EXPLORING THE UTILITY OF SEMANTIC WEB TECHNOLOGY IN BUILDING PERFORMANCE SIMULATION

Ardeshir Mahdavi[1], Ulrich Pont[1], Feryal Shayeganfar[1], Neda Ghiassi[1]
Amin Anjomshoaa[2], Stefan Fenz[2], Johannes Heurix[2], Thomas Neubauer[2], A Min Tjoa[2]
[1]Department of Building Physics and building Ecology, Vienna University of Technology, Vienna, Austria
[2]Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria

ABSTRACT

The simulation-based assessment and evaluation of alternative building design and retrofit options requires large amounts of information. Thus, efforts are necessary to support the process of data search and collection, such that building models for performance evaluation routines and applications could be efficiently populated. The web environment can provide large amounts of potentially useful information. In this context, the present contribution reports on the SEMERGY project, which explores the utility of the semantic web technologies toward populating the input data for building performance simulation models via the navigation of the extensive but currently ill-structured web-based information space pertaining to building materials, elements, components, and systems, as well as resources and documents concerning procedural, climatic, and financial information.

INTRODUCTION

Common use cases of building performance simulation involve the evaluation of alternative building design and retrofit options. Toward this end, simulation tools must be supplied with large amounts of information. Such information primarily includes buildings' geometry, building components' technical (physical and ecological) properties, occupants' presence and actions, micro-climatic data, etc. Moreover, design decision making requires also information with respect to applicable codes and standards as well as available financing and subsidy opportunities. Conventional approaches to gathering such information are often cumbersome and time-consuming. Hence, opportunities for in-depth simulation-supported exploration of design and retrofit options (in view of their relative functional, economical, and ecological advantages and disadvantages) may not be optimally exploited. Thus, efforts are necessary to support the process of data search and collection, such that building models for performance evaluation routines and applications could be efficiently populated. To make further progress in this area, the vast and potentially useful information in the web environment should be considered. However, this web-based potential remains 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 represents a promising opportunity to improve and expedite the process of information acquisition and collation toward population of simulation models.

The Architecture, Engineering and Construction (AEC) industry has already started to make relevant data such as building products data, applicable normative documents, procedural and financial information, etc. available on the web. 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.

Efficient generation of building simulation models is hampered in part due to 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. The SEMERGY project intends to bridge this gap by providing semantic links between real world products and building model’s abstract concepts and elements.

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. The key contribution of the research is the exploration and demonstration of the semantic web technologies toward populating the input data for building performance simulation models via the navigation of the extensive but currently ill-structured web-based information space pertaining to building materials, elements, components, and systems, as well as resources and documents concerning procedural, climatic, and financial (e.g., public funding) information that could be of value to designers and decision makers.

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 type of user (novice versus expert).

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. 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, both simplified calculation routines (e.g., those necessary to generate energy certificates for building projects or perform life-cycle analyses) and numeric simulation applications (e.g., thermal and lighting performance

