recursos: Wikipedia, Wikidata y la web del Museo del Prado<sup>1</sup>. La web del Museo del Prado aporta información sobre las obras (año de creación, autores, etiquetas, procedencia, etc.), además de una descripción textual de la misma. En el caso de Wikipedia, las descripciones de las obras contienen una serie de enlaces que aportan un valor semántico a los mismos. Por ejemplo, en la descripción de la obra “El Jardín de las Delicias” de El Bosco<sup>2</sup> se puede encontrar el siguiente fragmento:

*“A partir de Gibson y hasta la actualidad (Hans Belting, 2002) se ha conjeturado que se realizó para Enrique III de Nassau.<sup>2</sup> Los primeros poseedores de la obra fueron, pues, los miembros de la casa de*

---

<sup>1</sup> <https://www.museodelprado.es/>

<sup>2</sup> [https://es.wikipedia.org/wiki/El\\_jardín\\_de\\_las\\_delicias](https://es.wikipedia.org/wiki/El_jardín_de_las_delicias)

*Nassau, en cuyo palacio de Bruselas pudo ver el cuadro el primer biógrafo del Bosco, Antonio de Beatis, personaje que viajaba en el séquito del cardenal Luis de Aragón, en el año 1517 (...)."*

Como se puede ver, la información se presenta como un texto con una serie de conceptos con enlaces, que contienen información semántica de los mismos. Por ejemplo, se encuentran distintas fechas ("2002" y "1517"), varios personajes ("Enrique III de Nassau", "Antonio de Beatis" y "Luis de Aragón"), un lugar ("Bruselas") y el nombre de una dinastía o familia noble ("Casa de Nasau").

Por otro lado, la web del Museo del Prado también contiene información importante que permite descubrir conceptos previamente desconocidos de los textos. Por ejemplo, a través de las etiquetas asociadas a cada obra. Siguiendo el ejemplo anterior, a partir de la web del museo<sup>3</sup> se pueden encontrar conceptos como "Grisalia", "Antiguo Testamento", "Aves", etc.

Wikidata se utiliza como fuente para obtener información estructurada sobre las obras, autores y elementos relacionados, además de establecer la relación entre las obras de Wikipedia y la web del Museo del Prado.

### 1.3 ESTRUCTURA DEL DOCUMENTO

A continuación, se describe brevemente cada uno de los siguientes capítulos que conforman este documento:

Capítulo 2. Estado del Arte. En este capítulo se revisa el trabajo previo relacionado con modelos narrativos.

Capítulo 3. Creación del modelo narrativo. En este capítulo se describe el modelo narrativo propuesto en este trabajo.

Capítulo 4. Extracción y procesamiento de la información. En este capítulo se describe en detalle todo el proceso de extracción de la información del modelo narrativo.

Capítulo 5. Creación de la red conceptual y selección de tramas. En este capítulo se describe el proceso de creación de la red de conceptos, así como el proceso seguido para generar itinerarios automáticamente a partir de la información que contienen.

Capítulo 6. Evaluación. En este capítulo se describe la evaluación llevada a cabo y se realiza un análisis de los resultados.

Capítulo 7. Conclusiones y trabajo futuro. En este capítulo se enumeran las principales conclusiones obtenidas en este trabajo, así como una serie de líneas de investigación futuras.

---

<sup>3</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/triptico-del-jardin-de-las-delicias/02388242-6d6a-4e9e-a992-e1311eab3609>



## 2 ESTADO DEL ARTE

---

Dada la naturaleza de este trabajo, el estudio del estado del arte se abordará desde dos puntos de vista distintos. En primer lugar, una aproximación teórica o lingüística, que estudie los conceptos asociados a este trabajo y de dónde provienen, además de los distintos modelos narrativos existentes desde un punto de vista teórico. En segundo lugar, una aproximación computacional, que estudia cómo se ha llevado a cabo el traspaso de los modelos narrativos hacia el punto de vista computacional. Por último, se describirán las herramientas y recursos utilizados a lo largo de este trabajo.

### 2.1 APROXIMACIÓN TEÓRICA

El estudio de las distintas teorías lingüísticas, como el Estructuralismo, el Constructivismo o la Pragmática, permiten comprender cómo se organiza el lenguaje para permitir la comunicación entre individuos. El objetivo del trabajo es, en primer lugar, establecer un modelo narrativo que permita la creación, publicación y evaluación del sistema. En la bibliografía hay diferentes aproximaciones y definiciones de conceptos muy similares, por ello mientras se revisa en lo que sigue el estado del arte, se va definiendo y consolidando la forma de entender los elementos que rodean a las narrativas.

La palabra **historia** se utiliza coloquialmente para expresar cualquier representación de eventos, personas y/o cosas. Por ejemplo, una historia podría ser una representación textual explicando la vida del pintor Goya. De esta manera, según Brooks [5], la historia puede ser vista como un sistema de asociaciones entre personas, eventos y/o cosas. Chatman [6] planteó la historia como la selección y organización llevada a cabo por el discurso a través de un medio desde el punto de vista de un nivel abstracto.

Todorov en su libro "*Grammaire du Décaméron*" [7] distingue entre relato e historia. El **relato** es el elemento dominante de la obra en prosa; en éste se desenvuelve la historia, y le concierne el orden cronológico: en la historia se entretajan varios hilos que van rotándose el orden de importancia en la narración, cada uno a su tiempo, "*corresponde a una exposición pragmática de lo que sucedió*".

En 1992, Branigan [8] definió la **narrativa** como "una actividad perceptual cuyo objetivo es el análisis del tiempo y el espacio, la organización y la clasificación de información empírica". Su objetivo era crear una intersección entre los espectadores de las películas y los textos. Para Branigan, la información de la historia es una función de lo que ocurre y cómo se presenta y se da a conocer por los personajes y el espectador. De este modo, la narración es un componente de la historia que trata las discrepancias entre los personajes y el espectador, explicando los cambios argumentativos, introduciendo a los nuevos personajes, etc. Dicho de otro modo, puede verse como la distribución de conocimiento que determina cuándo y cómo el espectador va a adquirirlo a partir del texto. El objetivo final es que la construcción de la narrativa se lleve a cabo a través de la intersección entre el que la crea y quien la visualiza. Así, se puede definir la diferencia entre narrativa y narración. La **narrativa** representa la organización de la información y la forma de expresión para esa organización, es decir, el universo de elementos que conforman una historia. Por otro lado, una **narración** representa una navegación específica por ese universo. En su artículo sobre representación y administración de información narrativa compleja, Zarri [9] distingue entre **narrativa ficticia**, cuyo objetivo es entretener; y **narrativa no ficticia**, a la que atribuye un valor de sentido económico, en el que las personas pagan por obtener la información.

Las narrativas constituyen, desde un punto de vista teórico, el objeto de la disciplina denominada narratología. El término **Narratología** fue acuñado por primera vez por Todorov [7] en 1969 y se refiere tanto al estudio de la narrativa y la estructura de la misma, como a la manera en que ésta afecta a nuestra percepción, aunando conceptos del área de Estructuralismo, que estudia la naturaleza de la historia en su conjunto en lugar de estudiarla como partes de la misma y de la Semiótica o Semiología, que estudia los signos de comunicación entre sociedades humanas y su comportamiento. Por ello, el objetivo de la narratología es producir una descripción en profundidad de las estructuras sintácticas y semánticas de las narrativas.

El perfil experto que maneja las narrativas es el **narratólogo**, y se encarga de dividir las partes de la misma para establecer sus funciones, sus objetivos y las relaciones entre las mismas [9]. El modelo narrativo de Brooks [10] define también el término **agentes** (*story agents*) como una versión en digital de la figura de los cuentacuentos (*storytellers*). Los **cuentacuentos** tienen la capacidad de conectar con los espectadores, dando un valor adicional a la historia, adaptándose a los mismos en cada momento. Genette [11] define al **narrador** como “la voz del discurso narrativo”. Así, el narrador tiene la función de establecer el contacto comunicativo con el destinatario (“*addressee*”, o “*narratee*”), y se encarga de decidir qué contar y cómo contar la historia, además de seleccionar qué partes no contar de la misma.

Una vez definida qué es una narrativa, el siguiente paso es estudiar cómo se construyen.

Bal [12], divide las narrativas, desde un punto de vista estructuralista, en tres capas: el texto, la fábula y la historia. En este caso, una **fábula** es un conjunto de eventos relacionados de forma cronológica, la **historia** es un subconjunto de la fábula que tiene como objetivo informar sobre el contexto de una serie de eventos (por lo que, a partir de una fábula, se pueden obtener una serie de historias). Dicho de otro modo, una historia es una fábula contada desde un punto de vista. El **texto** lo conforman las unidades semióticas, es decir, utiliza una serie de signos de comunicación (un lenguaje) para expresar la historia. Una modificación de esta estructura, desde un punto de vista computacional, la ofrecen Swartjes y Theune [13], que investigan cómo se desarrollarían las historias simulando un mundo virtual habitado por personajes virtuales (denominadas “**narrativas emergentes**”), estudiando el proceso de generación de la historia como una secuencia de eventos emergentes. En este caso, se definen las historias como “*tramas*”, y el texto como “*capa de presentación*”.

Otro modelo narrativo a destacar surge del trabajo de Mulholland [14], que pretende utilizar las narrativas como medio para mejorar el aprendizaje colaborativo y la exploración del legado cultural. Mulholland parte de las definiciones de narrativa e historia de Brooks [5], donde una historia es el espacio conceptual que representa eventos, personas y objetos, y una narrativa es cualquiera de las formas de contar la historia. Esta forma de contar las historias se ajustan a diferentes patrones estructurales que enfatizan distintos aspectos de la historia con el fin de crear un mensaje, tal y como expone Schank [15]. Este autor identifica una serie de patrones denominados “**esqueletos de historia**”, que se categorizan por lo que quieren contar (en lugar de por la estructura narrativa), y que se utilizan para construir distintas narrativas a partir de una historia. Además, caracteriza las historias como la base para la creación de memoria y conocimiento, argumentando que la inteligencia humana se nutre, de alguna manera, de narrativas, partiendo del trabajo “*Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*” [16], que define un “**script**” como una serie de expectativas sobre lo que ocurrirá después en una situación conocida. Así pretende mostrar el modo de aprendizaje a partir de narrativas, creando una base de conocimiento y comprensión de las mismas, basándose en experiencias pasadas (*storytelling* y *story understanding*).

En la guía de la Teoría de la Narrativa, actualizada por Jahn [17] en 2017, la **narrativa** se define como “algo que cuenta o presenta una historia”, donde la **historia** es una secuencia de eventos con personajes involucrados y el **narrador** es quien cuenta la narrativa, “articulando” el texto narrativo. Partiendo de estas tres definiciones básicas, Jahn introduce el concepto de **discurso narrativo** describiéndolo, en primera instancia, como la voz narrativa que creamos de forma automática al leer un texto (donde se crea una entonación, señales de emoción, subjetividad, etc.). Este concepto ya lo introdujeron Genette [10] en 1980 con su ensayo “Discurso Narrativo” para representaciones textuales y Chatman [6] en 1978, “Historia y Discurso”. Genette sugirió diferenciar entre narrativa homodiegética (“*homodiegetic narrative*”) y narrativa heterodiegética (“*heterodiegetic narrative*”), donde “diegética” se refiere a “perteneciente a la narrativa”, “homo” a “de la misma naturaleza” y “hetero” a “de diferente naturaleza”. Así, en una **narrativa homodiegética** la historia la cuenta un narrador que forma parte de la historia y además es uno de los personajes que aparecen en ella. Por otro lado, en una **narrativa heterodiegética** el narrador no está presente en la historia.

Estos conceptos dan lugar a la definición de la técnica de **focalización**. Partiendo del modelo narrativo de Genette [10], la focalización representa la selección (y, por tanto, restricción) de información narrativa relativa a la percepción, conocimiento y punto de vista de una persona. Genette distingue tres tipos distintos. La primera, **no focalización** o **focalización cero**, donde los eventos se cuentan desde un punto de vista de narrador omnisciente (es decir, se presenta toda la información sin restricciones). La segunda, **focalización externa**, es la forma de presentación que muestra sólo el contenido visible y audible, donde el narrador no conoce los puntos de vista de los personajes. La tercera, **focalización interna** (“*internal focalization*”), donde el narrador es un personaje de la historia y la cuenta desde su punto de vista (**focalizador**). Esta técnica permite al lector formar parte de la historia. El término “**inmersión**” se refiere a la sensación del lector para sentirse en la historia. Por otro lado, por “**control**” se entiende a la capacidad del lector de influir en el curso de la historia, bien tomando decisiones sobre las acciones que ocurren en la trama, o bien tomando el control de alguno de los personajes. Murray [18] compara, en su libro “*Hamlet on the Holodeck: The Future of Narrative in Cyberspace*”, la instalación de realidad simulada de inmersión con hologramas que se encuentra en naves y bases estelares en el universo de Star Trek con la forma en que se producen narrativas que permitan a los agentes simular mundos ficticios y adentrarse en los mismos cambiando el curso de la historia. Sin embargo, el problema principal al intentar permitir al usuario tomar el control del curso de la historia, surge al intentar mantener la coherencia de esta.

Los conceptos narrativos explicados anteriormente se pueden englobar en el **marco narratológico** propuesto por Jahn [17]. Como se puede ver en la Ilustración 1, “quién” se puede referir, dentro del texto, al narrador, o fuera del mismo, al autor. El acto de narrar es el acto de contar el “qué” y “cómo” contarlos. La historia se compone de eventos y elementos existentes en la misma, como personajes y lugares. El cómo lo determinan el discurso narrativo (texto) y el resto de los recursos multimedia que lo conforman.



Ilustración 1 Marco narratológico planteado en Jahn [19]

Prácticamente todas las teorías narrativas distinguen entre el “qué” se cuenta (la historia) y el cómo se cuenta (el discurso). De acuerdo con el lingüista Saussure [20], padre del Estructuralismo, en “*Nature of the linguistic sign*”, cada **signo** consiste en un **significante** y un **significado**, donde el primero es la versión tangible (el discurso) y el segundo la historia. Por lo tanto, se pueden seguir dos orientaciones distintas en el estudio narratológico. En el primer caso, desde el punto de vista del **discurso narrativo**, que analiza la forma del texto y las características pragmáticas que lo contextualizan desde un punto de vista social y cultural. En el segundo caso, desde el punto de vista de la **historia narrativa**, que estudia la trayectoria de los eventos y motiva la trama. Como se verá, estos dos puntos de vista teórico conforman las dos capas de las que se compone el modelo narrativo propuesto.

Además de conocer desde qué punto de vista se cuenta la historia y qué parte de la historia se va a contar, hay conceptos relevantes relativos a cómo contarla: El análisis del tiempo narrativo, el análisis temporal, la duración y los modos narrativos.

El **análisis de tiempos narrativos** estudia el punto en el tiempo en que se encuentran la historia y los personajes. Hay dos tiempos narrativos principales: el **pasado narrativo** y el **presente narrativo**. Dentro del tiempo presente, Casparis [21] distingue el **presente narrativo**, es decir, el tiempo en el que habla el narrador crea el contexto, que se desarrolla en el tiempo actual del personaje, que es el punto actual de la historia. El **presente histórico** presenta un contexto temporal pasado con tiempo verbal presente. El **presente genérico** que muestra, en forma de proverbios, hechos de validez general (por ejemplo, comenzando con “es un hecho que...”), y el **presente sinóptico** usa el tiempo presente en los títulos de capítulos, resúmenes, etc. Otra forma de categorizar las narrativas con respecto al tiempo es planteando la distinción entre las relaciones de anterioridad y posterioridad del tiempo de la historia con respecto a la del personaje, como plantea Margolin [22]. Se puede distinguir entre **narrativa retrospectiva**, que usa verbos en pasado para eventos que ocurrieron en el pasado; **narrativa concurrente**, que produce tiempos narrativos en presente de acciones que ocurren al mismo tiempo (por ejemplo, un diario); y **narrativas prospectivas**, que producen tiempos narrativos futuros de eventos que todavía no han ocurrido.

El **análisis temporal** estudia el orden o la **cronología** de la historia, la **duración** de la historia y el discurso y la **frecuencia**. En el caso del orden, la historia puede seguir una consecución cronológica de eventos u **orden cronológico**, o no seguirla. La **anacronía** se define como una “secuencia de eventos temporales desordenados” [10]. En cuanto a la duración, se distingue en primer lugar el **tiempo del discurso del tiempo de la historia** en que ocurren los eventos. En el primer caso, el tiempo del discurso plantea cuestiones como: ¿Se puede leer todo el texto en una vez? ¿Cuánto tiempo se tarda en leer cada capítulo? ¿Cómo se relacionan el tiempo del discurso y el tiempo de la historia? Estas cuestiones las trataron Müller [23] en 1969, Genette [10] y Rimmon-Kenan [24] en 1983. Por lo tanto, la relación entre el tiempo del discurso y de la historia permiten estudiar el **tempo** de esta. Una **presentación isocrónica** (“*isochronous presentation*”) es aquella en la que ambos tiempos son similares. En caso de que el tiempo del discurso sea menor que el de la historia se considera una **aceleración**, y el caso contrario, una **deceleración**. Cuando el tiempo de la historia se detiene (por ejemplo, debido a una descripción o a un comentario), se considera una **pausa**. La frecuencia determina, por su parte, el número de veces que cuenta la historia. Por ello, puede ser **singular** (cuando se cuenta una sola vez), **repetitiva** (cuando se cuenta varias veces, pero ocurrió una vez) e **iterativa** (cuando se cuenta una vez, pero ocurrió varias veces).

Los modos narrativos determinan la forma en que se presenta un texto. Se distinguen dos modos narrativos principales: **presentación escénica** (isocrónica), donde se presentan una serie de eventos consecutivos en orden cronológico; y **resumen**, donde el narrador condensa una serie de eventos desde un punto de vista específico.

Una vez definidos los conceptos básicos que engloban a las narrativas y sus formas de presentación, el siguiente paso es estudiar la relación entre los textos narrativos y el acto de narrarlos.

La **teoría de la citación** (“*quotation theory*”) estudia la forma en que el discurso de los personajes (lo que hablan o piensan) se relaciona con el discurso de la historia. Un modo indirecto de citar el habla o el pensamiento de un personaje es el **discurso indirecto libre**. Este tipo de discurso es una mezcla de elementos deícticos<sup>4</sup> donde se representan las palabras o pensamientos del personaje utilizando pronombres y tiempos verbales de la cita alineados con la estructura narrativa. Por ejemplo: “¿Cómo podremos salir del cine ahora?”. En un estilo de **discurso directo** se presenta primero al personaje que va a hablar y después se cita entrecomillado lo que dice. Por ejemplo: José dijo: “Podemos salir por la primera puerta”. Por otro lado, en el **discurso indirecto** se presenta el habla o pensamiento del personaje de forma indirecta, pero nombrando al personaje. Por ejemplo “Juan dijo que la primera puerta estaba cerrada”.

Los distintos modelos teóricos asociados a las narrativas ofrecen diferentes formas de estructurar y encapsular la información de forma comprensible. Estos modelos de estructuración pueden ser trasladados a otros ámbitos, como al computacional.

## 2.2 APROXIMACIÓN COMPUTACIONAL

Un concepto muy importante que surge al trasladar la teoría narrativa al ámbito computacional es el diálogo. Un **diálogo** se define como “plática entre dos o más personas, que alternativamente manifiestan sus ideas o afectos” [25]. La noción de **acto de diálogo** la introdujo Bunt [26] en 1979 y su definición se recoge en el **estándar ISO 24617-2** de 2012 como “actividad comunicativa de un participante del diálogo, interpretado como tener una cierta función comunicativa y un contenido semántico”. Este estándar recoge con información sobre actos de diálogo, el marcado o anotación del diálogo hablado, escrito y multimodal. Las funciones comunicativas descritas en el estándar especifican cómo el contenido semántico del acto del diálogo modifica el estado de la información de dirección una vez entendido el comportamiento comunicativo del hablante [27]. Son funciones comunicativas, por ejemplo: informar, preguntar, pedir, responder, etc. El estándar recoge también la definición de **agente** [28], que en este caso comprende tanto a personas como máquinas; en el sentido de que no sólo soporta lenguaje natural, sino también lenguaje no verbal propias de humanos (como gestos y expresiones faciales) y presentes agentes conversacionales (como guiños, pitidos, etc.).

De acuerdo con Traum [29], en su artículo “Aproximaciones Computacionales al Diálogo” un **sistema de diálogo** se compone de modelado y gestión del diálogo. El **modelado del diálogo** consiste en el desarrollo de un modelo de diálogo con un propósito concreto. Por otro lado, la **gestión del diálogo** o **estrategia de diálogo** es el proceso de decidir qué hacer en el siguiente paso del diálogo, desde las perspectivas del participante y el observador. Existen dos tipos de diálogos, uno basado en interfaces y el otro en simulaciones. El **diálogo basado en interfaces** funciona como un traductor entre el usuario

---

<sup>4</sup> Proviene de “deixis” [100]: “Señalamiento a una persona, un lugar o un tiempo, o a una expresión lingüística mediante ciertos elementos gramaticales. Expresan deixis el adverbio ayer, el verbo venir y el posesivo tus en -Tus amigos vinieron ayer-.”

y el sistema; y el **diálogo basado en simulaciones** intenta simular el diálogo humano. Por su parte, la gestión del diálogo se puede dividir en dos tipos: aproximación basada en estructura y basada en marcos. La **aproximación basada en estructura** se compone de unos “*scripts*” (patrones) que deciden cuál será el siguiente paso del diálogo, es decir, qué ocurrirá después. Esta aproximación se puede subdividir, a su vez, en tres tipos: aproximación basada en principios, estructural o basada en aproximaciones gramaticales (“*dialogue grammars*”) y basada en pares de adyacencia (en la que una persona habla y otro reconoce lo que dice). La **aproximación basada en marcos (“frames”)** parte de modelos de árbol y de estado finito donde se requieren unos parámetros para formular una consulta. Las distintas aproximaciones que surgen a partir de ésta son: basada en planes (como la propuesta por García-Serrano et al. [30] para modelar componentes cognitivos), basada en estados de información (que trabajan con la coherencia en problemas de preguntas-respuestas, como la arquitectura Trindikit) y modelos estadísticos (como los modelos basados en aprendizaje por refuerzo [31]).

En este momento es necesario destacar la **Teoría de actos del habla**, que fue inicialmente propuesta, desde un punto de vista teórico, por Austin (1962) [32] y Searle (1969) [33] y que daría lugar, más adelante, a la aproximación computacional propuesta por Bunt (1979) [27] que, a su vez, da lugar al estándar ISO 24617-2 de 2012, citado anteriormente.

Las principales dificultades que surgen a la hora de crear los sistemas de diálogo son el desconocimiento del contexto, el diálogo no cooperativo, el diálogo multimodal y el diálogo multiparte. **Grounding** se refiere al desconocimiento del contexto, donde uno de los participantes del diálogo no tiene los conocimientos necesarios para su comprensión. Una solución propuesta para este problema es aplicar técnicas de aprendizaje por refuerzo [31] como las que se detallarán más adelante. **Diálogo no cooperativo**, donde al menos uno de los participantes no participa de forma activa en el diálogo. Para suplir este problema surge el concepto de **obligaciones del discurso**, que pretende motivar la coherencia del diálogo en ausencia de un participante colaborativo [34]. En el caso del **diálogo multimodal** se engloba no sólo texto, sino elementos multimedia. En este caso, los agentes conversacionales virtuales con apariencia humana sirven como ayuda aportando distintas capacidades (como la de parpadear, gesticular, etc.). Por último, los **diálogos multiparte**, donde hay un gran número de participantes y la gestión del sistema de diálogo es más compleja al necesitarse, por ejemplo, gestión de hilos.

El tipo de aplicaciones que se desarrollan con sistemas de diálogo van desde agentes conversacionales (como los asistentes de los bancos) con apariencia humana o no, o narrativas interactivas. Esto representa la diferencia entre el diálogo desde el punto de vista del Procesamiento del Lenguaje Natural, frente al diálogo como método narrativo.

Las **narrativas interactivas** son sistemas que usan técnicas de generación automática de narrativas para crear múltiples variantes de una historia [35]. Un ejemplo de narrativa interactiva es la aplicación móvil Capitán Pirata, desarrollada por ficcioninteractiva.com, y mostrada en la Ilustración 2 [36].

En el artículo de Endrass et al. [37], se presenta una aproximación experimental que refleja distintos diseños de narrativas interactivas (basados en turnos frente a interacción continua) e investiga cómo responde el usuario a las mismas. El **diseño basado en turnos** limita las posibilidades de decisión del personaje, dejando sólo elegir entre las conversaciones que el sistema tiene preestablecidas. En cambio, en un **diseño basado en diálogo o interacción continua**, el sistema permite al usuario detener la conversación en cualquier momento, aunque esto implique interrumpir mientras se cuenta la historia del personaje. En este caso, es más complejo mantener la coherencia de la historia, pero

permite una mayor autonomía al usuario. Los resultados mostrados indican que el usuario prefiere una interacción continua.

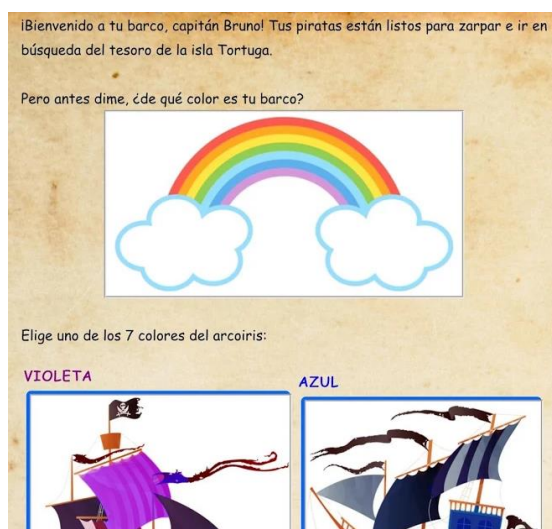


Ilustración 2 Imagen extraída de *ficcioninteractiva.com* correspondiente al juego *Capitán Pirata*

Podemos encontrar otras dos aproximaciones asociadas al dominio narrativo que han surgido partiendo de un punto de vista computacional: “*storytelling*” y “*eChronicles*”. **Storytelling** trata las diferentes formas de convertir las historias y eventos en palabras, imágenes y sonidos, con el objetivo de entretener, enseñar, explicar, etc. Al utilizar este término en un contexto digital, donde es posible contar las historias utilizando, por ejemplo, los videojuegos, televisión o realidad aumentada, surge el concepto de **inteligencia narrativa** (área de la IA que comenzó en la década de los 70 que explora la intersección de IA, los estudios multimedia y el diseño de sistemas de interacción persona-ordenador [38]). Por otro lado, un sistema **eChronicle** es un modelo que incluye formas de grabar, organizar y acceder a grabaciones de eventos multimedia, en un contexto no ficticio. En este sentido, las crónicas son las narrativas obtenidas que representan el comportamiento real de los personajes. Su objetivo principal es conseguir agregar nuevos eventos (o episodios) a las crónicas manteniendo la coherencia [39].

### 2.2.1 Tecnologías y estándares relacionados con la web.

El área de investigación de **Humanidades Digitales**<sup>5</sup> pretende integrar las disciplinas humanísticas (como historia del arte, lingüística, ...) con el uso de herramientas informáticas. En su etapa inicial, las humanidades digitales se centraron en el desarrollo de estándares que permitieran representar la información relativa al patrimonio cultural, y de ahí surgió el consorcio internacional **TEI** (“**Text Encoding Initiative**”) [40]. TEI desarrolló un estándar de marcado textual basado en el lenguaje de marcado **XML** [41] (“*Extensible Markup Language*”) para etiquetar los textos. XML es un metalenguaje<sup>6</sup> desarrollado en 1997 por la comunidad internacional **W3C** (“*World Wide Web*”).

<sup>5</sup> Se pueden encontrar distintas definiciones de Humanidades Digitales en <http://whatisdigitalhumanities.com>

<sup>6</sup> Un metalenguaje de marcado es un lenguaje para definir lenguajes de marcas.

*Consortium*") [42] a partir del lenguaje de marcado **SGML** ("Standard Generalized Markup Language") [43], con el objetivo de facilitar la interoperabilidad entre SGML y HTML. **HTML** ("Hypertext Markup Language") [44] es el lenguaje usado en la Web, diseñado inicialmente para describir semánticamente documentos científicos, actualmente permite describir la estructura de los documentos web a través de un conjunto de etiquetas y atributos definidos. La diferencia principal entre XML y HTML es que el segundo sirve para almacenar la estructura del texto, pero no permite almacenar la información semántica del mismo. Esta información semántica es posible almacenarla, sea cual sea el dominio, utilizando el metalenguaje XML, ya que permite definir las reglas que conformarán el lenguaje de marcado.

Además de comprender cómo almacenar la información textual de forma que sea posible preservar su información semántica subyacente, también es necesario plantear qué información se va a almacenar y cómo se presenta la misma. Los textos pueden ser estructurados, semiestructurados o no estructurados. La **información estructurada** se compone de colecciones y corpus<sup>7</sup> que contienen anotaciones y metadatos. Un **metadato** (del griego μετα, "meta", "después de, más allá de"<sup>8</sup> y dato, del latín datum, 'lo que se da'<sup>9</sup>) es un término acuñado en la década de los 60 del s.XX para describir las características de los datos (datos sobre datos) y permitir su recuperación e interoperabilidad. Por ejemplo, en el entorno bibliotecario, la catalogación es un proceso de generación de metadatos, que actualmente puede realizarse de forma manual o automática. La **información no estructurada** es la información presentada en texto plano, de forma que no se tiene, a priori, información semántica ni estructural del mismo. Por último, la **información semiestructurada** contiene parte de información estructurada y otra parte de información en texto plano. Para modelar los metadatos existentes en la web, W3C creó una serie de especificaciones denominada RDF. **RDF** ("Resource Description Framework") es un estándar de modelado para el intercambio de datos en la web, y permite mezclar, mostrar y compartir información estructurada y semiestructurada entre diferentes aplicaciones web [45]. El lenguaje utilizado para consultar información RDF es SPARQL. **SPARQL** es una especificación X3C se refiere tanto al protocolo remoto para realizar consultas (SPROT) [46] como a la especificación XML para representar los resultados de consultas (RESULTS) [47]. En su versión más reciente de 2013, **SPARQL 1.1** [48] distingue entre una serie de especificaciones que permiten consultar y modificar la información contenida en RDF, añadiendo, entre otras, la posibilidad de intercambiar la información en otros formatos, como **JSON**<sup>10</sup>.

Las principales áreas de investigación relacionadas con el objetivo de la extracción de información de la web son la minería de textos y la extracción y recuperación de información. **Minería de textos** es el descubrimiento por un computador de información previamente desconocida, extrayendo automáticamente información de diferentes lugares. En minería de textos los patrones se obtienen desde el texto en lenguaje natural [49]. **Extracción de información**, por su lado, se basa en analizar el lenguaje natural para extraer trozos de información. El proceso toma textos (o expresiones) y produce datos de salida no ambiguos que pueden ser directamente mostrados al usuario final, almacenados en una base de datos, analizados o usados para indexar en motores de búsqueda. La diferencia entre extracción de información y **recuperación de información** es que en el primer caso la aplicación

---

<sup>7</sup> El término corpus, referido a corpus lingüístico, ha sido prestado del inglés, que a su vez tomó el término del latín, cuerpo. Según el diccionario de la Real Academia Española, un corpus es un "conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación (<http://dle.rae.es/?id=AwTBMcs>).

<sup>8</sup> Meta: según la Real Academia Española (<http://dle.rae.es/?id=P4Cj7Rg>)

<sup>9</sup> Dato: según la Real Academia Española (<http://dle.rae.es/?id=Bskzsq5|BsnXzV1>)

<sup>10</sup> **JSON** [68] es un formato para serializar datos



analiza el texto y devuelve los trozos que necesita el usuario; mientras que en el segundo caso el sistema encuentra información relevante y la muestra al usuario [2].

### 2.2.2 Datos enlazados en abierto en la Web.

LOD (*“Linked-Open Data”*) surge como un nuevo paradigma donde la información se almacena en forma de datos estructurados, anotados y relacionando entre si los conceptos que lo conforman; permitiendo el acceso libre a la información y su interoperabilidad con otros formatos y sistemas a través del uso de distintos estándares de almacenamiento.

Algunos ejemplos de fuentes de recursos bajo el paradigma LOD en la web son:

- **Wikipedia** [50] es un proyecto de enciclopedia web multilingüe que comenzó en el año 2001 y que cuenta con más de un millón de artículos.
- **Wikidata** [51] es una base de conocimiento de acceso libre y gratuito, colaborativo y multilingüe que incluye datos estructurados para servir de soporte a Wikipedia [52], Wikimedia Commons [53] y otras comunidades de la Fundación Wikimedia [54]. Wikidata es un repositorio compuesto por una serie de ítems (que representan conceptos, temas u objetos), y cada uno de estos ítem está contiene una etiqueta, una descripción y una serie de sobrenombres (nombres alternativos para ese ítem). Para describir las características de los ítems se utilizan los enunciados, y están compuestos por una propiedad<sup>11</sup> y su valor. Para poder obtener la información existente en Wikidata, se pueden utilizar las herramientas web disponibles, como **Wikidata Query Service** [55], o bien a través de su API<sup>12</sup>, que permite obtener los datos en distintos formatos (como XML o JSON) [56]. Wikidata cuenta, además, con la información que existía en **Freebase** [57], una base de conocimiento colaborativa creada partiendo de distintas fuentes, incluyendo la colaboración de usuarios, que quedó en desuso en el año 2014.
- **DBpedia** [58] extrae información de Wikipedia y la organiza y muestra de forma estructurada, lo que permite realizar consultas complejas sobre la información existente en Wikipedia utilizando consultas SPARQL.

### 2.2.3 Estándares de metadatos en el ámbito de las Humanidades Digitales.

En el ámbito de las humanidades digitales, al igual que ocurre en otras áreas de investigación, no sólo es importante poder almacenar, consultar y modificar la información relativa a los conceptos, sino también sus relaciones. En este sentido, surge, en el ámbito computacional, la definición de **ontología**. Siguiendo la definición de la RAE: “En ciencias de la comunicación y en inteligencia artificial, red o sistema de datos que define las relaciones existentes entre los conceptos de un dominio o área del conocimiento” [59]. Las ontologías especifican el vocabulario y las relaciones entre los conceptos. Por

---

<sup>11</sup> La lista de propiedades disponibles en Wikidata se puede consultar en [https://www.wikidata.org/wiki/Wikidata:List\\_of\\_properties/es](https://www.wikidata.org/wiki/Wikidata:List_of_properties/es)

<sup>12</sup> Interfaz de programación de aplicaciones, es una capa de abstracción que permite comunicar dos aplicaciones utilizando una serie de funciones específicas.

ejemplo, una ontología médica podría modelar conceptos como enfermedades y síntomas, y sus relaciones existentes.

Desde finales de 2015, el Museo del Prado ofrece en su web información referente a las obras y artistas, todo esto bajo el paradigma de datos enlazados utilizando una serie de estándares asociados a los museos [60]. Esto permite la organización de todos los metadatos asociados a las obras en una red de conceptos y las relaciones que existen entre los mismos. Por ejemplo, se puede conocer la fecha de creación de las obras, y por ello, es posible realizar la búsqueda contraria: obtener qué obras se crearon en una fecha en concreto. La principal ontología o vocabulario específico que se ha utilizado en la Web del Museo del Prado ha sido el modelo de referencia CIDOC-CRM. Además, se utiliza el estándar FRBR para modelar los registros bibliográficos del museo. En la Ilustración 3 se presenta el modelo utilizado en el museo [61].

Es posible acceder a la información completa de la ontología utilizada en el Museo del Prado a través de la siguiente URL:

<https://www.museodelprado.es/pmartwork>

El modelo digital del Prado está organizado, en su nivel superior, alrededor de 18 Objetos de Conocimiento enumerados a continuación:

- [OC Obra de arte](#)
- [OC Autor](#)
- [OC Exposición](#)
- [OC Voz de enciclopedia](#)
- [OC Evento Cronología del museo](#)
- [OC Actividad](#)
- [OC Estudio/Restauración](#)
- [OC Oferta de empleo](#)
- [OC Beca](#)
- [OC Noticia](#)
- [OC Cita en medio](#)
- [OC Recursos multimedia \(vídeo, signoguía, audio, audioguía\)](#)
- [OC Boletín](#)
- [OC Licitación](#)
- [OC Adjudicación](#)
- [OC Post red social \(Facebook y Twitter\)](#)
- [OC Producto](#)
- [OC Recorrido \(Mi Prado\)](#)

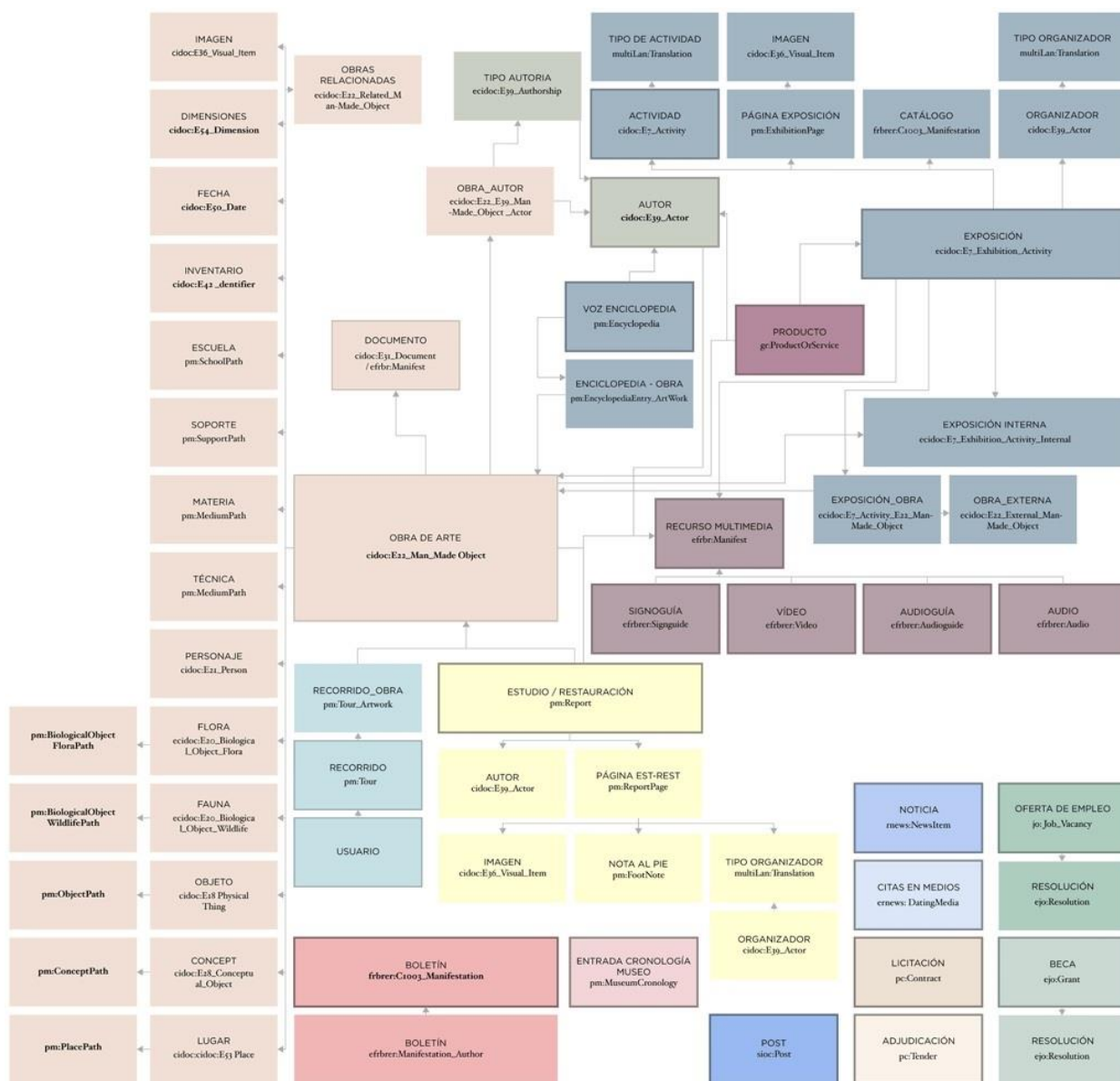


Ilustración 3 Modelo semántico extraído de la Web del Museo del Prado

**Estándar CIDOC para museos, bibliotecas y archivos. Ontología CIDOC-CRM.**

La dificultad a la hora de definir qué estructuras de metadatos se van a utilizar en el ámbito del museo, la encontramos en la variedad y ambigüedad de los datos de los que se dispone. Para suplir esta falta de normalización, el Grupo de normalización documental del Comité internacional para la documentación del Consejo internacional de los museos (ICOM-CIDOC) elaboró el modelo CIDOC-CRM (“Contextual Reference Model”), una ontología formal y un estándar ISO (ISO 21127:2006).

CIDOC Conceptual Reference Model [62] (CRM) cuenta con descripciones y una estructura formal para poder representar los procesos centrales del museo. Provee una forma de representar las propiedades cambiantes de los objetos culturales a lo largo del tiempo (por ejemplo, el cambio de propietario de

una pintura). Pretende facilitar la integración, mediación e intercambio de información heterogénea de patrimonio cultural. Los principios de diseño seguidos en su implementación son:

- Para cada concepto<sup>13</sup> debe haber evidencias de estructuras de datos extensamente utilizadas.
- El desarrollo se hace desde abajo hacia arriba y de manera incremental.
- Se concentra en la definición de las relaciones más que en la terminología en sí.
- Trata con la información del pasado y presente, dando por hecho que los datos vienen correctos desde la fase anterior.
- Trata con roles y entidades discretas, no con valores numéricos.

El concepto sobre el que se basa el estándar es el modelado de los eventos<sup>14</sup>, lo que permite la representación de metadatos y la descripción de contenido. Así, los cuatro objetivos en los que se centra el CRM son:

1. La posible ambigüedad de las relaciones entre entidades y los identificadores que se usan para referirse a las mismas (los distintos nombres que pueden tener las entidades).
2. Los sistemas de clasificación y tipado engloban tanto la estructura de información que expresa la realidad desde un punto de vista externo como parte de la realidad histórica.
3. La manera humana natural de analizar el pasado es dividir la evolución en eventos del espacio y el tiempo.
4. Los objetos inmateriales (“objetos conceptuales”) pueden residir en más de un portador a la vez. Puede haberse perdido el último portador que lo poseía, pero el objeto inmaterial no dejará de existir.

