Four-valued Knowledge Augmentation for Representing Structured Documents

Mounia Lalmas and Thomas Roelleke

Department of Computer Science, Queen Mary, University of London
London, E1 4NS, England
{mounia,thor}@dcs.qmul.ac.uk
tel: +44 (0)20 7882 52 00 fax: +44 (0)20 8980 6533
web: qmir.dcs.qmul.ac.uk

Abstract. Structured documents are composed of objects with a content and a logical structure. The effective retrieval of structured documents requires models that provide for a content-based retrieval of objects that takes into account their logical structure, so that the relevance of an object is not solely based on its content, but also on the logical structure among objects. This paper proposes a formal model for representing structured documents where the content of an object is viewed as the knowledge contained in that object, and the logical structure among objects is capture by a process of knowledge augmentation: the knowledge contained in an object is augmented with that of its structurally related objects. The knowledge augmentation process takes into account the fact that knowledge can be incomplete and become inconsistent. The model is based on the definitions of four truth values, modal operators, possible worlds, accessibility relations and truth value assignments used to characterise content knowledge, structural knowledge, augmented knowledge, incompleteness and inconsistency.

1 Introduction

We view a structured document as being composed of *objects* with a content and a logical structure [17, 1, 16, 18, 5]. These objects correspond to the document components. The content refers to the content of objects. For example, the content of an object can be described by the term "sailing". The logical structure refers to the way structured documents are organised. For example, a structured document may be composed of objects such as sections and figures, which can themselves be composed of other objects (e.g. a section and it subsections). The *root* object, which is unique, embodies the whole document. *Atomic* objects are document components that are not composed of other components. All other objects are referred to as *inner* objects.

With the widespread development of structured document repositories (e.g. CD-ROM, the Internet and digital libraries), there is a need to develop retrieval models that dynamically return objects of varying granularity (e.g. [8, 7, 13, 12, 3, 15, 10]). A retrieval result may then consist of entry points to a same document: an entry to the root object when the entire document is relevant to a query, an entry to an atomic object when only that object is relevant to the query, or an entry to an inner object when that object and its structurally related objects are relevant to the query. Returning objects of varying granularity requires sophisticated models that go beyond the purely term-based models of classical information retrieval, so that the relevance of an object is not solely based on its content, but also takes into account the logical structure among objects.

This paper proposes a formal model for representing structured documents that allows the retrieval of objects of varying granularity. The model exploits the content and the logical structure among objects to arrive at a representation of the structured document. The content of an object is viewed as the *knowledge* contained in that object, and the structure among objects is captured by a process of *knowledge augmentation*: the knowledge contained in an object is *augmented* with the knowledge contained in its structurally related objects. The knowledge augmentation process is a combination of knowledge specifically defined to provide for a representation of the object that is based on its own content and that of its structurally related objects. For instance, a non-atomic object composed of an object about "sailing" and a second object about "boat" can be considered as being about "sailing and boat". The knowledge contained in the non-atomic objects (its content description) is augmented with that of its related objects. It is this representation that can then be used to estimate the relevance of the object to a query. The estimate takes into account that non-atomic objects are structurally related to other objects. In our example, the non-atomic object should be ranked higher than its related objects for a query about "sailing and boats". In this paper, we are concerned with the representation of structured documents, in particular the knowledge augmentation process.

The knowledge augmentation process takes into account the fact that knowledge can be incomplete and become inconsistent. Incompleteness is due to the fact that knowledge that is not explicit in an object should not be considered false. For instance, an object composed of one object about "sailing" and a second object where nothing is said about "sailing" (incomplete description) should be considered as being about "sailing". Inconsistent arises when two objects upon which the augmentation process is based contain knowledge that is contradictory. For example, an object composed of one object about "sailing" and a second object about "not sailing" should not be considered as being about "sailing" nor "not sailing" since there is contradictory evidence.

The model is based on the definition of modal operators, referred to as *knowledge modal operators*, one for each of the objects forming a structured document. The knowledge modal operator associated with an object is used to model the knowledge contained in the object, which can be of two types: content knowledge (i.e. propositions describing the information content of the object) and structural knowledge (i.e. the fact that the object is structurally related to other objects). The semantics of the knowledge modal operators is based on an interpretation structure and an interpretation function defined upon four truth values, possible worlds, accessibility relations and truth value assignment functions [6, 9]. Possible worlds and truth value assignment functions reflect the knowledge contained in objects. The accessibility relations captures structural relationships between objects. Four truth values are used to capture incompleteness and inconsistency [4].

We follow two routes to formalise the knowledge augmentation process. The first approach defines a different set of modal operators, referred to as *augmented knowledge modal operators*, one for each of the objects forming a structured document. An augmented knowledge operator is used to model the content knowledge of an object augmented with that of its structurally related objects. The semantics of the augmented knowledge modal operators is defined upon the interpretation structure and interpretation function defining the semantics of the knowledge modal operators. The second approach, which follows the framework of [9] for modelling combination of knowledge, is based on the definition of *G-world trees* that formalise a graphical representation of the set of possible worlds and accessibility relations associated with the representation of objects. Special truth value assignment functions, referred to as *augmented truth value assignment functions*, defined upon G-world trees are used to characterise the augmented knowledge. We show that the two approaches are equivalent, that is, the same representation of objects is obtained after the knowledge augmentation process.

The outline of this paper is as follows. Section 2 describes the model for representing content and structural knowledge of objects. Section 3 and 4 present the two approaches for describing the knowledge augmentation process. Section 5 shows that the two approaches are equivalent. Section 6 concludes and discusses future work.

2 Representing content and structural knowledge

The syntax used to characterise the content and structural knowledge of a structured document is given in Figure 1. A structured document is described by a *program*, which is a set of *clauses* where a clause is either a *fact* or a *context*. A *context* consists of an object identified by an object name, in which a program is nested. The object name in a context clause is referred to as a *context name*. We use the context name "this" to refer to the context name associated with the global program (i.e. representing the structured document). The set of object names is called $\mathcal{C}$. A *fact* is a *proposition*, or a proposition preceded by "not" or *truth-list*. We call Φ the set of propositions. *truth-list* represents the four truth values: "1" means *true*, "0/1" means *false*, "0/0/1" means *inconsistent*, and "0/0/0/1" means *unknown*. In this paper, "1 proposition" and "0/1 proposition" are the same as "proposition" and "not proposition", respectively[1].

```
program     ::= clause | clause program
clause      ::= fact | context
fact        ::= proposition | 'not' proposition | truth-list proposition
proposition := NAME
truth-list  ::= '1' | '0/1' | '0/0/1' | '0/0/0/1'
context     ::= NAME '[' program ']'
NAME        ::= [a-z][a-zA-Z0-9_]*
```

Fig. 1. Syntax of programs

Let d and s be context names in $\mathcal{C}$ and "sailing" be a proposition in Φ. A context clause is called an *atomic context* clause when it contains only facts, but no context clauses. For instance, "d[sailing]" is an atomic context clause and expresses that the content of the object modelled by the context name d is described by the proposition "sailing". Otherwise, a context clause is called a *structured context* clause. For instance, "d[s[sailing]]" is a structured context clause and expresses that the object represented by d is composed of the object represented by s, and the content of the object modelled by s is described by the proposition "sailing". We refer to s as *a subcontext* of d, and d as *the supercontext* of s. A supercontext can have several subcontexts, whereas a subcontext has exactly one supercontext. Atomic context clauses characterise atomic objects, whereas structured context clauses characterise non-atomic objects (i.e. inner and root objects). Finally, a function $sub : \mathcal{C} \to \wp(\mathcal{C})$ yields the set of subcontexts of a supercontext for a given program.

