SEMERGY: semantic web technology support for comprehensive building design assessment

A. Mahdavi, U. Pont, F. Shayeganfar, N. Ghiassi
Department of Building Physics and Building Ecology, Vienna University of Technology, Vienna, Austria

A. Anjomshoaa, S. Fenz, J. Heurix, T. Neubauer, A M. Tjoa
Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria

ABSTRACT: This paper explores developmental opportunities toward effective evaluation environments for comparative assessment of alternative building design and retrofit options. Such options are to be benchmarked in view of their functional, ecological, and economical performance. Thereby, multiple use cases are to be accommodated, involving a) novice designers, b) experienced professionals, and c) area municipalities. Toward this end, the rather ill-structured nature of available design-relevant information needs to be taken into account. In this context, the web offers extensive amount of potentially useful information. This web-based potential remains, however, mostly unexploited, as its extraction is hampered by lack of sufficient structure in the encapsulation and presentation of the information. The SEMERGY project adapts the semantic web technology as a promising opportunity to improve and expedite the process of information acquisition and collation toward populating design analysis models.

1 INTRODUCTION

Applications for evaluation of alternative building design and retrofit options require a large amount of information (Mahdavi and El-Bellahy 2005). Instances of such information are building components’ cost and technical properties, relevant micro-climatic data, applicable codes and standards, as well as available financing and subsidy opportunities. Conventional methods toward collating such information are cumbersome, time-consuming, and error-prone (Pont et al. 2011). This circumstance can deter building professionals from in-depth exploration of the aforementioned design and retrofit options in view of their relative functional, economical, and ecological advantages and disadvantages. Consequently, the decision-making quality and the subsequent performance level of buildings could be negatively affected. Hence, efforts are necessary to support efficient information search and collation processes toward populating building models that would fit into performance evaluation routines and applications.

Previous efforts relevant to this problem – such as those related to IFC and IFD frameworks (Buildingsmart 2012) – have partially improved the circumstances. However, to make further progress in this area, the rather ill-structured nature of available design-relevant information needs to be taken into account. Specifically, the web undoubtedly contains extensive amount of potentially useful information. This web-based potential remains, however, mostly unexploited, as its extraction is hampered by lack of sufficient structure in the encapsulation and presentation of the information.

In this context, semantic web technology (Berners-Lee et al. 2001, Shayeganfar et al. 2008) represents a promising opportunity to improve and expedite the process of information acquisition and collation toward population of design analysis models. The present research represents an effort in this direction.

2 OBJECTIVES

The primary objective of our research is to explore developmental opportunities toward effective evaluation environments for comparative assessment of alternative building design and retrofit options. Such options are to be benchmarked in view of their functional, ecological, and economical performance. Thereby, multiple use cases are to be accommodated, involving a) novice designers (as well as laymen), b) experienced professionals, and c) area municipalities (or other local authorities). Accordingly, communication of design intention is to be supported both via simple web-based interfaces and high-resolution building informational models. Likewise, computational routines for option assessment shall involve both more simple (e.g. code-based and normative) procedures as well as advanced numeric analysis and simulation applications. The key feature
of the proposed environment is the incorporation of semantic web technology toward efficient search for and compilation of input information required for comprehensive analysis and evaluation of candidate design options supported by multi-objective decision support methods.

3 BACKGROUND & RELATED WORK

Currently there are a number of existing online platforms that support the energy-efficient building planning and refurbishment. Examples of such online services are the Energy Globe platform (Energyglobe 2012) and the energy-saving calculator of "Raiffeisen Bank" (Raiffeisen 2012) which are meant to support the optimization. These tools facilitate the input of a simplified building model, which includes building geometry and building components (walls, windows, heating, ventilation, etc.). These tools can accommodate simple use-cases, but limit the end-user to a predefined set of options, i.e., a list of generic templates.

The main problem of such systems is the missing link between users' simplified component representations (e.g., "external wall", "window") versus the complexity of specifications of real world products. In other words, it remains the task of end-users, to map such simple notions of building components to appropriate real-world products that meet calculation procedures' informational requirements. Accomplishing this task for the public user is a major barrier to the widespread and effective use of such tools.

The SEMERGY system is planning to bridge this gap by providing semantic links between real world products and building model's abstract concepts and elements. Section 5 of the present paper addresses this approach in more detail (Semantic Web concepts, Linked Open Data).

Another challenge pertains to capturing buildings' geometry as models' input information. The aforementioned web-based tools offer only a very limited set of possibilities for the definition of building geometry. To address this problem, SEMERGY provides two different input modalities for novice and expert users (see section 4).

