DOCTORAL DISSERTATION

Contribution to the use of Semantic Web technologies to promote reflection and situational curiosity in Cultural Heritage experiences

Ahmed Tarek Abdelfattah Mohamed Dahroug
2019
Ahmed Tarek Abdelfattah Mohamed Dahroug

DOCTORAL DISSERTATION

Contribution to the use of Semantic Web technologies to promote reflection and situational curiosity in Cultural Heritage experiences

Supervised by:
Dr. Martín López Nores
Dr. José Juan Pazos Arias

2019
Acknowledgements

After thanking almighty god, who blessed me with the grace of loving science, being fond of knowledge and education, I have to thank and admit the grace of Dr. Martín López Nores for not just being a very devoted and supportive supervisor, but also a great mentor and a caring friend who had always encouraged me to do my best, always pushed me to learn more in my field and made a great change to my way of thinking and interpreting things. I would also love to give credit where credit is due, Dr. José Juan Pazos Arias the man that runs the show smoothly and for all his efforts to facilitate all the academic and managerial procedures.

Second, I would love to give special thanks to my professors in Egypt for their continuous and precious support and for paving the way and putting me on the right track.

My greatest love, my continuous respect and everlasting gratitude goes to my mother for being such a superior mother, who had devoted every second in her life for me and for our family to grow, proceed and succeed. Mum had always been and will always be my great motivator, the positive energy of my life and my best friend.

To my great father, a “thank you” word is not enough for all what you have done for me since my childhood, for all the effort, love and care you have surrounded me with, a thank you that you deserve is to dedicate this work and all my future successes to you.

All the appreciation goes to University of Vigo management staff for their continuing support and endless assistance.

Last but not least, thanks are due to Arab Academy for Science and Technology the place where I belong to for giving me the chance to proceed with my academic career and to earn my PhD Degree.
Resumen

Los últimos veinte años han experimentado un cambio importante en la creación de conocimiento y en los procesos de comunicación, debido al rápido desarrollo de las tecnologías de la información y las comunicaciones (TIC). La naturaleza innovadora y global que poseen estas tecnologías tuvo un impacto directo en casi todos los aspectos de la vida, así como en la mayoría de las disciplinas académicas. Se convirtieron en herramientas para producir conocimiento, objetos de estudio y un nuevo dominio de la ciencia en sí misma. La integración entre humanidades y tecnologías dio forma a lo que se conoce como “Humanidades Digitales”, que abarcan todo un conjunto de cuerpos de conocimiento, metodologías y procedimientos articulados en torno a recursos tales como bibliotecas digitales, sistema de información geográfica, redes sociales, mecanismos de interacción humano-computador, etc. Esta noción tiene sentido en tanto que las “Humanidades”, por definición, se refieren a disciplinas académicas que tratan el estudio de diferentes aspectos de la sociedad y la cultura humana, incluyendo lenguas antiguas y modernas, literatura, filosofía, historia, geografía humana, derecho, política, religión, sociología y arte. El apodo “digital” se puede agregar a todas esas áreas, de modo que se cubren los campos tradicionales con apoyo en TIC.

Las Humanidades Digitales utilizan diversas tecnologías para analizar y representar registros culturales y humanos. Así, los científicos ahora tienen la capacidad de encontrar respuestas para datos analógicos que solían ser extremadamente laboriosos o rayanas en lo imposible, tales como el análisis de conjuntos complejos de datos y la reconstrucción de patrimonio tangible o intangible. En el sentido contrario, al promover y facilitar la interacción directa entre científicos de distintas disciplinas, las Humanidades Digitales contribuyen también a la perspectiva para el desarrollo y la mejora de las TIC.

En esta tesis nos centramos en un área concreta de las Humanidades Digitales: el patrimonio cultural. El desarrollo del patrimonio, su papel y el lugar que ocupó en la sociedad europea llevó a la nominación de 2018 como el Año del Patrimonio Cultural Europeo. Los ciudadanos tuvieron la oportunidad de comprender mejor su cultura con sus diferentes dimensiones, al tiempo que compartían la responsabilidad por ella. La Unión Europea desempeñó asimismo un papel importante en la sensibilización acerca de cuestiones de restauración...
y conservación del patrimonio. También apoyó la investigación tecnológica y el progreso científico en soluciones tecnológicas, al igual que enfatizó la necesidad de abordar el tráfico y el robo de elementos del patrimonio, así como el papel de este en el turismo sostenible.

En particular, nos ha sido de relevancia la consideración del patrimonio cultural en diferentes programas de financiación de la Unión Europea, que facilitaron a los Estados miembros tomar medidas para reactivar su patrimonio nacional o local y salvar sus tradiciones y sus respectivas identidades. El Parlamento Europeo ha adoptado resoluciones para hacer hincapié en los peligros de la protección del patrimonio cultural en la Unión Europea y en el mundo. En consecuencia, se han puesto a disposición varios fondos para que el patrimonio cultural alcanzase la llamada visión “Europa 2020”. Así, el programa “Europa Creativa” asignó un presupuesto de 1.460 millones de euros para el periodo 2014-2020, destinado a cubrir principalmente proyectos culturales y de medios, en tanto que el programa Horizonte 2020 ha apoyado proyectos con más de 100 millones de euros en el periodo 2016-2019. Los proyectos financiados en estos programas abordaron en diferentes medidas cuestiones relativas a las siguientes tres dimensiones:

• Humanidades: creación de contenido, agregación y curación, evaluación de experiencias, metodologías y técnicas relacionadas, participación del usuario, etc.

• Tecnología: adaptación de contenido, personalización, herramientas para ser usadas en el sitio y fuera del sitio, herramientas de autoría, soporte mejorado para la digitalización, interacción hombre-computadora, realidad virtual y aumentada, etc.

• Difusión y explotación: planes de negocios, derechos de propiedad intelectual, interacción en laboratorios vivientes (“Living Labs”), estrategias comunes, publicaciones, publicidad, etc.

Uno de los proyectos financiados dentro del programa Horizonte 2020 ha proporcionado el marco para el desarrollo de esta tesis doctoral. Se trata del proyecto CrossCult (“Promoción de la reutilización del patrimonio cultural digital en cortes transversales de la historia de Europa sensibles al contexto”), cuyo objetivo fue utilizar las TIC para transformar las visitas a museos, sitios arqueológicos o lugares históricos al aire libre en experiencias más interactivas, basados en las tecnologías maduras de la Web Semántica, los notables avances en Inteligencia Artificial y las muchas iniciativas recientes que van dando forma a las Humanidades Digitales.

El proyecto CrossCult buscaba eliminar las fronteras culturales entre las naciones europeas a través de la interconexión de los recursos históricos digitales actuales y la creación de otros nuevos, ya fuera por profesionales o por personas sin formación específica en cuestiones de patrimonio. Para lograr el objetivo
prometido, el consorcio CrossCult —liderado por el Instituto Luxemburgués de Ciencia y Tecnología, la Universidad del Peloponeso y la Universidad de Vigo, con la participación adicional del University College de Londres, la Universidad de Malta, la Universidad de Padua, la empresa GVAM Guías Interactivas SL, el Centro Nacional para la Investigación Científica de Francia, la Universidad del Ática Occidental y la National Gallery de Londres— propuso la creación de una plataforma de servicios de utilidad para el diseño y el despliegue de experiencias interactivas de contenido histórico-cultural que fuese más allá de las lecturas tradicionales. Así, se construyó un paquete sofisticado de software y aplicaciones listas para brindar servicios no solo a los visitantes actuales y futuros de los museos, sitios arqueológicos o lugares históricos (a través de aplicaciones Android o iOS) sino también a diferentes tipos de partes interesadas, incluidos curadores de museos, expertos en Humanidades, analizadores de datos, desarrolladores de aplicaciones y administradores de sistemas (a través de interfaces web a medida). El funcionamiento de las interfaces web y móviles está respaldado por una trastienda que proporciona infraestructura e instrumentos para alojar los componentes principales, incluidas la lógica de los respectivos servicios y una base de conocimiento global en la que se acumulan e interrelacionan los metadatos disponibles sobre las distintas localizaciones y el patrimonio que atesoran.

Los conceptos principales de CrossCult se aplicaron en un conjunto de experimentos piloto en diferentes sitios europeos, así como en un conjunto de experimentos menores que se concentraron en tecnologías específicas. En este marco, esta tesis doctoral comparte la motivación general y los objetivos del proyecto: utilizar los desarrollos recientes en las TIC para mejorar la valoración del patrimonio cultural europeo. Más específicamente, hemos investigado el potencial de la narración interactiva y sensible al contexto en juegos serios, como un medio para aumentar el interés en el patrimonio cultural y la historia. Además, hemos investigado si las tecnologías de la Web Semántica se pueden utilizar para impulsar el descubrimiento de asociaciones o temas transversales entre múltiples elementos y ubicaciones del patrimonio, personajes o eventos históricos, etc. El descubrimiento de dichas conexiones se considera como la semilla de la creación de narrativas que despiertan curiosidad para impulsar los juegos y el aprendizaje. Finalmente, hemos desarrollado una nueva vía para la personalización de las experiencias de patrimonio cultural, al vincular fechas históricas relevantes o conmemoraciones anuales a los intereses de cualquier usuario particular de tecnología móvil, con el objetivo de servir el propósito de ofrecer narrativas más profundamente adaptadas.

Los objetivos de la tesis pueden, por lo tanto, ser enumerados de la siguiente manera:

- En primer lugar, contribuir a la aplicación de las tecnologías de la Web Semántica en el área del patrimonio cultural, de manera que promueva el intercambio de información y la integración de múltiples recursos y, de
este modo, facilite el descubrimiento automatizado de conexiones significativas.

- En segundo lugar, integrar los recursos abiertamente disponibles de la iniciativa Linked Data —así como otros de nueva creación— para permitir el razonamiento sobre elementos del patrimonio, lugares o sitios culturales y el contexto espacio-temporal de los usuarios de las TIC.

- En tercer lugar, implementar sistemas de descubrimiento y recomendación de asociaciones para identificar conexiones relevantes entre los perfiles de usuarios, las bases de conocimiento del patrimonio cultural y un compendio semántico de fechas relevantes.

- En cuarto lugar, diseñar e implementar herramientas para expertos en patrimonio cultural, que les sirvan para depurar las asociaciones descubiertas automáticamente y, a partir de ellas, elaborar narrativas interactivas sobre temas de su interés.

- Finalmente, reducir la brecha y ayudar a las personas a comprender mejor su identidad nacional y encontrar asociaciones comunes con otras culturas.

El trabajo se inició paralelamente con el arranque del proyecto CrossCult en 2016 y finalizó también con el mismo, contando con la colaboración directa —en aspectos tanto técnicos como de humanidades— de personal del University College de Londres, la Universidad de Malta, la Universidad de Padua y la Universidad del Peloponeso, principalmente.

Las principales contribuciones resultantes del trabajo se pueden resumir como sigue:

- En primer lugar, con una orientación más próxima a las Humanidades, hemos ayudado en la formalización de estrategias de narración de historias que maximizan los resultados deseados en términos de reflexión, reinterpretación, curiosidad, etc. Así, hemos enfrentado las preguntas de cómo facilitar la creación de narrativas para captar la gran cantidad de conexiones o asociaciones entre eventos y personajes históricos, elementos del patrimonio cultural y los intereses, las preferencias y los contextos de las audiencias objetivo de dichas narrativas. Vemos esto como una forma de superar las narrativas predominantes impuestas de manera implícita por el contenido estático de los libros de texto que se usan en colegios e institutos, por la disposición única de exposiciones en diferentes salas de un museo (según ciertos criterios de clasificación: cronología, autoría, estilo de pintura, ...), por la estructura de navegación de la web de una institución cultural, por el sesgo editorial de los medios de comunicación, o por la disposición fija e inmutable del patrimonio tangible que se extiende alrededor de una ciudad.
• En segundo lugar, ya como aportación tecnológica, hemos contribuido a la creación de herramientas para expertos para crear narrativas basadas en las asociaciones descubiertas por la lógica mencionada anteriormente, así como una aplicación móvil que ofrece esas narrativas como el conductor de un juego de preguntas. También hemos creado y evaluado una muestra de narrativas que buscan promover el conocimiento mutuo entre los mundos europeo y árabe, con un nuevo enfoque pedagógico enraizado en las similitudes y diferencias históricas y culturales existentes entre ubicaciones, personajes, eventos y cualquier otro elemento de identidades ya establecidas y narrativas dominantes. Nuestro diseño integra diferentes medios para instigar la curiosidad, la reflexión y la reinterpretación desde un enfoque de múltiples perspectivas, y una evaluación preliminar del mismo ha arrojado resultados positivos que alientan a seguir mejorando la aplicación, resolviendo problemas de usabilidad y refinando los contenidos. La experiencia permite vislumbrar estudios posteriores centrados en el uso de la aplicación en entornos educativos, contribuyendo a planes de estudios que permitan a los estudiantes desafiar los estereotipos y desarrollar una comprensión profunda de temas históricos controvertidos.

• En tercer lugar, hemos contribuido al diseño, implementación y evaluación de un prototipo de bases de conocimiento semántico adecuada para la captura e interrelación de contenidos digitales relacionados con el patrimonio cultural y los metadatos que caracterizan a elementos concretos del mismo (o a colecciones completas). La ontología de CrossCult se definió, pues, como una estructura conceptual genérica de nivel superior que capture conceptos y relaciones comunes en los experimentos piloto del proyecto, con flexibilidad suficiente para acoger un conjunto creciente de sedes, colecciones y obras de arte. Como tal, la ontología ofrece formalismos que se adaptan a arreglos conceptuales comunes y permiten el aumento, la vinculación, el razonamiento semántico y la recuperación de información, incluyendo (i) una estructura única y genérica de nivel superior que actúa como una capa semántica de conceptos y relaciones comunes entre pilotos, (ii) una definición ontológica sólida que permite un razonamiento semántico rico y eficientes, (iii) una estructura escalable que se puede ampliar formalmente para satisfacer necesidades conceptuales especializadas cuando sea necesario, (iv) un modelo de representación de conocimiento que se basa en estándares ampliamente utilizados de la Web Semántica —tales como CIDOC-CRM, Dublin Core, SKOS y FOAF— para promover la interoperabilidad y la vinculación con recursos de terceros, y (v) un modelo de representación de conocimiento que habilita la reutilización de recursos.

• En cuarto lugar, hemos creado la lógica necesaria para ejecutar sistemas de descubrimiento y recomendación de asociaciones semánticas como
parte del catálogo de servicios ofrecidos por la plataforma CrossCult. Nuestro enfoque se basa en los estándares generales para el modelado semántico de la información del patrimonio cultural, que suplementamos con una selección de recursos adicionales más nuevas clases y propiedades para capturar la semántica de fechas especiales relevantes, así como para modelar temas de reflexión y narrativas. En esa línea, se construyó un almanaque semántico a través de la minería de datos en diferentes sitios de la World Wide Web, resolviendo un problema (la gestión de conmemoraciones periódicas y su significado) que no había sido abordado con anterioridad.

- Por último, hemos creado un sistema de personalización basado en modelos de espacios vectoriales de palabras, que clasifica las asociaciones descubiertas según los intereses y eventos personales de cualquier usuario específico. Nuevamente, se trata de una aproximación sin precedentes en la literatura, que hemos evaluado satisfactoriamente en el contexto del Museo Arqueológico de Trípoli (Grecia) de manos de visitantes potenciales y expertos arqueólogos, pudiendo concluir con cierta confianza que las narraciones recomendadas son útiles para promover la reflexión y la retención de conocimiento sobre importantes temas culturales e históricos. Vemos este trabajo como un punto estratégico de acción en el proyecto CrossCult, en línea con su objetivo final de interconectar recursos digitales culturales, lugares físicos y opiniones de los ciudadanos de manera que promuevan la reflexión sobre la historia y el patrimonio cultural.

A nuestro entender, el trabajo que hemos realizado en CrossCult ha revelado algunas de las muchas posibilidades de la Inteligencia Artificial para apoyar una narración más efectiva en las Humanidades Digitales aplicadas al patrimonio cultural. La experiencia obtenida a través de los diferentes pilotos y experimentos ha proporcionado evidencia de que diversas técnicas de Inteligencia Artificial —desde la Web Semántica y la teoría de grafos hasta el reconocimiento de entidades y vectores de párrafos— pueden proporcionar ayudas significativas para descubrir asociaciones y conexiones interesantes entre recursos digitalizados, perfiles de usuario y cualesquiera características se consideren relevantes sobre el contexto actual de este. Esas conexiones se pueden emplear directamente para llamar la atención de los usuarios de la tecnología móvil a piezas de contenido seleccionadas —ya sea de los repositorios de un museo determinado o de fuentes disponibles públicamente como Wikipedia o Europeana— o, de manera alternativa, pueden servir de base para crear narraciones más específicas, con contenido preciso, confiable y específico por parte de expertos.

Desde un punto de vista crítico, cabe destacar que, incluso si las conexiones son fuertes, las aplicaciones de patrimonio cultural pueden ser tan buenas como el contenido que sirven y los mensajes duraderos que envían. Este hecho resalta la importancia de integrar la Inteligencia Artificial en las herramientas de autoría
adecuadas, con características para apoyar la colaboración, la integración de los archivos del lugar y recursos de datos diversos, la creación colaborativa de gráficos para la narración, la visualización y la gestión de asociaciones, etc. Se trata de aprovechar al máximo el tiempo y el esfuerzo de los expertos, sin dejar a un lado que es posible recurrir a la sabiduría de la multitud, siguiendo las experiencias pioneras de lo que se conoce como “cómputo por personas” o “crowd computing” en el área del patrimonio cultural. Consideramos que este es un nicho de investigación que aún debe desarrollarse en el futuro, con mucho espacio para la innovación y un notable potencial de mercado.

En definitiva, creemos que contar con narrativas nuevas y no convencionales que se centren en aspectos interculturales y transfronterizos puede hacer una diferencia en la valoración de la historia y el patrimonio cultural como una experiencia compartida, toda vez que hemos constatado repetidamente —en diferentes escenarios internacionales, con diferentes conceptos de aplicaciones y en diversos grupos objetivo— que las simetrías históricas y culturales pueden apoyar la curiosidad, la reflexión y la comprensión. Desde el lado puramente tecnológico, hemos podido comprobar que las tecnologías de Web Semántica se encuentran en un punto de desarrollo y adopción que, efectivamente, brinda soluciones en diferentes dominios de aplicación, tras muchos años en los que se cuestionaba si el coste de la generación de anotaciones formales se vería compensado por los beneficios derivados de su explotación.

La tesis consta de cinco capítulos, agrupados en tres partes diferentes:

• La Parte I ("Introducción") contiene dos capítulos. En el Capítulo 1 ("Introducción") examinamos el impacto de las TIC en el área amplia de Humanidades Digitales en general, y en el ámbito del Patrimonio Cultural en particular. También se ilustraron los objetivos de investigación y las contribuciones. El Capítulo 2 ("Antecedentes y trabajo relacionado"), por su parte, se concentra en analizar los aportes de otros investigadores en el pasado, brindando una visión general del estado del arte en los enfoques de narración y las tecnologías aplicadas al patrimonio cultural, además de las diferentes tecnologías que se utilizaron en la investigación, presentando así una amplia información sobre diferentes conceptos y conceptos técnicos como la semántica, las ontologías, el descubrimiento de asociaciones y la personalización.

• La Parte II ("Contribuciones de la tesis") desgrana los aportes resultantes del trabajo de doctorado, en dos capítulos. En el Capítulo 3 ("Contribuciones al modelado semántico de conocimiento sobre el patrimonio cultural y a las narrativas apoyadas en tecnología") explicamos las estrategias de narración desarrolladas en el marco del proyecto CrossCult, destacando nuestras contribuciones a los diferentes enfoques, herramientas y experimentos. Por su parte, en el Capítulo 4 ("Contribuciones al descubrimiento de asociaciones, a la sensibilidad al contexto y a la personalización") presentamos
los algoritmos de descubrimiento y personalización de la asociación que coinciden con la información contextual de las fechas relevantes. También se incluyen los resultados del experimento realizado para el Museo Arqueológico de Trípoli.

• Finalmente, la Parte III (“Conclusiones y trabajo futuro”) contiene el Capítulo 5 (de idéntico título), que proporciona un resumen de las conclusiones resultantes de este trabajo de doctorado, junto con algunas instrucciones para el trabajo futuro, ya sea una continuación directa de nuestras contribuciones o sugerencias para explorar nuevos potenciales.
Abstract

The last twenty years have experienced a significant change in the creation of knowledge and communication processes, due to the rapid development of information and communication technologies (ICT). The innovative and global nature of these technologies had a direct impact on almost all aspects of life, as well as in most academic disciplines. They became tools to produce knowledge, objects of study and a new domain of science itself. The integration between humanities and technologies gave shape to what is known as “Digital Humanities”, which encompass a whole set of bodies of knowledge, methodologies and procedures articulated around resources such as digital libraries, system of geographic information, social networks, human-computer interaction mechanisms, etc. This notion makes sense insofar as the “Humanities”, by definition, refers to academic disciplines that deal with the study of different aspects of society and human culture, including ancient and modern languages, literature, philosophy, history, human geography, law, politics, religion, sociology and art. The nickname “digital” can be added to all these areas, so that traditional fields are covered with ICT support.

The Digital Humanities use diverse technologies to analyze and represent cultural and human records. Thus, scientists now have the ability to find answers for analog data that used to be extremely laborious or borderline on the impossible, such as the analysis of complex data sets and the reconstruction of tangible or intangible heritage. In the opposite direction, by promoting and facilitating direct interaction between scientists from different disciplines, the Digital Humanities also contribute to the perspective for the development and improvement of ICT.

In this thesis we focus on a specific area of Digital Humanities: cultural heritage. The development of heritage, its role and the place it occupied in European society led to the nomination of 2018 as the Year of European Cultural Heritage. Citizens had the opportunity to better understand their culture with its different dimensions, while sharing responsibility for it. The European Union also played an important role in raising awareness about heritage restoration and conservation issues. He also supported technological research and scientific progress in technological solutions, as well as emphasizing the need to address the traffic and theft of heritage elements, as well as its role in sustainable tourism.
In particular, the consideration of cultural heritage in different funding programs of the European Union, which facilitated the Member States to take measures to reactivate their national or local heritage and save their traditions and their respective identities, has been of relevance to us. The European Parliament has adopted resolutions to emphasize the dangers of the protection of cultural heritage in the European Union and in the world. As a result, several funds have been made available for cultural heritage to reach the so-called "Europe 2020" vision. Thus, the program “Creative Europe” allocated a budget of 1,460 million euros for the period 2014-2020, intended to cover mainly cultural and media projects, while the Horizon 2020 program has supported projects with more than 100 million of euros in the 2016-2019 period. The projects financed in these programs addressed in different measures issues related to the following three dimensions:

• Humanities: creation of content, aggregation and healing, evaluation of experiences, methodologies and related techniques, user participation, etc.

• Technology: content adaptation, personalization, tools to be used on the site and off site, authoring tools, improved support for digitization, human-computer interaction, virtual and augmented reality, etc.

• Dissemination and exploitation: business plans, intellectual property rights, interaction in living laboratories (“Living Labs”), common strategies, publications, advertising, etc.

One of the projects funded under the Horizon 2020 program has provided the framework for the development of this doctoral thesis. This is the CrossCult project (“Promotion of the reuse of digital cultural heritage in transversal sections of the history of Europe sensitive to context”), whose objective was to use ICT to transform visits to museums, archaeological sites or open-air historical places in more interactive experiences, based on the mature technologies of the Semantic Web, the notable advances in Artificial Intelligence and the many recent initiatives that are shaping the Digital Humanities.

