Semantic Description of 3D Environments: a Proposal Based on Web Standards

Fabio Pittarello*
Università Ca' Foscari di Venezia
Dipartimento di Informatica

Alessandro De Faveri†
Università Ca' Foscari di Venezia
Dipartimento di Informatica

Abstract

While the number of virtual environments available on the net is constantly increasing, most of them are composed by a wide number of low-level geometric objects that lack any semantic description. Such situation prevents advanced uses of the data contained inside the environments, such as selection and extraction of semantic objects or advanced queries that refer to high-level properties of the environment. Recently some attempts have been performed in order to annotate 3D environments using the descriptive capabilities of MPEG-7 standard. While the solutions proposed are interesting, there are still a number of issues to be solved, including the definition of scene-independent ontologies that can be useful in different situations (e.g., 3D world validation, semantic search through a set of worlds, etc.).

This work proposes an alternate approach for associating semantic information to 3D worlds based on the integration of two web standards: the X3D language and the semantic web. The approach is characterized also by the definition of scene-independent ontologies and by the definition of semantic zones that complement the role of semantic objects for giving a complete description of the environment.

In order to show the potentialities of such approach the paper will illustrate an application scenario characterized by the extraction of the semantic information from an X3D document and the associated ontology for generating a high-level and multilevel textual description of a tour through the 3D environment described by them.

CR Categories: H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing—Indexing Methods; I.3.6 [Computer Graphics]: Methodologies and Techniques—Standards; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality;

Keywords: 3D environments, ontologies, semantic web, X3D

1 Introduction

The authoring of most 3D environments is characterized by the modelling and composition of low-level geometric elements like polygonal meshes or, in the most advanced environments, by objects belonging to the family of NURBS surfaces; in both cases such elements represent the building blocks used by expert modelers for assembling more sophisticated objects to which the authors themselves associate implicitly a semantics that is shared with the visitors of the 3D world (of course a successful association of a semantics to the 3D shapes depends on different factors like the skill of the modeler, a common cultural background for the modeler and the visitors, etc.).

Unfortunately, while the final users recognize such associations examining the visual properties of the 3D objects, often there is no way to deduct a high-level description of them from the examination of the file where the 3D information is stored. The modelling tools themselves, mostly conceived for obtaining simply a sequence of rendered views from the models, provide limited means to give a structure to the 3D objects in relation to their semantics. Besides, current modelling practices don't encourage the semantic connotation of 3D objects; some specific yet limited notation opportunities such as the layer metaphor (derived from the traditional technical drawing and used for partitioning the 3D world in different groups of objects) or the object labelling are not consistently used for storing in a standard way the high-level structure and communicate it to other people as an additional result of the modelling work.

The lack of any standard high-level description of the elements of such environments represents a fundamental drawback for advanced uses of the 3D world description that require the knowledge of the objects' semantics. A range of possible uses for such knowledge includes:

- web search engines processing requests formulated in natural language referring to high-level features: (e.g., *find the buildings containing an inner court surrounded by columns*);
- extraction of semantic objects from huge files for easier examination and automatic creation of high-level libraries that can be used for creating different 3D environments;
- high-level textual descriptions of paths to a given location, generated on the basis of static or dynamically generated paths.

Of course, an advantageous use of a high-level description is not limited to 3D virtual environments, but includes all the realm of the so-called mixed reality [Milgram and Kishino 1994] (e.g., a high-level description of a virtual model mapping a real environment can be useful for helping navigation in the real scene or in the augmented environment resulting from the sum of real and virtual components).

In this work we'll analyze the existing solutions for the semantic description of multimedia, with a particular reference to 3D scene description. We'll consider their points of strength and drawbacks and we'll compare them with our proposal that is characterized by the following features:

- integration of two web standards: the X3D language [Web3D Consortium 2004] for the description of 3D environments and the semantic web [W3C 2001];
- the definition of scene-independent ontologies that are used

*e-mail:pitt@unive.it
†e-mail:adefaver@dsi.unive.it

Copyright © 2006 by the Association for Computing Machinery, Inc.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.
Web3D 2006, Columbia, Maryland, 18–21 April 2006.
© 2006 ACM 1-59593-336-0/06/0004 $5.00

for defining classes of objects and their relations;

- the definition of semantic zones as a means to enhance the high-level description of the environment (the notion of Interaction Locus [Pittarello 2003], defined by one of the authors, will be exploited in this work);

- a multilevel description of the environment, allowing to query the scene at different levels of semantic granularity.

The rest of the work is organized as follows: Section 2 will consider related works, with a particular reference to the high-level description of 3D environments; Section 3 will present our proposal for the semantic description of 3D environments, focusing on the definition of the semantic information to be included in the X3D files and in the associated ontology; an introductory example will be given to illustrate how complementary semantic information is distributed across such formats; Section 4 will focus on a complex application scenario: the high-level and multi-level description of a path through a virtual environment starting from a semantic description of the 3D environment itself and from the low-level definition of the path (e.g., a manually or automatically generated polyline); the description of the associated implementation architecture will conclude the section; in Section 5 the scenario considered in the previous section will be exemplified by a case study related to an urban environment with a VoiceXML compliant presentation of the guided tour through an ancient Venetian palace; Section 6 will conclude the paper, giving some hints of future development.

2 Related Work

Research related to the semantic description of multimedia documents has become increasingly important in the last few years. Most of the work done so far is related to two distinct areas: the semantic annotation of audio-video information and the semantic annotation of resources available on the web; attempts to describe semantically 3D worlds (including this work) refer to such research areas.

Concerning audio-video information, the Moving Picture Experts Group (MPEG) [MPG] has done the most relevant work, defining a set of standards for coding and describing such data. The most interesting standards in relation to this work are MPEG-4 [Pereira and Ebrahimi 2002] and MPEG-7 [Nack and Lindsay 1999a] [Nack and Lindsay 1999b]: the first specification considers a multimedia document as the summa of different objects the user can interact with; a recent addition to MPEG-4 standard (XMT-A) defines an XML based format containing also a subset of the X3D nodes; the latter specification allows to describe multimedia content specified using other standards (e.g., MPEG-4, SVG, etc.); such description is archived independently of the content data file itself.