4 SYSTEM DESIGN

The general design of the proposed building analysis and evaluation environment is schematically depicted in Figure 1. The overall structure of the use cases is as follows:

User provides initial information regarding the intended building activity (erection of a new building, additions or modifications to an existing building, etc.). This initial information contains geometry, building components and their properties, together with additional background information concerning available budget and/or desirable or intended performance objectives. Note that the level of detail and resolution of this input information depends on the previously mentioned type of user (novice versus expert).

![Figure 1. SEMERGY system design](image-url)
SEMERGY considers a number of semantic (non-geometric) permutations of this initial information. This means that next to the initial case, multiple alternatives are generated based on user-specific constraints regarding (i) energy efficiency, (ii) cost, and (iii) sustainability. Thereby, material, element, and component alternatives are considered. For example, different variants for windows, external wall systems, and roof constructions are combined in multiple ways to arrive at a larger set of possible design alternatives.

To accomplish this task, the SEMERGY system deploys two main strategies. First, using semantic web-technologies, information regarding building materials, elements, and components are obtained from the web environment. Likewise, additional relevant and necessary information concerning product prices, microclimatic boundary conditions, applicable legal and administrative constraints, and available subsidies and tax incentives are extracted from the web environment. This information allows filtering the possibilities in view of appropriate materials, elements, and components. Thus, the corpus of possible permutations of the initial design could be efficiently reduced to a computationally reasonable size. Once the ordered set of feasible alternatives is constructed, it is made subject to a comprehensive evaluation process. Thereby, the SEMERGY system refers again to the web environment to identify available computational tools and required resources and fitting for the evaluation task. Such tools and resources are, for example, calculation routines (e.g., those necessary to generate energy certificates for building projects or perform life-cycle analyses) and simulation applications (e.g., thermal and lighting performance simulation tools).

Upon completion of the assessment of the alternative designs via the aforementioned tools and applications, a ranking can be generated and – using appropriate visualization tools – presented to the user. Criteria for ranking could be based on default (locally applicable) benchmarking systems, or selected and weighted by the users.

In summary, the SEMERGY system has links for: (i) user interaction, (ii) applications, and (iii) sources of information.

The user interaction link involves both simple web-based templates for efficient communication of design intention as well as advanced building information models. The application link supports data exchange between the system and multiple analysis tools pertaining to energy calculation, life-cycle analysis, financial payback assessment, and optimization. The information link, which is the critical ingredient of the proposed architecture, is supported by semantic web technology.

To obtain semantic information in SEMERGY, a two-fold strategy is deployed. To identify the kinds of information required for the analyses, a reverse analysis approach is applied to the relevant applications (e.g., calculation tools, simulation routines). On the other side, web-based sources of potentially useful information (building products data, applicable normative documents, procedural and financial information, etc.) are studied in view of existing information representation practices and structures.

5 IDENTIFICATION OF THE REQUIRED INFORMATION FOR ANALYSES

The aforementioned SEMERGY use cases differ in view of user expertise degree and the intended resolution of the outcome. However, the analysis procedures are similar for all cases.

According to the use case, the data required for computation is to be provided by the user through a web-based Graphical User Interface, extracted automatically from an advanced building information model (CAD, BIM), or derived from GIS data.

SEMERGY is intended to incorporate both a simple normative procedure (the energy certificate calculation based on Austrian Standards and Guidelines (OIB 2011, OENORM 2011) and an advanced simulation engine (Energy plus) to arrive at the value of the pertinent performance indicators. Thereby, the informational requirements of the more detailed simulation engine are assumed to also cover those of the normative procedure.

To arrive at a working model of the input data required for SEMERGY, we have done the reverse-analysis of both a normative procedure and a detailed simulation application. As mentioned above, SEMERGY is mainly intended to be a decision support and optimization tool. The more complex and detailed the analyzed model is, the higher the number of variables in the optimization procedure. To avoid extremely high number of input data variations, some input variables of the simulation application (in this case the EnergyPlus IDF) will assume system-assigned default value (Energyplus 2011). The reverse-analysis process (together with the assumptions pertaining to variables that could accept default values) resulted in an appropriate building model (the SEMERGY Building Model or SBM), which includes the necessary data to run the simulation but excludes non-vital details or variables. The SEMERGY building model (SBM) is not identical to the IFC scheme, but it can be systematically mapped to it. SBM defines the building through its physical properties, surroundings and thermal specifications, as illustrated in figure 2.
The SEMERGY building model is sufficiently detailed to suit simulation purposes but also includes sets of derived values required for normative calculations (e.g., U-Values and surface areas instead of detailed material specifications and vertex coordinates).