The CrossCult project sought to eliminate cultural borders between European nations through the interconnection of current digital historical resources and the creation of new ones, whether by professionals or by people without specific training in heritage issues. To achieve the promised goal, the CrossCult consortium —led by the Luxembourg Institute of Science and Technology, the University of Peloponnese and the University of Vigo, with the additional participation of University College London, the University of Malta, the University of Padua, the company GVAM Guías Interactivas SL, the National Center for Scientific Research of France, the University of Western Attica and the National Gallery of London— proposed the creation of a platform of useful services for the design and deployment of interactive experiences of historical-cultural content that went beyond traditional readings. Thus, a sophisticated package of
software and applications ready to provide services not only to current and future visitors of museums, archaeological sites or historical sites (through Android or iOS applications) but also to different types of interested parties was built, including curators of museums, experts in Humanities, data analyzers, application developers and system administrators (through custom web interfaces).

The operation of the web and mobile interfaces is supported by a back room that provides infrastructure and tools to host the main components, including the logic of the respective services and a global knowledge base in which the available metadata are accumulated and interrelated on the different locations and the heritage that they treasure.

The main concepts of CrossCult were applied in a set of pilot experiments in different European sites, as well as in a set of smaller experiments that focused on specific technologies. In this framework, this doctoral thesis shares the general motivation and objectives of the project: to use recent developments in ICT to improve the appreciation of European cultural heritage. More specifically, we have investigated the potential of interactive and context-sensitive storytelling in serious games, as a means to increase interest in cultural heritage and history. In addition, we have investigated whether Semantic Web technologies can be used to promote the discovery of associations or cross-cutting themes among multiple elements and heritage locations, historical characters or events, etc. The discovery of these connections is considered as the seed of the creation of narratives that arouse curiosity to promote games and learning. Finally, we have developed a new way to personalize cultural heritage experiences by linking relevant historical dates or annual commemorations to the interests of any particular user of mobile technology, in order to serve the purpose of offering more deeply adapted narratives.

The objectives of the thesis can, therefore, be listed as follows:

• First, to contribute to the application of Semantic Web technologies in the area of cultural heritage, in a way that promotes the exchange of information and the integration of multiple resources and, in this way, facilitates the automated discovery of connections significant.

• Secondly, to integrate the openly available resources of the Linked Data initiative —as well as new ones— to allow reasoning about elements of heritage, places or cultural sites and the spatiotemporal context of users of ICT.

• Third, implement discovery systems and recommendation of associations to identify relevant connections between user profiles, knowledge bases of cultural heritage and a semantic compendium of relevant dates.

• Fourth, design and implement tools for cultural heritage experts, to help them debug associations automatically discovered and, from them, develop interactive narratives on topics of interest.
• Finally, reduce the gap and help people to better understand their national identity and find common associations with other cultures.

The work started in parallel with the start of the CrossCult project in 2016 and also ended with the same, counting on the direct collaboration —in both technical aspects and humanities— of staff of the University College of London, the University of Malta, the University of Padua and the University of the Peloponnese, mainly.

The main contributions resulting from the work can be summarized as follows:

• First of all, with a closer orientation to the Humanities, we have helped in the formalization of storytelling strategies that maximize the desired results in terms of reflection, reinterpretation, curiosity, etc. Thus, we have faced the questions of how to facilitate the creation of narratives to capture the large number of connections or associations between historical events and characters, elements of cultural heritage and the interests, preferences and contexts of the target audiences of those narratives. We see this as a way to overcome the predominant narratives imposed implicitly by the static content of textbooks used in schools and institutes, by the unique arrangement of exhibitions in different rooms of a museum (according to certain classification criteria: chronology, authorship, painting style, ...), by the navigation structure of the website of a cultural institution, by the editorial bias of the media, or by the fixed and unchanging disposition of tangible heritage that extends around a city.

• Secondly, already as a technological contribution, we have contributed to the creation of tools for experts to create narratives based on the associations discovered by the logic mentioned above, as well as a mobile application that offers these narratives as the driver of a game of questions. We have also created and evaluated a sample of narratives that seek to promote mutual knowledge between the European and Arab worlds, with a new pedagogical approach rooted in the similarities and historical and cultural differences between locations, characters, events and any other element of identities and established and dominant narratives. Our design integrates different means to instigate curiosity, reflection and reinterpretation from a multi-perspective approach, and a preliminary evaluation of it has yielded positive results that encourage us to continue improving the application, solving usability problems and refining the contents. The experience allows to glimpse later studies focused on the use of the application in educational environments, contributing to curricula that allow students to challenge stereotypes and develop a deep understanding of controversial historical issues.
• Thirdly, we have contributed to the design, implementation and evaluation of a prototype of semantic knowledge bases suitable for the capture and interrelation of digital contents related to the cultural heritage and the metadata that characterize specific elements of the same (or complete collections). CrossCult’s ontology was defined, then, as a higher-level generic conceptual structure that captures common concepts and relationships in the project’s pilot experiments, with sufficient flexibility to accommodate a growing set of venues, collections and works of art. As such, ontology offers formalisms that adapt to common conceptual arrangements and allow for the enhancement, linking, semantic reasoning, and retrieval of information, including (i) a unique, generic top-level structure that acts as a semantic layer of common concepts and relationships among pilots, (ii) a solid ontological definition that allows rich and efficient semantic reasoning, (iii) a scalable structure that can be formally expanded to meet specialized conceptual needs when necessary, (iv) a representation model of knowledge that is based on widely used standards of the Semantic Web—such as CIDOC-CRM, Dublin Core, SKOS and FOAF—to promote interoperability and linkage with third-party resources, and (v) a model of representation of knowledge that enables the reuse of resources.

• Fourth, we have created the necessary logic to run systems for discovering and recommending semantic associations as part of the catalog of services offered by the CrossCult platform. Our approach is based on the general standards for the semantic modeling of cultural heritage information, which we supplement with a selection of additional resources, new classes and properties to capture the semantics of relevant special dates, as well as to model reflection and narrative topics. In this line, a semantic almanac was built through the mining of data in different sites of the World Wide Web, solving a problem (the management of periodic commemorations and their meaning) that had not been addressed before.

• Finally, we have created a personalization system based on models of vector spaces of words, which classifies the associations discovered according to the interests and personal events of any specific user. Again, this is an unprecedented approach in the literature, which we have satisfactorily evaluated in the context of the Archaeological Museum of Tripoli (Greece) at the hands of potential visitors and expert archaeologists, and can conclude with some confidence that the recommended narratives are useful to promote reflection and retention of knowledge about important cultural and historical issues. We see this work as a strategic point of action in the CrossCult project, in line with its final objective of interconnecting digital cultural resources, physical places and citizens’ opinions in a way that promotes reflection on history and cultural heritage.
To our knowledge, the work we have done in CrossCult has revealed some of the many possibilities of Artificial Intelligence to support a more effective narrative in the Digital Humanities applied to cultural heritage. The experience gained through the different pilots and experiments has provided evidence that various techniques of Artificial Intelligence—from the Semantic Web and graph theory to the recognition of entities and paragraph vectors—can provide significant support for discovering interesting associations and connections between digitized resources, user profiles and any characteristics considered relevant to the current context of this. These connections can be used directly to draw the attention of mobile technology users to selected pieces of content—either from the repositories of a particular museum or from publicly available sources such as Wikipedia or Europeana—or, in an alternative way, can serve as a basis for creating more specific narratives, with precise, reliable and specific content by experts.

From a critical point of view, it is worth noting that, even if the connections are strong, cultural heritage applications can be as good as the content they serve and the lasting messages they send. This fact highlights the importance of integrating Artificial Intelligence into the appropriate authoring tools, with features to support collaboration, the integration of local archives and diverse data resources, the collaborative creation of graphics for storytelling, visualization and management of associations, etc. It is about making the most of the time and effort of the experts, without leaving aside that it is possible to resort to the wisdom of the crowd, following the pioneering experiences of what is known as "computation by people" or "crowd computing" in the area of cultural heritage. We consider that this is a research niche that still needs to be developed in the future, with a lot of space for innovation and a remarkable market potential.

In short, we believe that having new and unconventional narratives that focus on cross-cultural and cross-border aspects can make a difference in the valuation of history and cultural heritage as a shared experience, since we have repeatedly observed—in different international scenarios, with different application concepts and in different target groups—that historical and cultural symmetries can support curiosity, reflection and understanding. From the purely technological side, we have been able to verify that Semantic Web technologies are at a point of development and adoption that, effectively, provides solutions in different application domains, after many years in which they questioned whether the cost of generation of formal annotations would be offset by the benefits derived from its exploitation.

The thesis consists of five chapters, grouped into three different parts:

- Part I (“Introduction”) contains two chapters. In Chapter 1 (“Introduction”) we examine the impact of ICT in the broad area of Digital Humanities in general, and in the field of Cultural Heritage in particular. The research objectives and contributions were also illustrated. Chapter 2 (“Background
and related work”), on the other hand, focuses on analyzing the contributions of other researchers in the past, providing an overview of the state of the art in narrative approaches and technologies. Applied to cultural heritage, in addition to the different technologies used in research, thus presenting a wide range of information on different concepts and technical concepts such as semantics, ontologies, discovery of associations and personalization.

- Part II (“Contributions of the thesis”) breaks down the contributions resulting from doctoral work into two chapters. In Chapter 3 (“Contributions to semantic modelling of knowledge about cultural heritage and technology-supported narratives”) we explain the narrative strategies developed within the framework of the CrossCult project, highlighting our contributions to different approaches, tools and experiments. On the other hand, in Chapter 4 (“Contributions to the discovery of associations, sensitivity to context and personalization”) we present the algorithms of discovery and personalization of the association that coincide with the contextual information of the dates. Relevant Also included are the results of the experiment conducted for the Archaeological Museum of Tripoli.

- Finally, Part III (“Conclusions and future work”) contains Chapter 5 (of the same title), which provides a summary of the conclusions resulting from this PhD work, together with some instructions for the future work, either a direct continuation of our contributions or suggestions to explore new potentials.
## Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>xix</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>xxi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xxiii</td>
</tr>
</tbody>
</table>

## 1 Introduction

### 1.1 The growing area of the Digital Humanities

### 1.2 Overview of ICT in Cultural Heritage

### 1.3 The context of the CrossCult project

#### 1.3.1 Related projects

### 1.4 Objectives and contributions

### 1.5 Thesis organization

## 2 Background and related work

### 2.1 The foundations of Semantic technologies

#### 2.1.1 General-purpose semantic resources: Linked Data

#### 2.1.2 Semantic modelling of temporal information

#### 2.1.3 Semantics for Cultural Heritage

### 2.2 Interactive storytelling

#### 2.2.1 Storytelling for cultural heritage

#### 2.2.2 Semantic approaches for storytelling

### 2.3 Discovery of semantic associations

#### 2.3.1 Association discovery tools

### 2.4 Personalization

#### 2.4.1 Semantics-based personalization

#### 2.4.2 Personalization in cultural heritage

### 2.5 Summary
## Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

### 3 Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

#### 3.1 Introduction

#### 3.2 A semantic knowledge base for reflective Cultural Heritage

- **3.2.1 Population of the CCKB**
- **3.2.2 Modelling topics of reflection**
- **3.2.3 Modelling historical events and commemorations**

#### 3.3 Strategies for technology-supported storytelling

- **3.3.1 Associations through coinciding or contrary features**
- **3.3.2 Associations through universal themes**
- **3.3.3 Associations through the users’ context**
- **3.3.4 Associations through trending topics**
- **3.3.5 Involving experts and interested audiences**

#### 3.4 An experience with associations between Europe and the Arab world

- **3.4.1 App design and features**
- **3.4.2 Preliminary evaluation**

#### 3.5 Conclusions

### 4 Contributions to the discovery of associations, context awareness and personalization

#### 4.1 Introduction

#### 4.2 Association discovery

#### 4.3 Personalizing associations

- **4.3.1 Important personal events**
- **4.3.2 Personal interests**

#### 4.4 Case study: The Archaeological Museum of Tripoli

- **4.4.1 General calendar-based associations for the Archaeological Museum of Tripoli**
- **4.4.2 Filtered associations for three personas**
- **4.4.3 Evaluation by Humanities experts and potential museum visitors**

#### 4.5 Conclusions

### Conclusions and future work

#### 5 Conclusions and future work

- **5.1 Conclusions**
- **5.2 Future work**
# List of Figures

1.1 Block diagram of the CrossCult platform .......................................................... 9

2.1 The Linked Open Data cloud diagram ................................................................. 20

2.2 Core relationships in the upper-level ontology of CIDOC-CRM ....................... 25

2.3 Samples of EDM classes ....................................................................................... 27

2.4 A screen capture of RelFinder ............................................................................. 36

2.5 A screenshot of the RelClus user interface .......................................................... 37

3.1 The CCKB stack of ontological layers ................................................................ 47

3.2 Stages of the data modelling process followed in CrossCult to populate the CCKB . 48

3.3 Semantics of the *Reflective Topic* class, with the example of Apollo and Muses .................. 49

3.4 Semantics of the date of the first public performance of the opera *Sappho* ...... 51

3.5 A cross-border association through one common feature of three historical characters .......................................................... 53

3.6 A graph of connections among the four venues of CrossCult’s pilot on cross-border associations ................................................................................. 54

3.7 Screen captures of the CrossCult app created for the National Gallery ................................................................................................................. 55

3.8 EuroVoc topics under “Social questions” → “Migration” ................................ 57

3.9 Screen captures of the CrossCult app created for the Archaeological Museum of Tripoli ........................................................................................................ 58

3.10 EuroVoc’s semantics for “flood” ......................................................................... 61

3.11 A cluster of pictures related to the semantics of a trending topic ...................... 62

3.12 Screen captures of CrossCult tools to discover associations and develop narratives ........................................................................................................ 63

3.13 Screen captures of the CrossCult pilot on the urban exploration .................... 64

3.14 Initial screens ....................................................................................................... 66

3.15 A bit of a story and related visual contents ....................................................... 67

3.16 A mesh of associations and questions ................................................................ 67

3.17 A multiple-choice question ............................................................................... 68

3.18 Seeking quick feedback about reflection points ................................................. 68
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Block diagram of the CrossCult platform</td>
<td>73</td>
</tr>
<tr>
<td>4.2</td>
<td>Example of association discovery between a museum item and a special date via a DBpedia concept.</td>
<td>75</td>
</tr>
<tr>
<td>4.3</td>
<td>Flowchart of the association discovery process, linking museum items and special dates indirectly via a DBpedia concept, and the selection of appropriate associations based on personal events and interests.</td>
<td>76</td>
</tr>
<tr>
<td>4.4</td>
<td>All keywords used to find the associations between special dates and the reflective narratives from the Archaeological Museum of Tripoli.</td>
<td>80</td>
</tr>
<tr>
<td>4.5</td>
<td>Distribution of associations found between reflective topics of the Archaeological Museum of Tripoli and different types of events.</td>
<td>81</td>
</tr>
<tr>
<td>4.6</td>
<td>Distribution of Mata’s final associations (up to one per date).</td>
<td>83</td>
</tr>
<tr>
<td>4.7</td>
<td>Distribution of Uter and Irmgard’s final associations (up to one per date).</td>
<td>85</td>
</tr>
<tr>
<td>4.8</td>
<td>Tag counts among potential museum visitors.</td>
<td>89</td>
</tr>
<tr>
<td>4.9</td>
<td>Tag counts among Humanities experts.</td>
<td>89</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Summary of the users’ tagging of personalized and random daily highlights. .................................................. 59

4.2 Tag counts for the sample associations and the linked narratives. ......................................................... 87
4.1 A sample association description computed for the case of Üter and Irmgard. ........................................ 88
Part I

Introduction
Chapter 1

Introduction

The development of Cultural Heritage (hereafter, CH), its role and the place it took in the European society led to the nomination of 2018 as the year of European’s cultural heritage. The citizens were given a chance to better understand their culture with its all different dimensions, while sharing responsibility for it [1]. In spite of the fact that the EU owns limited resources regarding the cultural heritage —the role of European institutions is only concerned with financial support and coordination of joint projects— it played an important role in raising the awareness about preservation, restoration and conservation issues. It also supported the technological research and scientific progress in technological solutions [2].

Furthermore, cultural heritage has been considered into different EU funding programs that facilitated Member States to take actions to revive their national or local heritage and to save their traditions and identity [3]. Resolutions have been adopted by the European Parliament to stress on the dangers that cultural heritage in EU and world should be protected from, and emphasizing the need to address trafficking and robbing of cultural heritage items, as well as the role of cultural heritage in sustaining tourism [4].

Accordingly, several EU funds have been made available for cultural heritage to achieve the so-called “Europe 2020” vision. Thus, a 1.46 billion euros budget was assigned by the Creative Europe program for the period 2014-2020 to cover mostly cultural and media projects [5] and the Horizon 2020 program has supported projects with over 100 million euros in the period 2016-2019. These projects have been related to different areas such as heritage science, societal challenges and industrial leadership. Their scope varies from interpretation of the past and mutual understanding of troubled pasts to heritage-led urban regeneration.

The Horizon 2020 also thought of harnessing the current breakthrough that technology is causing in society, aiming to exploit it in the cultural heritage
area. The projects funded in this line paid different relative importance to the following three dimensions [6]:

- **Humanities**: content creation, aggregation and curation, experience evaluation, related methodologies and techniques, user engagement (pre/post/revisit).

- **Technology**: content adaptation, personalization, tools to be used on-site and off-site, authoring tools, enhanced support for digitization, human-computer interaction, augmented/virtual reality, etc.

- **Dissemination & Exploitation**: business plans, intellectual property rights, Living Labs, common strategies, publications, advertising, etc.

In this line, this doctoral thesis has been conducted in the context of the project “CrossCult: Empowering reuse of digital cultural heritage in context-aware crosscuts of European history”, which aimed at using Information and Communication technologies (ICTs) to transform the experiences of the people who visit museums, archaeological sites or historical outdoor locations into more interactive ones, grounded on the mature technologies of the Semantic Web, the notable advances in Artificial Intelligence, and the many recent initiatives that have shaped the new area of the “Digital Humanities”.

### 1.1 The growing area of the Digital Humanities

The last twenty years have experienced a major change in the creation of knowledge and the communication processes, due to the rapid development of ICTs. The innovative and global nature that these technologies own had a direct impact almost on every aspect of life as well as most of the academic disciplines. They became tools for producing knowledge, study objects and a new domain of science itself [7]. The integration between humanities and technologies shaped what is known as “Digital Humanities” (hereafter, DH), embracing a set of ICT-related spirits such as digital libraries, visualization, geographic information system, social networking as well as digital culture. This makes sense as “humanities”, by definition, refers to academic disciplines that deal with the study of different aspects of human society and culture. The humanities include ancient and modern languages, literature, philosophy, history, human geography, law, politics, religion, sociology and art. The “digital” moniker can be added to all of those areas, so we can conclude that DH covers them all as usual, but aided with technologies.

Old researches used to concentrate on historical content. The practices and methods that digital technologies provided have facilitated different varieties of theoretical practices, such as examining complicated data sets and to reconstruct either tangible or intangible heritage. It includes a significant number of applications that aid the investigation of the humanities traditional research.
Digital Humanities make use of technologies to analyze and represent cultural and human records in parallel. Scientists now have the ability to find answers for analogue data that used to be stumbling blocks in the way. Thanks to new technologies, simulation and prediction is enabled, which has provided a powerful means to analyze intangible events. The now available techniques facilitated a wide range of researches that used to be impossible. By direct engagement with scientists, Digital Humanities change the informed perspective to the development and improving of the new technologies.

As for arts and artists, digital technologies ease the formation of new artworks for anyone who has the basic knowledge of technology. Technology accelerates the process of creating and achieving new effects; it also rises and extends the imaginative chances of the artworks [8]. The transfer from narrow physical environment to the wide world of the Internet allows not only new inspiration for artworks but to preserve old ones as well.

DH play a vital role when it comes to education too. The interactive and creative ways it uses in presenting and interpreting information make the youth much more attracted and involved in the learning processes, in ways that the traditional techniques don’t provide [9]. In addition, Digital Humanities researchers implement broader skills —either technical or collaborative— than others to improve and enhance both confidence and employability skills [9].

DH finally contribute with big share in the media. The mining of social networks and communication systems such as Facebook and Twitter, generating visualized reports and other reporting methods that allow to understand complicated sociological and psychological data [10]. Taking these results into consideration and working accordingly might lead to the creation of a better social capital, strengthening and improving social cohesion.

1.2 Overview of ICT in Cultural Heritage

Cultural heritage domain and new technologies have always had a very complex and controversial relationship, usually recognized as the follow-up of technologies that can become a burden during users’ cultural experiences [11].

The strategic merge between ICTs and cultural heritage is not something new [12]. Cultural heritage, tourism industries and technology have gone in a parallel direction and hand-to-hand for many years, indeed [13]. The breakthrough was that technology took the lead and changed the way business and services, as well as cultural experiences, are developed causing new opportunities and challenges [14].

The use of technologies introduced new types of activities that can alter the traditional cultural experiences into new and more interactive visitor-based experiences. Technology can work in both ways either as mediator or as the base of the experience [15]. It is important to capture and record the occurring
changes as technology only can’t develop different experiences, but it is a result of technology and cultural changes integration [16].

Innovative technologies, applications and services can reduce —which is among our goals in this thesis— the gap between cultural spaces (e.g. museums, art exhibitions, historical centres and archaeological sites) and citizens. Technology facilitates all interactions and connections between all the involved parties; enjoyment can be improved and people can appraise their culture in a more rewarding manner. Yet, many challenges are still to be faced, such as assisting individuals and groups inside the cultural venues, designing interoperable ICT solutions, managing cultural knowledge properly and making the cultural experience much more interesting and exciting [11]. This, undoubtedly, opens a wide research area. State-of-the-art solutions were discussed and presented by experts, and all were related to new technologies applications, methodologies and techniques in cultural heritage.

Knowing that technology has a great impact on all humanities applications, it touches cultural heritage in so many different and unique ways. It affects museums very directly and in a number of different points. Museums have the chance to manage their collections better, offer exciting experience for the visitors that exceed their physical limits by using online distribution and communication channels [17].

Moreover, the great revolution of the Internet brings into scene many different solutions. Internet as a product and service distribution channel, or as a communication and promotion means, can be used also by museums [18]. Virtual visitation, access to different educational and informational resources can be offered online. Museums in some countries offer e-shops where several products are sold. The Internet is being used to promote for their mission and events and to encourage the attendance to take place in the public activities provided by the museum. This leads to higher public awareness of cultural heritage, community supports and donations [19].

The recent advent of the Internet of Things (IoT) is further supporting the shift from closed environment, where objects were defined as descriptors, to an open environment where objects freely deal with their surroundings, as they became intelligent. As a result, objects such as cars, televisions and artefacts will be connected to the Internet [20], and so can happen with cultural or historical exhibits, which can be non-invasively supplemented with sensors to allow the remote management and monitoring. IoT is playing a significant role in different research areas, and in the last few years cultural heritage became one of the most appropriate fields, where huge achievements can be accomplished. The unique features of the domain and the different areas of interests can be considered as a fruitful soil, as people who have passion with cultural heritage and artefacts always have direct contact with real objects [20].
Another dimension of technology is that of virtual environments and augmented reality, which entail a notable potential to improve cultural heritage experiences for sizable audiences, complementing the current practices based on tangible assets such as museums, exhibitions, books and visual content [21].