Se define el CRM a través de las relaciones que éste soporta:

- Identificación de los ítems a través de sus nombres reales.
- Clasificación de ítems del mundo real.
- Descomposición en partes de cosas materiales e inmateriales, grupos de personas, lugares y tiempos.
- Participación de ítems persistentes en entidades temporales.
- Localización de entidades temporales en espacio-tiempo y objetos físicos en el espacio.
- Influencia de objetos en actividades y productos, y viceversa.
- Referencias de los objetos de información a un ítem del mundo real.

Gracias a este modelo se pueden distinguir los nombres que representan un ítem real de los nombres que representan sólo nombres [63].

Existen herramientas en Internet que permiten detectar los posibles nombres que pueden englobar a la misma entidad nombrada. En la Ilustración 4 se muestra un ejemplo del artista Velázquez<sup>15</sup>.

---

<sup>13</sup> Concepto: Idea que concibe o forma el entendimiento [101]. En este ámbito, se puede referir a ideas, movimientos artísticos, objetos, personas, etc.

<sup>14</sup> Los eventos se definen como “eventualidad, hecho imprevisto, o que puede acaecer” [102], es decir, son los hechos que ocurren en la historia y alteran el curso natural de la misma (por ejemplo, una guerra).

<sup>15</sup> <http://isni-url.oclc.nl/isni/0000000121400618>

ISN:000000121400618 [search](#) [more](#)

Examples: Prince; Princeton University; Ken Follett; Follett, Ken; Follet,

shortlist **title data** search history

results search [or] ISN:000000121400618 | 1 hits

labels sources data marc21

**Please help us improve this record**

If you have any supplemental information about the identity listed here, please click in this box to go to the contribution form.

**Thank you in advance!**


<p> <b>ISNI:</b> 0000 0001 2140 0618</p> <p><b>Name:</b></p>	<p>Belasquez, Diego Rodríguez de Silva y De Silva, Diego</p> <p>De Silva Velaskes, Diego</p> <p>De Silva Velazquez, Diego Rodriguez</p> <p>De Velázquez y Silva, Diego</p> <p>Didacus Velázquez</p> <p>Diego Belazquez</p> <p>Diego de Silva Velázquez</p> <p>Diego de Silva y Velázquez</p> <p>Diego Rodríguez de Silva y Velasquez</p> <p>Diego Rodríguez De Silva Y Velazquez</p> <p>Diego Rodríguez de Silva y Velázquez (spanischer Maler des Barock)</p> <p>Diego Valasquez</p> <p>Diego Velasco Spagnolo</p> <p>Diego Velaskes</p> <p>Diego Velasquez</p> <p>Diego Velásquez (peintre espagnol)</p> <p>Diego Velázquez</p> <p>Diego Velázquez (Espaniol a pintor)</p> <p>Diego Velázquez (espanjalainen taidemaalari)</p> <p>Diego Velázquez (hiszpański malarz)</p> <p>Diego Velázquez (kunstschilder)</p> <p>Diego Velázquez (pintor espanhol, artista principal na corte de Filipe IV)</p> <p>Diego Velázquez (pintor español)</p> <p>Diego Velázquez (pittore spagnolo)</p> <p>Diego Velázquez (principal artista da corte de Filipe IV de Espanha e do século de Ouro)</p> <p>Diego Velázquez (Spanish painter)</p> <p>Diego Velázquez (Spanish painter who was the leading artist in the court of King Philip IV)</p> <p>Diego Velázquez (spansk maler)</p> <p>Diego Velázquez (spanyol barokk festő)</p> <p>Diego Villasco</p> <p>Diego Velaskes</p> <p>Djego Velaskess</p> <p>Don Diego Velasquez</p> <p>Rodríguez de Silva y Velásquez, Diego</p> <p>Rodríguez de Silva y Velázquez, Diego</p> <p>Rodríguez de Silva y Velázquez, Diego (d 1599-1660)</p> <p>Silva, Diego de</p> <p>Silva Velázquez, Diego de</p>
---	---

Ilustración 4 Imagen que representa los distintos nombres de artista Diego Velázquez

### Otros estándares de metadatos para los museos.

Además de los estándares de metadatos para museos integrados en el Museo del Prado, existen otros modelos que han surgido a partir de los anteriores.

- **ILFA** (“*International Federation of Library Associations and Institutions*”) [64] se fundó en Edimburgo, Escocia, en 1927. Se creó con la intención de revisar y regular distintos estándares bibliográficos.
- El modelo **FRBR** (“*Functional Requirements for Bibliographic Records*”) es el estándar utilizado para modelar los registros bibliográficos. El modelo FRBR fue diseñado como un modelo entidad-relación enfocado en ser independiente de dominio. Su mayor innovación es englobar las publicaciones y otros ítems alrededor de la noción de un origen conceptual en común para soportar la recuperación de información [63].
- El estándar **FRBR-OO** surge como una integración de los estándares FRBR y CIDOC-CRM, y el objetivo es intentar crear un estándar para recursos bibliográficos y museos orientado a

objetos. Es un modelo conceptual aún no implementado, por lo que no garantiza la representación de todos los tipos de datos, ni la resolución de errores. Por ejemplo, no se pueden representar datos temporales de forma directa, sino haciendo uso de distintas clases heredadas de CIDOC-CRM (*E65 Creation, E12 Production, E7 Activity, etc.*).

- El formato **MARC 21** [65] pretende crear una estructura para poder catalogar los archivos bibliográficos de forma que puedan ser interpretados por una máquina. Para ahorrar costes a la hora de compartir recursos y datos de los catálogos, surge BIBFRAME [66], que pretende acercar los estándares de descripciones bibliográficas a modelos de datos enlazados.
- **Schema.org** [67] es una comunidad fundada por Google, Yahoo y Yandex cuyo objetivo es crear, mantener y promocionar esquemas de metadatos. Incluye distintas codificaciones como RDFa, Microdata y JSON-LD.
- **Formato JSON-LD** [68]. JSON es un formato para serializar datos. JSON-LD se refiere a la especificación basada en JSON para datos enlazados. Sus objetivos son:
  - Simplicidad. Sin necesidad de librerías externas y con un lenguaje muy sencillo de aprender.
  - Compatibilidad. Satisface el estándar JSON.
  - Expresividad. La sintaxis serializa grafos dirigidos. Gracias a esto sabemos que podemos representar todos los modelos del mundo real que existen.
  - Brevedad. La sintaxis es breve y fácil de entender.
  - Sin ediciones la mayoría del tiempo, teniendo en cuenta siempre que el objetivo más importante es simplificar en lo posible.
  - Usable como RDF<sup>16</sup>, por lo que no hace falta conocer RDF para implementarlo.

#### 2.2.4 Modelos de organización del conocimiento asociados a las narrativas.

El proyecto DECIPHER [69] desarrolló en el año 2012 la ontología Curate [70], para describir las historias de museos en los que las relaciones entre las tramas (como la relación causa y efecto) puede expresarse entre los eventos, como se muestra en la Ilustración 5.

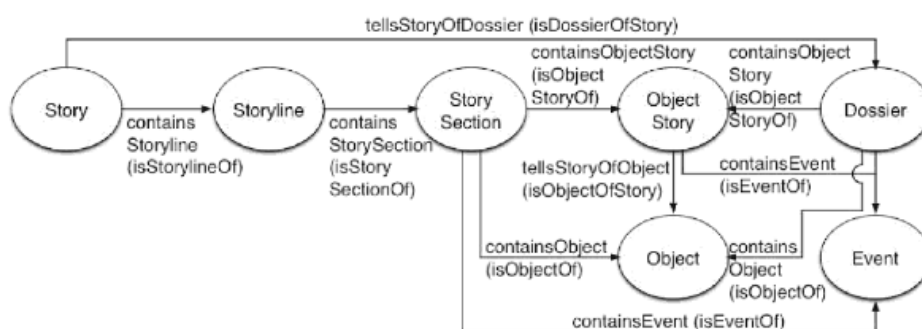


Ilustración 5 Estructura de clases y propiedades de la ontología Curate extraído de [70]

Existen varios modelos de organización creados a partir de esta ontología, entre los que destacan StorySpace y StoryScope. Además, se describe de forma general en qué consiste el proyecto CROSSCULT, que plantea una aplicación de CIDOC-CRM dentro del dominio del “storytelling”.

<sup>16</sup> RDF (Resource Description Framework) es un lenguaje para representar información en la Web.[103]

## Storyspace

**Storyspace** [4] es una API (Interfaz de Programación de Aplicaciones, en español) desarrollada para MacOS de la ontología *Curate*. Es utilizada por el personal de conservación de obras de dos museos para modelar narrativas y los procesos en los que están contruidos [70]. Una captura de ejemplo de este sistema se muestra en la Ilustración 6.

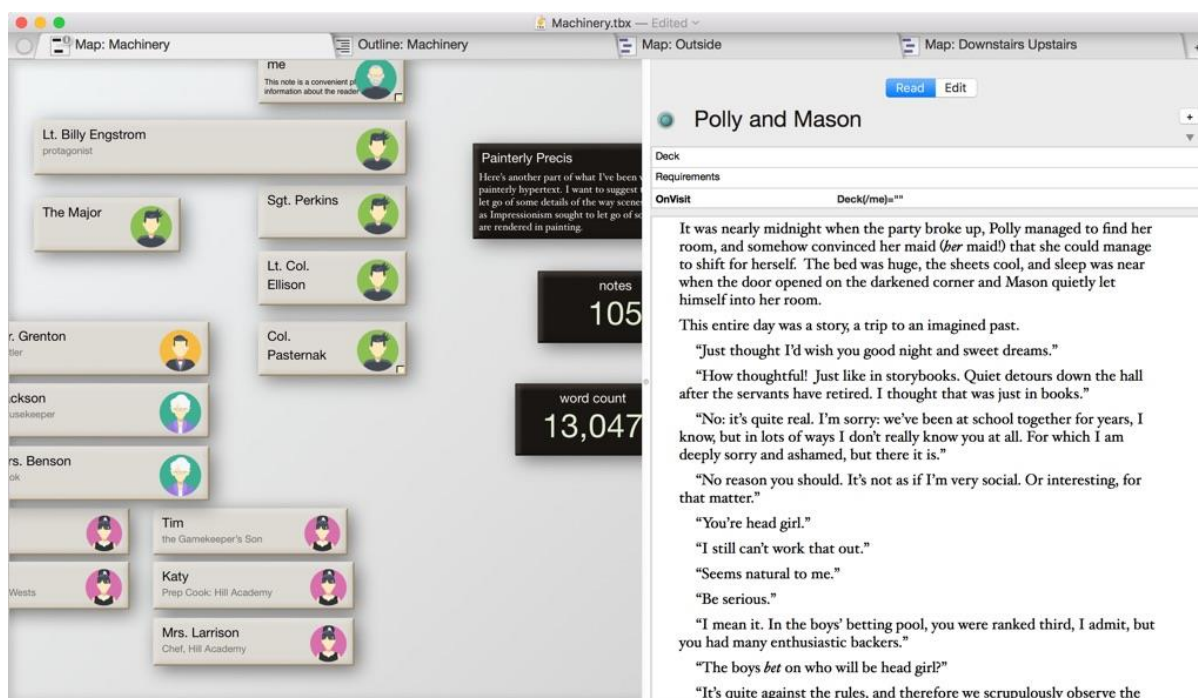


Ilustración 6 Imagen extraída de la Web de descarga de Storyspace [4]

Las características principales de esta herramienta son:

- Disponible para IOS: no es multiplataforma.
- Estructura de narración con hiperenlaces: No utiliza estándares, metadatos ni elementos narrativos.
- Creada para los trabajadores del museo que crean manualmente las narrativas.

Con respecto al modelo propuesto en este trabajo, StoryScope plantea un sistema en el que se anota la información textual de manera manual y sin plantear el uso de estándares. Esto implica que no existe, a priori, información semántica sobre la información obtenida.

## StoryScope

**Storyscope** [3] se desarrolló para crear y publicar historias de museos. Es un modelo, metodología y herramienta software para explorar y crear historias sobre patrimonio cultural y colecciones de museos. La ontología Storyscope fue desarrollada a través de la especificación SIOC del W3C<sup>17</sup>,

<sup>17</sup> SIOC (Semantically-Interlinked Online Communities) [104] provee conceptos y propiedades requeridas para describir información de comunidades online en la Web semántica.



partiendo de la ontología *Curate*, que modela formalmente la historia, trama y narrativa reflejando la estructura del entorno en Storyscope. Este sistema permite a los trabajadores del museo y a los visitantes investigar, desarrollar y presentar historias que conectan los distintos objetos del museo.

Está organizado en una serie de dossiers, y cada dossier se usa para englobar un conjunto de historias y contenido añadido asociado a un tema o un objeto. Cada historia contiene texto, una galería de imágenes y anotaciones semánticas.

A continuación, se describe más en detalle el modelo de StoryScope. Podemos definir, en este contexto, los siguientes conceptos obtenidos de la versión 7 online de Storyscope [71]:

- **Objetos.**

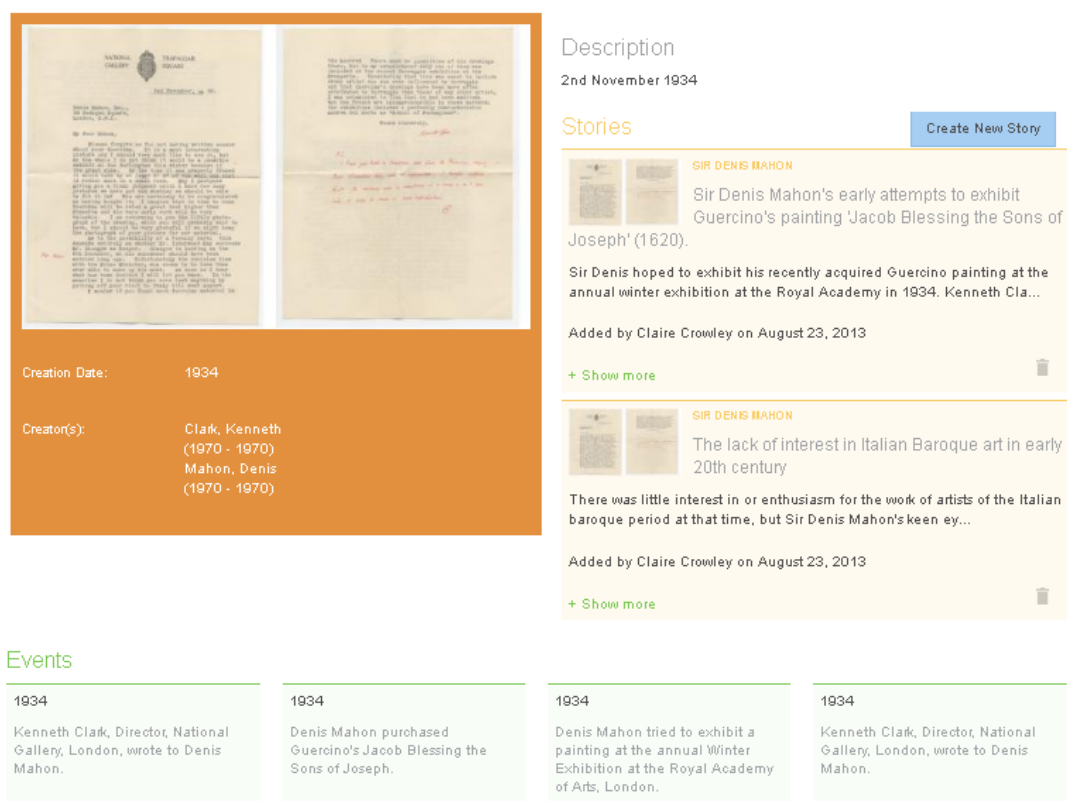
Los objetos son ítems de significancia cultural o histórica que conforman una colección del museo. Un objeto puede tener muchas historias diferentes. Un ejemplo se puede visualizar en la Ilustración 7.

Letter to Denis Mahon, from Kenneth Clark, Director, National Gallery, London

---

*Objects are digitised items of cultural and/or historical significance that are part of a museum's collection. Objects might originate in the user's own collection and be imported to Storyscope, from their Collection Management system. Alternatively Objects can be imported using the the CA. Objects contain media and metadata associated with an artefact such as a painting.*

*If you are logged in as an administrator of this dossier you will be able to create a new Story for this Object by hitting the 'Create new story' button located below.*



The screenshot displays a digital record for an object in Storyscope. At the top, it shows the title 'Letter to Denis Mahon, from Kenneth Clark, Director, National Gallery, London' and a description of what objects are in the system. Below this is a note for administrators. The main content area is divided into two parts: a document viewer on the left showing a scanned letter, and a metadata and stories section on the right. The metadata includes the creation date (1934) and creator(s) (Clark, Kenneth (1970 - 1970) and Mahon, Denis (1970 - 1970)). The stories section contains two entries, both titled 'SIR DENIS MAHON', with brief descriptions and 'Show more' links. At the bottom, there is an 'Events' section with four entries, each representing a specific event related to the object in 1934.

Creation Date: 1934

Creator(s): Clark, Kenneth (1970 - 1970)  
Mahon, Denis (1970 - 1970)

**Description**  
2nd November 1934

**Stories** Create New Story

**SIR DENIS MAHON**  
Sir Denis Mahon's early attempts to exhibit Guercino's painting 'Jacob Blessing the Sons of Joseph' (1620).  
Sir Denis hoped to exhibit his recently acquired Guercino painting at the annual winter exhibition at the Royal Academy in 1934. Kenneth Cla...  
Added by Claire Crowley on August 23, 2013  
Show more

**SIR DENIS MAHON**  
The lack of interest in Italian Baroque art in early 20th century  
There was little interest in or enthusiasm for the work of artists of the Italian baroque period at that time, but Sir Denis Mahon's keen ey...  
Added by Claire Crowley on August 23, 2013  
Show more

**Events**

1934	1934	1934	1934
Kenneth Clark, Director, National Gallery, London, wrote to Denis Mahon.	Denis Mahon purchased Guercino's Jacob Blessing the Sons of Joseph.	Denis Mahon tried to exhibit a painting at the annual Winter Exhibition at the Royal Academy of Arts, London.	Kenneth Clark, Director, National Gallery, London, wrote to Denis Mahon.

Ilustración 7 Ejemplo de objeto extraído de la versión 7 de Storyscope<sup>18</sup>

- **Historias.**

Una historia es una pieza de información sobre un objeto de patrimonio cultural. Puede relatar cómo se hizo el objeto o quién lo hizo, lo que representa o lo que le ocurrió desde que se hizo. Una historia

<sup>18</sup> <http://storyscope7test.ssl.co.uk/dossiers/1304/objectstories/1129/objects/5>



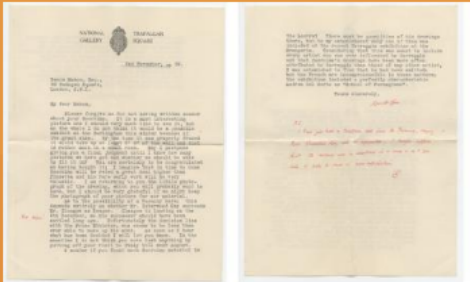
puede tener eventos y referencias, y muchas historias pueden unirse en una narrativa. Un ejemplo se puede visualizar en la Ilustración 8.

SIR DENIS MAHON > STORIES >

Sir Denis Mahon's early attempts to exhibit Guercino's painting 'Jacob Blessing the Sons of Joseph' (1620). View Edit

---

**Object**



Letter to Denis Mahon, from Kenneth Clark, Director, National Gallery, London

Creation Date: 1934

Creator(s): Clark, Kenneth (1970 - 1970)  
Mahon, Denis (1970 - 1970)

Notes: 2nd November 1934

Added by Claire Crowley on August 23, 2013

[view all](#)

Sir Denis hoped to exhibit his recently acquired Guercino painting at the annual winter exhibition at the Royal Academy in 1934. Kenneth Clark, wrote to Mahon explaining that he was unsure if they would be able to include a painting of such scale.

Added by Claire Crowley on August 23, 2013

[add similar +](#)

Add a Resource to this Story

Event
Reference

**Events**

**1934**

Kenneth Clark, Director, National Gallery, London, wrote to Denis Mahon.

**1934**

Denis Mahon purchased Guercino's Jacob Blessing the Sons of Joseph.

**1934**

Denis Mahon tried to exhibit a painting at the annual Winter Exhibition at the Royal Academy of Arts, London.

*Ilustración 8 Ejemplo de historia extraído de la versión 7 de Storyscope*

- **Eventos.**

Un evento es una situación, acción u ocurrencia que tuvo lugar en un cierto tiempo o lugar. Un evento pudo ser llevado a cabo por un individuo o un grupo. Pueden visualizarse como una línea temporal y un lugar. Se puede ver un ejemplo en la Ilustración 9.

Kenneth Clark, Director, National Gallery, London, wrote to Denis Mahon.

**Properties**

Activity: Write

Event start time: 1934

Agent(s)  
Creator(s): Clark, Kenneth (1903 - 1983)  
Mahon, Denis (1970 - 1970)

Locations: London

In the letter he notes that the Italian artists of the 17th century have fallen out of favour with art historians but would become popular in the future

Added by Claire Crowley on August 23, 2013

**Stories**

Sir Denis Mahon's early attempts to exhibit Guercino's painting 'Jacob Blessing the Sons of Joseph' (1620).

Sir Denis hoped to exhibit his recently acquired Guercino painting at the annual winter exhibition

The lack of interest in Italian Baroque art in early 20th century

There was little interest in or enthusiasm for the work of artists of the Italian baroque period

Ilustración 9 Ejemplo de evento extraído de la versión 7 de Storyscope<sup>19</sup>

- **Dossiers.**

Un dossier es una carpeta virtual, privada o compartida, que permite a los usuarios trabajar solos en un proyecto o colaborar en proyectos de otros. Pueden contener objetos, historias, eventos, referencias y narrativas.

Cada objeto en un dossier tiene una o más historias asociadas con él, viéndolas desde distintas perspectivas. Un ejemplo se puede visualizar en la Ilustración 10.

<sup>19</sup> <http://storyscope7test.ssl.co.uk/dossiers/1304/objectstories/1129/events/927>

## Sir Denis Mahon

In 2010, Sir Denis Mahon, one of the most distinguished art historians and collectors of the twentieth century, presented his personal library and archive to the National Gallery of Ireland. Of Irish descent, Sir Denis had a long and close association with the Gallery having donated a gift of eight outstanding seventeenth-century Baroque pictures to the collection. His munificence is considered the most important to the Gallery in recent years and has added significantly to its holdings of Baroque art and its research collections. Sir Denis's remarkable library reflects his scholarly interests covering art in the Western European tradition from classical times onwards. Italian art is strongly represented, particularly the baroque period, while British, Dutch, French and Spanish art also feature prominently. Many rare and antiquarian volumes, which are regarded as key sources in the study of art, have also been included in this gift.

Added by Claire Crowley on August 23, 2013

### Narratives

Add a narrative

#### Test Narrative

Cras mattis consectetur purus sit amet fermentum. Fusce dapibus, tellus ac cursus commodo, tortor mauris condimentum nibh, ut fermentum massa justo sit amet risus.

View Microsite

### Stories


Add a story



Sir Denis Mahon's early attempts to exhibit

Guercino's painting 'Jacob Blessing the Sons of Joseph' (1620).

Sir Denis hoped to exhibit his recently acquired Guercino painting at the annual winter exhibitio




Sir Denis Mahon and the National Gallery, London

Twice trustee of the National Gallery, London, Mahon crusaded tirelessly on its behalf.



Orlandi's publication of international artists' biographies

Pellegrino Antonio Orlandi's most influential publication, 'L' abecedario pittorico dall'autore r



Giuseppe Ghezzi and the Accademia di S. Luca

Ghezzi was a painter, illustrator and administrator.

### Objects

Add an object



Letter to Denis Mahon, from Kenneth Clark, Director, National Gallery, London



Jacob Blessing the Sons of Joseph



Letter from Denis Mahon to Lord Silkin



L'abecedario pittorico

[view all](#)

### Events

Add an event

1389

Cosimo de' Medici was born.

---

1992

Denis Mahon identified a painting by Caravaggio in Ireland.

---

1815

1989

Denis Mahon catalogued the Royal Collection of Guercino drawings

---

1986

Michael Wynne wrote to Denis Mahon.

---

1954



### References

Add a reference

Sir Denis Mahon obituary

Obituary published in The Guardian newspaper, April 28 2011

TWO FREAK PICTURES, The Irish Times, 23 October, 1923

Transcription: TWO FREAK PICTURES ART-

Sir Denis Mahon advocating free admission to museum

A letter from Julian Treuherz, keeper, Walker Art Gallery, Liverpool, to The Guardian newspaper i

Ilustración 10 Ejemplo del contenido de un dossier extraído de la versión 7 de Storyscope<sup>20</sup>

<sup>20</sup> <http://storyscope7test.ssl.co.uk/dossiers/1304>

- **Referencias.**

Las referencias son artículos o textos útiles (PDFs, enlaces, vídeos, etc.) que pueden compartir los guías o guardar los usuarios como material de referencia. Un ejemplo se puede visualizar en la Ilustración 11.



*Ilustración 11* Ejemplo de una referencia extraída de la versión 7 de Storyscope<sup>21</sup>

- **Narrativas.**

Las narrativas son una manera de enlazar objetos culturales con sus historias asociadas. Están compuestas de secciones similares a los capítulos de un libro o las páginas de un sitio Web, donde cada sección contiene un número de objetos y sus historias asociadas. A partir de un dossier se pueden crear múltiples narrativas y se pueden procesar como Micrositios. Un ejemplo se puede visualizar en la Ilustración 12.

---

<sup>21</sup> <http://storyscope7test.ssl.co.uk/dossiers/1304/references/175>

## Test Narrative

### Summary

Cras mattis consetetur purus sit amet fermentum. Fusce dapibus, tellus ac cursus commodo, tortor mauris condimentum nibh, ut fermentum massa justo sit amet risus. Sed posuere consetetur est at lobortis. Donec ullamcorper nulla non metus auctor fringilla. Cras justo odio, dapibus ac facilisis in, egestas eget quam. Sed posuere consetetur est at lobortis. Etiam porta sem malesuada magna mollis euismod.

Vestibulum id ligula porta felis euismod semper. Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum. Cras justo odio, dapibus ac facilisis in, egestas eget quam. Integer posuere erat a ante venenatis dapibus posuere velit aliquet.

[View Microsite](#)

*A Section functions much like a chapter of a book or a page of a website - it is a block of information; a number of Sections combine to create the overall Narrative. Each Section contains Objects and their associated Stories.*

[Re-order Sections](#)

[Add new Section](#)

Added by admin on November 7, 2013

#### SECTION

Cormac's Section



Sed posuere consetetur est at lobortis. Morbi leo risus, porta ac consetetur ac, vestibulum at eros. Curabitur blandit tempus porttitor. Fusce dapibus, tellus ac cursus commodo, tortor mauris condimentum nibh, ut fermentum massa justo sit amet risus.

Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum. Fusce dapibus, tellus ac cursus commodo, tortor mauris condimentum nibh, ut fermentum massa justo sit amet risus. Cras mattis consetetur purus sit amet fermentum. Aenean eu leo quam. Pellentesque ornare sem lacinia quam venenatis vestibulum.

#### SECTION

Added by admin on November 11, 2013

Cormac's second section



Added by admin on November 20, 2013

Added by admin on November 20, 2013

#### SECTION

Audio Section



Cras justo odio, dapibus ac facilisis in, egestas eget quam. Morbi leo risus, porta ac consetetur ac, vestibulum at eros. Nulla vitae elit libero, a pharetra augue. Praesent commodo cursus magna, vel scelerisque nisl consetetur et. Etiam porta sem malesuada magna mollis euismod. Donec sed odio dui.

Etiam porta sem malesuada magna mollis euismod. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Cras mattis consetetur purus sit amet fermentum. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus.

[Re-order Sections](#)

[Add new Section](#)

[Recommended Sections](#)

[Recommended Trails](#)

*Ilustración 12* Ejemplo de narrativa extraída de la versión 7 de Storyscope<sup>22</sup>

- **Agentes** [72]. En este contexto, son las personas u organizaciones que interactúan con el sistema. En concreto, un agente puede estar asociado a varios eventos, como se ve en las ilustraciones 13 y 14.

<sup>22</sup> <http://storyscope7test.ssl.co.uk/dossiers/1304/stories/5391>

## AGENT

Accademia di San Luca (3)  
Ashmolean Museum (2)  
Baldinucci, Filippo (5)  
Barry Flanagan (8)  
Bartolozzi, Francesco (7)

Ilustración 13 Ejemplo de agentes extraídos de Storyscope 7

SIR DENIS MAHON › ALL DOSSIER EVENTS ›

Marco Dente da Ravenna engraved a copy of Raphael's Saint Cecilia after the engraving by Marcantonio Raimondi

Added by Claire Crowley on August 23, 2013

Properties	
Activity:	Engrave
Event start time:	1514
Event end time:	1520
Agent(s) Creator(s):	Marco Dente da Ravenna (1493 - 1527) Raimondi, Marcantonio (1970 - 1970) Raphael (1970 - 1970) Ravenna, Marco Dente da (1493 - 1527)
Locations:	Rome
Genres:	Christian Art
Styles and Movements:	Renaissance

Ilustración 14 Ejemplo de evento con varios agentes extraído de Storyscope 7

- **Micrositios.**

Los Micrositios son las páginas usadas para visualizar las narrativas, por lo que permiten al usuario ver su narrativa de un modo visual. Cada sección de la narrativa puede verse como un enlace separado del micrositio. Un ejemplo se puede visualizar en la Ilustración 15.

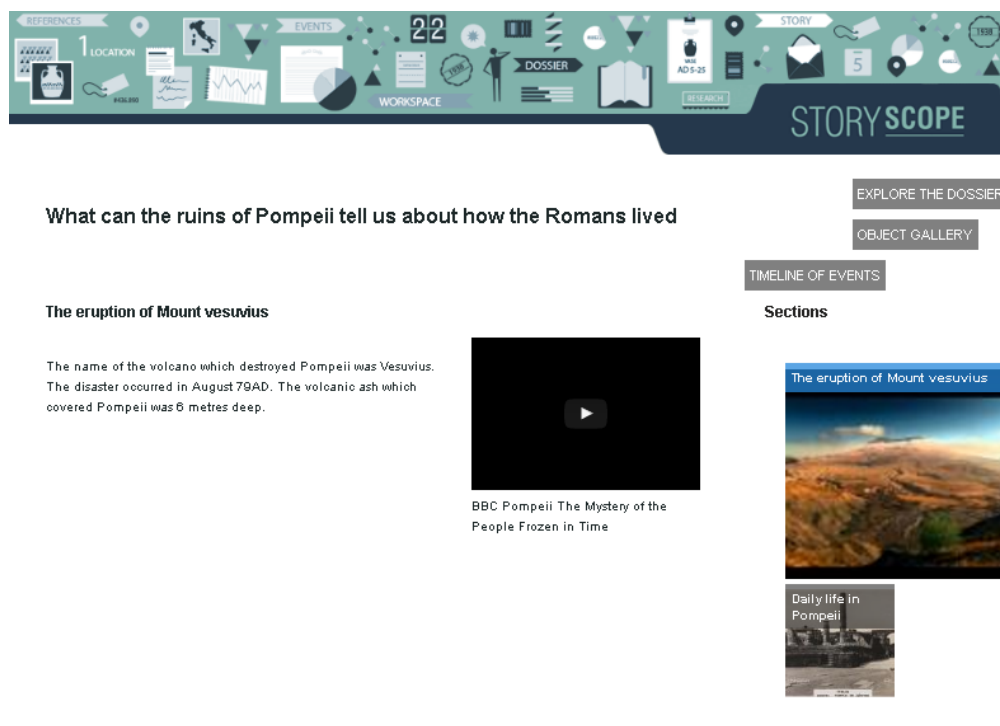


Ilustración 15 Ejemplo de micrositio extraído de la versión 7 de Storyscope<sup>23</sup>

Para crear una narrativa en StoryScope, se deben seguir los siguientes pasos [73]:

1. Investigación. Donde se deciden qué conceptos e ideas se van a explorar para construir la narrativa. Por ejemplo, se pueden escoger qué eventos históricos van a tratarse para una obra.
2. Creación de dossier. Es la fase de creación de dossier, donde se añaden objetos (obras), eventos, referencias e historias al mismo.
3. Exploración de narrativas y recursos, investigación de relaciones entre objetos, creación de la narrativa. En esta fase se investigan otras narrativas, además de otros recursos (eventos, historias y referencias), previamente creados. Así, se puede comprender cómo es el proceso de selección y desarrollo, además de poder aprovechar otros recursos previamente creados. Por otro lado, se investigan las relaciones entre los objetos del dossier.
4. Evaluación y documentación. Donde se añade material adicional a disposición del público.

Es también de interés el método de representación de eventos utilizado en StoryScope. Para permitir interpretar los objetos del museo y representar historias, se utilizan etiquetas, normalmente obtenidas de una base de datos externa, como Freebase<sup>24</sup>. Si la etiqueta es una persona, entonces los datos representados serán, entre otros, lugar y fecha de nacimiento. Los eventos se representan usando un evento de esquema simple con la **ontología LODE** [3]. Un ejemplo de etiquetas se puede visualizar en la Ilustración 16.

<sup>23</sup> <http://storyscope7test.ssl.co.uk/dossiers/6393/stories/6468/sections/6469/output>

<sup>24</sup> Freebase [57] es una base de datos creada de manera colaborativa usada para estructurar conocimiento humano en general.

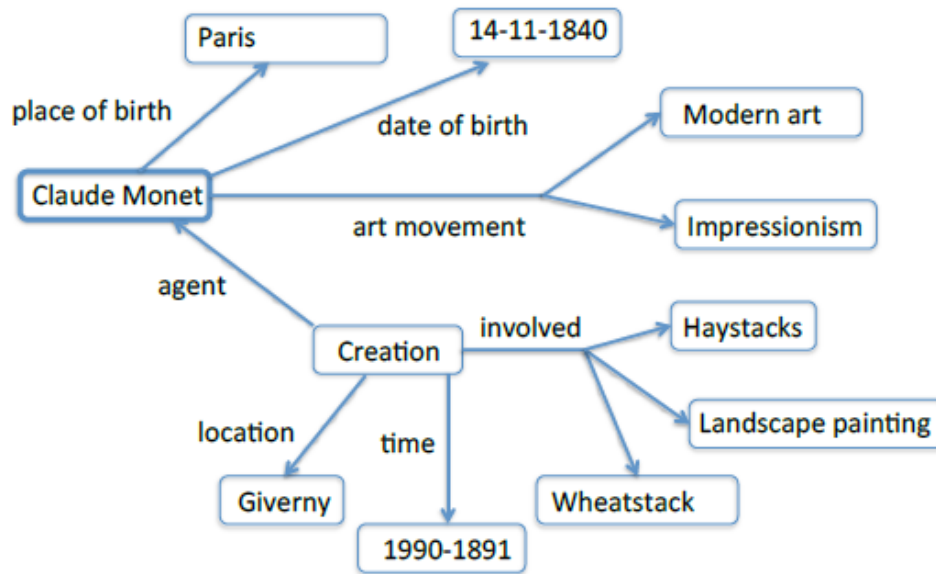


Ilustración 16 Representación de eventos y hechos relativos a Claude Monet[3]

Las propiedades de los eventos son:

- Tipo
- Agente
- Localización
- Fecha
- Involucrado (incluye otro concepto asociado con el evento)

El esquema usado para describir los eventos lo rellenan los trabajadores del museo. Este tipo de modelado puede seguir el estándar CIDOC-CRM detallado anteriormente.

Los temas (o tópicos) son conceptos centrales que unen otros elementos de la historia, donde la historia ya tiene un conjunto de estos en forma de etiquetas.

Así, los conceptos pueden tener más importancia atendiendo a dos medidas:

1. Cobertura. El total de etiquetas de historia que están asociadas a cada hecho o coocurren en un evento con la etiqueta de historia.
2. Frecuencia. El total de veces que el concepto aparece en un hecho o en un evento asociado con la etiqueta de historia.

Los temas candidatos del espacio de conceptos son ordenados en primer lugar por su cobertura, y en segundo lugar por su frecuencia. Los conceptos los genera el agente y están “encapsulados” en propiedades de los eventos y propiedades de cualquier hecho (como los movimientos de arte asociados) que son especificados como relevantes para el tema en el panel de configuración de Storyscope, como puede verse en la Ilustración 17.



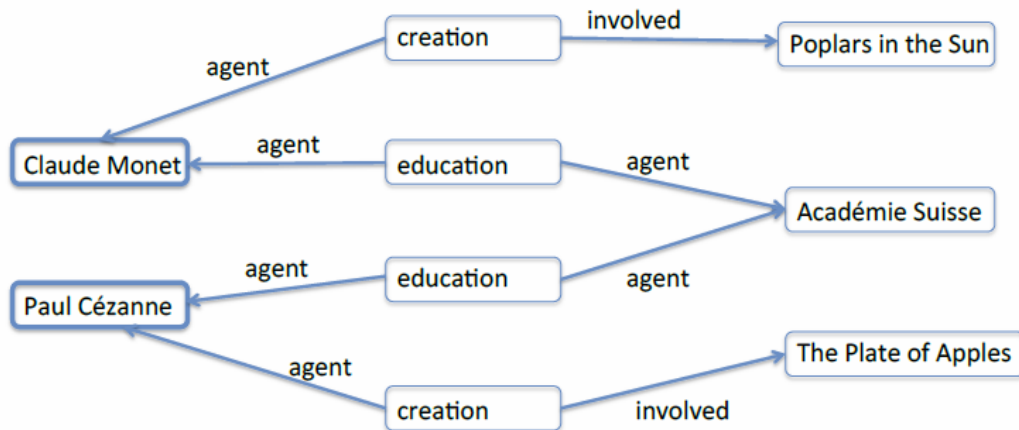


Ilustración 17 Temas desde las etiquetas de historia de Claude Monet y Paul Cézanne [3]

Los escenarios son generados por el espacio de eventos asociado con una (o todas) las etiquetas de historia. Los escenarios candidatos de un espacio de eventos son todas las parejas localización-fecha que pueden ser derivadas desde las propiedades de tiempo y lugar de los eventos. Por ello, incluyen un punto temporal (fecha de inicio o de fin) o un período temporal (fecha de inicio y de fin).

La evaluación de los escenarios se hace de la misma manera que en el caso de los conceptos: cobertura y frecuencia; pero en el caso de los escenarios, la frecuencia ocupa un segundo lugar, por ser menos prioritario.

Los conceptos<sup>25</sup> se calculan en dos fases:

1. Calcular la frecuencia y la cobertura de cada escenario único.
2. Solapamiento de los escenarios que ocurren en el mismo lugar y coinciden con el mismo tiempo. Por ejemplo, los escenarios “Paris 1990-1992” y “1992-1994” se convierten en un único escenario “Paris 1990-1994”.

Seleccionar un tema, entonces, producirá un espacio de eventos y hechos relativos al concepto. Por ejemplo, si seleccionamos el tema “Claude Monet”, permitimos buscar información acerca de qué otras cosas ha hecho el pintor. Por otro lado, seleccionar un escenario produce un conjunto de eventos que ocurren en el mismo. Por ejemplo, seleccionar “Giverny 1991” permite buscar qué otros hechos ocurrieron el lugar y fecha indicados.

Como se puede ver, este modelo supone un avance con respecto al anterior, por lo que sirve de punto inicial en el diseño del modelo narrativo propuesto en este trabajo. Las características básicas con las que cuenta son:

- Sistema de información basado en Web.
- Incluye modelo, metodología y colecciones de obras.
- La narración se escribe manualmente sobre la obra.
- Los elementos narrativos están separados de la narración, pero asociados a la obra.

<sup>25</sup> Concepto: Idea que concibe o forma el entendimiento [101].

- No hay relaciones entre los elementos narrativos, por lo que no es posible utilizar este modelo para obtener una serie de tramas automáticamente.

### *Proyecto CROSSCULT*

El proyecto CROSSCULT [74] busca crear una plataforma en abierto para aplicaciones móviles y web uniendo las áreas de razonamiento semántico, recomendación y contar historias (“*storytelling*”) en formato digital. El objetivo principal del proyecto es crear un base de conocimiento que relacione los recursos de datos enlazados con nuevos recursos basados en el estándar CIDOC-CRM, destacando relaciones de entidades de distintas áreas de interés.

En el artículo [75], los autores presentan una herramienta para ayudar a los expertos a desarrollar redes de asociaciones semánticas que sirven para conducir la jugabilidad al contar historias en las aplicaciones. En este sentido, utilizan distintos algoritmos de recomendación basados en *word2vec* [76] con conjuntos de datos importados de noticias de Google y utilizando la distancia coseno para calcular las distintas entidades recomendadas para una entidad dada.

Este proyecto resulta de especial interés dentro de este trabajo, puesto que la búsqueda de tramas se realizará de un modo similar, pero utilizando el grafo semántico creado en la base de preprocesamiento y adaptándolo al modelo narrativo propuesto, planteando, entre otras, la medida de distancia coseno para encontrar conjuntos de elementos con segmentos similares.

## 2.3 HERRAMIENTAS Y RECURSOS

Las fuentes desde las que se ha obtenido la información en este trabajo son:

- **Web del Museo del Prado:** Descripciones y metadatos asociados a las obras y los autores.
- **Wikidata:** Metadatos asociados a obras y autores. Además, establece las relaciones entre los datos del Museo del Prado y Wikipedia.
- **Wikipedia:** Información textual (descripciones) de obras, autores y palabras clave extraídas en las mismas.

Por otro lado, en este trabajo se han creado tres aplicaciones distintas. La primera, crea la estructura del modelo narrativo y procesa la información. La segunda, un sistema web creado para mostrar los itinerarios y evaluarlos. La tercera, para extraer la información y crear los grafos a partir de las sugerencias de las encuestas.