Some interesting proposals for describing the semantics of a 3D scene take advantage of the MPEG-7 standards ([Bilasco et al. 2005], [Halabala 2003], [Mansouri 2005]). Halabala uses MPEG-7 to store semantic graphs related to a 3D scene. A prototype application developed by Halabala allows the user to select a high-level description of a specific object on an semantic hierarchy of objects obtained from an MPEG-7 description; the selection process highlights on a VRML browser the associated low-level geometry (described using VRML language); on the other side, the user is enabled to select a piece of the 3D geometry on the VRML browser in order to obtain the associated semantic objects (a hierarchy of semantic objects exists, so the user is given the chance to select the appropriate level; e.g., the click on a polygon associated to a wall enables the user to select the semantic objects wall, room or building).

An interesting problem considered by this work is that of the shared objects such as the wall shared between different rooms; the author states that the VRML/X3D hierarchy used for describing the geometry is not well suited to the description of the semantics, being conceived for modelling virtual reality rather than for a high-level description of the scene. That is the reason why the author proposes to use MPEG-7 for overcoming such deficiency and obtaining a satisfactory result.

While the work described above represents an interesting proposal, complemented with a number of application scenarios, a significant drawback is that the semantic description of the 3D world is related to the specific scene; in other words is not available a scene-independent ontology defining which are the valid relations for the objects belonging to a certain domain.

Another interesting work [Mansouri 2005] uses MPEG-7 for describing the semantics of virtual worlds, enhancing both querying and navigation inside of them (e.g., the system can return virtual words after semantic queries such as *I am looking for a tall tree*). The author of this work considers the issue of enhancing the description of existing 3D worlds with domain-oriented annotation, using a modified version of OntoWorld [Ont], a tool designed for generating virtual environments from the semantic description of the environments themselves; the modified tool generates MPEG-7 compliant documents where semantic information added to existing 3D files is saved.

Again, while the Mansouri work is important for considering the effects of the application of a semantic descriptive layer on the query and navigation phases of virtual worlds and for the proposal of an annotation system for existing 3D scenes, it still lacks in generality. While the solution described minimizes the impact on required modifications to existing VRML/X3D files, the associated MPEG-7 description is not scene-independent and therefore can't be reused as a guide for generating or validating virtual worlds, nor as a description that can be profitably used by search-engines capable of performing semantic searches across a set of virtual worlds.

Concerning annotation of resources available on the web, the World Wide Web Consortium promotes the definition of a set of languages, rules and tools allowing a high-level description of such information: the semantic web. The semantic web is composed by different layers, where the lower one is occupied by the data themselves (expressed in XML) and the higher ones describe the semantic properties of such data. While the highest level layers focusing on logic properties and trust are not completely defined, the second and the third layers are available for use:

- the RDF (Resource Description Framework) [RDF 2004] for creating a set of semantic rules to create a description and the RDF Schema for combining such descriptions to create a vocabulary;

- the OWL (Web Ontology Language) [OWL 2004] for supplying ontologies.

Such layers will be used in the context of this work for contributing to the creation of a semantic description for the 3D environments.

Most of the research related to the semantic notation of 3D environments has focused on the high-level description of objects available on the scene and/or the description of higher-level semantic objects given by the composition of existing objects (e.g., the higher level semantic object *room* is described as the *sum* of a number of *walls* that are lower-grade semantic objects).

While such notation is important, it lacks to explicitly identify the partitions of space that are perceived as a morphologic unit and are assigned to specific uses. Such morphologic and functional units are a fundamental element for allowing the user to build a mental map of the environment; users *attach* lower-level semantic objects to such partitions and such association grants a preferential way to recall the objects themselves, as demonstrated also by mnemonic techniques developed by western civilization starting from the Renaissance [Yates 1966].

The designer of the 3D environment implicitly assigns a semantics to the space using elements that surround completely or mark partially the zone; such semantics will then be recognized by the final user that will derive the identity of the space examining the visual properties of the 3D objects and will associate it to a specific function. But the lack of an explicit high-level description of the morphologic and functional role of such zones will prevent a full understanding of the environment and of its use.

The Interaction Locus [Pittarello 2003] approach represents a contribution for enhancing the description of a 3D space through the definition of the zones (called Interaction Loci) that are semantically relevant for the users; such approach will be considered in the next section as a component of the paper proposal.

The approach shares some features with the Action Spaces proposal [Dachselt 2000], such as the recognition of the relation between places and tasks performed inside of them (i.e., both approaches identify spatial partitions and associate to them subsets of the actions allowed in the whole environment, in order to avoid visual clutter and occlusion). In spite of that, while the Action Spaces approach is more task-oriented and less anchored to geometry (i.e., an action space may be defined also by the viewpoint of the user, one or more interactive objects and a set of associated interaction widgets), the Interaction Locus approach is more focused on evidencing the morphologic features of the environment through a hierarchy of places. The underlying idea is that, even though spaces can be associated to classes of actions, morphology precedes function and it is an universal value that can be recognized even when the relations between locations and associated actions are lost, such as in the ruins of ancient buildings. That is the reason why the Interaction Locus approach considers as primary semantic components the partitions of spaces based on morphology.

3 A Proposal for the Semantic Description of 3D Environments

The requirements we started from for the definition of our proposal included:

- the reference to existing web standards, in order to maximize the generality of the solution, to minimize the effort related to the creation of solutions and tools for manipulating data and, possibly, to ease the conversion of the vast amount of 3D environments available on the net;
- the definition of a scene-independent reusable ontology, as an improvement to the previous solutions that packaged in the MPEG-7 files the semantic description of a specific environment;
- the definition of a complete multi-level description of the environment including not only the 3D semantic objects but also the semantic zones that characterize the environment.

Concerning the definition of scene independent ontologies, we considered to take advantage of MPEG-7 for such purpose, but finally we shifted towards an orthogonal solution entirely based on the semantic web. Such solution allows a better integration with the other web standards and, in particular, with X3D; besides, the upcoming definition of the highest level layers of the semantic web, focusing on logic and trust, seems an interesting opportunity for future improvements of the semantic scene description.

In our proposal semantic information is split across the two different file formats we have chosen for storing data and associated descriptions:

- the X3D file, containing the description of the geometry of the 3D environment and the metadata for adding semantics to the objects of the environment;
- the RDF Schema file, for storing the relations between the different classes of semantic objects; such relations are scene-independent and generally are valid for a specific domain (e.g., the built environment, automotive, boats, etc.).

3.1 Describing Semantic Objects Using X3D

A significant part of the description of the semantics of the virtual environments is based on the concept of object. Two kinds of objects are defined: the *geometric objects* (GO) and the *semantic objects* (SO). The first class of objects represents both the VRML/X3D primitives (such as the *box* node) or composite objects obtained aggregating such basic components. One or more geometric objects can be associated to a semantic object, an entity that has a specific meaning shared by the author and by the final user of the environment. Semantic objects can be organized in a hierarchical structure, originating a multi-level description.