The notion of serious games (SG) has also appeared as a new tool to appraise cultural (also health-related or socially-relevant) content in an engaging way. Such videogames, created from the grounds up with educational purposes in mind, provide amusing engagement and compelling experiences, which may retain the player’s attention for long-lasting sessions [22]. Many SGs help players to accomplish learning goals through fun, which is heavily influenced by different factors such as usability, visual appeal, competition, content depth and time requirements, just to name a few. The pedagogical approach is implemented by structuring educational content and presentation organization. As a result, the design process in SG is different from the common e-learning applications, as it has to balance between learnability and playfulness. Of course, learning content dominates in the game play, but researchers have agreed over the years that the interactions and mechanisms of the games shouldn’t be a funny layer added on top of a digital learning tool [23].

Not only does technology play an important role with increasing the number of visitors to museums; it also affects all cultural heritage stakeholders. Indeed, technology can enhance the learning offered to students at all levels in the fields of art, history and cultural heritage. Nowadays technologies are considered as convenient tools in the multifaceted sector of cultural heritage for renewing and enhancing education [24]. Thus, educational technology affects cultural heritage education in various aspects; current perspectives are considered when looking at cultural heritage pieces, when they are shown in digital form. Educational technology also reflects on the new opportunities offered to Cultural Heritage Education by ICT-enabled educational approaches and methods [24].

The role of ICTs passes to reach historians and Art experts to better do their jobs by facilitating classification or cataloguing task, or the identification of artworks related to their research from among huge collections, or by identifying commonalities between sets of pictures (e.g. same style or theme, or coinciding dates, etc) [25]. Even journalists and influential media can make use of technology in the field of cultural heritage. Any assistance to write a new article (e.g. about curious links between the date of today and some piece of heritage relevant to the readers) could be much appreciated, just like the automatic identification of historical events linked to any topic in preparation [26].

Beside all of the above, technology supports policymakers, as we can think of politicians and investors interested in assessing the potential benefit of promoting/investing one way or the other, on the grounds of substantial reports that not only provide bulk statistics of visitors but also insight into their interests, preferences and needs [27].
The CrossCult project lasted from March 2016 to February 2019, as a joint initiative led by the Luxembourg Institute of Science and Technology, the University of Peloponnese and the University of Vigo, with the aid of other partners like University College London, the National Gallery, the University of Malta and the University of Padova. Overall, the consortium started out with 11 partners and 14 associate partners —including social science, computer scientists, humanity researchers and historians— from 7 European countries. It was one of the four selected proposals in the REFLECTIVE-6 call of the Horizon 2020 program of the European Union.

The main goal of CrossCult was to change the way in which history and cultural heritage are appraised by Europeans, through the understanding and reinterpretation of cultural digital resources, peoples’ viewpoints and physical sites. This implied taking away the cultural barriers between EU countries in order to create unique perspectives that cross the borders, by linking existing digital historical resources and developing new ones, either by experts or by open public participation.

In order to accomplish that goal, the CrossCult consortium bet on using cutting-edge technology to link existing digital cultural assets and to merge them in different types of interactive experiences, most often delivered by mobile apps, in a way that stimulates reflection, increases retention, and increases appreciation for the links between past and present. The partners proposed the use of digital cultural assets to motivate a new shared cultural European history, to move from difference to unification.

The core ideas of CrossCult were tried in four real-world pilots involving 8 different sites in Europe, plus a number of side experiments focused on specific technologies. The consortium tried to ensure that stakeholders would be always involved through a Living Lab approach, and it also aimed to ensure future prospects by defining a concrete business plan, that would favour the creation of new companies and the development of new services related to European cultural heritage.

The core asset to realize this vision was the CrossCult Platform, which is a sophisticated bundle of software and applications ready to deliver services not only to current and future museum visitors (through Android or iOS apps) but also to different types of stakeholders, including museum curators, Humanities experts, data analysers, app developers and system administrators (through tailor-made web-based frontends). The operation of the web-based and mobile frontends is supported by a backend that provides infrastructure and instrumentation for hosting the core components, as presented in the overall block diagram of Fig. 1.1.
1.3 The context of the CrossCult project

At the core of the platform, the CrossCult Knowledge Base (hereafter, CCKB) is a repository for storage, management and retrieval of semantic information, i.e. information represented in formats that enable some interpretation of its meaning by computers. The CCKB implements a single and generic upper-level structure that acts as a layer of common concepts and relationships, building on standard technologies to facilitate interoperability and linking with third-party resources and open repositories such as Wikipedia and Europeana.

Next to the CCKB in Fig. 1.1, a number of software modules provide high-level, application-oriented services covering six major functional areas, namely "Association Discovery", "User Profiling", "Recommendation", "Context Awareness", "Social Networking" and "User Experience". Each functional area is covered by one or more technological modules, which offer distinct services within the functional area (e.g. chatting and micro-blogging) or address different facets of a single issue in a complementary fashion (e.g. carousel-based profiling vs interaction-based profiling, item recommendation vs path recommendation, etc.). More detailed information about the modules and services related to this doctoral work—as well as about their underlying technologies—will be provided in subsequent chapters of this thesis.

As a final note, it is worth highlighting that the CrossCult project took part in the Open Research Data Pilot, which targets to make the data generated by selected Horizon 2020 projects accessible with very few restrictions, while at the same time protecting sensitive data from unauthorized access. Information on what data has been made available—including some of the data gathered as part of this doctoral work—can be found at the project’s web site [www.crosscult.eu](http://www.crosscult.eu).
1.3.1 Related projects

The following is a list of EU-funded projects whose implementation overlapped and influenced CrossCult to different extents:

- **EMOTIVE**[^1] (*Storytelling for cultural heritage*): this project was seeking verifiable, generalisable triggers for emotive experiences in relation to cultural heritage, meaning acts of reciprocity, imitation, replication via verbalisation, personalization / familiarization of experience, legitimate decision-making, humour, conversation, body themes and social interaction between museum visitors. The project developed authoring tools for mobile on-site and web-based virtual museum experiences, as well as low-cost and high-quality 3D reconstructions and creation of 3D printed objects.

- **meSch**[^2] (*Material encounters with digital cultural heritage*): this project focused on co-designing novel platforms for the creation of tangible exhibits at heritage sites, offering visitors new interactive experiences by means of material interaction with smart objects. The project sought to make innovate use of existing digital repositories by integrating them in tools to create adaptive smart exhibits.

- **WhoLoDance**[^3]: This project aimed at investigating bodily knowledge, thus helping to preserve the cultural heritage embedded in dance representations, innovate the teaching of dance and revolutionize the discipline of choreography. The consortium worked to create and enrich a digital library of movements from ballet, contemporary, flamenco and Greek dance using motion capture to reuse in educational contexts. They also designed and developed whole-body interaction experiences for learning and teaching dance, including uses of Augmented Reality and Virtual Reality.

- **iMARECulture**[^4] (*Advanced VR, immersive serious games and Augmented REality as tools to raise awareness and access to European underwater cultural heritage*): this project aimed in raising public awareness of European identity by focusing in maritime cultural heritage, with pilot experiences developed in underwater archaeological sites and notable shipwrecks in Cyprus, Malta and Italy.

- **INCEPTION**[^5] (*Inclusive cultural heritage in Europe through 3D semantic modelling*): the main goal of this project was to realise innovation in 3D modelling of cultural heritage through an inclusive approach for

[^1]: [www.emotiveproject.eu](http://www.emotiveproject.eu)
[^2]: [www.mesch-project.eu](http://www.mesch-project.eu)
[^3]: [www.wholodance.eu](http://www.wholodance.eu)
[^4]: [imareculture.eu](http://imareculture.eu)
[^5]: [www.inception-project.eu](http://www.inception-project.eu)
time-dynamic 3D reconstruction of artefacts, buildings, sites and social environments. The partners developed an open-standard Semantic Web platform for accessing, processing and sharing interoperable models resulting from 3D survey and data capturing, including means for representation and consolidation of imperfect data and spatio-temporal analysis, as well as uses of Augmented Reality and Virtual Reality.

• **POEM**\(^6\) (*Participatory memory practices*): This project developed concepts, strategies and media infrastructures for envisioning socially-inclusive futures of European societies through culture, empowering membership and contribution of people from diverse social and cultural backgrounds. The key concept was connectivity across sectors built by memory institutions, people and groups, and memory modalities, under consideration of open knowledge media literacy and data competence, social entrepreneurship and innovation, empowerment, organization and leadership in the non-profit sector.

• **GIFT**\(^7\) (*Meaningful personalization of hybrid virtual museum experiences through gifting and appropriation*): this project aimed to establish meaningful user experiences that allow for personal, complex and emotional encounters with art and cultural heritage, through digital gifting and emotional appropriation. The former concept is based on personalization of museum experiences by allowing visitors to create their own museum tours as digital *mixtapes* and share them as gifts. The latter, in turn, is based on a playful re-appropriation of museum spaces by enabling strong identification with museum collections and sharing emotional responses to artwork.

• **ArchAIDE**\(^8\): This project aimed to support the classification and interpretation work of archaeologists with innovative computer-based tools, able to provide the users with features for matching of each discovered sherd over the huge existing ceramic catalogues. The central concepts were the automatic recognition of archaeological pottery by deep learning, the automatic digitization of text, the conceptual and linguistic mapping of archaeological pottery and the *datafication* of archaeology.

• **ARCHE**\(^9\) (*Accessible resources for digital heritage ecosystems*): This project aimed to create more inclusive cultural heritage environments for people through an in-depth research analysis and the development of innovative and accessible applications and experiences. The key assets included exploration groups in six European museums, a fully-participatory

\(^6\)www.poem-horizon.eu
\(^7\)https://gifting.digital/
\(^8\)www.archaide.eu
\(^9\)www.arches-project.eu
research methodology, accessible platforms and apps, sign language avatars and serious games, and tactile reliefs and audioguides.

- **DigiArt**\(^{10}\) (*The Internet of historical things and building new 3D cultural worlds*): This project sought to provide a new cost-efficient solution to capture, process and display cultural artefacts, featuring innovative 3D capture systems and methodologies, including aerial capture via drones, automatic registration and modelling techniques to speed up post-capture processing, semantic image analysis, a storytelling engine offering pathways for deeper understanding of art, and Augmented/Virtual Reality technologies for viewing and interacting with the 3D models.

- **I Media Cities**\(^{11}\): This project aimed to create an online environment where different target users (researchers, archivists and the general public) could view, research, annotate and interact with digital cultural audiovisual heritage about European cities. The partners worked with visualization of contents on interactive maps and timelines, analysis by target users and automatic tools, discovering similarities in European history through the audiovisual materials, and a Linked Open Data approach as the basis for annotations and multilingualism.

- **NM** (*QoE-driven museum touring - Human-in-the-loop approaches*): This project aimed at creating autonomic visitor-centric management and optimization of cultural heritage tours, dealing automatically with congestion management to maximize Quality of Experience (QoE) metrics. The consortium worked with machine learning techniques for intelligent recommendation selection, as well as game theory for QoE-driven optimization in a distributed manner.

- **Pluggy**\(^{12}\) (*Pluggable platform for heritage awareness and participation*): This project aimed to create an online platform to boost the participation of citizens in creating, modifying and safeguarding heritage, focusing on the heritage that is all around us, everyday and non-prominent. The technological innovations had to to with geolocation, social networks, 3D modelling and gamification concepts.

Overall, these projects conform a good picture of what the area of Digital Humanities is nowadays in relation to cultural heritage, highlighting the main technological trends, most of which are surveyed as relevant underpinnings for this doctoral work in Chapter 2.

\(^{10}\) digiart-project.eu
\(^{11}\) imediacities.eu
\(^{12}\) www.pluggy-project.eu
1.4 Objectives and contributions

As this doctoral work was conducted under the umbrella of the CrossCult project, it shares overall motivation and goals with it: using recent developments in information and communication technologies to improve the appraisal of European cultural heritage. More specifically, we have investigated the potential of interactive, context-aware storytelling in serious games, as a means to raise interest in cultural heritage and history. Furthermore, we have investigated whether and how Semantic Web technologies can be used to drive the discovery of associations or crosscutting topics among multiple heritage items and locations, historical characters or events, etc. The discovery of such connections is seen as the seed for the creation of curiosity-raising narratives to drive the abovementioned serious games. Finally, we have developed a new avenue for the personalization of cultural heritage experiences, by linking any relevant historical dates or yearly commemorations to the interests of any particular user of mobile technology, aiming to serve the purpose of delivering more deeply-tailored narratives.

The objectives of the thesis can, therefore, be enumerated as follows:

(i) Contribute to the application of Semantic Web technologies in the area of cultural heritage, in a way that eases the exchange of information and the integration of multiple resources and, thus, facilitate the automated discovery of meaningful connections.

(ii) Integrate existing Linked Data resources and newly-created ones to enable reasoning about heritage items, cultural venues/sites and the spatio-temporal context of ICT users.

(iii) Implement association discovery and recommendation systems to identify relevant connections among user profiles, cultural heritage knowledge bases and a semantic almanac of dates.

(iv) Design and implement tools for cultural heritage experts to curate the automatically-discovered associations and, therefrom, design interactive narratives about topics of their interest.

(v) Tighten the gap and helping people to better understand their national identity and to find common associations with other cultures.

The main contributions resulting from the work, in turn, would be as follows:

• We have aided in the formalization of storytelling strategies that maximize the desired outcomes in terms of reflection, reinterpretation, curiosity raising, etc. Furthermore, we have contributed to the creation of tools for experts to create narratives based on the associations discovered by the
aforementioned logic, as well as a mobile app that delivers those narratives as the driver of a serious quiz game. We have also created and evaluated one sample of narratives that seek to promote mutual knowledge between the European and Arab worlds.

• We have produced a prototype of databases and logic needed to run association discovery and recommendation systems as part of the catalogue of services offered by the CrossCult platform, plus an annex to the CCKB in the shape of a newly-created semantic almanac of historical dates and yearly celebrations, obtained through data mining on different sites of the World Wide Web: OnThisDay.com, the web site of UNESCO, Wikipedia, etc.

• Finally, we have created a personalization system based on vector space models that rates the discovered associations according to the interests and personal events of any specific user. This system was evaluated in a pilot study on the collection of the Archaeological Museum of Tripoli (Greece), finding that the discovered associations are meaningful and usable to trigger curiosity and deliver unexpected learning.

1.5 Thesis organization

The thesis consists of five chapters, grouped into three different parts:

• Part I contains two chapters:
  - In Chapter 1, we have looked at the impact of ICTs in the broad area of Digital Humanities in general, and in the Cultural Heritage realm in particular. The research objectives and contributions were illustrated as well.
  - Chapter 2 concentrates on related work, providing an overview of the state-of-the-art in storytelling approaches and technologies applied to cultural heritage, plus the different technologies that were used in the research, thus presenting a wide background about different concepts and techniques such as semantics, ontologies, association discovery and personalization.

• Part II contains the contributions of the doctoral work, in two chapters:
  - In Chapter 3, we explain the storytelling strategies developed in the framework of the CrossCult project, highlighting our contributions to the different approaches, tools and experiments.
  - In Chapter 4, we present the association discovery and personalization algorithms that match the contextual information of relevant
dates with (i) the cultural heritage represented in the collection of a given venue, and (ii) the interests captured in a user’s profile. The results from the experiment conducted for the Archaeological Museum of Tripoli are included, too.

- Part III contains chapter 5, which provides the summary of the conclusions resulting from this doctoral work, together with some directions for future work, either direct continuations of our contributions or suggestions to explore new potentials.
Chapter 2

Background and related work

This chapter provides an overview of the state-of-the-art in the main areas related to the accomplishment of the goals stated in Chapter 1. Thus, we look at the main concepts and precedents in terms of Semantic Web technologies, storytelling, association discovery and personalization, first in general terms and then with specific attention to uses in the realm of Cultural Heritage.

2.1 The foundations of Semantic technologies

Thanks to the World Wide Web, people nowadays can share information and data easily and globally. The huge volume of data and the enormous increase in the number of web pages made the traditional search engines ineffective and unsuitable anymore [28]. Needless to say that search engines are the most powerful tools for the World Wide Web discovery, in order to overcome their weaknesses, semantic technologies came to light as a means to enable computers to look at the meaning of the information, and thus maximize the value of the search results [29].

As the Internet is fast developing, it gave users an easy and effective way to access information and services. The enormous semi-structured database of the World Wide Web offers massive volumes of data and information. With its increasing size, however, new challenges arise not for only for detecting relevant information, but also for retrieving information automatically from various resources [28]. Semantic search engines aim to reason about the meaning of the resources, so as to yield only results that the user is targeting, as they use well-defined vocabularies and terms to accurately describe the information [30].

The Semantic Web developed new definitions for the web contents, allowing both humans and machines understand the information [31]. It is considered as extension of the World Wide Web, where the meaning of information and services semantics are defined. Semantics make it possible to understand and satisfy people’s requests and machines to make use of the web content [32]. The
term first appeared in the beginning of the 2000s to refer to this extension; since then it became a fertile soil for research providing a huge number of relevant results [33], and it enables new, unforeseen uses of the technology in recommendation systems, syndication tools, resource management systems, etc.

The role of semantics also extends to providing solutions to problems of interoperability, as it supports data discovery, integration and reuse. These problems were obstacles that used to face web technologies [32]. It is possible to integrate different content, information and applications for purposes of publishing, blogging, etc. [34]

The “Semantic Web” term is usually called on the technologies that empower it. Linked data collection and structuring is supported by technologies that give formal descriptions of concepts and relationships in certain knowledge domains. W3C standards have identified these technologies to include, among others, the Resource Description Framework (RDF), RDF Schema (RDFS), the Simple Knowledge Organization System (SKOS), the SPARQL query language for RDF, and the Web Ontology Language (OWL) that is used for knowledge representation [35].

Now, we would like to explain some terms: ontologies, taxonomies, thesauruses and vocabularies, in order to know the role of each and understand their differences, even if the frontiers have become somewhat fuzzy as more and more resources have been publish that fall in intermediate positions among concepts that were clearly different at first.

In this line, the word ontology is often used to describe different things such as glossaries, data dictionaries, thesauri and taxonomies, as well as data models and, properly, formal ontologies [36]. To be more specific with the definition, a formal ontology is a controlled vocabulary, subject to certain rules of use, that serve to represent concepts from a specific domain [37]. Formal ontologies have their own grammatical formal constraints, such as the specification of what a well-formed statement is, and the same goes for assertions and queries. There are foundational ontologies, used to provide a general framework for reasoning, and domain ontologies, used to model entities from specific areas.

Since we’ve said that a formal ontology is a controlled vocabulary, it’s necessary to explain that a controlled vocabulary is, basically, a set of words, i.e. a closed list of terms or expressions that one can choose from. Examples for controlled vocabularies could be “Yes, No” and “Boy, Girl, Man, Woman”. To be considered a controlled vocabulary, there must be a kind of authority in order to agree upon the measurements and procedures if any list requires changing [38].

Ontologies and annotations are the key foundations of the Semantic Web. RDF and OWL languages are used to represent ontologies and to describe shared conceptualizations of a domain [39]. On the other hand, annotations bind ontology-based descriptions to existing resources, which are usually described in RDF too [40].
A controlled vocabulary is just a list of terms without a specific order; therefore, the terms may be displayed in alphabetical order, sorted by length, etc. If there is any precise order for those lists, that could create metadata about their relationships, then we will be talking about a taxonomy and not a controlled vocabulary anymore [41]. A taxonomy, indeed, is often described as a set of controlled vocabulary terms that have a hierarchical structure. Actually, the terms of a taxonomy can have one or more parent-child relationships with one another, and some taxonomies allow terms to have multiple parents [38].

The last term we have here to identify is the thesaurus. A thesaurus is a network or mesh of controlled vocabulary terms. It has more metadata than a taxonomy and makes use of both associative as well as parent-child relationships [42]. Thesauruses can express associative relationships in different and simple ways; also terms can be co-related to one another [43].

Besides search engines, Semantic Web resources can be useful in so many different applications, such as in publishing and media [44], oil & gas supply and operations management [45], life sciences and pharma [46], etc. For the purposes of this thesis, they provide the underpinnings required by the goal of modelling the knowledge available about individual exhibits or the overall collections owned by different types of cultural venues, and then also for matching that knowledge against the information available about the visitors and their interests. These tasks require a level of automated processing that can’t be achieved with traditional relational databases, plain text processing and hyperlinking.

2.1.1 General-purpose semantic resources: Linked Data

The need of humans and machines to exchange information in an accurate and effective way was the reason that made ontologies and semantic services more widely used in knowledge sharing processes. Next, we will focus on the most famous semantic resources, many of which have been used in our work. These are all part of the so-called Linked Data initiative: a massive collection of semantic resources, more or less tightly related to one another, than can be reused for many different purposes within a common framework that is curated over time by thousands of contributors worldwide. Figure 2.1 shows a representation of the Linked Data cloud, with each circle representing a different resource.

When it comes to integrating heterogeneous knowledge coming from several resources, then upper-layer ontologies (also upper ontologies, in short, or foundational ontologies) are the answer. Upper ontologies define general-purpose classes and properties that may be used by different parties as a reference, common model of reality. In [47], an upper ontology is “an attempt to create an ontology which describes very general concepts that are the same across all domains, with the aim of having a large number of ontologies accessible”. Arguably, the most popular such ontology is the Suggested Upper Merged Ontology (SUMO) [48], which is —
Figure 2.1: The Linked Open Data cloud diagram. [Credit: Richard Cyganiak and Anja Jentzsch, http://lod-cloud.net]
according to the working groups in charge—“the largest formal public ontology in existence today”, used for plenty of applications in information search, document processing, reasoning, etc. SUMO is extended with many domain ontologies and a complete set of links to many other resources. SUMO proper consists of 1000 terms and 4000 axioms; counting also its domain ontologies, however, it goes beyond the dimension of 20000 terms and 60000 axioms.

Another popular upper ontology is the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), which aims at “capturing the ontological categories underlying natural language and human commonsense” [49]. Using this ontology, it is possible to co-locate different entities in overlapping regions of space and overlapping periods of time. The taxonomy of the most basic categories assumed in DOLCE includes, for instance, abstract quality, abstract region, amount of matter, agentive and non-agentive physical object, physical quality, physical region, process, temporal quality, etc. Altogether, it contains around 200 terms and axioms.

At the centre of Fig. 2.1 appears DBpedia which is a huge compendium of semantic knowledge extracted from the articles published in Wikipedia. DBpedia defines classes and properties of its own, which are called the DBpedia Ontology. This is a cross-domain ontology, created manually from the most commonly used infoboxes within Wikipedia. As of 2018, the DBpedia ontology covers 685 classes, which are described by 2,795 different properties that, in turn, have been related to SUMO and other Linked Data resources. One remarkable feature is that the knowledge base comes along with tools that allow users to type queries for semantic relationships and properties of Wikipedia resources, including links to other related datasets (which serves as a source of supplementary knowledge in many applications) and to perform entity recognition in text, aiming to identify which concepts mentioned in the paragraphs have a corresponding entry in DBpedia (ergo, in Wikipedia too).

YAGO shares some of the goals of DBpedia. This is a sizable semantic knowledge base, derived not only from Wikipedia, but also WordNet (an enhanced thesaurus, to be described below) and GeoNames (a collection of hierarchically-organized names of locations worldwide). As of 2018, YAGO contains semantic characterizations of more than 10 million entities (e.g. people, organizations, cities, etc.) and contains more than 120 million facts about them. It is advertised as “an ontology that is anchored in time and space” by its authors, in the sense that it does attach temporal and/or spatial dimensions to many of its facts and entities.

WordNet is a large lexical database of English, commonly used for computational linguistics and natural language processing. This database brings
together different types of words (mainly nouns, adjectives, verbs and adverbs) into the so-called *synsets*, on the grounds of synonymy relationships. The synsets are interlinked, giving information about inclusion or subsumption of meaning, partial overlaps, etc. This, together with the fact that WordNet labels the semantic relations, goes beyond the scope of a thesaurus that simply groups words together based on their meanings, thus allowing advanced forms of semantic disambiguation.