[1] We use both "0/1" and "not" because the proposed representation is currently being extended to allow for a probabilistic-based representation of objects, thus capturing the uncertainty inherent to the information retrieval process.

Figure 2 is a commented example of a program (the comments appear after %). The structured document is composed of two objects represented by the structured context d and the atomic context s. d is a subcontext of the global context "this" representing the whole document. s is a subcontext of d. The context s has two propositions "not boats" and "sailing". Finally, $sub(\text{this}) = \{\text{d}\}$ and $sub(\text{d}) = \{\text{s}\}$.

```
d[       % structured context
   s[    % atomic context
      not boats              % same as 0/1 boats
      sailing                % same as 1 sailing
   ]     % end of context s
]        % end of context d
```

Fig. 2. Example of a program

The semantics of programs is defined upon an *interpretation structure* denoted M and an *interpretation function* denoted $\models$ that assigns truth values to programs, clauses, contexts, facts and propositions. We consider contexts as "agents" that possess *knowledge* concerning their content and their structure and use a Kripke structure (see [9]) to define the semantics of programs. Different from [9], we use the terminology "context" instead of "agent" and we consider four truth values. We first define an interpretation structure with respect to a set of propositions Φ and a set of context names $\mathcal{C}$.

Definition 1. (Interpretation structure M): *An* interpretation structure M *for a set Φ of propositions and a set $\mathcal{C}$ of context names is a tuple $M = (W, \pi, \mathcal{R})$ where*

- $W := \{w_d : d \in \mathcal{C}\} \cup \{w_{slot}\}$ *is a finite set of possible worlds. For each context name, a possible world is defined. The set includes the possible worlds w_{this} for the global context "this" and w_{slot} to model the access to the "this" context.*
- $\pi : W \rightarrow (\Phi \rightarrow \{\text{true}, \text{false}, \text{inconsistent}, \text{unknown}\})$ *is a function on W that yields a truth value assignment for all world w in W, which is a function $\pi(w) : \Phi \rightarrow \{\text{true}, \text{false}, \text{inconsistent}, \text{unknown}\}$ that assigns a truth value to each proposition in Φ, where the four truth values are defined as* true$:=\{t\}$, false$:=\{f\}$, inconsistent$:=\{t, f\}$, *and* unknown$:=\{\}$.
- $\mathcal{R} := \{R_d : \text{d} \in \mathcal{C}\}$ *is a finite set of binary relations on W, called accessibility relations, one for each context name. For any supercontext* d *and subcontext* s, *i.e.* $\text{s} \in sub(\text{d})$, $\{(w_d, w_s)\} \in R_s$ *where R_s is the accessibility relation associated to* s, *and w_d and w_s are the possible worlds associated with* d *and* s, *respectively. We say that context* s accesses *or* reaches *world w_s from world w_d. The accessibility relation for the global context "this" is defined as $R_{this} := \{(w_{slot}, w_{this})\}$.*
- *The function $R_d(w) := \{w' | (w, w') \in R_d\}$ for $w \in W$ and $R_d \in \mathcal{R}$ yields the set of worlds that can be reached from world w by the context* d.

Consider the program "d[s[sailing]]". d and s are context names and their respective possible worlds are w_d and w_s. To represent that the proposition "sailing" appears within s but neither in d or the global context "this" (i.e. the root object), the truth values assigned to sailing are as follows: $\pi(w_s)(\text{sailing}) = \textit{true}$, $\pi(w_d)(\text{sailing}) = \pi(w_{this})(\text{sailing}) = \pi(w_{slot})(\text{sailing}) = \textit{unknown}$. *unknown* captures incompleteness since it is not explicitly stated that sailing does not describe the content of d and the global context. The logical structure is represented through the accessibility relations $R_s = \{(w_d, w_s)\}$, $R_d = \{(w_{this}, w_d)\}$ and $R_{this} = \{(w_{slot}, w_{this})\}$.

The semantics of the content and structural knowledge of objects is based upon considering context names as modal operators, referred to as *knowledge modal operators*. The atomic context clause "d[sailing]" becomes interpreted as "d knows sailing" and captures the content knowledge of the corresponding object. The structured context clause "d[s[sailing]]" becomes interpreted as "d knows that s knows sailing" and captures the structural knowledge of the object represented by d. With this interpretation in mind, we define the interpretation function $\models$ that assigns truth values to propositions, facts, atomic and structured context clauses, and programs with respect to the interpretation structure defined above.

Definition 2. (Interpretation function $\models$): *Let $M = (W, \pi, \mathcal{R})$ be a interpretation structure as defined in Definition 1. Let φ be a proposition in Φ and w a world in W. Let* d *be a context name in $\mathcal{C}$ and $R_d \in \mathcal{R}$ its corresponding accessibility relation. Let* s *be a context name in $\mathcal{C}$. The interpretation of facts is defined as follows:*

$$(M, w) \models 1\ \varphi :\Longleftrightarrow\ \pi(w)(\varphi) = \text{true}$$
$$(M, w) \models 0/1\ \varphi :\Longleftrightarrow\ \pi(w)(\varphi) = \text{false}$$
$$(M, w) \models 0/0/1\ \varphi :\Longleftrightarrow\ \pi(w)(\varphi) = \text{inconsistent}$$
$$(M, w) \models 0/0/0/1\ \varphi :\Longleftrightarrow\ \pi(w)(\varphi) = \text{unknown}$$

The interpretation of atomic context clauses is defined as follows:

$$(M, w) \models \mathrm{d}[1\ \varphi] :\Longleftrightarrow\ \forall w' \in R_d(w) : (M, w') \models 1\ \varphi$$
$$(M, w) \models \mathrm{d}[0/1\ \varphi] :\Longleftrightarrow\ \forall w' \in R_d(w) : (M, w') \models 0/1\ \varphi$$
$$(M, w) \models \mathrm{d}[0/0/1\ \varphi] :\Longleftrightarrow\ \forall w' \in R_d(w) : (M, w') \models 0/0/1\ \varphi$$
$$(M, w) \models \mathrm{d}[0/0/0/1\ \varphi] :\Longleftrightarrow\ \forall w' \in R_d(w) : (M, w') \models 0/0/0/1\ \varphi$$