Figure 1 illustrates a hierarchy of semantic objects; the lowest level semantic objects (e.g., the wall, the ceiling and the pavement) are associated to geometric objects (i.e., not represented in the scheme), while higher-level semantic objects like the room or the house are composed by other semantic objects.

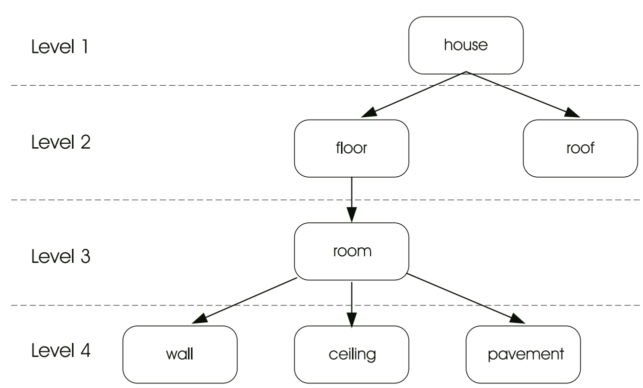

Figure 1: A hierarchy of semantic objects.

The transformation of a GO into a SO is realized associating to the existing geometry nodes a *MetadataSet* node that describes the semantic properties of the objects. The *MetadataSet* field *name* is used for identifying the class of the object (e.g. the class *wall*), while the field *reference* specifies the identity of the object itself (e.g., *wall01*) and must be unique in the whole X3D file. Other semantic properties can be added using nested *MetadataString* or *MetadataSet* nodes. Such information can be a property of the object itself or the declaration of a relation with other objects of the

3D environment (such declaration of course needs to be compliant with the scene-independent ontology that will be illustrated in Section 3.3).

Figure 2 shows a simple semantic notation of the semantic object *wall*. The metadata represent both the properties of the object but also the relations with the higher level semantic objects; in particular the *MetadataString* declares the relation of containment with the higher-level object *bedroom01*.

```
<Transform DEF='dad_Box4' ...>
 <MetadataSet name="wall" reference="wall01">
   <MetadataString name="containedBy"
    value="bedroom01"/>
   ...
  </MetadataSet>
 <Shape DEF='Box4'>
 ...
 </Shape>
</Transform>
```

Figure 2: X3D code for describing the semantic object *wall*.

Borrowing the terminology from [Halabala 2003], we'll distinguish *real* and *virtual semantic objects*. In the case of real semantic objects the SO has a counterpart in a geometry node available in the scene (e.g., a *Box* node or a *Transform* node containing one or more geometric objects); in such situations the metadata nodes can be added as children to the geometry node.

In the case of virtual semantic objects the SO is not directly represented using geometry nodes; therefore there is an additional problem of finding an appropriate location to store the associated metadata, possibly without altering the hierarchy of the scene geometry. In such case we can identify two different situations: semantic objects represented by a set of unrelated geometry objects (i.e., not grouped) and semantic objects based on lower-level semantic objects.

In both cases the solution tries to minimize the impact on the scene, associating to the *WorldInfo* node a *MetadataSet* containing a set of *MetadataSet* nodes; each *MetadataSet* will contain the semantic information related to a specific virtual semantic object:

- in the case of virtual semantic objects represented by a set of unrelated geometry objects, the latter objects will declare their relation with the virtual object using a simple *MetadataString* node;

- in the case of virtual semantic objects based on lower-level semantic objects, the latter objects will declare their relation with the virtual object using a *MetadataString* node nested into the *MetadataSet* associated to them, as shown in Example 1.

Finally, the semantic annotations introduced in the document will be associated to the external domain-specific ontology through a meta tag inserted into the head section of the X3D file, as shown in Figure 3.

```
<head>
    <meta name="ontology"
     content="urban_environment.rdfs"/>
</head>
```

Figure 3: Associating X3D documents to an external ontology.

3.2 Describing Semantic Zones Using X3D

While a high-level description of the geometry of a given environment is useful for cataloging the parts the 3D scene is made of and can be very useful for searching semantic objects at different levels of granularity, such world annotation is not complete because it doesn't take into account explicitly the use of space.

A simple example is the element *door* that it is a semantic object composed by other lower-level semantic objects (e.g., the handle), but it is also an artifact that generates a space that has the important function of connecting two different zones of a 3D environment.

Therefore, even though objects generate spaces, there is a strong semantic difference between semantic objects and spaces that must be considered by all the ontologies aiming to give a complete description of the environment.

The Interaction Locus (IL) concept, defined by one of the authors of this work, satisfies the need for an explicit definition of the partitions of space that are recognized as morphologic units and have a specific functional role. The IL was originally conceived for enhancing the navigation capabilities inside 3D environments for the net and was progressively extended as a fundamental component for describing the dynamics of interactive 3D environments [Pittarello and Fogli 2005].

The semantic zones defined for this paper share all the features of the ILs definition:

- they are mid-level semantic elements positioned between the whole 3D scene and the low-level semantic objects;

- they are explicitly identified through the statement of their boundaries and have a number of associated properties;

- they can be nested, originating a hierarchy of spaces.

Concerning implementation, the ILs, starting from their initial definition [Pittarello 2003], have been associated to *ProximitySensor* nodes. While this choice limits the definition of their boundaries to 3D boxes, such morphology is sufficient for most situations (i.e., of course a *Prototype* may be defined for overcoming such limits); on the other side such solution enables to monitor easily the navigation of the user inside different locations and to send to internal or external engines information for activating suddenly information and actions compliant with the semantics of such spaces during the interactive use of the environment. The association to *ProximitySensors* will be maintained even for the semantic zones of this work.

The semantic zones are referenced in the X3D code by metadata associated to real or virtual semantic objects (e.g., objects belonging to the class *room*, *door*, etc.) and contain a set of metadata for storing their properties.

```
<ProximitySensor ...>
    <MetadataSet name="doorSpace"
     reference="door1Space">
    ...
    <MetadataString name="type"
     value="connection space"/>
    ...
    </MetadataSet>
    ...
</ProximitySensor>
```

Figure 4: Describing semantic zones in X3D.

