Modeling language and tools for the semantic link network

Hai Zhuge*,†, Kehua Yuan, Jin Liu, Junsheng Zhang and Xiaofeng Wang

China Knowledge Grid Research Group, Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, P.O. Box 2704-28, 100080 Beijing, China

SUMMARY

The semantic link network (SLN) is the extension of the hyperlink Web by attaching semantics to hyperlinks. It is an approach to construct a semantic overlay on Web resources. The specification of the semantics of the SLN model is an essential issue of SLN application. This paper proposes a modeling language for SLN consisting of an algebraic definition for SLN, a SLN metamodel and a Unified Modeling Language (UML) profile for SLN. The SLN metamodel specifies the primitives of the modeling language. The UML profile for SLN defines the specific syntax on SLN to make the modeling language understandable and usable. The development of the SLN builder implementing this language and the graphical SLN browser is introduced. This work is a part of the SLN model. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: browser; metamodel; modeling language; semantic link network; UML

1. INTRODUCTION

The simple hyperlink design is the key to the success of the World Wide Web. It enables any page to link to any page. Humans’ operating behaviors lead to the network effect and the uneven distribution of links. Another key to the success of the Web is its easy utility mode. It enables ordinary people to use it without any previous training. But the hyperlink is limited in its ability to support relation reasoning because it does not contain any semantics.

*Correspondence to: H. Zhuge, China Knowledge Grid Research Group, Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, P.O. Box 2704-28, 100080 Beijing, China.
†E-mail: zhuge@ict.ac.cn

Contract/grant sponsor: National Basic Research Program; contract/grant number: 2003CB317001
Contract/grant sponsor: International Cooperation Project of Ministry of Science and Technology of China; contract/grant number: 2006DFA11970
Contract/grant sponsor: EU 6th Framework Project GREDIA; contract/grant number: IST-FP6-034363

Copyright © 2007 John Wiley & Sons, Ltd.
The Semantic Web upgrades the Web to support intelligent applications by establishing machine-understandable semantics [1]. Much research has been done toward this goal by developing semantic-richer markup languages and domain ontologies. XML reflects the structural information of documents (www.w3.org/XML). The resource description framework (RDF, www.w3.org/RDF) has been proposed and used to describe relations between objects, attributes and values. The Web ontology language (OWL) has been developed to support description of relations and logical reasoning. It is designed for use by applications that need to process the content of information instead of just presenting information to humans, and it facilitates greater machine interpretability on Web content than those supported by XML and RDF by providing additional vocabulary together with a formal semantics (www.w3.org/TR/owl-features). More powerful languages are expected to emerge in the near future.

Rethinking the success of the Web, the simplicity and easy-to-use features can help evaluate the current efforts toward the Semantic Web.

Can we develop a Semantic Web by inheriting the successful features of the Web [2]?

Research on the semantic link network (SLN) model is increasing in recent years [3–9]. The model is the extension of the current hyperlink Web by attaching semantics to hyperlinks. It is a simple way to realize the Semantic Web by establishing a semantic linking overlay on Web resources including Web pages and underlying structures [1]. The SLN enriches the semantics of the Web and supports semantic calculus.

Compared with other Semantic Web solutions, the major advantages of SLN are its simplicity and the nature of semantic self-organization: any node can link to any semantic relevant node. It supports the implementation of an autonomous and self-organized Semantic Web [4,7].

Pons applies the semantic link technique to improve Web page prefetching [6,7]. He proposes a prefetcher utilizing this semantic link information associated with the current Web page’s hyperlink set to predict which Web objects to prefetch during the limited view time interval of the current Web page. The research result shows that the semantic link approach is effective in improving Web browser’s cache-hit percentage while significantly lowering Web page-rendering latency [3].

A SLN builder was implemented to facilitate the construction of SLN [6]. It focuses on defining the relations on documents. An intelligent browser was also developed to enable users to foresee several steps ahead while browsing the semantically linked document network. But the builder and browser have three shortcomings: lack of metamodel, lack of graphical display and lack of a distributed cooperative building mechanism. A cooperative mechanism is very important for a SLN to evolve from small and simple to large and rich.