La primera herramienta se ha implementado en Python 2.7<sup>26</sup>[77][77][77], utilizando el motor de indexación Solr 7.2.1<sup>27</sup> para las distintas fases de procesamiento de la información y MySQL<sup>28</sup> para almacenar los índices y las relaciones. Para crear el grafo se utiliza la librería de Python Networkx<sup>29</sup>. El capítulo 4 detalla la implementación de esta herramienta. El código es libre y puede encontrarse en la siguiente URL:

<sup>26</sup> <https://www.python.org/>

<sup>27</sup> <http://lucene.apache.org/solr/>

<sup>28</sup> <https://www.mysql.com/>

<sup>29</sup> <https://networkx.github.io/>

[https://github.com/alicialara/TFM\\_PythonBackend](https://github.com/alicialara/TFM_PythonBackend)

El sistema web se ha implementado utilizando PHP<sup>30</sup> y JavaScript<sup>31</sup> como lenguajes principales. Para su construcción, se ha utilizado el marco de trabajo Laravel<sup>32</sup>, además de una serie de librerías de ayuda para mostrar la información. La información detallada sobre esta herramienta se encuentra en el capítulo 5. El código es libre y puede encontrarse en la siguiente URL:

<https://github.com/alicialara/TFM>

La tercera aplicación se ha implementado en Python 2.7, utilizando la herramienta CoreNLP de Stanford [78] para extraer los sustantivos y adjetivos, la librería Networkx de Python para crear los grafos y la librería TextRank [79] para establecer un peso en los nodos. TextRank [80] es un algoritmo que toma como entrada un conjunto de documentos y obtiene como salida un grafo de co-apariciones de términos. Cada nodo se corresponde con un término y cada enlace entre dos nodos se construye teniendo en cuenta las co-apariciones de dichos términos dentro de una ventana de texto. En el caso de este trabajo, se establece un valor de ventana 2, si bien los experimentos se realizan también sobre valores de ventana 3 y 4. La implementación de este trabajo se encuentra en Github, en la siguiente URL:

[https://github.com/alicialara/IR\\_SurveySuggestions](https://github.com/alicialara/IR_SurveySuggestions)

---

<sup>30</sup> <http://php.net>

<sup>31</sup> <https://www.javascript.com/>

<sup>32</sup> <https://laravel.com/>



### 3 CREACIÓN DEL MODELO NARRATIVO

Los modelos descritos anteriormente sirven como base para la creación de narrativas, pero no permiten integrar toda la información semántica contenida en el propio texto. Por ello, se propone el diseño de un nuevo modelo narrativo, dividido jerárquicamente en seis capas que permiten la obtención de un itinerario, como se puede ver en la Ilustración 18.

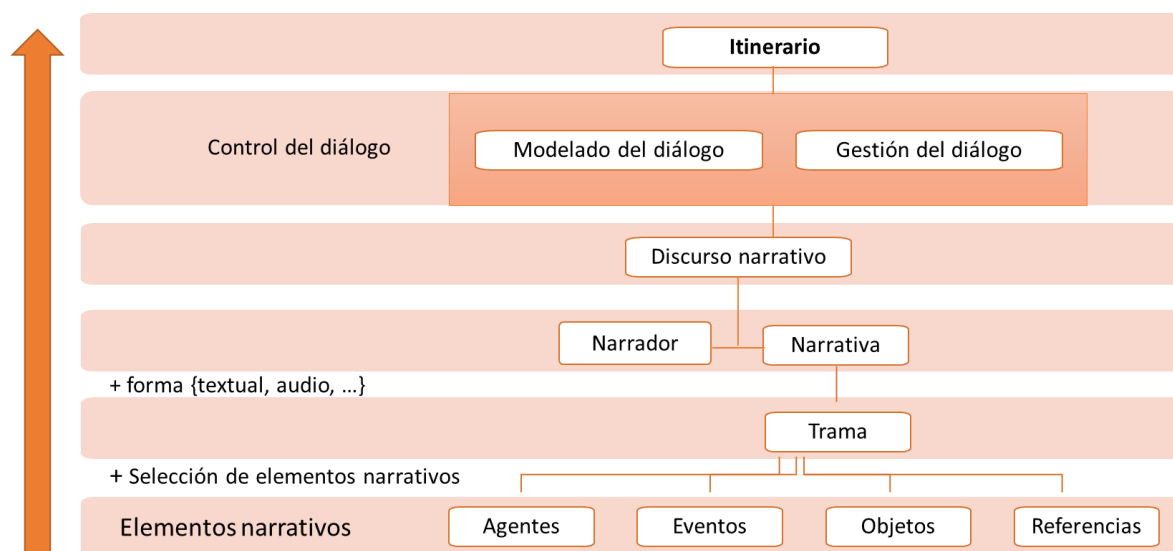


Ilustración 18 Modelo narrativo creado para este trabajo.

El modelo narrativo propuesto parte de las teorías aportadas por distintos lingüistas y expertos en el campo, como Brooks, Chatman, Todorov y Bal; y trasladadas a un punto de vista computacional por Paul Mulholland, como ya se ha visto anteriormente. Mulholland [5] pretende utilizar las narrativas como medio para mejorar el aprendizaje colaborativo y la exploración del legado cultural, partiendo de la idea de Brooks, quien plantea que la historia puede ser vista como un sistema de asociaciones entre personas, eventos y/o cosas, y una narrativa es cualquiera de las formas de contar la historia. Esta forma de organización de conceptos llega a formar la base del sistema narrativo propuesto en este trabajo, que contiene los distintos elementos narrativos: agentes (personajes o personas), eventos o hechos ocurridos en el sistema en un momento dado, objetos relacionados con el sistema y referencias (elementos externos al sistema que aportan valor semántico).

La selección de un subconjunto de elementos narrativos forma la trama, creando una red de conceptos y relaciones incluidos en el texto, la narrativa. Al asignar un narrador, esto es, el agente encargado de narrarla, se genera el discurso narrativo. Llegados a este punto, el modelo narrativo organiza la información como un texto anotado por los elementos existentes en la red semántica preexistente. El sistema del diálogo es la capa intermedia encargada de dividir las unidades textuales en segmentos (por ejemplo, oraciones), y plantear las posibles combinaciones que se puedan crear entre los mismos. Una de esas posibles combinaciones creará un itinerario, que será un conjunto de segmentos extraídos con la información relevante para el sistema. Dicho de otro modo, los itinerarios

son planes de recorridos con alternativas variables. En caso de no existir usuarios, se puede ver como una historia multimedia. A continuación, se muestran las definiciones de los conceptos:

A continuación, se detalla cada capa del modelo propuesto.

### 3.1 ELEMENTOS NARRATIVOS

Los elementos narrativos conforman el nivel más básico del modelo. Son los elementos anotados en las descripciones que conforman los itinerarios. En este trabajo: personajes, eventos, referencias y obras. Se trata de un grafo formado por nodos de tipo personajes, referencias, obras y eventos:

- **Personajes:** Pueden ser reales o ficticios, y pueden estar relacionados directamente con la obra (como el autor de esta), como indirectamente (si tiene alguna relación con la misma, como un Felipe IV en Las Meninas<sup>33</sup>). Por ejemplo, “Velázquez”, “Maestro de San Esteve de Andorra”, “Eliseo”, etc.
- **Eventos:** Son los acontecimientos que aportan el contexto histórico o político a los itinerarios. Por ejemplo, “Creación obra El pelele”, “Nacimiento de Ramón Despuig y Martínez de Marcilla, Gran maestro de la Orden de Malta.”, “Segunda Guerra del Opio”, “Guerra de las Malvinas”, etc.
- **Referencias:** Son todos los conceptos con información semántica anotados en las descripciones de las obras. Por ejemplo, “perro”, “bóveda”, “Manierismo”, “Barroco”, etc.
- **Obras:** Son todas las obras pertenecientes al Museo del Prado de las que se ha obtenido la información.

### 3.2 TRAMA

Una trama es una selección de un conjunto de elementos narrativos. Por ejemplo, una trama puede estar formada por “Velázquez”, “Barroco”, “perro” y “Nacimiento de Velázquez”. Son los elementos que conformarán el itinerario, es decir, los distintos temas de los que tratará el mismo. Una trama sencilla, por ejemplo, puede ser la formada por un solo elemento narrativo, como “Peter Paul Rubens”.

Por lo tanto, una trama puede ser vista como el conjunto de temas de los que trata el itinerario, y estos temas, a su vez, son elementos narrativos del modelo.

### 3.3 NARRATIVA

Una vez que la trama toma forma, ya sea textual, en vídeo, audio, audio descriptivo, etc. Se crea la narrativa. En este paso, se crea una representación textual de la trama. Por ejemplo, una narrativa para los elementos narrativos “Carlos IV”, “Napoleón” y “1799”:

“La familia de *Carlos IV* perteneció a la magnífica serie de retratos reales iniciada por Goya en septiembre de 1799, en las vísperas del Consulado de *Napoleón*, que, en un principio, prometía una pacificación de la tumultuosa década pasada.”

---

<sup>33</sup> [https://es.wikipedia.org/wiki/Las\\_meninas](https://es.wikipedia.org/wiki/Las_meninas)

### 3.4 DISCURSO NARRATIVO

Cuando a una narrativa se le asigna un narrador, se convierte en discurso narrativo. Un narrador es el personaje, real o ficticio, encargado de contar el itinerario. Por ejemplo, un guía turístico, o un personaje de realidad aumentada creado para esto.

### 3.5 SISTEMA DE DIÁLOGO

Esta fase del modelo narrativo es la encargada de procesar, categorizar y generar los itinerarios. Utiliza los nodos del grafo existentes en la capa inferior del mismo, relacionándolos con los segmentos que contienen estos elementos. Los segmentos son las unidades textuales en las que se dividen los itinerarios. En este caso, cada segmento corresponde con un párrafo de las descripciones. Así, es posible obtener, por ejemplo, una lista de todos los párrafos correspondientes a las obras que tratan sobre una trama concreta. Esto permite obtener tramas automáticas y generar los itinerarios sin necesidad de seleccionarlos manualmente.

### 3.6 ITINERARIO

Un itinerario es un conjunto de narraciones acerca de los elementos que lo conforman. Cada itinerario es el resultado del procesamiento del sistema de diálogo. Por ello, un itinerario está formado por un conjunto de segmentos ordenados que representan una trama.

En el Capítulo 5 se detallará en profundidad cómo se estructura el grafo y cómo realiza el gestor del diálogo la selección de tramas a partir del mismo.





## 4 EXTRACCIÓN Y PROCESAMIENTO DE LA INFORMACIÓN

---

El proceso de extracción de información sobre las obras del Museo del Prado se realiza a partir de distintas fuentes, organizando y mezclando los metadatos obtenidos. Las etapas llevadas a cabo en este objetivo son:

1. Extracción de información desde distintas fuentes y almacenamiento local.
2. Preprocesamiento e integración de metadatos y textos descriptivos desde distintas fuentes.
3. Segmentación de los textos obtenidos e identificación de los elementos narrativos.

### 4.1 EXTRACCIÓN DE INFORMACIÓN DESDE DISTINTAS FUENTES.

El proceso de extracción de información se ha llevado a cabo siguiendo los siguientes pasos:

1. **Obtención de una lista inicial de obras en Wikidata.**

A través de la API<sup>34</sup> que ofrece Wikidata para realizar consultas SPARQL, se realiza una búsqueda de todas las obras que pertenecen al Museo del Prado en idioma español. La lista de propiedades disponibles en Wikidata se puede consultar en la web [81]. En este caso, la propiedad *P195* representa una colección (de arte, un museo, o una colección bibliográfica). Por otro lado, los ítems se representan con identificadores que comienzan con el carácter "Q". En este caso, el Museo del Prado es el ítem *Q160112*<sup>35</sup>. Por lo tanto, la consulta realizada es la mostrada en la Ilustración 19:

```
SELECT ?obra WHERE {  
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],es". }  
  ?obra wdt:P195 wd:Q160112.  
}
```

*Ilustración 19 Consulta SPARQL para obtener las obras del museo*

Una vez obtenidos los enlaces a las obras de Wikidata, se almacenan en la base de datos para poder obtener la información de cada una de ellas más adelante.

2. **Obtención de la información asociada a cada una de las obras obtenidas en el paso anterior.**

En este caso, además de obtener los metadatos asociados a cada obra y su autor, se obtiene, en el caso de existir, el enlace a la obra en el Museo del Prado (a través de la propiedad *P973*) y el enlace a la obra en Wikipedia.

---

<sup>34</sup> Interfaz de programación de aplicaciones

<sup>35</sup> <https://www.wikidata.org/wiki/Q160112>

De nuevo, la información se obtiene a través de consultas SPARQL. Por ejemplo, para obtener la información relativa a la obra “Danza de aldeanos”, de Peter Paul Rubens (con identificador de ítem Q5799148), se realiza la siguiente consulta:

```
SELECT * WHERE {
  wd:Q5799148 rdfs:label ?label.
  OPTIONAL { wd:Q5799148 wdt:P361 ?forma_parte_de. }
  OPTIONAL { wd:Q5799148 wdt:P18 ?imagen. }
  OPTIONAL { wd:Q5799148 wdt:P571 ?fecha_de_fundacion_o_creacion. }
  OPTIONAL { wd:Q5799148 wdt:P135 ?movimiento. }
  OPTIONAL { wd:Q5799148 wdt:P17 ?pais. }
  OPTIONAL { wd:Q5799148 wdt:P276 ?ubicacion. }
  OPTIONAL { wd:Q5799148 wdt:P136 ?genero_artistico. }
  OPTIONAL { wd:Q5799148 wdt:P170 ?creador. }
  OPTIONAL { wd:Q5799148 wdt:P495 ?pais_de_origen. }
  OPTIONAL { wd:Q5799148 wdt:P2048 ?altura. }
  OPTIONAL { wd:Q5799148 wdt:P2049 ?ancho. }
  OPTIONAL { wd:Q5799148 wdt:P186 ?material_employado. }
  OPTIONAL { wd:Q5799148 wdt:P180 ?representa_a. }
  OPTIONAL { wd:Q5799148 wdt:P195 ?coleccion. }
  OPTIONAL { wd:Q5799148 wdt:P217 ?numero_de_inventario. }
  OPTIONAL { wd:Q5799148 wdt:P973 ?descrito_en_la_URL. }
  FILTER(LANGMATCHES(LANG(?label), "ES"))
}
```

*Ilustración 20 Consulta SPARQL para obtener la información de una obra del museo*

Los datos obtenidos en este caso para cada obra son:

- Nombre de la obra: “Las Meninas”, “La Fragua de Vulcano”, ...
- Colección en la que se encuentra: “Los cinco sentidos”, “Pinturas negras”, ...
- Movimiento al que pertenece: “Renacimiento italiano”, “Romanticismo”, ...
- País: “España”, “Francia”, ...
- Ubicación: “Real Academia de Bellas Artes de San Fernando”, “Egipto”, ...
- Género artístico: “pintura floral”, “retrato pictórico”, ...
- Creador: “Carlos de Haes”, “Giuseppe Antonio Petrini”, ...
- País de origen: “Francia”, “Alemania”, ...
- Altura de la obra (en centímetros)
- Ancho de la obra (en centímetros)
- Material empleado: “papel”, “pintura al temple”, ...
- Qué representa: “paisaje”, “Coronación de la Virgen”, ...
- Colección: “Museo del Prado”, “Museo de El Greco”, ...
- Número de inventario o identificador de la obra dentro de la colección.
- Enlace de la obra al Museo del Prado

Puesto que los datos que Wikidata realmente devuelve son los identificadores de los ítems, para finalizar esta parte se obtienen, para cada resultado obtenido, los nombres de estos. Por ejemplo, Wikidata devuelve el identificador *Q610687* perteneciente a un movimiento, por lo que es necesario obtener el nombre del ítem. En este caso, se refiere a “Escuela veneciana”.

### 3. Obtención de los enlaces a Wikipedia a partir de cada obra de Wikidata.

Para obtener el nombre de las obras de Wikidata tal y como aparecen en Wikipedia, se realiza a través de la API REST<sup>36</sup>. La documentación referente a ésta se puede consultar en [82].

Por ejemplo, para obtener el nombre de la obra de Claudio de Lorena “El vado”, se realiza la consulta a través de su identificador de ítem de Wikidata:

[“https://www.wikidata.org/w/api.php?action=wbgetentities&props=sitelinks&format=json&language=es&sitefilter=eswiki&ids=Q27701629”](https://www.wikidata.org/w/api.php?action=wbgetentities&props=sitelinks&format=json&language=es&sitefilter=eswiki&ids=Q27701629)

Esta consulta devuelve, en formato JSON, el nombre de la obra “El vado (Claudio de Lorena)”. Así, se puede obtener el enlace a la obra en Wikipedia concatenando “https://es.wikipedia.org/wiki/” con el nombre obtenido:

[“https://es.wikipedia.org/wiki/El vado \(Claudio de Lorena\)”](https://es.wikipedia.org/wiki/El_vado_(Claudio_de_Lorena))

De forma automática, Wikipedia modifica el enlace cuando se hace la consulta a la web redirigiendo al enlace final de la obra, que en este caso sería:

[https://es.wikipedia.org/wiki/El vado \(Claudio de Lorena\)](https://es.wikipedia.org/wiki/El_vado_(Claudio_de_Lorena))

- Al finalizar este paso, se tienen almacenados los enlaces del Museo del Prado, Wikidata y Wikipedia que las obras que se han obtenido desde Wikidata.

### 4. Extracción de la información asociada a partir de cada obra de Wikipedia

Wikipedia permite obtener información de forma automática, ya no sólo de la obra, sino de referencias externas a las mismas (personajes, lugares, fechas, etc.). El proceso de obtención de la información, en este caso, se realiza obteniendo el HTML de cada enlace, y extrayendo los datos asociados al mismo.

Para obtener la información de la forma más completa posible se almacenan, en primer lugar:

1. Información en texto plano de cada obra.
2. Información en texto plano de cada sección de cada obra (acontecimientos, historia, etc.).
3. Información en formato HTML de cada obra.
4. Información en formato HTML de cada sección de cada obra (acontecimientos, historia, etc.).

Además, cada enlace encontrado dentro de las descripciones de las obras puede contener información valiosa para este trabajo, como personajes, fechas, lugares, y otros elementos. Por ello, se obtiene

---

<sup>36</sup> REST hace referencia a las interfaces que permiten conectar diferentes aplicaciones web utilizando el protocolo HTTP

también la información textual (en este caso, en texto plano) de cada uno de estos enlaces. Por ejemplo, partiendo de la descripción del cuadro *“La infanta Margarita”*<sup>37</sup>:

*“Este es el primer cuadro de los cinco que Velázquez realizó a la infanta Margarita Teresa de Austria, aunque en un principio se consideró que el personaje retratado era su hermanastra la infanta María Teresa de Austria.”*

Es posible obtener los enlaces a Wikipedia de los personajes Margarita Teresa de Austria y María Teresa de Austria. Por ello, se obtiene la información de las descripciones, en este caso, de los dos personajes y después se almacena una versión del texto de la obra modificado que contiene el enlace al identificador dentro del proyecto, lo que permite enlazar tanto los conceptos como sus descripciones.

En el caso de que el enlace obtenido se refiera a fechas (años), se obtienen desde Wikipedia una lista de eventos ocurridos en ese año (nacimientos, fallecimientos, guerras, etc.).

Además, se obtiene también la descripción textual referente a la técnica usada en el cuadro (por ejemplo, óleo) y al movimiento al que pertenece (por ejemplo, Manierismo).

## 5. Extracción de la información de las obras del Museo del Prado

La web del Museo del Prado, como ya se comentó anteriormente, contiene información estructurada de las obras y autores (metadatos), además de información no estructurada (descripciones de las obras y autores). La información se obtiene, o bien a través de los datos en formato JSON-LD (si existen), o bien extrayendo la información a partir del HTML.

Los datos obtenidos de cada obra son:

- Fecha de publicación de la obra (en el museo).
- Fecha de creación de la obra.
- Nombre de la obra.
- Enlace a la imagen de la obra.
- Número del catálogo o inventario (que coincide con el encontrado en Wikidata).
- Altura de la obra.
- Ancho de la obra.
- Técnica utilizada.
- Superficie utilizada.
- Procedencia de la obra.
- Palabras clave asociadas.
- Enlace a los autores de la obra.
- Descripción (en texto plano) de la obra.

Una vez obtenidos los datos referentes a las obras, se obtienen de forma similar los datos de los autores de estas:

- Nombre del autor.
- Fecha de nacimiento.

---

<sup>37</sup> [https://es.wikipedia.org/wiki/La\\_infanta\\_Margarita](https://es.wikipedia.org/wiki/La_infanta_Margarita)

- Lugar de nacimiento.
- Fecha de fallecimiento.
- Lugar de fallecimiento.
- Descripción (en texto plano) del autor.
- Palabras clave asociadas.

Además, en cada obra se buscan todos los enlaces de otras obras del Museo del Prado y se añade a la lista de la base de datos.

Los resultados obtenidos en esta parte del trabajo son un conjunto de datos organizados y preparados para procesarse. El total de datos en cada parte de detalla en la tabla 1:

Nombre	Total encontrados
Obras (Wikidata)	2061
Obras (Wikipedia)	371
Obras (Museo del Prado)	4830
Autores (Museo del Prado)	776
Técnicas (Wikipedia)	6
Movimiento/Estilo (Wikipedia)	12
Personajes (Wikipedia)	825
Referencias externas (Wikipedia)	2311
Año (Wikipedia)	301

Tabla 1 Número de elementos obtenidos de cada tipo en el proceso de extracción de información

## 6. Almacenamiento de los datos obtenidos

Los datos obtenidos se almacenan en el motor de indexación Solr teniendo en cuenta los tipos mostrados en la tabla 2:

Tipo de dato	Nombre
0	Obras (Wikidata)
1	Obras (Wikipedia)
2	Obras (Museo del Prado)
3	Autores (Museo del Prado)
4	Técnicas (Wikipedia)
5	Movimiento/Estilo (Wikipedia)
6	Personajes (Wikipedia)
7	Referencias externas (Wikipedia)
8	Año (Wikipedia)

Tabla 2 Tipos de colecciones de datos obtenidos en la etapa de extracción de información en Solr

Cada uno de estos tipos contienen una serie de metadatos asociados, que se almacenan teniendo en cuenta qué tipo de dato se obtiene en cada caso (si es una cadena de caracteres, un número entero, una fecha, etc.). Por ejemplo, los datos obtenidos desde el Museo del Prado sobre la obra “El lavatorio” de Jacopo Robusti Tintoretto son los mostrados en la Ilustración 21:

```

    "datePublished_d":2015.0,
    "name_s":"El lavatorio",
    "image_s":"https://content3.cdnprado.net/imagenes/Documentos/imgsem/77/77d1/77d1fd63-1918-40b7-a79e-6d427e19bed8/305bc206-418c-4fc8-b21f-bdcac970a7f4.jpg",
    "artEdition_s":"P002824",
    "id_author_is":[4402],
    "description_txt_es":"<p>En 1547 la Scuola del Santísimo Sacramento de la iglesia de San Marcuola en Venecia encargó a Jacopo Tintoretto <em>El Lavatorio</em> y una <em>Última Cena</em>, aún «in situ». Estas «scuole» estaban dedicadas a fomentar el culto a la Eucaristía, y en sus reuniones, el guardián ofrecía a los cofrades agua bendita, imitando el gesto de Jesús al lavar los pies a los apóstoles. Esta exaltación de la humildad era recurrente en textos devocionales de la época como<em> I Quattro libri de la humanita di Christo</em> de Pietro Aretino (1539), (...) </p>",
    "height_f":210.0,
    "dateModified_d":2018.0,
    "dateCreated_d":1549.0,
    "width_f":533.0,
    "procedencia_s":"San Marcuola, Venecia; colección de Carlos I de Inglaterra; adquirido en la Almoneda de la Commonwealth, de los bienes de Carlos I de Inglaterra, por Houghton, 1651, quien lo vendió en 1654 a don Alonso de Cárdenas para don Luis Méndez de Haro, quien lo regaló a Felipe IV; Colección Real (Salas Capitulares de El Escorial, 1656-1936); depositada en el Museo del Prado por la Junta Delegada de Incautación, Protección y Conservación del Tesoro Artístico Nacional, 1936; por Decreto de 02-03-1943 quedó establecido su depósito en el Museo del Prado",
    "artform_s":"Óleo",
    "keywords_ss":["Óleo",
    "Lienzo",
    "Jesús y los Apóstoles",
    "Perro (Canis lupus familiaris)",
    "Jesús",
    "San Pedro",
    "Vaso / copa",
    "Cuenco",
    "1548",
    "1549",
    "Tintoretto, Jacopo Robusti",
    "Pintura"],
    "type":[2],
    "id":"2529",
    "artworkSurface_s":"Lienzo",
    "_type_s":"VisualArtwork",
    "_context_s":"http://schema.org"

```

Ilustración 21 Ilustración que muestra los datos obtenidos de la obra "El Lavatorio" desde la web del Museo del Prado

Como se puede ver en la figura anterior, en este caso se ha obtenido información sobre la fecha de publicación de la obra en la web del museo, nombre de la obra, imagen, número de catálogo, identificador del autor, descripción, tamaño del lienzo, palabras clave, procedencia, técnica y superficie utilizada, fecha de creación de la obra, fecha de la última vez que se modificaron los datos

de la obra en la web del museo y el identificador de la obra. También es posible acceder a los datos obtenidos sobre el autor desde el museo, como muestra la Ilustración 22.

```

"name_s": "Tintoretto, Jacopo Robusti",
"keywords_ss": ["Tintoretto, Jacopo Robusti",
  "1518/19",
  "1594",
  "Escuela Italiana",
  "Escuela Veneciana"],
"p96_E67_p7_gave_birth_place_s": "Venecia",
"description_txt_es": "\nJacopo Comin. Conocido de joven como Jacopo Robusti.",
"url_s": "https://www.museodelprado.es/coleccion/artista/tintoretto-jacopo-robusti/134e0f91-ba9a-4fab-a3e4-d49573c318e7",
"p3_has_note_s": "\nJacopo Comin. Conocido de joven como Jacopo Robusti.",
"p96_E67_p4_gave_birth_date_d": 1518.0,
"p100i_E69_p7_death_place_s": "Venecia",
"p100i_E69_p4_death_date_d": 1594.0,
"type": [3],
"id": "4402"

```

Ilustración 22 Ilustración que muestra la información obtenida desde la web del Museo del Prado referente al autor "Jacopo Robusti Tintoretto"

Es importante resaltar que, en esta parte del proceso, se almacena la máxima cantidad de información posible de forma automática. Por ello, las claves de cada metadato se almacenan acordes a los campos dinámicos de Solr [83]. Estos campos sirven para indicar a Solr de forma automática cómo indexar cada metadato sin tener información previa sobre qué metadatos se obtienen. Por ejemplo, si el metadato contiene un texto en español, la clave del dato terminará en “\_txt\_es”, si es una cadena de caracteres, terminará en “\_s”; si es una lista de cadenas de caracteres, en “\_ss”, etc.

- Un problema que no se ha abarcado hasta este punto del trabajo es la relación entre los autores desde las distintas fuentes. Por ejemplo, una vez obtenidos los datos, se obtienen distintos identificadores para el mismo autor, como muestra la tabla 3:

id	url	type
4402	<a href="https://www.museodelprado.es/coleccion/artista/tintoretto-jacopo-robusti/134e0f91-ba9a-4fab-a3e4-d49573c318e7">https://www.museodelprado.es/coleccion/artista/tintoretto-jacopo-robusti/134e0f91-ba9a-4fab-a3e4-d49573c318e7</a>	3
8294	<a href="https://es.wikipedia.org/wiki/Tintoretto">https://es.wikipedia.org/wiki/Tintoretto</a>	6

Tabla 3 Fuentes desde las que se extrae información referente a " Jacopo Robusti Tintoretto"

## 4.2 PROCESAMIENTO E INTEGRACIÓN DE METADATOS Y TEXTOS DESCRIPTIVOS DESDE DISTINTAS FUENTES

Para la etapa de preprocesamiento de los datos se define una estructura de metadatos concreta que permita juntar toda la información de las obras a partir de las distintas fuentes e indexar en Solr los datos de forma correcta.

Siguiendo el modelo narrativo definido en este trabajo, los elementos narrativos que conforman la base son: personajes, obras, eventos y referencias. Por lo tanto, el siguiente paso es detallar de qué datos se dispone, cómo se almacenará la información de cada tipo y qué contienen.

## Referencias

Las referencias son todos los conceptos y/o enlaces asociados a las narrativas que aportan información complementaria sobre el cuadro, autor, personaje o evento. Por ejemplo, el cuadro “Niños con perros de presa”<sup>38</sup> pintado por Goya en 1786 muestra a dos niños llevando con sus correas a dos perros de presa. Las palabras clave que se pueden extraer de forma automática desde la web del Museo del Prado son:

- Comedor de los Príncipes de Asturias, Palacio de El Pardo, Madrid, 1786-87
- Óleo
- Lienzo
- Juegos y diversiones
- Perro (canis lupus familiaris)
- 1786
- Goya y Lucientes, Francisco de
- Cartón para tapiz

Como se puede ver, estas etiquetas, consideradas como referencias en este modelo narrativo, aportan información semántica sobre el cuadro. Esta información permite, entre otras cosas, establecer relaciones sobre otros cuadros que guarden similitudes (por ejemplo, permite encontrar otros cuadros en los que aparezcan perros).

Otros tipos de referencias que podemos encontrar son enlaces a vídeos, audiodescripciones, libros, revistas, artículos de prensa, etc. Este tipo de referencias se utilizarán como material complementario en las narrativas, pero no se obtiene información semántica de los mismos.

## Obras

La selección de los metadatos que se realiza pretende agrupar toda la información disponible desde las tres fuentes de información, dando prioridad a la información extraída desde Wikidata y Wikipedia, y complementándola con la información de la web del Museo del Prado.

Los metadatos asociados a cada obra son:

- Nombre
- Identificador del cuadro del Museo del Prado
- Identificador del cuadro de Wikipedia
- Identificador del cuadro de Wikidata
- Género artístico
- Entidad de Wikidata de la que es instancia (“cuadro”)
- Lista de palabras clave de la web del Museo del Prado
- Lista de palabras clave de Wikipedia
- Material empleado
- Movimiento
- País de origen
- Procedencia del cuadro

---

<sup>38</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/nios-con-perros-de-presa/4168eefd-fdce-4b9b-8df8-275dca330287?searchid=44c08032-3156-74a9-b44d-fd7d3cac26e6>



- Superficie
- Técnica
- Ubicación
- Imagen
- Estilo
- Alto
- Ancho
- Lista de identificadores de autores
- Identificador del catálogo
- Colección

Además de los metadatos, se almacenan dos descripciones de cada obra:

- Descripción (texto plano) de la obra del Museo del Prado
- Descripción (con enlaces y dividida en secciones) de la obra en Wikipedia

### Personajes

Los personajes, en este caso, se refiere tanto a los autores de las obras como a todos los personajes que aparecen de forma indirecta en los mismos. Los metadatos básicos asociados a éstos son:

- Nombre
- Fecha de nacimiento
- Fecha de fallecimiento
- Lugar de nacimiento

Como ocurre en el caso anterior, también se obtienen dos descripciones asociados a cada personaje:

- Descripción (texto plano) del personaje del Museo del Prado (si es autor)
- Descripción (en texto plano y dividida en secciones) del personaje en Wikipedia

### Eventos

Los eventos se asocian a los años que aparecen de forma directa (por ejemplo, el año de creación de una obra) o indirecta (por ejemplo, en la descripción de Wikipedia) de las obras.

Los eventos pueden ser de cuatro tipos:

- Nacimientos
- Fallecimientos
- Acontecimientos
- Creación de obra
- Inicio o finalización de una guerra

La información de los eventos que no se extraen de las obras ni personajes, se extrae directamente desde Wikipedia, buscando todo lo referente a ese año. Por ejemplo, el mostrado en la Ilustración 23:

<https://es.wikipedia.org/wiki/1423>

# 1423

---

## Acontecimientos [\[ editar \]](#)

---

- Nuno Álvares Pereira ingresa en la orden [carmelita](#).
- En la actual España, el rey [Carlos III de Navarra](#) instituye el título de [príncipe de Viana](#).
- [14 de noviembre](#) (11/10/30, según el calendario [Oei](#)): en [Ugo](#) ([Japón](#)) a las 8:00 (hora local) se registra un terremoto de 6,5 grados en

## Nacimientos [\[ editar \]](#)

---

- [3 de julio](#): [Luis XI](#), rey francés entre 1461 y 1483.

## Fallecimientos [\[ editar \]](#)

---

- [Per Afán de Ribera el Viejo](#), adelantado mayor de la frontera de Andalucía y notario mayor de Andalucía.


*Ilustración 23 Captura extraída de Wikipedia sobre los eventos asociados al año 1423*

Por motivos de eficiencia, toda la información en este caso se almacena en texto plano, por lo que no se tiene información semántica (estructurada) sobre los eventos salvo que hayan sido extraídos a partir de obras o personajes.

### 4.2.1 Enriquecimiento automático de la red conceptual utilizando las herramientas LOD presentes en Wikipedia y Wikidata

En el punto anterior se definieron los distintos elementos narrativos presentes en el trabajo, sin embargo, es necesario obtener la información semántica de los conceptos de las descripciones de las obras y autores, y de las palabras clave no enlazadas directamente desde las fuentes de información.

Wikipedia aporta, gracias a sus datos enlazados, la información semántica en las descripciones de los conceptos existentes. Por ejemplo, en la página de [El Lavatorio](#), de [Tintoretto](#), se pueden ver una serie de conceptos enlazados, como muestra la [Ilustración 24](#).

-  Como se puede ver, gracias al paradigma de datos enlazados presente en Wikipedia y Wikidata, es posible acceder directamente a la descripción de cada concepto que se encuentra en las obras y personajes, lo que permite aportar mayor riqueza a la información obtenida.

Cada una de las páginas a las que enlazan contienen una descripción del concepto, y, además, es posible obtener el enlace a la página de Wikidata a la que hace referencia. Esto permite, también, obtener la información sobre a qué tipo de elemento se refiere (a partir del metadato presente en Wikidata “[instancia\\_de](#)”). Es decir, es posible encontrar información acerca de los personajes que guardan alguna relación con la obra, porque aparecen enlazados en la descripción y, además, en su página de Wikidata son una instancia de “[human](#)” (humano).

Dado al valor añadido que aporta la posibilidad de obtener esta información, se almacena en la base de datos la información enlazada de las descripciones de las obras y los autores, además de las descripciones de esos conceptos y el nombre de la instancia a la que pertenecen. Así, se pueden categorizar automáticamente como personajes a todos los que se encuentren en estas definiciones y los autores de las obras. También, se encuentran los eventos asociados a las obras cuando las instancias se refieren a fechas. Del resto de referencias se almacena la información sobre su descripción, nombre y a qué instancia hace referencia, para en un futuro poder realizar búsquedas

complejas sobre las mismas (por ejemplo, la de obtener todos los segmentos narrativos que hagan referencia al género pictórico).

## El Lavatorio (Tintoretto)



Este artículo o sección necesita **referencias** que aparezcan en una **publicación acreditada**.

Puedes avisar al redactor principal pegando lo siguiente en su página de discusión: `{{sust:Aviso referencias|El Lavatorio}}`

Este aviso fue puesto el 18 de noviembre de 2017.

**El Lavatorio** es un cuadro del pintor italiano **Tintoretto**, considerado una de sus obras más importantes. Fue pintado entre 1548-1549, en España, como depósito de **Patrimonio Nacional**.

Se representa aquí una escena narrada por el **Evangeli**o de Juan, en la que se describe cómo durante la **Última Cena**, Jesús se levantó y echó **agua** en un recipiente, se puso a lavar los pies a sus discípulos. **Simón Pedro** pretendió negarse, pero al insistir Jesús en que de o

El artista representa el episodio antedicho con Cristo y san Pedro en un extremo de la composición. La mayor parte del lienzo está ocupada por los discípulos en torno a ella. En el centro destaca un **perro**, y detrás los apóstoles descalzándose o en diversas posturas y **escorzos**. En un **Jacobo de la Vorágine** en su **Leyenda dorada**. El extremo de la izquierda está dominado por otro **apóstol** que se está desatando el calzado. **Ángel**.

La composición parece descentrada, con el episodio principal desplazado a un lado del cuadro. Esto se explica por el emplazamiento original: los creyentes verían más cerca precisamente la parte donde estaba Jesús. Además, la mesa está orientada hacia esa zona, de modo que viéndose desde una **perspectiva**. A ello también contribuye el pavimento de losas con formas geométricas.

*Ilustración 24 Imagen extraída de Wikipedia sobre el cuadro El Lavatorio, de Tintoretto<sup>39</sup>*

**Enriquecimiento automático de eventos: guerras y nacimientos y fallecimientos de personajes.**

Hasta ahora, los eventos existentes en la red conceptual son:

- Nacimientos y fallecimientos de autores
- Creación de obras
- Acontecimientos ocurridos en cada año (en texto plano, sin información semántica)

Por ello, en este paso del proceso se pretende añadir información que pueda resultar de interés a la hora de contextualizar las obras o sus descripciones.

En primer lugar, utilizando la API SPARQL de Wikidata, se busca información acerca de las guerras, como ejemplo de eventos de interés general a través del tiempo. Para obtener resultados de calidad, que puedan resultar útiles, la búsqueda se acota a las guerras de las que se disponga información sobre, como mínimo, su fecha de inicio y ubicación. La consulta SPARQL generada es la mostrada en la Ilustración 25<sup>40</sup>.

```
SELECT * WHERE {
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],ES". }
  ?guerra wdt:P31 wd:Q198.
  OPTIONAL { ?guerra wdt:P31 ?instancia_de. }
  ?guerra wdt:P580 ?fecha_de_inicio.
  OPTIONAL { ?guerra wdt:P582 ?fecha_de_fin. }
}
```

*Ilustración 25 Consulta SPARQL para obtener la información sobre guerras de Wikidata*

<sup>39</sup> [https://es.wikipedia.org/wiki/El\\_Lavatorio\\_\(Tintoretto\)](https://es.wikipedia.org/wiki/El_Lavatorio_(Tintoretto))

<sup>40</sup> <http://tinyurl.com/y77gotzu>

La información se almacena como eventos de tipo “guerra”. En total, se almacena información de 221 guerras. Una muestra de las mismas se puede ver en la Ilustración 26.

Guerra de los Treinta Años - 1618  
Guerra de Babilonia - 0310  
Guerras berberiscas - 1801  
Guerra civil camboyana - 1967  
Guerra entre Etiopía y Eritrea - 1998  
Guerras Indias - 1622  
Guerra del rey Felipe - 1675  
Guerra Civil de Laos - 1953  
Guerra de Halcón Negro - 1832  
Guerra negra - 1820  
Crisis samoana - 1886  
Conquista musulmana de Persia - 0633  
Invasiones de tártaros en Rusia en el siglo XVI - 1507  
Conquista musulmana del Magreb - 0643  
Primera Guerra Siria - 0273  
Guerra polaco-lituano-teutónica - 1409  
Guerras del Bacalao - 1958  
Tercera Guerra Carlista - 1872  
Guerra de Abjasia de 1998 - 1998  
Guerra turco-veneciana - 1570  
Guerras de Carnatic - 1746

*Ilustración 26 Lista de guerras existentes en Wikidata*

Además de la información sobre las guerras, se almacena información sobre los nacimientos y fallecimientos de todos los personajes existentes en la base de datos. Dado que ya está almacenada dicha información como metadatos, el proceso consiste en crear los eventos a partir de los mismos. Algunos datos de ejemplo aparecen en la Ilustración 27.

1802 - Fallecimiento de María Teresa de Silva Álvarez de Toledo en Madrid en 23 July 1802  
1762 - Nacimiento de María Teresa de Silva Álvarez de Toledo en Madrid en 10 June 1762  
1802 - Fallecimiento de María Teresa de Silva Álvarez de Toledo, duquesa de Alba en Madrid en 23 July 1802  
1762 - Nacimiento de María Teresa de Silva Álvarez de Toledo, duquesa de Alba en Madrid en 10 June 1762  
1523 - Fallecimiento de David, Gérard en Brujas en 13 August 1523  
1661 - Fallecimiento de Luis Méndez de Haro, Madrid en Madrid en 26 November 1661  
1598 - Nacimiento de Luis Méndez de Haro, Madrid en Valladolid en 17 February 1598  
1576 - Fallecimiento de Tiziano en Venecia en 6 September 1576  
1488 - Nacimiento de Tiziano en Pieve di Cadore en 1488

*Ilustración 27 Lista de nacimientos y fallecimientos de personajes*

#### 4.2.2 Enriquecimiento automático de la red conceptual a partir de las obras del Museo del Prado

El problema encontrado en la web del Museo del Prado es que esta información aparece como texto plano, por lo que no se tiene, en principio, la relación directa con la información semántica que aporta. Por ello, antes de crear la red de conceptos, se lleva a cabo un paso previo que consiste en extraer toda la información posible de forma automática a partir de las palabras clave de las obras del museo.

Para solventar este problema, se realiza una búsqueda de cada concepto en distintas fuentes. El algoritmo utilizado en este caso sigue los siguientes pasos mostrados en la Ilustración 28.

```
Buscar_informacion_de_concepto (label) :  
  if existe_entidad_en_bbdd(label):  
    informacion_asociada = obtener_datos(label)  
  else:  
    if existe_concepto_en_solr(label):  
      informacion_asociada = obtener_datos(label)  
    else:  
      if existe_concepto_en_wikidata(label):  
        informacion_asociada = obtener_datos(label)  
  return informacion_asociada
```

Ilustración 28 Pseudocódigo que muestra el proceso de búsqueda de nuevas relaciones entre conceptos previamente desconocidos

Siguiendo el ejemplo obtenido en este paso referente a la obra de *El Lavatorio*, de *Tintoretto*<sup>41</sup>, el primer concepto desconocido es “*San Marcuola, Venecia*”, y se puede encontrar entre los lugares en los que ha estado la obra expuesta a lo largo del tiempo (procedencia).