The interpretation of structured context clauses is defined as follows::

$$(M, w) \models \mathrm{d}[\mathrm{s}[\mathrm{fact}]] :\Longleftrightarrow\ \forall w' \in R_d(w) : (M, w') \models \mathrm{s}[\mathrm{fact}]$$

The interpretation of a program within a context d *is defined as follows, where* program* *is the set of clauses given in the program:*

$$(M, w) \models \mathrm{d}[\mathrm{program}] \iff \forall \mathrm{clause} \in \mathrm{program}^* : (M, w) \models \mathrm{d}[\mathrm{clause}]$$

An interpretation structure M is called a model *of a program P iff* this[P] *is true in all worlds with respect to M. We use "valid" for "true in all worlds". The notation $M \models$* this[P] *reads "*this[P] *is valid".*

Returning to our example, for an interpretation M to be a model of the program "this[d[s[sailing]]]", we need to show that for all worlds $w \in W$, $(M, w) \models$ this[d[s[sailing]]]. These worlds are w_{slot}, w_{this}, w_d and w_s. We have $\pi(w)(\text{sailing}) = \textit{true}$ in w_s and unknown in all other worlds. We start with w_{slot}. For $(M, w_{slot}) \models$ this[d[s[sailing]]] to hold, we must verify that $(M, w_{this}) \models$ d[s[sailing]] holds since w_{this} is the only world accessible from w_{slot} with respect to R_{this}. For $(M, w_{this}) \models$ d[s[sailing]] to hold, we must verify that $(M, w_d) \models$ s[sailing] holds, which holds if $(M, w_s) \models$ sailing, which is holding from the truth value assignment function $\pi(w_s)$. In all other worlds w_{this}, w_d and w_s, the context "this" cannot reach another world (i.e., through the accessibility relation R_{this}), hence "this[d[s[sailing]]]" is true in these worlds. Consequently, M is a model of the above program.

Two properties often associated to knowledge are (see [9]) the *knowledge generalisation rule* and the *truth axiom of knowledge*. The knowledge generalisation rule states that knowledge present in all worlds is knowledge of all contexts. Formally:

$$\text{if } M \models \varphi \text{ then } M \models \mathrm{d}[\varphi]$$

where $M \models \varphi$ means that φ is true in all worlds (φ is valid). If φ is valid, then d knows it to be true in all worlds. Let M be a model of the program "d[sailing s[boats]]" (i.e. $M \models$ this[d[sailing s[boats]]]). From Definition 2, we have $M \models$ d[sailing]. Using the knowledge generalisation rule, we could derive that $M \models$ s[d[sailing]], which does not reflect the logical structure of the program. To overcome this problem, we introduce the notion of *context-validity* for structured context clauses:

$$M \models_C \mathrm{d}[\mathrm{s}[\varphi]] \iff M \models \mathrm{d}[\mathrm{s}[\varphi]] \text{ and } \mathrm{s} \in sub(\mathrm{d})$$

A structured context is *context-valid* iff the structured context clause is valid and the logical structure of the program is reflected. The context-valid programs constitute a subset of the valid programs. The Definition 2 is extended as follows.

Definition 3. (Interpretation function $\models$ - Extended Definition): *An interpretation structure M is a model of a program iff all context clauses of the program are valid and all structured context clauses are context-valid.*

In the remainder of this paper, for simplicity, we use $\models$ to refer to a context-valid interpretation function.

The truth axiom of knowledge states that a context can only "know" what is true. Formally:

$$\text{if } (M, w) \models \mathrm{d}[\varphi] \text{ then } (M, w) \models \varphi$$

Consider the program "d[sailing]" and M as a model of this program. We have then from Definition 1 a set of possible worlds $W = \{w_{slot}, w_{this}, w_d\}$ and the accessibility relations $R_{this} = \{(w_{slot}, w_{this})\}$ and $R_d = \{(w_{this}, w_d)\}$, and truth value assignments $\pi(w_d)(sailing) = \mathit{true}$ and $\pi(w_{slot})(sailing) = \pi(w_{this})(sailing) = \mathit{unknown}$. It can be shown that by applying Definition 2 $(M, w_{this}) \models$ d[sailing] although $(M, w_{this}) \not\models$ sailing. Our knowledge modal operators do not follow the truth axiom of knowledge[2].

This concludes our formalism (its syntax and semantics based on a Kripke structure) for representing the content and structural description of the objects forming a structured document. Objects are viewed as agents, referred to as contexts in this paper, that possess knowledge about their own content and their structure. The formalism is next extended so that the knowledge of an object is augmented with that of its structurally related objects. We followed two routes, which we show to lead to equivalent formalisms in Section 5. In the first one, Section 3, two new modal operators are defined, one to characterise the knowledge associated with a group of context, and the second to characterise the knowledge associated with a context and its (group of) sub-contexts. In the second one, Section 4, special truth value assignments are defined based on a formalism of the graphical representation of the possible worlds and accessibility relations used to model a structured document.

[2] The truth axiom of knowledge is satisfied if the accessibility relations are reflexive. Our accessibility relations are not reflexive from the way they are constructed (see Definition 1).

3 Knowledge augmentation using modal operators

The semantics of programs as defined in the previous section formalises the content and structural knowledge of objects. The present section describes the formalisation of the knowledge augmentation process, necessary to allow for the retrieval of objects at varying granularity.

Consider the program "d[s1[sailing] s2[boats]]". A model for the program is based on an interpretation structure M defined upon a set of possible worlds $W = \{w_{slot}, w_{this}, w_d, w_{s1}, w_{s2}\}$, the accessibility relations $R_{this} = \{(w_{slot}, w_{this})\}$, $R_d = \{(w_{this}, w_d)\}$, $R_{s1} = \{(w_d, w_{s1})\}$ and $R_{s2} = \{(w_d, w_{s2})\}$, and the truth value assignments $\pi(w_{s1})(\mathit{sailing}) = \pi(w_{s2})(\mathit{boats}) = \mathit{true}$, and *unknown* in all other cases. Definition 2 yields $(M, w_{w_d}) \models$ s1[sailing] and $(M, w_d) \models$ s2[boats] characterising the knowledge content of s1 and s2 (s1 knows sailing and s2 knows boats), and $(M, w_{this}) \models$ d[s1[sailing]] and $(M, w_{this}) \models$ d[s2[boats]] characterising the structural knowledge of d (d knows that s1 knows sailing and s2 knows boats). Contexts s1 and s2 would be relevant to queries about "sailing" and "boats", respectively, but none of the contexts s1 and s2 would be relevant to a query about "sailing and boats". The supercontext d should be relevant to such a query because it knows that one of its subcontext contains sailing and the other contains boats. This could be inferred if the knowledge content of d is augmented with that of its subcontexts.