The following are some additional resources that, despite not being so widely used as the aforementioned, have been relevant to this doctoral work:

- **Dublin Core (DC)** [50] is a small set of vocabulary terms that can be used to describe digital resources (video, images, web pages, etc.), as well as physical resources such as books and objects like artworks.

- **Friend of a Friend (FOAF)** [51] is an ontology used to describe people, their activities and their relations to other people and objects, acting as a common foundation for online social networks.

- The **Simple Knowledge Organization System (SKOS)** is a W3C recommendation “designed for representation of thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary” [52]. Its main goal is to enable easy publication and use of such resources as linked data. It defines classes and properties to represent common features found in a thesaurus, based on a concept-centric view of the vocabulary. Each concept can be attached to RDF properties, including one or more preferred index terms, alternative terms or synonyms, definitions and notes involving definitions from other ontologies, etc. Besides, concepts can be organized in hierarchies using broader-narrower relationships, or linked by non-hierarchical relationships.

- **EuroVoc** [53] is the European Union’s multilingual and multidisciplinary thesaurus. Formally a thesaurus ontology schema that extends SKOS and Dublin Core, it contains keywords, organized in 21 domains and 127 subdomains, which are used to describe the content of documents. Firstly, it was used with EU legal documents, but it has been applied to other areas, since it provides an extensive and fine-grained classification of transverse societal topics. EuroVoc features a solid indexing tool that enables the effective management of its documentary resources.

### 2.1.2 Semantic modelling of temporal information

Since one of the contributions of this doctoral work relates to the management of dates as contextual information for cultural experiences, we now pay some attention to how temporal information can be managed by different Linked Data resources. This aspect is not properly catered for by the ontologies
2.1 The foundations of Semantic technologies

mentioned above. Indeed, DBpedia and SUMO per se don’t have any entities that can handle the temporal events. In turn, YAGO can model some information as it has a specific class called `YagoDate` that defines time points, plus properties `StartsExistingOnDate` and `EndsExistingOnDate` that define start and end points [54], but these constructs are insufficient to go deep in capturing the knowledge about historical events and such relevant dates as yearly commemorations. Finally, DOLCE can merely use `endurant` and `perdurant` to represent entities that fully exist in an instant of time and those which may have temporal parts, respectively.

The most noteworthy upper-level ontologies that were developed with concepts for events and temporality include the Unified Foundational Ontology (UFO)\(^4\) and the Event ontology\(^5\). The latter seems to be the most famous resource for modelling events, as it can have place, time, factors, agents and products, and in addition each event can consist of sub-events. The Linking Open Description of Event (LODE)\(^6\) has the same purpose, yet it focuses more on describing historical events and adds properties to link things with events for better illustration. The Simple Event Model (SEM)\(^7\) ontology is yet another alternative, with very detailed links to other ontologies and vocabularies.

Although the aforementioned can identify when and where a certain event took place, who acted in this event and what the relevant factors were, they can’t identify what kind of event or action it was. This can be done by defining subclasses, and there are different resources specializing those constructs for each specific domain. For example, SUMO contains a taxonomy of social events under the class `Social interaction` as well as many types of personal events by means of mappings to WordNet synsets, which enable reasoning because of the explicit modeling of hypernyms, hyponyms, part meronyms, derivationally related concepts, etc. Likewise, DBpedia and YAGO contain taxonomies that include many types of personal and societal events.

There are several choices, too, for the actual representation of temporal information. The most common is the OWL-Time ontology\(^8\) which specifies the class `TemporalEntity` with subclasses for both instants and intervals. It can specify their relationships using properties such as after and before. These core entities basically have `DateTimeDescription` or `DurationDescription` which have the date/time or duration specified with a set of datatype properties. The SEM ontology defines its own set of properties, for example, to define open-ended and uncertain intervals.

\(^4\)https://oxygen.informatik.tu-cottbus.de/drupal7/ufo/
\(^5\)https://lov.linkeddata.es/dataset/lov/vocabs/event
\(^6\)http://linkedevents.org/ontology/
\(^7\)http://semanticweb.cs.vu.nl/2009/11/sem/
\(^8\)https://www.w3.org/TR/owl-time/
It is also worth mentioning Basic Formal Ontology (BFO), whose main relationships are defined in the Relation Ontology (RO) [55]. BFO make use of the terms continuant and occurrence.

2.1.3 Semantics for Cultural Heritage

In order to speak about semantics in the specific realm of cultural heritage, we have to differentiate between the roles of the previously-mentioned instruments and how they are used to manage the information in the hands of cultural institutions. Several semantic instruments take part in the field of cultural heritage, but the most important ones are the Conceptual Reference Model (CRM) of the International Council Of Museums (commonly referred to as CIDOC-CRM) and Europeana Data Model (EDM). The former is the most famous and widely used domain ontology in Digital Humanities, currently used by a growing number of museums worldwide; the latter, in turn, is widely used by cultural stakeholders explaining how they are being developed and aid the cultural heritage area.

2.1.3.1 CIDOC-CRM

CIDOC-CRM is a standardized abstract model (ISO 21127:2006) that was developed from scratch as a tool to represent both implicit and explicit knowledge of cultural heritage and to provide the semantic connection between different cultural heritage information sources [56]. It is the dominant ontology in cultural heritage field, representing the categorical knowledge of this domain. Although being abstract and flexible, its use has seen some difficulties, which are nonetheless exceeded by its advantages [57]. Taking a deeper look into the abstract concepts it defines to describe cultural items and artefacts, we find that it consists of classes and logical groups of properties, plus a set of entities with different types, that represent real-world things connected by relationships. The reasoning depends on using the right relationships with the right entities, which is the reason why it’s a must to understand the initial mapping process [58].

CIDOC-CRM entities and relationships are general abstractions for huge number of cultural heritage data models, using data collected from number of experts through meetings and workshops and by exchanging knowledge with different institutions. Two main concepts are persistent and temporary things, as well as the hierarchy of entities. Persistent things are those that don’t change over time or along with events such as physical things and concepts, while temporal concepts are those that happen in certain time frame or period and not permanently exist (for example, events or activities like historical periods). Second, the hierarchy of entities gives different levels of generalization as we can’t always describe everything precisely [58]. Figure 2.2 illustrates the upper level of CIDOC-CRM top classes.
2.1 The foundations of Semantic technologies

All the instances in CIDOC-CRM, furthermore, can be annotated with types, which most commonly determine the set of properties that refer in general to things of a certain kind. Besides, any instance of a class can be linked to instances of Appellation (E41), which can be used for example to model names, labels, titles and common tags in the historical context.

Regarding the way in which CIDOC-CRM handles temporal information, it takes an event-based approach, so Temporal Entities (E2) exist at its core, representing matters that took place at a given moment or during a certain period in the past. Thus, Temporal Entities can have Time Spans (E52), while Objects—either Conceptual (E28) or Physical (E18)—can’t have a direct connection with time. Actors (E39) and Places (E53) work the same as objects. Geographical places and locations, as well as specific places such as a court or play field, are included in Place Entity (E53). Individuals or groups are modelled as Actors (E39), who can interact with both tangible things called Physical Things (E18) and with intangible things called Conceptual Things (E28). Type Entity (E55) is for classifications as organizations use different classifications systems, any number of types can be applied as many times as needed [59].

2.1.3.2 The Europeana Data Model

Another important tool in the cultural heritage domain is the European digital library called Europeana, a digital access point for the European multilingual cultural heritage. It collects metadata from different cultural organizations and facilitates the search and retrieval of information about cultural artefacts of many different types [60]. This repository is based on the Europeana Data Model (EDM), which specifies a number of classes and properties that can be used to describe the cultural items.
Europeana’s specifications describe different vocabularies and ontologies used in its model in order to represent cultural heritage artefacts and to grant full access for their digital representations. The EDM is founded on the Dublin Core definitions, plus a set of different metadata models that allow providers and organizations to use their preferred metadata standards [61], so it is common to find uses of SKOS, DC and CIDOC-CRM, as well as thesauri and other resources like the Art & Architecture Thesaurus (AAT) [62], WordNet and DBpedia. As a result, the description of the cultural objects has extreme heterogeneity, but users are still able to perform much more accurate queries and to retrieve better results thanks to the complex ontological metadata model used by Europeana’s search engine.

The EDM seeks full data integration as the objects can get more and more complex, and so can different organizations view them differently. The main aim is to work on Linked Data both exposing record sets [62] and using Linked Data resources [63] in order to increase Europeana content.

Europeana makes use of Dublin Core elements and terms, such as dc:subject, dc:contributor, dcterms:created and dcterms:alternative, as they are considered the most popular means for cataloguing and description [64]. As for SKOS, it’s essential for Europeana to classify the published resources according to a classification schemes; the main class is skos:Concept, that is identified as a specific type of edm:NonInformationResource to introduce an idea or notion [64].

Notably, some of the Europeana classes and properties are aligned to CIDOC-CRM specifications. Thus, the class edm:Event is equal to E4 Period, while the class edm:InformationResource is a subclass of Information Object and the property edm:wasPresentAt is equivalent to P12 (Occurred in the presence of). On the other hand, Europeana also has its own core classes, such as edm:ProvidedCHO for the source cultural heritage object, edm:WebResource for the digital object being submitted, and edm:Aggregation for the overall package [60]. Figure 2.3 provides samples of EDM classes.

### 2.2 Interactive storytelling

Storytelling goes almost into all aspects of humans’ daily lives, either it is to inform, convince, inspire, educate or motivate people [65]. Nearly everyone tells stories in one way or another. However in spite of their power, not all stories can be effective, grab the attention of their intended audiences or deliver their messages.

In an era marked by the extensive use of ICTs, digital storytelling is considered to be the most effective way to deliver content in digital era, as it combines different multimedia elements and features (e.g. text, slide shows, animations and videos in addition to audio), following a given narrative [66]. The term storytelling implicitly indicates that users won’t just get information or data, but
2.2 Interactive storytelling

Figure 2.3: Samples of EDM classes.

they will get a full story with its all parts, serving the purpose of delivering clear messages related to whichever aspects and goals, from politics to education, from science to cultural heritage.

Story narrating is well-known (ever since Antiquity) as a means for grabbing audiences engagement and attention [67]. Accordingly, many studies have suggested that digital stories can boost the engagement of people to content, even across different devices, from tablets to desktops and using smartphones [68]. Of course, different styles of narratives have been tried, including immersive environments such as those provided by Augmented Reality and Virtual Reality.

For the purposes of this thesis, the most relevant works to survey involve multimedia storytelling with an extra feature: interactivity. Merely videos or movie clips are not considered, as they have linear flow: a starting point and an ending point, and only one way to progress between them. Users are expected to follow such stories in one way only, and the narrative writers count on the fact that stories have linear sequence and that minute $x + 1$ depends on minute $x$. Linear stories recently can have partial interactivity through modern devices, as users can navigate forward and backward in the plot or skip unwanted parts, but the possibilities end there [67].
Things become totally different when it comes to interactive narratives. A narrative is considered to be interactive if it’s made of semi-independent fragments, which may also be called episodes. Fragments consist of multiple pieces of text and multimedia contents, but each fragment has its own meaning and can be understood by itself. Different plot styles can be used to put several episodes together, for example, following linear, inter-linked, hyper-linked or hierarchical structures [69]. However, the organization of the episodes is flexible, meaning that the user can exercise features such as systematic browsing, linear walkthrough, random selection, exploration and search. All of these make the experience attractive for different users with their different needs [67][69].

The following are some hints that can be useful for writers to develop effective interactive narratives [67]:

- **Raw content.** This can be described as the materials that create the stories. It includes documents, books, images and videos. The critical point is that generic information does not build up a story: a story must be interesting and engaging, in content and in presentation. So, the task is getting more complex and harder if the starting point is a huge amount of raw materials, as one has to decide what would be interesting for the users or what should be selected and presented to them. To overcome this predicament, experts are often interviewed to collect bulks of required information focusing on what is really valuable.

- **Fragmentation.** Interactive stories are composed of fragments or episodes. A fragment is a short tale that has enough length to say something interesting and useful, not so short to lose users’ concentration and long enough to deliver certain information. The author has to deliver an idea in approximately 120 words, and also must put into consideration that each fragment has to be reasonably consistent by itself.

- **Media.** It is important to create stories that can be presented in different means: they have to be short, with easy syntax. Text can be displayed on mobiles and also can be converted into audio. This allows users to listen to stories while they are visiting a museum or taking a walk outdoors. Combining images with audio strengthens the message. Images are rich with captions, which make the stories much more understandable. Videos also are good medium but expensive, that’s the reason why videos are often generated by playing images in sequence with audio accompanied with audio.

- **Structure.** Probably, the hardest task is to structure a narrative from fragments or episodes. The basic idea is to create linear narrative which is a set of fragments played in order or randomly. A playlist, in general, should not to exceed 8 fragments. Another option is to have different playlists that are coordinated in meaningful ways, where fragments can be
interlinked to make use of semantic references. One more possibility is to create a hierarchical structure, for example a playlist of 6-8 fragments, so that each of them is linked to 4-5 more fragments with supplementary information. The user then can select the short version, containing the main fragments only, or the long version that contains all of the episodes. Finally, it is also possible to add together fragments from different playlists in an unstructured manner, allowing the users to directly access any fragment.

- **Adaptivity.** A fragment should be designed for a certain purpose and context, which has to be made explicit in order to control the possibilities for reuse in broader, narrower or non-overlapping situations. A simple technique is to add prefixes or suffixes to fragments, to facilitate classification and search.

### 2.2.1 Storytelling for cultural heritage

Multimedia communication is seen by most stakeholders in the cultural heritage area as an essential aid to reach wider audiences, making use of the features offered by smart devices to integrate cultural and tourism, and to convey the messages deemed relevant for educational purposes, or even political ones (e.g. related to the consolidation of some national identity) [70]. In this line, storytelling is defined as the way of getting audiences or visitors engaged, as a means to ensure the delivery of such messages [71].

In particular for the historical and cultural scope, narratives are considered an effective way to improve learning capabilities, increasing retention thanks to emotional impact [72]. Their relevance increases because the transmission of event-related historical data is facilitated by the fact that the users already have a mind-structure for historical knowledge [73]. It has been proved, for example, that drama-based narratives containing accurate references to a site’s cultural content have the power to shape museum visitors experiences, encourage them to repeat the visits and facilitate direct interaction and deep knowledge transfer [74]. This phenomenon gets deeper knowing that the members of the audience are facing real facts, or at least one of their possible interpretations. As a result, telling narrations about historical events not only improves the learning about those events, but enhances the natural narrative process as well [75]. One disadvantage, though, is that in some historical places, narratives usually have a narrow scope: they are used as a means to deliver the findings and research conducted by the experts of the domain in the location, which reflects only their own experience and most often sacrifices emotions or impact for objectivity, neutrality and historical rigour [72].

When it comes to visiting museums, heritage items or art pieces in public places, three time frames can be identified: before, during and after the visit. ICT developments have combined these three moments, in order to explain the physical and virtual visit in a related way, in site or on distance [70]. Digital
storytelling tools and services tend to remove and diminish the gap between the virtual and on-site visits [74]. The virtual visits allow visitors to know the museum even before they enter it physically, and they can use social networks and web services to interact with other visitors and the institution as well [76]. At the time of the physical visits, they enter a cultural environment in which all objects and art pieces are connected into narratives, that they can explore using different digital devices such as smartphones and touchscreens. Some authors have called this approach “a world of narratives”; the ARTbabble [71], for example, has become a huge collection of stories of many kinds.

Overall, digital interactive storytelling is a very effective mean to increase people’s attention toward cultural heritage and to enrich visitors’ experience by allowing them to engage and participate with the surrounding cultural environment. Managers of cultural heritage have to keep track of the changes in order to provide innovative services and to ensure the attractiveness of the cultural venues over time [77].

### 2.2.2 Semantic approaches for storytelling

In this section we address the question of how technologies more or less closely related to the Semantic Web have been used as enablers of storytelling, first in general settings and then in the specific realm of cultural heritage.

Good storytellers frequently adjust their narratives and content to match their audience level and age. They may repeat the same story over and over again, but in slightly different ways according to the listener, achieving the goal of maximizing the level of engagement and retention of the intended messages. This type of content adaptation is actually one foundational form of interactivity, since the audience inadvertently influences the development of the story.

Adaptation can be facilitated by having a semantic representation for the overall structure of the story to be told, for the fragments and the pieces of content that may be put together, for the key messages that must be delivered and how they relate to the aforementioned pieces, etc. Researchers have been working for years on methods and techniques to link stories structures to the scenes descriptions.

Narratives have three layers according to Bal’s layered view [78]:

- The lowest level is called the **fabula**, where the raw chronological events are represented and drawn. When it comes to representing this level, it is necessary to solve the problem of the distribution of the open-world scenario.

- The second level is the **story**, which is a part of the fabula that takes a new structure for a specific purpose. At this level, story **grammars** can be used. Genuine grammars include rules about how story parts are organized for a certain category; those rules determine a solid structure that
must be generated by items of the fabula. A grammar based on the Toulmin model [79], for example, has been implemented to generate rhetorical argumentation for video documentaries.

- The last and highest level is the *narrative*, which is a story with all its semantics; any story can be converted to several narratives such as a monologue, a novel or a multimedia presentation. The ontological model and semantics of the narrative will affect the presentation media. Upper level semantic will depend mainly on the form and may be derived from the story category.

Traditional event models have been traditionally unable to model temporal-spatial artefacts for narratives well. Therefore, specific narrative models have received attention during the recent years. The *Event* ontology —mentioned in Section 2.1.2— allows sub-events to be used to represent relationships between events, with unlimited levels of nesting. It can also be merged with the *Timeline* ontology for signals or temporal objects annotation. The *Stories* ontology is precisely based on *Event* and *Timeline*, its main classes being *Story*, *EventSlots* as well as *Substory* [80].

Another narrative ontology was proposed in [81], containing as core elements *Actant*, *Action*, *Actor*, *Event*, *Quality* and *Object*. However, it lacked related properties to represent the continuity of narratives. Sharing some constructs, a different ontology was proposed in [82], including all the information that a narrative or an event should contain, yet it also didn’t model properly the relationships between sub-stories and sub-events. In order to overcome this problem, the *ABC* ontology [83] put forward a similar model, in which sub-event relationships are represented by using situations as intermediaries, and properties like *follows* and *proceeds* to express structural relationships. Finally, the BBC’s *Storyline* ontology [84] has been used for the organization of news events, representing temporal relationships between *StorylineSlots* by using a handful of different relationships [84].

On the grounds of any of the aforementioned semantic definitions, different types of intelligent agents have been proposed to identify and suggest to people and experts different story paths that could be used to structure an existing story or to propose a new content [85]. For example, Natural Language Generation (NLG) mechanisms can automatically create story event structures, working on the problem of mapping the story events and their semantic representations so that a story can be narrated in a natural language [85]. Recently, researches have focused on creating dialogues based on turn-by-turn approaches, and on scaling up text planners in order to generate larger text compositions [86]. Other techniques focused on providing authoring tools based on simple story trees or underlying plan-based representations [87]. A remarkable approach is that of *algorithmic storytelling*, whose purpose is to generate a story by explicitly linking
queries with a path that contains semantically-linked resources, starting from a set of keywords and given entities [88]. The most widely-used algorithm to find such links between different resources is A*, a graph-based algorithm that represents foundational data in the shape of nodes and edges, and then applies graph traversal strategies aiming to minimize whichever cost metrics when trying to connect any two resources [89]. Other solutions fall within the category of semantic association discovery, which will be the subject of Section 2.3.

**Semantic approaches for storytelling in cultural heritage.** The use of narratives is widespread in cultural heritage in various forms, in an attempt to link the historical artefacts and sites to the real-life experiences of the visitors. Narratives can be created by the venue curators or by the visitors who could contribute their personal narratives and experiences [90]. The use of social media and their popularity allows the collection of the personal visitor narratives and can enrich the cultural content, too [91]. Venue curators usually develop the narratives for the venue visitors and lately there are specific technological tools for the task, like emplotment web-based systems [92]. It is also possible to deliver location-specific stories, through systems that use spatial information, maps and synchronized content [93].

Narratives have been used in different settings and with different technologies (e.g. videos, augmented reality, etc.). It was found that despite the medium, narratives have been a very effective way to increase the level of immersion of the visiting experience and increase the overall quality of the cultural experience [94]. Since narratives have been proven so useful, researchers also developed frameworks that allow the quick and cost effective delivery of narrative applications for cultural heritage, across different end-user devices [95].

In addition, also recognizing the fact that museum visitors often seek an educational as well as a social experience, there have been recent developments that allow social experiences through informative narratives. For example, group-based digital narratives and storytelling have been studied and not only cultural understanding increased but also visitor cooperation was significantly enhanced. Furthermore, there is evidence that interacting with friends can also provide valuable information to enhance the cultural narratives [96][97].

In what relates to the use of Semantic Web artefacts to formally model narratives, researchers in the cultural heritage area have used different ontologies than the ones we mentioned in the preceding sections, which have been applied in broader contexts. Specific ontologies have been proposed that tackle two main goals: the classification of story types and the formal annotation of the narratives. An example can be found in the work presented in [82], that was in turn inspired by Propp [98][99], an ontology for fairy tales encoded in OWL. Given a story plan, the system presented in [100] searches for a similar plot in the ontology, driven by metrics of semantic similarity. Following that, it generates a text version of the resulting plot, adapted to the characters, situations, length
and structure indicated by the user. Likewise, the authors of [101] presented an non-supervised approach to classify plot types. However, these approaches not have gained significant adoption, since it is perceived that they are not flexible enough to meet the expectations of a population that consumes vast amounts of information through the new media [102,103]. The main challenge that researchers face is how to overcome differences between media types and genres. One famous media–independent model is provided by the OntoMedia ontology, which is being used in projects such as Contextus [104].

2.3 Discovery of semantic associations

Association discovery in semantic knowledge bases is a recently-established area of research, facilitated since the consolidation of the adoption of Semantic Web technologies in different fields. Usually, there are four general approaches to find the different relationships among chosen entities:

- **Graph-based techniques** work on finding paths and links between entities using entity-relation graphs, seeking from shorter to longer sequences of chained triples involving the selected entities, either allowing artefacts such as loops or diverging paths in between or not. The most relevant work on the characterization of types of associations was presented by Anyanwu and Seth in [105][106], drawing their motivation from cultural heritage scenarios and using a mathematical notation that is nonetheless beyond the scope of this thesis.

- **Data mining techniques** use different classification and clustering algorithms so as to find hidden relations between data and objects, often supplementing the explicit semantic relationships with statistical data about the relative usage of different terms in written text, plus grammatical rules and other types of information closer to the area of natural language processing. Data mining for crowdsourcing, for example, has been implemented (see [107]) to get sizable amounts of input from humans, mining data primarily out of heterogeneous and unstructured information (e.g. short textual or oral comments) provided by venue visitors.

- **Semantic reasoning engines** can be employed to check the validity of existing knowledge bases and to draw further inferences from the RDF datasets and OWL ontologies. The solutions include the many Description Logics reasoners enumerated in [http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/](http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/) or rule-based reasoning engines like Bossam [108].

- **Pattern mining and machine learning** offer consolidated techniques (unsupervised and, potentially, also supervised) that can be used to search
for unconventional links, based on historic facts which can be tractable by a computer (e.g. characters linked by the same birthplace or events happening concurrently) or informed by the arrangement of data in the repositories (e.g. relationships based on the number of common edges in a semantic network).

