Enabling the Use of Real World Objects to Improve Learning

Katja Niemann, Martin Wolpers
Fraunhofer Institute for Applied Information Technology FIT
Sankt Augustin, Germany
{katja.niemann, martin.wolpers}@fit.fraunhofer.de

***Abstract*—In the MACE project we aim to support architecture students while searching for the right learning materials by connecting several repositories containing architectural learning objects and by linking these objects beyond repository boundaries. We use real world object representations which serve as connection between learning materials to allow the students the exploration of new and more complete learning paths. In this paper we outline the generation and usage of real world object representations within the MACE system and evaluate our approach.**

Keywords-real world objects; learning objects; repositories

I. INTRODUCTION

Architecture education, specifically in higher education, relies on the usage of existing entities like buildings and projects for inspiration [1] [2]. These entities that we call real world objects provide an insight into what has been done in architecture, what is possible and what has not yet been explored. Therefore, real world objects provide excellent study objects for architecture students, supporting the paradigm of “learning by example” but also by providing examples for theoretical concepts and calculations. Consequently, students need to get access to learning material about the real world objects and related architectural concepts, architects, legislative information, construction, design, etc.

Today, relevant information and learning materials are available in rather distributed repositories that, even worse, are not related with each other. Students searching for learning material need to access all repositories to find the relevant learning material. For example, educational material is scattered over many repositories like the Dynamo repository (http://dynamo.asro.kuleuven.be/dynamovi/) providing information about architectural projects or ICONDA (http://www.iconda.org/) providing access to legislative documents important to building construction and design.

In addition, simple keyword search and result link presentation are not sufficient for architecturally motivated information search. Instead, students need simple and personalized access to vast amounts of architectural information using advanced, visually based discovery oriented mechanisms for access to the learning material [3]. Examples might be image and location based search and classification browsing. Such advanced methods of access require rich information about the learning resources.

Within the European project MACE (Metadata for Architectural Contents in Europe, http://www.mace-project.eu) [4], we enable searching through and finding appropriate learning resources in a more discovery oriented way. By automatically and manually linking related architecture-related learning resources of various non-related repositories with each other, we establish relations among them. These relations enable simple and unified access to architectural learning resources scattered throughout repositories world-wide. Finally, users are able to discover new learning resources that can serve as additional sources of inspiration and help them to reach their learning goals. The relations must have a high precision, as wrong relations impede the learner in finding the right learning paths.

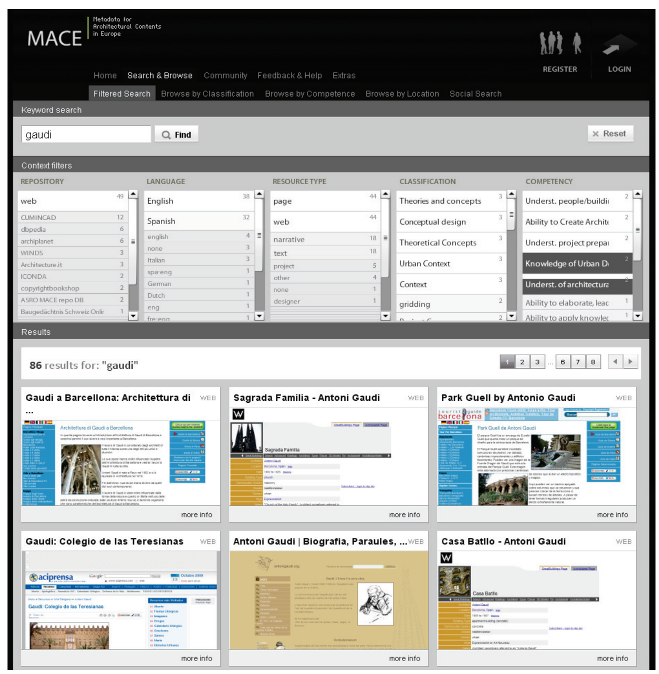

Figure 1. Filtered search interface of the MACE system - searching for learning resources related to the architect "Antonio Gaudi"

MACE also provides the necessary tools to setup and maintain communities around the topic of architecture learning material. The communities in turn contribute by annotating learning resources with tags, comments and ratings, they can build up personal portfolios, and they contribute new learning resources. Exploiting this user-generated metadata to its full extend enables richer descriptions of resources and "social browsing", i.e., new ways to navigate the learning resources.