1. En primer lugar, el sistema busca el concepto entre los que ya se han extraído previamente de Wikidata a partir de los textos de todas las obras, autores y personajes de Wikipedia. En este caso, no se encuentra, dado que no es un concepto que haya aparecido hasta ahora.
2. Dado que no se ha podido obtener la información, realiza una nueva búsqueda en Solr, donde pueden aparecer además todos los conceptos y las referencias, añadiendo los encontrados en el Museo del Prado que cuenten con información de datos enlazados. En este caso tampoco encuentra resultados.
3. Si no encuentra información asociada en los casos anteriores, la última opción es buscar el concepto directamente a través de la API de Wikidata. La búsqueda se realiza a través del mismo servicio web detallado en los puntos anteriores, pero realizando una búsqueda exacta del concepto, a partir de la siguiente URL:

<https://en.wikipedia.org/w/api.php?action=query&list=search&srsearch=San%20Marcuola&srwhat=nearmatch>

---

<sup>41</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/el-lavatorio/77d1fd63-1918-40b7-a79e-6d427e19bed8>

- Para hacer las búsquedas más exactas y evitar ambigüedades, se toma la decisión de buscar utilizando solamente los caracteres que se encuentran antes del carácter “,”, y eliminando la información entre paréntesis.
4. En este caso, la búsqueda sí arroja resultados, como muestra la Ilustración 29.

```
{
  "query": {
    "search": [
      {
        "ns": 0,
        "title": "San Marcuola",
        "pageid": 36358142,
        "size": 3332,
        "wordcount": 382,
        "snippet": "[[File:Chiesa di San Marcuola interno Canal Grande Venezia.jpg|thumb|left|The interior of San Marcuola]]<br>{{Infobox religious building",
      }
    ]
  }
}
```

Ilustración 29 Resultados extraídos de la consulta mostrada anteriormente

Como se resalta en la figura anterior, la búsqueda devuelve el identificador de la página en Wikipedia que contiene información acerca del concepto buscado, por lo que, a partir del mismo, se puede acceder a la descripción del concepto y a su enlace en Wikidata (para conocer de qué elemento es instancia). En este caso, es una instancia de “iglesia”, como muestran las ilustraciones 30<sup>42</sup> y 31<sup>43</sup>.

Como se puede ver en este caso, no existe información en español de este concepto, por lo que la descripción no se almacena en este caso.

## San Marcuola (Q1585514)

church

[In more languages](#) Configure

Language	Label	Description
English	San Marcuola	church
Spanish	No label defined	No description defined
Catalan	No label defined	No description defined
Galician	No label defined	No description defined

[All entered languages](#)

### Statements

instance of	church
-------------	--------

[0 references](#)

Ilustración 30 Imagen extraída de Wikidata referente al concepto "San Marcuola"

<sup>42</sup> <https://www.wikidata.org/wiki/Q1585514>

<sup>43</sup> <https://en.wikipedia.org/?curid=36358142>

# San Marcuola

From Wikipedia, the free encyclopedia



The church of **San Marcuola**, a church dedicated to the saints [Hermagoras and Fortunatus](#) ("Marcuola" is a [Venetian Grand Canal](#) and located in the [sestiere](#) of [Cannaregio](#) in [Venice](#), Italy.

The present church was first erected in the 12th century. Major reconstruction was designed by [Antonio Gaspari](#), never completed. The church has a large collection of statues by [Gaetano Susali](#), and paintings by [Francesco Missoni](#).

For the right side, [Tintoretto](#) painted *Christ Washing the Disciples' Feet* but it is now at either [Museo del Prado](#) in Madrid. Its place at San Marcuola is occupied now by a copy by [Carlo Ridolfi](#).

The grave of German composer [Johann Adolf Hasse](#) (1699-1783) is located in San Marcuola.

The church gives its name to the [San Marcuola vaporetto](#) stop on the Grand Canal.

*Ilustración 31 Imagen extraída de Wikipedia referente al concepto "San Marcuola"*

## 4.2.3 Enriquecimiento automático de la red conceptual integrando los autores del Museo del Prado con sus datos en Wikipedia y Wikidata

Como se mencionó anteriormente, un problema a abordar es que no se tiene relación previa entre los autores de los cuadros del Museo del Prado y los mismos en Wikipedia o Wikidata. Se pueden dar los siguientes casos:

1. Existe enlace de la obra a Wikidata y el Museo del Prado, y contiene un solo autor.
2. Existe enlace de la obra a Wikidata y el Museo del Prado, y contiene varios autores.
3. No existe enlace a Wikidata.

En el primer caso, el enlace es directo. En el segundo caso, es necesario buscar qué autor de la web del museo se refiere a cuál de Wikidata. Dado que los nombres de los autores extraídos desde el Museo del Prado suelen tener un formato "apellidos, nombre" (como los ejemplos mostrados en la siguiente figura), se realiza una búsqueda primero con el nombre sin modificar, y, en caso de no obtener resultados, se cambia el orden y se elimina el carácter ",". Se puede ver un ejemplo en la Ilustración 32.

```
{"name_s": "Escalante, Juan Antonio de Frías y"},  
{"name_s": "Coello, Claudio"},  
{"name_s": "Murillo, Bartolomé Esteban"},  
{"name_s": "Tintoretto, Jacopo Robusti"},  
{"name_s": "Velázquez, Diego Rodríguez de Silva y"},  
{"name_s": "Rafael"},  
{"name_s": "El Bosco"},  
{"name_s": "El Greco"},  
{"name_s": "Tiziano, Vecellio di Gregorio"},  
{"name_s": "Van Dyck, Antonio"}
```

*Ilustración 32 Extracción de nombres de autores de la Web del Museo del Prado a través de Solr*

En caso de no obtener la información, se utiliza una función auxiliar que busca la página de Wikipedia más similar al concepto dado, para obtener su información, utilizando la API de Wikidata:

<https://es.wikipedia.org/w/api.php?action=query&list=search&format=json&srwhat=nearmatch&srsearch=concepto>

```

Buscar concepto en Wikidata y Wikipedia (concepto, idioma="es"):
  consulta_api = url api con concepto
  si consulta_api no devuelve datos -> self (concepto_sin_parántesis,idioma)
  si existe redirección -> Buscar concepto en Wikidata y Wikipedia (redireccion, idioma)
  si consulta_api devuelve datos:
    url_wikipedia = 'https://' + idioma + '.wikipedia.org/wiki/' + concepto
    return Obtener y almacenar datos de Wikidata y Wikipedia(url_wikipedia)
  return False

```

*Ilustración 33 Búsqueda de un concepto en Wikidata y Wikipedia*

El algoritmo, mostrado en la Ilustración 33, busca primero el concepto en español. En caso de no encontrar resultados, modifica el concepto (elimina paréntesis) y vuelve a realizar la búsqueda. Si la consulta devuelve una página que contiene una redirección, vuelve a realizar la búsqueda con la redirección. Finalmente, en caso de devolver la página de Wikipedia encontrada, extrae todos los elementos asociados de Wikipedia y Wikidata.

Por ejemplo, si buscamos el concepto “Diego\_Rodríguez\_de\_Silva\_y\_Velázquez”, se obtienen como resultados indicando que la página en Wikipedia es una redirección a:

[https://es.wikipedia.org/wiki/Diego\\_Velázquez](https://es.wikipedia.org/wiki/Diego_Velázquez)

Por lo que así se puede obtener el nombre del autor en Wikidata, y todos los datos relativos en las dos fuentes de información.

Se pueden ver dos ejemplos de resultados obtenidos en la tabla 4.

Nombre	Enlace al Museo del Prado	Enlace a Wikipedia
El Bosco	<a href="https://www.museodelprado.es/coleccion/artista/el-bosco/c9716e4a-4c24-44dd-ac65-44bc4661c8b5">https://www.museodelprado.es/coleccion/artista/el-bosco/c9716e4a-4c24-44dd-ac65-44bc4661c8b5</a>	<a href="https://es.wikipedia.org/wiki/Jheronimus_Bosch">https://es.wikipedia.org/wiki/Jheronimus_Bosch</a>
El Greco	<a href="https://www.museodelprado.es/coleccion/artista/el-greco/b031da57-6a7e-43f2-a855-293275efc340">https://www.museodelprado.es/coleccion/artista/el-greco/b031da57-6a7e-43f2-a855-293275efc340</a>	<a href="https://es.wikipedia.org/wiki/El_Greco">https://es.wikipedia.org/wiki/El_Greco</a>

*Tabla 4 Enlaces obtenidos entre dos autores distintos entre Wikipedia y el Museo del Prado*

#### 4.2.4 Descubrimiento de información automática de las descripciones de las obras del Museo del Prado

Las obras del Museo del Prado se obtienen en texto plano, por ello, es necesaria una etapa de anotación automática de los mismos. Para ilustrar el proceso de anotación llevado a cabo en este trabajo, se parte con el ejemplo de la obra “El Lavatorio”, de Tintoretto, cuya descripción se puede ver en la página web<sup>44</sup>.

<sup>44</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/el-lavatorio/77d1fd63-1918-40b7-a79e-6d427e19bed8>



La descripción HTML ofrece, gracias a las etiquetas “<em>” que contiene, conceptos candidatos a ser procesados como conceptos, pero no información semántica de los mismos, como se puede ver en la Ilustración 34:

```
<p>En 1547 la Scuola del Santísimo Sacramento de la iglesia de San Marcuola en Venecia encargó a Jacopo Tintoretto <em>El Lavatorio</em> y una <em>Última Cena</em>, aún «in situ». Estas «scuole» estaban dedicadas a fomentar el culto a la Eucaristía, y en sus reuniones, el guardián ofrecía a los cofrades agua bendita, imitando el gesto de Jesús al lavar los pies a los apóstoles. Esta exaltación de la humildad era recurrente en textos devocionales de la época como<em> I Quattro libri de la humanita di Christo</em> (...) </p>
```

*Ilustración 34 Información resaltada en los textos del Museo del Prado*

En este caso, aparecen tres posibles conceptos que pueden corresponder con algún tipo de elemento narrativo. Por otro lado, como se puede observar, hay otra serie de conceptos que no aparecen etiquetados pero que ya han sido anotados previamente en descripciones de Wikipedia, como “1547”, “San Marcuola” o “Venecia”. Así que este proceso tiene como objetivo principal encontrar la mayor cantidad de elementos narrativos y etiquetarlos correctamente.

Para lograr este objetivo, se aprovechará toda la información del grafo semántico obtenido con Wikipedia, donde ya se han encontrado la mayor parte de los conceptos importantes, poniendo en evidencia las ventajas que ofrece la información de datos enlazados en abierto para etiquetar los textos. Además, se utilizará el identificador de entidades nombradas de la Universidad de Stanford [84] para buscar nuevos conceptos. Finalmente, se compararán los conceptos obtenidos con Wikipedia frente a los obtenidos con la herramienta de Stanford.

#### *Extracción de información semántica a partir de elementos ya conocidos*

En primer lugar, se realiza una búsqueda de cada concepto etiquetado en la descripción del Museo del Prado en el grafo semántico, buscándolos como obras, personajes, eventos o referencias. En caso de no encontrarlo, busca el concepto en Wikipedia y crea la referencia. Este proceso es el mismo que se utilizó anteriormente para extraer información semántica de las palabras clave de las obras del Museo del Prado, descrito anteriormente.

Siguiendo con el ejemplo anterior, en este proceso encuentra varios resultados, como muestra la Ilustración 35.

```
<p>En 1547 la Scuola del Santísimo Sacramento de la iglesia de San Marcuola en Venecia encargó a Jacopo Tintoretto <a id_artwork="3303">El Lavatorio</a> y una <a id_reference="8830">Última Cena</a>, aún «in situ». Estas «scuole» estaban dedicadas a fomentar el culto a la Eucaristía, y en sus reuniones, el guardián ofrecía a los cofrades agua bendita, imitando el gesto de Jesús al lavar los pies a los apóstoles. Esta exaltación de la humildad era recurrente en textos devocionales de la época como <a id_reference="10989">I Quattro libri de la humanita di Christo</a> (...) </p>
```

*Ilustración 35 Información anotada de los textos del Museo del Prado*

En segundo lugar, para complementar la información, se crea una lista con todos los elementos narrativos existentes hasta el momento. Después, se busca cada uno de esos conceptos en los textos, y se etiquetan los que se encuentren.

En este paso del proceso, no se tienen en cuenta los eventos, ya que el evento correspondiente a cada fecha dependerá de la obra en cuestión. Por ello, el etiquetado de fechas se realizará en el siguiente paso del proceso.

Siguiendo con el ejemplo anterior, en este proceso encuentra varios resultados, como muestra la Ilustración 36.

En 1547 la Scuola del Santísimo Sacramento de la **iglesia** de **San Marcuola** en **Venecia** encargó a Jacopo **Tintoretto** **El Lavatorio** y una Última Cena, aún «in situ». Estas «scuole» estaban dedicadas a fomentar el culto a la Eucaristía, y en sus reuniones, el guardián ofrecía a los cofrades agua bendita, imitando el gesto de Jesús al lavar los pies a los apóstoles. Esta exaltación de la humildad era recurrente en textos devocionales de la época como I Quattro libri de la humanita di Christo (...)

*Ilustración 36 Información anotada de los textos del Museo del Prado*

- Como se puede ver en este ejemplo, la suma de ambos etiquetados produce una anotación muy completa, donde todos los elementos narrativos principales quedan recalcados, además de los elementos resaltados por el Museo del Prado (escribiendo en negrita las palabras clave).

### *Descubrimiento de entidades nombradas usando la herramienta Stanford CoreNLP.*

El conjunto de herramientas para Procesamiento del Lenguaje Natural (Stanford CoreNLP) [78] que ofrece la Universidad de Stanford, sirve en este trabajo como ayuda para poder reconocer automáticamente entidades nombradas en los textos. Esto permite descubrir nueva información que, a priori, no se conocía. Esta herramienta dispone, además, de una versión en español<sup>45</sup>, capaz de reconocer expresiones idiomáticas (por ejemplo, “no obstante”), nombres de personas, fechas y expresiones cuantificadoras. Por ello, puede aportar mucha información relevante para los textos, como nuevos personajes, eventos o referencias.

Dado que la extracción de información que realiza la herramienta desarrollada en su versión en Python<sup>46</sup> se realiza palabra a palabra (por lo que palabras clave de mayor longitud quedarían divididas), se implementa un proceso que une las entidades nombradas encontradas por la herramienta. Por ejemplo, en lugar de tener dos entidades “Jacopo” y “Tintoretto”, se encuentra como entidad de tipo PERSON a “Jacopo Tintoretto”.

<sup>45</sup> <https://nlp.stanford.edu/software/spanish-faq.shtml>

<sup>46</sup> <https://github.com/Lynten/stanford-corenlp>

Dado el texto mostrado anteriormente, encuentra las entidades, como muestra la Ilustración 37.

```
[['1547', 'NUMBER'],  
 ['Scuola', 'ORGANIZATION'],  
 ['Santísimo Sacramento', 'ORGANIZATION'],  
 ['San', 'CITY'],  
 ['Marcuola', 'LOCATION'],  
 ['Venecia', 'CITY'],  
 ['Jacopo Tintoretto El Lavatorio', 'PERSON'],  
 ['Última Cena', 'ORGANIZATION'],  
 ['Eucaristía', 'ORGANIZATION'],  
 ['Jesús', 'PERSON'],  
 ['I Quattro libri', 'ORGANIZATION'],  
 ['Christo', 'ORGANIZATION'],
```

*Ilustración 37 Entidades nombradas existentes en una descripción.*

Las expresiones encontradas con esta herramienta sirven de guía para decidir si se encuentran, o no, como elemento narrativo en el texto. Por ello, el siguiente paso es procesarlas dependiendo de su tipo:

- Si corresponden a nombres de persona (*'PERSON'*).

En este caso, primero se realiza una búsqueda de la expresión exacta en la base de datos. Si se encuentra, y corresponde a un personaje, se etiqueta como el personaje encontrado. En caso de no encontrarlo, se realiza una búsqueda en Wikipedia a partir del nombre, como se ha realizado en los puntos anteriores. Si se encuentran resultados, y además corresponden al identificador de Wikidata “Q5” (“ser humano”), se añade a la tabla de personajes y se etiquetan todas las ocurrencias que aparezcan en el texto con esa expresión exacta.

En caso de no encontrar resultados, no se etiqueta la entidad.

- Si corresponden a números (*'NUMBER'*) o a fechas (*'DATE'*).

En este caso, y para simplificar el trabajo, sólo se tienen en cuenta los casos que correspondan a años (“1956”, “1970”, ...). Como ocurre en la búsqueda de eventos, teniendo en cuenta el identificador de obra y la fecha, se busca un evento que haga referencia al mismo. Más adelante, se creará un proceso que enriquezca los datos que falten, como, por ejemplo, si esa fecha corresponde con el nacimiento o fallecimiento del autor de esa obra.

- En el resto de los casos.

En este caso, se realiza la misma búsqueda que si fuera un personaje, pero buscando elementos narrativos que sean referencias. Sólo se creará una referencia si ya existe previamente, o se encuentra en Wikipedia.

Siguiendo el ejemplo comentado en este punto, la anotación automática quedaría:

En 1547 la Scuola del [Santísimo Sacramento](#) de la iglesia de San Marcuola en [Venecia](#) encargó a Jacopo [Tintoretto](#) El Lavatorio y una Última Cena, aún «in situ». Estas «scuole» estaban dedicadas a fomentar el culto a la Eucaristía, y en sus reuniones, el guardián ofrecía a los cofrades agua bendita, imitando el gesto de Jesús al lavar los pies a los apóstoles. Esta exaltación de la humildad era recurrente en textos devocionales de la época como I Quattro libri de la humanita di [Christo](#) (...)

*Ilustración 38 Texto anotado del Museo del Prado*

Como se puede ver en la Ilustración 38, la ventaja de utilizar este anotador y las funcionalidades incorporadas es que encuentra nuevas referencias desconocidas previamente, y las filtra aprovechando Wikipedia y Wikidata, por lo que se añaden finalmente los elementos narrativos que aporten información relevante.

#### *Selección de palabras clave y anotación automática de los textos del Museo del Prado*

Una vez obtenida una lista completa de palabras clave, catalogadas según el elemento narrativo al que se refiere en cada caso, se procede a la selección de las palabras que se anotarán en el texto.

En primer lugar, se descartan como palabras clave las entidades que no es posible encontrar en Wikidata (por ejemplo, las oraciones en letra negrita que se obtenían de los textos del Museo del Prado). Esta información puede ser de interés para el lector, pero no aporta información semántica a la red.

Al descartar estos datos, siguiendo con el ejemplo anterior, se obtienen los siguientes resultados, como muestra la Ilustración 39.

`<p>En 1547 la Scuola del Santísimo Sacramento de la iglesia de San Marcuola en Venecia encargó a Jacopo Tintoretto El Lavatorio y una Última Cena, aún «in situ». Estas «scuole» estaban dedicadas a fomentar el culto a la Eucaristía, y en sus reuniones, el guardián ofrecía a los cofrades agua bendita, imitando el gesto de Jesús al lavar los pies a los apóstoles. Esta exaltación de la humildad era recurrente en textos devocionales de la época como I Quattro libri de la humanita di Christo (...)`

*Ilustración 39 Texto anotado del Museo del Prado.*

#### 4.2.5 Resultados obtenidos de los puntos anteriores

Hasta este momento, el trabajo se centra en encontrar, obtener y procesar toda la información relativa a los elementos narrativos referentes a las obras del Museo del Prado. La información de las descripciones (tanto las obtenidas desde Wikipedia como las obtenidas desde el Museo del Prado) de las obras se anota automáticamente siguiendo el modelo narrativo propuesto en este trabajo. Por ejemplo, los textos obtenidos y anotados desde el Museo del Prado y Wikipedia de la obra “Los borrachos” de Diego Velázquez son los mostrados en las ilustraciones 40 y 41.

```
<p>El pago, en julio de 1629, de 100 <a id_reference="2600">ducados</a> a <a id_character="11543">Velázquez</a> <em>por cuenta de una pintura de <a id_reference="8872">Baco</a> </em>que había hecho para el rey nos informa sobre la fecha aproximada de la obra e identifica a su destinatario. Estamos en las vísperas del primer viaje del <a id_reference="2715">pintor</a> a Italia, cuando llevaba poco más de un lustro de trabajo al servicio del rey y acababa de conocer a Rubens. En esa época se estaba especializando en la pintura de retratos, aunque hacía poco que había realizado un afamado <a id_reference="1945">cuadro</a> de carácter histórico, <em>La expulsión de los moriscos, </em> (...)</p>
```

*Ilustración 40 Texto anotado del Museo del Prado sobre "Los Borrachos" de Velázquez*

```
<p>El triunfo de Baco es una <a id_reference="1961">pintura</a> del <a id_reference="1967">español</a> <a id_character="203">Velázquez</a>, conservada en el <a id_reference="1942">Museo del Prado</a> y creada en <a id_event="12133">1629</a>. Es conocida popularmente como Los borrachos.</p><p>El cuadro lo pintó algunos años después de su llegada a <a id_reference="8772">Madrid</a> (...)</p>
```

*Ilustración 41 Texto anotado de Wikipedia sobre "Los Borrachos" de Velázquez*

Además de las descripciones de los textos, se obtienen los metadatos de las obras, y las descripciones en texto plano de los elementos narrativos (personajes, referencias y eventos) obtenidos en el proceso.

Los metadatos obtenidos de las obras se almacenan en Solr, creando previamente los tipos de datos que permiten la optimización de futuras consultas. Por otro lado, las descripciones en texto plano de los elementos narrativos obtenidos de Wikipedia se almacenan en Solr, utilizando los tipos de datos dinámicos para catalogar automáticamente las secciones que tengan en cada caso. Por ejemplo, la descripción de la referencia “Parque del Retiro de Madrid”<sup>47</sup>, cuenta con una descripción y un índice que divide todo el texto en distintas secciones, como muestra la Ilustración 42.

En este caso, se muestra un extracto de textos obtenidos y almacenados en Solr, como muestra la Ilustración 43.

<sup>47</sup> [https://es.wikipedia.org/wiki/Parque\\_del\\_Retiro\\_de\\_Madrid](https://es.wikipedia.org/wiki/Parque_del_Retiro_de_Madrid)

Índice [ocultar]	
1	Toponimia
2	Marco legislativo
3	Historia
3.1	Siglo XVII
3.2	Siglo XVIII
3.3	Siglo XIX
3.4	Siglos XX y XXI
3.5	Evolución histórica en la pintura
4	Arquitectura
4.1	Puertas monumentales
4.2	Edificios

Ilustración 42 Índice extraído de Wikipedia que muestra las distintas secciones textuales del Parque del Retiro

```
"toponimia_string_txt_es": "Los topónimos de Retiro y Buen Retiro hacen referencia (...),
"marco_legislativo_string_txt_es": "El parque del Retiro está protegido desde 1935 (...),
"arquitectura_string_txt_es": "El Parque del Retiro cuenta con diecisiete entradas (...),
"senda_botanica_y_arboles_singulares_string_txt_es": "Con el nombre de Senda Botánica (...),
"flora_string_txt_es": "El Retiro tiene una variada vegetación, en la que confluyen (...),
"description_string_txt_es": "El parque del Retiro o parque del Buen Retiro (...),
"siglo_xviii_string_txt_es": "Tras el incendio del Real Alcázar en 1734 (...)
```

Ilustración 43 Almacenamiento en Solr de textos de Wikipedia

- Una vez procesados todos los datos, se obtienen un total de 4823 obras, 1504 personajes, 14166 referencias y 25569 eventos. De los eventos encontrados, 2799 están relacionados con obras, 2090 con personajes y 221 con referencias.

#### 4.3 SEGMENTACIÓN DE LOS TEXTOS OBTENIDOS E IDENTIFICACIÓN DE ELEMENTOS NARRATIVOS

Como se describió anteriormente, existen dos tipos de descripciones procesadas y etiquetadas (o que precisan estar etiquetadas):

1. Descripciones obtenidas desde la web del Museo del Prado
2. Descripciones obtenidas desde Wikipedia

La etapa de segmentación tiene como finalidad extraer los párrafos (denominado segmento) de cada texto, asignarle un identificador único, extraer los elementos narrativos que contiene, añadir la información de metadatos de la obra a la que pertenecen e indexar toda la información en Solr.

Gracias a esta segmentación, es posible realizar consultas que devuelvan segmentos con uno o varios elementos en común. Por ejemplo, si queremos extraer una lista de segmentos en los que aparezca el personaje “Jacopo Robusti Tintoretto”, y se refieran en el texto a la referencia “Venecia”, como puede verse en la Ilustración 44.

Los segmentos se identificarán de forma unívoca estableciendo una clave con el siguiente patrón:

“Identificador de obra\_Tipo de texto\_Número de segmento”,

Donde el tipo de text puede ser “wp” (si viene de Wikipedia) o “mp” (si viene del Museo del Prado), y el número de segmento indica el número, en este caso, de párrafo en el texto. Por ejemplo, “3\_wp\_2” se refiere a la obra con identificador “3”, correspondiente a un texto extraído de Wikipedia y el segundo párrafo.

El almacenamiento de la información de los segmentos se realiza en Solr, de forma que no sólo se tiene el identificador de segmento y el segmento (en HTML) en si. Además, se almacenan las listas de identificadores narrativos que aparecen en el segmento, y las listas de elementos narrativos que aparecen en la obra. Esto servirá para calcular en el siguiente paso el peso de los nodos de la red conceptual.

1. Resultado:
  - Segmento:"<p>En 1547 la Scuola del <a id\_reference="11243">Santísimo Sacramento</a> de la <a id\_reference="7442">iglesia</a> de <a id\_reference="1970">San Marcuola</a> en <a id\_reference="8851">Venecia</a> encargó a Jacopo (<a id\_character="201">Tintoretto</a> (...)) </p>"
  - Obra:"El lavatorio"
2. Resultado:
  - Segmento:"<p>El caballero de la cadena de oro es un cuadro del pintor <a id\_reference="1969">italiano</a> <a id\_character="201">Tintoretto</a> realizado en <a id\_reference="1966">óleo sobre lienzo</a>. (...) Probablemente fue una de las pinturas que <a id\_character="203">Diego Velázquez</a> compró en <a id\_reference="8851">Venecia</a> durante su segunda visita a <a id\_reference="1969">Italia</a>.</p>"
  - Obra:"El caballero de la cadena de oro"
3. Resultado:
  - Segmento:"<p>José y la mujer de Putifar es un cuadro del pintor <a id\_reference="1969">italiano</a> <a id\_character="201">Tintoretto</a> realizado en <a id\_reference="1966">óleo sobre lienzo</a>.(...) Este cuadro fue comprado por <a id\_character="203">Diego Velázquez</a>, durante su segunda visita a <a id\_reference="8851">Venecia</a>, para <a id\_character="250">Felipe IV</a> junto con otras cinco pinturas de temática <a id\_reference="5177">bíblica</a>.</p>"
  - Obra:"José y la mujer de Putifar"

*Ilustración 44 Segmentación de textos de distintas obras referentes a "Venecia"*

- **Una vez terminado el proceso, se obtienen 11017 segmentos.**





## 5 CREACIÓN DE LA RED CONCEPTUAL Y SELECCIÓN DE TRAMAS

---

La creación de la red conceptual permitirá visualizar, analizar y extraer la información necesaria sobre la información obtenida anteriormente. La red que se obtendrá servirá para analizar y encontrar las tramas que conformarán los itinerarios, utilizando el grado de los nodos (es decir, el número de las relaciones entre los nodos, obtenidas previamente) para encontrar nuevos temas, teniendo en cuenta distintas variables.

Es importante destacar que el objetivo de la construcción del grafo es detectar, a partir del mismo, una serie de temas que conformarán la trama del modelo narrativo. Para conseguirlo, se parte del trabajo realizado por Coursey y Milhacea en 2009 [85], que planteaban un método no supervisado para la identificación de temas usando un grafo creado a partir de Wikipedia. Para encontrar los temas relevantes, en este caso, se propone una variación del algoritmo PageRank [86] basado en un algoritmo de grafo centralizado sesgado aplicado al grafo de Wikipedia. En este trabajo, se modifica la estructura del grafo planteado, de manera que las aristas de este no representan las relaciones de proximidad entre los documentos de Wikipedia, sino entre los segmentos del modelo narrativo.

### 5.1 CREACIÓN DE LA RED. SELECCIÓN DE NODOS Y ARISTAS.

La red conceptual (o grafo) se compone de nodos y aristas. Los nodos se corresponden con los elementos narrativos del modelo narrativo propuesto en este trabajo, asignándoles un identificador unívoco a cada uno que permita su fácil visualización y comprensión. Los identificadores de los elementos narrativos se crean siguiendo el siguiente patrón:

- Referencias: Se crean concatenando “REF”, identificador de la referencia y su nombre. Por ejemplo, “REF8925\_infierno”.
- Personajes: Se crean concatenando “CH”, identificador del personaje y su nombre. Por ejemplo, “CH241\_maria\_luisa\_de\_borbon”.
- Eventos: Se crean concatenando “EV”, identificador del evento y su nombre. Por ejemplo, “EV16914\_fallecimiento\_de\_francisco\_i\_de\_austria\_en\_viena\_en\_2\_marzo\_1835”.
- Obras: Se crean concatenando “ARTW”, identificador de la obra y su nombre. Por ejemplo, “ARTW2\_el\_lavatorio”.

Una vez creados los nodos que corresponden a los elementos narrativos, se crean los nodos de segmentos.

- Segmentos: Se crean concatenando “SEG” con el identificador de segmento. Por ejemplo, “SEG\_3\_wp\_2”.

Las aristas se corresponden con las relaciones existentes entre los elementos narrativos y sus entidades en Wikidata, y con los segmentos en los que se encuentran:

- Los elementos narrativos se pueden corresponder con entidades de Wikidata, y éstas, con instancias de entidades de Wikidata, lo que aportaría categorías semánticas a los elementos. Por ello, en primer lugar, se crean las relaciones entre los elementos narrativos y sus entidades e instancias de entidades.

- Las entidades se crean concatenado “E”, el identificador de la entidad y el nombre.
- Las instancias de entidades se crean concatenando “IO”, el identificador de la instancia de entidad y el nombre.
- Los segmentos se relacionan con la obra a la que pertenecen, y con los elementos narrativos que contienen.

Un ejemplo, mostrado en la Ilustración 45, es la red de la siguiente imagen, que contiene dos segmentos de dos obras distintas: “Felipe IV” y “La Fragua de Vulcano”. Ambas obras tienen una referencia en común, “Pareja”. Además, el segmento correspondiente a la obra “Felipe IV” tiene como referencia “Caballo (Equus ferus caballus)”, que se corresponde con la entidad de Wikidata “Caballo”, y a su vez, con la instancia de Wikidata, “Nombre común”.

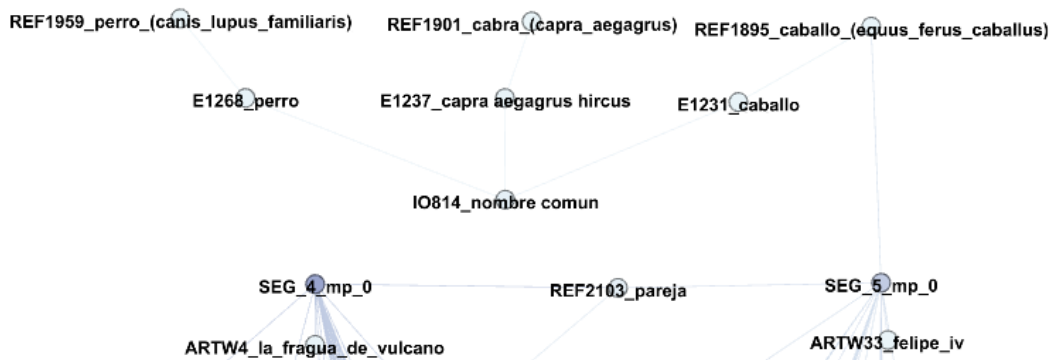


Ilustración 45 Captura de la red conceptual generada en este trabajo.

Otro ejemplo es el mostrado en la Ilustración 46, que contiene parte de la información relativa a un segmento. Aparecen distintos personajes, cuya instancia es “Ser humano”, referencias, como la Guerra de Flandes y un evento, correspondiente al año 1568, relacionado con la obra.

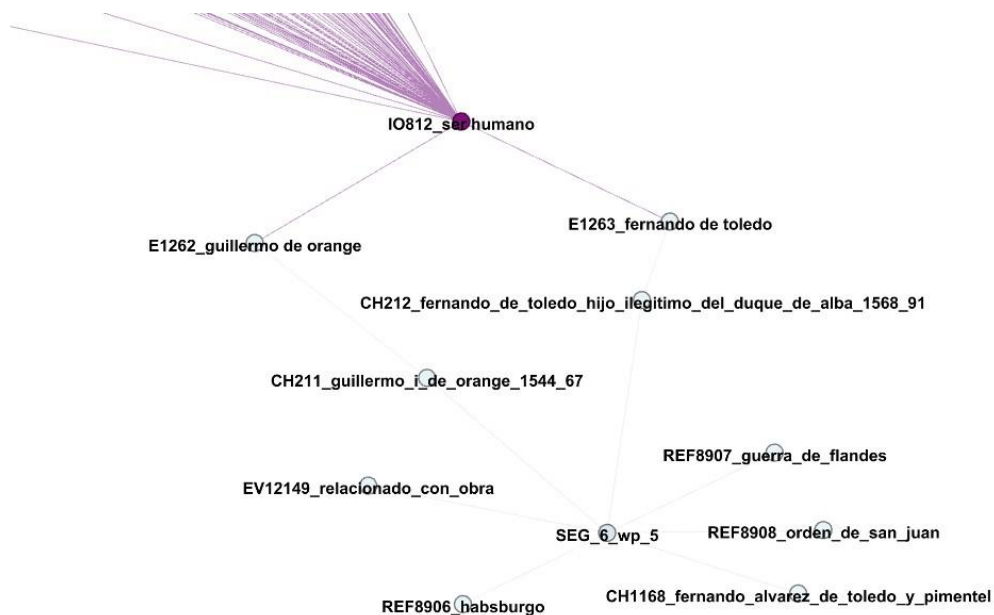


Ilustración 46 Captura de la red conceptual generada en este trabajo.

## 5.2 SELECCIÓN DE TRAMAS ASOCIADAS A UN CONJUNTO DE OBRAS

Dada la cantidad de información, para seleccionar las tramas es necesario, en primer lugar, obtener una serie de temas que aporten un valor de interés al itinerario. Los temas se extraen de los elementos narrativos. Por ejemplo, pueden ser temas: “Cuadros asociados al personaje Velázquez”, “Cuadros pertenecientes al Manierismo” o “Cuadros que hablen del Toisón de Oro”.

La selección de tramas se realiza sobre el grafo creado en el punto anterior de este trabajo. En primer lugar, utilizando la herramienta de análisis y visualización de grafos Gephi [87], se pueden intuir algunos segmentos con información relacionada que podrían constituir un itinerario. Por ejemplo, la siguiente captura se muestra una parte del grafo generado en este trabajo tras ejecutar el algoritmo Force Atlas 2 [88], implementado de forma nativa en Gephi y que tiene como objetivo ayudar en la visualización del grafo, como muestra la Ilustración 47.

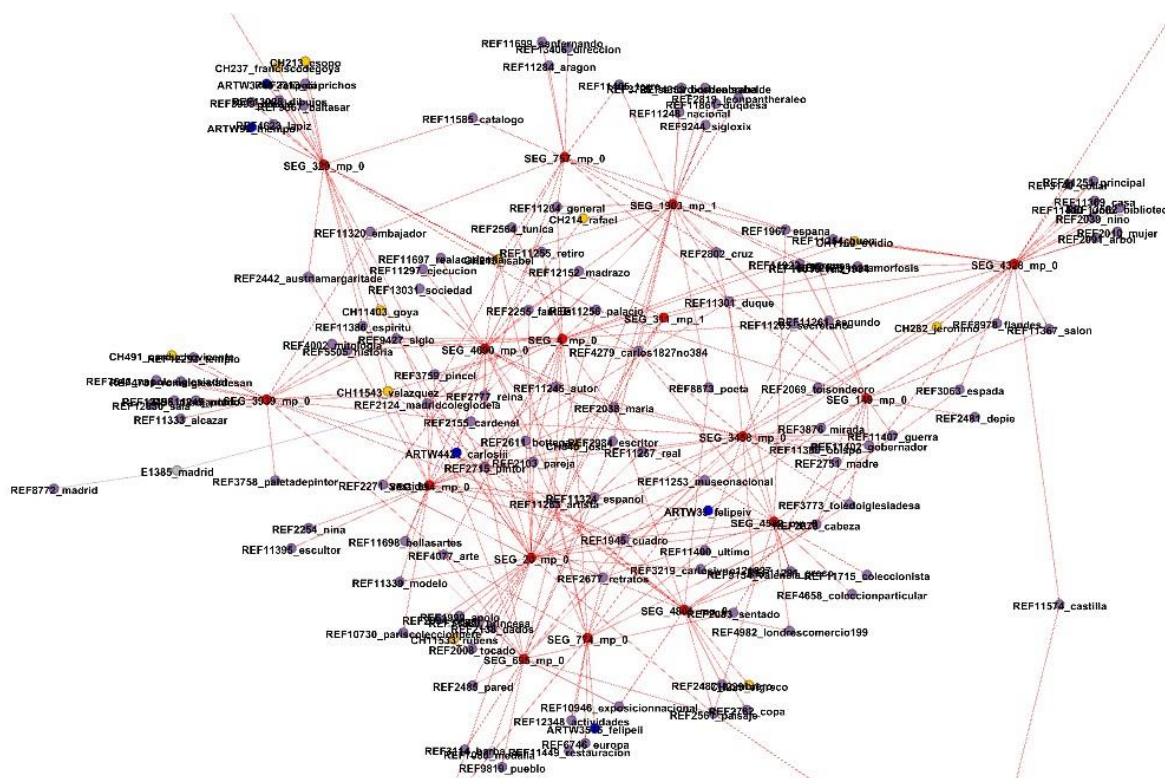


Ilustración 47 Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo

Si seleccionamos algunos de los nodos en color rojo, que corresponden a los segmentos, se pueden ver las distintas relaciones entre los mismos, a través de personajes, referencias o eventos, como muestra la Ilustración 48.

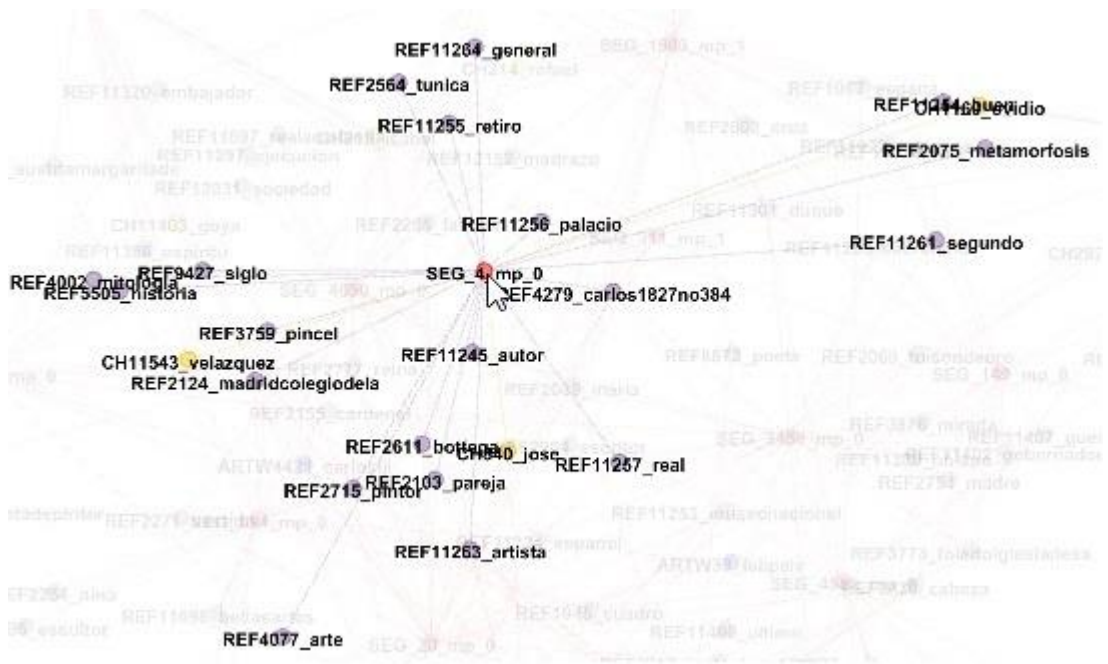


Ilustración 48 Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo

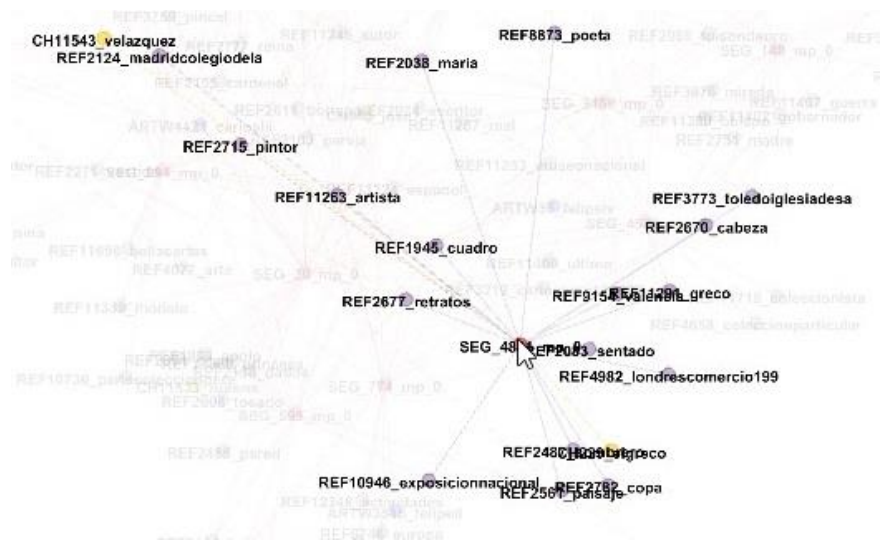


Ilustración 49 Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo

En este caso, se puede ver en la Ilustración 49, por ejemplo, la relación a través del personaje “Velázquez”. Por lo tanto, en este caso, un tema candidato podría ser “Cuadros relacionados con Velázquez”.

Para abordar esta parte del trabajo, se realiza siguiendo las siguientes etapas, que se detallarán en los puntos siguientes:

1. Filtrado de información irrelevante para el grafo.
2. Descomposición del grafo, según las dos aproximaciones utilizadas en este trabajo para seleccionar los temas.
3. Selección de elementos para la creación de itinerarios.
4. Asignación de elementos a itinerarios.