The unified modeling language (UML) is a collection of diagram notations for software modeling [10–17]. Although accurate definition of its semantics is impossible [18], its syntax is well known and its meaning is well understood in the software community. Using UML to explain SLN enables users in the software community to understand SLN.

This paper proposes a SLN modeling language consisting of the algebraic definition for SLN, the SLN metamodel and the UML profile for SLN. The SLN metamodel specifies the primitives of modeling SLN. As the specific syntax, the UML profile for SLN is defined for users who are familiar with the UML to use the SLN model to describe domain knowledge. A SLN builder implementing the SLN modeling language is developed to help users easily define a SLN. A graphical SLN browser is also developed to visualize SLNs. A distributed SLN building and sharing prototyping system is developed to realize a cooperative SLN application environment.
2. DESIGN PRINCIPLE AND SLN MODELING TECHNIQUES

The design of the SLN modeling language follows the following four principles:

(1) **Extensibility**: The SLN model consists of two modules: the core module and the extension module. The core module consists of a set of semantic nodes, the semantic links between nodes and the linking rules (SLN rules). It can be extended by adding more reasoning and mapping features [9].

(2) **Compatibility**: The SLN metamodel should comply with an existing metamodel specification such as the MOF (meta-object facility) to ensure interoperability between SLN and other relevant techniques in the modeling architecture of the metamodel.

(3) **Well defined**: The metamodel should be well defined to guarantee the correctness of the SLN modeling language and the valid transformation from the logic model to its physical counterpart.

(4) **Easy to use**: The specific syntax of the SLN modeling language should adopt a well-known style to enable users to easily use the language.

The UML modeling community provides two built-in extension mechanisms for specific modeling needs, the MOF and the UML profile. The former is a meta metamodel that can be used to define specific metamodels such as the SLN metamodel, while the latter can be used for a 'lightweight' extension by attaching stereotypes or tags to a UML model structure and introducing semantics for specific modeling purpose [10].

A metamodel outlines the abstract syntax of a modeling language. It can be used to specify several specific syntaxes. The UML profile is eligible to define a specific syntax that is widely acceptable [11]. In the MOF metamodeling architecture, each model layer employs the same syntax style to provide a coherent meaning for designers.

Figure 1 illustrates the SLN modeling technique. The SLN metamodel adopts the MOF as the basic metamodeling language to describe its core semantic elements [16,17,19–22]. Just as the

![Figure 1. Techniques on the SLN modeling language.](image-url)
UseCase model in requirement modeling, the SLN metamodel demarcates the basic boundary of SLN modeling. The UML profile for SLN describes the syntax of SLN [12,13]. The modeling tool implementing the modeling language facilitates the SLN model design. The transformation mechanism from SLNs to the XML schemata enables the SLN browser to understand the UML profiles and enables users to intuitively observe the SLN structure.

3. SLN MODELING LANGUAGE

A modeling language consists of a set of syntactic notations and the meaning of these notations.

3.1. The general algebra semantics of SLN

A SLN consists of a semantic node set SemanticNodes, a semantic link set SemanticLinks and a reasoning rule set SLNRules, denoted as (SemanticNodes, SemanticLinks, SLNRules). Any semantic link in the SemanticLinks is a binary relation between two semantic nodes in the SemanticNodes. For any three semantic nodes A, B and C in the semantic node set, if there exist two semantic links A—→x B and B—→β C in the semantic link set, and there exists a semantic link rule X—→x Y, Y—→β Z ⇒ X—→γ Z (denoted as x • β = γ in short) in the SLNRules, then A—→γ C can be derived and added to SemanticLinks [5]. Two SLNs can be merged into one by common nodes or by adding semantic links between nodes of different networks.

SLN can be extended by including richer rules in the SLNRules to support inductive reasoning and analogical reasoning [9].