The "filtered search" interface of the MACE system is shown in Figure 1 as an example of how a user can search for learning resources. The user is able to qualify the search with several additional facets that describe the context of the learning resource(s) in question: the repositories in which to search, the language of the results, the resource media type, the resource classification, and the associated competency. When choosing a respective facet, the interface is dynamically updated by providing the numbers of results for each facet that match the selected criteria.

The results of a search are shown below the context filters. A small overview for each result provides further selection criteria, e.g. the resource title, a short description, and the repository. The user can decide to either immediately go to the result, or to view more metadata about the resource on the respective MACE detail page, shown in Figure 2.

In this paper, we will focus on the creation and usage of real world object representations as one of the main tools for bridging repositories and thus for the provision of a unified view on available related learning resources. In [13] we provide a preliminary description of our overall approach. In this paper, we will elaborate in more detail on the creation, usage, evaluation and the implications of real world objects within the MACE system. Therefore, section 2 will describe the nature of real world object representations in MACE, while section 3 deals with their creation. Section 4 outlines how real world objects link learning resources with each other. Section 5 will present the evaluation of our approach and section 6 concludes the paper.

II. Real World Objects in MACE

Objects in the real world form one of the cornerstones of education in architecture. They can be buildings, building materials, construction plans, projects or architects. Teachers use them to demonstrate architectural concepts, ideas and examples. Students explore them to understand concepts, for inspiration, and also for demonstration purposes. Consequently, objects of the real world must be represented in any learning environment used for architecture education. Therefore, the MACE system includes digital representations of objects in the real world. Real world object representations (RWO) are used to store contextual information of learning resources similar to the context model described in [6]. Within MACE, RWOs are used to tie learning material about the respective RWO to the RWO as well as add further contextual information about the RWO, e.g. its location or building classification. Learning material is thus ordered according to its content automatically, simplifying discovery and access to all relevant information and thus reducing the learner from the cognitive burden of finding appropriate learning material.

The RWO approach relies on the idea that each object in the physical world has exactly one digital representation in MACE, which serves as a reference between reality and the MACE system. The references enable the MACE system to include descriptions of objects of the real world within its virtual realm and therefore provide the bridge between digitally represented architecture learning resources in MACE with objects of the real world.

Within MACE, real world objects are used as bags which collect all related learning resources (that we call media objects as they most often have some type of media directly affiliated with them). For example, the real world object representing the famous church "La Sagrada Familia" references all media objects that deal with related topics.

The nature of the real world object therefore allows us to add specific metadata to each real world object that cannot be added to the related media objects. For example, the geographical location of a building like "La Sagrada Familia" exists only at one specific location, while media learning resources might deal with an architectural concept like "parametric models" which has been used (or is suggested to be used) for the two columns of "La Sagrada Familia". Such a media learning resource cannot have one geographical location but needs to reference all those RWOs that are related to its content. The respective RWOs feature the geographical location if they represent objects that have a fixed geographical location.