• **Exploratory association search** (Explass) deals with complex and undefined information needs, assuming that search systems will enable cognitive processing of the results returned in response to any particular query [109]. Facets and clustering are two common approaches for recognizing this [110]. Facets are usually statically defined, while clustering depends on results. The two approaches are being used widely in web searches and, especially, in entity search [111][112]. The output is usually a list of clusters and facet values, that allow to re-focus and refine the originating query. Each cluster is labelled with an ontological pattern, which gives a conceptual summary of the associations it includes. In order to recommend frequent, informative and small-overlapping patterns and facet values, different approaches may rely on query context and information theory.

One common research question is how to rank the discovered associations as per their relevance with the user-provided queries [113]. Current ranking techniques make use of different structural aspects of an association [105], put in consideration the specificity of the query [114] and give personalized results [113].

The discovery of associations is driven by metrics of **novelty** and **typicality**, which are two of the three criteria that are used to assess whether artefacts generated by machines can be deemed stemming from a genuinely creative process [115]:

• **Novelty** takes into account distance within a conceptual network as well as difference in tags associated with the artefact.

• **Typicality** is ensured as a constraint on the human readability of the associations (e.g. via constraints that events take place on similar time frames or at similar locations).

Feedback mechanisms are generally used to validate the discoveries, helping to discard false, irrelevant or trivial connections. This is important because **value** (the last criterion for attributing creativity) can't be quantified and used as a search criterion: only human evaluation of the associations (via rating, e.g. likes/dislikes or ranking) can be used to estimate the value of the links provided by algorithms. Human feedback can be used, too, to come up with training sets for supervised machine learning in order to improve the criteria of association discovery.
2.3 Discovery of semantic associations

2.3.1 Association discovery tools

Lately, the approaches mentioned above have been implemented in association discovery tools [116]. Outstanding among current efforts, RelFinder [117] utilizes classes of entities and relations that appear in associations to refine the search and filtering processes. RelClus [118], in turn, arranges associations into a hierarchy of clusters for refocusing, where each cluster is labelled with a pattern. Since the graph that visualizes the relationships can still become large, these tools come with interactive features and filtering options in the user interfaces, thus enabling a reduction of displayed nodes and facilitating understanding.

RelFinder. RelFinder (Figure 2.4) works on every RDF knowledge base that provides standardized SPARQL access [119]. It is implemented using Adobe Flex and runs on different web browsers. The user can search for specific terms which get mapped to unique objects of knowledge base. Filters can be used to increase or decrease the number of the appearing relationships and to emphasize over a specific aspect of interest [120].

RelFinder uses a query building process that consists of different SPARQL queries that search for associations between the entities or objects chosen by the user. As the shortest connection or link is unknown, the process searches for connections with incremental length starting from zero. One restriction is that the direction of the property relations with other connection chains can just change for one time. This condition was implemented for performance reasons and because —according to the authors— the continuous change in edge directions are hard to be traced and understood by the users.

The algorithm used has different parameters, it can be set up to remove cycles from the results, so that all objects can occur just once in connecting/linking relationships. Regular expressions patterns can be used to ignore objects and properties in case someone is not interested in specific objects or properties. In addition, structural relations that exist between objects can be taken away, such as if two objects belong to one class or a part of a class hierarchy (e.g. to neglect rdf:type and rdfs:subClassOf properties). It is also possible to define the maximum length of the relationships to be returned.

RelClus. RelClus works exclusively with DBpedia, and it is composed of four main elements: keyword mapping, path finding, relationship clustering and result presentation (Figure 2.5).

In keyword mapping, the user interaction begins with two keywords or phrases that describe two entities whose relationships must be found. RelClus has an auto-complete function —developed based on Apache Lucene— that helps the user to map from keyword phrases to actual semantic entities. Also a drop-down list is available to aid the user to adjust the length limit of the returned relationship. The path finding mechanism works once the search button
Figure 2.4: A screen capture of RelFinder.
2.3 Discovery of semantic associations

Figure 2.5: A screenshot of the RelClus user interface.
is clicked: it starts to find all the paths between the two given entities, and paths are found by using bi-directional breadth-first search (bi-BFS). It operates two searches at the same time from the two given entities, and the returned paths are found when the two intersect or meet in the middle. Finally, RelClus clusters all the relationships found into a hierarchy: at first, the relationships are grouped according to their length; then each group is processed on its own. Finally, the tool presents the results as a hierarchical cluster, either collapsed or expanded tree [121].

2.4 Personalization

Personalization means tailoring something so that it matches the interests, preferences and needs of a specific person or group. It is about delivering content or services to fit individuals, built on knowledge about their preferences and behaviour [122]. The first area of research here was web personalization, i.e. tailoring web resource features such as content, links and the structure of web pages. One common problem in web usage that users can't differentiate between useful and not useful information; personalization aimed to improve the users’ web experience and to allow them to locate the most relevant information [123]. Personalization systems make use of each user's preferences in order to filter the content and show only results that may be interesting to him/her. This idea has been generalized to many domains of application, being at the core of the extensive literature of recommender systems. The hardest part in the process is to understand each user preference and to use it in an effective way for content filtering [122][123].

Several approaches are integrated during the personalization process in order to match users’ preferences as much as possible. They are often classified according to the data they use, the way the data is processed and user interaction [124]. There are two main classes:

- The individual (also known as content-based) approach tries to predict a user’s preferences based on data from earlier interactions only with that specific user. The main problem here occurs when a new user registers as he/she has empty records and the system has no information to match against the available items [124]; likewise, it is common that the recommendations computed by this method become overspecialized around the user’s own preferences, failing to extend the scope to unknown items in order to increase diversity.

- Differently, the collaborative approach depends on data about the user’s neighbourhood, which contains users with similar preferences. A traditional collaborative filtering strategy makes use of ratings from user’s neighbours who have matching or similar ratings for the same items. Items are recommended to the user based on the overall rating of his/her neighbours.
The problem here is just the opposite of the previous technique: it arises when the system adds a new item, as there are no previous ratings for it [124]. For the same reason, it is often difficult to delimit neighbourhoods properly, because when the number of items is large, it is hard to find a set of users who have rated the same ones.

At the same time it is possible to differentiate proactive approaches that try to gather information about the users automatically, whereas reactive approaches required data to be provided by the users themselves [125]. The related literature has plenty of hybrid approaches that mix all of the above to different extents and with different purposes in mind [123][126]. Collaboration via content is one hybrid technique, where the users’ profiles contain item descriptions based on computations of users’ similarity.

### 2.4.1 Semantics-based personalization

Semantic technologies have been used for personalization purposes practically since the advent of the discipline, as they enable insight into the characteristics, meaning, purpose, use, … of the items that may be offered to the users, and also to finely describe the interests, preferences, needs, behaviour and even context of the users.

Semantic expansion approach to content-based filtering, for example, has been used to to improve document recommendation [127], using semantic trees (a kind of taxonomy) to represent existing correlation among domain concepts. This approach applied a semantic network activation spreading model, starting from concepts related to the users’ interests and spreading progressively to concepts of possible appeal. Spreading inside the semantic network is based on relationships between concepts and principles of generalization and specialization. The semantic expansion approach models user’s preferences in more precise way rather than the traditional content-based approaches, but the problems with new users remain unsolved [123].

Another significant approach was based on using FOAF [128] for web personalization. This approach offers an enhancement to content-based filtering based on gathering information from the users’ FOAF profiles, which were popular at the time and, therefore, avoided enforcing the users to rate content items. A critical disadvantage in FOAF, however, was related to privacy, as the profiles can hold private data that should be kept away from unauthorized third parties [128].

A hybrid approach based on semantic reasoning was presented in [129]. Content-based filtering and collaborative filtering were merged with ontological domain modelling technologies, and a personalization technique was applied where the information about items and users was processed altogether in order to find hidden connections between users and contents. The ontology profile concept was proposed to be the best solution for item and user modelling, too [129].
Another relevant semantic approach is one that merges users’ information gathered from social media networks and rating records to enhance the user behaviour prediction [130]. It was found that the social relations between users was highly beneficial in the personalization process, as it is believed that users’ social connections have a direct effect on their actions; indeed, several experiments revealed that when users compare recommendations made by their friends and by recommender systems, users tend to choose those from their friends [130].

2.4.2 Personalization in cultural heritage

Personalization technologies have been applied in cultural heritage for more than a decade now, making stable progress since. A detailed survey was presented by Ardissono et al. in [131], dating back to 2012. The amounts of information in the hands of the cultural venues, the many stories that can be told thereof, and the possibility of retrieving additional contents from online repositories and social networks, altogether yield a landscape in which it is strategically important to deliver information in the way that best matches the interests, preferences, needs and context of each individual user (a student, a museum visitor or a tourist in a city) or each group (a class, a family, a community of interest, a group of friends or the crowd in general). In doing so, the role of Humanities experts as mediators between the knowledge bases and the users’ end devices is also magnified, since they are the key actors in ensuring that the ensembles of contents presented to the audiences are meaningful, historically rigorous and relevant. Accordingly, the experts must be given proper tools not only to develop catalogues and knowledge repositories for cultural heritage venues, but also to curate the interconnections among these and with the aforementioned online repositories. Those tools can also become the medium to systematize the creation of stories as drivers of the users’ experiences.

By 2012, Ardissono et al. highlighted “the lack of common standards of infrastructure, data structure and user model modules, that would ease the implementation and evaluation of novel ideas, allowing researchers to focus on the research and innovation aspects”. Up to then, most projects had had their own strategies for research, using their own data and developing complete technological solutions with little efforts to reuse parts from previous works. However, more recent initiatives (like many of the H2020 projects mentioned in Chapter 1) have slightly changed the scenario, thanks to the progressive adoption of semantic standards for cultural documentation (remarkably, CIDOC-CRM), the availability of architectural frameworks like Apache Spark, and the existence of toolkits and libraries like LensKit [132], LibRec [133] or CARSKit [134], that provide off-the-shelf building blocks for recommender systems: interfaces, algorithms, data models, etc. Nevertheless, the use of these resources is still far from systematic in the cultural heritage domain (not so much in other commercial settings), so the promise is
not yet materialized that standards will ease the implementation and evaluation of novel ideas.

Another trend of the last decade relates to the fact that museums and cultural institutions have been increasingly using online social networks to support the sharing of experiences among visitors and the publication of user-generated material, aiming to establish long-term relationships with people. The trends about group personalization and social networks have indeed intensified since 2012, with numerous variations on the idea of how to build and manage group profiles [135] and how to feed profiles from information posted on Facebook, Twitter and other sites [136]. Recently, a system was presented in [137] that enhances cultural recommendations through social media and Linked Open Data resources, harnessing the new tools provided by graph databases such as Neo4j to bring together user profiles, social relationships, semantic knowledge about exhibits, etc. into a unique formalism.

Traditionally, the delivery mode in personalized cultural heritage experiences has been based on the metaphor of hypermedia with content organized in pages and links, followed by multimedia presentations. In recent years, this approach has expanded largely, thanks to the noticeable increase in the computing, visualization and interaction capabilities of smart mobile devices. Augmented Reality and Virtual Reality have consolidated, and several systems have been presented and deployed for education, exhibition enhancement, exploration, reconstruction and virtual museums [138][139]. Nevertheless, as argued by [140], there is a general lack of support for personalization and communication activities in these “augmented” systems, failing to enable features such as context-aware bookmarking, curation of augmentations for the exhibits, new forms of body interaction for embodiment, etc. Besides, the quality of support for analytical activity is low, as the applications typically fail to show interesting information that is there but cannot be seen by the naked eye. In sum, it may be said that the limiting factor nowadays is not in the devices, but in the human-content interaction aspects.

### 2.5 Summary

In this chapter, we have surveyed the main technologies that have been used in our research. The core concepts of semantics have been illustrated along with the reference standards, including samples of upper-layer ontologies, domain ones, vocabularies and thesauri that have provided essential underpinnings for our work. We then exposed storytelling with its various approaches, discussing how it can enrich the users' experience in many different areas, but with a closer look at the realm of cultural heritage, which was actually the inspiration for several ontologies used in the recent past to model narratives. Finally, the main association discovery and personalization techniques were also presented,
as they are essential technologies for narrative creating and storytelling in the CrossCult project and, thereby, in this doctoral work.

Nowadays, the situation is such that there is a vast amount of information available from different venues and sites in online repositories, built according to semantic standards that are gaining worldwide adoption, and it is essential to provide cultural heritage experiences that involve only reasonable amounts of carefully-selected information. The retrieved contents, in addition, may give foot to tens or hundreds of different stories, which may be targeted to individuals or groups, and again proper personalization and recommendation processes are needed in order to ensure effective delivery of the intended messages. Advanced adaptive storytelling is probably the key ingredient, also, to ensure a meaningful interaction continuum from the moment that a user expresses interest in visiting a cultural or historical place, and after the visit to fully integrate information, observations and personal experience. The persistence of the user profiles could also realize the vision of systems that deliver successive related narratives that link together different cultural heritage experiences –most likely, at different places– in a lifelong chain.
Part II

Thesis contributions
Chapter 3

Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

In this chapter, we describe the development of the semantic models used in the CrossCult project, which support the operation of most of the services and apps of the platform, and how they were used to create a framework for storytelling in cultural heritage. We describe the methodological aspects and the different ways in which narratives can be developed and tailored to different users and contexts, from a theoretical point of view but with real examples derived from our work in the project. Furthermore, we present the development efforts aimed at creating a mobile app that fully implements one of the storytelling strategies, plus one particular case study about historical and cultural associations between the European and the Arabic and Muslim worlds, and the results from preliminary evaluation.

3.1 Introduction

The central motivation for this thesis derived from the general verification that the masses experience history and cultural heritage in ways that reinforce siloed and one-sided interpretations. It was our vision that a more plural, diverse and integrative approach could be achieved by harnessing recent advances in Artificial Intelligence, principally grounded on the technologies we have surveyed in Chapter 2. Formal, explicit and machine-readable specifications of a shared conceptualization (i.e. an ontology) can be employed for delivering proper content to users of ICTs, in ways that encourage curiosity about intercul-
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

tural aspects, promote mutual awareness and understanding, provoke thought to the point of reconsidering biases and misconceptions, and deliver long-lasting learning.

In this chapter, we discuss how the recent advances in knowledge management can improve the pedagogy of History and Cultural Heritage, by helping to apprehend connections among different characters, locations, historical events, cultural or archaeological sites, etc. Through such a discussion, the CrossCult consortium has defined a set of technology-supported storytelling strategies, that have been evaluated through a number of experiments involving different European cultural venues.

In the forthcoming sections, then, we describe the construction of the semantic knowledge base with a focus on the most relevant artefacts for this thesis (Section 3.2) and the storytelling strategies, along with some illustrative examples and summaries of evaluation results that reveal traces of the desired outcomes in terms of fostering reflection about cultural heritage topics, enhancing retention of learnt material, triggering curiosity as a means to attract people into the narratives, … (Section 3.3). Finally, in Section 3.4 we present the design of an application of our own that fully implements one of the storytelling strategies, exemplified with the development of an interactive narrative that connects the European and the Arab world through the lives of the historical characters of Salah ad-din and Richard I of England. An experiment is described, too, that provides evidence that the app attains its goal in relation to the Humanities objective of fostering mutual knowledge and understanding, with a cross-border perspective and notable parallelisms between past and present.

3.2 A semantic knowledge base for reflective Cultural Heritage

The CrossCult Knowledge Base is a comprehensive, standards-based structure of semantic artefacts, created with the aim of easing interoperable connections between multiple sources of information related to History and Cultural Heritage. Its architecture is shown in Fig. 3.1:

- The bottom layer contains four generic schemas used by all the layers above. CIDOC-CRM, as as the most widely adopted semantic standard for cultural documentation, provides the general scaffolding with more than 80% of the definitions of classes and properties, guaranteeing the usage of well-defined and interoperable semantics. SKOS, Dublin Core and FOAF are used to serve their common purposes, that we described in Chapter 2.

- The middle layer contains project-specific semantics that could not be found in existing solutions, including classes and properties needed for
3.2 A semantic knowledge base for reflective Cultural Heritage

Figure 3.1: The CCKB stack of ontological layers.

storytelling and reflection purposes, which essentially come down to more fine-grained definitions of types, terms and appellations than can be achieved with the generic concepts and relationships of CIDOC-CRM. The upper-level ontology (written in OWL) captures common concepts and relationships in the realm of Cultural Heritage, enhancing the capabilities for semantics-based reasoning, linking, augmentation and information retrieval. This is supplemented with CrossCult-specific definitions, as well as with links to concepts from external resources such as DBpedia, YAGO or Europeana, which extend the reasoning capabilities further thanks to their additional sets of classes and properties.

- The box on the right hand side refers to the supplementary vocabulary called the CrossCult Classification Scheme (CCCS), which brings together terminology from popular thesauri like Getty’s Arts and Architecture Thesaurus, the UNESCO Thesaurus, EuroVoc and the vocabulary created by the Library of Congress Subject Authorities (LC).

- The top layer contains the ontologies used to represent the knowledge about cultural venues and users, aiming to describe the collections, purposes and premises of the former, as well as the preferences, interests, needs, capabilities, behaviour, personality, educational background... of the latter. These ontologies have been specified in a way that is compliant with the guidelines of CIDOC-CRM and FOAF.

3.2.1 Population of the CCKB

The process followed to model the data in CrossCult —led by staff from University College London— can be abstracted into three main stages, as shown in Fig. 3.2:

- First of all came the selection of data sources, which involved a range of unstructured text written by staff of the venues collaborating in the
Figure 3.2: Stages of the data modelling process followed in CrossCult to populate the CCKB.

project, or retrieved from any of their documentation archives. The text contained in these sources commonly referred to specific museum exhibits, describing their physical characteristics, location, purpose, historical context, cultural meaning, etc.

Due to the lack of structure, it was necessary to search for semantic entities that might be meaningful and relevant in order to connect the knowledge of the various venues, considered in isolation or not. Thus, using entity recognition tools like DBpedia Spotlight[^1] it was possible to identify mine textual instances, temporal information, location names, related characters or deities, etc.

- The next process involved aligning the data to the constructs provided by the upper-level ontology. This could be done in a semi-automatic fashion, with formats provided by the tables of a custom-made relational database. Data received in the shape of informal sheets in Excel format were added, too, following conversion to comma-separated values for easier processing by command-line scripts.

- The final step was to generate OWL statements to populate the ontology from the relational database. The different tables and fields were mapped to convenient ontology classes, properties and assertions.

[^1]: https://www.dbpedia-spotlight.org/
3.2.2 Modelling topics of reflection

As part of the project-specific elements included in the CCKB, we agreed on introducing the class called *Reflective Topic* to model the subjects of the reflection pursued by a narrative delivered through whichever of the storytelling strategies (to be presented in Section 3.3). Subjects such as “Power structures in society”, “Women and marriage”, “Health and rituals in Antiquity”, and “Colonial powers of the 19th century” can be modelled as instances of this class and annotated with types, appellations and any other semantics enabled by the resources of Fig. 2.2.

Specifically, the class is defined as an extension of *E89 Propositional Object* — defined by CIDOC-CRM to model “items that are ‘about’ something in the sense of a subject, ... such as expressions of psychological value”— linked to any entity in the ontology through the project-specific property “reflects”. For example, Fig. 3.3 shows how a man-man object (specifically, a marble plaque depicting the nine Greek Muses with Apollo, with reference ckb:E22/MT0034 in the CCKB) takes the reflective topic titled “Education Apollo and the Muses” as the primary subject of reflection. This reflective topic is linked to CCCS terms expressed as instances of `skosConcept`, and also appears linked to media objects and a narrative (just a short piece of text).

![Figure 3.3: Semantics of the Reflective Topic class, with the example of rt_0053 Education Apollo and Muses. [Credit: Andreas Vlachidis (University College London), in collaboration during the CrossCult project.]](image-url)
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

A broad reflective topic can be decomposed into more specific ones by means of the P148 has component property. In Fig. 3.3 the topic of “Education Apollo and the Muses” shows a kind of recursive composition that allows a user to articulate a narrative as a sequence, defined by the has first, has next and has last properties.

The reflective narratives, in their most basic form as in Fig. 3.3, are short stories authored by Humanities experts and aimed at contextualizing a topic of reflection with inspiring historical/social facts and relevant associations, complemented with links to digital resources. As such, the narratives revolve around a particular exhibit or a collection thereof, promoting reflection and reinterpretation as intended by the storytelling strategies. The semantics of the CCKB allow interweaving multiple pieces of content in two (non-exclusive) ways:

- First, via manual assignment of CCCS concepts to physical items and reflective topics.
- Second, by an automatic process of entity recognition and resolution. This way, for instance, the reflective narrative cckb:E73/rn_0053 illustrated in Fig. 3.3 was annotated in relation to the DBpedia instances of db:Apollo, db:Sappho, and db:Kithara using DBpedia Spotlight.

3.2.3 Modelling historical events and commemorations

Motivated by some of the goals pursued in this thesis (which will be thoroughly developed in Chapter 4), the goal of modelling special dates such as historical events or periodic commemorations is to enable the discovery of calendar-based associations across one or several collections of cultural heritage items, and also to serve as a source of information for personalization purposes. The special dates follow class and property assertions that enable connections with other ontology classes, such as reflective topics and cultural items.

The reflective narratives are fully interwoven within the CCKB and enjoy a rich set of semantics, which enable connections with other ontology classes, such as reflective topics and cultural items. In order for the instances of the special dates to become fully integrated and to enjoy the same level of semantic interoperability with the rest of the resources, they had to adopt the semantics of the CCKB. In this sense, special dates should follow class and property assertions that would allow them to connect within the semantic environment of the knowledge base.

Accordingly, the special dates captured in the CCKB are attached to instances of E73 Information Object to describe matters—in textual form or, in greater detail, with instances of E89 Propositional Object—that occur periodically or that occurred on specific dates/periods in the past, and which somehow had a relevant effect from some point of view. Thus, dates can be modelled as
instances of the class Event as in Fig. 3.4, where the description is linked to CCCS concepts by properties P67 refers to—just like it is done for the reflective narratives (see Section 3.2.2).

Figure 3.4: Semantics of the date of the first public performance of the opera “Sappho”. [Credit: Andreas Vlachidis (University College London), in collaboration during the CrossCult project.]

Figure 3.4 also represents the E52 Time Span instance that specifies the date of the event in two different ways: with a literal of xsd:dateTime and with an instance of the class E50 Date. The information conveyed by these two artefacts is the same, but it is advisable to implement consolidation processes to enforce a unique way in order to facilitate the semantic reasoning and association discovery processes. In this regard, the implementation of the CCKB includes scripts that turn xsd:dateTime and other specifications of time instants or periods into appellations modelled by E50 Date and related properties.

3.3 Strategies for technology-supported storytelling

The various research and development teams in the CrossCult consortium took all the literature different directions enabled by the CCKB to create a framework of technology-supported strategies for storytelling. As participants in the process, we contributed to the definition process that reshaped—to a significant extent—the visions in the minds of the project proponents back in September 2014. Next, we present the strategies along with some examples and shallow experimentation results. More detailed material will follow in Section 3.4, which

2The textual note on Fig. 3.4 was retrieved from www.onthisday.com.
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

is devoted to our full-fledged experience with creating the narrative bridging the European and the Arab world.

3.3.1 Associations through coinciding or contrary features

Essentially, the CCKB is a graph of labelled nodes and connecting labelled properties. Therefore, as explained in Chapter 2, it is possible to apply different algorithms from the realm of graph theory in order to search for coinciding or contrary values for metadata fields linked to any given set of entities. In other words, discovering associations comes down to encountering semantic paths that take from one of those entities to another — opposing features take values that appear in the knowledge base linked to any semantic properties that express gradable, complementary or relational antonymy.