### 5.3 DESCOMPOSICIÓN DEL GRAFO

En el grafo creado anteriormente, se mezcla información de distintos tipos: Existen nodos de distintas clases (personajes, eventos, referencias, obras y segmentos) y aristas de dos tipos (los que relacionan los segmentos y sus elementos narrativos, y los que relacionan los elementos narrativos con sus entidades de Wikidata, y éstos, a la instancia de Wikidata que referencian).

Por lo tanto, dado que la información contenida en cada caso ha de tratarse de forma distinta, el primer paso es crear los dos grafos a partir del grafo principal. El primer grafo contendrá los elementos narrativos y los segmentos, además de las relaciones entre éstos. El segundo grafo contendrá los elementos narrativos y sus entidades e instancias de entidades de Wikidata.

#### 5.3.1 Primer grafo. Segmentos y elementos narrativos.

Las características principales del grafo son:

- Se trata de un grafo dirigido formado por 18516 nodos y 67852 aristas.
- El grado medio es de 3.6

Como se puede ver en la Ilustración 50, esta representación del grafo permite visualizar rápidamente los elementos narrativos que existen en cada segmento.

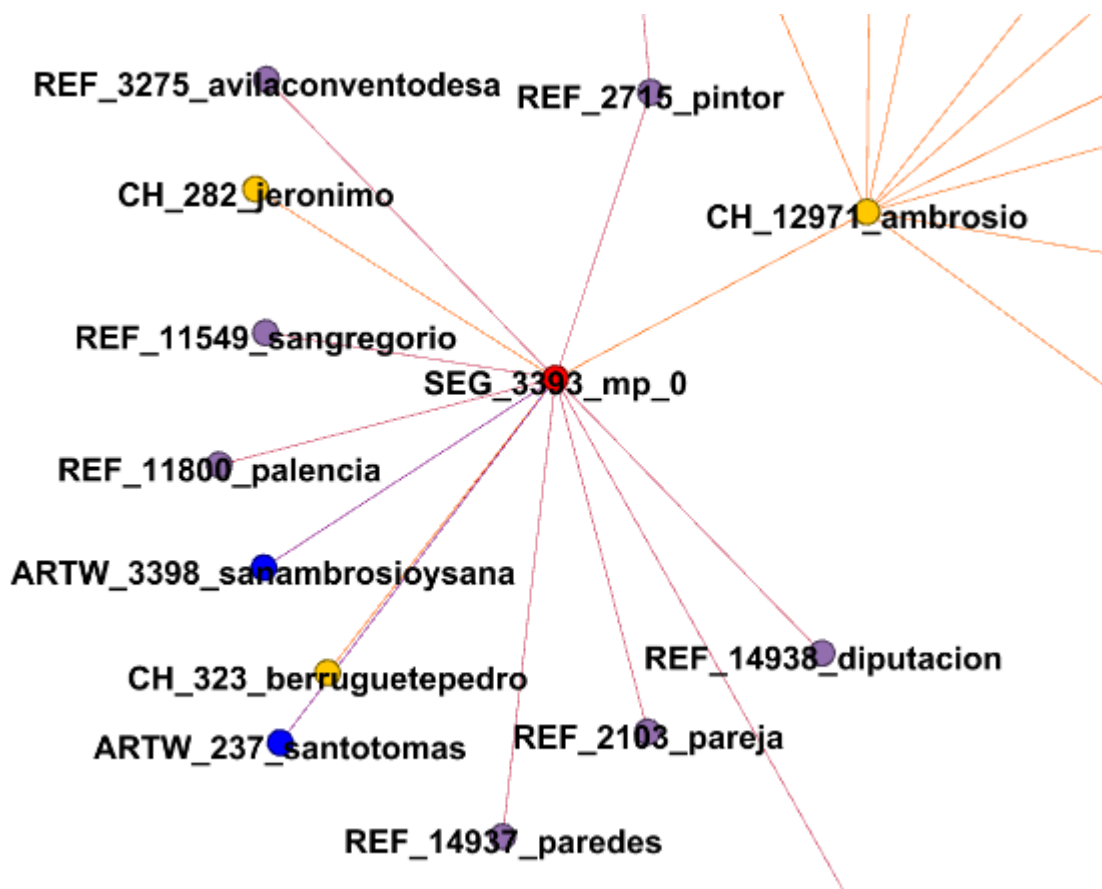


Ilustración 50 Captura del grafo correspondiente con los elementos narrativos existentes en un segmento.

### 5.3.2 Segundo grafo. Elementos narrativos y entidades de Wikidata.

Las características principales del grafo son:

- Se trata de un grafo no dirigido formado por 22664 nodos y 15342 aristas.
- El grado medio es de 0.6

Las entidades de Wikidata aportan posibles formas de categorizar los segmentos. Por ello, son candidatos como elementos de tramas. Un ejemplo, es el subgrafo creado a partir del grafo que contiene elementos narrativos y entidades de Wikidata mostrado en la Ilustración 51.

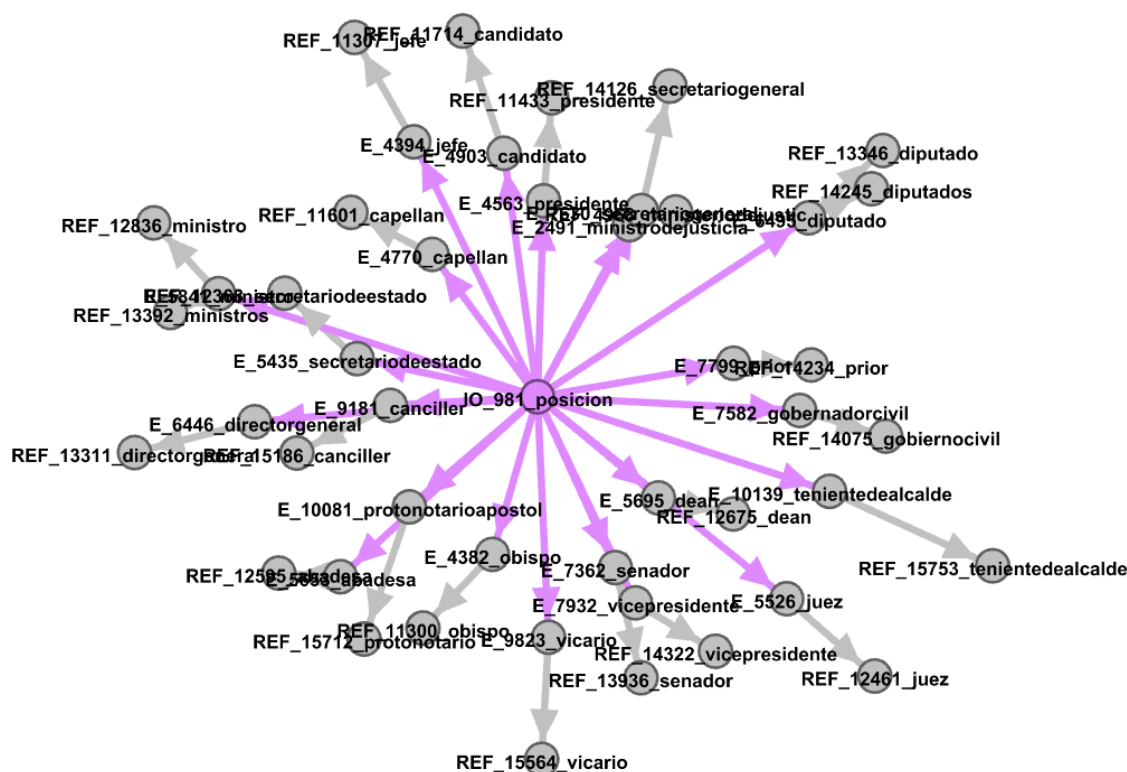


Ilustración 51 Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo

Se puede ver, por ejemplo, que la instancia de “posición”, apunta o señala a una serie de entidades, como por ejemplo “diputado”, “secretario”, “ministro de justicia”, etc. A su vez, estas entidades apuntan a su referencia dentro del modelo narrativo de este trabajo. En este caso, los temas candidatos podrían ser sobre cada una de las entidades de Wikidata, como “cuadros donde aparezca un presidente”. O, también, podrían ser la intersección de entidades de Wikidata que contengan los suficientes segmentos como para crear un itinerario.

Otro ejemplo, mostrado en la Ilustración 52, es el que se crea a partir de un metadato almacenado en los cuadros, y que se ha almacenado en el grafo como una entidad más: el tema artístico.

Como se puede comprobar, el proceso de desambiguación llevado a cabo utilizando la API de Wikipedia, da resultados interesantes. En este caso, las referencias “Madonna col Bambino” y “La Virgen con el niño” pertenecen a la misma entidad de Wikidata.



Una vez separados los grafos y analizado, a alto nivel, la información que contiene y cómo se estructuran, el siguiente paso es decidir qué información del grafo es relevante y cuál no.

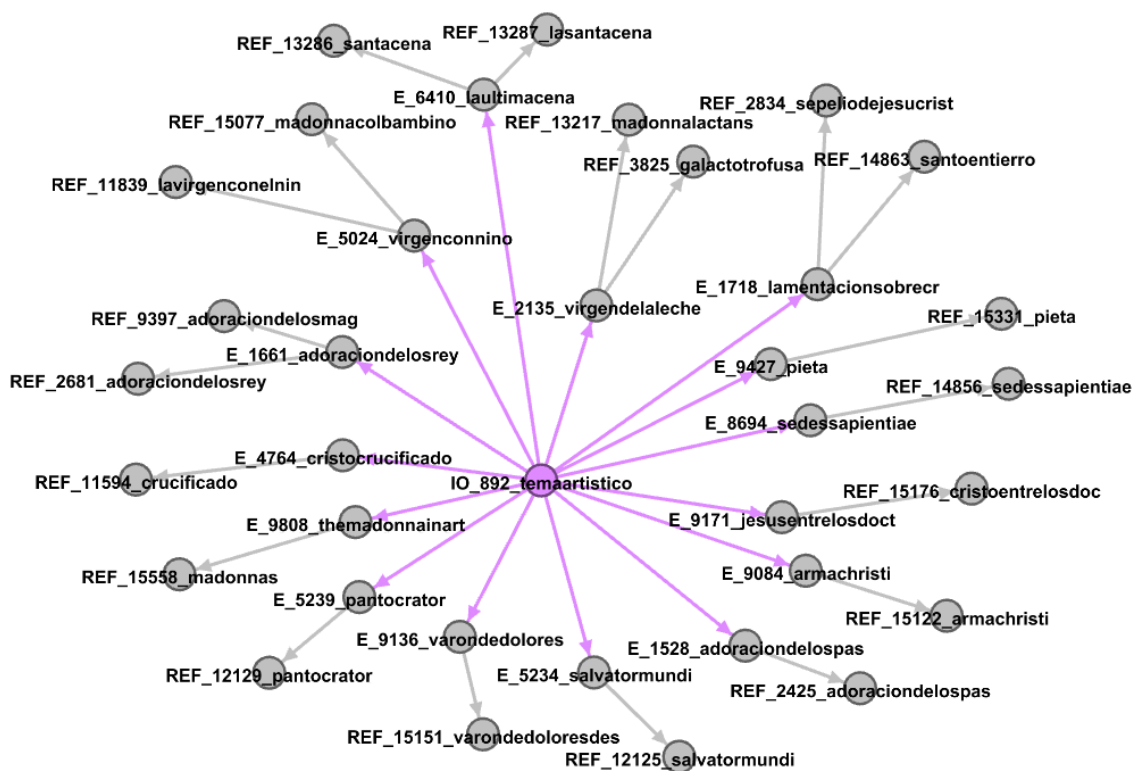


Ilustración 52 Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo

## 5.4 FILTRADO DE INFORMACIÓN

Un problema que ya se puede apreciar en la etapa de visualización del grafo es el exceso de información, que en algunos casos no aporta información relevante para el usuario. Por ejemplo, la palabra “cuadro” es un elemento narrativo de referencia, pero a la hora de crear los itinerarios, dado que todos son cuadros, no aporta información de interés. Por ello, es necesario trabajar con un algoritmo que tenga en cuenta que, si una palabra clave aparece en muchos segmentos, no puede considerarse como candidata para el itinerario.

Partiendo del grafo que relaciona los segmentos con los elementos narrativos, y seleccionando únicamente las unidades narrativas de tipo referencias, obtenemos un total de 15913 nodos y 55041 aristas. Como se puede ver, el número de relaciones es muy superior al de nodos, debido a la existencia de casos como el que se acaba de mostrar de “cuadro”. Por ello, la etapa de filtrado pretende buscar un método para eliminar los nodos de tipo referencia que no resulten relevantes para el objetivo de este trabajo, es decir, que no permitan crear itinerarios de interés por el hecho de ser muy genéricos.

El primer paso es seleccionar qué información es relevante y cuál irrelevante. Dado que esta etapa del proceso pretende seleccionar elementos narrativos como candidatos para la creación de itinerarios, se tendrán varias consideraciones:

1. Un elemento narrativo (de tipo referencia o personaje) es irrelevante si aparece en muchos segmentos.
2. Un elemento narrativo de tipo obra se considera relevante, pese a que aparezca en muchos segmentos.
3. Un elemento narrativo de tipo evento no se considera relevante ni irrelevante en esta etapa de filtrado, puesto que los eventos estarán directamente asociados al resto de elementos narrativos, y su ocurrencia o no dependerá de éstos.
4. Todos los elementos narrativos se consideran irrelevantes si no existen en, al menos, una cantidad de segmentos por itinerario dada como parámetro.

Teniendo en cuenta los puntos anteriores, el primer paso es plantear cuántos segmentos deben conformar un itinerario. Dicho de otro modo, cuántos párrafos crean un itinerario. El sistema debe ser capaz de permitir que el usuario seleccione este valor, siendo el mínimo de 3 (para obtener un mínimo de información) y el máximo de 15. La idea es que un itinerario debe tener un mínimo de segmentos o párrafos, pero no debe ser demasiado extenso.

La siguiente figura representa en una gráfica el valor de DF (*Document Frequency*), es decir, para cada elemento narrativo de tipo referencia o personaje, en cuántos segmentos aparece, como muestra la Ilustración 53.

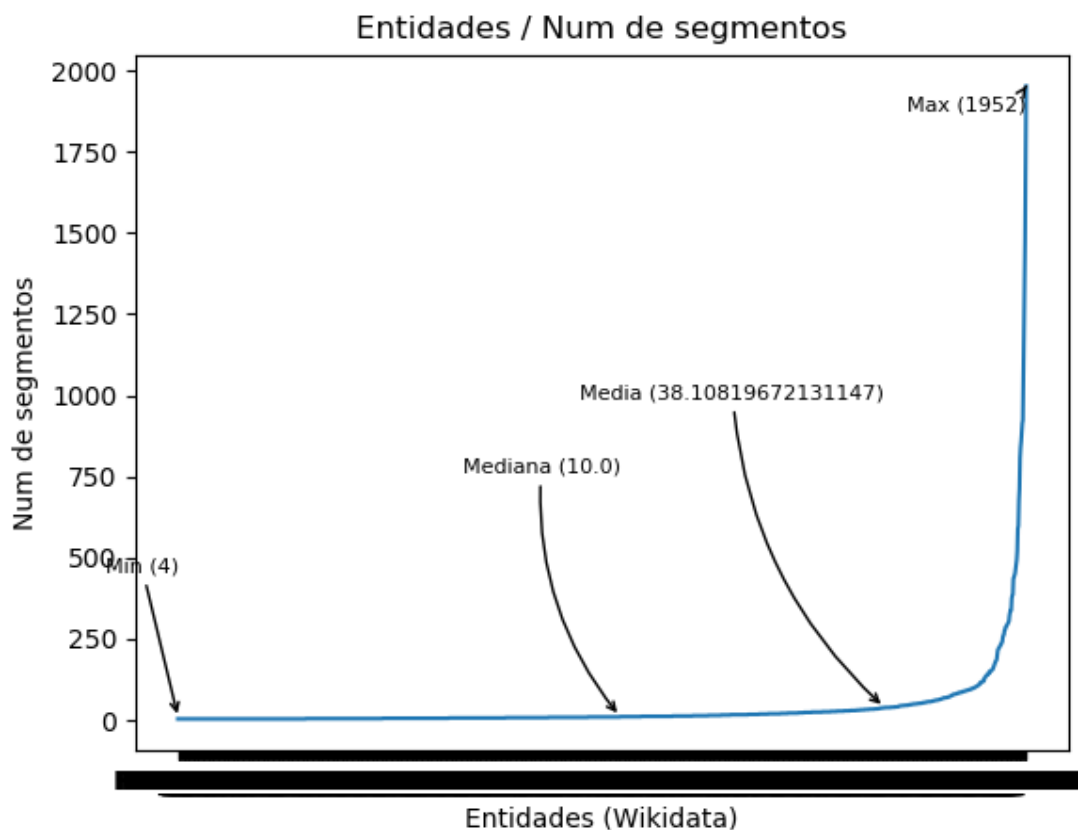


Ilustración 53 Gráfica que representa el DF (frecuencia de elementos narrativos / segmentos) de la información obtenida

Como se puede ver en la figura anterior, la media de segmentos por elemento narrativo es de 38 y el elemento narrativo que aparece en mayor cantidad de segmentos, aparece en 1952 segmentos. La idea de esta etapa de filtrado es eliminar los itinerarios que corresponden con elementos narrativos que no se consideren relevantes. Para ello, se llevan a cabo las siguientes aproximaciones:



#### 5.4.1 Aproximación 1. Cálculo de umbral no supervisado.

Partiendo del trabajo realizado en [89], que buscaba resolver un problema similar dentro del dominio de Twitter, la primera aproximación consiste en calcular un umbral que desestime los elementos narrativos que aparezcan en muchos documentos. Este umbral se calcula en dos fases:

1. El valor mínimo de segmentos por cada elemento narrativo lo establece el propio usuario, como se dijo anteriormente, siendo un valor entre 3 y 15. Si el usuario selecciona, por ejemplo, que cada itinerario tenga más de 10 segmentos, el valor mínimo del umbral sería 11.
2. El valor máximo del umbral se calcula como la media de los valores más el valor máximo, dividido por la mitad.

$$\gamma = \frac{\bar{x} + \max(x)}{2}$$

Además, se realiza el cálculo teniendo en cuenta, en primer lugar, todos los elementos narrativos con, al menos, 3 segmentos (filtro general). Después, se realiza el mismo cálculo teniendo en cuenta el valor mínimo de segmentos dado por el usuario (en este ejemplo, 10)

En la tabla 5 se muestra el valor del umbral calculado en caso de número de segmentos mayor que 3 y mayor que 10, trabajando sobre toda la información en ambos casos (es decir, tomando como valor máximo 1952).

Tipo de filtro	Media	Mediana	Valor máximo	Valor mínimo	Valor de umbral máximo
Filtro general	38,11	10,00	1952,00	4,00	995,05
Filtro 1 - Segmentos/itinerario	72,93	26,00	1952,00	11,00	1012,47

Tabla 5 Primera aproximación del cálculo del umbral no supervisado.

El total de elementos narrativos que se obtienen dentro del umbral en el caso de utilizar el filtro general es de 1525. En caso de aplicar el primer filtro, se obtienen 1519 (sólo se eliminan 6 elementos narrativos).

#### 5.4.2 Segunda aproximación. Cálculo del umbral con dos iteraciones.

En esta iteración, se vuelve a realizar el filtro sobre los datos obtenidos a partir del filtro general. Así, se pretende eliminar en primer lugar la información más irrelevante, y, en segundo lugar, la información irrelevante partiendo del valor mínimo de segmentos por itinerario.

Tipo de filtro	Media	Mediana	Valor máximo	Valor mínimo	Valor de umbral máximo
Filtro general	38,11	10,00	1952,00	4,00	995,05
Filtro 1 - Segmentos/itinerario sin filtro general	72,93	26,00	1952,00	11,00	1012,47
Filtro 2 - Segmentos/itinerario a partir del filtro general	61,46	25	924	11	492,73

Tabla 6 Segunda aproximación del cálculo del umbral no supervisado.

Como se puede ver en la tabla 6, en esta segunda aproximación, el valor máximo queda en 924, lo que supone haber eliminado una cantidad importante de elementos narrativos considerados irrelevantes.

Ejemplos de los elementos eliminados gracias a esta segunda iteración, son las referencias “Madrid”, “Pinturas”, “España”; o el personaje “José”. En este caso, con el segundo filtro, se obtienen un total de 724 elementos narrativos, por lo que se podrían crear, al menos, 724 itinerarios de cada tema.

Al representar la información seleccionada con los filtros aplicados en este caso, se crea el gráfico mostrado en la Ilustración 54.

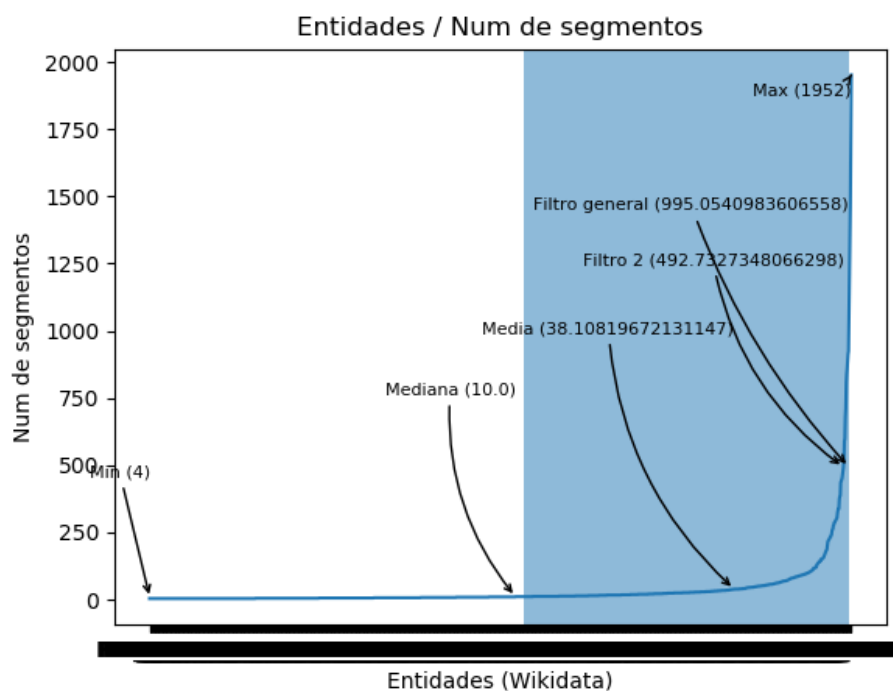


Ilustración 54 Gráfica que representa el DF (frecuencia de elementos narrativos / segmentos) de la información obtenida aplicando los filtros.

### 5.4.3 Tercera aproximación. Cálculo del umbral basado en el número de obras.

Esta tercera aproximación pretende mejorar la riqueza de los itinerarios, filtrando la información anterior a partir de la variedad de obras, no sólo por la cantidad de segmentos. Cada segmento pertenece a una obra, y varios segmentos pueden pertenecer a una misma obra. En la aproximación anterior, puede darse el caso de que un itinerario tenga segmentos de una misma obra y se considere información relevante. En esta aproximación, se busca que los segmentos pertenezcan, además, a obras distintas. Así, los itinerarios serán más variados.

De este modo, obtenemos una nueva versión de la tabla anterior. En este caso, el valor del umbral es menor que en el caso anterior, incluso en la etapa de filtrado general. Esto se debe a que existen muchos elementos narrativos que coinciden, en muchos segmentos, pero además en muchas obras iguales. La primera iteración del filtrado ya elimina toda esta información, mostrada en la tabla 7.

Tipo de filtro	Media	Mediana	Valor máximo	Valor mínimo	Valor de umbral máximo
Filtro general	33,25	9,00	1528,00	4,00	780,63
Filtro 2 - Segmentos/itinerario a partir del general	55,37	23	736	11	395,69

Tabla 7 Filtrado final de la información

La gráfica generada en este caso se muestra en la Ilustración 55.

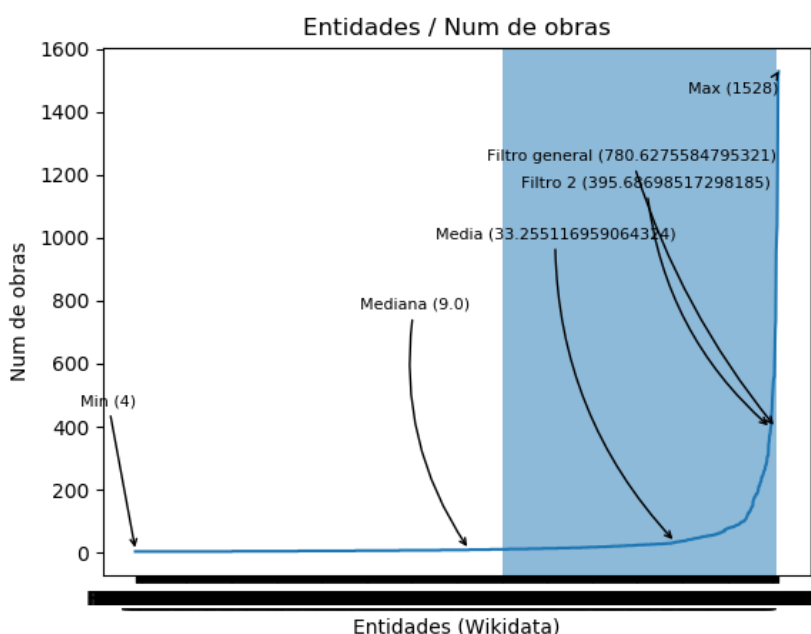


Ilustración 55 Gráfica que representa el DF (frecuencia de elementos narrativos / obras) de la información obtenida aplicando los filtros.

En esta etapa de filtrado se obtienen un total de **607 elementos narrativos** candidatos a convertirse en itinerarios.

## 5.5 SELECCIÓN DE TRAMAS.

La selección de tramas puede ser vista como un proceso de extracción de temas a partir del grafo creado, como proponen Rada Milhacea [85] y Laura Plaza [90]. En el punto anterior se eliminaron los elementos narrativos que se consideran irrelevantes en la etapa de selección, obteniendo una lista de candidatos a convertirse en itinerarios.

En total, se obtienen **10457 segmentos** que contienen 607 elementos narrativos, creando un grafo con 11064 nodos y 42470 aristas, y con un grado medio de 3.8.

Un problema a tener en cuenta en este punto es que al seleccionar los elementos narrativos que formarán parte de los itinerarios, sus segmentos correspondientes pueden contener información que no se considere útil. Por ejemplo, si pretendemos crear un itinerario a partir del elemento narrativo “Montero” (<https://www.wikidata.org/wiki/Q990706>), podemos obtener segmentos de, al menos,

tres cuadros distintos: “Abraham y los tres ángeles”<sup>48</sup>, “La prudente Abigail”<sup>49</sup> y “Triunfo de la Fe sobre los Sentidos”<sup>50</sup>. En los tres cuadros, se puede encontrar exactamente el mismo segmento o párrafo:

*“Forma parte de la serie de dieciocho lienzos de temas bíblicos relacionados de algún modo con la Eucaristía, que pintó Escalante entre 1667 y 1668, para la sacristía de la Merced Calzada de Madrid, donde los citan todos los escritores clásicos con sumo elogio. La serie se completaba con un lienzo de Montero de Rojas, El paso del río Jordán con el Arca de la Alianza (P4697).”*

Por lo tanto, un problema es que, en caso de crear un itinerario a partir de esta referencia, se mostraría el mismo segmento todas las veces.

La obtención de las tramas a partir del grafo se realiza obteniendo los elementos narrativos seleccionados como candidatos, creando una trama por cada elemento narrativo seleccionado.

### 5.5.1 Selección de tramas a partir de elementos narrativos.

En este caso, el objetivo es seleccionar, ordenar y almacenar el texto en formato HTML de los segmentos que conformarán cada trama a partir de cada elemento narrativo. Para ilustrar esta parte del proceso, se propone como ejemplo el itinerario sobre el personaje Luis Tristán<sup>51</sup>.

Para llevar a cabo este objetivo, en primer lugar, se obtiene el subgrafo local generado para cada elemento narrativo a partir del grafo filtrado obtenido en el punto anterior. En este caso, el subgrafo generado para Luis Tristán se puede ver en la Ilustración 56.

Como se puede ver en la figura anterior, el grafo contiene una serie de segmentos y los elementos narrativos que contienen los mismos. Para saber qué segmentos seleccionar, se calcula el **coeficiente de Jaccard**, que devuelve la similitud de las aristas teniendo en cuenta su grado, tal y como se describió anteriormente. Dada la naturaleza del problema, se tienen en cuenta las similitudes de las aristas que relacionan los nodos de tipo segmento, y, además, se eliminan los segmentos cuya similitud es mayor que 0.85 (es decir, se eliminan las relaciones entre los segmentos que son muy similares). Así, es posible obtener una lista de pares de segmentos, ordenada por sus similitudes, como se puede ver en la Ilustración 57.

- La ventaja principal que ofrece el uso del coeficiente de Jaccard como medida de similitud entre cada par de segmentos del grafo, es que se elimina la posibilidad de escoger, ya no sólo dos segmentos iguales, sino también, dos segmentos aparentemente distintos (que vengan de distintas obras), pero que realmente contengan la misma, o casi la misma, información.

---

<sup>48</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/abraham-y-los-tres-angeles/6c994788-69e4-4a19-96ad-e930d34713e3>

<sup>49</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/la-prudente-abigail/c33679b4-bbe2-48e1-bd3e-b01c0adadc2b>

<sup>50</sup> <https://www.museodelprado.es/coleccion/obra-de-arte/triunfo-de-la-fe-sobre-los-sentidos/da77520b-a6b7-492f-b17d-b300f51d268c>

<sup>51</sup> [https://es.wikipedia.org/wiki/Luis\\_Tristán](https://es.wikipedia.org/wiki/Luis_Tristán)



Ilustración 57 Imagen extraída utilizando Gephi sobre el grafo generado en este trabajo correspondiente a Luis Tristán

SEG_32_wp_3 - SEG_3798_mp_0	->	0.0833333333333333
SEG_32_wp_3 - SEG_1003_mp_0	->	0.0555555555555556
SEG_32_wp_3 - SEG_1002_mp_1	->	0.142857142857
SEG_32_wp_3 - SEG_1002_mp_0	->	0.111111111111111
SEG_32_wp_3 - SEG_1002_mp_2	->	0.166666666666667
SEG_32_wp_3 - SEG_3803_mp_0	->	0.1
SEG_32_wp_3 - SEG_1622_mp_0	->	0.0625
SEG_32_wp_3 - SEG_3685_mp_0	->	0.066666666666667
SEG_32_wp_3 - SEG_3757_mp_2	->	0.0625
SEG_32_wp_3 - SEG_1004_mp_0	->	0.125
SEG_32_wp_3 - SEG_3795_mp_0	->	0.105263157895
SEG_32_wp_3 - SEG_3757_mp_0	->	0.0769230769231
(...)		

Ilustración 56 Valores de similitud correspondientes a un conjunto de segmentos

Para calcular la ruta de segmentos que conforman el itinerario, se siguen los siguientes pasos:

1. Calcular la media de los coeficientes de Jaccard, para obtener un valor intermedio de las similitudes.
2. Seleccionar y añadir a la ruta de segmentos seleccionados la primera pareja de segmentos cuyo coeficiente de Jaccard sea igual o mayor a la media calculada. Siguiendo el ejemplo

anterior, la media obtenida es de 0.1793. Así, se seleccionaría la pareja de segmentos “SEG\_3225\_mp\_2 - SEG\_3218\_mp\_0” cuya media es de 0.18.

3. Eliminar de la lista de parejas de segmentos, los seleccionados en el paso anterior, para evitar seleccionar varias veces el mismo segmento para el mismo itinerario. Para ello:
    - a. Se eliminan como siguiente valor ambos segmentos (es decir, se elimina el segundo valor de la pareja)
    - b. Se eliminan las combinaciones que involucren ambos segmentos.
  4. Hasta completar la lista de segmentos para crear el itinerario:
    - a. Escoger todos los segmentos cuyo primer valor en la tupla corresponda con el anterior seleccionado para el camino. La idea es que se escoja en todo momento el valor más central (calculado a partir de la media) del subgrupo de parejas de segmentos creado a partir de los segmentos relacionados con el último segmento añadido al itinerario. Por ello:
      - i. En caso de haber varias parejas de segmentos disponibles, seleccionar el valor medio.
      - ii. En caso de no existir parejas de segmentos que comiencen por el segmento seleccionado, seleccionar el segmento siguiente cuya media sea mayor o igual a la de los segmentos seleccionados. En caso de que no haya nuevos segmentos que cumplan esta condición, se selecciona un segmento aleatorio de entre los disponibles.
- Además de la ventaja comentada anteriormente, medir la similitud entre cada par de nodos permite también seleccionar los nodos que no sean “demasiado distintos” ni “demasiado iguales”. Esto se consigue seleccionando en cada caso la pareja de segmentos central, es decir, la que corresponde a la media de los coeficientes.

Así, siguiendo de nuevo con el ejemplo, se seleccionarían los siguientes segmentos para el itinerario (suponiendo que el número de segmentos que conforman un itinerario es de 10), mostrado en la Ilustración 58.

Siendo cada segmento:

“[SEG]\_[ id obra]\_[ procedencia (Museo del Prado o Wikipedia)]\_[número de párrafo]”

```
['SEG_3225_mp_2',  
'SEG_3218_mp_0',  
'SEG_3145_mp_4',  
'SEG_1005_mp_5',  
'SEG_1005_mp_6',  
'SEG_1005_mp_3',  
'SEG_3220_mp_2',  
'SEG_3217_mp_1',  
'SEG_3224_mp_2',  
'SEG_3220_mp_0',  
'SEG_3222_mp_0']
```

*Ilustración 58 Segmentos seleccionados para un itinerario.*

Una vez obtenidos los segmentos, se crea el itinerario almacenándolo en un fichero de texto. El fichero de texto generado para este ejemplo sería el mostrado en la Ilustración 59.

#### **<title>Cuadros relacionados con el personaje Luis Tristán</title>**

```
<p id_artwork="1002">Representación del <a id_reference="11387">Santo</a> meditando sobre las  
sagradas escrituras. (...) </p>  
<p id_artwork="3797">En este <a id_reference="2677">retrato</a> de extrordinaria calidad, el (...) </p>  
<p id_artwork="1005" >La colección <em>Cuadros selectos de la <a id_reference="11697">Real  
Academia</a> de (...) </p>  
<p id_artwork="1005">Una vez ejecutadas las láminas, (...) </p>  
<p id_artwork="3809">Pintada por Tristán en 1616, esta <a id_reference="11725">pintura</a> (...) </p>  
(...)
```

*Ilustración 59 Creación final de un itinerario en HTML.*

### 5.5.2 Resultados obtenidos

Partiendo del estudio realizado por Aletras et al. [91], que muestra que las etiquetas textuales son más fáciles de interpretar para los usuarios que las listas de términos y etiquetas de imágenes, la selección final de las tramas se presentará al usuario utilizando estas etiquetas, es decir, cada tema se representará con una oración representativa del mismo, en lugar de una lista de palabras clave.

Una vez finalizada esta etapa, **se obtienen un total de 581 itinerarios generados automáticamente.**





## 6 EVALUACIÓN

---

Para evaluar los resultados obtenidos, en primer lugar, es necesario el diseño y la implementación de un sistema de información basado en web, que permita a los distintos usuarios interactuar con los itinerarios obtenidos. Una vez creado el sistema web y definidos los distintos roles de usuario, se elabora el cuestionario que apoyará la evaluación y se planteará una tarea a los usuarios a realizar antes del cuestionario. Finalmente, se muestran los resultados obtenidos en la misma, las conclusiones y las propuestas de mejora.

### 6.1 DISEÑO E IMPLEMENTACIÓN DEL PROTOTIPO PARA LA VISUALIZACIÓN DE ITINERARIOS

Para poder visualizar los resultados y evaluar el sistema obtenido con usuarios, se diseña e implementa un prototipo de sistema web, llamado MPOC. Se puede acceder al mismo a través de la siguiente URL:

<http://aluned.laraclares.com/>

El registro se encuentra abierto al público y es posible visualizar los itinerarios existentes en el sistema. La documentación referente al sistema se ha detallado acorde a los estándares vigentes en cada caso:

- El manual de usuario se encuentra en el ANEXO I “**Manual de usuario IEEE 1063-2001**”.

### 6.2 EVALUACIÓN CON USUARIOS

El estándar **ISO 9241-11**, cuya última versión ha salido publicada en marzo de 2018<sup>52</sup>, explica cómo la usabilidad se puede interpretar en términos de rendimiento y satisfacción. Define la usabilidad como *“la medida en que un sistema, producto o servicio puede ser utilizado por usuarios específicos para alcanzar objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso específico”*. Dicho de otro modo, la usabilidad es la facilidad de uso de un producto o servicio.

Este estándar relaciona directamente al usuario con los objetivos propios del producto. Por lo tanto, para poder medir la usabilidad de un producto, es necesario descomponer los objetivos del producto en términos de **efectividad, eficiencia y satisfacción**. Efectividad es la “capacidad de lograr el efecto que se desea o se espera”<sup>53</sup>. Eficiencia es la “capacidad de disponer de alguien o de algo para conseguir un efecto determinado”<sup>54</sup>. Satisfacción, por su parte, es el “cumplimiento del deseo o del gusto”<sup>55</sup>.

En la revisión del estándar realizada en 2015 por Bevan et al. [92], se definen tres conceptos relacionados directamente con el estudio de la usabilidad:

- Calidad centrada en el usuario
- Accesibilidad

---

<sup>52</sup> <https://www.iso.org/standard/63500.html>

<sup>53</sup> <http://dle.rae.es/?id=EOjKmrI>

<sup>54</sup> <http://dle.rae.es/?id=EPVwpUD>

<sup>55</sup> <http://dle.rae.es/?id=XLJpCvk>

- Experiencia de usuario

Además, replantea el ámbito de la usabilidad hacia los siguientes aspectos:

- Se centra en los resultados de la interacción más que en el diseño de la interfaz de usuario y en los atributos del producto resultantes que hacen que un producto sea utilizable.
- Define la usabilidad como un concepto de alto nivel en lugar de referirse al uso normal de un producto en contraste con el aprendizaje de su uso o reutilización.

El proceso de evaluación de usuarios se llevará a cabo siguiendo los siguientes pasos:

1. Planteamiento de una tarea a realizar a los usuarios sobre el sistema, que les permita familiarizarse con el mismo y su objetivo.
2. Evaluación de la experiencia de usuario.
3. Estudio de comportamiento de los usuarios en el sistema, a través de los logs.
4. Obtención de valoraciones y sugerencias.

En primer lugar, se planteará a los usuarios la tarea de buscar dos itinerarios y visualizarlos. Los itinerarios seleccionados son:

- Cuadros relacionados con el personaje Diego Velázquez.
- Cuadros relacionados con Adoración de los Reyes Magos.

A continuación, se describirán en detalle cómo se ha llevado a cabo la evaluación en cada una de las tres vías planteadas.

### 6.2.1 Preguntas cerradas. Evaluación de la experiencia de usuario

Partiendo del estándar descrito anteriormente, Srchep et al. [93] comenzaron a desarrollar en 2005 la elaboración de un cuestionario para evaluar la experiencia de usuario, dando lugar a un marco de trabajo completo donde propone un cuestionario que pretende medir la experiencia de usuario de un producto de forma cuantitativa. Para ello, parte de una interpretación diferente de experiencia de usuario, definiéndolo como un conjunto de criterios de calidad: eficiencia, controlabilidad, habilidad de aprender, estimulación, novedad, etc.

El cuestionario utilizado en este trabajo se planteó por primera vez por expertos en usabilidad, partiendo de un conjunto de 229 conceptos relacionados con la experiencia de usuario [94]. Tras validar el cuestionario en distintos estudios, se obtiene finalmente un conjunto de 26 conceptos que evalúan 6 baremos [95]:

- **Atracción.** Impresión global del producto. ¿Gusta a los usuarios?
- **Perspicacia o transparencia.** ¿Es fácil familiarizarse con el producto?
- **Eficiencia.** ¿Pueden los usuarios resolver las tareas sin realizar un esfuerzo innecesario?
- **Fiabilidad o controlabilidad.** ¿El usuario siente el control en la interacción?
- **Estimulación.** ¿El producto resulta motivante para el usuario?
- **Novedad.** ¿Es el producto novedoso y creativo?

La consistencia de esta encuesta se ha validado en 11 pruebas de usabilidad con un total de 144 participantes, además de una encuesta online con 722 participantes [95]. Los resultados de esta encuesta muestran una consistencia de escala suficientemente alta medida con el test de **Alfa de**

**Cronbach** [96], que tiene en cuenta la correlación existente entre cada par de conceptos para estimar el límite inferior de fiabilidad de los resultados. La definición formal de este test puede verse en la Ilustración 60. El valor  $k$  indica el número de ítems de la prueba (en este caso, 26),  $S_{sum}^2$  es la varianza de la prueba, y  $S_i^2$  es la varianza con respecto a cada ítem.

$$\alpha = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum S_i^2}{S_{sum}^2} \right)$$

Ilustración 60 Definición del test Alfa de Cronbach

Este cuestionario utiliza distintos criterios de calidad, categorizados como muestra Ilustración 61. Los baremos descritos anteriormente se dividen en dos categorías: calidad pragmática y calidad hedónica o de percepción del usuario.