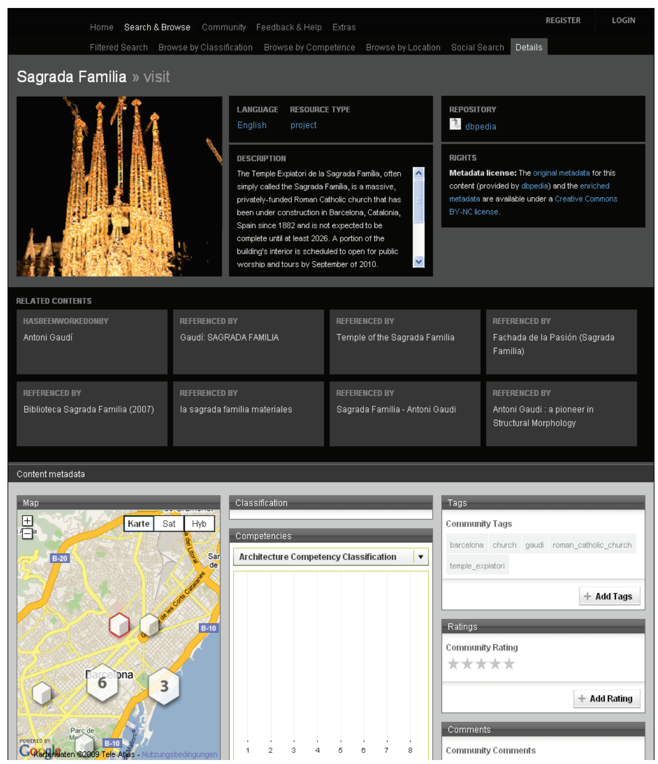

Figure 2. The real world object representation of "La Sagrada Familia"

Finally, RWOs do not only relate to respective learning resources. They also relate to each other as real world objects might feature relations among them. For example, the church "La Sagrada Familia" was built by the architect "Antonio Gaudi". Both, the church and the architect, are real world objects. A relation between their RWOs is the "hasWorkedOn" relation to describe that the architect Antonio Gaudi worked on the construction of the church "La Sagrada Familia". The example is shown in Figure 2 and describes the metadata that forms the digital representation of “La Sagrada Familia”.

With RWOs, users are able to navigate from the real world object to all related aspects, e.g. the architect, the materials used, the legislative documentation, building plans, etc. As the related aspects are described in either other real world objects or media learning resources, users can navigate to them. This simple structure provides the ability to build additional access paths to the learning material.

III. Generating of Real World Object Representations

A. Adaption of the MACE Application Profile for Describing Real World Objects

In MACE the metadata of learning objects, i.e. media objects (digital or non-digital learning objects like an exercise or a figure) and real world objects, is represented using the MACE application profile (http://www.mace-project.eu/index.php?option=com_docman&task=cat_view&gid=58&Itemid=154) which extends the LOMv1.0 standard [7]. Given that real world objects have certain characteristics that distinguish them from media objects, the MACE application profile defines several rules for representing them. The LOMv1.0 general category is used to distinguish between media objects and real world objects. The LOMv1.0 educational category states whether a RWO describes a “designer” or a “project”. We choose the term “project” as generic term for architectural projects like buildings and bridges and the term “designer” for all persons working in the architectural domain like architects and engineers. Additional vocabulary values can be used to describe the real world object depending on its type. For example, the status of a project in the LOMv1.0 lifecycle category is “built”, “demolished”, “rebuilt”, “renovated” or “unbuilt” and the contributors of a project have the roles “architect”, “constructor”, “engineer” or “owner”. Additionally, only representations of digital media objects can have a technical location like a URI where the media object can be accessed. Instead, RWOs describing a project can have geographical coordinates that are stored according to the OGC KML standard [8]. Relationships among real world objects and between real world objects and learning objects are expressed through the LOMv1.0 relation category which is extended with values like “has worked on”, “has collaborated with” or “references”. These relations can be used to express facts describing the relation between the learning objects. For example, an architect has worked on a project (Gustave Eiffel has worked as engineer and architect on the Eiffel Tower), two designers collaborated with each other (Renzo Piano and Richard Rogers worked together on the Centre Georges Pompidou in Paris) or a media object references a real world object (a text about the world famous buildings contains a section about the Piazza del Duomo in Pisa).

B. Repositories for Real World Object Generation

Currently the MACE repository contains RWOs describing designers and projects generated from data offered by DBpedia [9], the UNESCO World Heritage List (http://whc.unesco.org/en/list) and Mimoa (MI MOdern Architecture, http://www.mimoa.eu).