However, a knowledge augmentation process can lead to inconsistent knowledge. Consider the program "d[s1[sailing] s2[not sailing boats]]". Subcontext s1 knows sailing; subcontext s2 knows the opposite. The example makes evident that augmenting the content knowledge of d by that of s1 and s2 leads to inconsistent knowledge regarding the proposition "sailing", since we have evidence for true from s1 and false from s2. In the augmented context d(s1,s2), "sailing" is therefore inconsistent.

To distinguish between the content knowledge of a context and its *augmented content knowledge*, we introduce the terminology of *augmented context* as opposed to *basic context*. Basic contexts are context names (e.g. d, s1 and s2), whereas *augmented contexts* consist of a supercontext name and a list (group) composed of augmented contexts or basic contexts (e.g. d(s1(s11),s2)). A context clause with only basic contexts is called a *basic context clause* (e.g. "d[sailing]" is a basic atomic context clause and "d[s[sailing]]" is a basic structured context clause). A context clause with augmented contexts is called an *augmented context clause* (e.g. "d(s1,s2)[sailing]").

The knowledge augmentation process *combines* knowledge of a supercontext with that of its subcontexts. Modal operators have been defined to formalise specific combination of knowledge (e.g. common knowledge and distributed knowledge [9]). However, as discussed in [16], these operators are not appropriate for a combination of knowledge that arises from a knowledge augmentation process. We therefore define other modal operators.

The first operator is the *united knowledge modal operator* denoted U_G, which is used to represent the combined knowledge of a group of context $G = $ (s1,...,sn) referred to as a *united context*. Here, we are aiming at capturing the combined knowledge of a group of context, not the knowledge of an augmented context.

Definition 4. (United knowledge operator U_G): *Let $M = (W, \pi, \mathcal{R})$ be an interpreteation structure defined upon a set Φ of propositions and a set $\mathcal{C}$ of context names. Let w be a world in W, φ a proposition in Φ, and G a united context from*

$\mathcal{C}$.

$$
\begin{aligned}
(M,w) \models U_G[1\ \varphi] :\Longleftrightarrow\ & \exists \mathrm{s}\ \in G : (M,w) \models \mathrm{s}[1\ \varphi]\ \ \textit{and} \\
& \forall \mathrm{s}\ \in G : ((M,w) \models \mathrm{s}\ [1\ \varphi]\ \textit{or}\ (M,w) \models \mathrm{s}\ [0/0/0/1\varphi]) \\
(M,w) \models U_G[0/1\ \varphi] :\Longleftrightarrow\ & \exists \mathrm{s}\ \in G : (M,w) \models \mathrm{s}[0/1\ \varphi]\ \ \textit{and} \\
& \forall \mathrm{s}\ \in G : ((M,w) \models \mathrm{s}\ [0/1\ \varphi]\ \textit{or}\ (M,w) \models \mathrm{s}[0/0/0/1\varphi]) \\
(M,w) \models U_G[0/0/1\ \varphi] :\Longleftrightarrow\ & (\exists \mathrm{s}\ \in G : (M,w) \models \mathrm{s}[1\ \varphi]\ \textit{and} \\
& \ \exists \mathrm{s} \in G : (M,w) \models \mathrm{s}[\ 0/1\ \varphi])\ \ \textit{or} \\
& \exists \mathrm{s}\ \in G : (M,w) \models \mathrm{s}[0/0/1\ \varphi] \\
(M,w) \models U_G[0/0/0/1\ \varphi] :\Longleftrightarrow\ & \forall \mathrm{s}\ \in G : (M,w) \models \mathrm{s}[0/0/0/1\ \varphi]
\end{aligned}
$$

The proposition φ is true in a united context G iff there exists a context s in G that knows φ to be true and all contexts know φ to be true or unknown. The definition for false is analogous. The proposition φ is inconsistent iff there exist contexts that know φ to be true and false, or there exists a context in G that knows φ to be inconsistent. The proposition φ is unknown iff all contexts in G know φ to be unknown. For our example, we would obtain $(M,w_d) \models U_{(\mathrm{s1},\mathrm{s2})}$ [boats], $(M,w_d) \models U_{(\mathrm{s1},\mathrm{s2})}$ [sailing].

We define now the knowledge of augmented contexts through the introduction of an *augmented knowledge modal operator* $A_{\mathrm{d}G}$ where dG is an augmented context (d is the supercontext and G is the united context formed with the subcontexts of d, i.e. $G =(\mathrm{si} : \mathrm{si} \in sub(\mathrm{d}))$). The knowledge of the united context is combined with the knowledge of the supercontext. The augmented knowledge modal operator is therefore defined upon the united knowledge modal operator.

Definition 5. (Augmented knowledge operator $A_{\mathbf{d}G}$): *Let $M = (W,\pi,\mathcal{R})$ be an interpretation structure defined upon a set Φ of propositions and a set $\mathcal{C}$ of context names. Let w be a world in W, φ a proposition in Φ,* d, s1, ..., sn *contexts in $\mathcal{C}$ such that $sub(\mathrm{d}) = \{\mathrm{s1},\ldots,\mathrm{sn}\}$, and R_d the accessibility relation in $\mathcal{R}$ associated with the supercontext* d.

$$
\begin{aligned}
& (M,w) \models A_{\mathrm{d}(\mathrm{s1},\ldots,\mathrm{sn})}[1\ \varphi] :\Longleftrightarrow \\
& \quad ((M,w) \models \mathrm{d}[1\ \varphi]\ \textit{and} \\
& \quad\quad \forall w' \in R_d(w) : ((M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[1\ \varphi]\ \textit{or}\ (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[0/0/0/1\ \varphi]))\ \textit{or} \\
& \quad ((M,w) \models \mathrm{d}[0/0/0/1\ \varphi])\ \textit{and} \\
& \quad\quad \forall w' \in R_d(w) : (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[1\ \varphi])
\end{aligned}
$$

$$
\begin{aligned}
& (M,w) \models A_{\mathrm{d}(\mathrm{s1},\ldots,\mathrm{sn})}[0/1\ \varphi] :\Longleftrightarrow \\
& \quad ((M,w) \models \mathrm{d}[0/1\ \varphi]\ \textit{and} \\
& \quad\quad \forall w' \in R_d(w) : ((M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[0/1\ \varphi]\ \textit{or}\ (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[0/0/0/1\ \varphi]))\ \textit{or} \\
& \quad ((M,w) \models \mathrm{d}[0/0/0/1\ \varphi])\ \textit{and} \\
& \quad\quad \forall w' \in R_d(w) : (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[0/1\ \varphi])
\end{aligned}
$$

$$
\begin{aligned}
& (M,w) \models A_{\mathrm{d}(\mathrm{s1},\ldots,\mathrm{sn})}[0/0/1\ \varphi] :\Longleftrightarrow \\
& \quad ((M,w) \models \mathrm{d}[0/0/1\ \varphi]\ \textit{or} \\
& \quad\quad \forall w' \in R_d(w) : (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[0/0/1\ \varphi])\ \textit{or} \\
& \quad ((M,w) \models \mathrm{d}[1\ \varphi]\ \textit{and} \\
& \quad\quad \forall w' \in R_d(w) : (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[0/1\ \varphi])\ \textit{or} \\
& \quad ((M,w) \models \mathrm{d}[0/1\ \varphi]\ \textit{and} \\
& \quad\quad \forall w' \in R_d(w) : (M,w') \models U_{(\mathrm{s1},\ldots,\mathrm{sn})}[1\ \varphi])
\end{aligned}
$$

$$(M, w) \models A_{\text{d(s1,...,sn)}}[0/0/0/1\ \varphi] :\Longleftrightarrow$$
$$(M, w) \models \text{d}[0/0/0/1\ \varphi]\ \mathit{and}$$
$$\forall w' \in R_d(w) : (M, w') \models U_{\text{(s1,...,sn)}}[0/0/0/1\ \varphi]$$