3.2. The SLN metamodel

A metamodel at the abstraction level describes a specific modeling domain such that newly developed models are the instances of the metamodel. To define a metamodel for the SLN modeling language, essential semantic elements should be defined and extended to overcome the structural constraints of the modeling language. Figure 2 shows the MOF-based SLN metamodel.

SLN elements are defined as follows [4–9]:

SemanticNode: A SemanticNode describes the characteristics of the specific semantic nodes representing a type of resources. SemanticNodes are linked by SemanticLinks. A SemanticNode owns several properties, together with property constraints defined by SLNNodeConstraints.

SemanticLink: A SemanticLink represents a binary association between two SemanticNodes as its source end and target end. With the extensibility of the SLN metamodel, users can define new semantic link types to extend the core metamodel.

SLNProperty: A SLNProperty represents the properties of the SemanticNode. It inherits MOF element property and owns two sub-elements: DatatypeProperty and SLNNodeProperty. A SemanticNode owns several SLNNodeProperties. The AnnotationPropertyValue represents a property value. For an AnnotationPropertyValue, its subject is an AnnotateableElement and its object is a DataValue inherited from the MOF Annotation.
**SLNType:** A **SLNType** can have several **DatatypeProperties**. It can be an integer, a string or a specific semantic node.

**SLNConstraint:** A **SLNConstraint** inherited from the MOF **Constraint** restricts SLN properties. It owns a **LiteralString** expression as its value. Three categories of constraints are **GlobalConstraint**, **DatatypeConstraint** and **SLNNodeConstraint**. For example, the constraint forEachNode(node.source LinkList.size()+node.targetLinkList.size()>0) indicates that no isolated node exists in a SLN. **SLNNodeConstraint** specifies the property type and values. For example, constraint node.targetLink List.size()>0 means that the semantic node is accessible from at least one node.

**SLNRule:** A **SLNAgebraRule** extends the **SLNRule** to capture the features of the algebra-based reasoning. Each rule owns several reasoning expressions according to the MOF **LiteralString**. Figure 3 illustrates the formal expression of reasoning rules. For example, the algebra-based specific reasoning rule: \( r \rightarrow ce \rightarrow r1, r1 \rightarrow imp \rightarrow r2 \Rightarrow r \rightarrow ce \rightarrow r2 \) can be represented as a **LiteralString** \( ce^{*}imp \Rightarrow ce \), where \( r, r1 \) and \( r2 \) represent semantic node, \( ce \) represents cause-effective link and \( imp \) represents implication link. Users can extend the **SLNRule** according to domain requirements.

### 3.3. UML profile for SLN

A UML profile provides a generic extension mechanism for building UML models in particular domains. UML profiles are based on additional stereotypes and tagged values that are applied to elements, attributes, methods, links and link ends. A UML profile is a collection of such extensions and restrictions for describing some particular modeling problems and for facilitating modeling constructs in a domain.
The UML profile for SLN employs stereotypes and tagged values to annotate SLN semantics in the SLN modeling syntax \cite{14,15}. Table I indicates the mapping between SLN features, UML features and UML profile features.

We have used the UML profile for SLN to construct a Dunhuang cave culture SLN model as shown in Figure 4. Two UML packages with stereotype ‘⟨⟨SLN⟩⟩’ were used to organize the SLN model DunhuangSLNExample and the inference rule set DunhuangSLNRule on this model. The package of the SLN model depends on the package of the rule set.

The package of the SLN model records dynasty information and Dunhuang cave information. The node Dynasty is the supertype of Ming, Tang and Wei dynasties. The node DunhuangCave is the aggregation of Wallpainting, Architecture and Sculpture. The node Feitian is a subtype of Wallpainting, and it owns the constraint FeitianConstraint on its properties. The semantic
Table I. The mapping between SLN features, UML features and UML profile features.