Figure 4 shows the ProximitySensor associated to the semantic zone *doorSpace*. The field *value* declares the primary role of such zone: a space for connecting other zones (*connection space*); an alternative role, as will be shown in Subsection 3.3, is that of a space for performing different kinds of residential activities (*action space*); such information can be useful in a number of application scenarios (e.g., for guiding the extraction of semantic information from the environment itself, for planning paths across the environment, etc.).

3.3 An RDF Schema Based Ontology

Association of metadata to the X3D scene is only the first part of the semantic annotation of the 3D environment. Such association permits to add scene-specific high-level information that describes the properties of the objects and the relations existing between them (e.g., the *wall01* is contained inside the *room01*). The description of scene-independent properties, defining how can be composed a scene belonging to a specific domain (e.g., the class *wall* can be contained inside the class *room* or inside the class *loggia*) is specified in a separate RDF Schema file.

The resulting scene-independent ontology can be used for different aims, such as building worlds conforming to specific rules (e.g., worlds compliant with classic Greek architecture), validating 3D environments or helping a semantic search engine to understand which kind of relations can be asked in a specific domain.

Generally speaking, one or more domain-oriented ontologies can be associated to a specific scene, depending on the objects contained inside the 3D world (e.g., a 3D environment containing cities, humans and flowers will need to refer to three different ontologies related to the specific domains); of course we'll need to add a specific *meta* tag to the X3D file for each additional ontology.

3.4 The Shared Walls World

To exemplify the concepts discussed so far, we'll consider a simple semantic world, compliant with a basic ontology (i.e., the *Shared Walls* ontology). The classes of semantic objects are organized in three levels, as shown by Figure 5.

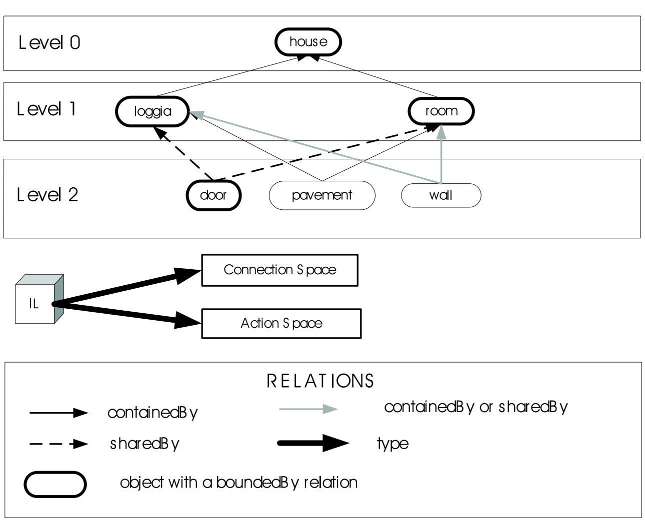

Figure 5: The *Shared Walls* ontology.

In spite of the simplicity, such ontology considers also the case of shared semantic objects; as stated in Section 2, shared objects cause a representational problem using X3D standard, that is conceived for the modelling of the scene rather than for its semantic annotation. That is the reason why Halabala describes the semantics of shared objects in an external MPEG-7 file. Our solution shares with such approach the adoption of an external file, even though this file is used only for describing the possibility of having certain relations between classes, in order to maintain the independence of the ontology from the scene; complementary information related to the specific objects is stored inside the X3D file.

The ontology illustrated in Figure 5 allows to define different relations between the semantic objects:

- *containedBy* relation, for expressing that the object contained is a component of an higher-level semantic level;
- *sharedBy*, for expressing that the object is shared between two higher-level semantic objects;
- *boundedBy* relation, for expressing that the semantic object has an associated semantic space; such space can belong to the type *connection space* or *action space*, depending on its primary role in the 3D world.

Let us consider a world compliant with the *Shared Walls* ontology: an house formed by a bedroom and a loggia, composed by walls, pavements and a door; some of the lower-level semantic objects are shared (i.e., the wall and the door connecting the room and the loggia).

The geometry of the objects and part of the semantics of such world are stored in the X3D file, shown in Figure 6. The environment is partially described by real semantic objects associated to *Transform* nodes. The first *Transform* node describes one of the three walls that have a relation of containment with the semantic object *bedroom*, expressed in the metadata with the relation *containedBy*. The second *Transform* node is shared by two different higher level semantic objects; therefore the associate metadata declare such relation (i.e., *sharedBy*) and the names of the higher level objects (*bedroom* and *loggia1*).

While the lower-level semantic objects are associated to existing geometry nodes, the mid-level nodes don't have such counterpart to attach metadata; therefore, according to what stated in Section 3.1, a set of metadata is included inside the node WorldInfo for describing such semantic objects. The example shows a *MetadataSet* describing the properties of the semantic object *bedroom*, its association with the semantic space *bedroomSpace* and the relation of containment with the semantic object *house1*.

All the semantic spaces are described using *ProximitySensors* and associated metadata; in particular the semantic space *roomSpace* is described at the end of the example; its primary role of *action space* is stated using the field *value*.

The RDF Schema document describes the classes of the semantic objects contained in the X3D document and their relations. The name used for identifying the class of a semantic object in the X3D document (i.e., the *name* field of MetadataSet node) corresponds to the *ID* of the class in the RDF Schema file.

Moreover, the RDF Schema document describes also the relations between classes (for example, the relation *containedBy* that creates a bottom-up structure from the object at lower level to the top level (the root semantic object that contains all the others). The OWL elements that permit the use of restrictions have been used for some classes and relations.

Figure 7 shows a fragment of the RDF Schema defining the *Shared Walls* ontology; the code shows the definitions of the class *wall*

```
<head>
<meta name="ontology" content="rooms.rdfs"/>
</head>
<WorldInfo ... >
 <MetadataSet name="virtual objects">
  <MetadataSet name="room" reference="bedroom">
   <MetadataString name="description"
    value="room for sleeping"/>
   <MetadataString name="boundedBy"
    value="bedroomSpace"/>
   <MetadataString name="containedBy"
    value="house1"/>
  </MetadataSet>
    ...
</MetadataSet> </WorldInfo>
...
<Transform ...>
    ....
 <MetadataSet name="wall"
  reference="wall01">
  <MetadataString name="containedBy"
   value="bedroom"/>
 </MetadataSet>
    ...
<Shape DEF='Box4'> ... </Shape>
</Transform>
  ...
<Transform ...>
    ....
 <MetadataSet name="wall" reference="sharedWall01">
  <MetadataString name="sharedBy" value="bedroom"/>
  <MetadataString name="sharedBy" value="loggia1"/>
 </MetadataSet>
    ...
 <Shape DEF='Box10'> ... </Shape>
 </Transform>
...
<ProximitySensor>
 <MetadataSet name="roomSpace" reference="bedroomSpace"/>
  <MetadataString name="type" value="action space"/>
 </MetadataSet>
</ProximitySensor>
```

Figure 6: An X3D world compliant with the *Shared Walls* ontology.

and of the relations *containedBy* and *sharedBy*; the OWL restrictions permit to express that the containment relation for an objects belonging to the class *wall* can be defined towards a single object belonging to the class *loggia* or to the class *room* (*maxCardinality*=1), while the sharing relation for the same object can be defined towards two objects belonging to the higher-level classes (*maxCardinality*=2).