The shared/opposite values may represent associations that an expert could turn into a reflective narrative, relating to observations that have more or less straightforward readings depending on the number of semantic hops involved. Thus, the coinciding birthplaces of two or more historical characters would be a direct association, whereas the fact that two pictures were created by Polish painters would be a more indirect one. Depending on the audience and the type of cultural experience sought, it is possible to exploit longer semantic paths, as long as the successive properties are meaningful enough; for example, it might be interesting in a given setting to bring together a number of cathedrals in different locations because they were made of stone from, say, the Carrara marble quarries in Italy, or several writers because happened to participate in wars against Russians or Ottomans.

As explained in [141], we have helped implement different combinations of breadth-first and depth-first graph traversal algorithms for association discovery, which give priority to certain sets of semantic properties that we deemed more relevant than others. Figure 3.5 shows the snapshot of one finding displayed in one of the CrossCult apps, namely one association between the Palace of Versailles (France), the Sisi Museum (Austria) and the National Gallery (United Kingdom) through a shared trait of three different historical characters represented in their collections: Prince Louis XVII, Empress Elisabeth and King Henry VIII. When clicking on the shared feature at the centre, a dialogue pops up would ask the user about the common feature. The association was found due to the inclusion of the three characters in the Wikipedia category of “Deaths from tuberculosis”[3]. The missing information sought by the quiz question causes the well-

---

[3]Wikipedia categories (aka “wikicats”) are groups of links that bring together pages on related subjects or concepts. The categories attached to a given article (in turn, a DBpedia entity) appear listed at the bottom of the Wikipedia pages. The main purpose is to help the readers navigate around a certain subject, but clearly the categories also provide semantic classifications for other purposes.
known effect of the information gap, which can work as an element of a narrative to instigate curiosity and facilitate the appraisal of bits of knowledge [142][143].

Figure 3.5: A cross-border association through one common feature of three historical characters.

The associations given by semantic paths can be put together into extensive meshes that bear reflective narratives with a cross-border perspective. We helped to do so in the context of a CrossCult pilot that involved four archaeological venues from the South of Europe. With the aid of Humanities experts from the University of Vigo, the University of Padova and the Greek Ephorate of Antiquities, we highlighted cultural similarities and differences between the four sites before, during and after the Roman Empire. Then, following the process of Section 3.2.1 we turned that knowledge into semantic artefacts for inclusion in the CCKB, and this in turn yielded graphs of associations like that of Fig. 3.6.

These graphs were presented to users of the pilot app to visualize mind maps of the shared/opposing aspects of the four places in relation to selected historical, social and cultural reflective topics. The users were given aids to browse the graphs along with successive bits of a story —the story would be tailored to the specific reality of the closest location, thanks to appellations attached to the Information Objects in the CCKB.

An experiment carried out during the second half of 2018 following a think-aloud protocol [144] revealed that a quiz game linked to the associations can promote self question reflection about the selected topics, and that the connecting of stories serves to trigger cognitive effects and emotional responses, most often related to acts of discovery [145]. In another experiment, having the mesh of associations on the screen, the users could click on certain nodes to display selected items of information, and then steer to other concepts driven by curios-
Figure 3.6: A graph of connections among the four venues of CrossCult’s pilot on cross-border associations.
3.3 Strategies for technology-supported storytelling

ity. Thus, for example, they could learn about the evolution of rituals connected to “eating blessed food” in Christianity, or about other characters listed along with Apollo under the Wikipedia categories “Deities in the Iliad” or “Dragon slayers”.

Taking a different approach, the value of the semantic annotations of the CCKB has been assessed in another CrossCult pilot that took place in the National Gallery, London, which looked at associations among paintings, and artists across European history. The team in charge created a mobile app that allows the users to express their preferences for a tour in the Gallery, using a Tinder-like interface to indicate whether they like paintings from a given carousel. Following the construction of a user profile — kept in a properly-protected part of the CCKB, with the artefacts of the user ontology of Fig. 3.1, the app recommends itineraries that would take the target user to paintings that share features (typically, of theme or style) with the liked ones. This way, the users can reflect on the diverse works based on their own knowledge, choices and experience, rather than following a random route or being forced along a single predetermined path.

Figure 3.7 displays a few screen captures of the mobile app created for the National Gallery, showing the rooms where the selected pictures can be found and snippets of information retrieved according to their semantic characterization in the CCKB.

Figure 3.7: Screen captures of the CrossCult app created for the National Gallery.

In the experiments reported at [145], it was found that 80% of the users agreed with the statement that “Each painting in the recommended group was according to my visit objectives”, whereas 69% agreed with “I think I reached my visit objectives following the app recommendations”, notwithstanding the fact that the recommen-
dation logic left aside the most popular masterpieces in order to distribute visitor flows in the venue and promote the “second row” paintings. Furthermore, it was noticed in some cases that the personalized routes encouraged the users to think about Art from non-classical points of view, causing deep reflection.

3.3.2 Associations through universal themes

Traversing the semantic graph of the CCKB is likely to succeed at finding meaningful associations if the characterization of the selected entities is sufficiently detailed. This was the case for all the paintings owned by the National Gallery due to the intense cataloguing and classification work they conduct on a regular basis. Hundreds of paintings, with tens of semantic appellations each, facilitate the task of finding connecting paths that could be turned into interesting stories. On the contrary, in smaller collections, with shallower and uncurated annotations, and/or relating to locations or periods of History that are poorly represented in Linked Data resources, the outputs from the association discovery algorithms are often limited to a handful of vague or trivial connections. If the metadata is not in English, the outputs are even more reduced, inasmuch as Linked Data resources receive most of their contributions in that language.

This situation was addressed in the CrossCult pilot developed in the Archaeological Museum of Tripoli, which houses more than 7,000 findings from recent excavations in the Greek region of Arcadia, of which only a few tens are in display. Next to each exhibit there is a little information card that includes shallow and unconnected pieces of text.

A storytelling strategy for this environment couldn’t be based on bringing together groups of exhibits according to common or opposing features. However, it could look at their relation to broad, universal themes, such as “Urban development”, “Rituals”, “Daily life”, “Parent-child relationships”, “Human rights”, “Power structures in society”, etc. To this aim, the EuroVoc thesaurus became a particularly valuable resource, as it provides an extensive hierarchical classification of themes in all the EU languages and connected to other Linked Data resources (see Fig. 3.8 for a sample view).

By integrating the EuroVoc thesaurus in the CCKB, it is possible to apply deep learning methods such as paragraph vectors [146][147] to estimate the relatedness of the exhibits from a given collection to any of the themes in the hierarchy, given only their textual descriptions and following a training phase (for example, fed by Wikipedia articles in the corresponding language). Therefrom, it is possible to derive subsets of items that may be connected in crosscutting narratives. The subsets and themes should, nonetheless, be presented to Humanities experts for validation and completion.

Following this approach, the CrossCult team from the University of Pelleponnese —under the supervision of a historian from the University of Vigo— created a number of narratives (in Greek and English) that brought together
2811 migration

internal migration
  NT1 commuting
  RT length of journey [4811]
  NT1 interurban migration
  NT1 intraurban commuting
  NT1 migration from the countryside to the town
  RT concentration of the population [2816]
  RT town-country relationship [2816]
  NT1 nomadism
  RT Roma [2821]
  NT1 rural migration
  RT underpopulation [2816]
  NT1 seasonal migration
  RT seasonal worker [4411]
  RT tourism [2826]

migration
  RT International Organisation for Migration [7621]
  NT1 Community migration
  NT1 diaspora
  RT geographical distribution of the population [2816]
  NT1 emigration
  NT2 brain drain
  NT1 family migration
  RT rights of aliens [1231]
    NT2 child of migrant
    RT cultural difference [2831]
  NT1 female migrant
  NT1 forced migration
  RT freedom of movement [1236]
  RT refugee [2816]
  NT2 transfer of population

Figure 3.8: EuroVoc topics under “Social questions”→“Migration”.

exhibits under different themes, dealing with similarities and differences between the past represented by the items of the Archaeological Museum of Tripoli and the present reality of Western society. The same team created another mobile app (a screen capture is given in Fig. 3.9) that profiles the user through simple questionnaires and social media games, and thereby list the narratives ordered by decreasing level of appeal. Yet, the user is free to browse and select any of the narratives. The app displays the contents associated to each exhibit and/or each narrative fragment, as well as the corresponding locations of on an interactive plan of the museum.

The narrative fragments finish with a question, e.g. “Do you think the picture above does portray women in a disadvantageous position? Or is it liberating them?"
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

Figure 3.9: Screen captures of the CrossCult app created for the Archaeological Museum of Tripoli.

These questions are intended to instigate reflection and emotional responses. In the think-aloud experiments reported in [145], it was confirmed that this type of interactive narratives allow the users to articulate viewpoints and feelings. Furthermore, it was found that the users realized that the museum was worth visiting several times, inasmuch as every new visit could result in an entirely different intellectual journey driven by a new narrative.

3.3.3 Associations through the users’ context

The strategies described in Sections 3.3.1 and Sec. 3.3.2 can render effects of curiosity and serendipity in cultural heritage experiences, owing to the fact that the rich semantic annotations of the exhibits allow the target users to find interesting, unexpected and personalized content. Next, we explain that the same effects can be intensified by exploiting any associations between the venues’ collections and the visitors’ context, meaning any transitional circumstance that affects them.
3.3 Strategies for technology-supported storytelling

For example, on any 18th of January, a football fan could find interest in a reflective narrative regarding sports in history upon receiving a notification on his/her smartphone indicating that “today is Pep Guardiola’s birthday”. Similarly, someone who has recently read articles about the role of women in power could accept a recommendation to open a certain narrative because Margaret of Parma died on January 18th (Isabella of Austria, dead on January 19th, would be close enough, too). In the same line, someone who might be interested in colonialism could be notified because of the celebration of the Flag Days of Honduras and Tunisia on that same day. Despite the subtlety of some of these links, it was worth formalizing this strategy because it has been repeatedly observed in literature that humans retain information more easily when it is presented in connection with unexpected or familiar stuff [148].

During the development of the CrossCult project, we helped with some early experiments with this type of connections, through an app that would send messages to the user with the caption “Did you know that today is <whatever>?”, and references to a handful of selected Wikipedia articles. The app selected materials driven by the initialization of a user profile through a questionnaire on historical/cultural knowledge. Table 3.1 summarizes the appreciations of this approach (selected from five predefined and non-exclusive responses) compared to the same app delivering random articles chosen from the front page of Wikipedia, which is updated every day. It can be seen that exploiting any associations with the users results in greater interest, reflection and novelty, as well as less indifference.

Table 3.1: Summary of the users’ tagging of personalized and random daily highlights.

<table>
<thead>
<tr>
<th></th>
<th>Personalized</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>That’s interesting</td>
<td>24%</td>
<td>18%</td>
</tr>
<tr>
<td>That makes me think</td>
<td>19%</td>
<td>12%</td>
</tr>
<tr>
<td>That’s surprising</td>
<td>27%</td>
<td>15%</td>
</tr>
<tr>
<td>I knew that already</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>Who cares?</td>
<td>6%</td>
<td>21%</td>
</tr>
</tbody>
</table>

The calendar-based associations need not be related to personal preferences, but rather to historical events, which can provide the criteria to sort the lists of narratives presented to the users. For example, the date of January 18th is celebrated by the English to commemorate the arrival of James Cook as the first European to the Hawaiian Islands, which can give foot to narratives on discoveries and sea travel, among others. The approach is valid not only for events that happened at a given moment or period in the past, but also for periodic commemorations, holidays or observances like the aforementioned Flag Days or the “Winnie the Pooh Day”, which —surprisingly enough— can lead to narratives on Friedrich Nietzsche’s philosophy. This idea was fully developed in
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

In the context of this doctoral work, implying contributions to the state-of-the-art in semantic reasoning and association discovery as we thoroughly explain in Chapter 4.

In applications used to present cross-border associations (like the one we developed, described in Section 3.4), dates can be one of the main drivers of the selection of the venues involved. For instance, the days surrounding every 18th of January or every 10th of July may be propitious contexts to deliver narratives connecting such distant venues as the ancient theater of Epidaurus in Greece and the Roman Army Museum of Northumberland in the United Kingdom, knowing that the aforementioned were the birth and death dates of the Roman Emperor Hadrian, who once visited Epidaurus and had a wall built that set the northern limit of Britannia (Hadrian’s wall). Similarly, the 24th of March (“World Tuberculosis Day”) yields every year the opportunity to talk about cultural heritage items related to characters who suffered from that illness, like was the case of Empress Elisabeth of Austria, King Henry VII of England and Prince Louis XVII of France.

### 3.3.4 Associations through trending topics

The fourth storytelling strategy formalized and evaluated in CrossCult relates to exploiting the reasoning enabled by semantics to search for associations between historical/cultural contents and trending topics in the social networks of a particular user. Knowing that he/she has liked or shared news about a given topic can be harnessed as the starting point to recommend a related narrative. Of course, this is even more so if the user has participated in debates about the topic, in which case the pieces of text he/she has provided can drive the selection of contents.

Figure 3.11 shows an example of this idea, obtained in October 2018 for a Spanish user of a CrossCult app who was in the city of Vigo at the time. Those were days of significant fuss in the media about a cyclone that hit the Atlantic coast of Spain. In this context, the trending topic analysis services of CrossCult (presented in detail in [149]) noticed a relation to EuroVoc’s entry “flood” (Figure 3.10), which in turn led to the discovery of an association between three pictures about the story of Noah’s ark from the Musée des Beaux-Arts d’Arras (France), the National Museum of Warsaw (Poland) and Museo Lázaro Galdiano (Spain).

This selection of artworks could be extended with other paintings representing similar scenes owned by other venues, and in particular by venues that got increased relevance due to geographical proximity to the user (which could be measured thanks to the knowledge captured in GeoNames). This way, the fourth picture on Fig. 3.11 comes from the collection of the Museum of Urbano Lugris in A Coruña, not far (around 160 Km) away from Vigo, in the same region of
3.3 Strategies for technology-supported storytelling

Spain. Here, the topics for reflection are implicit, but preliminary experience shows that many users can identify them correctly.

3.3.5 Involving expects and interested audiences

The strategies presented hitherto to develop narratives in relation to historical/cultural associations lend themselves to many potential stories. However, the experience gathered by the CrossCult consortium showed that only a small percentage of the semantic paths that join any two entities of interest serve to deliver worthy messages. Likewise, the classification offered by wikicats contains plenty of mistakes, whereas social networks are known to contain noisy information and to promote fake associations [150]. In consequence, the association discovery processes can’t be left to work unsupervised, and it is necessary to create tools for Humanities experts and/or interested audiences to easily validate the findings that will make it into the CCKB.

Figure 3.12 shows snapshots of one tool we have contributed to implement in the project, which not only supports the enrichment of the knowledge base but also offers interfaces to ease the creation of narratives, like the one we present in Section 3.4.
The involvement of interested audiences of non-experts was the core idea of the last CrossCult pilot, in which a team led by the University of Luxembourg created a mobile app to share and gather stories from the users as they walk around a city. The app took inspiration from developments in the realm of serious games to promote participation. In this line, the app uses story fragments (see Fig. 3.13) to challenge the users to rethink about their local environment and everyday experiences. The points of interes (POIs) displayed on a map direct them to new locations where new fragments can be activated.

The users can rate the stories and provide comments in response to questions provided by experts. Besides, the users can provide new stories at any time. The app is coupled to a web frontend that allows the experts to moderate the inputs, to perform entity recognition on the approved ones, to curate and interrelate the stories, etc. In the end, the user-provided are connected together —and with the expert-provided ones— in the CCKB.

The experiments presented in [145] revealed that location-based stories can promote re-interpretation of the cultural heritage that one usually takes for granted, also increasing the users’ connectedness to the POIs and openness to express/share personal experiences. In doing so, the users created valuable knowledge, in what can be seen as a *crowdsourcing* experience, even if the num-
Figure 3.12: Screen captures of CrossCult tools to discover associations and develop narratives.
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

The number of participants in the experiments was lower than 50. Indeed, we see crowdsourcing as a technology with huge potential, since interested and motivated people can become a major actor towards enriching semantic knowledge bases and constructing collective memories. The state-of-the-art in user profiling and semantic reasoning technologies should be advanced enough nowadays to identify the users whose contributions could be more valuable in relation to any topics.

3.4 An experience with associations between Europe and the Arab world

In the new societies that have emerged due to globalization and the migration patterns of the post-colonial period, history and cultural heritage are often presented in strongly-politicized ways that promote visions of homogeneous nations [151][152][153][154]. The reinforcement of such one-sided structures of collective memory is seen as one of the most pressing contemporary challenges [155]. Accordingly, many scholars and institutions advocate the promotion of “transnational understandings of History, embedding principles of inclusion, inquiry, diversity and freedom of expression” [156][157].

Since the textbooks used at schools and universities have fixed structures, just like museum exhibitions are almost static, and the majority of the mass media nowadays have notable ideological bias, it seems to us that only the Internet offers the flexibility needed to realize a shift to these new models, helping peo-
people develop awareness of their own personal and national identities through an understanding of their own and others’ histories and cultures [158][159][160].

In this line of thinking, we have developed one mobile application that delivers narratives and quiz games focused on the relationships between Europe, on the one hand, and the Arab world, on the other. The goal is to foster reflection and re-interpretation of history and heritage driven by the commonalities and disparities between two notable historical characters, namely Richard I of England and Salah ad-Din, who have been traditionally regarded as heroes by one side and villains by the other because of their roles in the Third Crusade. These two characters help to develop an interactive experience grounded on the three core principles highlighted in:

- Reflecting the diversity of Europe and the Arab world.
- Acknowledging the positive aspects of intercultural encounters.
- Developing a multi-perspective approach to the study of the past.

The following subsection presents the main features of the app, along with a set of snapshots. Then, we summarize the results of preliminary evaluation, conducted by means of questionnaires on a small sample of volunteers from Spain and Egypt.

### 3.4.1 App design and features

Figure 3.14 shows the starting screen of the app, which provides brief descriptions and visual contents about the different locations around the Mediterranean Sea that were relevant to the lives of Richard I and Salah ad-Din, as well as to the historical event of the Third Crusade. These locations were chosen to quickly highlight the diverse cultural landscape of the 12th century and the notable differences to the political maps of the 21st.

The app offers successive rounds of storytelling, gameplay and reflection, each one focused on a different reflective topic: “Life in year 1180/575”, “Political, military and religious leaders”, “Alliances around the Mediterranean”, “Fighting and co-existing”, “811/835 years later”. These topics are weaved together by successive bits of a story (see Fig. 3.15) that serve to confront the points of view of an English peasant and a shepherd from Jerusalem, about their societies, their political leaders, their culture, their faith, their involvement in the Battle of Jaffa, and the aftermath of the Crusade, when neither side was entirely satisfied with the results. The writing was supervised by Humanities experts in order to reinforce the increasingly-recognized capabilities of storytelling to integrate multiple viewpoints in cultural settings.

The gameplay features of the app lie within the exploration of the associations among the selected locations and the historical characters, filling in some questions left blank by the experts.
Figure 3.16 shows a mesh of concepts and connecting properties related to the reflective topic of "Political, military and religious leaders". We developed these associations, again with the aid of Humanities experts, in order to delve into the common and opposing features between Richard I and Salah ad-Din—particularly the former, since we expect most of the users to be surprised by shared personality traits and family-related facts, besides the well-documented respect that the two characters had for one another despite their rivalry.

Likewise, within the reflective topic of "811 or 835 years later", we have developed a symmetry between the history of Richard I of England and Salah ad-Din and the modern-day events of the Gulf Wars, trying to favour retention thanks to associations with the characters of Tony Blair (who, just like Richard I, had
3.4 An experience with associations between Europe and the Arab world

power in England and was educated in Oxford) and Saddam Hussein (a ruler from the Arab world, born in Tikrit, Iraq, just like Salah ad-Din).

When browsing the associations, the user can drag the nodes around, zoom in/out and click on the pale blue nodes to display a gallery of visual contents related to the corresponding characters, locations or events. If he/she clicks on the question nodes (labelled Q1 to Q4 in Fig. 3.16), a pop-up appears including a question title and a set of choices (Figure 3.17). The user must reply correctly to all the questions before the next reflective topic is unlocked.

In relation to each reflective topic, the user can also display a handful of highlights, chosen by Humanities experts according to the messages that we...
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

aim to convey and the reflection we aim to instigate, with a particular focus on acknowledging the positive aspects of intercultural encounters before and after the 12th century. By clicking on each piece of text, the user can again display a gallery of pictures and/or videos related to each fact. The user can give some quick feedback too, by clicking on one of the emojis of Fig. 3.18 to indicate different emotions: “that’s good to know”, “that makes me think”, “that’s surprising”, “I knew that already” and “who cares!”. This feedback is one of the sources of information for the experiments to conduct in the future for the full validation of our proposal. Prior to that, we have resorted to volunteers and questionnaires for the preliminary evaluation of the next section.

Figure 3.17: A multiple-choice question.

Figure 3.18: Seeking quick feedback about reflection points.
3.4.2 Preliminary evaluation

The first version of our app has been assessed in collaboration with 16 volunteers from Spain and Egypt (9+7) through experiments conducted in January and February 2019. The assessment looked at the feedback gathered through the emojis interface of Fig. [3.18] and the replies to a paper-based questionnaire that would take up to 15 min to fill in.

The volunteers tried the app in one only session each, lasting for an average of 35 min. Then, they proceeded to fill in the questionnaire, which contained 31 statements to rate on a Likert scale, plus 8 questions requiring textual answers (about the user experience, the features they liked the most and the least, suggested improvements, etc.). The results yield the following core observations:

- While the users were faced with an average of 13.2 highlights in the app, they rated only 4.7 with one of the emojis, either because they did not read the rest of the highlights or they did not find any of the emojis to be representative enough. From among the replies, 62% of the users would express interest, thinking or surprise; 32% would click on the smirking face, and only 6% would express indifference or lack of interest.

- Nearly 1/4 of the users had negative opinions about the app, and more than 1/3 did not perceive much added value in the graph visualizations of concepts and associations. This datum correlates with their suggestion that it would be better to use the app on a larger screen than offered by a smartphone, and that things would be much easier to understand with a teacher or guide browsing/explaining the contents. In contrast, the features related to the bits of the stories, the multiple-choice questions and the reflection points were appealing to nearly 88% of the users.

- 37% of the users indicated that they ended up with a richer understanding of the happenings of the 12th century, while 56% had a neutral opinion. It is likely that the statement in the questionnaire was too ambitious for the relatively short time spent with the app.

We plan to use the feedback gathered during these experiments to update the implementation of the mobile app and to complete its contents. More extensive evaluation will be conducted once the new version has been finalized, with a broader spectrum of participants and an explicit effort to make the app available to primary and secondary education.

3.5 Conclusions

The context of the CrossCult project provided plenty of opportunities to brainstorm about the possibilities for cultural heritage storytelling enabled by
Contributions to semantic modelling of Cultural Heritage knowledge and technology-supported storytelling

automated reasoning on semantically-annotated collections. We became familiar with the constructs during the specification and construction of the CCKB, and then contributed to the definition of the set of storytelling strategies that shape one of the most important outcomes from the project.

Within that framework, we participated in the development efforts to create a mobile app implementing one of those strategies entirely, and also led the creation of one particular experience for that app that aims to contribute to bridging the gap between Europe and the Arab world through a new pedagogical approach, rooted on the historical and cultural similarities and differences existing among locations, characters, events and any other elements of already-established identities and dominant narratives. Our design integrates different means to instigate curiosity, reflection and re-interpretation from a multi-perspective approach.