<table>
<thead>
<tr>
<th>SLN features</th>
<th>UML features</th>
<th>UML profile features</th>
<th>UML diagram syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Namespace</td>
<td>Package</td>
<td>⟨⟨SLN⟩⟩</td>
<td></td>
</tr>
<tr>
<td>SemanticNode</td>
<td>Class</td>
<td>⟨⟨SemanticNode⟩⟩</td>
<td></td>
</tr>
<tr>
<td>SemanticLink</td>
<td>Binary</td>
<td>⟨⟨SemanticLink⟩⟩</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Direct</td>
<td>X−z → Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Association</td>
<td>⟨⟨SemanticLink⟩⟩</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pre 1 suc</td>
<td></td>
</tr>
<tr>
<td>Property</td>
<td>Attribute</td>
<td>⟨⟨SLNProperty⟩⟩</td>
<td></td>
</tr>
<tr>
<td>Constraint</td>
<td>Constraint</td>
<td>⟨⟨SLNConstraint⟩⟩</td>
<td></td>
</tr>
<tr>
<td>Rule</td>
<td>Association</td>
<td>⟨⟨SLNRule⟩⟩</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class</td>
<td>⟨⟨SLNRule⟩⟩</td>
<td></td>
</tr>
</tbody>
</table>

nodes of SLN models are represented by the UML classes with the stereotype \langle\langle SLNNode \rangle\rangle. The SLN properties can be recorded as the properties of these classes. The semantic link ‘subtype’ in SLN models is represented by the UML inheritance relation, while the other semantic links should be represented by the UML association with the stereotype \langle\langle SemanticLink \rangle\rangle. The \langle\langle SLNConstraint \rangle\rangle constraint is connected with the restricted \langle\langle SLNNode \rangle\rangle via the UML association.

In the package of SLN reasoning rules, a rule is represented by the class marked with \langle\langle SLNRule \rangle\rangle. The semantic links of the SLN model are represented as the association classes and marked with the stereotype \langle\langle SemanticLink \rangle\rangle. Several \langle\langle SemanticLink \rangle\rangle classes as the rule parameters are associated with a \langle\langle SLNRule \rangle\rangle class. A rule instance exists if a set of nodes and semantic links satisfies the precondition of a rule. There may be several rule instances in a SLN.
model for the same rule. In this example, ‘Ming—sequential → Tang, Tang—sequential → Wei ⇒ Ming—sequential → Wei’ is a rule instance of the rule ‘X—sequential → Y, Y—sequential → Z ⇒ X—sequential → Z’.

4. TOOLS

4.1. The modeling tool

The aforementioned metamodel and UML profile are used to develop a modeling tool on the eclipse platform. The tool facilitates the SLN modeling process. Each component is implemented as a service. The model information and the graphical information are maintained in separate files. The tool supports synchronization of the SLN models and its physical counterparts.

When the canvas is initialized, a root controller and two command stacks are assigned. One stack saves ‘undo’ command and the other saves ‘redo’ command. Actions of users are translated into the corresponding requests and processed by the editpolicy module. Each model element owns its own model, controller and render. When the ‘create’ command is triggered, the root controller initializes the corresponding controllers to render the model elements on the tool canvas according to the model definition. The SLN on Dunhuang culture shown in Figure 5 is established by this visual modeling tool.

Figure 5. The SLN modeling tool based on the SLN metamodel and UML profile.
The following is the XML schema used as the common format for the SLN tools:

```xml
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified">
  <xs:element name="graph">
    <xs:complexType>
      <xs:all>
        <xs:element name='semantic_node'>
          <xs:complexType>
            <xs:attribute name="id" type="xs:integer" use="required"/>
            <xs:attribute name="name" type="xs:string" use="required"/>
            <xs:attribute name="label" type="xs:string" use="optional"/>
            <xs:anyAttribute namespace="##any"/>
          </xs:complexType>
        </xs:element>
        <xs:element name="semantic_link">
          <xs:complexType>
            <xs:attribute name="label" type="xs:string" use="required"/>
            <xs:attribute name="source" type="xs:integer" use="required"/>
            <xs:attribute name="target" type="xs:integer" use="required"/>
            <xs:anyAttribute namespace="##any"/>
          </xs:complexType>
        </xs:element>
      </xs:all>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

On the basis of validation mechanism of the SLN models, the modeling tool supports the transformation mechanism from the SLN model to the SLN XML file. The following algorithm checks the validity of a SLN model. If a SLN is valid, then the semantic nodes are unique and each semantic link has a source node and a target node.

Algorithm 1: validate the UML profile of SLN