4 An Application Scenario: Extracting Semantic Information for Guiding the User through a Complex Environment

Adding semantic information to raw geometric shapes allows to explore a number of interesting scenarios. This Section will consider in detail an application scenario characterized by the extraction of the semantic information from a X3D document and the associated ontology for generating a high-level and multilevel textual description of a tour through the environment. Such description can be

```
<owl:Class rdf:about="#wall">

    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#containedBy"/>
        </owl:onProperty>
        <owl:someValuesFrom>
          <owl:Class>
            <owl:oneOf rdf:parseType="Collection">
              <owl:Class rdf:ID="room"/>
              <owl:Class rdf:about="#loggia"/>
            ...

    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:ID="sharedBy"/>
        </owl:onProperty>
        <owl:someValuesFrom>
          <owl:Class>
            <owl:oneOf rdf:parseType="Collection">
              <owl:Class rdf:about="#room"/>
              <owl:Class rdf:about="#loggia"/>
           ...

    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:cardinality rdf:datatype="&xsd;int">
        1</owl:cardinality>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#containedBy"/>
        </owl:onProperty>
      </owl:Restriction>
    </rdfs:subClassOf>

    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty>
          <owl:ObjectProperty rdf:about="#sharedBy"/>
        </owl:onProperty>
        <owl:cardinality rdf:datatype="&xsd;int">
        2</owl:cardinality>
      ...
```

Figure 7: A fragment of the RDF Schema defining the *Shared Walls* ontology.

useful in a number of situations, including the automatic creation of personalized guided tours through a 3D real environment, created on the basis of a specific path; besides the solution proposed can be an interesting starting point for supporting blind users moving in real environments mapped by their 3D virtual semantic counterparts.

The implementation architecture proposed uses as starting points the semantic description of the environment and a low-level representation of the path (i.e., a *polyline2D* node) that can be created from scratch or obtained by the observation of the behavior of the users inside the environment [Chittaro and Ieronutti 2004].

We define the *path* as a connected sequence of segments having a starting and an ending location (see Figure 8(a)), that can be optionally enhanced (see Figure 8(b)) by the presence of interesting locations (points of interest or POIs, borrowing the terminology from GPS navigators technology). Such POIs can be positioned by the

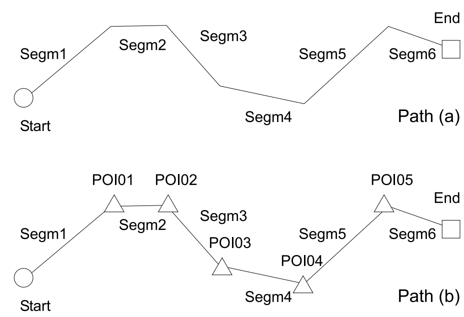

Figure 8: A simple (a) and a POI-enhanced (b) path.

modeler as a part of the creation of the path, or more interestingly, can be automatically derived from the observation of the behavior of previous users inside the environment (e.g., putting a POI on the path where the users stand for a certain amount of time) for creating a tour of the most visited locations of a certain environment.

While in our scenario the second version of the path can help to extract the appropriate information from the semantic description of the environment (e.g., avoiding details in the locations that lack any POI and extracting complete multi-level information where a POI is available), the first version of the path, used for our current implementation architecture, allows to extract all the semantic information at different levels; such solution, as shown by the case study in Section 5, leaves to the user the freedom to *zoom in* for having in-depth information or to *zoom out* for having information about the higher-level semantic zones.

4.1 Implementation architecture

Figure 9 describes the implementation architecture for the application scenario described above. The architecture is based on a coordination component (*main process*, Figure 9(1)) that elaborates the results of three specialized components:

- *the metadata extraction component* (Figure 9(2)), that extracts metadata from the geometry using an external XML stylesheet [XSLT 1999];
- *the structure and information extraction component* (Figure 9(3)), that extracts semantic information from the ontology using RDQL queries and creates a table with a set of rules (e.g., object X can be contained in object Y, etc.);
- *the positional information creation component* (Figure 9(4)), that uses the low level path and the geometry of the X3D file for defining the relations between the position of the user and objects (e.g., in front of, on the left, etc.), changes of direction (e.g., turn right, etc.) and partial distances between the starting and end points of the segments of the path.

The coordination component outputs its result in the form of a structured XML file that is post-processed using XSLT or alternative components that generate files compliant with a number of presentation formats (e.g., XHTML, VoiceXML, etc.).

Therefore, while the implementation architecture is designed for a specific scenario (i.e., guiding the user through a complex environment), it offers a significant flexibility for presenting monomodal or multimodal information in a variety of situations, ranging from assistance in real environments to navigational support in desktop virtual reality.

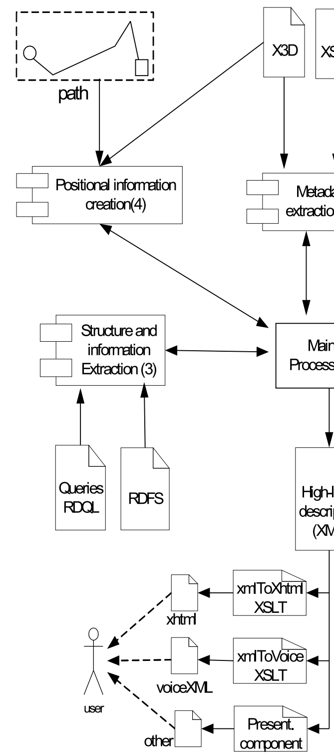

Figure 9: An implementation architecture for extracting multilevel semantic information.

Besides, while the current architecture uses as initial input the static description of a path through a given environment and outputs navigational information that is valid only if the user doesn't depart from the path itself, its components can be used for an extended architecture (i.e., that goes beyond the scope of this work) capable of processing dynamically new user positions outside the predefined navigational steps and presenting to the user information updated in real-time.