SBM specifies the kinds of input data required for the initial performance analysis and subsequent optimization. Thus, properties of web-based design intent communication tools as well as data import requirements from BIM environments are clearly defined. Moreover, the SBM provides the basis for the definition of ontologies that are needed to search the web environment for building product information and populate the simulation model with required semantic information.

6 APPLICATION OF SEMANTIC WEB TECHNOLOGIES TOWARDS POPULATING THE ANALYSIS MODEL

The Architecture, Engineering and Construction (AEC) industry is composed of multiple knowledge domains. Sharing and exchanging knowledge is the key factor to success in such a collaborative environment. However, the distributed nature of AEC information has lead to knowledge gaps between AEC related domains (Shayeganfar 2008). Such information is often either trapped in customized data structures of software vendors or locked in domain-specific databases which makes data reuse very difficult if not impossible.

More recently, World Wide Web has facilitated data publishing and information sharing. The AEC industries have used this opportunity to make relevant data such as building products data, applicable normative documents, procedural and financial information, etc. available on the web. These web sources are commonly used by architects to facilitate decision making in different AEC use-cases. Such information resources are easily accessible and contain helpful data, but they necessitate heavy user interaction with distributed resources and data integration overhead. The main reason for this is the fact that the underlying language of World Wide Web is basically designed for data presentation, styling, and rendering and not for machine-to-machine interactions.

In this context, Semantic Web aims to amend the existing Web with Web of Data, where machines are able to "comprehend" data in order to facilitate logical inferences. Semantic Web thus targets machine processing and services automation on a global scale (Berners-Lee et al. 2006).

The gap between required and available AEC data sets is hypothesized to be bridgeable based on two main pillars: First, a set of compact and versatile ontologies should be created that serve as a shared standard vocabulary of AEC concepts. Secondly, the scattered information resources on the web should be mapped to these ontologies and linked with other data sources. Fortunately, recent advances in BIMs have introduced mature data models such as IFC that capture detailed information from building industry knowledge domains. Such models facilitate the specification of required conceptualization for AEC ontologies.

As such, the scattered information on the web may form a global data graph that connects distributed resources and facilitates the discovery of new resources. This approach, which follows a set of best practices for publishing and connecting structured data on the web is known as "Linked Data" (LinkedData 2011) and has gained momentum in the last few years. Linked Data provides a publishing paradigm in which not only documents but also data can be a first class citizen of the Web, thereby enabling the extension of the Web with a global data space based on open standards - the Web of Data (Heath 2011). Depending on the published data formats and readiness of data providers, the Linked Open Data (LOD) may be modeled, published, and reused in different ways. In this regard, the five-star schema of the LOD has been introduced to score the quality of shared LOD (Berners-Lee 2009). The increasing quality levels of this schema are defined as follows:

- Data availability on the web with an open license.
- Data availability as machine-readable structured data.
- Data availability in non-proprietary formats such as CSV, XML, etc.
- Using W3C’s open standards (RDF and SPARQL) to identify entities.
- Linking data to other people’s data to provide data context.

The SEMERGY project aims to reach the top level of this schema by establishing a solid basis for mapping and reuse of AEC-related data. In this regard, a collection of interfaces, parsers, connectors, and concept/attribute identification methods will be developed that facilitate the linking process between available structured and unstructured data resources on the web.

Based on this interlinked pool of resources, elaborate use-cases can be developed which address the specific requirements of alternative building design evaluation. Thus, both novice and experts in the building delivery process can be supported with regard to design decision making. The performance of the resulting evaluation environment will be tested in view of usability, coverage, and reliability.

As proof of concept, here a simple building element such as a plywood slab is considered. There is information available on the web that describes this building element but, as explained before, machines cannot collate this information and answer sophisticated queries. For this use-case we illustrate these challenges and the potential of LOD to address them.

First, the relevant information resources for the given building element should be identified. As the next step toward creating LOD, we need a machine-readable form of such information. Finally this information is semantically interconnected to facilitate answering the domain queries. For this purpose, we use the following resources:

The Wikipedia page for plywood (Plywood 2012) provides a generic definition of the product type, its history, and its applications. Fortunately, the machine-readable form of Wikipedia pages are already provided via DBpedia (DBpedia 2012) project, which allows to ask sophisticated queries against Wikipedia, and to link other data sets on the Web to Wikipedia data. Figure 3, shows the relevant DBpedia page for Plywood, which interestingly includes also some links to equivalent pages in other known resources such as Freebase (freeBase 2012) and OpenCyc (OpenCyc 2012). The raw data of DBPedia pages can be extracted in variety of formats such as RDF, CSV, and JSON. It is also important to note that a given element can also be queried via its alternative labels in other languages (see rdfs:label in figure 3).