Our preliminary evaluation has yielded positive findings that encourage us to keep improving the app, dealing with usability issues and refining the contents. Subsequent studies can focus on investigating the use of the app in educational settings, contributing to a curriculum that enables students to challenge stereotypes and to develop deep understanding of controversial historical issues. On the technological side, we are interested in introducing means to personalize the contents of the app according to the age and the level of historical knowledge of each student, aiming to tailor the complexity of the writing and the difficulty of the quiz game accordingly.

Following this part of the doctoral work, we reinforced the believe that, all over History, common grounds are to be found between human beings, regardless of nationality, religion and any other traits. As illustrated by the lives of Salah ad-din and Richard I —paralleled with the lives of the peasants of the driving story—, we all express the same feelings and emotions, and the human experience is more or less the same in the essential aspects. We all have our unique circumstances, yet they all go in the same frame in what concerns, for example, family relationships, childhood memories, the sadness for losing a close one or the happiness of having a new family member. We are pretty much the same when it comes to our struggles to defend ourselves, our lands or our dignity. The world unfortunate events make all of us anxious and worried, peace unites us all everywhere and any brutal attack in any place cause pain to us all.
Chapter 4

Contributions to the discovery of associations, context awareness and personalization

In this chapter, we present the technical solutions we have created, on top of the knowledge models described in Chapter 3, in order to implement the storytelling strategies that depend on matching the information available about venues and their collections, the target users and their context. Our proposal for this personalized storytelling system is based on graph-traversal and word vector techniques that can be applied to general features of context. However, we will pay specific attention to temporal aspects, linking experiences to any commemorations or personal events that become relevant on a given period. In addition to the technological solution, we present the results of preliminary evaluation carried out with narratives developed in the CrossCult project for the Archaeological Museum of Tripolis, in Greece.

4.1 Introduction

As we explained in Chapters 1 and 2, the evolution of Information and Communication Technologies is progressively transforming the way that the public can appraise Cultural Heritage. The inception of the World Wide Web first made loads of information available online; then, the social revolution of the Web 2.0 brought in opportunities for people to share cultural experiences, and the more recent developments of the Semantic Web are enabling computers to process growing amounts of metadata, e.g. to automatically assemble personalized groups of heritage resources, that can be offered to the target users as multimedia galleries in an app, or tailor-made itineraries to follow in a museum. In this chapter, we build on the state-of-the-art in semantics and Linked Data resources to systematize the representation and exploitation of calendar-based associations between people and cultural heritage items, as a means to instigate curiosity, reflection and serendipitous learning.
Contributions to the discovery of associations, context awareness and personalization

Our approach connects personal calendar events in user profiles to cultural heritage items in a communal knowledge base of one or more cultural heritage sites. The connections are drawn based on coincidences, overlaps or proximity of dates which are relevant to a specific user (e.g. graduation day) and concurrent historical events relevant to the cultural heritage site (e.g. the user’s birthday coincides with the day a painting was first displayed or an earthquake brought about major damage). The connections may be more subtle, drawn from the semantic labels in the knowledge base: for instance, the user is pregnant and some items are directly related to maternity in Antiquity. Whenever these calendar-based associations between a specific user and a site do not exist, connections can be drawn between exhibits and users based on the current date.

Our proposed idea is to identify topics that are worth some attention on a given day: for instance, March 22nd (the World Water Day) would be a relevant date for a museum with several items associated to the concept of water (e.g. wells or rituals for rain). Beyond the current date and location, discovery can look into the future and for cross-border associations: on March 22nd, the framework can discover that in two days (March 24th, the World Tuberculosis Day) there is a possible narrative connecting the lives of notable people who suffered from tuberculosis at some point in their lives, as we saw examples in Chapter 3. In cases where several such connections between the knowledge base and the current date (or dates in the near future) are discovered, they can be sorted based on their relevance to the interests captured in a user’s profile.

Our approach—which can be seen as a technical realization of the storytelling strategy presented in Section 3.3.3, uses the well-defined and rigorous semantics of the CrossCult Knowledge Base, capable of supporting the conceptual and vocabulary needs of connections to temporal dimensions. From among the semantic resources mentioned in Section 3.2, we adopt the Getty Arts and Architecture Thesaurus (AAT) as an appropriate hub vocabulary for accommodating semantic labels in form of vocabulary concepts. Such concepts are used for cross-connecting and linking the user profile to subtle subjects of interest. In addition, the temporal construct requirements of our work are supported by the event-driven model defined by CIDOC-CRM, which adequately captures the semantics behind the most common data structures of the cultural heritage domain.

In the following sections, we present the details of the components of the CrossCult platform (depicted again in Fig. 4.1) that implement our proposal for the management of calendar-based associations, seeking to identify the most relevant narratives for museum visitors. These are the following:

- An “Association Discovery” microservice that can provide connections between the heritage items of any venue registered in the CrossCult platform and the special dates modeled in the CCKB (Section 4.2).
• A “Recommendation” microservice that sorts out a set of narratives according to personal events and interests recorded in the profile of any given visitor (Section 4.3).

4.2 Association discovery

As explained in Section 3.2, each cultural heritage item in the CCKB relates to one or more reflective topics, and through them to reflective narratives, which are enriched with links to CCCS and DBpedia concepts. Special date entries are similarly linked to CCCS and DBpedia concepts, enabling techniques for cross-searching and association discovery via a common layer of semantics. Using SPARQL queries we can identify associations between museum items and special dates, by generating subject-based matches via a common layer of concepts applicable to both.

The SPARQL query below is one sample from a catalogue of association discovery queries we wrote for use in the CrossCult microservices. It exploits DBpedia enrichments (?extVocabularyURI) of the special date descriptions (?infoObject) by connecting them to the narratives modelled in the CCKB, which are linked to cultural heritage items represented by instances of crm:E22_Man-Made_Object. It is a composite query of two separate SELECT clauses, where the results of the inner clause feed the outer SELECT clause. The inner clause retrieves the DBpedia concepts that each special date is associated with. Then the DBpedia concepts of the inner clause are matched against the same DBpedia concepts linked to the range of the museum exhibits in the CCKB.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
Contributions to the discovery of associations, context awareness and personalization

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX crm: <http://erlangen-crm.org/160714/>
PREFIX cckb: <http://kb.crosscult.eu/>
SELECT DISTINCT ?item ?reflective_topic ?extVocabularyURI ?date
WHERE {
  ?reflective_topic crm:P67_4_has_narrative ?narrative.
  ?narrative crm:P3_has_note ?narrative_note.
  ?narrative crm:P67_refers_to ?extVocabularyURI.
  {
    SELECT ?extVocabularyURI ?date
    WHERE {
      ?specialDate crm:P2_has_type <kb.crosscult.eu/skosConcept/1652>.
      ?specialDate crm:P4_has_time-span ?timespan.
      ?timespan crm:P82_at_some_time_within ?date.
      ?infoObject crm:P67_refers_to ?specialDate.
      {
        ?infoObject crm:P67_refers_to ?extVocabularyURI.
        MINUS {?infoObject rdf:type crm:E73_Information}
      }
      ORDER BY ?item
    }
  }
}
}

Figure 4.2 illustrates one association discovered between an exhibit from the Archaeological Museum of Tripoli and a special date. The museum item, that we already mentioned when explaining the semantics of reflective topics in Section 3.2.2, can be used to instigate reflection on the topic of Education, hence it is connected to the reflective topic rt_0053, which is in turn realized by the narrative rn_0053. The narrative tells the story of Apollo and the Muses and how music played an important role in the education of Ancient Greeks, particularly of women. It then moves into highlighting the role of the female poet Sappho in the music education of women in ancient Greece.

It is worth noting that, through the appellation that links to the DBpedia entity db:Sappho, an ancient artefact depicting Apollo and the Muses can be related to a 19th-century tragedy inspired by an ancient Greek female poet. Both ends support and stimulate a reflection about women’s education.

4.3 Personalizing associations

Through the process described in Section 4.2, the number of candidate associations can be overwhelming once the CCKB contains annotations for more than a few hundred special dates. Depending on the terms provided by DBpedia Spotlight, associations can be found between one date (often based on the same historical event) and most of the items. In order to both limit the volume of associations presented to a visitor, and to provide only associations appropriate to
4.3 Personalizing associations

Figure 4.2: Example of association discovery between a museum item and a special date via a DBpedia concept. [Credit: Andreas Vlachidis (University College London), in collaboration during the CrossCult project.]

each individual, a number of steps are taken as described below and visualized in Fig. 4.3.

4.3.1 Important personal events

User profiles in CrossCult can contain data gathered about each visitor in different ways (explicit or implicit), and accumulate the annotations resulting from the use of different apps or web-based questionnaires. One of the profiling features of the apps allows users to provide the dates for personal events that have importance to them, which are modelled by the user ontology of the CCKB (see Section 3.2). Examples of such events are provided in Section 4.4.2, including dates of birth, marriage or graduation. Specifically, the personal events are stored as instances of the xsd:dateTime datatype, similarly to special dates. We do not require users to specify the meaning of each event, but each personal event must have a day, a month and a year.

Each personal event is used to filter associations based on different combinations of the year, month and day information. For the sake of clarity, we assume an example personal event on February 14th, 1976. Initially, associations are sought related to the exact day, month and year (e.g. a historical event on 14 February 1976, such as a US nuclear test at the Nevada Test Site, or the establishment of Foundation Vasarely museum in Aix-en-Provence), then based on
Figure 4.3: Flowchart of the association discovery process, linking museum items and special dates indirectly via a DBpedia concept, and the selection of appropriate associations based on personal events and interests.
4.3 Personalizing associations

the exact day and month (e.g. all associations on 14 February, such as the annual event of “St. Valentine’s day”), then based on the year (e.g. all historical events of 1976), then based on the month (e.g. all annual and historical events on February) and finally based on the day (e.g. all events on any month’s 14th day). The order of the filtering plays a role, as more exact matches (e.g. exact date) will be preferred for showing to the user than broader matches. The broadest match would be any date with the same day, but could still be presented in an appealing way to the user based on the proximity of the upcoming personal event: e.g. on 14 December, the application could show an association with the introductory message: “Your personal event is exactly two months away! Today, . . .”.

4.3.2 Personal interests

The user profiles can also capture personal interests, which in our implementation are chosen from lists of reflective topics or keywords —again, these may be provided explicitly by the users, or learned (without user intervention) by the profiling services of the CrossCult platform by observing their actions in any apps. As explained in Section 3.2.1 reflective topics and keywords are curated by Humanities experts, and they can be used to further filter associations based on their similarity to the CCCS or DBpedia concepts that bring about the associations.

While there is a variety of ways to find which concepts are relevant to any personal interests provided by the user, we have resorted to Word2Vec [161] as the most general and scalable approach. Word2Vec uses artificial neural networks to reconstruct linguistic contexts of words.

The model used in this work relies on a pre-trained Google News corpus \((3 \cdot 10^9\) running words) word vector model (3 million 300-dimension English word vectors). This model can calculate a similarity score (based on the vector distance) between two words, provided that both exist in the corpus. The similarity score is between 0 (no connection) and 1 (the words are identical). Since both concepts and personal interests can consist of multiple words (e.g. “Ancient Greece”), we took the similarity score as the average score of all pairwise similarity scores between all words in either the concept or the personal interest —certainly, this is an aspect that could be sophisticated much further. For any words that are not found in the corpus, the concepts or personal interests that contain them are omitted from the filtering process. The current implementation of Word2Vec automatically downsamples frequent words, thus minimizing the impact of stop words.

Using Word2Vec we compute similarity scores for each pair of association-related concepts and a user’s personal interests. For each association, we consider the closest personal interest to be the one for which the corresponding pair of association-related concept and the user’s interest has the maximum similarity score. We then take two steps to filter associations based on this information:
Contributions to the discovery of associations, context awareness and personalization

(i) **Remove trivial similarities.** If the maximum similarity score of the association is below a specific threshold, then the similarity is considered trivial and the association is removed from the list. This threshold is needed in part because Word2Vec returns non-zero similarity scores between most words in its corpus.

The choice of threshold affects the volume of associations removed; for our tests, we have chosen arbitrarily based on the use cases of Section 4.4. Indicatively, the similarity score of “family” with “families” is 0.55; with “child rearing” it is 0.26; with “nuns” it is 0.16.

Based on these value ranges and following some experimentation regarding the volume of associations removed with different thresholds, we chose to omit all associations where the maximum similarity score between concept and personal interest vector is below 0.2.

(ii) **Find best matching event per date.** After all trivial associations have been removed, one association is chosen for each day-month combination, so that the user is not overwhelmed daily by numerous associations. Instead, on any given day they may receive either one association (the one closest to their personal interests) or none (because no association exists on this date at all, or because this date is not relevant to their personal events, or because the association is only trivially connected to their personal interests).

In this vein, associations found on the same day and month are sorted based on the maximum similarity score between their keyword and the user’s personal interest vector. For each day and month, the association with the highest similarity score is displayed to the user.

4.4 Case study: The Archaeological Museum of Tripoli

We have tested our methods in the context of one of the CrossCult pilot experiments, titled “One venue, non-typical transversal connections”, that takes place in the Archaeological Museum of Tripoli. The starting point for this museum is representative of the current situation of thousands of small and medium-sized cultural venues around Europe, which suffer from very little traffic and whose treasures are unknown to the vast majority of citizens, partly because they fail to deliver even one of the many stories they could tell.

Our hypothesis was that, if those stories were developed and annotated properly—including the management of calendar-based associations we advocate in this chapter—then it would be possible to deliver interesting content linked to events that are meaningful to each visitor. In this line, a team of humanities experts from the CrossCult consortium developed a set of 75 reflective narratives about life in Antiquity involving the heritage items of the museum, using a
controlled vocabulary about appearance, mortality, religion, rituals, goddesses, humans, amazons, nudity, social status, education, daily life, weaving, dowries, food, names, wild animals and healing practices. These narratives were associated (via human curation) to 17 archaeological items displayed physically at the museum. These narratives provided a wealth of opportunities to automatically identify the most interesting stories to offer to any visitor, enabling synthesized views on reflective topics that could hardly be conveyed before.

For our experiment, the CCKB was populated with 60,525 special dates. Most of them were created from the online resource [www.onthisday.com](http://www.onthisday.com), passing the short textual descriptions of each historical event through DBpedia Spotlight, with default settings. In addition, we compiled a list of annual commemorations observed by the United Nations and UNESCO, which convey global signification. We also processed the National Days and Flag Days listed in Wikipedia, which are useful for filtering associations based on the users’ nationality (in general, people are more interested in historical facts involving their own country than in others). These commemorations were annotated manually with AAT and EuroVoc concepts, and any textual descriptions (or even only the title of the commemoration) were passed through DBpedia Spotlight too.

It is worth noting that all the knowledge about dates was pre-computed and stored in the CCKB. Therefore, association discovery could work fast and did not result in any web traffic with external websites.

In the following subsections, we first present the output of the discovery of special dates associations for the museum. Then, we analyze the results provided by the personalization mechanisms for three synthetic profiles. Finally, we present the results of an evaluation poll conducted with a team of experts and a set of potential museum visitors to appraise the wisdom, interest and value of the associations in relation to the intended phenomena of curiosity, reflection and retention.

### 4.4.1 General calendar-based associations for the Archaeological Museum of Tripoli

Performing association discovery, as described in Section 4.2, between the semantic annotations of the reflective narratives created for the Archaeological Museum of Tripoli and the two types of special dates results in an extensive set of matches. In total, 3,856 associations are found, out of which the majority (3,544) is with historical events and 312 are with annual commemorations. This imbalance is hardly surprising, as the volume of historical events is massive (with more than 160 events daily). Similarly, the common concepts found between the narratives and the events are different based on repository: historical events are linked to the set of reflective narratives through 57 DBpedia entities (such as “db:Ancient_Greece”, “db:London” and “db:Track_and_field”) while annual events are linked predominantly through AAT concepts (16 of them, including “fertility”,...
“health” or “women”), which are usually broader. This does not mean that there
is no overlap, however: theatre, slavery, Greece and Cyprus are found as links
in both sources.\[1\] As expected, some links appear far more often than others: in
associated historical events, the most popular links are Greece, London, Athens,
Ancient Greece and slavery, which account for 79% of all associations with his-
torical events. Associations with annual commemorations are mostly found with
the following concepts: men (30%), health (20%), family (16%) and breast (11%). All
the concepts present in associations are shown in Fig. 4.4.

Figure 4.4: All keywords used to find the associations between special dates and
the reflective narratives from the Archaeological Museum of Tripoli (size and
color both denote prominence). [Credit: Antonios Liapis (University of Malta),
in collaboration during the CrossCult project.]

Considering other aspects of the associations discovered, Fig. 4.5 shows how
associations are distributed in terms of the concepts used as links, the distinct
dates (i.e. day-month combinations) and the reflective narratives and museum
items that are linked. Evidently, much of the information comes from historical
events, although annual events have associations with most narratives and mu-
seum items. Notably, there are two narratives where an association is found only
with annual events, but not with historical events:

• One is associated with four different dates on the topics of fertility, family
  (twice), and women.

• The other is associated with men, which links it with every day of Novem-
  ber (due to the “Movember” month, dedicated to men’s health).

\[1\] However, some of these words are formatted differently (“Theater” vs “Theatre”), so manual
verification was needed for this finding.
4.4 Case study: The Archaeological Museum of Tripoli

Due to the latter association, November dominates (at 53%) the associations with annual events; however, every other month except for April and July is also represented with at least one annual event in the associations found. By comparison, almost every day has at least one association due to a historical event.

Considering both historical events and annual commemorations, 360 distinct dates within the year are represented. This finding gives significant leeway in terms of finding events on a specific date important to one person, but also comes with a daunting task of filtering the most relevant personal associations from a vast pool of almost 4,000 candidates.

4.4.2 Filtered associations for three personae

In order to evaluate how the numerous associations can be tailored based on a specific user’s personal events and interests, we will now explain the cases of three personae (synthetic profiles) representing potential visitors to the Archaeological Museum of Tripoli. Each subsection comes with its own analysis of the findings.
4.4.2.1 Persona: Mata

Mata is a 18-year old girl from Tripoli, Greece. She has lived in Tripoli her whole life, and since her parents got divorced she has joined the goth subculture and the preference towards mysticism and the morbid. In a CrossCult app, Mata has provided her birthday (August 31st, 1999), the date her father left the house (February 13th, 2013) and the day she finished her final school year (June 16th, 2017). In terms of interests, Mata chooses: “Veils”, “Cloaks”, “Talismans”, “Mortality”, “Funerary sculpture”, “Cemeteries” from among the choices contained in the CCKB.

Mata provided three personal events, which are used to first filter only associations related to those dates. For her birthday, four associations are found with August 31st; 14 with the year she was born (1999); 446 with the month of August, and 55 with the 31st of any month. Using all three dates in a similar way, 1,280 associations on 119 distinct dates throughout the year are found without taking into account her interests. Unsurprisingly, most of these dates are in August (26%), February (24%) and June (24%).

Given the large number of associations, it is important to filter them further taking into account Mata’s interests. Among them, “Cloaks” and “Talismans” are not found in the Word2Vec database and are thus we ignored them. Using a similarity threshold of 0.2 between Mata’s four remaining interests and the concepts linking dates and reflective narratives, 418 associations remain.

As a final step, associations are filtered by similarity and the most similar to Mata’s interests (based on the closest word among those interests) for each date is chosen. The result is 58 associations, all on distinct dates. The distribution of these associations is shown in Fig. 4.6.

Most of the associations derive from Mata’s interest in veils (67%) which, interestingly enough, appears most often associated with the concept of slavery (the Word2Vec similarity between “veils” and “slavery” is 0.31). This explains why slavery is often prominent among the associations (45%). It is also interesting to note that Mata’s interest for cemeteries is used to filter associations based on the concepts “village”, “archaeology” and “health”.

Mata’s interest in funerary sculpture, however, is only exploited to find one relevant date: 24th of August (7 days before her birthday), the day that on 79 AD “Mt Vesuvius erupts, buries Roman Pompeii and Herculaneum, 15,000 die”. This day is associated with a plaque depicting Apollo and the nine Muses, and is linked due to the term “Pompeii” (in the narrative, the muse Sappho is also shown in an image of a painting from Pompeii). Unsurprisingly, again, of the 58 dates relevant to Mata, most are on the three months of her special dates.
4.4 Case study: The Archaeological Museum of Tripoli

Figure 4.6: Distribution of Mata's final associations (up to one per date).
Contributions to the discovery of associations, context awareness and personalization

Finally, it should be noted that Mata’s associated dates refer to 15 annual events, while many of the historical events date prior to the first millennium A.D. and very few are in the 20th century.

4.4.2.2 Personae: Üter and Irmgard

Üter and Irmgard are a naturist German couple in their 50s, stopping in Tripoli on their way through clubs and beaches from Kefalonia down to Kalamata and then to Corinth. While searching for an afternoon activity on his phone, Üter found a link to the Archaeological Museum of Tripoli and filled in an online form under the title “Let us personalize your visit”. When asked for three relevant dates, he provided his birthday (March 27th, 1960), Irmgard’s (July 18th, 1963) and the date they got married (December 24th, 1980) and as topics of interest he chose “Nudity”, “Marriage” and “Mythology”.

As in the previous example, the couple provided three dates which are used to first filter only associations related to those dates. Based on these dates, a total of 1,319 associations on 132 distinct dates throughout the year are found before taking into account Üter’s and Irmgard’s interests. Most of these dates are in March (32%), the month Üter was born, and December (20%); interestingly, July and August are almost equally represented (despite the former being Irmgard’s birthday), at 13% and 12% respectively. Since the couple has a narrow set of interests, it is expected that with a similarity threshold of 0.2 most of the trivial associations will be removed. Indeed, 354 associations are close enough to the couple’s interests.

As a final step, associations are filtered by similarity and the most similar to the couple’s interests for each date is chosen. The result is 66 associations, on distinct dates. The distribution of these associations is shown in Fig. 4.7.

It should be expected that an Archaeological Museum has more links to the couple’s interest in mythology (58%) than nudity (11%). Interestingly, the concepts used do not refer to “Ancient Greece”, “Greece” or “archaeology” (only two associations are linked to Greece), despite the fact that these topics were generally prominent in the complete pool of associations. “Marriage” is often used to choose associations, but all those associations are based on “slavery” (the two terms have a Word2Vec similarity of 0.35). On the other hand, “nudity” is used to choose associations based on the concepts “bikini”, “men”, “nun” and “water”.

Most of the 66 dates are on a month of someone’s birthday (March, July) and to a lesser extent on the month of their marriage (December). Unlike Mata, the couple’s associations are not often with annual events; moreover, there are more events referring to the 1900s and 2000s. The most recent event, for example, is on the 31st of March (4 days after Irmgard’s birthday) of 2012, when “Fiji floods
4.4 Case study: The Archaeological Museum of Tripoli

Figure 4.7: Distribution of Üter and Irmgard’s final associations (up to one per date).
Contributions to the discovery of associations, context awareness and personalization

kill 2 people and force thousands to be evacuated”. This event is linked through the DBpedia concept “db:force” with a tondo depicting Heracles and Auge from the 3rd century AD, and its narrative on how Heracles forced himself on Auge, the daughter of his host king Aleus of Tegea.