The proposition φ is true in an augmented context d(s1,...,sn) iff (i) φ is true in the supercontext d and the united context (s1,...,sn) gives evidence for true or unknown, (ii) φ is unknown in the supercontext and the united context gives evidence for true. The definition for false is analogous. The proposition φ is inconsistent in d(s1,...,sn) iff (i) φ is inconsistent in the supercontext or the united context gives evidence for inconsistent, (ii) φ is true (false) in the supercontext and the united context gives evidence for false (true). The proposition φ is unknown in d(s1,...,sn) iff φ is unknown in the supercontext and the united context gives evidence for unknown.

Applying Definition 5 to our program "d[s1[sailing] s2[boats]]", we obtain $(M, w_{this}) \models A_{\text{d(s1,s2)}}$ [sailing] and $(M, w_{this}) \models A_{\text{d(s1,s2)}}$ [boats] because $(M, w_d) \models U_{\text{(s1,s2)}}$ [sailing], $(M, w_d) \models U_{\text{(s1,s2)}}$ [boats]. By augmenting the content knowledge of d by that of its subcontexts s1 and s2, we arrive at a representation of the object represented by the context name d that can be assessed relevant to a query about "sailing and boats".

The definitions of the united and augmented knowledge modal operators allow the propagation of inconsistent knowledge. Inconsistency in a subcontext of a group is propagated to the united knowledge of the group, and inconsistency in the supercontext or the group of an augmented context leads to inconsistent knowledge in the augmented context. For example, a proceedings (the supercontext) may contain two documents (the subcontexts) that contradict each other. The augmented knowledge of the supercontext is inconsistent. Combining this proceedings with another (e.g. to form a digital library) preserves this inconsistency.

4 Knowledge augmentation using truth value assignment functions

In the previous section, the knowledge augmentation process was defined through the introduction of modal operators. We defined two modal operators for representing united knowledge and augmented knowledge. The semantics of the knowledge modal operators is based on an interpretation structure defined upon a set of possible worlds, truth value assignment functions and accessibility relations. In this section, we define the knowledge augmentation process based on a graphical representation of worlds and accessibility relations formalised through the notions of G-world-trees. This follows the approach adopted in [9], where combination of knowledge was defined upon the notion of "G-reachability". Two truth value assignment functions are defined upon G-world trees to characterise united knowledge and augmented knowledge.

The logical structure of a non-atomic object can be described as trees, referred to as *G-world-trees*. These tree structures are also reflected in the way possible worlds and accessibility relations are defined with respect to contexts.

An empty united context is denoted "()", and for a context name s in $\mathcal{C}$, s is treated as augmented context with an empty united context (e.g. s is the same as s()).

Definition 6. (G-world-trees): *Let* $M = (W, \pi, \mathcal{R})$ *be an interpretation structure defined upon a set* Φ *of propositions and a set* $\mathcal{C}$ *of context names. Let* d *be a context in* $\mathcal{C}$ *with accessibility relation* R_d *in* $\mathcal{R}$. *Let* w *and* w_0 *be worlds in* W. *Let* G_n *be the united context* (s1,...,sn) *and let* G_{n+1} *be the united context* (s1,...,sn+1), *for* s1, ..., sn+1 *in* $\mathcal{C}$. *The set of G-world-trees associated with a united context is defined inductively as follows:*

$$\begin{aligned} \mathit{trees}(w, G_{n+1}) := \{(w, S) | \exists S_n, t : (w, S_n) \in \mathit{trees}(w, G_n) \wedge \\ t \in \mathit{trees}(w, s_{n+1}) \wedge S = S_n \cup \{t\}\} \end{aligned}$$

A tuple (w, S) *is a* G-world-tree *of a world* w *and a united context* G_{n+1} *if there exists a set* S_n *such that* (w, S_n) *is a G-world-tree of the world* w *and the united context* G_n, *there exists a G-world-tree* t *of the world* w *and the context* sn+1, *and* $S = S_n \cup \{t\}$.

The G world-tree of a world w *and an empty united context* () *is* $(w, \{\})$.

The set of trees associated with an augmented context is defined inductively as:

$$\mathit{trees}(w_0, \mathrm{d}G) := \{(w, S) | w \in R_d(w_0) \wedge (w, S) \in \mathit{trees}(w, G)\}$$

A tuple (w, S) *is a* G-world-tree *of a world* w_0 *and an augmented context* dG *if* $w \in R_d(w_0)$ *and* (w, S) *is a G-world-tree of the world* w *and the united context* G.

Using our example program "d[s1[sailing] s2[boats]]" and the interpretation structure obtained in Section 3, we obtain the following G-world trees:

$$\begin{aligned}
\mathit{trees}(w_{s1}, ()) &= \{(w_{s1}, \{\})\} \\
\mathit{trees}(w_d, s1) &= \{(w, S) | w \in R_{s1}(w_d) \wedge (w, S) \in \mathit{trees}(w, ())\} \\
&= \{(w_{s1}, \{\})\} \\
\mathit{trees}(w_d, (s1)) &= \{(w_d, S) | \exists S_n, t : (w_d, S_n) \in \mathit{trees}(w_d, ()) \wedge \\
&\quad t \in \mathit{trees}(w_d, s1) \wedge S = S_n \cup \{t\}\} \\
&= \{(w_d, \{(w_{s1}, \{\})\})\} \\
\mathit{trees}(w_d, (s1, s2)) &= \{(w_d, S) | \exists S_n, t : (w_d, S_n) \in \mathit{trees}(w_d, (s1)) \wedge \\
&\quad t \in \mathit{trees}(w_d, s2) \wedge S = S_n \cup \{t\}\} \\
&= \{(w_d, (w_{s1}, \{\}), (w_{s2}, \{\}))\} \\
\mathit{trees}(w_{this}, d(s1, s2)) &= \{(w, S) | w \in R_d(w_{this}) \wedge (w, S) \in \mathit{trees}(w, (s1, s2))\} \\
&= \{(w_d, (w_{s1}, \{\}), (w_{s2}, \{\}))\}
\end{aligned}$$

The G-world tree of an augmented context d(s1,...,sn) formalises the accessibility of possible worlds associated with the subcontexts si from the possible world associated with the context d, and for each si, the accessibility of the possible worlds associated with the subcontexts of the si from the possible world associated with the context si, etc. This reflects the logical structure among d, its subcontexts si, the subcontexts of si, etc. We call $\mathcal{T}$ the set of G-world trees associated with an interpretation structure M.