In addition to DBpedia page which provides only the generic material information, there are also some pages which contain more technical information about plywood and real world products made of plywood material. As an example, we have taken two sample pages from Baubook (Baubook 2012) and MASEA (Masea 2012) platforms.

The Baubook web platform is a database of building materials and their ecological and physical properties. Building product manufacturers can define their products which will be evaluated and classified according to their properties and compliance with ecological regulations such as global warming potential, acidification potential, etc. Furthermore, the cross reference between building products, open biddings, and governmental funding facilitate the material navigation and extends the visibility of building elements in the building industry.

**Figure 3.** DBPedia description of Plywood including connections to freeBase and openCyc
The Baubook information meets the requirements of the first three levels of the five-star schema that was defined before by publishing raw XML data of the products on the web. In order to semantify the data, we have converted the data via a dedicated RDFizer component, which generates RDF versions of raw data according to the SEMERGY ontology. Listing 1 shows a sample RDF output for a plywood product of Baubook.

Listing 1. RDF version of Baubook data for a plywood product

```xml
<owl:NamedIndividual rdf:about="#solidwood_2142705277">
  <rdf:type rdf:resource="#SolidWood"/>
  <rdfs:label l:lang="de">Brettsperrholzplatte</rdfs:label>
  <rdfs:comment xml:lang="de">
    Großformatige Massivholzpanele aus Kreuzweiseverleimten Brettlagen ...
  </rdfs:comment>
  <semergy:heatTransferCoefficient>0.12</semergy:heatTransferCoefficient>
  <semergy:acidificationPotential>0.00341</semergy:acidificationPotential>
  <semergy:notRenewablePEI>8.04</semergy:notRenewablePEI>
  <semergy:globalWarmingPotential>-1.26</semergy:globalWarmingPotential>
</owl:NamedIndividual>
```

The MASEA platform is an open material database that covers the requirements of use-cases such as moisture prevention and thermal retrofit of existing buildings. With respect to our example, MASEA provides also the product information in XML format, which can be also transformed to SEMERGY-compliant RDF resources as shown in listing 2.

Listing 2. RDF output of MASEA for a plywood product

```xml
<owl:NamedIndividual df:about="#masea_sperrholz_1">
  <rdf:type rdf:resource="#BuildingMaterial"/>
  <rdf:type rdf:resource="#Wood"/>
  <rdf:type rdf:resource="#BuildingPanel"/>
  <rdfs:label xml:lang="de">BFU 100 Sperrholz</rdfs:label>
  <semergy:bulkDensity>427</semergy:bulkDensity>
  <semergy:specificHeatCapacity>1600</semergy:specificHeatCapacity>
  <semergy:thermalConductivity>0.11</semergy:thermalConductivity>
  <semergy:diffusionResistanceFactor>188</semergy:diffusionResistanceFactor>
  <semergy:freeWaterSaturation>572.6</semergy:freeWaterSaturation>
  <semergy:waterAbsorptionCoefficient>0.13</semergy:waterAbsorptionCoefficient>
  <semergy:openPorosity>66</semergy:openPorosity>
</owl:NamedIndividual>
```

After having the RDF version of resources in place, they can be linked together via appropriate semantic predicates. As it can be seen from Figure 4, the plywood product of Baubook is the same product that is documented under MASEA platform.

![Figure 4. Linking to LOD resource via well-known semantic predicates](image-url)
on such semantic links, the distributed data can be uniformly used in domain queries. Further items such as EU and national certificates can also be connected to the building products. This has been done for the Baubook’s plywood product, which is now related to the corresponding certificate on the web in PDF format.

The relations between information resources of the given example can be simply captured in RDF format as shown in Listing 3.

```
<owl:NamedIndividual
   rdf:about="#solidwood_2142705277"/>
<rdf:type rdf:resource="yago:Certified_wood"/>
<rdf:type rdf:resource="dbpedia:Plywood"/>
<owl:sameAs rdf:resource="semergy:#masea_sperrholz_1"/>
<semergy:hasCertificate rdf:resource="#M1BSPerrossplan_certificate"/>
</owl:NamedIndividual>

<bibo:LegalDocument
   rdf:about="#M1BSPerrossplan_certificate">
   <bibo:uri>
   </bibo:uri>
</bibo:LegalDocument>
```

Listing 2. Linking Baubook product to other data resources

7 MULTIOBJECTIVE OPTIMIZATION

Based on the SBM, the optimization problem can be specified. The optimization procedure’s main goal is to identify alternative building configurations, which are to be compared with the initial design encoded in the SBM. These alternative configurations, in the following denoted as ‘solutions’, are generated by (non-geometric) permutation with different pre-defined alternatives concerning material, elements and components, categorized into different ‘classes’ (e.g. framed windows, wall compositions). These classes encode the necessary meta-information to ensure that only feasible solutions are created.