4.4.3 Evaluation by Humanities experts and potential museum visitors

In order to appraise the ability of associations with dates to foster reflection, retention, curiosity and other cognitive phenomena, we asked four experts in Humanities and 81 other users (of a broad age range, and potentially interested in visiting the Archaeological Museum of Tripoli) to tag as many associations as they could from among the sets computed for the personae of Sections 4.4.2.1 and 4.4.2.2. The associations were automatically formulated in a way that presents the date, personal context, museum item, reflective narrative and associated event (see Table 4.1 for an example). The evaluation was conducted in two rounds between March and July 2018, recruiting non-expert users from among students of diverse degrees in the University of Vigo in Spain and the Arab Academy for Science, Technology and Maritime Transport in Egypt. Feedback from expert users was solicited via direct contacts from within the University of Vigo.

To begin with, the participants in the experiment were asked to assign any of the following tags (possibly none, possibly several) to the associations and the linked narratives:

- **Informative**: the association/narrative provides new knowledge, even if it is not closely related knowledge to the museum item.

- **Thought-provoking**: the association/narrative makes me reflect on the association itself.

- **Memorable**: the association/narrative is likely to be remembered.

- **Curious**: the association triggers curiosity to make the narrative attractive.

- **Personal**: the association/narrative is connected to the user’s interests or dates.

- **Funny**: the association/narrative can be perceived in a humorous way.

Clearly, the criteria for success would be to get many associations tagged as “Thought-provoking” (as it relates to reflection), “Memorable” (retention), “Curious” (curiosity). The number of “Informative” tags provides a measure of interest in the associations, whereas “Personal” aimed to assess the value of sorting associations according to personal dates and interests. The number of associations tagged as “Funny” was a secondary aspect.
In addition, the participants had to choose one of the following mutually-exclusive tags for each association:

- **Notable**: the association is close to museum items or its narrative.
- **Indirect**: the association has some sort of connection, but this connection has several degrees of separation.
- **Irrelevant**: the association is circumstantial, uninteresting, misleading (due to incorrect interpretation of the meaning of a term) or unclear.

Finally, the four Humanities experts were asked to indicate whether the associations that they had found to be “notable” or “indirect” could also be described as:

- **Valuable**: the association is worth showing to the museum visitors.
- **Useful**: the association can increase the visibility and/or the understanding of the museum items.
- **Potentially offensive**: the association involves terms that could be offensive to some potential visitors, and should therefore be filtered.

Table 4.1 shows a sample of the association descriptions that were provided for review along with the persona descriptions.

The associations were mixed and distributed randomly. The final tag counts are shown in Table 4.2 and represented in Figs. 4.8 and 4.9.

<table>
<thead>
<tr>
<th>Tag</th>
<th>4 experts</th>
<th>81 potential visitors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informative</td>
<td>16/136 (12%)</td>
<td>105/323 (33%)</td>
</tr>
<tr>
<td>Memorable</td>
<td>17/136 (13%)</td>
<td>75/323 (23%)</td>
</tr>
<tr>
<td>Thought-provoking</td>
<td>19/136 (14%)</td>
<td>81/323 (25%)</td>
</tr>
<tr>
<td>Curious</td>
<td>12/136 (9%)</td>
<td>40/323 (12%)</td>
</tr>
<tr>
<td>Funny</td>
<td>3/136 (2%)</td>
<td>13/323 (4%)</td>
</tr>
<tr>
<td>Personal</td>
<td>18/136 (13%)</td>
<td>35/323 (11%)</td>
</tr>
<tr>
<td>Notable</td>
<td>24/136 (18%)</td>
<td>68/323 (21%)</td>
</tr>
<tr>
<td>Indirect</td>
<td>76/136 (56%)</td>
<td>201/323 (62%)</td>
</tr>
<tr>
<td>Irrelevant</td>
<td>36/136 (26%)</td>
<td>54/323 (17%)</td>
</tr>
<tr>
<td>Valuable</td>
<td>45/136 (33%)</td>
<td>N/A</td>
</tr>
<tr>
<td>Useful</td>
<td>66/136 (49%)</td>
<td>N/A</td>
</tr>
<tr>
<td>Potentially offensive</td>
<td>17/136 (13%)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 4.2: Tag counts for the sample associations and the linked narratives.
Contributions to the discovery of associations, context awareness and personalization

**Association**

Due to your interest in **Women** and **Mythology**, let us tell you about an event that happened **exactly 104 years ago**:

10 March – 1914: Suffragettes in London damage Rokeby’s painting *Venus of Velasquez*.

**Local exhibit**

This date, linked to **Venus**, takes us to this item from our collection:

*Marble tombstone with a representation of a woman and a young athlete (mother and son). The female figure bears a reaching to the feet sleeved chiton and a cloak covering her head and the lower part of her body. The young man is represented nude, in an enface position and holding a strigil in his right hand. The heads of the figures are broken. Originated from Mantinea and dated in the end of the 5th century B.C.*

Height: 0.95m
Width: 0.47m
Location: Room 15, 1st floor.

**Reflective narrative**

On 10 March 1914, the suffragette Mary Richardson walked into the National Gallery and attacked Velázquez’s canvas with a meat cleaver and left 7 slashes on the painting, the most notable between the figure’s shoulders. The incident has come to symbolize a particular perception of feminist attitudes towards the female nude (and, in a sense, a stereotypical image of feminism as a whole).

Do you think the picture above does portray women in a disadvantageous position? Or is it liberating them? What about our marble statue, that depicts women in the opposite way, and that was also vandalized in the past?

Table 4.1: A sample association description computed for the case of Üter and Irmgard.
4.4.3.1 Analysis of results

Based on the total tags provided by Humanities experts and potential visitors, summarized in Table 4.2, several conclusions about the quality and usefulness of our approach can be drawn.

Considering first the mutually-exclusive tags, we observe that of the 136 and 323 associations rated by experts and potential visitors respectively, 26% are deemed irrelevant by experts and 17% by potential visitors. While this is a promising finding, participants also predominantly considered the associations made to be indirect (56% for experts, 62% for potential visitors). Even indirect associations are deemed meaningful, however, since many participants tagged such associations as “informative”, “memorable”, “thought-provoking” and “curious”.

Notably, potential visitors were more prone to use such tags than experts, as they are likely less knowledgeable of the topics discussed in the narratives of the
museum. Since the tools are intended towards attracting potential visitors, this is a positive finding. Even though it is not overwhelming, the presence of the “memorable”, “thought-provoking” and “curious” tags reinforces the intended value of our approach in terms of raising curiosity to deliver more information about cultural heritage in a way that increases reflection and retention of knowledge.

The analysis of co-occurrence of tags, correlations with user data and other measurements such as inter-rater agreement ratio, is left for future work, since the size of the current sample was not sufficient and the associations were randomly distributed among participants.

The predominance of “indirect” associations (compared to “notable” ones) suggests that connecting cultural heritage to dates could also be exploited to promote serendipity, in the sense of learning about valuable or agreeable things not initially sought by the museum visitors.

It is worthwhile to investigate results regarding the tag “personal” in particular, which had a total of 53 occurrences among 369 “notable” or “indirect” associations. We believe that this low number of “personal” tags is partially caused by the limited amount of personal information in the persona profiles, with a few relevant dates and interests from a closed vocabulary. The design of the experiment could also have an influence as the “personal” tag was optional and, ultimately, it is questionable that one might be able to properly evaluate whether something is personal when it was selected for another person (in this case, a synthetic persona). Nevertheless, based on users’ feedback (through tagging and later discussions) we can conclude that many historical events would typically be outside the interests of most visitors, e.g. being only relevant to certain nationalities. For example, events such as “Great Storm of 1987: hurricane force winds hit the South of England killing 23 people” and “Lady Godiva rides naked on horseback through Coventry, to force her husband to lower taxes” could be relevant to someone from Great Britain, whereas a Spaniard would probably be more interested in a less tragic event happening in his/her country, or in characters from local history.

All in all, this suggests that additional personal information (such as country of origin) may be an appropriate additional filter to find matchings with the set of dates from the CCKB.

The experts’ side As explained above, Humanities experts additionally tagged the 100 associations that they considered not to be “irrelevant” as both “useful” (in 60% of cases) and “valuable” (41%).

Their verbal feedback about the associations they found “valuable” and/or “useful” revealed that, even in the cases of “notable” associations, it would be necessary to add one or two sentences to bring all aspects together, so that the associations could be properly understood in the end. In other words, experts could be put in the loop to reinforce the link between dates and narratives. For
example, the historical event from 1959, “1st known radar contact is made with Venus” can be related to mythology, but only after explaining why planets were named after gods. This can be considered in terms of explaining the associations to the user, which was not the main focus of this work.

In relation to the “potentially offensive” tag, the experts indicated that some discovered associations touch on issues that might be regarded as sensitive and controversial. Visitors, depending on their cultural background and beliefs, might feel less comfortable with associations exploring certain historical and social matters. Overall, 17 associations were identified as such, for instance involving headlines about military confrontations (e.g. “Fighting breaks out between Turks and Greeks over dispute islands in Cyprus and 16 are killed”). While controversy has been known to act as a powerful trigger of reflection [162][163], the experts pointed out that museums should probably avoid unnecessary controversy, especially when the association of certain events to the cultural heritage of the place is indirect. Notwithstanding, they argued that the attempt to recommend (push) one or several narratives to visitors is most likely to have a beneficial effect.

Finally, the experts also noticed that the association discovery algorithms can provide them with clues to develop new reflective narratives to enrich a venue's contents. This is a promising direction which deviates from the original goal of showing content directly to users, and incorporates association discovery as part of a curator's workflow of providing appropriate links with popular, commemorative, or indirectly-related events and concepts. The additional step of a curator-driven enrichment can alleviate automatically-discovered (and sometimes irrelevant or potentially offensive) associations, by removing or rephrasing such associations, and then applying the personalization algorithms to select among the curated associations.

4.5 Conclusions

In this chapter, we have presented how calendar-based associations between cultural heritage items (and collections thereof) and historical events or annual commemorations can be discovered automatically, as a means to bring specific attention to some out of the many narratives that may be linked to a venue's collection. We also proposed a way in which those associations/narratives can be prioritized according to information captured in the profiles of any venue visitors, plus an important feature of context: the day it is today.

The proposal aims to make it easier for visitors to grasp the stories that the venue can tell, so that they can promptly decide which itinerary to follow. To the best of our knowledge, there have been no aids for such a decision in state-of-the-art projects about storytelling applied to cultural heritage experiences.
Contributions to the discovery of associations, context awareness and personalization

Our approach relies on mainstream standards for the semantic modelling of cultural heritage information, which we enhanced with a selection of additional resources plus new classes and properties in order to capture the relevant special dates, and to place reflective topics and narratives as the key mediating element in the association discovery and personalization processes. It is worth noting that the construction of a thorough compendium of historical events is an open research problem [164][165], with notable contributions nonetheless in [166][167][168][169][170]. The management of periodic commemorations and their meaning, in contrast, has not been systematized before. In turn, the reasoning features of our system combine an ontological approach and a vector space model which can handle any type of user interests (including, for example, free-form text). Recommender systems in the cultural heritage area (such as those of [171][172][173] have not previously investigated such a vector space model.

In our view, special dates can unfold an unprecedented range of possibilities for personalized and context-aware cultural heritage experiences. Namely, they allow exhibits to be rearranged in countless ways and can promote all the narratives written for a given venue throughout the year, thanks to links to universal topics and intra-venue to cross-border associations.

Our experiment with the Archaeological Museum of Tripoli shows that, in general, there may be a plethora of possibilities to promote reflection on a collection of cultural heritage items. Both the Humanities experts and the potential museum visitors confirmed that the associations and the linked narratives can deliver new information, helping to retain cultural knowledge and inspiring further thought. To a lesser extent, we have noticed that the approach can trigger curiosity, even though some associations are weak, many are indirect, and there is much research to be done on disambiguation, misleading words and relevance thresholds.

Based on the evaluation results and the feedback from experts, we can conclude with some confidence that the recommended narratives are useful for promoting reflection and retention over significant cultural and historical topics. We see this work as a strategic point of action in the CrossCult project, in alignment with its ultimate goal of interconnecting cultural digital resources, physical venues and citizen's opinions in ways that promote reflection on history and cultural heritage.
Part III

Conclusions and future work
Chapter 5

Conclusions and future work

In this chapter we provide a wrap-up with a critical eye on all of the challenges addressed in this doctoral work, as well as the outcomes generated both in the Humanities and the Computer Science domains, rising a question whether they managed to provide solution to some of the current challenges in research, development and innovation in technologies applied to Cultural Heritage. Finally, we will discuss the possible future directions that our research and results can be improved.

5.1 Conclusions

This doctoral work started from the analysis and understanding of the current trends and challenges in the area of technologies applied to the realm of Cultural Heritage. Digital cultural heritage, as a specific area within the broader concept of the Digital Humanities, is now a mature field, in which novel information and communication technologies are used in the service of preserving cultural heritage and facilitating its discovery and valorization by the public. Following the extensive digitization efforts conducted since the mid 1990s, and the many uses of supporting technologies that followed since the 2000s in education and tourism, we found ourselves in an era where state-of-the-art ICT enabled the delivery of personalized content to visitors of cultural heritage sites in a context-aware manner. However, back then in 2015 when we started the literature review, there were plenty of ways in which new developments could redefine the way cultural heritage information can be managed, stored and delivered to the target audiences.

From the viewpoint of the Humanities, the last few years have witnessed new approaches to the study and presentation of culture and history, with much greater emphasis in developing storytelling strategies with which to stimulate reflection, curiosity and even reinterpretation, rather than just presenting aggregations of multimedia contents, selected with greater or littler care. From the point of view of Computer Science, in turn, the trend is determined by the
ever-increasing reasoning and analysis power that stems from innovations such as deep learning, developments in smart interconnected devices, and advances in Human-Computer interaction technologies that open new ways to the delivery of content. Finally, from the point of view of cooperation between Humanities and ICT, the experience in the CrossCult project and the knowledge about other concurrent EU-funded initiatives have served to ascertain that we live in the era in which the former boundaries—that kept experts in one domain isolated from experts in the other—are starting to fall, that bridges are being formed and robust synergies appearing leading to new opportunities and perspectives.

The most appealing works for us in the context of a Doctoral Program in ICTs related to the existence of huge amounts of digitized cultural information in the hands of the cultural heritage venues (many, freely available online), continued with the realization that it is theoretically possible to tell thousands of stories thereof, and ended with the possibility of retrieving additional contents from online repositories (e.g. Wikipedia, Europeana and other Linked Open Data resources) and social networks in increasingly-standardized ways. Altogether, these facts yield an environment in which it is of utmost importance to deliver not just information, but rather content and messages, in the way that best matches the interests of each individual user (a student, a museum visitor or a tourist) or each group (a class, a family, a community of interest, or the crowd in general). In doing so, the role of Humanities experts as mediators between the knowledge bases and the users’ end devices is also magnified: they are the essential actors towards the goal of ensuring that the ensembles of contents offered to the target people are meaningful and historically rigorous, and at the same time serve the purposes of the stakeholders (either public or private institutions).

Aiming to contribute to pushing forward in this field, this thesis focused on the four main topics surveyed in Chapter 2, namely: semantics, storytelling, association discovery and personalization. These topics relate to strategical approaches, technological development and the necessary standardization that underpins everything else if we really want the developed systems to scale, providing solutions to the whole of Europe and beyond. Upon this belief, CrossCult took all the literature different directions and combined them in one framework of technology-supported strategies for storytelling; as explained in Chapter 3, we contributed not only our thoughts in the process, but also aided in the development efforts to create a mobile app implementing one of those strategies, started and led the creation of one particular experience for that app (with the case study about Salah ad-din and Richard the I of England) and, finally, conducted the preliminary evaluation that returned the sought positive findings, as most of the participants reflected about the point that the aspirations, motivations and behaviour of humans remain almost untouched across time and space, though affected by the circumstances surrounding them (but, again, at the societal level, the core dynamics have not changed much over millennia). As a matter of fact, we had a faith this work had a further humanity dimension rather than
just technological interest, helping people to better understand their national identities as well as the pervading links with other cultures.

All of the aforementioned strategies were successfully accomplished and demonstrated in CrossCult, leaning on a common framework for knowledge modelling, to which we contributed by proposing new semantic approaches for narratives as discussed in Chapter 3, too. The CrossCult Knowledge Base (CCKB) became the central resource to apply queries that would return content to feed the development of new narratives, supplemented with external online resources under the umbrella of the Linked Data initiative. We consider our contributions to the CCKB sufficiently validated by the outcomes of the four pilots developed in the project, plus the additional experiments that we have reported here.

Chapter 4 dealt with how the semantic underpinnings of the CCKB could enable sophisticated reasoning about cultural venues and their collections, multimedia contents available through whichever repositories, user profiles and different features of context (with specific attention to temporal information). From among many other possibilities, we focused on the uses of the reasoning for association discovery and personalization, targeting both experts who want to create new narratives and end users who want to discover the most appealing narratives for them. One topic of particular interest was the design, implementation and validation of semantics-based mechanisms that aim to link cultural heritage storytelling to dates (historical events or periodical commemorations), owing to their connections to a collection of items and to the visitors’ interests and meaningful personal dates. The technical solution involved different forms of reasoning (graph traversal and vector space models) over the CCKB constructs, and it was evaluated in a pilot study on the collection of the Archaeological Museum of Triápoli (Greece), taking advantage from the fact that a team of humanities experts had written a set of narratives about its exhibits in the context of one of the CrossCult pilots beforehand.

Through the proposed mechanism, a year-round calendar was crafted so that certain narratives would be more or less relevant on any given day. Expanding on this calendar, personalized recommendations could be made by sorting out the relevant narratives according to personal events and interests recorded in the profiles of the target users. Evaluation of the associations by experts and potential museum visitors showed that the approach can discover relevant connections, while many others that are more incidental can still contribute to the intended cognitive phenomena of raising curiosity, focusing reflection, increasing retention, and even revealing that it might be worth visiting the same venue again to learn about new aspects, following other narratives. Just like the experience of Chapter 3 showed that a properly-designed mobile app can increase the visibility of small and medium-sized museums by highlighting cross-border associations with the ones that the user visits, this experience with semantic reasoning and relevant dates yielded the finding that technology entails a huge
potential to increase the appeal of museums and archaeological sites that are not so well known among both locals and tourists. The collaborating experts strongly believed that the matching algorithms could provide very good clues to implement new reflective narratives that would improve the venues' contents, and to recommend different narratives to visitors that would have highly-appreciated impact.

Finally, we believe that ICTs have enormous potential for enhancing every aspect of people's life nowadays. It has a major impact in shaping their personalities, feeling and thoughts, by channelling much of their access to information. They also shape their work and socialization activities, either with social media or different online applications, and have the potential to become a major vessel for learning and, in particular, for apprehending cultural heritage. In this line, ICTs are managing to add high value for teaching and learning, by increasing the effectiveness of learning and the quality of the materials given. It made the process much more interactive, which totally fit the new generations, for whom the traditional way that history was taught at schools or universities does not work anymore. The lines of research, development and practice started/enhanced by CrossCult and alike projects may have a great impact in the medium term on how these audiences perceive and create opinions about what they have been told.

5.2 Future work

The work presented in this thesis may be continued along several lines, either closer to the realm of Humanities or to different areas of Computer Science. The main possibilities that come to our mind upon finishing this report are as follows:

- First of all, there is ample space to investigate the application of the storytelling strategies put forward in Chapter 3, inasmuch as the resources and duration of the CrossCult project merely allowed testing them in a reduced number of scenarios. A deep study could be conducted, for example, checking the applicability of the strategies we called “Symmetries via common or opposing features” and “Symmetries via universal themes” in relation to the type of cultural venues, considering size, predominant type of artworks, monolithic collections vs wide-ranging ones, predominant visitor profiles, etc. Likewise, it would be interesting to assess the appreciations of experts when using the authoring tools to discover associations and create narratives, measuring the effort required by the different storytelling strategies, for single venues or multiple ones, working in isolation or in groups, etc. The value resulting from that effort should be measured, too, in terms of end-user satisfaction and metrics of venue attendance, repercussion in online social networks, etc.
• The mention of online social networks calls for an analysis regarding the opportunity, cost and effectiveness of creating narratives in relation to trending topics. There may be implicit thresholds regarding those three parameters, as well as many opportunities to systematize the reuse of bits of narratives (plus any of the linked multimedia assets) thanks to the underlying semantic modelling. There are also technical challenges to the identification of the trending topics themselves — which has been a hot research theme for nearly a decade — when a more or less subtle link to cultural heritage is sought.

• Regarding the strategy of “Symmetries via current context”, it is our vision that this thesis has simply begun to scratch the surface of an idea with great potential to yield many new improvements to further improve the contextualization of the cultural experiences, or in other words, to maximize the effectiveness of the narratives. The context dimension provided by the user’s current location has been addressed many times in the past, so the possibilities for improvements may be limited. In contrast, the temporal dimensions provided by dates were almost unexplored before our work, and the same goes for meteorological conditions (which are present in so many daily conversations after all), traffic conditions, the informational context defined by the TV series the user is watching during a given period or the last blog entries he/she has read on the Internet, etc.

• There is likely a lot of work to do in terms of the consideration of new types of information about the users in the computation of recommendations. We have seen that CrossCult manages semantic user profiles with the ability to capture much more than demographical information, lists of topics of interest, levels of knowledge about whichever topics, preferences about crowded spaces or accessibility concerns. Variables such as mood, transient or chronic health conditions, daily routines, beliefs (e.g. religion) and superstitions (e.g. astronomy or numerology), ... may well act as triggers of curiosity if properly captured and processed.

• Association discovery is another field that, in our opinion, remains heavily unexplored. Our proposal to use both graph traversal and vector space models brings in many new possibilities for technical improvements and, probably, foundational innovations. For instance, we would like to evaluate the impact of adding to the CCKB a new knowledge layer with the information about birth and death dates of notable people, another one of sport-related events or musical works, and the largely anecdotal commemorations listed in websites like Days of the Year (www.daysoftheyear.com). We believe these are propitious grounds to come up with new types of associations, that may work well at least with specific audiences (e.g. sports or music fans, procrastinators at work, etc.). Parallely, it sounds interesting to check whether replacing/supplementing the pre-trained Google
News corpus of Word2Vec with a new corpus trained on a collection of documents from the cultural heritage domain (e.g. Europeana) would improve the experts’ perceptions of the association discovery tools.

Of course, we also leave the recommendation of conducting more extensive experiments in order to fully evaluate the ideas presented in this thesis. Calendar-based associations, for example, should be investigated as per their ability to contribute to long-lasting learning about cultural heritage. For this purpose, it might be good to set the experiments in such a way that the same participants are interviewed after a few months, with very specific questions aiming to assess how much they retain from what they said they had learnt. If the CrossCult platform becomes widely adopted in the near future, that would pose a great opportunity to conduct this kind of studies, involving different audiences, venues and contents.

Finally, whereas Google has experimented with the notion of $x$ degrees of separation of items in museum collections based on their visual similarity [165], we find the notion of $x$ degrees of semantic separation far more stimulating. We suggest that, given any pair of cultural items, we can find meaningful links between them with a reasonably small number of steps that go through facts related not only to history or art, but also to popular culture and any available information about the target user’s memories and interests. Experiments reported in [166] provide evidence that user-specific intermediate entities can be used as elements of the path linking two nodes in a semantic network, and we would like to further explore this issue through experiments specifically targeted to the aforementioned aspects: popular culture, user’s memories and interests.
Bibliography


Conclusions and future work

humanities scholarship. First Monday journal, 15(10).
5.2 Future work


5.2 Future work


Conclusions and future work


5.2 Future work


Conclusions and future work

variation in interactive fiction. University of Pennsylvania: Philadelphia (PA), USA.


5.2 Future work


Conclusions and future work


5.2 Future work


[129] Blanco-Fernández, Y., Pazos-Arias, J.J., Gil-Solla, A., Ramos-Cabrer, M.,


5.2 Future work


5.2 Future work


Conclusions and future work