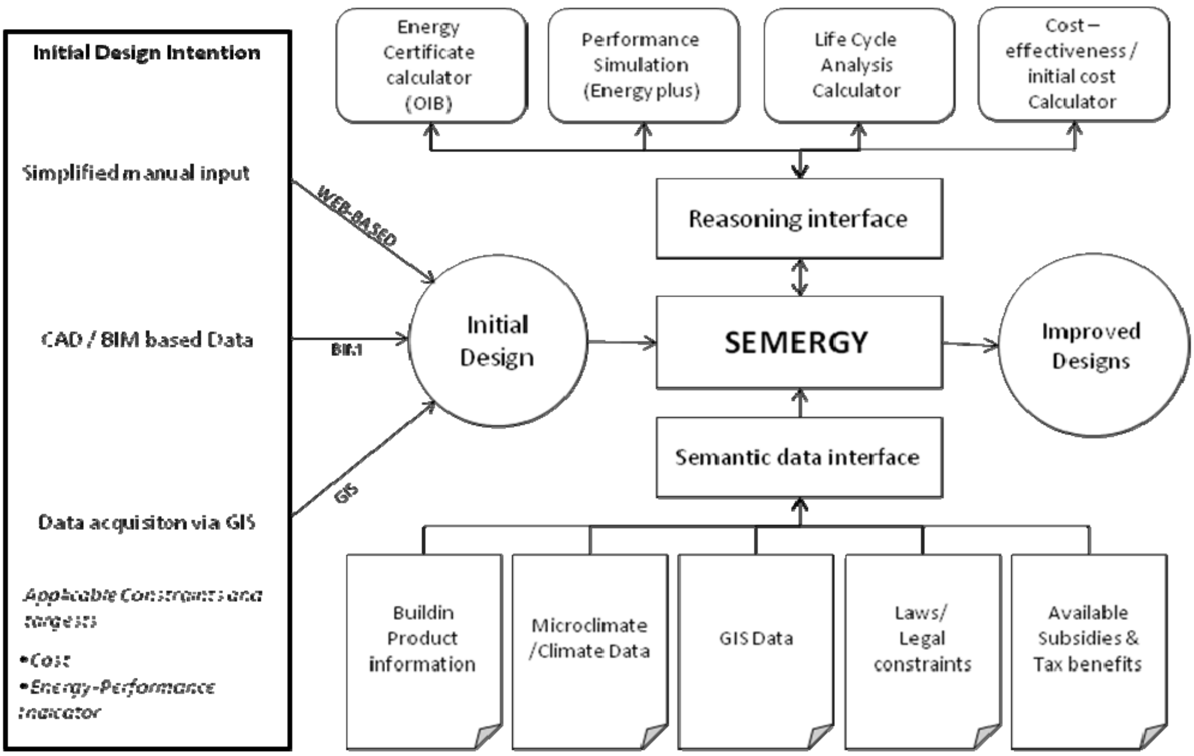


Figure 1 The general design of the SEMERGY environment

simulation tools) could be deployed. 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 and 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.

BUILDING INFORMATION REQUIREMENTS

For the purposes of the prototypical implementation, SEMERGY shall incorporate both a simple normative procedure (OIB 2012, ÖNORM 2011) and an advanced simulation engine (EnergyPlus 2011) to arrive at the value of the pertinent performance indicators. To arrive at a working model of the input data required for SEMERGY, the application was "reverse-engineered" to define an appropriate building information schema (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 (Buildingsmart 2012), 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. SBM specifies the kinds of input data required for the initial performance analysis and subsequent optimization. Moreover, the SBM provides the basis for the definition of ontologies that are needed to search the web environment for building product information und populate the simulation model with required semantic information.

SEMANTIC WEB AND ANALYSIS MODEL

The Architecture, Engineering and Construction (AEC) industry is composed of multiple knowledge domains. However, AEC-related information is often either trapped in customized data structures of software vendors or locked in domain-specific databases, which makes data reuse very difficult. More recently, the AEC industries have used the World Wide Web potential to provide relevant data such as building products data, applicable normative documents, procedural and financial information, etc. Such information resources are easily accessible and contain helpful data, but they necessitate heavy user interaction with distributed resources and data integration overhead. 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 (Berners-Lee et al. 2006).

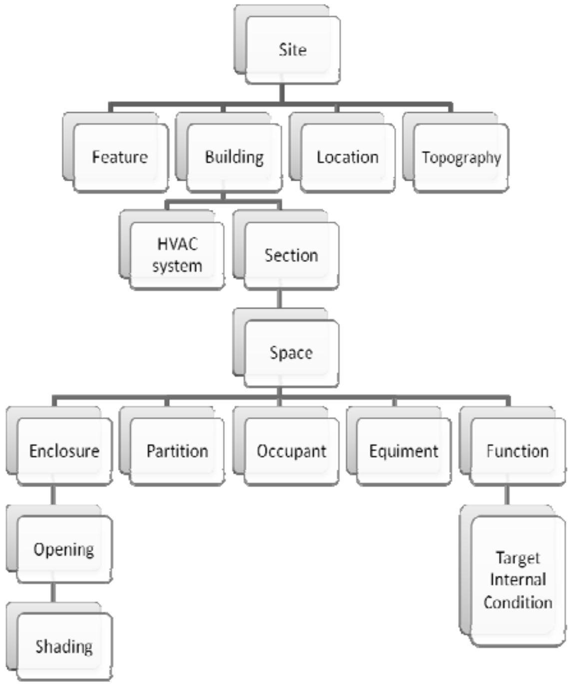


Figure 2 Building Data Model of SEMERGY (SBM)

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 Building Information Modeling (BIM) 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.