The DBpedia data set is a large multi-domain ontology which consists of RDF triples that have been derived from Wikipedia. Currently, the DBpedia knowledge base describes more than 2.6 million entities and provides different classification schemata, e.g. the Wikipedia categories which are represented using the SKOS vocabulary [10] and the Yago Classification [11]. These classification schemata are used to find the entities that describe designers or projects. For each of these entities containing at least a label, a real world object representation is created and added to the MACE repository. For many entities, DBpedia offers multilingual titles and descriptions as well as geographical coordinates for architectural projects. Furthermore, DBpedia offers several relations like „significant building“ or „significant project“ between designers and projects that can be derived and used to connect real world objects in MACE.

The UNESCO World Heritage List includes 890 items forming part of the cultural and natural heritage which the World Heritage Committee considers as having outstanding universal value. An XML file containing the name, an English description and geographical coordinates for all items is accessible at the UNESCO World Heritage website and is used to generate real world object representations in the MACE repository for every described item.

Mimoa is a free online architecture guide containing about 3000 architectural projects in Europe. Mimoa offers an XML feed that contains the name, an English description and an English textual description of the location for each project. The geographical coordinates can be derived from the textual description and assigned to the real world object representations that are generated using the Mimoa XML feed.

C. Uniqueness of Real World Objects

As we use several sources for the generation of real world objects, we need to deal with the problem of duplicate entries. For example, there is an entry for the Cathedral of Notre-Dame in the UNESCO World Heritage List and in DBpedia. To avoid such duplicates, new RWOs are compared to already existing ones before they are added to the MACE repository. We use the titles and geographical coordinates of the RWOs for this comparison. First, the RWOs that match the geographical location of the new RWO, including all RWOs that don’t hold geographical coordinates yet, are selected. According to the precision of the geographical coordinates (which depends on the respective source repository) are compared on street or city level. To compare the titles of RWOs, the title is tokenized, stopwords removed and compared. If all tokens of the title of the RWO are contained in the title of another one or vice versa, they are considered to be similar. This means that the titles don’t need to match completely and also the ordering of the tokens can be different. Hence, we ensure that related titles are considered to reference the same building e.g. two RWOs with the titles “Centre Pompidou”, resp. “Centre Georges Pompidou” as well as two real world objects representation with the titles “Airport Brussels”, resp. “Brussels Airport”. When the system assumes that a potential new RWO already exists in the MACE repository, its new information is added to the associated RWO. Otherwise it is added as a new RWO.

IV. Connecting Learning Objects

Our approach relies on named entity recognition. We automatically extract names of persons, locations and buildings from the learning resource’s metadata. The person and building names are used to connect learning objects, i.e. media objects and real world objects.

A. Named Entity Recognition

Names of locations and persons in the learning resource’s metadata are identified through a set of modules adapted from the ANNIE information extraction system which consists of a set of IE components included in GATE (General Architecture for Text Engineering) [12]. The Tokenizer splits the text into individual tokens and classifies them into words, numbers, punctuation, symbols and spaces while the Sentence Splitter splits the text into individual sentences. The Part-of-Speech Tagger produces a part-of-speech tag like noun or adjective for each recognized token and the ANNIE Gazetteer tags proper names using over 100 predefined lists of names and keywords. These tags are then used by the Semantic Tagger which annotates the text with new information such as entity types.

Occurrences of building names are identified by a gazetteer that uses a predefined list of building names which comprises all multilingual titles of the real world object representations that are already stored in the MACE repository. If a title contains brackets, it is added twice, one time with and one time without the brackets, e.g. „Lynn Lake (Eldon Lake) Water Aerodrome“ is also added as „Lynn Lake Water Aerodrome“. Currently, the list comprises 66,893 names of buildings.

B. Connecting Learning Objects and Real World Objects

The system performs the named entity recognition process for each learning object representation stored in the MACE repository. When the name of a person or an architectural project is recognized, the system tries to find the associated real world object through comparing the name to all titles of RWOs that represent designers, resp. projects. This process is similar to the process of finding duplicate RWOs. The titles don’t need to match completely and also the order of the tokens is not considered, but all tokens of one title devoid of stopwords and abbreviations need to be contained in the other title. This means, that a learning object containing “Mr. Eiffel” in its textual description is assumed to reference “Gustave Eiffel” since “Mr.” is an abbreviation and is not considered. All titles of a RWO irrespective of their language are considered to ensure that the associated real world object is found independent of the language of the reference.