Each alternative component, in the following denoted as ‘candidate’, is rated in multiple ‘categories’, contributing either to benefits or resources. For example, (initial or running) costs are classified as resources which obviously need to be minimized, while quality indicators of components or subsidies for particular construction methods are benefits and thus should be maximized. The individual category values of the candidates involved in a solution are then aggregated to calculate the solution’s category values which are then evaluated against other generated solutions. This aggregation may be done by simple summation of individual values as with the costs or may involve more complex calculations, e.g., energy efficiency indicators.

Handling the contradictory properties of benefit and resource categories can be achieved by different approaches. The straightforward approach is to apply suitable weights and aggregate the individual category values into a single scalar value, which can then be easily compared to the aggregated value of other candidate solutions. As this scalar value cannot properly represent the actual distribution of the category values, SEMERGY follows an alternative approach relying on Pareto-dominance, which means that the optimization procedure seeks for non-dominated solutions. A solution x is declared as non-dominated if there is no other solution that is better in all categories than x, or in other words, a solution x dominates another one if it is better in at least one category and not worse in all the others (see, for example, Silva et al. 2004). The outcome is a set of non-dominated solutions of the objective space denoted as Pareto front. Basically, this optimization problem refers to a discrete multi-objective combinatorial optimization problem and produces no single-best solution but different ‘tradeoffs’ between the objectives. This is especially important when no preferable ordering of the categories can be found.

As a combinatorial problem, iterating through all possible combinations to identify the Pareto-optimal ones is usually impossible to do for realistic numbers of candidate classes and members within each class. Even with filtering of candidates and thus reducing the corpus of possible combinations, evaluating the complete remaining solution space may still be computationally infeasible. However, in the majority of optimization problems, determining all Pareto-optimal solutions is actually not required. Decision makers are usually satisfied with a reasonably-accurate approximation of the Pareto front. Therefore, SEMERGY relies on meta-heuristics such as evolutionary algorithms – e.g., genetic algorithms (Holland 1992) – which only evaluate a fraction of the complete search space, but still produce reasonable results. Evolutionary algorithms are naturally inspired and mimic evolutionary processes such as crossover and mutation to generate potentially superior solutions where the inferior solutions are replaced with the superior ones in the population. With each iteration of creating new solutions and evaluating them, the population gradually approaches the Pareto-front, where the selection of the initial population and other mechanics prevent that the current population reaches the global maximum instead of being stuck in local ones. Examples of popular genetic algorithms that are able to handle multiobjective combinatorial problems include NSGA2 (Srinivas and Deb 1994) and SPEA2 (Zitzler et al. 2001).
7.1 Evaluation and Decision Making

The final component of SEMERGY is the evaluation and decision making. The solutions, i.e., alternative building configurations, produced in the optimization process can be evaluated and compared to the initial design in the valuation categories defined in the optimization problem. As the multi-objective optimization process usually produces a large number of solutions representing different tradeoffs between the category values, they need to be further filtered according to the decision maker’s preferences. SEMERGY provides an interactive mechanism where the decision maker is able to modify ‘hard limits’, e.g., setting an upper bound for maximum costs, and thus reduce the number of valid solutions to a manually manageable size. These remaining solutions can then be further analyzed in depth concerning all the building properties (material, components, etc.) the decision maker is interested in.

8 CONCLUSION
To be filled!

9 ACKNOWLEDGEMENT:

This project is funded under the FFG Research Studios Austria programme (grant No. 832012) by the Austrian Federal Ministry for Economy, Family and Youth (BMWFJ).

10 REFERENCES

Baubook 2012.
[http://www.baubook.at/], last visited March 2012
Buildingsmart 2012.
DBpedia 2012.
[http://dbpedia.org/About] last visited March 2012
FreeBase 2012.
Holland J. 1992, Genetic algorithms, Scientific American, p.66-72
Energy Plus 2011: Input Output Reference
Linkeddata 2012.
[http://linkeddata.org/], last visited March 2012

Masea 2012.
[http://www.masea-ensan.de/], last visited March 2012
OIB 2011: Richtlinie Energieeinsparung und Wärmeschutz & Leitfaden zur Berechnung von Energiekennzahlen [http://www.oib.or.at/], last visited March 2012,
Raiffeisen 2012.
[http://www.raiffeisen-baupsparen.at/energiesparrechner/], last visited March 2012