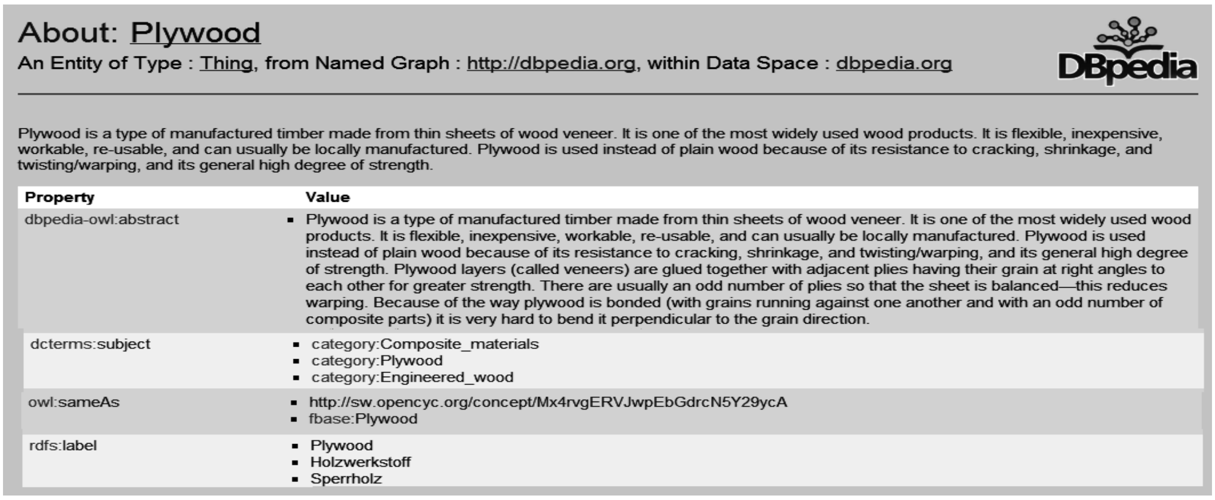
As such, the scattered information on the web may form a global data graph that connects distributed resources and facilitates the discovery of new resources. This approach, which follows a set of best practices for publishing and connecting structured data on the web is known as "Linked Data" (Linkeddata 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 and Bizer 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: *i)* Data availability on the web with an open license; *ii)* Data availability as machine-readable structured data; *iii)* Data availability in non-proprietary formats such as CSV, XML, etc.; *iv)* Using W3C's open standards (RDF and SPARQL) to identify entities; *v)* 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.

To understand this point, consider a simple building element such as a plywood slab. 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. The machine-readable form of Wikipedia pages are already provided via DBpedia (DBpedia 2012) project, which allows to launch 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 with 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.


About: Plywood

An Entity of Type : Thing, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Plywood is a type of manufactured timber made from thin sheets of wood veneer. It is one of the most widely used wood products. It is flexible, inexpensive, workable, re-usable, and can usually be locally manufactured. Plywood is used instead of plain wood because of its resistance to cracking, shrinkage, and twisting/warping, and its general high degree of strength.