Of course, for such extended architecture, an additional component monitoring the user motion through the environment will be needed in order to discover deviations from the path, computing and presenting the appropriate navigational support.

Therefore, while the current architecture offers support for interactive experiences developing through a predefined path (i.e., a number of works have emphasized the importance of offering navigational support through constrained navigation [Haik et al. 2002]), the extended architecture would allow to enlarge its usefulness also to interactive worlds where the user is given higher degrees of freedom, including guided tours based on dynamic paths and free motion across the environment.

Concerning the current implementation, we have chosen Java for its portability and the presence of many libraries for manipulating the XML files (e.g., the Jena libraries [McCarthy 2004] for making queries in RDQL language [RDQL 2004] and the Jdom libraries [JDO] for working with the X3D file).

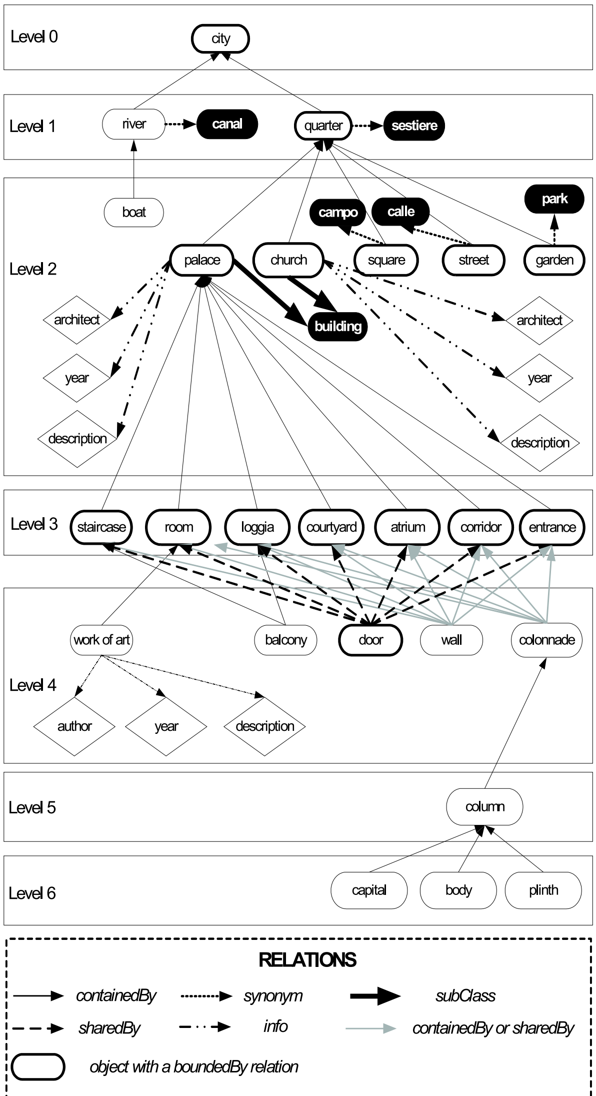

Figure 10: The ontology for the urban environments domain. A view centered on semantic objects and their relations.

5 A Voice-guided Tour through a Venetian Palace

In this section we'll describe a case study related to a particular urban environment: the city of Venice. We'll show how to obtain, using the concepts and the implementation architecture described above, a natural language description of a tour through an ancient Venetian palace hosting an art exhibition. Such description, obtained from the semantic description of the building (compliant with an ontology for urban environments) and from a low-level description of a path inside the palace, will guide the user inside the different rooms of the environment. The multi-level description will enable the user to perform semantic *zoom in* and *zoom out* operations, for having more detailed information about the lower-level semantic objects contained inside the rooms (e.g., detailed descriptions of the architectural components of a column) or to zoom out for having general information about the context (e.g. the palace or the *sestiere* (i.e., a synonym for quarter, used in Venice) to which the building belongs). Besides, the user will be given high-level information for navigating across the different rooms following the path.

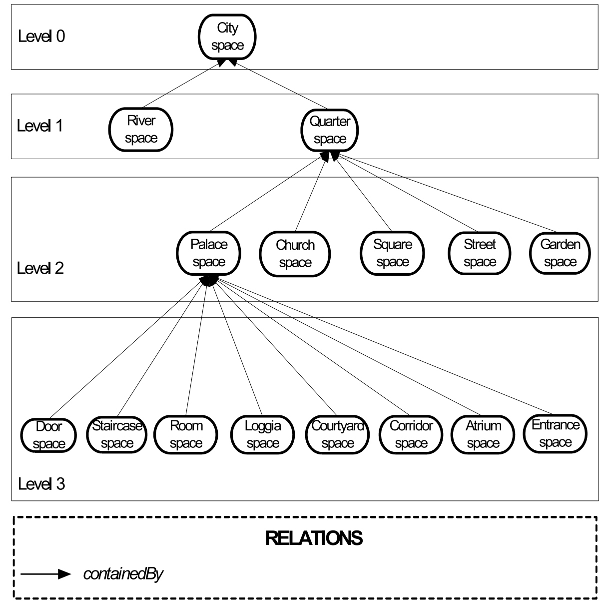

Figure 11: The ontology for the urban environment domain. A view centered on semantic zones and their relations.

Figures 10 and 11 illustrate two views related to the urban environments ontology that we have built and we have used for the case study. Generally speaking, an ontology is described by a graph that is not easy to see as a whole. That is the reason why, for the sake of simplicity, we have described our ontology using two partial views: the first one (Figure 10) showing the relations between the semantic objects and their relations with the semantic spaces; the second one (Figure 11) illustrating the relations between such semantic spaces. Besides, for the sake of clarity, not all the relations have been evidenced, but only those necessary to the understanding of the example, the Venetian palace illustrated in Figure 8.

Figure 10 shows that an high number of semantic levels are needed to describe the elements belonging to a built environment; besides, a variety of relations among the different semantic objects populates the scheme. In particular, comparing such ontology with that one illustrated in Figure 5, the following additional relations have been defined: *info* (for descriptive content of the semantic objects), *synonym* (for storing the synonyms of a certain class), *subClass* (for assessing that some items are subclasses of a more general class of semantic objects) and *boundedBy* (for marking the semantic objects with an associated semantic zone).