```
Procedure validate()
{
  Initialize the umlfile of current SLN;
  Sparse the umlfile to get SLN resources;
  for ( each element N in resources )
  {
    if ( type(N) == ClassImp )
    {
      Generate semantic node with nodeID and node information;
      Add semantic node N to node_list;
      Add semantic node N to semantic node set node_set;
      for ( i=0; i < N.sourcelinklist.size()+N.targetlinklist.size(); i++ )
        add semantic node N to mapping_node_list;
    }
  }
}
```
// assure the names of semantic nodes are unique
If ( size(node_list) <> size(node_set) ) return false;
for ( each element L in resources )
{
    If ( type(L) == AssociationImp )
    {
        Get source node and target nodes of L;
        Remove source node of mapping from mapping_nodelist;
        Remove target node of mapping from mapping_nodelist;
    }
}

// assure the SLN is connected
If ( size(mapping_nodelist) <> 0 ) return false;
return true;

The following algorithm checks the validity of a SLN and stores the valid SLN into an XML file. The XML contains the semantic node and semantic link information.

Algorithm 2: Store the SLN model with XML
Procedure transform()
{
    validate();
    Initialize the umlfile of current SLN;
    Sparse the umlfile to get SLN resources;
    for ( each element N in resources )
    {
        if ( type(N) == ClassImp )
        {
            Generate semantic node with nodeID and node information;
            Add semantic node N to node_list;
        }
    }
    for ( each element L in resources )
    {
        if ( type(L) == AssociationImp )
        {
            Generate semantic link of L;
            Insert the semantic link into link_list;
        }
    }
    Store semantic nodes and semantic links into the XML file;
}
The following algorithm constructs a tree in depth first order and shows the semantic relevant nodes and links.

Algorithm 3: Construct Tree in Depth First Order

Procedure treeConstruction(SemanticNode root)
{
    for (each node n links to root)
    {
        add n to the node_list;
        add links between n and root to the link_list;
    }
    // assure the nodes and links construct a tree
    if ( size(node_list) - size(link_list) <> 1 ) return false;
    else
    {
        for ( each node K in node_list )
        {
            if ( outdegree(K) <> 1 ) return false;
            else
            {
                if ( indegree(K) == 0 ) return false;
                else
                {
                    dfTreeEstablish(K);
                }
            }
        }
        return true;
    }
}

Procedure dfTreeEstablish(SemanticNode activeNode)
{
    for ( each link L in link_list )
    {
        if ( targetNodeOF(L) == activeNode )
        {
            dfTreeEstablish ( sourceNodeOF(L) );
            Remove semantic link L from link_list;
        }
    }
}
4.2. The SLN browser

The SLN browser adopts a loosely coupled architecture for multiple types of display and interactions as shown in Figure 6. This architecture consists of three layers according to the MVC (Model, Controller and Viewer) design pattern. In this case, the SLN model represents the MVC Model. The MVC Controller translates the MVC Model into the runtime rendering, i.e. the MVC View. The display module of the SLN browser realizes the rendering of graphs. The user interface (UI) module deals with user interactions and forwards the interaction requests to the listeners of the SLN browser [23]. These listeners associate with the MVC viewers and reply to specific requests from the UI module in the run time.

The SLN browser adopts visualization technologies such as degree-of-interest (DOI) tree and hyperbolic tree [12,24].

5. APPLICATION

The SLN modeling language and support tool have been applied to establish the semantic overlay on various Dunhuang cultural resources in forms of image, text, video and audio in the Dunhuang Cultural Exhibition System. As shown in Figure 7, resource entities are located at the bottom, while the concept schema at the top layer annotates these resource instances. Here are 76 concepts and 126 resource instances. An additional link belongTo was introduced to extend the primitive links to help researchers study the content similarities and identify the implied links between resources.