The next step is to define truth value assignment functions with respect to G-world trees to capture the logical structure among the contexts. First we define the truth value value assignment function, referred to as *united truth value assignment function* associated with a united context to model the united knowledge of the united context. Here, the definition of the four truth values as sets (see Definition 1) is exploited to arrive at the *united truth value*.

Definition 7. (United truth value assignment function π_U): *Let $M = (W, \pi, \mathcal{R})$ be an interpretation structure defined upon a set Φ of propositions and a set $\mathcal{C}$ of context names. Let $\mathcal{T}$ the set of G-world trees associated with M. Let (w, S) be a G-world-tree in $\mathcal{T}$ of world $w \in W$ and a united context G from $\mathcal{C}$. The united truth value function $\pi_U : \mathcal{T} \to (\Phi \to \{\text{true}, \text{false}, \text{inconsistent}, \text{unknown}\})$ of a G-world-tree (w, S) is defined as the union of the truth value functions $\pi(w')$ where the worlds w' are the roots of the subtrees $(w', \{\})$ in the set S. Formally, for all proposition φ in Φ:*

$$\pi_U((w, S))(\varphi) := \bigcup_{(w', \{\}) \in S} \pi(w')(\varphi)$$

For an empty set S, the united truth value $\pi_U((w, \{\}))(\varphi) := \text{unknown}$.

Consider the united truth value of the united context (s1,s2). Given the accessibility relations $R_{s1} = \{(w_d, w_{s1})\}$ and $R_{s2} = \{(w_d, w_{s2})\}$, we obtain for instance[3]:

$$\begin{aligned}\pi_U((w_d, \{(w_{s1}, \{\}), (w_{s2}, \{\})\}))(\mathit{sailing}) &= \pi(w_{s1})(\mathit{sailing}) \cup \pi(w_{s2})(\mathit{sailing}) \\ &= \mathit{true} \cup \mathit{unknown} = \mathit{true}\end{aligned}$$

This means that with respect to the logical structure between s1 and s2, the united truth value of sailing in w_d is *true*.

We define now the truth value assignment function, referred to as *augmented truth value assignment function*, associated with an augmented context to model its augmented knowledge. Analogously to united truth value, the function is defined upon G-world-trees thus capturing the logical structure between the contexts, and the definition of the four truth values as sets is again exploited to arrive at the *augmented truth value*.

Definition 8. (Augmented truth value assignment function π_A): *Let $M = (W, \pi, \mathcal{R})$ be an interpretation structure defined upon a set Φ of propositions and a set $\mathcal{C}$ of context names. Let $\mathcal{T}$ the set of G-world trees associated with M. Let d be a context in $\mathcal{C}$ and G a united context in $\mathcal{C}$. Let (w, S) be a G-world-tree in $\mathcal{T}$ of world w_0 in W and the augmented context $\mathrm{d}G$. The augmented truth value function $\pi_A : \mathcal{T} \to (\Phi \to \{\text{true}, \text{false}, \text{inconsistent}, \text{unknown}\})$ of a G-world-tree (w, S) is defined as the union of the truth value function $\pi(w)$ of world w and the united truth value function $\pi_U((w, S))$ of the G-world-tree (w, S). Formally, for all φ in Φ:*

$$\pi_A((w, S))(\varphi) := \pi(w)(\varphi) \cup \pi_U((w, S))(\varphi)$$

Consider the G-world-tree $S = (w_d, \{ (w_{s1}, \{\}), (w_{s2}, \{\})\})$ for the world w_{this} and the augmented context d1(s1,s2). We obtain for instance

$$\begin{aligned}\pi_A((w_d, S))(\mathit{sailing}) &= \pi(w_d)(\mathit{sailing}) \cup \pi_U((w_d, S))(\mathit{sailing}) \\ &= \pi(w_d)(\mathit{sailing}) \cup \pi(w_{s1})(\mathit{sailing}) \cup \pi(w_{s2})(\mathit{sailing}) \\ &= \mathit{unknown} \cup \mathit{true} \cup \mathit{unknown} = \mathit{true}\end{aligned}$$

World w_d is associated with the context d. For the basic context d, sailing is *unknown* in w_d, whereas, for the augmented context d(s1,s2), sailing is *true* in w_d, thus modelling augmented knowledge.

[3] We recall that $\mathit{true} = \{t\}$ and $\mathit{unknown} = \{\}$ so $\mathit{true} \cup \mathit{unknown} = \{t\} \cup \{\} = \{t\} = \mathit{true}$.

5 Equivalence of the two approaches

Sections 3 and 4 present two approaches that formalise the knowledge augmentation process. The two approaches are based on different formalisations of united knowledge (Definitions 4 and 7) and augmented knowledge (Definitions 5 and 8). In the present section, we formally show that the two approaches are equivalent, that is, they lead to the same knowledge augmentation process. First, we show that the definitions of united truth value assignment function (Definition 7) and united knowledge modal operator (Definition 4) lead to the same truth value assignment to a proposition in a united context. For space constraint, we only provide the proof for the truth value *true*.

Theorem 1. (United knowledge): *.Let $M = (W, \pi, \mathcal{R})$ be an interpretation structure defined upon a set Φ of propositions and a set $\mathcal{C}$ of context names. Let $\mathcal{T}$ the set of G-world trees associated with M. In a world $w \in W$, a proposition $\varphi \in \Phi$ is true in a united context G iff for each tree (w, S) of world w and united context G, the united truth value equals* true. *Formally* $\forall$ s *in* $G : R_s(w) \neq \{\}$*:*

$$(M, w) \models U_G[\varphi] \iff \forall (w, S) \in \mathit{trees}(w, G) : \pi_U((w, S))(\varphi) = \text{true}$$

We prove the theorem by contradiction and induction over united contexts.