When an associated RWO is found, the system adds a relation to the RWO that states that it is referenced by the learning object. When no corresponding real world object is found, the person resp. building name is neglected. A textual reference can also be ambiguous and more than one RWO of the MACE repository matches the found name. For example, the name „Millennium Tower“ can refer to several real world objects, as there are RWOs for the Millennium Towers in Amsterdam, Dubai, London and Tokyo stored in the MACE repository. If a reference is ambiguous, a disambiguation is needed to find the associated real world object. For the disambiguation process, the system searches through the matching RWOs to find further entities, i.e. person, building and location names found in the actual learning object. If only one RWO contains at least one of these entities or if only one RWO contains all of these entities, the systems assumes that this real world object is referenced, otherwise the reference is neglected. For example, if a learning object’s textual description contains the project name “Millennium Tower” and the location name “Tokyo” the system will assume that the Millennium Tower in Tokyo is referenced.

V. Evaluation

Currently, the MACE repository holds 52022 RWOs from which 3366 represent designers and 48672 represent architectural projects.

The RWOs are interconnected with 3646 relations that are derived from DBpedia. Furthermore, they are connected to learning objects of 6 different repositories (about 24.000 learning objects) with 4598 relations that were automatically assigned. For 16% of the learning objects at least one reference to a RWO that is stored in the MACE repository was found.

When connecting real world objects we aim for a high precision because faulty connections impede the learners in finding the right learning paths. Therefore, we apply two strategies. We start with the named entity recognition process where we mainly aim for a high recall so to not loose possibly valuable information. Thereafter, we use the extracted entities to find the representations of the referenced real world objects to connect them to the learning objects. In this step we mainly aim for a high precision, hence we don’t use references that cannot be clearly disambiguated.

For our evaluation, we considered the 100 most used learning objects. Each learning object was assigned with tags that represent the referenced designers and projects. The tagging process was performed by three experts of the architectural domain and only tags that were assigned by at least two of the experts were considered for the evaluation to ensure the quality of the tags. 67 of the 100 learning objects hold at least one tag after this process.

A. Completeness of the MACE repository

We start with evaluating the completeness of the MACE repository. We check for how many of the projects and designers mentioned in the learning object’s metadata the MACE repository holds RWOs.

The considered learning objects reference 73 designers and 61 architectural projects. The MACE repository holds RWOs for 40 of the designers (54.8%) and 21 of the architectural projects (34.4%).

Referenced real world objects that don't have a representation in the MACE repository yet are e.g. metro or bus stations, designs of not yet built constructions as well as only locally known buildings or architects.

We will integrate further repositories to increase the completeness of the MACE repository. Additionally, we plan to allow logged in users to add representations for real world objects that seem important to them.

B. Connecting Learning Objects and Real World Objects

To evaluate the relations between learning objects and real world objects, we consider the completeness and the correctness of these relations.

We reach a precision of 90.6% and a recall of 72.5% for the identification of real world objects that represent designers and that are referenced by the considered learning objects. The precision is quite encouraging due to the fact that in the second step, whereas the referenced real world objects are identified, most of the wrongly identified person names are filtered out. Not all person names are found by the named entity recognition process since the metadata sometimes consists only of single keywords. We plan to extend this process with a gazetteer containing all names of designers for which a representation is stored in the MACE repository similar to the gazetteer for building names to further increase the recall.

For the identification of real world objects that represent architectural projects we achieved a precision of 92.9% and a recall of 61.9%. Again, the precision is high due to the filtering process. Several buildings are not found by the gazetteer as they were referred to using abbreviations (e.g. "HUHKA" instead of the full name "Het Museum van Hedendaagse Kunst"), imprecise descriptions (e.g. "Schlikker'sche villa" instead of the official name "Villa Schlikker") or partial names (e.g. "Guggenheim" instead of "Guggenheim Museum").