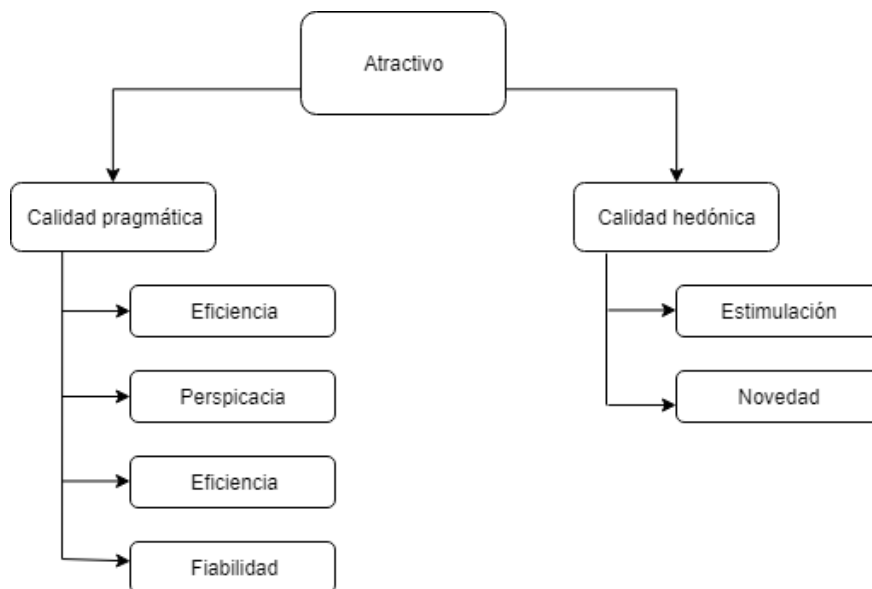


Ilustración 61 Criterios de calidad contemplados en el cuestionario

La utilidad de este cuestionario se ha planteado y demostrado, además, en diferentes escenarios [97], como comparar la experiencia de usuario entre distintos productos o determinar áreas de mejora. En el caso de este trabajo, dado que no es comparable con otro proyecto similar, se centra en el estudio de qué aspectos son mejorables en el sistema.

La documentación correspondiente con esta encuesta se encuentra en la siguiente URL:

<https://www.ueq-online.org/>

Las preguntas cerradas realizadas en este trabajo se pueden ver en el **ANEXO II “Preguntas cerradas”**.

## 6.2.2 Preguntas abiertas. Obtención de valoraciones y sugerencias.

Además de las preguntas referentes a la experiencia de usuario, es necesario obtener una serie de sugerencias directamente de los usuarios. Además, se obtiene información sobre la edad, el nivel de estudios y el conocimiento previo sobre las obras del museo. Esta información permitirá categorizar las sugerencias atendiendo a distintos tipos de usuarios.

Las preguntas abiertas realizadas en este cuestionario son:

1. ¿Qué es lo que más te ha gustado de este sistema y por qué?
2. ¿Qué es lo que menos te ha gustado de este sistema y por qué?
3. ¿Cómo crees que se podría mejorar el sistema? ¿Te gustaría añadir alguna funcionalidad en concreto?

Las respuestas obtenidas por los usuarios en este trabajo se pueden ver en el **ANEXO III “Respuestas a preguntas abiertas”**.

## 6.2.3 Estudio de comportamiento

Otra tarea fundamental para obtener información acerca del sistema es el estudio de los logs. Se almacena información cada vez que el usuario abre un enlace dentro del dominio, utilizando como medio de identificación la variable de sesión asociada a cada usuario, o el identificador de usuario, en caso de haber iniciado sesión.

El objetivo es, en este caso, averiguar qué hacen los usuarios cuando acceden al sistema: si ven un itinerario, si se quedan a mitad de uno o lo ven completo, si hacen clic en los elementos narrativos, cuánto tiempo (aproximado) permanecen en la página, etc.

## 6.3 RESULTADOS OBTENIDOS

La evaluación de resultados obtenidos se realiza en dos partes. La primera parte, a partir de los resultados mostrados en los logs. La segunda parte, a partir de las encuestas.

### 6.3.1 Análisis de las encuestas

El trabajo de las encuestas se divide, a su vez, en dos partes. Por un lado, el análisis de las respuestas a las preguntas abiertas. Por otro lado, el análisis de las respuestas de las preguntas cerradas.

El resumen de resultados obtenidos de la evaluación del prototipo de navegación y visualización de los itinerarios construidos por la herramienta desarrollada se puede visualizar en la página del prototipo MPOC<sup>56</sup>, como muestra la Ilustración 63.

---

<sup>56</sup> <http://aluned.laraclares.com/encuestas/resultados>

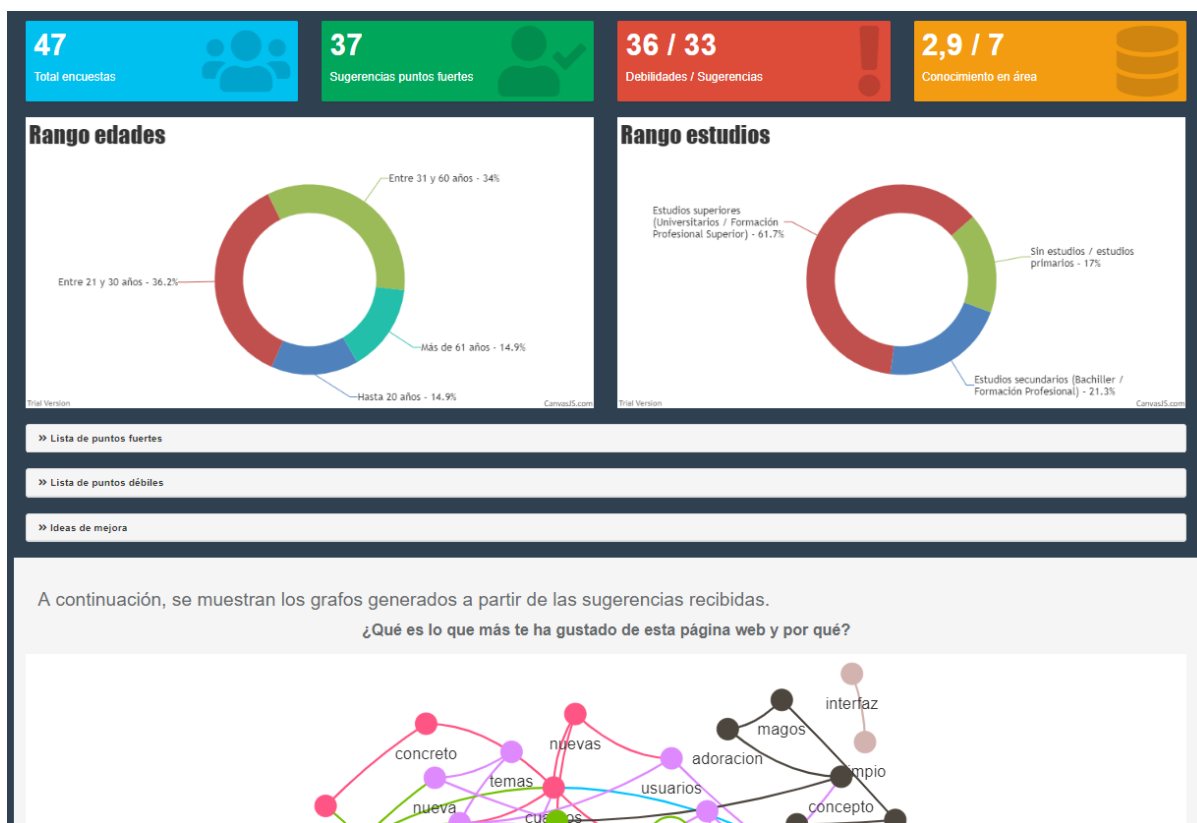


Ilustración 62 Captura de la página de MPOC con el resumen de resultados de las encuestas

### Análisis de las respuestas a preguntas cerradas

Para evaluar las respuestas obtenidas a las preguntas cerradas, se utiliza la herramienta Excel propuesta por Schrepp [98]. Las escalas que se miden en esta encuesta, descritas anteriormente, son:

- Atracción.
- Transparencia.
- Eficiencia.
- Controlabilidad.
- Estimulación
- Novedad.

Esta herramienta tiene las siguientes ventajas:

1. Muestra gráficamente los resultados de las encuestas.
2. Permite detectar casos sospechosos, o encuestas que se responden de forma aleatoria, teniendo en cuenta resultados incongruentes.
3. Estudia el **intervalo de confianza** de cada escala. El intervalo de confianza mide la precisión de la estimación de la media, por lo que un intervalo de confianza menor indica que los datos son más consistentes.
4. Calcula la **correlación** de los ítems de la encuesta que pertenecen a una misma escala, utilizando el **coeficiente Alfa de Cronbach** [96].

5. Calcula una aproximación del tamaño de la muestra que sería necesaria para minimizar el error de los resultados en cada caso.

La metodología llevada a cabo para realizar el análisis de las respuestas en este caso es:

1. Detección de datos sospechosos o posibles respuestas aleatorias.
2. Análisis de los resultados de alto a bajo nivel: (a) por categorías de escalas, (b) por escalas y (c) por ítems de la encuesta.
3. Análisis de consistencia de datos.
4. Cálculo del tamaño de la muestra para mejorar la evaluación.
5. Obtención de conclusiones.

El número de encuestas total es de 48. El primer paso es detectar y sacar de la lista las encuestas detectadas como sospechosas, es decir, que se han respondido de forma aleatoria. Para ello, se utiliza una heurística muy sencilla, que detecta cuán distinto es el valor obtenido en los distintos ítems de cada escala. Si la diferencia es muy grande, probablemente se trate de una respuesta aleatoria. El tamaño final de la muestra es de 40.

Una vez eliminadas las encuestas detectadas como aleatorias, se visualizan los resultados a nivel de categorías de escalas: atracción, calidad pragmática y calidad hedónica. Como muestra la Ilustración 62, en general se obtienen buenos resultados, especialmente teniendo en cuenta que se trata de la primera versión del prototipo. En concreto, los peores resultados se obtienen en cuanto a calidad hedónica o de percepción del usuario. Para entender mejor estos resultados, el siguiente paso es estudiarlos a un nivel intermedio, es decir, a nivel de escalas.

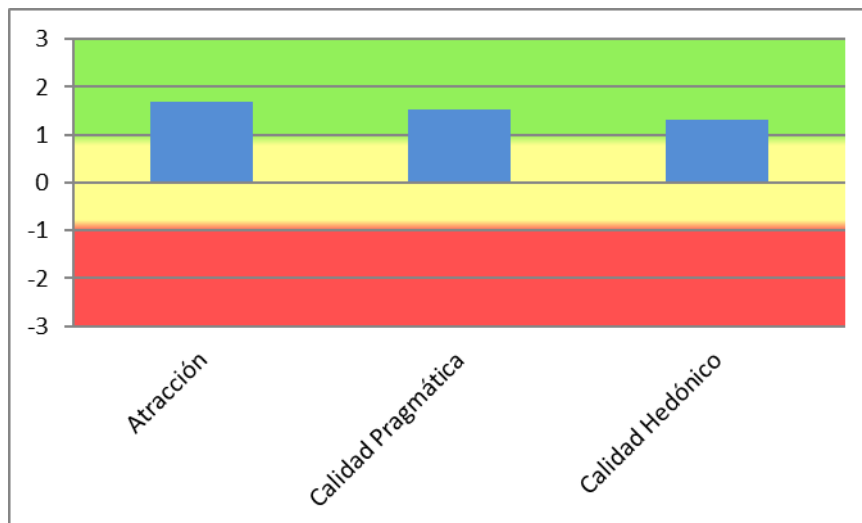


Ilustración 63 Resultados de las encuestas por categorías de escalas

Los resultados a nivel de escalas se muestran en la Ilustración 64. Como se puede ver, las tres escalas con menor puntuación son controlabilidad (calidad pragmática), estimulación (calidad hedónica) y novedad (calidad hedónica).

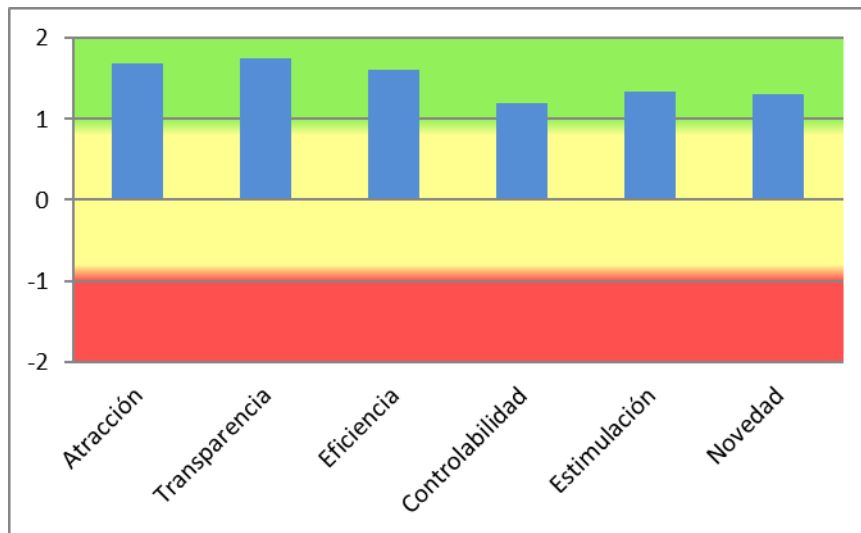


Ilustración 64 Resultados de las encuestas por escalas

Para encontrar dónde están las debilidades del prototipo identificados por los usuarios, se estudian los resultados obtenidos en la encuesta a través de la evaluación de los ítems. Como puede verse en la tabla 8, hay un aspecto que plantea una necesidad de mejora: Impredicible/Predecible. Es decir, si es o no predecible para el usuario qué va a ocurrir al hacer un clic.

Mean	Variance	Std. Dev.	No.	Left	Right	Scale
↑ 1,8	1,6	1,3	40	desagradable	agradable	Atracción
↑ 1,7	2,1	1,5	40	no entendible	entendible	Transparencia
↑ 1,0	2,5	1,6	40	creativo	sin imaginacion	Novedad
↑ 1,6	2,6	1,6	40	facil de aprender	dificil de aprender	Transparencia
↑ 1,7	1,7	1,3	40	valioso	de poco valor	Estimulación
↑ 0,9	1,2	1,1	40	aburrido	emocionante	Estimulación
↑ 1,5	1,4	1,2	40	no interesante	interesante	Estimulación
→ 0,2	2,2	1,5	40	impredicible	predecible	Controlabilidad
↑ 1,5	2,7	1,6	40	rapido	lento	Eficiencia
↑ 1,7	1,7	1,3	40	original	convencional	Novedad
↑ 1,4	0,9	1,0	40	obstructivo	impulsor de apoyo	Controlabilidad
↑ 2,1	1,0	1,0	40	bueno	malo	Atracción
↑ 1,9	1,5	1,2	40	complicado	facil	Transparencia
↑ 1,3	1,6	1,3	40	repeler	atraer	Atracción
↑ 1,4	2,0	1,4	40	convencional	novedoso	Novedad
↑ 1,7	1,2	1,1	40	incomodo	comodo	Atracción
↑ 1,6	1,8	1,3	40	seguro	inseguro	Controlabilidad
↑ 1,4	1,8	1,4	40	activante	adormecedor	Estimulación
↑ 1,6	1,2	1,1	40	cubre expectativas	no cubre expectativas	Controlabilidad
↑ 1,6	1,1	1,0	40	ineficiente	eficiente	Eficiencia
↑ 1,8	1,4	1,2	40	claro	confuso	Transparencia
↑ 1,2	1,7	1,3	40	no pragmatico	pragmatico	Eficiencia
↑ 2,1	0,9	1,0	40	ordenado	sobrecargado	Eficiencia
↑ 1,7	1,1	1,1	40	atractivo	feo	Atracción
↑ 1,7	0,8	0,9	40	simpatico	antipatico	Atracción
↑ 1,2	2,0	1,4	40	conservador	innovador	Novedad

Tabla 8 Resultados de la encuesta realizada para MPOC

Por otro lado, es necesario estudiar la consistencia de los datos, ya que indica si los resultados obtenidos son o no fiables. Para ello, se tiene en cuenta el coeficiente Alfa de Cronbach, que mide la correlación entre los ítems pertenecientes a una misma escala. Una escala se considera, en general, suficientemente consistente si su valor Alfa es mayor de 0.7. Los resultados obtenidos en este sentido se muestran en la tabla 9.

Transparencia		Eficiencia		Controlabilidad		Estimulación		Novedad	
Items	Correlation	Items	Correlation	Items	Correlation	Items	Correlation	Items	Correlation
2, 4	0,39	9, 20	0,60	8, 11	0,02	5, 6	0,40	3, 10	0,22
2, 13	0,60	9, 22	0,05	8, 17	-0,04	5, 7	0,50	3, 15	0,17
2, 21	0,71	9, 23	0,50	8, 19	0,03	5, 18	0,41	3, 26	0,41
4, 13	0,43	20, 22	0,46	11, 17	0,28	6, 7	0,73	10, 15	0,57
4, 21	0,43	20, 23	0,67	11, 19	0,56	6, 18	0,38	10, 26	0,49
13, 21	0,69	22, 23	0,28	17, 19	0,36	7, 18	0,34	15, 26	0,62
<b>Average</b>	0,54	<b>Average</b>	0,43	<b>Average</b>	0,20	<b>Average</b>	0,46	<b>Average</b>	0,41
<b>Alpha</b>	<b>0,83</b>	<b>Alpha</b>	<b>0,75</b>	<b>Alpha</b>	<b>0,50</b>	<b>Alpha</b>	<b>0,77</b>	<b>Alpha</b>	<b>0,74</b>

Tabla 9 Resultados de consistencia obtenidos para cada escala

La única escala que se encuentra por debajo de 0.7 es controlabilidad, lo que quiere decir que los usuarios han respondido con valores incongruentes o aleatorios dentro de los ítems de esta escala (por ejemplo, han respondido que es predecible saber qué ocurrirá al hacer un clic, pero inseguro).

También es importante calcular el valor Alfa para la encuesta general, mostrada en la tabla 10. Como puede verse, los resultados de la encuesta son consistentes, y, por lo tanto, se pueden obtener conclusiones a partir de la misma.

Atracción	
Items	Correlation
1, 12	0,64
1, 14	0,76
1, 16	0,69
1, 24	0,65
1, 25	0,57
12, 14	0,57
12, 16	0,64
12, 24	0,58
12, 25	0,44
14, 16	0,65
14, 24	0,63
14, 25	0,51
16, 24	0,62
16, 25	0,40
24, 25	0,42
<b>Average</b>	0,59
<b>Alpha</b>	<b>0,89</b>

Tabla 10 Cálculo de la consistencia general de la encuesta



La tabla 11 muestra el intervalo de confianza para cada escala. El nivel de confianza es una medida para calcular la precisión de la estimación de la media, donde un menor intervalo de confianza indica una mayor precisión, por lo que los resultados son más fiables. Está relacionada directamente con la medida de consistencia, donde una mayor consistencia implica un menor intervalo de confianza, y, por lo tanto, una mayor precisión.

Confidence intervals (p=0.05) per scale						
Scale	Media	Desv.est.	N	Confianza	Intervalo de conf.	
Atracción	1,692	0,902	40	0,280	1,412	1,971
Transparencia	1,744	1,100	40	0,341	1,403	2,085
Eficiencia	1,600	0,918	40	0,284	1,316	1,884
Controlabilidad	1,200	0,749	40	0,232	0,968	1,432
Estimulación	1,344	0,947	40	0,293	1,050	1,637
Novedad	1,300	1,065	40	0,330	0,970	1,630

Tabla 11 Desviación típica e intervalos de confianza para cada escala

Por último, es importante calcular qué tamaño de la muestra sería necesaria para minimizar el error obtenido. Estos valores dependen de la desviación típica obtenida anteriormente, y de los valores de precisión y probabilidad de error que se quieran alcanzar. La precisión mide la desviación entre la media de la población y la media de la muestra, y la probabilidad de error mide la probabilidad de que los valores reales se encuentren fuera del intervalo de confianza. Teniendo esto en cuenta, se muestra la tabla 11. Como puede verse, para optimizar completamente estos valores, harían falta 805 encuestas, si bien ya se obtienen datos aceptables teniendo en cuenta el ámbito de este trabajo.

Condición	Atracción	Transparencia	Eficiencia	Controlabilidad	Estimulación	Novedad
Prec.=0.5, Prob.Err.=0.1	9	13	9	6	10	12
Prec.=0.5, Prob.Err.=0.05	13	19	13	9	14	17
Prec.=0.5, Prob.Err.=0.01	32	32	22	15	24	30
Prec.=0.25, Prob.Err.=0.1	35	53	37	24	39	49
Prec.=0.25, Prob.Err.=0.05	50	74	52	35	55	70
Prec.=0.25, Prob.Err.=0.01	87	129	90	60	95	121
Prec.=0.1, Prob.Err.=0.1	222	329	229	153	244	309
Prec.=0.1, Prob.Err.=0.05	313	465	324	216	344	436
Prec.=0.1, Prob.Err.=0.01	542	805	561	374	597	756

Tabla 12 Tamaño de la muestra estimado para obtener ciertos valores de precisión y probabilidad de error.

En resumen, se obtienen valores en general, consistentes. Se han detectado una serie de puntos sobre los que hacer hincapié, como indica la Ilustración 65. La escala de controlabilidad aparece debajo de la media, si bien es inconsistente y no es fiable el resultado mostrado. Por otro lado, la estimulación (que mide si es valioso, aburrido, interesante o adormecedor), es un valor consistente y mejorable en este trabajo.

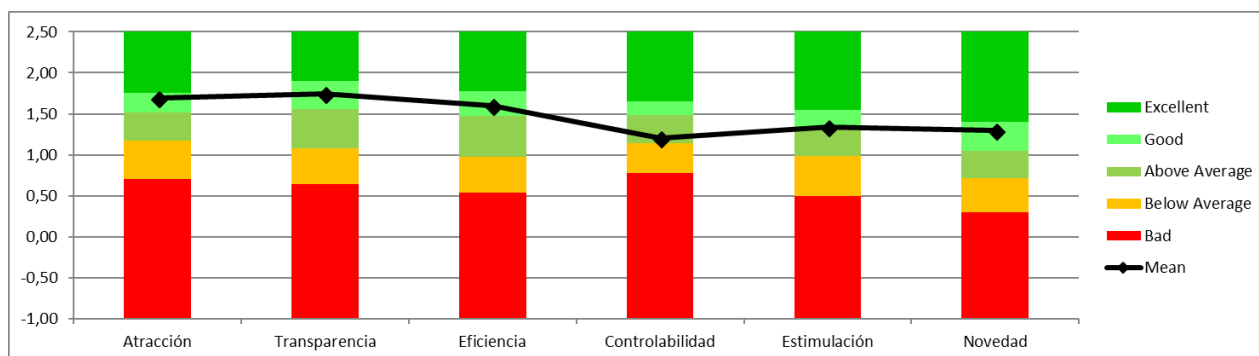


Ilustración 65 Resultados finales de la encuesta

### Análisis de sugerencias

Sobre las sugerencias textuales, se han aplicado técnicas de extracción de palabras clave y creado un grafo con pesos para cada una de las tres preguntas. La intención es encontrar una forma sistemática de analizar las sugerencias textuales aportadas por los usuarios, ayudándose de técnicas de extracción de la información existentes y automatizando parte del proceso. Los pasos seguidos para cada cuestión son los siguientes:

1. Utilizando CoreNLP de Stanford [78], comentado anteriormente en la propuesta, se extraen los sustantivos y adjetivos correspondientes a los textos de las sugerencias.
2. Seleccionando una ventana 2 (palabras que aparecen juntas en los textos), se crea un grafo relacionado, y se utiliza el algoritmo utilizando anteriormente, TextRank [80], para establecer un peso en los nodos. Las Ilustraciones 64, 65 y 66 muestran el grafo creado en cada uno de los tres casos.
3. Visualizando el grafo con Gephi [87], se extraen las comunidades existentes en cada caso. Para ello, se calcula la Modularidad<sup>57</sup>, utilizando el método Louvain [99]. La modularidad de una partición es un valor escalar entre -1 y 1 que mide la densidad de enlaces dentro de las comunidades en comparación con los enlaces entre comunidades. El método consta de dos fases. En primer lugar, busca comunidades "pequeñas" optimizando la modularidad de forma local. En segundo lugar, agrega nodos de la misma comunidad y construye una nueva red cuyos nodos son las comunidades. Estos pasos se repiten iterativamente hasta alcanzar un máximo de modularidad.
4. Se extrae el subgrafo correspondiente a cada comunidad, y se evalúa manualmente, obteniendo las conclusiones pertinentes.

El grafo obtenido en los comentarios referentes a la pregunta “¿Qué es lo que más te ha gustado de esta página web y por qué?” se muestra en la Ilustración 64. Como se puede ver, se resaltan los conceptos considerados como beneficiosos para el sistema: “interfaz limpio”, “información interesante”, “itinerarios interesantes”, “intuitivo”, “sencillo”, etc. En este caso, no se considera necesario estudiar las comunidades encontradas por separado.

<sup>57</sup> <https://github.com/gephi/gephi/wiki/Modularity>



Ilustración 66 Grafo correspondiente a la primera pregunta de la encuesta en MPOC

En el caso de las sugerencias obtenidas a partir de la pregunta “¿Qué es lo que menos te ha gustado de esta página web y por qué?”, el grafo obtenido se muestra en la Ilustración 65. Se detectan siete comunidades distintas, correspondientes a los siete colores del grafo.

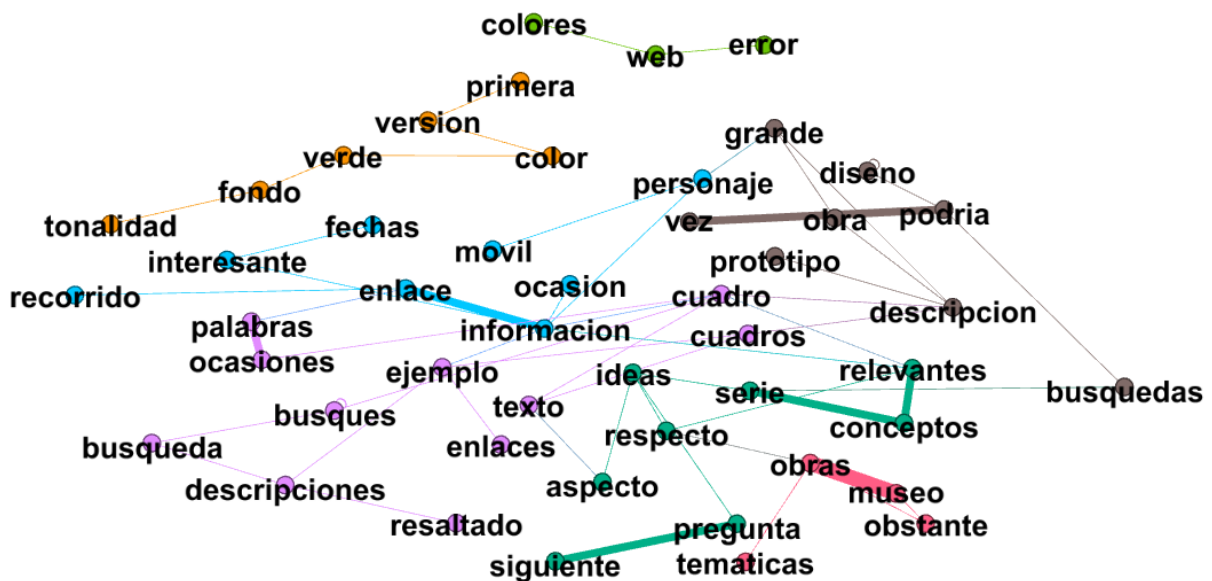


Ilustración 67 Grafo correspondiente a la segunda pregunta de la encuesta en MPOC.

Se pueden obtener las siguientes sugerencias en este caso:

1. Modificar los colores de la página web. En concreto, eliminar el fondo verde de las imágenes.
2. Reducir el tamaño de los textos en la versión para ordenador (pantallas grandes).
3. Obligar a seleccionar un solo tema en la búsqueda de itinerarios. Puesto que se muestra sólo un tema, no se debe permitir escoger más de uno.
4. En ocasiones, el tema del itinerario no está centrado en el concepto seleccionado.
5. En ocasiones, las palabras clave anotadas no corresponden con el concepto existente en el texto. Por ejemplo, cuando se anota “Carlos” en lugar de “Carlos IV”.

6. En la visualización por móvil, el texto de la página de inicio aparece en gris, sería conveniente cambiarla a color negro.
7. Es necesaria mejorar la visualización para móviles.

En el caso de las sugerencias obtenidas a partir de la pregunta “¿Cómo crees que se podría mejorar la página web? ¿Te gustaría añadir alguna funcionalidad en concreto?”, el grafo obtenido se muestra en la Ilustración 66.

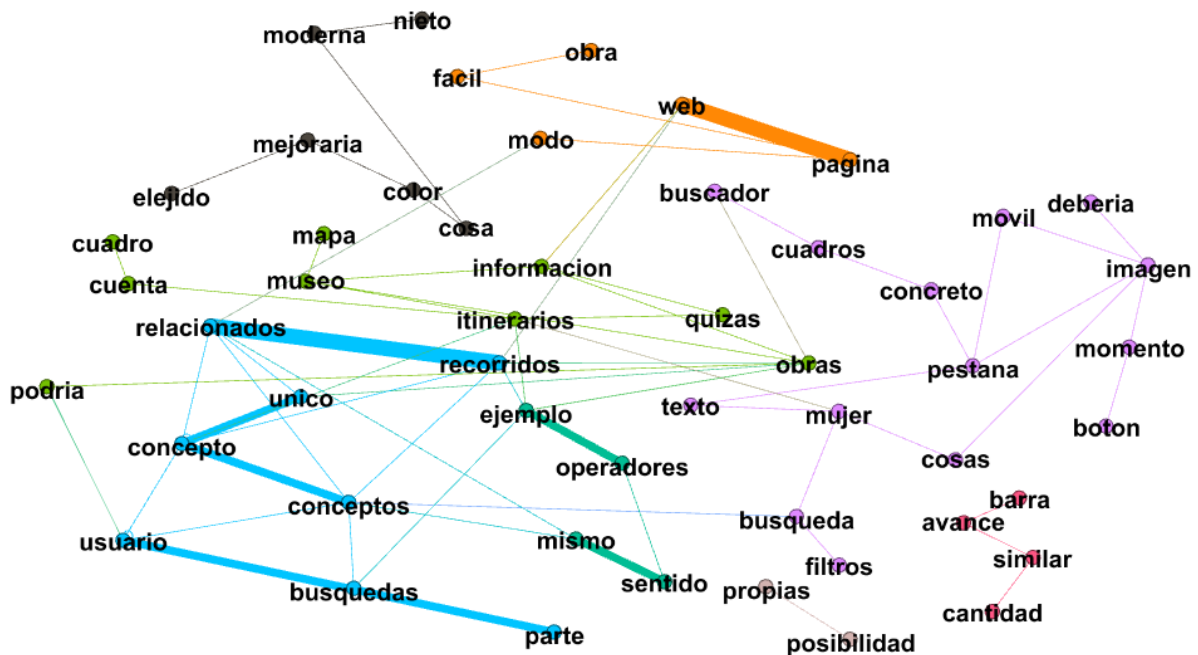


Ilustración 68 Grafo correspondiente a la primera pregunta de la encuesta en MPOC

Las sugerencias de mejora, en este caso, son las siguientes:

1. Permitir personalizar las búsquedas. Además, permitir buscar personajes, eventos, referencias y obras.
2. Permitir crear y editar conceptos e itinerarios.
3. Añadir un enlace al manual de usuario en todas las páginas.
4. Crear los itinerarios teniendo en cuenta la disposición de los cuadros en el museo.
5. Mejorar la interactividad de la página.
6. Añadir material multimedia.
7. Añadir una barra que muestre el porcentaje del itinerario visualizado.
8. Crear itinerarios categorizados por épocas concreta.
9. Mejorar la visualización de las tablas en móvil.

Tras finalizar las encuestas, se llevan a cabo las siguientes mejoras en el sistema:

1. Eliminación del fondo verde las imágenes de las obras.
2. Se reduce el tamaño de los textos para la versión de pantalla grande.
3. Se arregla el problema de selección de varios temas del itinerario.
4. Se modifica el color de la página de inicio en la visualización con móvil.
5. Se arregla el problema de visualización de tablas de las obras en la versión con móvil.

6. Se añade en todas las páginas un enlace a preguntas frecuentes y otro al manual de usuario.
7. Se añade información sobre el porcentaje del itinerario visualizado en cada paso.

El resto de las tareas se plantean como trabajo futuro, ya que demuestran un claro interés por parte del usuario y resulta viable llevarlas a cabo.

### 6.3.2 Análisis de comportamiento en logs

Para analizar el comportamiento de los logs, se ha implementado una página web que muestra los resultados en tiempo real, como muestra la Ilustración 61:

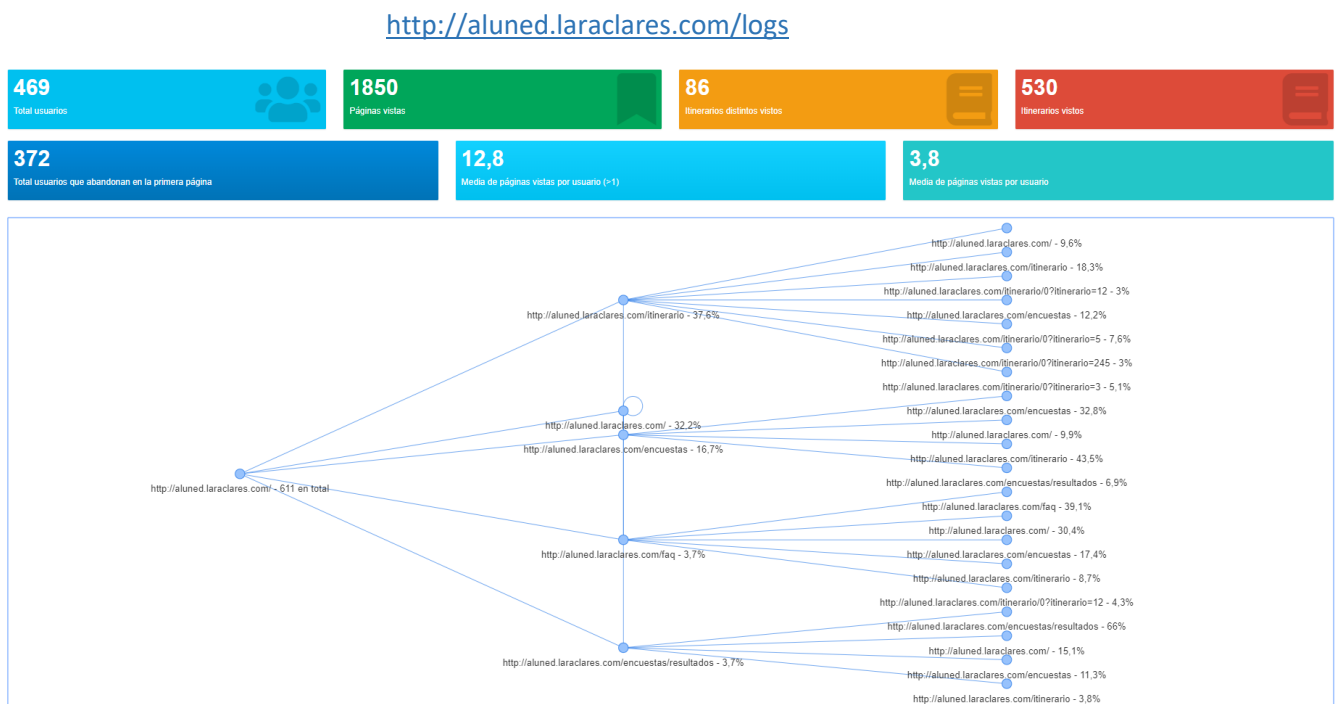


Ilustración 69 Imagen extraída de los logs correspondiente a MPOC

Como puede verse en la Ilustración 62, se han visto en total **1850 páginas** correspondientes a **469 usuarios**. Estos usuarios han visualizado **530 itinerarios** correspondientes a **86 itinerarios distintos**.

Además, 372 usuarios abandonan la página antes de hacer un segundo clic. Esto se debe, principalmente, a las pruebas realizadas en el desarrollo del trabajo sobre el sistema real. Si no se tiene en cuenta a los usuarios que abandonan en el primer clic, encontramos una media de **12,8 clics por usuario**. Esto muestra que los usuarios se interesan, visitan itinerarios y navegan por la página.

Por otro lado, el grafo mostrado analiza el comportamiento de los usuarios. Comenzando a la izquierda del grafo con el nodo de la página de inicio, calcula el porcentaje de clics que se realizan en las páginas en el siguiente paso, terminando en el tercer clic. En este trabajo sólo se analiza el comportamiento a tres niveles, si bien sería recomendable, en un futuro, realizarlo al menos a cuatro niveles, como muestra la media de páginas vistas por usuario.

De las 611 visualizaciones de la página de inicio, el 37.6% de los usuarios acceden a la página de itinerarios, pero sólo un 3.7% de los usuarios acceden a la página de preguntas frecuentes, y menos de un 3% de los usuarios leen el manual. El itinerario más visualizado en primer lugar es “Cuadros relacionados con el personaje Diego Velázquez”.

Por ahora, los datos encontrados en este análisis muestran que los usuarios acceden a una media bastante alta de URLs, es decir, que visualizan, generalmente, más de un itinerario. También muestra que el porcentaje de páginas visualizadas frente a las correspondientes a itinerarios es del 28.6%, un porcentaje más bajo del esperado teniendo en cuenta que la única funcionalidad, además de la creación de las encuestas, es la de ver itinerarios.

## 7 CONCLUSIONES Y TRABAJO FUTURO

---

La generación automática de itinerarios supone, hoy en día, un gran reto debido a la escasez de información estructurada que permita extraer contenido más allá de las descripciones de los objetos. Si pensamos, por ejemplo, en realizar una visita por una ciudad, es fácil encontrar una lista de lugares que visitar e información sobre los mismos, pero no es tan sencillo relacionar las historias y las anécdotas que los envuelven. Ya existen distintas aproximaciones, como la que propone Mulholland [1], creando un entorno web donde los expertos humanistas pueden escribir las historias desde su punto de vista y añadiendo elementos como personajes o eventos relacionados con las mismas.

El objetivo principal de este trabajo es proponer un modelo narrativo y desarrollar y evaluar un prototipo demostrativo para generar automáticamente itinerarios a partir de sus componentes. El modelo narrativo sirve como estructura de datos, que relaciona toda la información existente y permite trabajar con ella a distintos niveles. Por lo tanto, no sólo relaciona los textos con los conceptos, también permite trabajar esa información y utilizarla para distintos fines, como es en este caso la generación automática de itinerarios basados en temáticas para las obras del Museo del Prado. Cada nivel del modelo narrativo plantea una capa de abstracción de la información que contiene. En su nivel más bajo, contiene un grafo con los conceptos (personajes, referencias, obras y eventos) y las relaciones existentes entre éstos. El análisis del grafo, utilizando distintas aproximaciones, permite encontrar automáticamente las temáticas de las que tratan las descripciones de las obras, por lo que es posible reorganizar toda la información existente en las mismas para mostrar al usuario itinerarios basados en temáticas en lugar de en las obras.

La información con la que se cuenta en este trabajo proviene desde fuentes de recursos en abierto como Wikipedia o Wikidata. Gracias a esto, es posible extraer, además de las descripciones, información semántica y conceptos existentes en éstas, como personajes, referencias o eventos. El modelo propuesto utiliza la información de las descripciones y sus conceptos para descubrir, seleccionar y categorizar las temáticas existentes en ellas. En este trabajo se han generado automáticamente 581 itinerarios a partir de Wikipedia, Wikidata y el Museo del Prado. Estos itinerarios, a su vez, están formados por 4823 obras, 1504 personajes, 14166 referencias y 25569 eventos. De los eventos encontrados, 2799 están relacionados con obras, 2090 con personajes y 221 con referencias.

Para evaluar con usuarios los itinerarios generados, se crea un sistema web que permite buscarlos y visualizarlos (en varios soportes, incluyendo el móvil). Se plantea a los usuarios una serie de tareas para familiarizarse con el entorno antes de realizar la encuesta. La evaluación se realiza analizando tres factores: (a) las preguntas cerradas, (b) las preguntas abiertas y (c) el análisis de comportamiento en la página web. Los resultados muestran un claro interés por parte de los usuarios, que plantean una serie de mejoras y nuevas funcionalidades. Una parte de las mejoras se realiza dentro del ámbito de este trabajo, dejando las demás como trabajo futuro.

Un aspecto importante que destacar es que el sistema web permite generar nuevos itinerarios a partir de los párrafos existentes relativos a las temáticas encontradas, por usuarios expertos o especialistas. Esta parte del sistema se encuentra completamente implementada, pero su publicación para el público y su evaluación se plantea como trabajo futuro.

Hay varias líneas de trabajo futuro para mejorar los métodos propuestos:

1. Generalización del modelo narrativo, modificando su estructura de forma que pueda resultar útil para otros fines, como, por ejemplo, la generación de resúmenes extractivos. Esta idea surge a partir de trabajos como el propuesto por Plaza et al. (2011) [90], donde se propone un método para la generación de resúmenes basado en grafos.
2. Estudio de las características propias de la narrativa y el discurso descritas a nivel teórico en el estado del arte (como focalización, cronología, temporización, etc.) para enriquecer el modelo narrativo.
3. Generación de tramas a partir del peso de las aristas, utilizando, por ejemplo, el coeficiente de Jaccard. Esto permitiría crear itinerarios a partir de conjuntos de temáticas, no sólo a partir de los temas encontrados inicialmente.
4. Estudio de cómo mejorar las anotaciones automáticas realizadas sobre los textos, además de encontrar nueva información en los mismos.
5. Implementación en el sistema web de las mejoras planteadas por los usuarios, como la transformación en una red social donde los usuarios puedan crear, editar y publicar en abierto sus propios itinerarios, lo que serviría, indirectamente, para retroalimentar el grafo de conceptos, y, por lo tanto, mejoraría también los resultados obtenidos.
6. Mejora del sistema para analizar el comportamiento de los usuarios, lo que permitiría personalizar las búsquedas de los usuarios y crear un sistema de recomendación de itinerarios.



## PUBLICACIONES

---

- Lara-Clares, A., García-Serrano, A., & Rodrigo, C. (2018, January). **Enrichment of Accessible LD and Visualization for Humanities: MPOC Model and Prototype**. In Metadata and Semantic Research: 11th International Conference, MTSR 2017 Tallinn, Estonia, November 28-December 1, 2017, Proceedings (Vol. 755, p. 327). Springer.
- García-Serrano, A, Lara-Clares, A., & Rodrigo, C. (2017). **Narrativas con acceso a datos en abierto: MPOC, un sistema de información web sobre el Museo del Prado**, III Congreso de la Sociedad Internacional Humanidades Digitales Hispánicas Sociedades, políticas, saberes. Málaga, 18(20).
- (*Enviado, esperando revisiones*) Lara-Clares, A., García-Serrano, A., & Rodrigo, C. (2018). **Un sistema de información Web para el diseño de itinerarios en un museo con datos en abierto: El museo del Prado**. Artnodes Journal, Digital Humanities. Hispanic societies, politics and knowledge.