The SLN browser loads the Dunhuang SLN segment file and provides the navigating and searching services for users. As shown in Figure 8, it displays SLN in the main portion and the relevant content on the right column. The user can search by inputting keywords in the...
bottom line and selecting the scale of SLN relevant to the input by moving the distance pointer. It suggests heuristic clues and verifies the rationality of suggestions according to semantic link reasoning. It enables users to see the result in the form of a network of highlighted semantic linked resources.

6. DISTRIBUTED SLN BUILDING AND SHARING MECHANISM

The establishment of a large-scale SLN is a social behavior; hence, it is very important to develop a mechanism to support the building and sharing of SLNs. Sharing extends the individual ability of building a SLN and helps interconnect different SLNs. Figure 9 presents a client/server networking infrastructure that facilitates the cooperative SLN development and sharing [5].

Users can download the SLN modeling tool from the host server by searching the software list in the central directory server and use it to define SLNs, store them at the client and upload the
defined SLNs to the host servers and register them in the directory server for sharing with others. Users can also download the SLN browsers from the host server in the same way and use them to browse local SLNs. Any newly uploaded SLN and software need to be registered in the server. For large-scale applications, the central resource lists can be distributed onto several servers. Moreover, the SLN servers can be organized in a fully decentralized manner [25].

With the upload and download behaviors, the global and local SLNs become richer and richer. The SLN browser has the following functions:

1. **The editing and reasoning functions**: When several SLNs are loaded in the SLN builder, SLNs with common nodes are merged. Therefore, new semantic links can be derived according to the semantic linking rules and added to the merged SLN. Users can edit the composed SLN according to their requirements.

2. **The searching and recommendation functions**: The SLN browser will analyze the description of the nodes, find relevant SLNs in the resource list and recommend an appropriate one to the user by matching the context of the node (the neighboring concepts and the connected

---

**Figure 8.** The Dunhuang culture SLN shown by the SLN browser.
semantic links) and SLNs. The user can store the recommended SLNs at the client and link them to the local SLNs.

Ontology is the basis of connecting SLNs. Figure 10 shows the interface of an OWL-based SLN builder. Users create SLNs with the builder at the client side. The loaded SLNs are displayed in the builder’s window. SLNs designed by different people can be merged according to the ontology and saved as a new SLN. Node information is listed in the right portion. The node and the semantic link information can be edited by users. Nodes can be labeled with concepts in domain ontology. If different ontologies are used, mapping between ontologies is needed [26]. After the definition of SLNs, users can upload them onto a SLN server to share with others.

7. CONCLUSION

The proposed modeling language consists of the SLN algebraic definition, the SLN metamodel and the UML profile for SLN. The metamodel specifies the primitives of SLN modeling. The UML profile enables users familiar with UML to represent domain knowledge by using SLN.
The advantage is that people with basic UML knowledge can use SLN even without knowing extra notation. Given a set of domain-specific semantic link and linking rules, the modeling language supports the definition of any domain-specific SLN. The SLN builder implements the modeling language and facilitates the definition of SLN. The graphical SLN browser enables users to intuitively observe the semantic relations. The distributed SLN building and sharing mechanisms suggest a solution to realize a semantic-linked Web.

ACKNOWLEDGEMENTS

The research work is supported by the National Basic Research Program (973 project no. 2003CB317001), the International Cooperation Project of Ministry of Science and Technology of China (Grant no. 2006DFA11970) and EU 6th Framework Project GREDIA (IST-FP6-034363).
REFERENCES