C. Uniqueness of Real World Object Representations in the MACE Repository

Finally, we evaluate the uniqueness of RWOs in the MACE repository by testing the performance of the RWO merging algorithm. We used 100 randomly chosen entries from the UNESCO World Heritage List to evaluate our approach. For 14 of these 100 objects the MACE repository already holds a representation. While integrating the objects, we reached a precision of 85.7% and a recall of 57.1%. Due to the fact that we do not only use the titles but also the geographical coordinates when matching real world objects, the precision shows that in our experiment only one pair RWOs were wrongly merged. However, we aim for a higher precision therefore we plan to make use of location names contained in the descriptions, when no geographical coordinates are given. To avoid duplicates, we plan to weight the geographical coordinates more if available while slacken the comparison of the titles.

VI. Conclusion

We present the generation and usage of RWOs within the MACE system in this paper. RWOs provide the connection between related learning materials that are used in higher education in architecture. They offer the possibility to explore new learning paths while browsing through the MACE portal. We describe how our approach generates the representations automatically while ensuring that there is a representation for each object of the real world mentioned in the learning resource descriptions that are included in MACE. The overall evaluation of the MACE system in respect to learning improvements is given in [5]. Here, we show which technology base is needed to enable the real world object experience in MACE and its successful application. We plan to further investigate the identification of referenced real world objects by extending the lists used by the gazetteers. Furthermore, we plan to include more repositories and to offer users the possibility to add RWOs themselves to complete the MACE repository.

References

[1] J. Beckmann, Virtual Dimension: Architecture, representation, and crash culture. Princeton Architectural Press, 1998.

[2] M. Condotta, and I. D. Ponte, "Digipolazione Architettonica, nuovi software convertiti". Master's thesis, Universita IUAV di Venezia, 2002.

[3] G. Marchionini, "Exploratory search: from finding to understanding". Communication ACM, ACM Press, 2006, 49, pp. 41-46.

[4] M. Stefaner, E. Dalla Vecchia, M. Condotta, M. Wolpers, M. Specht, S. Apelt, and E. Duval, "MACE – enriching architectural learning objects for experience multiplication". Proc. 2nd European Conference on Technology Enhanced Learning (EC-TEL 2007), Crete , Greece , September 2007, in Creating New Learning Experiences on a Global Scale, LNCS, Vol. 4753.

[5] M. Wolpers, M. Memmel, and A. Giretti, "Metadata in architecture education - first evaluation results of the MACE system". In Proc. of EC-TEL 2009 - Learning in the Synergy of Multiple Disciplines. Fourth European Conference on Technology Enhanced Learning, volume 5794 of Lecture Notes in Computer Science, pp. 112-126.

[6] A. Zimmermann, A. Lorenz, and R. Oppermann, „An operational definition of context". In Proc. of 6th International and Interdisciplinary Conference on Context, volume 4635 of Lecture Notes in Computer Science, 2007, pp. 558-571.

[7] Institute of Electrical and Electronics Engineers Learning Technology Standards Committee (2002). "IEEE standard for learning object metadata". IEEE standard 1484.12.1.

[8] T. Wilson, *OGC® KML*. http://www.opengeospatial.org/standards/kml. 2008.

[9] C. Bizer, J. Lehmann, G. Kobilaro, S. Auer, C. Becker, R. Cyganiak, and S. Hellman, "DBpedia – a crystallization point for the web of data". Journal of Web Semantics: Science, Services and Agents on the World Wide Web, Issue 7, Pages 154–165, 2009.

[10] A. Miles, and S. Bechhofer, "SKOS Simple Knowledge Organization System Reference". http://www.w3.org/TR/skos-reference/. 2009.

[11] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge". WWW, 2007.

[12] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, "GATE: a framework and graphical development environment for robust NLP tools and applications". Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). Philadelphia, July 2002.

[13] M. Wolpers, K. Niemann, and C. Prause, "Metadata representation of real-world objects for architectural education". In Proc. of 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009, July 15-17, 2009, Riga, Latvia