For clarity, the relations between the semantic zones are described in Figure 11. Concerning the relation of containment (*containedBy*), the semantic zones are organized in a hierarchy. Please note that a relation of containment defined for a semantic object doesn't necessarily imply the same relation for the associated semantic zone; for example the class *door* is contained by the class *room*, but the class *door space* is not contained by the class *room space*; instead it is contained by the higher level class *palace space*.

Figure 12 illustrates the plan of the Venetian palace used for the example, showing the space organization and the locations of the works of art exposed. The dotted path leads the visitor from the entrance facing an external square to the interior of the palace. The morphology of the palace is complex and includes a variety

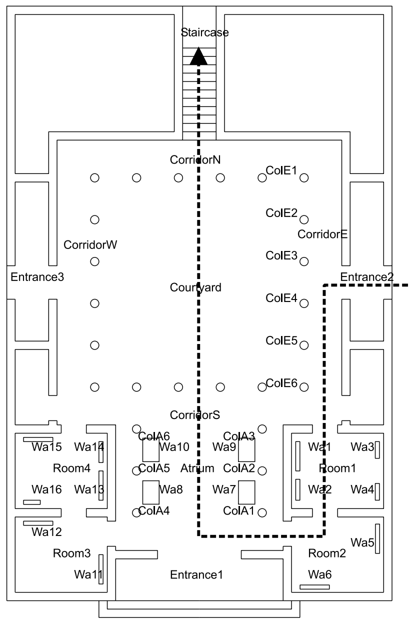

Figure 12: The plan of the Venetian palace and the low-level path.

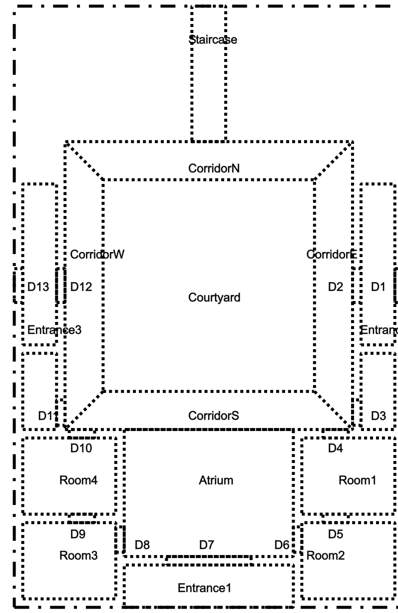

Figure 13: A scheme showing the set of nested semantic zones (ILs) associated to the building.

of spaces, ranging from a magnificent atrium to an ample set of rooms that are used for exhibiting the collection of works of art.

Figure 13 evidences the semantic spaces that characterize the palace; the lower-level spaces such as the *rooms* and the *corridors* are contained by the semantic space associated to the palace itself; all the semantic spaces are categorized as *connection* or *action space*, according to the ontology adopted. In some cases, such as the *atrium*, the role is twofold (i.e., the atrium is used both for exposing works of art and for connecting other rooms).

5.1 XML and VoiceXML Descriptions

According to the implementation architecture described in Section 4.1, the main process of the application will generate a high-level description of the tour. The code fragment shown in Figure 14 is the result of the processing of data related to the semantic description of the Venetian palace and of the low-level description of the path evidenced in Figure 12 (i.e., of course a different semantic environment and/or a different path would have originated a totally different XML file). The high-level XML description begins with a nested declaration of the semantic zones that characterize the ontology, starting from the highest level (i.e., the *city space*); a sequence of mid-level semantic zones (e.g., the *door space*, the *room space*, etc.) follows. Each zone embeds: the information related to the associated mid-level semantic object (e.g., the object *atrium* with associated description for the semantic zone *atrium space*), the description of the navigation steps for moving across the zone according to the low-level path requirements and the description of the low-level semantic objects contained (e.g., *works of art*, *colonnade*, etc.).

The description of the navigation steps includes values such as *right* or *straight* followed by a measure of distance expressed in meters; such information is not directly available in the semantic description of the environment but is computed by one of the application components on the basis of the X3D geometry and the specific low-level path (see Figure 9(4)).

```
<?xml version="1.0"?>
<citySpace name="Venice" type="space" ...>
 <city description="" type="object"
  containedBy="Venice">
 <quarterSpace name="San Marco" type="space" ...>
  <quarter description="" type="object"
   containedBy="San Marco">
  <palaceSpace name="Palazzo Grande"
    type="space" ...>
   <palace description="" type="object"
    containedBy="Palazzo Grande" >
   <doorSpace name="D6" type="space"
    kind="connection space">
    <door description="" type="object"
     containedBy="D6">
     <navigation direction="straight"
      lenght="0.5" unit="meters"/>
    </door>
   </doorSpace>
   <atriumSpace name="Atrium" type="space"
    kind="connection space | action space">
    <atrium description="" type="object"
     containedBy="Atrium">
     <navigation direction="straight"
      lenght="4" unit="meters"/>
     <navigation direction="right"/>
     ...
     <workOfArt position="left" name="Wa7"
      description="A statue" type="object"/>
     ...
     <colonnade position="left"
      description="" type="object">
      <column name="ColA1" description=""
       type="object">
       <capital description="marble Doric capital"
        type="object"/>
       <body description="" type="object"/>
       <plinth description="" type="object"/>
```

```
      </column>
      ...
     </colonnade>
     <colonnade position="right" description=""
      type="object">
     ...
     </colonnade>
     ...
     <wall position="left" name="Wall01"
      description="" type="object"/>
     ...
     <navigation direction="straight"
      lenght="7.5" unit="meters"/>
    </atrium></atriumSpace>
   </palace></palaceSpace>
  </quarter></quaterSpace>
 </city></citySpace>
```

Figure 14: XML description of a tour through a Venetian palace.

The description of the semantic objects (e.g., *work of art*, *wall*, etc.) includes an annotation (e.g., *left*, *infrontof*, etc.) about their position in relation to the user. Also in this case such annotation is computed by the application component (Figure 9(4)) on the basis of the user position and orientation along the low-level path. Therefore the current limit of the approach is that such information is valid for a specific position, reached by the user that has followed the specific sequence of navigation steps suggested by the tour.

Finally, the description of semantic objects is multi-level; related information is nested following the hierarchy determined by the *containedBy* relation of the ontology (e.g. a *capital* is contained by a *column*, that is contained by a *colonnade*).

