Modeling Contextualized Knowledge

Martin Homola[1,2], Luciano Serafini[1], and Andrei Tamilin[1]

[1] FBK-IRST, Via Sommarive 18, 38123 Trento, Italy
[2] Comenius University, FMFI, 84248 Bratislava, Slovakia
`homola|serafini|tamilin@fbk.eu`

Abstract. Most of the knowledge available in the Semantic Web is context dependent. Examples of contextual information that is associated with knowledge are time, topic, provenance, reliability, etc. Recently, several paradigms, tools and languages have been proposed with the aim of adding context awareness into the Semantic Web. That is, enabling representation and reasoning not only with the knowledge alone, but also with the associated contextual information. Examples include RDF quadruples, named graphs, annotated RDF, and contextualized knowledge repositories. These new paradigms introduce a new dimension into knowledge engineering: in addition to individuals, concepts, properties and their relations, we also need to define the set of contexts, and we need to "split" the knowledge between these contexts. In this paper, we propose a modeling exercise with one of the tools, for which we choose the contextualized knowledge repository. The example is complex enough to highlight many issues connected with contextualized knowledge representation, and it could possibly become the first benchmark for contextual knowledge representation tools.

1 Introduction and Motivation

The Wikipedia Infobox of the term "Italy" states that Italian President is "Giorgio Napolitano"; clearly, this fact holds only during the current legislature. Similarly, from Freebase one can see that Homer Simpson is a Nuclear Safety Inspector and that John McCarthy is a professor of Stanford University, these two facts hold under different circumstances, the former holds in the Simpsons fictional world while the latter holds in the current real world. Searching for Diego Maradona in the SIGMA[3] semantic search engine, one obtains that he is an attacking midfielder, and he coaches Argentina national team. These two facts cannot hold at the same time.

These are just simple examples showing that most of the knowledge retrievable from the Semantic Web is context dependent. Nevertheless, the information about the context is usually not specified in Semantic Web resources, and when it is so, e.g., by adding attributes like `rdfs:comment`, `owl:AnnotationProperty`, etc., this information is completely ignored in the reasoning process. The importance of contextualized knowledge has been widely recognized and this has

---

[3] `http://sig.ma`

motivated proposals for extending Semantic Web languages with the possibility of qualifying knowledge w.r.t. some specific contextual dimension. For instance [1, 2] focus on the representation of knowledge provenance; [3] allows the representation of propositional attitudes, [4] covers the dynamic aspect of knowledge and knowledge about events. In the last years there have been also proposals for a general context representation framework for the Semantic Web [5–7].

Recently we have proposed a new architecture which accommodates contextual knowledge within the current state-of-the-art semantic web technology [8, 9]. The architecture, called *Contextualized Knowledge Repository* (CKR), has been also implemented on top of one of the state-of-the-art RDF triple stores. Main features of CKR are as follows:

1. knowledge is contextualized relying on the well studied theories of context [10–15] and this contextualization is implemented inside the current Semantic Web languages without any semantic extension. This is an advantage since we want to rely on the plethora of existing Semantic Web tools;
2. a context is treated as a theory—set of sentences in a logical language, closed under logical consequence—and it is associated with a "point" in a dimensional space. Contexts are also first class objects, and the logic provides terms to denote them;
3. knowledge propagates across contexts according to schematic patterns. This is done through so called *qualified concepts and roles* which constitute a semantic bridge between different contexts. Lifting axioms are thus hidden from the user and they work automatically. This transfers part of the complexity of the modeling task from the user to the system.

In this paper, we show a practical modeling scenario on which the CKR architecture is applied. We demonstrate the modeling capabilities and main advantages of CKR. The proposed modeling scenario is from the FIFA World Cup domain which has been chosen for its inherent contextual nature and it is complex enough to highlight many issues connected with contextualized knowledge representation. We believe that this scenario can be possibly remodeled also in any other contextual Semantic Web framework thus constituting a modeling benchmark in this area.