Proof: Let G be (s1), i.e. a united context with one context. Assume that there exists a tree (w, S) for the world w and the united context (s1) such that the united truth value $\pi_U((w, S))(\varphi) \neq$ *true*. According to Definition 7 (united truth value), the union $\bigcup_{(w', \{\}) \in S} \pi(w')(\varphi) \neq$ *true*. According to Definition 6 (G-world-tree), for a united context with one context, we obtain:

$$\begin{aligned} \mathit{trees}(w, (\text{s1})) &= \{(w, S) | \exists t : t \in \mathit{trees}(w, \text{s1}) \wedge S = \{t\}\} \\ \mathit{trees}(w, \text{s1}) &= \{(w', \{\}) | w' \in R_{s1}(w)\} \end{aligned}$$

The set S contains one tree $(w', \{\})$, $w' \in R_{s1}(w)$. Since the union $\bigcup_{(w', \{\}) \in S} \pi(w')(\varphi) \neq$ *true*, there exists a world w' such that $w' \in R_{s_1}(w)$ and the truth value $\pi(w')(\varphi) \neq$ *true*. According to Definition 2 (interpretation function), s1$[\varphi]$ is not true in world w. According to Definition 4 (united knowledge), $U_{(\text{s1})}[\varphi]$ is not true in world w.

We prove now the reverse. Assume that $U_{(\text{s1})}[\varphi]$ is not true in world w. According to definition 4 (united knowledge), s1$[\varphi]$ is not true in world w. According to Definition 2 (interpretation function), there exists a world w' such that $w' \in R_{s_1}(w)$ and the truth value $\pi(w')(\varphi) \neq$ *true*. According to Definition 6 (G-world-tree), for a united context with one context, we obtain:

$$\begin{aligned} \mathit{trees}(w, (\text{s1})) &= \{(w, S) | \exists t : t \in \mathit{trees}(w, \text{s1}) \wedge S = \{t\}\} \\ \mathit{trees}(w, \text{s1}) &= \{(w', \{\}) | w' \in R_{s1}(w)\} \end{aligned}$$

The set S contains one tree $(w', \{\})$, $w' \in R_{s1}(w)$. Since there exists a world w' such that $w' \in R_{s_1}(w)$ and the truth value $\pi(w')(\varphi) \neq$ *true*, the union $\bigcup_{(w', \{\}) \in S} \pi(w')(\varphi) \neq$ *true*. According to Definition 7 (united truth value), there exists a tree (w, S) for the world w and the united context (s1) such that the united truth value $\pi_U((w, S))(\varphi) \neq$ *true*.

Assume that the theorem holds for a united context G_n with n contexts.

$$(M, w) \models U_{G_n}[\varphi] \iff \forall (w, S) \in \mathit{trees}(w, G_n) : \pi_U((w, S))(\varphi) = \mathit{true}$$

Adding one context sn+1 leads to the united context G_{n+1}, and we prove:

$$(M,w) \models U_{G_{n+1}}[\varphi] \iff \forall (w,S) \in \mathit{trees}(w, G_{n+1}) : \pi_U((w,S))(\varphi) = \mathit{true}$$

Assume that there exists a tree (w,S) for the world w and the united context G_{n+1} such that the united truth value $\pi_U((w,S))(\varphi) \neq \mathit{true}$. From Definition 6 (G-world-tree), we obtain the G-world-trees of the united context G_n and context sn+1.

$$(w,S) \in \mathit{trees}(w, G_{n+1}) \iff \\ (w,S_n) \in \mathit{trees}(w, G_n) \wedge t \in \mathit{trees}(w, \mathrm{sn{+}1}) \wedge S = S_n \cup \{t\}$$

According to Definition 7, $(\pi_U((w,S_n)) \cup \pi_U((w,\{t\})))(\varphi) \neq \mathit{true}$ since the united truth value $\pi_U((w,S))(\varphi) \neq \mathit{true}$. From Definition 4 (united knowledge), we obtain that for $U_{G_n}[\varphi]$ being true, φ is neither true nor unknown in (sn+1), and for φ being unknown in G_n, φ is not true in (sn+1). Thus, $U_{G_{n+1}}[\varphi]$ is not true in world w.

Assume that $U_{G_{n+1}}[\varphi]$ is not true in world w. From Definition 4 (united knowledge), we obtain that for $U_{G_n}[\varphi]$ being true, φ is neither true nor unknown in (sn+1), and for φ being unknown in G_n, φ is not true in (sn+1). From Definition 6 (G-world-tree), we obtain the G-world-trees of the united context G_n and context sn+1.

$$(w,S) \in \mathit{trees}(w, G_{n+1}) \iff \\ (w,S_n) \in \mathit{trees}(w, G_n) \wedge t \in \mathit{trees}(w, \mathrm{sn{+}1}) \wedge S = S_n \cup \{t\}$$

According to Definition 7, the united truth value $\pi_U((w,S))(\varphi) \neq \mathit{true}$, since $(\pi_U((w,S_n)) \cup \pi_U((w,\{t\})))(\varphi) \neq \mathit{true}$. Thus, there exists a tree (w,S) for the world w and the united context G_{n+1} such that the united truth value $\pi_U((w,S))(\varphi) \neq \mathit{true}$. □□□

We have shown the equivalence between the united knowledge modal operator (Definition 4) and the united truth value assignment function (Definition 7). We turn now to augmented knowledge, where we show that the definitions of the augmented truth value assignment (Definition 8) and the augmented knowledge modal operator (Definition 5) lead to the same truth value of a proposition in an augmented context. As for united knowledge, we prove the equivalence for the truth value *true* only.

Theorem 2. (Augmented knowledge): *Let $M = (W, \pi, \mathcal{R})$ be an interpretation structure defined upon a set Φ of propositions and a set $\mathcal{C}$ of context names. Let $\mathcal{T}$ the set of G-world trees associated with M. In a world w_0 in W, a proposition φ in Φ is true in an augmented context* dG *iff for each tree (w,S) of world w_0 and an augmented context* dG*, the augmented truth value equals* true*.*

$$(M, w_0) \models \mathrm{d}G[\varphi] \iff \forall (w,S) \in \mathit{trees}(w_0, \mathrm{d}G) : \pi_A((w,S))(\varphi) = \mathrm{true}$$

We prove the theorem by contradiction for an augmented context dG.

Proof: Assume that there exists a tree $(w,S) \in \mathit{trees}(w_0, \mathrm{d}G)$ for the world w_0 and the augmented context dG such $\pi_A((w,S))(\varphi) \neq \mathit{true}$. According to Definition 8 (augmented truth value), $(\pi(w) \cup \pi_U((w,S)))(\varphi) \neq \mathit{true}$. It follows that if $\pi(w)(\varphi) = \mathit{true}$, then $\pi_U((w,S)))(\varphi)$ is neither *true* nor *unknown*, and if $\pi(w)(\varphi) = \mathit{unknown}$, then $\pi_U((w,S))(\varphi) \neq \mathit{true}$. The same applies analogously for switched roles of $\pi(w)$ and $\pi_U((w,S))$. It follows that if φ is true in d for world w_0 ($(M,w_0) \models \mathrm{d}[1\ \varphi]$), then there exists $w \in R_d(w_0)$ such that φ is neither *true* nor *unknown* in the united context G for world w, and if φ is *unknown* in d for world w_0 ($(M,w_0) \models \mathrm{d}[0/0/0/1\varphi]$), then there exists $w \in R_d(w_0)$ such that $U_G[\varphi]$ is not true in world w.