The XML file contains an abstract representation of the guided tour that at the end of the process must be presented on a device (a smartphone, a voice browser, etc.). Therefore the XML file needs to be converted in a suitable format. In our case study we'll convert the XML description into a set of VoiceXML files for presentation on a voice browser (i.e., VoiceXML is an XML language for writing Web pages users can interact with by listening to spoken prompts, and control by means of spoken input; further information can be found in [Raggett 2001]).

```
<vxml>
 ...
<form>
 <block>
 <prompt> You are entering into the Atrium.</prompt>
 </block>
</form>

<menu>
 <prompt>
  Do you want to hear the description of the Atrium?
 </prompt>
 <choice
  next="http://www.unive.it/voiceXML/atrium.vxml">
  yes</choice>
 <choice
  next="http://www.unive.it/voiceXML/goon.vxml">
  no</choice>
 </menu>

<form>
<block>
<prompt> Go straight for about 4 meters. </prompt>
<prompt> Turn right. </prompt>
</block>
</form>

<menu>
 <prompt> The Atrium is composed by
          one, a colonnade;
          ...
          three, a work of art Wa07;
          ...
          Do you want some information about them?
          Please say the corresponding number
          or say 'go on' if you want to proceed.
 </prompt>
 <choice
  next="http://www.unive.it/voiceXML/colA1.vxml">
  one</choice>
 ...
 </menu>
</vxml>
```

Figure 15: VoiceXML description of a tour through a Venetian palace: the atrium.

Figure 15 shows a fragment of the VoiceXML file related to the navigation and description of the atrium; the user, arriving from the previous room, is informed that s/he's entering a new zone; then s/he's invited to do some navigation steps in order to face the center of the zone; then the user can choose to have detailed information about the objects contained in the atrium or to proceed. If the user pronounces the word *colonnade* a new VoiceXML file containing the description of the *colonnade* and of its components (*columns*) will be presented. A request of additional details will generate the presentation of a new file with information about the *column* and its components (*capital*, *body* and *plinth*).

6 Conclusion

Work performed so far has led to the proposal of an approach for describing semantic 3D worlds using available web standards. Such approach is characterized by the use of scene-independent ontologies, by the complete description of the environment thanks to the notion of semantic objects and semantic zones and by the multilevel description of the scene.

An application scenario has been illustrated, showing how semantic description of 3D worlds can be used for offering navigational support to users. Future work will consider the refinement of the implementation architecture proposed in this paper (including usability tests) and its extension for enlarging its usefulness to a wider number of navigation metaphors. While the current solution offers navigational support for guided tours generated on the basis of a static initial path, future work will consider support for guided tours based on dynamic paths and for navigation of users wandering through the environment with no constrain or specific target. In particular future work related to dynamic paths will consider support for different classes of users, including those categories that don't depart intentionally from the predefined path and need assistance for recovering the situation (e.g., blind users moving in a real environment).

Besides, the definition of a scene-independent ontology allows the exploration of a variety of scenarios that will be considered in future work, including authoring and validation of 3D worlds conforming to specific ontologies or the realization of search engines capable of querying sets of semantic 3D worlds on the net.

References

BILASCO, I. M., GENSEL, J., VILLANOVA-OLIVER, M., AND MARTIN, H. 2005. On indexing of 3d scenes using mpeg-7. In *Proceedings of the 13th annual ACM international conference on Multimedia*, ACM Press, 471–474.

CHITTARO, L., AND IERONUTTI, L. 2004. A visual tool for tracing users behavior in virtual environments. In *Proceedings of AVI 2004: 7th International Conference on Advanced Visual Interfaces*, ACM Press, 40–47.

DACHSELT, R. 2000. Action spaces - a metaphorical concept to support navigation and interaction in 3d interfaces. In *Proceedings of the Workshop on Usability Centred Design and Evaluation of Virtual 3D Environments*, Shaker Verlag.

HAIK, E., BARKER, T., SAPSFORD, J., AND TRAINIS, S. 2002. Investigation into effective navigation in desktop virtual interfaces. In *Proceedings of Web3D 2002: 7th International Conference on 3D Web Technology*, ACM Press, 5966.

HALABALA, P. 2003. Semantic metadata creation. In *Proceedings of CESCG 2003: 7th Central European Seminar on Computer Graphics*, 15–25.

JDOM Homepage. http://www.jdom.org/.

MANSOURI, H. 2005. *Using Semantic Descriptions for Building and Querying Virtual Environments*. PhD thesis, Vrije Universiteit Brussel.

MCCARTHY, P., 2004. Introduction to Jena. http://www-128.ibm.com/developerworks/java/library/j-jena/?ca=degr-LXXjenaintro.

MILGRAM, P., AND KISHINO, F. 1994. A taxonomy of mixed reality visual displays. *Transactions on Information Systems*, 12, 1321–1329.

MPEG Homepage. http://www.chiariglione.org/mpeg/.

NACK, F., AND LINDSAY, A. T. 1999. Everything you wanted to know about mpeg-7 - part 1. *IEEE Multimedia 6*, 3, 65–77.

NACK, F., AND LINDSAY, A. T. 1999. Everything you wanted to know about mpeg-7 - part 2. *IEEE Multimedia 6*, 4, 64–73.

ONTOWORLD Homesite. http://www.ontoworld.org/.

OWL, 2004. Web Ontology Language Guide. http://www.w3.org/TR/owl-guide/.

PEREIRA, F., AND EBRAHIMI, T. 2002. *The MPEG-4 Book*. Prentice-Hall.

PITTARELLO, F., AND FOGLI, D. 2005. Modelling complex user experiences in distributed interaction environments. In *Proceedings of DMS 2005: 11th International Conference on Distributed Multimedia Systems*, Knowledge Systems Institute, 184–189.

PITTARELLO, F. 2003. Accessing information through multimodal 3d environments: towards universal access. *Universal Access in the Information Society 2*, 2, 189–204.

RAGGETT, D., 2001. Getting started with voicexml 2.0. http://www.w3.org/Voice/Guide/.

RDF, 2004. RDF Primer W3C Recommendation. http://www.w3.org/TR/rdf-primer/.

RDQL, 2004. RDQL - A Query Language for RDF. http://www.w3.org/Submission/RDQL/.

W3C, 2001. Semantic Web. http://www.w3.org/2001/sw/.

WEB3D CONSORTIUM, 2004. Extensible 3D (X3D) ISO/IEC 19775:2004. http://www.web3d.org/x3d/specifications/ISO-IEC-19775-X3DAbstractSpecification/.

XSLT, 1999. XSL Transformations (XSLT) Recommendation. http://www.w3.org/TR/xslt/.

YATES, F. A. 1966. *The Art of Memory*. University of Chicago Press.