## REFERENCIAS

---

- [1] P. Mulholland, A. Wolff, E. Kilfeather, M. Maguire, and D. O'Donovan, "Modelling museum narratives to support visitor interpretation," in *Artificial Intelligence for Cultural Heritage*, L. Bordoni, F. Mele, and A. Sorgente, Eds. Cambridge Scholars Publishing, 2016, pp. 3–22.
- [2] H. Cunningham, "Information extraction, automatic," *Encycl. Lang. Linguist.*, pp. 665–677, 2005.
- [3] P. Mulholland, A. Wolff, and E. Kilfeather, "Storyscope: Supporting the authoring and reading of museum stories using online data sources," in *Proceedings of the ACM Web Science Conference*, 2015, p. 23.
- [4] "Storyspace: Storyspace." [Online]. Available: <http://www.eastgate.com/storyspace/>. [Accessed: 14-Mar-2017].
- [5] K. M. Brooks, "Do story agents use rocking chairs? The theory and implementation of one model for computational narrative," in *Proceedings of the fourth ACM international conference on Multimedia*, 1997, pp. 317–328.
- [6] S. B. Chatman, *Story and discourse: Narrative structure in fiction and film*. Cornell University Press, 1980.
- [7] T. Todorov, *Grammaire du "D{é}cam{é}ron": par Tzvetan Todorov*, vol. 3. Mouton, 1969.
- [8] E. Branigan, *Narrative comprehension and film*. Routledge, 2013.
- [9] G. P. Zarri, "Representation and Management of Complex 'Narrative' Information," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8002, pp. 118–137, 2014.
- [10] G. Genette, "Narrative Discourse," 1980.
- [11] G. Genette, *Narrative discourse revisited*. Cornell University Press, 1988.
- [12] M. Bal, *Narratology: Introduction to the theory of narrative*. University of Toronto Press, 2009.
- [13] I. Swartjes and M. Theune, "A fabula model for emergent narrative," *Technol. Interact. Digit. Storytell. ...*, pp. 49–60, 2006.
- [14] P. Mulholland and T. Collins, "Using digital narratives to support the collaborative learning and exploration of cultural heritage," *Proc. - Int. Work. Database Expert Syst. Appl. DEXA*, vol. 2002–Janua, pp. 527–531, 2002.
- [15] R. C. Schank, *Tell me a story: A new look at real and artificial memory*. Scribner New York, 1990.
- [16] R. C. Schank and R. P. Abelson, *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press, 1977.
- [17] M. Jahn, "Manfred Jahn Narratology: A Guide to the Theory of Narrative," 2017.
- [18] J. H. C. N.-C. L. B. Q. 7. I. M. 1998 D. U. E. 11-04-08 Murray, "Hamlet on the holodeck : the future of narrative in cyberspace," vol. 45, no. Summer, p. xii, 324 p., 1998.
- [19] Manfred Jahn, "Narratology: A Guide to the Theory of Narrative," 2017. [Online]. Available: <http://www.uni-koeln.de/~ame02/pppn.htm>. [Accessed: 09-Jan-2018].

- [20] F. De Saussure, "Nature of the linguistic sign," *Course Gen. Linguist.*, 1916.
- [21] C. P. Casparis, *Tense without time: the present tense in narration*, vol. 84. Francke, 1975.
- [22] U. Margolin, "Of what is past, is passing, or to come: Temporality, aspectuality, modality, and the nature of literary narrative," *Narrat. New Perspect. Narrat. Anal.*, pp. 142–166, 1999.
- [23] G. Muller, "Morphologische Poetik: Gesammelte Aufsätze," Ed. Elena Muller. *Tübingen Niemeyer*, 1968.
- [24] S. Rimmon-Kenan, *Narrative fiction: Contemporary poetics*. Routledge, 2003.
- [25] "DLE: diálogo - Diccionario de la lengua española - Edición del Tricentenario." [Online]. Available: <http://dle.rae.es/?id=DetWqMJ>. [Accessed: 10-Jan-2018].
- [26] H. C. Bunt, A. F. V Van Katwijk, L. F. Muller, and F. L. Van Nes, "Dialogue control acts," *Inst. Percept. Res. Annu. Prog. Rep.*, vol. 15, pp. 95–99, 1980.
- [27] H. Bunt, V. Petukhova, D. Traum, and J. Alexandersson, "Dialogue Act Annotation with the ISO 24617-2 Standard," in *Multimodal Interaction with W3C Standards*, Springer, 2017, pp. 109–135.
- [28] H. Bunt, "Guidelines for using ISO standard 24617-2," 2017.
- [29] D. Traum, "Computational Approaches to Dialogue," *Routledge Handb. Lang. Dialogue*, p. 143, 2017.
- [30] A. Garcia-Serrano and J. Calle-Gómez, "A cognitive Architecture for the design of an Interaction Agent," in *International Workshop on Cooperative Information Agents*, 2002, pp. 82–89.
- [31] O. L. (auth. . Verena Rieser, *Reinforcement Learning for Adaptive Dialogue Systems: A Data-driven Methodology for Dialogue Management and Natural Language Generation*, 1st ed. Springer-Verlag Berlin Heidelberg, 2011.
- [32] J. L. Austin, *How to do things with words*. Oxford university press, 1975.
- [33] J. R. Searle, *Speech acts: An essay in the philosophy of language*, vol. 626. Cambridge university press, 1969.
- [34] D. R. Traum and J. F. Allen, "Discourse obligations in dialogue processing," in *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, 1994, pp. 1–8.
- [35] J. Porteous, F. Charles, C. Smith, M. Cavazza, J. Mouw, and P. Van Den Broek, "An Interactive Narrative Platform for Story Understanding Experiments (Demonstration)," *Proc. 16th Int. Conf. Auton. Agents Multiagent Syst. (AAMAS 2017)*, S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil., pp. 1808–1810, 2017.
- [36] "FicciónInteractiva.com - Capitán Pirata (iOS, Android)." [Online]. Available: <http://ficcióninteractiva.es/capitan-pirata.html>. [Accessed: 10-Jan-2018].
- [37] B. Endrass, C. Klimmt, G. Mehlmann, E. Andre, and C. Roth, "Designing user-character dialog in interactive narratives: An exploratory experiment," *IEEE Trans. Comput. Intell. AI Games*, vol. 6, no. 2, pp. 166–173, 2014.
- [38] M. & Mateas and P. Sengers, "Narrative Intelligence," *AAAI Fall Symp. Narrat. Intell.*, pp. 1–10, 1999.
- [39] G. P. Zarri, "Modeling and Advanced Exploitation of eChronicle Information."

- [40] "TEI: Text Encoding Initiative." [Online]. Available: <http://www.tei-c.org/index.xml>. [Accessed: 07-Feb-2018].
- [41] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, and F. Yergeau, "Extensible markup language (XML).," *World Wide Web J.*, vol. 2, no. 4, pp. 27–66, 1997.
- [42] "World Wide Web Consortium (W3C)." [Online]. Available: <https://www.w3.org/>. [Accessed: 07-Feb-2018].
- [43] "Overview of SGML Resources." [Online]. Available: <https://www.w3.org/MarkUp/SGML/>. [Accessed: 07-Feb-2018].
- [44] "W3C HTML." [Online]. Available: <https://www.w3.org/html/>. [Accessed: 07-Feb-2018].
- [45] "RDF - Semantic Web Standards." [Online]. Available: <https://www.w3.org/RDF/>. [Accessed: 17-Oct-2017].
- [46] "SPARQL Protocol for RDF." [Online]. Available: <https://www.w3.org/TR/rdf-sparql-protocol/>. [Accessed: 27-Feb-2018].
- [47] "SPARQL Query Results XML Format (Second Edition)." [Online]. Available: <https://www.w3.org/TR/rdf-sparql-XMLres/>. [Accessed: 27-Feb-2018].
- [48] "SPARQL 1.1 Overview." [Online]. Available: <https://www.w3.org/TR/sparql11-overview/>. [Accessed: 17-Oct-2017].
- [49] M. Hearst, "What is text mining," *SIMS, UC Berkeley*, 2003.
- [50] "Wikipedia:Acerca de - Wikipedia, la enciclopedia libre." [Online]. Available: [https://es.wikipedia.org/wiki/Wikipedia:Acerca\\_de](https://es.wikipedia.org/wiki/Wikipedia:Acerca_de). [Accessed: 27-Feb-2018].
- [51] "Wikidata." [Online]. Available: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page). [Accessed: 27-Feb-2018].
- [52] "Wikipedia, la enciclopedia libre." [Online]. Available: <https://es.wikipedia.org/wiki/Wikipedia:Portada>. [Accessed: 27-Feb-2018].
- [53] "Wikimedia Commons." [Online]. Available: <https://commons.wikimedia.org/wiki/Commons:Welcome>. [Accessed: 27-Feb-2018].
- [54] "Fundación Wikimedia." [Online]. Available: <https://www.wikidata.org/wiki/Help:Wikimedia>. [Accessed: 27-Feb-2018].
- [55] "Wikidata Query Service." [Online]. Available: <https://query.wikidata.org/>. [Accessed: 27-Feb-2018].
- [56] "Wikibase/API - MediaWiki." [Online]. Available: <https://www.mediawiki.org/wiki/Wikibase/API>. [Accessed: 27-Feb-2018].
- [57] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge," in *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 2008, pp. 1247–1250.
- [58] "es.dbpedia.org." [Online]. Available: <http://es.dbpedia.org/>. [Accessed: 27-Sep-2017].
- [59] "DLE: ontología - Diccionario de la lengua española - Edición del Tricentenario." [Online]. Available: <http://dle.rae.es/?id=R5BOYYh>. [Accessed: 27-Feb-2018].
- [60] "El Museo del Prado se sitúa a la vanguardia de los museos en la Red con una web que transforma por completo la experiencia de su versión digital - Noticia - Museo Nacional del

- Prado,” 2015. [Online]. Available: <https://www.museodelprado.es/actualidad/noticia/el-museo-del-prado-se-situa-a-la-vanguardia-de/ac800655-62f6-41d9-ba1b-5907726bff07>. [Accessed: 06-Mar-2018].
- [61] “Modelo Ontológico - Museo Nacional del Prado.” [Online]. Available: <https://www.museodelprado.es/modelo-semantic-digital/modelo-ontologico>. [Accessed: 14-Mar-2017].
- [62] M. Doerr, C.-E. Ore, and S. Stead, “The CIDOC Conceptual Reference Model: A New Standard for Knowledge Sharing,” in *Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling - Volume 83*, 2007, pp. 51–56.
- [63] M. Doerr and P. LeBoeuf, “Modelling Intellectual Processes: The FRBR - CRM Harmonization,” in *Digital Libraries: Research and Development: First International DELOS Conference, Pisa, Italy, February 13-14, 2007, Revised Selected Papers*, C. Thanos, F. Borri, and L. Candela, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 114–123.
- [64] The IFLA.ORG domain, “IFLA -- Creating New Standards,” 2016. [Online]. Available: <https://www.ifla.org/node/8719>. [Accessed: 10-Dec-2017].
- [65] B. N. de España, “MARC 21 para Registros Bibliográficos,” 2009.
- [66] “BIBFRAME - Bibliographic Framework Initiative (Library of Congress).” [Online]. Available: <https://www.loc.gov/bibframe/>. [Accessed: 15-Dec-2017].
- [67] “Home - schema.org.” [Online]. Available: <http://schema.org/>. [Accessed: 15-Dec-2017].
- [68] W. W. W. Consortium and others, “JSON-LD 1.0: a JSON-based serialization for linked data,” 2014.
- [69] “DECIPHER Project | IMMA BLOG.” [Online]. Available: <https://immablog.org/2014/01/13/decipher-project/>. [Accessed: 16-Mar-2017].
- [70] P. Mulholland, A. Wolff, and T. Collins, “Curate and storyspace: An ontology and web-based environment for describing curatorial narratives,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7295 LNCS, pp. 748–762, 2012.
- [71] “1. Getting Started | Storyscope 7 Test.” [Online]. Available: <http://storyscope7test.ssl.co.uk/faq>. [Accessed: 14-Mar-2017].
- [72] “DLE: agente - Diccionario de la lengua española - Edición del Tricentenario.” [Online]. Available: <http://dle.rae.es/?id=14q5hDO>. [Accessed: 14-Mar-2017].
- [73] Rainer TYPKE, “DECIPHER Integrated Narrative Interfaces,” 2013.
- [74] “EU Project CROSSCULT,” 2016. [Online]. Available: <http://www.crosscult.eu/en/about/background/>. [Accessed: 15-Dec-2017].
- [75] A. D. B, A. Dahroug, A. Gil-solla, M. Ramos-cabrer, and Y. Blanco-fern, “Developing Quiz Games Linked to Networks of Semantic Connections Among Cultural Venues A Tool to Design CROSSCULT Experiences,” vol. 1, no. iii, pp. 239–246, 2017.
- [76] “Word2vec - Google Code Archive - Long-term storage for Google Code Project Hosting.” [Online]. Available: <https://code.google.com/archive/p/word2vec/>. [Accessed: 15-Dec-2017].
- [77] “Python.org.” [Online]. Available: <https://www.python.org/>. [Accessed: 21-May-2017].
- [78] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. Mcclosky, “The Stanford CoreNLP Natural Language Processing Toolkit.”

- [79] “textrank 0.1.0 : Python Package Index.” [Online]. Available: <https://pypi.python.org/pypi/textrank/0.1.0>. [Accessed: 01-Jun-2017].
- [80] R. Mihalcea and P. Tarau, “TextRank: Bringing order into texts,” *Proc. EMNLP*, vol. 85, pp. 404–411, 2004.
- [81] “Lista de propiedades - Wikidata.” [Online]. Available: [https://www.wikidata.org/wiki/Wikidata:List\\_of\\_properties/es](https://www.wikidata.org/wiki/Wikidata:List_of_properties/es). [Accessed: 06-Mar-2018].
- [82] “MediaWiki API help - Wikidata.” [Online]. Available: <https://www.wikidata.org/w/api.php>. [Accessed: 06-Mar-2018].
- [83] “Dynamic Fields | Apache Solr Reference Guide 6.6.” [Online]. Available: [https://lucene.apache.org/solr/guide/6\\_6/dynamic-fields.html](https://lucene.apache.org/solr/guide/6_6/dynamic-fields.html). [Accessed: 12-Mar-2018].
- [84] Stanford NLP Group, “Software - The Stanford Natural Language Processing Group.” [Online]. Available: <https://nlp.stanford.edu/software/>. [Accessed: 12-Apr-2018].
- [85] K. Coursey, R. Mihalcea, and W. Moen, “Using encyclopedic knowledge for automatic topic identification,” in *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, 2009, pp. 210–218.
- [86] S. Brin and L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Computer Networks and ISDN Systems, vol. 30, 1998.
- [87] “Gephi - The Open Graph Viz Platform.” [Online]. Available: <https://gephi.org/>. [Accessed: 23-Apr-2018].
- [88] M. Jacomy, S. Heymann, T. Venturini, and M. Bastian, “ForceAtlas2, A Continuous Graph Layout Algorithm for Handy Network Visualization,” 2012.
- [89] A. D. Delgado, R. M. Unanue, A. P. García-Plaza, and V. Fresno, “Unsupervised Real-Time company name disambiguation in {Twitter},” in *ICWSM Workshop on Real-Time Analysis and Mining of Social Streams, June 4, 2012, Dublin, Ireland, 2012*, pp. 25–28.
- [90] L. P. Morales, P. G. Gómez-Navarro, and A. D. Esteban, *Uso de grafos semánticos en la generación automática de resúmenes y estudio de su aplicación en distintos dominios: biomedicina, periodismo y turismo*. Universidad Complutense de Madrid, 2011.
- [91] N. Aletras, T. Baldwin, J. H. Lau, and M. Stevenson, “Evaluating topic representations for exploring document collections,” *J. Assoc. Inf. Sci. Technol.*, vol. 68, no. 1, pp. 154–167, 2017.
- [92] N. Bevan, J. Carter, and S. Harker, “ISO 9241-11 revised: What have we learnt about usability since 1998?,” in *International Conference on Human-Computer Interaction*, 2015, pp. 143–151.
- [93] M. Schrepp, A. Hinderks, and J. Thomaschewski, “Applying the user experience questionnaire (UEQ) in different evaluation scenarios,” in *International Conference of Design, User Experience, and Usability*, 2014, pp. 383–392.
- [94] B. Laugwitz, M. Schrepp, and T. Held, “Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten.,” in *Mensch & Computer*, 2006, pp. 125–134.
- [95] B. Laugwitz, T. Held, and M. Schrepp, “Construction and evaluation of a user experience questionnaire,” in *Symposium of the Austrian HCI and Usability Engineering Group*, 2008, pp. 63–76.
- [96] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, no. 3, pp. 297–334, Sep. 1951.

- [97] M. Schrepp, "User Experience Questionnaire Handbook," *All you need to know to apply UEQ successfully your Proj.*, 2015.
- [98] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Construction of a Benchmark for the User Experience Questionnaire (UEQ).," *IJIMAI*, vol. 4, no. 4, pp. 40–44, 2017.
- [99] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. theory Exp.*, vol. 2008, no. 10, p. P10008, 2008.
- [100] "DLE RAE: deixis." [Online]. Available: <http://dle.rae.es/?id=C5EQPe1>. [Accessed: 10-Jan-2018].
- [101] "DLE: concepto, ta - Diccionario de la lengua española - Edición del Tricentenario." .
- [102] "DLE: evento - Diccionario de la lengua española - Edición del Tricentenario." [Online]. Available: <http://dle.rae.es/?id=H9JpZQS>. [Accessed: 14-Mar-2017].
- [103] O. Lassila and R. Swick, Eds., "Resource Description Framework (RDF) Model and Syntax Specification." 1999.
- [104] D. Berrueta *et al.*, "SIOC Core Ontology Specification. W3C Member Submission 12 June 2007," *World Wide Web Consort.*, 2007.







## MANUAL DE USUARIO

IEEE 1063-2001

ALICIA LARA CLARES  
UNED

# CONTENIDO

---

Índice de ilustraciones .....	115
Índice de tablas .....	116
1 Uso del documento .....	117
2 Introducción .....	118
3 Ámbito del trabajo .....	119
4 Funcionamiento de MPOC .....	121
4.1 Navegación por la web .....	121
4.2 Manual de uso de la web .....	126
4.2.1 Visualización de itinerarios existentes .....	126
4.2.2 Creación y modificación de itinerarios (sólo autores y administradores) .....	128
4.2.3 Edición de elementos narrativos .....	132
4.2.4 Visualización de logs .....	133
4.2.5 Administración de la encuesta de evaluación .....	133
5 Reporte de errores y resolución de problemas .....	134
6 Glosario de términos .....	134
Referencias .....	135

## ÍNDICE DE ILUSTRACIONES

---

Ilustración 1 Captura de la página inicial de MPOC.....	118
Ilustración 2 Captura del formulario de registro de usuario en MPOC .....	121
Ilustración 3 Pantalla de inicio de MPOC.....	122
Ilustración 4 Pantalla de inicio de MPOC para autores .....	123
Ilustración 5 Pantalla de inicio de MPOC para administrador .....	124
Ilustración 6 Menú superior para usuarios visitante .....	125
Ilustración 7 Menú superior para usuarios autores.....	125
Ilustración 8 Menú superior para usuarios administradores.....	125
Ilustración 9 Visualización o búsqueda de un itinerario existente .....	126
Ilustración 10 Ejemplo de segmento de itinerario.....	127
Ilustración 11 Visualización del personaje Miguel Ángel.....	128
Ilustración 12 Visualización de la referencia Manierismo .....	128
Ilustración 13 Búsqueda de segmentos .....	129
Ilustración 14 Búsqueda y selección de segmentos.....	129
Ilustración 15 Ejemplo de segmentos combinados .....	130
Ilustración 16 Ejemplo de selección de segmentos .....	130
Ilustración 17 Asignación de título y creación de itinerario .....	131
Ilustración 18 Imagen correspondiente con la ventana de edición de itinerario.....	131
Ilustración 19 Selección de personaje .....	131
Ilustración 20 Ventana emergente de selección de elemento narrativo .....	132
Ilustración 21 Selección de personaje para edición.....	132
Ilustración 22 Edición de personaje.....	133

## ÍNDICE DE TABLAS

---

Tabla 1 Permisos de las funcionalidades por roles de usuario en MPOC .....	120
Tabla 2 Relación entre funcionalidades y botones mostrados en la página de inicio .....	124
Tabla 3 Relación entre funcionalidades y enlaces directos a la página web .....	125

# 1 USO DEL DOCUMENTO

---

MPOC es un sistema de información basado en web creado como parte del trabajo “Generación automática de itinerarios basados en LOD”.

La creación de este manual se ha realizado conforme al estándar para la creación de manuales de usuario **IEEE 1063-2001** (“*IEEE Standard for Software User Documentation*”) [1].

La estructura del documento es el siguiente:

- Capítulo 2. Introducción.

En este capítulo se introduce brevemente la motivación del problema y el caso de estudio en el que se trabaja.

- Capítulo 3. Ámbito del trabajo.

En este capítulo se presenta una visión general del contexto asociado al trabajo y encuadrando el objetivo de la web presentada en este manual dentro de los objetivos del trabajo final.

- Capítulo 3. Funcionamiento de la web.

Este capítulo muestra los tutoriales asociados a cada rol de usuario, incluyendo descripciones de las funcionalidades de cada sección y sus limitaciones.

- Capítulo 4. Reporte de errores y resolución de problemas.

En este capítulo se describe brevemente el método a seguir en caso de errores.

- Capítulo 5. Glosario de términos.

En este capítulo se definen los conceptos asociados a este proyecto.

- Capítulo 6. Referencias.

En este capítulo se muestran las referencias a los que se hace mención a lo largo del presente manual.

## 2 INTRODUCCIÓN

---

Hoy en día, gracias a fuentes de recursos web, es posible obtener información acerca de casi cualquier lugar o monumento antes de ir a visitarlo. Pese a esto, es imposible obtener los beneficios que aportan los guías turísticos a las visitas, ya que un guía turístico no sólo ofrece esta información estática, sino también su contexto, las relaciones que existen entre los objetos de la visita y las anécdotas que surgen en relación con éstos. En este trabajo se propone una puerta de enlace para aportar ese dinamismo a la información existente en recursos enlazados en abierto en la web, mostrando la información como algo moldeable, que permita establecer relaciones y aporte una nueva forma de preparar una visita al Museo del Prado, tanto presencial como virtual.

La web asociada a este trabajo se encuentra en la siguiente URL, y su página principal tiene el aspecto mostrado en la Ilustración 1:

<http://aluned.laraclares.com>



*Ilustración 70 Captura de la página inicial de MPOC*

El objetivo de este sistema es permitir visualizar los itinerarios creados sobre las obras del Museo del Prado, permitiendo al usuario conocer las obras desde nuevos puntos de vista, centrándose en la información semántica contenida acerca de las mismas.

### 3 ÁMBITO DEL TRABAJO

---

El trabajo en el que se enmarca este sistema web propone la **creación un sistema basado en un modelo narrativo propuesto para la gestión de itinerarios**. Un **itinerario** se puede ver como un conjunto de narraciones acerca de los elementos que lo conforman, que comparten un tema en común. Además de plantear un sistema capaz de generar de forma automática una serie de itinerarios, se destaca el uso de herramientas y recursos en abierto, aprovechando la información que ya existe, tanto en forma de metadatos como en forma textual, para extraer, organizar, seleccionar y publicar los itinerarios clasificándolos por temáticas. Por ello, la información se obtiene de fuentes de recursos en abierto, como Wikidata y Wikipedia.

El caso de estudio tiene por objetivo diseñar una serie de itinerarios o recorridos por las obras del Museo del Prado, permitiendo al usuario final seleccionar la temática de este. Está enmarcado en el proyecto MUSACCES, museología e integración social: la difusión del patrimonio artístico y cultural del Museo del Prado a colectivos con especial accesibilidad (invidentes, sordos y reclusos). El Consorcio MUSACCES propone acciones de museología e integración social en un entorno de trabajo interdisciplinar. Se busca acercar el Museo del Prado a colectivos de especial accesibilidad (invidentes, personas sordas y reclusos) con apoyo de las Tecnologías de la Información y la Comunicación, así como redefiniendo los recursos tradicionales. Los trabajos de investigación comenzaron el 1 de enero de 2016 con participantes de ámbitos multidisciplinares pertenecientes a grupos de investigación de la Universidad Complutense de Madrid, de la Universidad Nacional de Educación a Distancia (UNED) y de la Universidad Autónoma de Madrid (UAM), sobre información del Museo del Prado y otras empresas e instituciones.

Los elementos que conforman el modelo narrativo en este dominio son: personajes, obras, referencias y eventos. Los **personajes** pueden ser reales o ficticios, y pueden estar relacionados directamente con la obra (como el autor de la misma), como indirectamente (si tiene alguna relación con la misma, como un Felipe IV en Las Meninas<sup>58</sup>). Las **obras** son todas las pertenecientes al Museo del Prado de las que se ha obtenido la información. Las **referencias** son todos los conceptos con información semántica anotados en las descripciones de las obras (por ejemplo, “perro”, “bóveda”, “Manierismo”, “Barroco”, etc.). Los **eventos**, por su parte, son los acontecimientos que aportan el contexto histórico o político a los itinerarios.

Para llevar a cabo el trabajo propuesto, se plantean tres objetivos:

- (1) la creación de un modelo narrativo, que sirva como estructura para almacenar y trabajar con la información desde las distintas fuentes de recursos. Un modelo narrativo establece los conceptos y las relaciones entre los mismos involucrados en todo el proceso de la creación y visualización de las narrativas asociadas a una visita turística.
- (2) extracción de la información desde las fuentes de recursos en abierto, así como la anotación de las narrativas y su integración en el modelo narrativo planteado anteriormente.
- (3) diseño e implementación de una web para la visualización de los resultados. Este sistema planteará, haciendo uso de la información extraída en el punto anterior, un sistema integral de gestión de itinerarios.

---

<sup>58</sup> [https://es.wikipedia.org/wiki/Las\\_meninas](https://es.wikipedia.org/wiki/Las_meninas)



El sistema web al que hace referencia este manual de usuario forma parte del tercer objetivo. Los detalles técnicos se especifican en el ANEXO II “Especificación de requisitos IEEE-830”. A continuación, se describen brevemente los usuarios identificados en el mismo y las funciones que aborda el sistema.

Los usuarios existentes en este sistema web son:

- ✚ Usuario administrador. Posee todos los permisos de acceso, tanto a la web como a los datos almacenados en el sistema.
- ✚ Usuario autor. Es el usuario experto encargado de la creación y edición de itinerarios y elementos narrativos asociados al mismo.
- ✚ Usuario invitado. Es el visitante. Usuario con permisos de visualización de los itinerarios y elementos narrativos existentes.

Las funciones principales que aborda este sistema son:

1. Creación y modificación de itinerarios.
2. Edición de elementos narrativos (obras, personajes y referencias).
3. Visualización de itinerarios existentes.
4. Visualización de logs.
5. Administración de los usuarios.
6. Administración de la encuesta de evaluación.

La siguiente tabla muestra los permisos de los usuarios para utilizar las funcionalidades descritas anteriormente:

	<b>Administrador</b>	<b>Autor</b>	<b>Visitante</b>
<b>Creación y modificación de itinerarios</b>	Si	Si	Si
<b>Edición de elementos narrativos</b>	Si	Si	No
<b>Visualización de itinerarios existentes</b>	Si	Si	No
<b>Visualización de logs</b>	Si	No	No
<b>Administración de los usuarios</b>	Si	No	No
<b>Administración de la encuesta de evaluación</b>	Si	No	No

*Tabla 13 Permisos de las funcionalidades por roles de usuario en MPOC*

## 4 FUNCIONAMIENTO DE MPOC

---

Esta sección se divide en dos partes. En primer lugar, se muestra la navegación a lo largo de la web, con las distintas secciones que ofrece para cada tipo de usuario. En segundo lugar, se muestran los tutoriales para cada funcionalidad de las descritas en la sección anterior.

Para distinguir entre las funcionalidades para los tipos de usuario, se utilizará un sistema de colores. En caso de tratarse de una funcionalidad restringida para usuarios visitantes, se utilizará el **color azul**. En caso de ser una funcionalidad restringida para usuario administrador, se utilizará el **color rojo**.

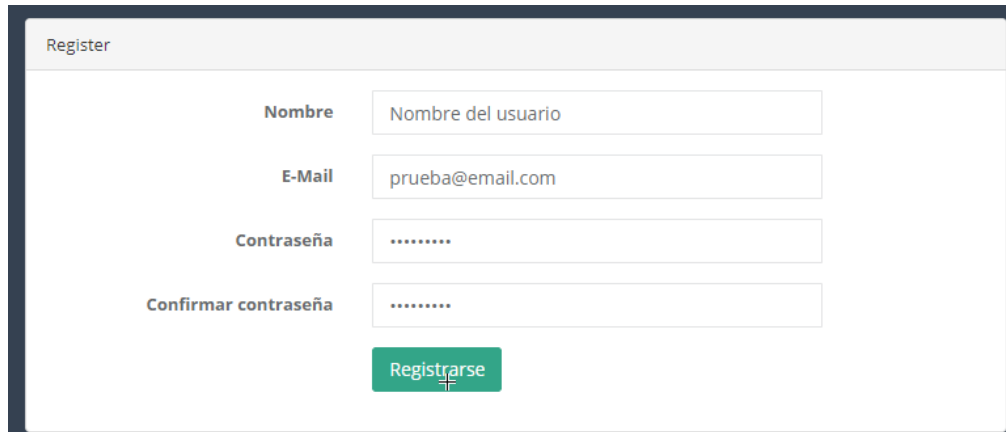
### 4.1 NAVEGACIÓN POR LA WEB

Para acceder a la página, se accede mediante la siguiente URL:

<http://aluned.laraclares.com/>

El primer paso para acceder por primera vez a la página es registrarse en la misma. Para ello, hay que hacer clic en la esquina superior derecha “Registro”, o acceder a la siguiente URL y completar el formulario mostrado en la Ilustración 2:

<http://aluned.laraclares.com/register>



*Ilustración 71 Captura del formulario de registro de usuario en MPOC*

- El registro se encuentra abierto a cualquier persona, realizándose de manera inmediata al rellenar el formulario.

En caso de estar ya registrado, para iniciar sesión se puede acceder a través del botón “iniciar sesión” o la URL siguiente y rellenando los campos de email y contraseña:

<http://aluned.laraclares.com/login>

Una vez registrado el usuario, se accede de nuevo a la página principal, como se muestra en la Ilustración 3:

**MPOC**  
Sistema web de visualización de itinerarios - UNED

**¿Sirve Wikipedia para algo más que mostrar información?**

- Los recursos en abierto no son sólo una fuente de información de conceptos.
- Los datos enlazados permiten relacionar, categorizar y descubrir información en los textos.

Este trabajo presenta la primera versión de un prototipo de generación de itinerarios que parte con **581 itinerarios obtenidos automáticamente** a partir de [Wikipedia](#), [Wikidata](#) y el [Museo del Prado](#). Estos itinerarios, a su vez, están formados por **4823 obras**, **1504 personajes**, **14166 referencias** y **25569 eventos**. De los eventos encontrados, 2799 están relacionados con obras, 2090 con personajes y 221 con referencias.

Tarea a realizar para familiarizarse:

1. Busque información sobre varios itinerarios. Por ejemplo:
  - Busque y visualice el itinerario "Cuadros relacionados con el personaje Diego Velázquez".
  - Busque y visualice el itinerario "Cuadros relacionados con Adoración de los Reyes Magos".
2. Rellene la encuesta.

Para cualquier duda relacionada con el sistema, puede acceder al [Manual de usuario](#)

**Búsqueda de itinerarios**

**Realizar encuesta**

**Consultar las preguntas frecuentes**

 **Museología e integración social: la difusión del patrimonio artístico y cultural del Museo del Prado a colectivos con especial accesibilidad (invidentes, sordos y reclusos)**  
[Web oficial Musaces](#)

Ilustración 72 Pantalla de inicio de MPOC

En caso de tratarse de un usuario autor, se mostrará información adicional, correspondiente con las funcionalidades especificadas en el capítulo anterior, como muestra la Ilustración 4:



MPOC

Sistema web de visualización de itinerarios - UNED

#### ¿Sirve Wikipedia para algo más que mostrar información?

- Los recursos en abierto no son sólo una fuente de información de conceptos.
- Los datos enlazados permiten relacionar, categorizar y descubrir información en los textos.

Este trabajo presenta la primera versión de un prototipo de generación de itinerarios que parte con **581 itinerarios obtenidos automáticamente** a partir de [Wikipedia](#), [Wikidata](#) y el [Museo del Prado](#). Estos itinerarios, a su vez, están formados por **4823 obras**, **1504 personajes**, **14166 referencias** y **25569 eventos**. De los eventos encontrados, 2799 están relacionados con obras, 2090 con personajes y 221 con referencias.

Tarea a realizar para familiarizarse:

1. Busque información sobre varios itinerarios. Por ejemplo:
  - Busque y visualice el itinerario "Cuadros relacionados con el personaje Diego Velázquez".
  - Busque y visualice el itinerario "Cuadros relacionados con Adoración de los Reyes Magos".
2. Rellene la encuesta.

Para cualquier duda relacionada con el sistema, puede acceder al [Manual de usuario](#)

 Búsqueda de itinerarios

 Realizar encuesta

Consultar las preguntas frecuentes

Edición de itinerarios



Gestión de referencias



Gestión de personajes



Gestión de obras



Gestión de itinerarios

Ilustración 73 Pantalla de inicio de MPOC para autores

Además, el usuario administrador podrá acceder a más funcionalidades, como muestra la Ilustración 5.

¿Sirve Wikipedia para algo más que mostrar información?

- Los recursos en abierto no son sólo una fuente de información de conceptos.
- Los datos enlazados permiten relacionar, categorizar y descubrir información en los textos.

Este trabajo presenta la primera versión de un prototipo de generación de itinerarios que parte con **581 itinerarios obtenidos automáticamente** a partir de [Wikipedia](#), [Wikidata](#) y el [Museo del Prado](#). Estos itinerarios, a su vez, están formados por **4823 obras**, **1504 personajes**, **14166 referencias** y **25569 eventos**. De los eventos encontrados, 2799 están relacionados con obras, 2090 con personajes y 221 con referencias.

Tarea a realizar para familiarizarse:

1. Busque información sobre varios itinerarios. Por ejemplo:
  - Busque y visualice el itinerario "Cuadros relacionados con el personaje Diego Velázquez".
  - Busque y visualice el itinerario "Cuadros relacionados con Adoración de los Reyes Magos".
2. Rellene la encuesta.

Para cualquier duda relacionada con el sistema, puede acceder al [Manual de usuario](#)



Ilustración 74 Pantalla de inicio de MPOC para administrador

Los enlaces o botones que se encuentran en las ilustraciones anteriores se corresponden con las funcionalidades del sistema, como muestra la tabla 2:

	Botones o enlaces de la página
<b>Creación y modificación de itinerarios</b>	Gestión de itinerarios
<b>Edición de elementos narrativos</b>	Gestión de obras, gestión de personajes, gestión de referencias
<b>Visualización de itinerarios existentes</b>	Búsqueda de itinerarios
<b>Visualización de logs</b>	Gestión de logs
<b>Administración de los usuarios</b>	-
<b>Administración de la encuesta de evaluación</b>	Gestión de encuestas

Tabla 14 Relación entre funcionalidades y botones mostrados en la página de inicio

 La gestión de usuarios no se realiza desde la página web en esta versión de la web.

Para acceder directamente a las funcionalidades, se puede realizar a través de las URL mostradas en la tabla 3:

	Botones o enlaces de la página
--	--------------------------------

<b>Creación y modificación de itinerarios</b>	<a href="http://aluned.laraclares.com/segmentos">http://aluned.laraclares.com/segmentos</a>
<b>Edición de elementos narrativos</b>	<a href="http://aluned.laraclares.com/artworks">http://aluned.laraclares.com/artworks</a> <a href="http://aluned.laraclares.com/referencia">http://aluned.laraclares.com/referencia</a> <a href="http://aluned.laraclares.com/personaje">http://aluned.laraclares.com/personaje</a>
<b>Visualización de itinerarios existentes</b>	<a href="http://aluned.laraclares.com/itinerario">http://aluned.laraclares.com/itinerario</a>
<b>Visualización de logs</b>	<a href="http://aluned.laraclares.com/logs">http://aluned.laraclares.com/logs</a>
<b>Administración de los usuarios</b>	-
<b>Administración de la encuesta de evaluación</b>	<a href="http://aluned.laraclares.com/encuestas">http://aluned.laraclares.com/encuestas</a>

Tabla 15 Relación entre funcionalidades y enlaces directos a la página web

Además de acceder a las distintas funcionalidades a partir de la página de inicio, es posible hacerlo a partir del menú existente en el resto de páginas de la web.

En caso de ser usuario visitante, el menú se presenta como muestra la Ilustración 6:

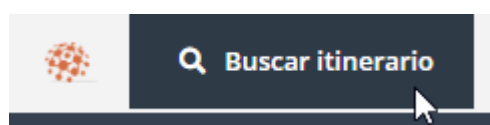


Ilustración 75 Menú superior para usuarios visitante

En caso de ser autor, se presenta como muestra la Ilustración 7:

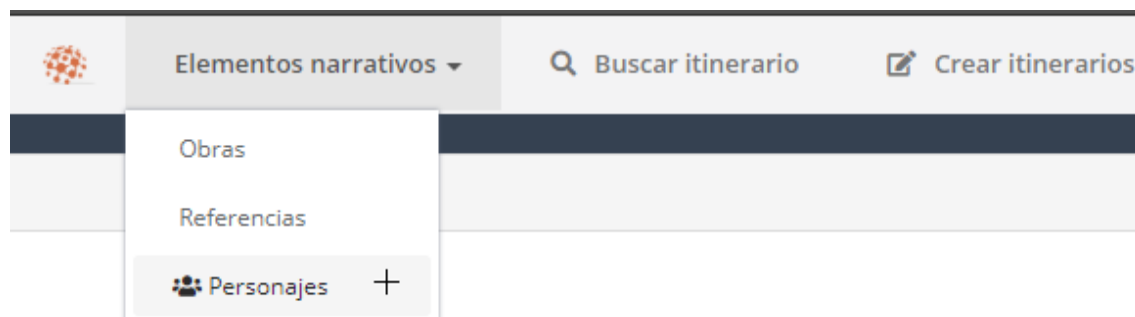


Ilustración 76 Menú superior para usuarios autores

En caso de ser administrador, se muestran los enlaces adicionales, como muestra la Ilustración 8:



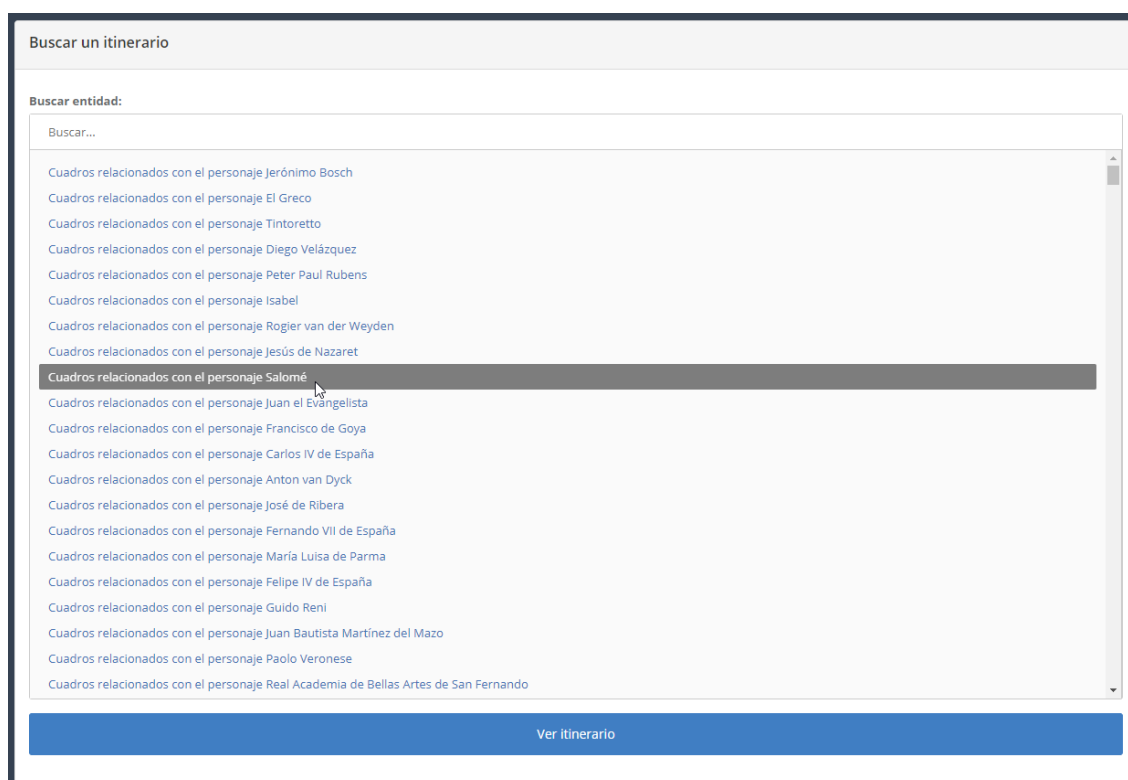
Ilustración 77 Menú superior para usuarios administradores

## 4.2 MANUAL DE USO DE LA WEB

En este punto se mostrará el tutorial relativo a cada funcionalidad.

### 4.2.1 Visualización de itinerarios existentes

La página mostrada al entrar en la búsqueda o visualización de itinerarios es la mostrada en la Ilustración 9:



*Ilustración 78 Visualización o búsqueda de un itinerario existente*


Como puede verse, es posible realizar una búsqueda por nombre. Una vez seleccionado el itinerario que se quiere visualizar, se hace clic en el botón "Ver itinerario", como muestra la Ilustración 10.