The same applies analogously for switched roles of $\mathrm{d}[\varphi]$ and $U_G[\varphi]$. Thus, $\mathrm{d}G[\varphi]$ is not true in world w_0.

Assume that $\mathrm{d}G[\varphi]$ is not true in world w_0. According to Definition 5 (augmented knowledge), if φ is true in d for world w_0 ($(M, w_0) \models \mathrm{d}[1\ \varphi]$), then there exists $w \in R_d(w_0)$ such that φ is neither *true* nor *unknown* in the united context G for world w, and if φ is *unknown* in d for world w_0 ($(M, w_0) \models \mathrm{d}[0/0/0/1\varphi]$), then there exists $w \in R_d(w_0)$ such that $U_G[\varphi]$ is not true in world w. The same applies analogously for switched roles of $\mathrm{d}[\varphi]$ and $U_G[\varphi]$. It follows that if for all $w \in R_d(w_0)$, $\pi(w)(\varphi) = \mathit{true}$, then there exists a tree (w, S) for the united context G such that $w \in R_d(w_0)$ and $\pi_U((w, S))(\varphi)$ is neither *true* nor *unknown*, and if for all $w \in R_d(w_0)$, $\pi(w)(\varphi) = \mathit{unknown}$, then there exists a tree (w, S) for the united context G such that $\pi_U((w, S))(\varphi) \neq \mathit{true}$. The same applies analogously for switched roles of $\pi(w)$ and $\pi_U((w, S))$. It follows that there exists a world $w \in R_d(w_0)$ such that $(\pi(w) \cup \pi_U((w, S)))(\varphi) \neq \mathit{true}$. Thus, there exists a tree $(w, S) \in \mathit{trees}(w_0, \mathrm{d}G)$ for the world w_0 and the augmented context $\mathrm{d}G$ such $\pi_A((w, S))(\varphi) \neq \mathit{true}$. □□□

The section shows, via (a sample of) its proofs, that our first formalism (Section 3) is compatible with the well-known framework defined in [9] for modelling the combination of knowledge (Section 4). Further studies are however necessary to investigate expressiveness and complexity issues when probabilities are added to the representation to capture the uncertainty inherent in describing content and its effect in retrieval.

6 Conclusion and future work

This paper describes a formal model for representing the content and logical structure of structured documents for the purpose of their retrieval. To obtain a representation that allows for the retrieval of objects of varying granularity, the content of an object is viewed as the knowledge contained in that object, and the structure among objects is captured by a process of knowledge augmentation, which leads to a representation of the object that is based on its own content and that of its structurally related objects.

The model is based on a Kripke structure in which the structure of a document is reflected. The objects composing the document are viewed as contexts (agents) that possess knowledge about their own content and their structure. This knowledge is augmented with that of its structurally related objects. The model is based on the definitions of knowledge modal operators, augmented knowledge operators and augmented truth value assignment functions, formalised upon an interpretation structure (possible worlds, truth value assignments and accessibility relations). The model uses four truth values to capture incompleteness and inconsistency. The knowledge augmentation process is formalised in two equivalent ways: (1) our first approach uses modal operators to formalise the content knowledge of an object augmented with that of its structurally related objects (2) our second approach, based on the framework defined in [9] to model the combination of knowledge, formalises the logical structure of an object through G-world-trees upon which augmented truth values are defined.

The formalism described in this paper is the basis of the development of a model for structured document retrievals that allows for the retrieval of objects of varying granularity. Future work will present the representation of the uncertainty inherent to the information retrieval process [2, 11], which will be used to estimate the degree to which an object is relevant to a query.

References

1. Abiteboul, S., Cluet, S., Christophides, V., Milo, T., Moerkotte, G., and Simeon, J. Querying documents in object databases. *International Journal on Digital Libraries 1* (1997), 1–9.
2. Baeza-Yates, R., and Ribeiro-Neto, B. *Modern Information Retrieval*. Addison Wesley, 1999.
3. Baumgarten, C. A probabilistic model for distributed information retrieval. In *Proceedings of ACM-SIGIR Conference on Research and Development in Information Retrieval* (Philadelphia, USA, 1997), pp. 258–266.
4. Belnap, N. A useful four-valued logic. In *Modern Uses of Multiple-valued Logic*, J. Dunn and G. Epstein, Eds. Reidel, Dordrecht, 1977.
5. Callan, J. Passage-level evidence in document retrieval. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland, 1994), pp. 302–310.
6. Chellas, B. *Modal Logic*. Cambridge University Press, 1980.
7. Chiaramella, Y. Browsing and querying: two complementary approaches for multimedia information retrieval. In *Proceedings Hypermedia - Information Retrieval - Multimedia* (Dortmund, Germany, 1997).
8. Chiaramella, Y., Mulhem, P., and Fourel, F. A model for multimedia information retrieval. Tech. Rep. Fermi ESPRIT BRA 8134, University of Glasgow, 1996.
9. Falgin, R., Harpen, J., Moses, J., and Vardi, M. *Reasoning about Knowledge*. MIT Press, Cambridge, Massachusetts, 1995.
10. Frisse, M. Searching for information in a hypertext medical handbook. *Communications of the ACM 31*, 7 (1988), 880–886.
11. Fuhr, N. Probabilistic models in information retrieval. *The Computer Journal, Special Issue 35*, 3 (1992), 243–255.
12. Fuhr, N., Goevert, N., and Roelleke, T. Dolores: A system for logic-based retrieval of multimedia objects. In *Proceedings of ACM-SIGIR Conference on Research and Development in Information Retrieval* (1999).
13. Lalmas, M., and Moutogianni, E. A Dempster-Shafer indexing for the focussed retrieval of hierarchically structured documents: Implememtation and experiments on a web museum collection. In *RIAO* (2000).
14. Lalmas, M., Roelleke, T., Turra, F., and Fuhr, N. Concepts for a graphical user interface for hypermedia retrieval. In *Flexible Question-Answering Systems (FQAS 2000)* (2000).
15. Myaeng, S., Jang, D. H., Kim, M. S., and Zhoo, Z. C. A flexible model for retrieval of SGML documents. In *Proceedings of ACM-SIGIR Conference on Research and Development in Information Retrieval* (Melbourne, Australia, 1998), pp. 138–145.
16. Roelleke, T. *POOL: Probabilistic Object-Oriented Logical Representation and Retrieval of Complex Objects - A Model for Hypermedia Retrieva*. PhD thesis, University of Dortmund, Germany, 1999.
17. Roelleke, T., and Fuhr, N. Retrieval of complex objects using a four-valued logic. In *Proceedings of ACM-SIGIR Conference on Research and Development in Information Retrieval* (Zurich, Switzerland, 1996), pp. 206–214.
18. Wilkinson, R. Effective retrieval of structured documents. In *Proceedings of ACM-SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland, 1994), pp. 311–317.