Property	Value
dbpedia-owl:abstract	▪ Plywood is a type of manufactured timber made from thin sheets of wood veneer. It is one of the most widely used wood products. It is flexible, inexpensive, workable, re-usable, and can usually be locally manufactured. Plywood is used instead of plain wood because of its resistance to cracking, shrinkage, and twisting/warping, and its general high degree of strength. Plywood layers (called veneers) are glued together with adjacent plies having their grain at right angles to each other for greater strength. There are usually an odd number of plies so that the sheet is balanced—this reduces warping. Because of the way plywood is bonded (with grains running against one another and with an odd number of composite parts) it is very hard to bend it perpendicular to the grain direction.
dcterms:subject	▪ category:Composite_materials ▪ category:Plywood ▪ category:Engineered_wood
owl:sameAs	▪ http://sw.opencyc.org/concept/Mx4rvgERVJwpEbGdrcN5Y29ycA ▪ fbase:Plywood
rdfs:label	▪ Plywood ▪ Holzwerkstoff ▪ Sperrholz

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.

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.

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.

Based 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="#SolidWood"/>
            <rdfs:label
l:lang="de">Brettsperrholzplatte</rdfs:label>
    <rdfs:comment xml:lang="de">
        Großformatige Massivholzpaneele aus
        Kreuzweiseverleimten Brettlagen ...
    </rdfs:comment>
    <semergy:heatTransferCoefficient>
        0.12</semergy:heatTransferCoefficient>
    <semergy:acidificationPotential>
        0.00341</semergy:acidificationPotential>
    <semergy:notRenewablePEI>
        8.04</semergy:notRenewablePEI>
    <semergy:gloablWarmingPotential>
        -1.26</semergy:gloablWarmingPotential>
</owl:NamedIndividual>
```

Listing 1 RDF version of Baubook data for a plywood product

```
<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:sorption>
        69.9</semergy:sorption>
    <semergy:freeWaterSaturation>
        572.6</semergy:freeWaterSaturation>
    <semergy:waterAbsorbtionCoefficient>
        0.13</semergy:waterAbsorbtionCoefficient>
    <semergy:openPorosity>
        66</semergy:openPorosity>
</owl:NamedIndividual>
```

Listing 2. RDF output of MASEA for a plywood product

```
<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=
        "#M1BSPcrossplan_certificate"/>
</owl:NamedIndividual>