Cuadros relacionados con el personaje Salomé

[Editar itinerario](#)

**Retablo de San Juan Bautista y Santa Catalina**

[Saber más sobre la obra](#)



Retablo de San Juan Bautista y Santa Catalina

El conjunto consta de cinco tablas: la central con los santos titulares, San Juan y Santa Catalina, de pie con seis figuras de guarnición, cuatro de Apóstoles, reconociéndose sólo a San Bartolomé (segunda a la derecha) y dos de los santos Lorenzo y Antón. En las tablas laterales se representa, a la derecha, el martirio de Santa Catalina, donde se ve a los ángeles rompiendo las ruedas de cuchillas y la degollación de la santa. En la parte de la izquierda se puede observar a Salomé ante su padrastro Herodes y la degollación de San Juan Bautista.

[Anterior](#)   [Siguiente](#)

*Ilustración 79 Ejemplo de segmento de itinerario<sup>59</sup>*

Como puede verse, es posible hacer clic en los distintos elementos narrativos en el texto, lo que llevará a la visualización de los mismos. Además, los botones de navegación “anterior” y “siguiente” conducen a los segmentos anterior y posterior de dicho itinerario respectivamente.

- ✚ En caso de seleccionar varios itinerarios, se mostrará únicamente el itinerario correspondiente al último valor seleccionado.

Los elementos narrativos que pueden encontrarse son personajes, referencias y eventos. En el caso de personajes, se muestra la información de metadatos existente y sus descripciones, tanto del Museo del Prado como de Wikipedia. En caso de referencias o eventos se muestran las descripciones de ambas fuentes. A continuación, se muestran dos ejemplos: en el primer caso, correspondiente al personaje de Miguel Ángel (Ilustración 11), y en segundo lugar correspondiente al Manierismo (Ilustración 12).

<sup>59</sup> <http://aluned.laraclares.com/itinerario/2?itinerario=10>



Miguel Ángel



- Nacimiento: 1475-03-06, Caprese Michelangelo
- Fallecimiento: 1564-02-18, Roma

Editar personaje

### Descripción

En 1488 entró como aprendiz en el taller de Domenico Ghirlandaio (1448/49-1494) en Florencia. Al año siguiente ingresó en la academia de jóvenes escultores en el jardín de los Médicis, donde probablemente recibió las enseñanzas de Bertoldo di Giovanni (ca. 1430/40-1491). Sin embargo, Miguel Ángel fue siempre autodidacta y gozó del mecenazgo de Lorenzo el Magnífico hasta el fallecimiento de éste en 1492. A continuación regresó al hogar paterno y, tras otra breve estancia en la casa de los Médicis, salió de Florencia antes de que éstos fueran expulsados de la ciudad en 1494; residió algún tiempo en Venecia y en Bolonia y regresó a Florencia en 1495.

### Descripción (Wikipedia)

#### Primera estancia en roma

Su salida hacia Roma tuvo lugar el 20 de junio de 1496. La primera obra que realizó fue un Baco con un sátiro de medida natural, con gran parecido a una estatua clásica, y encargada

Ilustración 80 Visualización del personaje Miguel Ángel<sup>60</sup>

manierismo

### Descripción

#### Centroeuropa escandinavia y el baltico

La corte de Rodolfo II en Praga (desde 1583) se convirtió en el principal centro artístico al norte de los Alpes (Giuseppe Arcimboldo, Bartholomäus Spranger, intentó traer a Giambologna, cuya obra admiraba, pero al no conseguirlo, contrató a Adriaen de Vries). Su Wunderkammer ("cámara" o "cuarto de maravillas"), el más famoso de Europa, reunía objetos artísticos y curiosidades procedentes de todo el mundo. Para etiquetar a algunos pintores de este ámbito, como Hans von Aachen o Joseph Heintz, se utilizan la expresiones "arte rudolfino"<sup>[84]</sup> y "Manierismo rudolfino" o "Manierismo de Praga"<sup>[85]</sup>El centro artístico del Báltico en la época manierista fue la ciudad de Danzig, con un programa constructivo municipal (el Langer Markt<sup>[92]</sup> "mercado largo", el Artushof<sup>[93]</sup> y el edificio del ayuntamiento<sup>[94]</sup>). Otro centro importante fue Lemberg (plaza del mercado<sup>[95]</sup> y capilla Boimów<sup>[96]</sup>).<sup>[97]</sup>En el reino de Dinamarca se realizó un ambicioso programa artístico manierista con motivo de la remodelación del Castillo de Kronborg.En el reino de Suecia algunos castillos, como el de Vadstena<sup>[98]</sup> o el de Kalmar<sup>[99]</sup> fueron remozados en la segunda mitad del siglo XVI con criterios renacentistas o manieristas.

#### Artes plásticas

En el arte italiano se considera que el inicio del Manierismo parte de un violento acontecimiento histórico: el Saco de Roma de 1527, que puso fin a la centralidad que la ciudad papal tuvo en las primeras décadas del siglo, dispersando a los artistas que hasta entonces habían trabajado en ella por otros centros artísticos (como la escuela veneciana, la escuela parmesana o la escuela veronesa); mientras que algunos de los más importantes que había reunido Julio II en su pontificado (de 1503 a 1513) habían ya muerto (Donato Bramante en 1514, Rafael Sanzio en 1520) y uno de los supervivientes, el longevo Miguel Ángel, evolucionó profundamente en su arte en las décadas que transcurrieron hasta su muerte (1564). En el asedio que mantuvo a Clemente VII dentro de la fortaleza de Castel Sant'Angelo estuvo luchando el artista Benvenuto Cellini, uno de los máximos exponentes del nuevo estilo.<sup>[24]</sup> También en 1527 el poeta Pietro Aretino y los Sansovino (el arquitecto Jacopo Sansovino y su hijo Francesco Sansovino) se establecen en Venecia, donde contactan con Tiziano, a quien introducen en las novedades del gusto de la Italia central, el "manierismo brutal" de Giulio Romano y el impactante cartón de Rafael (Conversión de San Pablo, no conservado), que conoce por un cartón propiedad del cardenal Grimani<sup>[100]</sup> a época coincide con la extensión de la influencia del Renacimiento italiano por toda Europa occidental, de modo que el

Ilustración 81 Visualización de la referencia Manierismo<sup>61</sup>

## 4.2.2 Creación y modificación de itinerarios (sólo autores y administradores)

La creación de itinerarios se realiza a partir de la siguiente URL:

<http://aluned.laraclares.com/segmentos>

Al cargar la página, se ve el cuadro mostrado en la Ilustración 13:

<sup>60</sup> <http://aluned.laraclares.com/personaje/973>

<sup>61</sup> <http://aluned.laraclares.com/referencia/1963>

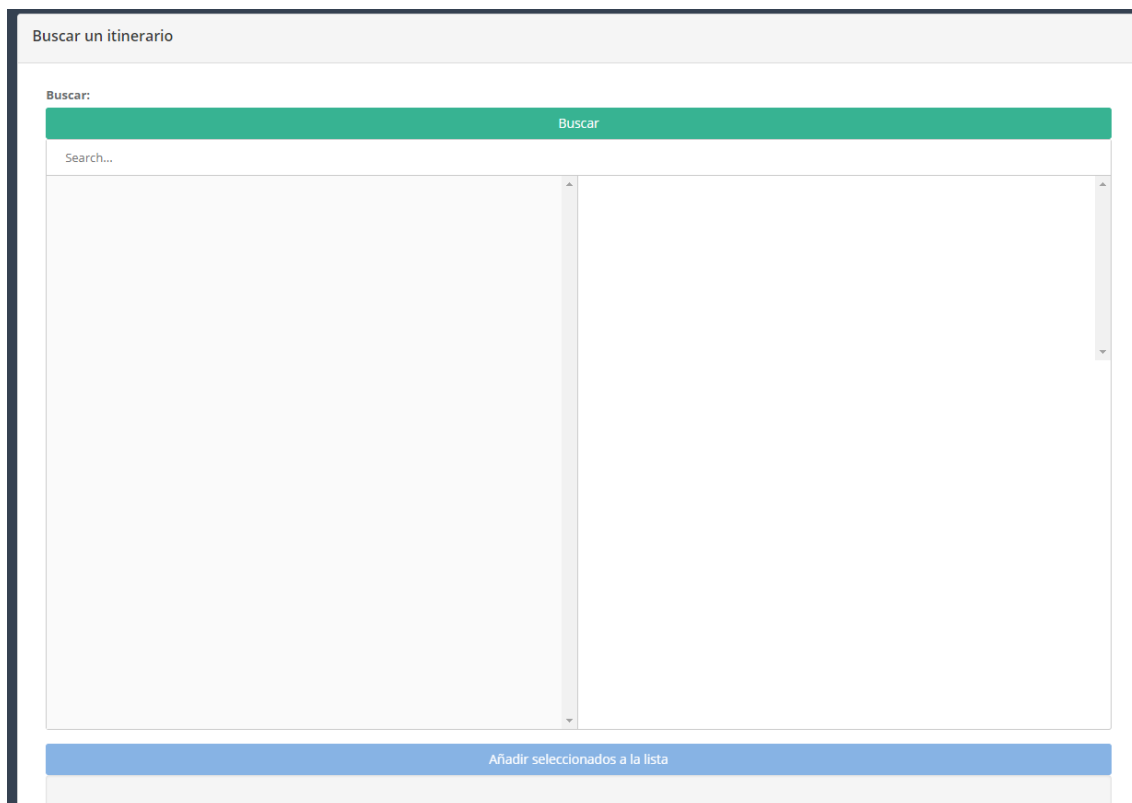


Ilustración 82 Búsqueda de segmentos

El sistema permite realizar la búsqueda por palabras, de manera que no es necesario introducir la palabra exacta (permite omitir tildes, signos de puntuación y no distingue mayúsculas y minúsculas).

Una vez realizada la búsqueda y seleccionado cada concepto que se quiere buscar, el siguiente paso es hacer clic en “Añadir seleccionados a la lista”, como muestra la Ilustración 14.

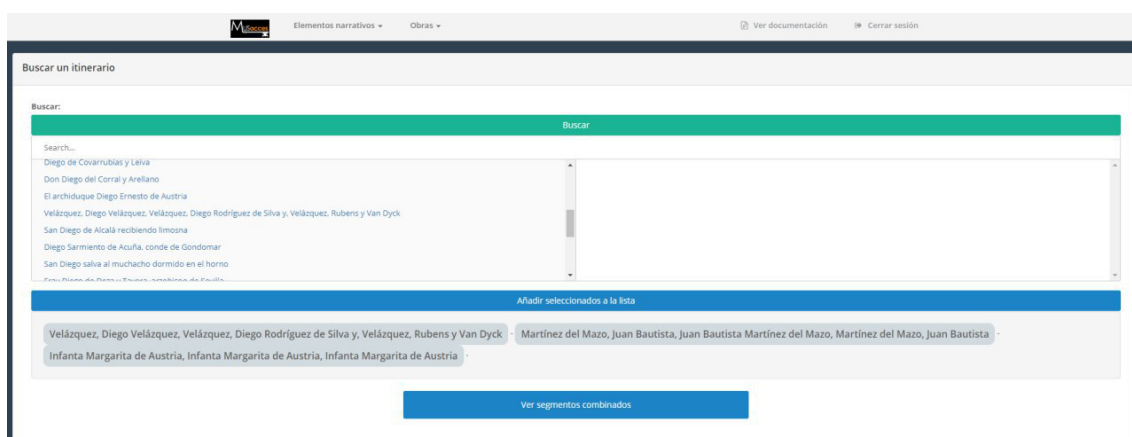


Ilustración 83 Búsqueda y selección de segmentos

Una vez seleccionados los temas o elementos narrativos con los que se quiere trabajar, se hace clic en “Ver segmentos combinados”.

La pantalla de la Ilustración 15 muestra un formulario donde se pueden seleccionar, aportando un valor numérico de orden, los segmentos que pueden conformar el itinerario:

**Combinación de segmentos**

Segmentos que pertenecen a todos los elementos narrativos seleccionados:

Id	Obra	Texto	Editar	Orden
35_wp_0	La infanta doña Margarita de Austria	La infanta doña Margarita de Austria es un retrato de dicha infanta que se conserva en el Museo del	<a href="#">Editar</a>	0

Segmentos que pertenecen a Velázquez, Diego Velázquez, Velázquez, Diego Rodríguez de Silva y, Velázquez, Rubens y Van Dyck

Id	Obra	Texto	Editar	Orden
3_wp_1	Los borrachos	El cuadro lo pintó algunos años después de su llegada a Madrid procedente de Sevilla, poco antes	<a href="#">Editar</a>	0
3_wp_0	Los borrachos	El triunfo de Baco es una pintura del español Velázquez, conservada en el Museo del Prado y creada	<a href="#">Editar</a>	0
4_wp_0	La fragua de Vulcano	La fragua de Vulcano es una obra de Diego Velázquez pintada en Roma en 1630, según informa Antonio	<a href="#">Editar</a>	0
5_wp_0	La reina Isabel de Borbón, a caballo	El retrato de La reina Isabel de Francia a caballo fue pintado por Velázquez entre 1628-36 y se con	<a href="#">Editar</a>	0
7_wp_0	Bufón con libros	El bufón don Diego de Acedo, «el Primo», es un retrato pintado por Velázquez y conservado en el	<a href="#">Editar</a>	0
10_wp_1	Carlos IV	Fue hecho con motivo de la proclamación de Carlos IV, en 1789. El nuevo rey gustaba mucho de las ob	<a href="#">Editar</a>	0

*Ilustración 84 Ejemplo de segmentos combinados*

Como puede verse, en primer lugar, realiza una búsqueda de segmentos que contengan a los elementos seleccionados, y después muestra un desglose de los segmentos correspondientes a cada elemento dado.

En la Ilustración 16 se muestra un ejemplo de selección de segmentos, aportando un orden en los seleccionados y manteniendo el valor “0” en los que no resultan seleccionados.

Id	Obra	Texto	Editar	Orden
35_wp_0	La infanta doña Margarita de Austria	La infanta doña Margarita de Austria es un retrato de dicha infanta que se conserva en el Museo del	<a href="#">Editar</a>	1

Segmentos que pertenecen a Velázquez, Diego Velázquez, Velázquez, Diego Rodríguez de Silva y, Velázquez, Rubens y Van Dyck

Id	Obra	Texto	Editar	Orden
3_wp_1	Los borrachos	El cuadro lo pintó algunos años después de su llegada a Madrid procedente de Sevilla, poco antes	<a href="#">Editar</a>	4
3_wp_0	Los borrachos	El triunfo de Baco es una pintura del español Velázquez, conservada en el Museo del Prado y creada	<a href="#">Editar</a>	0
4_wp_0	La fragua de Vulcano	La fragua de Vulcano es una obra de Diego Velázquez pintada en Roma en 1630, según informa Antonio	<a href="#">Editar</a>	2
5_wp_0	La reina Isabel de Borbón, a caballo	El retrato de La reina Isabel de Francia a caballo fue pintado por Velázquez entre 1628-36 y se con	<a href="#">Editar</a>	0
7_wp_0	Bufón con libros	El bufón don Diego de Acedo, «el Primo», es un retrato pintado por Velázquez y conservado en el	<a href="#">Editar</a>	3
10_wp_1	Carlos IV	Fue hecho con motivo de la proclamación de Carlos IV, en 1789. El nuevo rey gustaba mucho de las ob	<a href="#">Editar</a>	<input type="text" value="0"/>

*Ilustración 85 Ejemplo de selección de segmentos*

Una vez seleccionados los segmentos, se inserta un título y se hace clic en “Crear itinerario con los segmentos seleccionados”, como muestra la Ilustración 17.

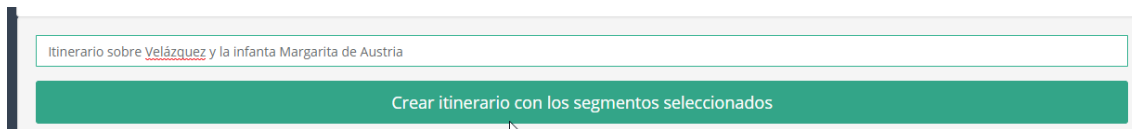


Ilustración 86 Asignación de título y creación de itinerario

Una vez creado el itinerario, se creará un botón para acceder a la visualización de este, y será posible encontrarlo en la búsqueda de itinerarios.

En la visualización del itinerario aparece un botón para editarlo en la esquina superior derecha. En la ventana de edición de itinerario, aparece un área de edición de texto por cada segmento correspondiente al itinerario, por lo que los segmentos se editarán por separado, como muestra la Ilustración 18.

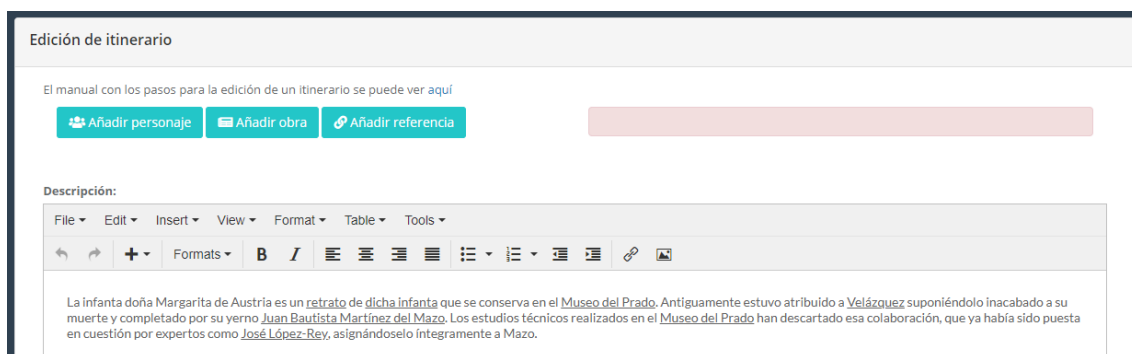


Ilustración 87 Imagen correspondiente con la ventana de edición de itinerario

Para poder añadir personajes, obras o referencias a las descripciones, se realiza siguiendo el mismo proceso en cada caso, una vez hecho clic en el botón correspondiente. Se muestra como ejemplo añadir un personaje al texto del primer segmento.

Al hacer clic en “Añadir personaje”, aparece la siguiente ventana mostrada en la Ilustración 19.

Lista de referencias

Show 10 entries

Id	Título	Descripción	Descripción (Wikipedia)	Ver	Selec
201	Tintoretto, Jacopo Robusti	Si	Si	Q Ver	<input type="radio"/>
202	Maestro de San Esteve de Andorra	No	No	Q Ver	<input type="radio"/>
203	Velázquez, Diego Rodríguez de Silva y	Si	No	Q Ver	<input type="radio"/>
204	Rubens, Pedro Pablo	Si	Si	Q Ver	<input type="radio"/>
205	Vos, Cornelis de	No	No	Q Ver	<input type="radio"/>
206	Bayeu y Subías, Francisco	Si	No	Q Ver	<input type="radio"/>
207	Beuckelaer, Joachim	Si	No	Q Ver	<input type="radio"/>
208	Isabel de Borbón (reina de España)	No	Si	Q Ver	<input type="radio"/>
209	El Bosco	Si	Si	Q Ver	<input type="radio"/>
210	René de Châlon, 1538-44	No	No	Q Ver	<input type="radio"/>

Showing 1 to 10 of 1,504 entries

Previous 1 2 3

Ilustración 88 Selección de personaje

Para seleccionar el personaje deseado, se hace clic en el botón de la derecha de la fila correspondiente, y aparecerá una ventana emergente como muestra la Ilustración 20.

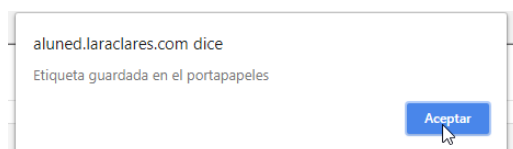


Ilustración 89 Ventana emergente de selección de elemento narrativo

Una vez creado el elemento, se almacena en el portapapeles, y además se copia la etiqueta en la página, para permitir pegarla dentro del texto, lo que creará la relación del cuadro con el personaje correspondiente.

Una vez editados los segmentos se pueden guardar los cambios haciendo clic en el botón “Guardar cambios”.

#### 4.2.3 Edición de elementos narrativos

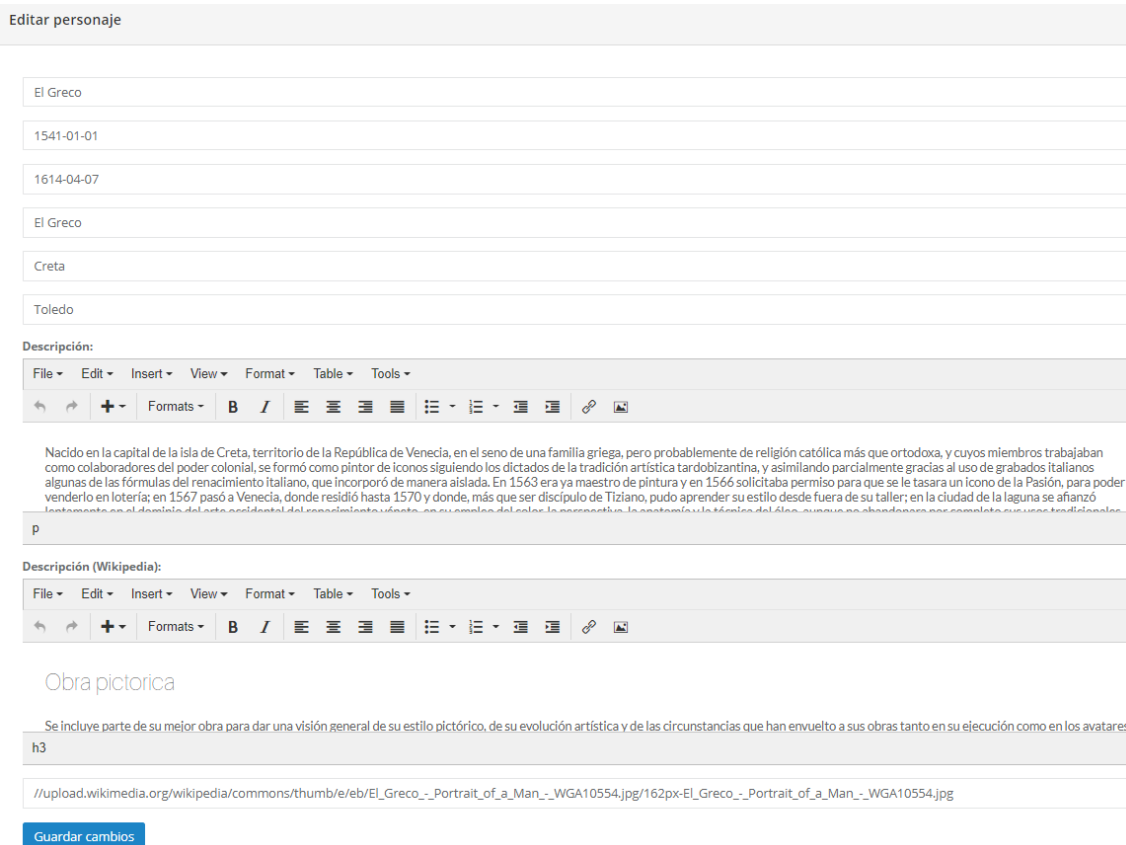
Todos los elementos narrativos se editan de forma similar, habiendo un botón de edición dentro de la visualización de cada uno, y permitiendo editar los campos necesarios. En este manual se muestra un ejemplo de edición de personaje.

En primer lugar, se accede a la búsqueda de personajes del menú superior y se selecciona el personaje a editar, como muestra la Ilustración 21.

225		Goya y Lucientes, Francisco de	Si	No	<a href="#">Q Ver</a>
226		Carlos IV de España	No	Si	<a href="#">Q Ver</a>
227		Tiziano, Vecellio di Gregorio	Si	No	<a href="#">Q Ver</a>
228		Van Dyck, Antonio	Si	No	<a href="#">Q Ver</a>
229		El Greco	Si	Si	<a href="#">Q Ver</a>
230		Bruegel el Viejo, Pieter	Si	No	<a href="#">Q Ver</a>

Ilustración 90 Selección de personaje para edición

Al hacer clic en “Editar personaje” aparece una ventana con los distintos campos que se pueden modificar, como muestra la Ilustración 22.



*Ilustración 91 Edición de personaje*

En el caso de la edición de elementos narrativos, los textos no están anotados, y por lo tanto, no contiene enlaces a otros elementos narrativos.

#### 4.2.4 Visualización de logs

La información referente a los logs puede consultarse en la siguiente URL:

<http://aluned.laraclares.com/logs>

#### 4.2.5 Administración de la encuesta de evaluación

El acceso a la encuesta se encuentra en la URL:

<http://aluned.laraclares.com/encuestas>

Por otro lado, los resultados de la encuesta pueden encontrarse en la siguiente URL:

<http://aluned.laraclares.com/encuestas/resultados>

## 5 REPORTE DE ERRORES Y RESOLUCIÓN DE PROBLEMAS

---

En caso de encontrar un error, basta con enviar un mail al administrador del sitio web con la siguiente información:

Asunto: Reporte de error en MPOC

Descripción: Fecha en la que se produjo el error y breve descripción.

## 6 GLOSARIO DE TÉRMINOS

---

- Itinerario. Es un conjunto de narraciones acerca de los elementos que lo conforman.
- LOD. *“Linked Open Data”*, o Datos Enlazados en Abierto. Surge como un nuevo paradigma donde la información se almacena en forma de datos estructurados, anotados y relacionando entre si los conceptos que lo conforman; permitiendo el acceso libre a la información y su interoperabilidad con otros formatos y sistemas a través del uso de distintos estándares de almacenamiento.
- Elementos narrativos. Son los elementos anotados en las descripciones que conforman los itinerarios. Son: personajes, eventos, referencias y obras.
- Personaje. Pueden ser reales o ficticios, y pueden estar relacionados directamente con la obra (como el autor de la misma), como indirectamente (si tiene alguna relación con la misma, como un Felipe IV en Las Meninas<sup>62</sup>).
- Evento. Son los acontecimientos que aportan el contexto histórico o político a los itinerarios
- Referencia. Son todos los conceptos con información semántica anotados en las descripciones de las obras (por ejemplo, “perro”, “bóveda”, “Manierismo”, “Barroco”, etc.)
- Obras. Son todas las pertenecientes al Museo del Prado de las que se ha obtenido la información.
- Segmento. Son las unidades textuales en las que se dividen los itinerarios. En este caso, cada segmento corresponde con un párrafo de las descripciones.
- Logs. Son los archivos que almacenan la información referente al uso por parte de los usuarios de la página web.

---

<sup>62</sup> [https://es.wikipedia.org/wiki/Las\\_meninas](https://es.wikipedia.org/wiki/Las_meninas)

## REFERENCIAS

---

- [1] “IEEE Standard for Software User Documentation,” *IEEE Std 1063-2001*, pp. 1–24, 2001.



## ANEXO II - PREGUNTAS CERRADAS

Menú

**Edad:**

**Estudios:**

**En general, ¿cuál es su nivel de conocimiento previo sobre las obras del museo?**

Casi nada  1  2  3  4  5  6  7 Soy experto

**Las siguientes preguntas evalúan la página web. Por favor, responda de la forma más espontánea posible, valorando entre 1 y 7 los siguientes parámetros.**

Desagradable	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Agradable
Difícil de entender	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Fácil de entender
Creativo	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	No creativo
Fácil de aprender	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Difícil de aprender
Con valor	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Sin valor
Aburrido	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Emocionante
No interesante	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Interesante
Impredecible	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Predecible
Rápido	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Lento
Original	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	No original
Obstrutivo	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Impulsor de apoyo
Bueno	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Malo
Complicado	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Fácil de usar
No atractivo	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Atractivo
Convencional	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Novedoso
Incómodo	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Cómodo
Seguro	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Inseguro
Activador	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Adormecedor
Cubre expectativas	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	No cubre expectativas

Ineficiente	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Eficiente
Claro	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Confuso
No pragmático	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Pragmático
Ordenado	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Sobrecargado
Atractivo	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Feo
Simpático	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Antipático
Conservador	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5	<input type="radio"/> 6	<input type="radio"/> 7	Innovador

## ANEXO III - RESPUESTAS A PREGUNTAS ABIERTAS

---

### ¿Qué es lo que más te ha gustado de esta página web y por qué?

- Lo que más me ha gustado es poder aprender sobre distintos temas sin necesidad de leer todas las obras.
- porque me da una nueva visión de los cuadros y puedo navegar por otros conceptos
- Es muy claro y sin mucha información que confunda, directo.
- Facilidad de uso, no todas las páginas son tan fáciles de entender y navegar
- La innovación en el uso de recorridos por temáticas. Se puede aprender mucho de un vistazo.
- Resulta fácil de leer, me gusta que cuente sobre la historia del cuadro y menos en referencia a "lo técnico".

Y me encantan las explicaciones de las palabras resaltadas.

- La página web es muy fácil de usar y muestra información muy interesante sobre diferentes conceptos y obras del museo. Considero que hay un gran trabajo detrás.
- Destacaría dos cosas:

1. Diseño sencillo e intuitivo.

2. Los recorridos de las obras permiten relacionar diferentes conceptos. Por ejemplo, si se buscan obras relacionadas con el concepto "Adoración de los Reyes Magos", el recorrido no se limita a enumerar obras tituladas de este modo, sino que además muestra otras obras de los pintores de esos cuadros. De este modo, se ofrece información adicional a los usuarios y nuevas maneras de explorar las obras del museo.

- Es una página web interesante, con contenido ilustrativo y lleno de referencias por las que puedes enlazar temas de un modo sencillo.

La explicación de las obras es entretenida, aunque en algunos casos puede ser un poco extensa, sobre todo pensando en personas que les gusta el arte, pero no son apasionados.

La navegación es simple e intuitiva.

El diseño es sencillo y no cansa leer el contenido, ni te "distraes" mirando otros menús, imágenes o secciones

- La información y su estructura, además de la capacidad de poder consultar más sobre un término concreto en la descripción de las obras.
- Es rápida, fácil de entender
- Me ha gustado la claridad para exponer y describir el cuadro y la facilidad para acceder a las rutas buscadas
- Muy interesante los itinerarios además de la información y cómo se navega entre ella, buena usabilidad
- Esta web resulta muy útil para la comprensión e historia de cada obra y artista, con una navegabilidad simple pero efectiva.
- Interfaz limpio, funcionamiento claro, rapidez.
- Todo
- La información aportada es completa y fácil de comprender, además de ser una forma original de conocer las obras del museo según su autor.
- La disponibilidad de la información y su presentación.
- Me gusta la exposición de los itinerarios. Fácil entendimiento de las obras
- Me ha parecido una web muy interesante. La considero una forma de acercar los museos a gente no experta. Destaca su sencillez y claridad.

- Lo que más me gusta es que al pasar el ratón por el título de un cuadro aparece la foto del cuadro
- Navegar por los cuadros de forma temática. Otra forma de pensar El Prado
- La idea del itinerario a elección personal
- Carga muy rápido, es intuitiva, fácil de usar.

Ayuda a aprender y hace que uno vaya de página en página sin parar de buscar más.

- Fácil de usar

No había usado nada igual

- Es fácil de usar, ordenada y te permite encontrar bastante información rápidamente.
- El hecho de que la página web una conceptos e información me parece algo novedoso y además es sencillo de usar.
- Una forma diferente de aprender sobre el museo
- fácil de usar y rápida
- Que hace una recopilación de información varias fuentes.
- Parece haber bastante información sobre obras del Museo del Prado y recopiladas por personajes, obras, autores, etc., que puede estar bien para realizar itinerarios por el museo. Además, se puede acceder fácilmente a una temática en concreto.
- sobre todo me ha gustado poder ver los cuadros ordenados por temas, es una nueva forma de verlos
  - - El interfaz me ha parecido sencillo e intuitivo, pues en pocos clicks pude realizar una búsqueda de itinerario y navegar viendo las obras del mismo. Además, pude encontrar descripciones de conceptos que no conocía de una manera muy familiar, a la que Wikipedia me tiene acostumbrado.
    - En mi caso, no estando familiarizado con el mundo de la pintura ni siendo gran forofo de los museos en general, me parece una manera entretenida de descubrir artistas y obras sin tener que visitar el museo. Sobre todo si puedo buscar itinerarios relacionados con algún objeto en concreto (como por ejemplo, el fuego) que lo hace más curioso.
    - Me ha parecido una buena sensación el poder elegir qué aprender/visualizar sin tener que ir "pasando" por todos los cuadros que no me transmiten tanto interés.
- Ver los cuadros y su explicación

## ¿Qué es lo que menos te ha gustado de esta página web y por qué?

- Lo que menos me ha gustado es no poder modificar las obras para crear mis propias guías y poder enseñárselas a más gente.
- nada
- teniendo en cuenta el objetivo del mismo, no ha habido nada fuera de lugar que me haya disgustado.
- los textos, algunos son repetitivos y no centrados en el tema de la búsqueda, es decir... las explicaciones del cuadro no varía según la búsqueda realizada
- Los espacios vacíos de algunos enlaces. Hay que terminarlos. Por ejemplo: Si escojo el recorrido de " El Greco", en el enlace sobre "mitras", no se encuentra información.
- Aun gustandome más todo el contexto y las fábulas faltan algunos datos de fechas.
- Los recorridos no muestran todas las obras del museo de un determinado pintor.
- Las búsquedas se limitan a una serie de conceptos preestablecidos. Dichos conceptos son los más relevantes con respecto a las obras del museo y además cubren una amplia gama de temáticas sobre las obras. No obstante, esta limitación no permite personalizar las búsquedas a los usuarios. En la siguiente pregunta propongo algunas ideas que pienso que pueden permitir mejorar este aspecto.
- El texto de la portada es de un gris demasiado claro y hay que dar mucho brillo en el móvil para leerlo bien. En el pc se ve bien.

Cuando pasas del cuadro 1 al 2, no se puede volver al 1. Cuando del 2 al 3, y del 3 al 2 sí se puede.

El texto, en algunas secciones, para pc es muy grande (descripción de un obra, grande. Descripción de un personaje, bien). Para móviles es perfecto.

- No se adapta de forma óptima a Chrome de Android y algunos elementos podrian ser mas intuitivos.
- No entiendo de diseño pero parece funcional
- Que sea leída. Sería interesante que hubiese otra opción para escuchar para poder admirar la obra a la vez
- Se podría mejorar el diseño
- El diseño y la apariencia de la web es un poco monótona y de poca variedad. Estaría bien un aspecto más atractivo.
- Que las opciones de itinerario estén limitadas, pero imagino que eso es por ser una primera versión.
- Me ha gustado todo
- El color verde UNED como fondo para las imágenes (aunque no sea exactamente la tonalidad uned, es muy cantoso) ; )
- Los colores de la web.
- Quizas los colores de la web.
- Es un listado de enlaces que recuerda a los buscadores antiguos o alternativos a google
- Las descripciones no adaptadas a la búsqueda, busques lo que busques a un mismo cuadro una misma descripción
- Faltan itinerarios, por ejemplo, por épocas, escuelas pictóricas...
- Hay un typo: en la página principal: protitipo --> prototipo

En la descripción de los cuadros, el texto no está justificado a ambos lados.

Algún enlace no funciona, pero está muy bien para ser un primer prototipo.

- Me gustaría que fuera más interactivo
- Tal vez podría ser más atractiva e incluir categorías más generales en las que se incluyan otras más específicas
- En algunas ocasiones hay demasiadas palabras con enlace a información que no son relevantes en cuanto al cuadro sobre el cual se informa, y en bastantes ocasiones estas palabras destacadas no contienen información al pinchar sobre ellas. Creo que es interesante cuando se enlazan fechas, nombres o conceptos relevantes pero es confuso cuando hay muchas palabras "comunes" con enlace a su significado también. Un ejemplo: "Carlos III" aparece en una de las

descripciones pero solo "Carlos" aparece resaltado (cuando pinchas el enlace no te ofrece información). En otra ocasión, sin embargo, "Carlos III" aparece resaltado en su totalidad y en esta ocasión si te lleva a información sobre el personaje.

- No parece muy optimizada para móvil.
- no dice el tiempo que se tardaría ni el recorrido en metros, esto es importante para los que usamos muletas o tenemos problemas de movilidad

me gustaria con mas colores

- - He encontrado un error en la web: al ir a pulsar sobre Las Meninas (dentro de la ventana del cuadro de La familia de Carlos IV), salta un error:  
ErrorException in ItinerarioController.php line 126:  
Undefined index: itinerario
- En algún cuadro la información se superpone y no se lee bien. Por ejemplo, en el de Cadmo y Minerva.
- al buscar los cuadros, se pueden escoger varios temas y luego sólo se selecciona uno
- 1. Cuando hay mucho texto descriptivo de una obra, puede ser tedioso tener que subir y bajar scroll para poder contrastar la descripción con el cuadro. No me ha parecido demasiado tedioso, pero me dio la sensación de que podría seguir mejor la descripción con audio, sin despegar los ojos de la imagen.
  2. Algunos enlaces a conceptos no incluían la descripción del contexto, y en algunos casos, el texto se presentaba con, para mi gusto, poca separación de párrafos dentro de una misma sección.
  3. Una vez que visualizo un cuadro y selecciono un concepto de su descripción. Me esperaba, no sólo una descripción del concepto, sino obtener un listado de obras que también estuvieran relacionadas con ese concepto, o incluso itinerarios relacionados con el concepto.

## ¿Cómo crees que se podría mejorar la página web? ¿Te gustaría añadir alguna funcionalidad en concreto?

- Me gustaría añadir la posibilidad de crear mis propias guías del museo.
- está bien
- Depende de la profundidad de la información que se quiera enseñar. Quizás clasificar cierta información importante o relevante al principio del texto relacionado. Por ejemplo, "otras obras que haya realizado el pintor".
- Añadiría un botón de búsqueda abajo junto a "siguiente ".

Para salir del manual del usuario y volver al menú principal añadiría un enlace directo

- Pensando en que se está hablando de una imagen debería poderse ampliar la imagen o, al utilizarla desde el móvil poder tener una pestaña en la que ver la imagen en miniatura con rapidez, sin ser así, cuando se comentan ciertas cosas que aparecen en la imagen tienes que subir para verlas y después volver a buscar en el texto. Si bajase una pestaña emergente conforme baja la pantalla podrías ver la imagen para relacionar en cualquier momento sin perder el hilo de lo que lees.
- Ampliar información de las obras con diversas fuentes o enlazando con referencias externas.
- Considero que las siguientes funcionalidades pueden mejorar la página web:

1. Ampliación de búsquedas. Por ejemplo, se pueden realizar búsquedas relacionadas con varios conceptos. La versión actual de la página web solamente permite realizar recorridos relacionados con un único concepto. Si esto se amplía a varios conceptos, los recorridos pueden ser más especializados o incluso pueden permitir que el usuario vea si algunos conceptos están relacionados entre sí. En este mismo sentido, se pueden incluir operadores de búsqueda. Por ejemplo, itinerarios relacionados con el concepto "Francisco de Goya", pero NO relacionados con "Pinturas Oscuras".

2. Búsqueda libre por parte del usuario. Permitir que el usuario busque el concepto que quiera y mostrar los recorridos relacionados. De este modo, la página web puede mostrar información del museo más personalizada con los intereses del usuario.

3. Itinerarios reales. Conociendo la disposición real de las obras en el museo (distancia entre las mismas, etc.), los itinerarios pueden tener en cuenta este factor. De este modo, se ayudaría a los usuarios a planear una visita presencial más eficiente.

- He echado en falta un buscador de cuadros concreto.

Algunas secciones, como tablas, no se ven bien en el móvil.

Preferiría que no se abra, casi en cada click, una nueva pestaña.

- No se me ocurre nada en concreto. Mucho ánimo.
- Explicación de cada obra
- Un poco más interactiva
- Ubicación de la obra de arte en el museo o añadir un mapa con los diferentes itinerarios dentro del museo
- Iniciando los itinerarios a partir de épocas o artistas, además de hacerlo más visual utilizando imágenes de tales y haciendo que la estructura de la web sea más llamativa.
- Quizá un contador, barra de avance o similar que nos indique la cantidad de elementos del itinerario y por cual vamos.
- No, está muy bien
- Sería interesante saber de antemano la longitud del itinerario. Por lo demás está muy bien.
- En el buscador de obras se podría poner una vista preliminar de la obra, así sería más fácil de descubrir la obra que se busca (visualmente).
- La página debería de ofrecer la opción de cambiar de itinerario o de generar feedback durante el transcurso del mismo.
- Que fuera más atractiva

- - ¿Búsqueda por filtros? ¿Resultados en formato cuadrícula ? ¿Ver los recorridos sobre el plano del Prado?

- La búsqueda sobre "mujer" devuelve cosas que contienen "mujer" en el texto pero que no necesariamente están relacionados con el campo "mujer" (quizá pase también con otros itinerarios).

- Quizás sería interesante mostrar videos de los cuadros, si es posible.
- juegos sobre cuadros tipo trivial sobre lo que he buscado
- Si accedes a una categoría y luego quieres probar con otra, hay que desmarcarla para que te permita entrar, si no, vuelve a entrar en la que tenías seleccionada de primeras.
- Creo que la página está bien diseñada y es facil de usar y entender, pero como sugerencia revisaría los términos que aparecen resaltados en las descripciones.
- un mapa que marque por donde va el recorrido elegido

creo que mejoraría con algo más de color

- - El botón de ayuda debería estar siempre disponible.

- Al finalizar el itinerario debería aparecer un mensaje de fin y un botón que te lleve al inicio, sino, te quedas un poco esperando a ver qué pasa.

- Podría estar bien incluir filtros, para buscar por autores, cuadros, etc. y que no esté todo mezclado, ya que si no sabes qué quieres ver, o quieres curiosear un poco, te pierdes entre tantas opciones.

- Hacer un poco más intuitivo que se puede pulsar sobre el cuadro que se está viendo, no me he dado cuenta hasta después de un rato y ha sido al pasar el ratón por encima...

- me gustaría poder crear mis propios itinerarios
  - Sobre la presentación en la web, quizás añadiría un botón de audio al lado de las obras que reprodujera la descripción textual.
  - Sobre la navegación, incluiría enlaces a itinerarios u obras una vez que se ha seleccionado un concepto.
  - Como funcionalidad, añadiría mayor capacidad de búsqueda de itinerarios. Quizás en base a fechas, estilos de arte o etapas de historia de arte, pues me ha parecido que la curiosidad que transmiten los itinerarios son un gran arma de aprendizaje y ha hecho que despertara mi interés.