<bibo:LegalDocument
            rdf:about="#M1BSPcrossplan_certificate">
    <bibo:uri>
      http://www.mm-
kaufmann.com/fileadmin/ablage/
      dokumente/MMK_M1_BSP_crossplan_De.pdf
    </bibo:uri>
</bibo:LegalDocument>
```

Listing 3 Linking Baubook product to other data resources

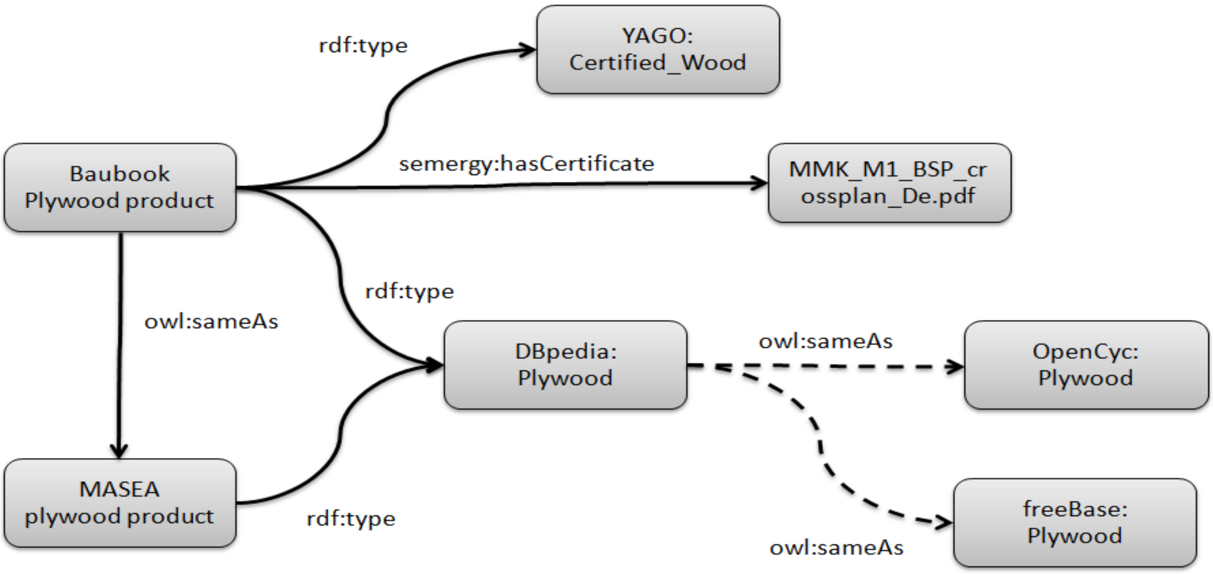


Figure 4. Linking to LOD resource via well-known semantic predicates

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. Thereby, the selection of the initial population and other mechanics ensure 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 multi-objective combinatorial problems include NSGA2 (Srinivas and Deb 1994) and SPEA2 (Zitzler et al. 2001).

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.

CONCLUSION

We reported on the initial stages and interim findings of the SEMERGY project. The overarching objective of the project is to contribute to the development of effective simulation-supported evaluation environments for comparative performance assessment of alternative building design and retrofit options. We considered different user types and associated modes of design intent communication. A compact yet versatile building data model (SBM) was derived via reverse-engineering of both normative calculation routines and numeric simulation tools. A constraint-based approach to the generation of semantic (non-geometric) variations of the initial design was used to provide a rich option space for the optimization process. The key contribution of the research is the exploration and demonstration of the semantic web technologies toward populating the input data for calculations and simulations via the navigation of the extensive but currently ill-structured web-based information space pertaining to building materials, elements, components, and systems, as well as resources and documents concerning procedural (e.g., legal), climatic, and financial (e.g., public funding) information that could be of value to designers and decision makers.

ACKNOWLEDGMENT

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

REFERENCES

Baubook 2012. *http://www.baubook.at/*.

Berners-Lee, T., Hendler, J., Lassila, O. 2001. The Semantic Web, Scientific American.

Berners-Lee, T., Gödel, K., Turing, A. 2006. Thinking on the Web. ISBN-13: 978-0471768142, Publisher: Wiley-Interscience.

Berners-Lee, T. 2009. Linked data. *http://www.w3.org/DesignIssues/LinkedData.html*.

Buildingsmart 2012. *http://www.buildingsmart.com/*.

DBpedia 2012. *http://dbpedia.org/About*.

FreeBase 2012. *http://en.wikipedia.org/wiki/Freebase*.

Heath, T., Bizer, C. 2011. Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool, ISBN: 9781608454303.

Holland J. 1992. Genetic algorithms, Scientific American, p. 66-72.

EnergyPlus 2012. *http://apps1.eere.energy.gov/buildings/ energyplus/pdfs/inputoutputreference.pdf*.

Linkeddata 2012. *http://linkeddata.org/*.

Masea 2012. *http://www.masea-ensan.de/*.

OIB 2012. Richtlinie Energieeinsparung und Wärmeschutz & Leifaden zuer Berechnung von Energiekennzahlen *http://www.oib.or.at/*.

ÖNORM 2011. ÖNORM B8110 1-6 & ÖNORM H5050 – H5060, Austrian Institute of Standardisation. Vienna.

OpenCyc 2012. *http://en.wikipedia.org/wiki/Cyc*.

Plywood 2012. *http://en.wikipedia.org/wiki/Plywood*.

Silva, J.D.L., Burke, E.K., Petrovic, S. 2004. An introduction of multiobjective metaheuristics for scheduling and timetabling. In Metaheuristics for Multiobjective Optimisation, Springer.

Holland J., 1992, Genetic algorithms, Scientific American, p.66-72.

Srinivas, N., Deb, K. 1994. Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms, Evolutionary Computation 2 (3), p.221-248.

Zitzler, E., Laumanns, M., Thiele, L. 2001. SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization, Evolutionary Methods for Design Optimization and Control with Applications to Industrial Problems, p. 95-100.