2 The Modeling Domain

As a case study for contextualized knowledge representation we propose the domain of football and in particular we focus on the FIFA World Cup 2010. The reasons for choosing this domain are the following:

1. it is a structured domain, and most of the information available, such as teams, matches, scores, etc., can be easily represented with the standard Semantic Web languages as RDF/OWL.
2. large part of the knowledge about this event is context dependent. For instance, the players of a team can be different for each match. Two teams

can play one against the other in two different matches obtaining different scores, a player can have a different role in different matches, etc.

3. the domain presents a high level of interconnection between knowledge contained in different contexts. For instance the players of a team are always the same along the entire competition, players do not change their shirt number, etc.

Already from the web site of the FIFA World Cup 2010[4], one can recognize a certain level of contextual representation of the information, and how contexts are used for a better presentation of the information. For instance, pages are organized by team, by player, by single match. The information contained in each single page focuses on a given topic, and it is supposed to hold within the contextual boundaries of the page. The contextual structure is mainly driven by the sake of effective organization of the web site. We will analyze the page content in detail, starting from more specific pages, proceeding to the more general, finally reaching the home page.

Among the most specific pages, there are the pages associated to each single match. They contain information which holds only in that particular match. For instance the lineup formation, the yellow/red cards, the substitutions, the goals, etc. For instance the fact that "Gilardino was replaced by Di Natale" is contained in the page for Match 11 of Group F, and it is supposed to hold in this match (and not in other matches). In RDF, this information is represented by a triple of the form "⟨Alberto_Gilardino⟩.⟨is_substituted_by⟩.⟨Antonio_di_Natale⟩". In this case, the triple is valid only in this particular context (i.e., that of the Match 11). However, not all of the information contained in a page is context dependent.

For instance, the pages of matches contain more general information about the teams, the players, etc. which also holds in a broader context. The page related to the match also lists general information about players such as their height, the club where they currently play, etc. This information is "imported" and it is apparently listed here because it is considered to be relevant for the context. It is important to notice that, only a part of all information that is possibly available is "imported" in this context. For instance, the information about the club is relative only on the current year, and not all the clubs a player has ever played for.

There are other pages associated with different phases of the competition. For instance, the page entitled "group stage"[5] summarizes the information of the initial phase of the competition, such as the composition of groups, and the matches taken within each group. Clearly, this information forms a broader context that encompasses all of the matches of the group stage. Some of the data listed cannot be associated with any of the particular matches, such as for instance the composition and the final ranking of each group.

The broadest context is the one associated with the whole competition. It contains information about the teams, the players, the stadiums, the schedule,

---

[4] `www.fifa.com`

[5] `www.fifa.com/worldcup/matches/groupstage.html`

etc. Such a context is not the broadest one can imagine, indeed one could consider all the different editions of the FIFA world cup, and the knowledge about all the football leagues in each country. As a consequence the context "FIFA world cup 2010" should be inserted in a larger contextual structure that involves all the other mentioned contexts.

3 Contextualized Knowledge Repository

According to the *context as a box* metaphor [16] a context can be viewed as a box. *Inside the box* is a set of logical statements representing the information associated with this particular context. This is the contextual data. The boundaries of the box are then determined by a set of dimensional values. This information is the contextual meta data, often simply marked as the *outside of the box*. For instance, a match between Italy and Paraguay during the FIFA World Cup can be represented by the following context:

$$\mathcal{C} = \overset{\mathsf{time}(\mathcal{C}, \mathsf{2010\text{-}06\text{-}14}),\ \mathsf{location}(\mathcal{C}, \mathsf{World}),\ \mathsf{topic}(\mathcal{C}, \mathsf{FIFA\_WC\_Match\_11})}{\boxed{\begin{array}{l} \mathsf{TeamA}(\mathsf{Team\_Italy}) \\ \mathsf{TeamB}(\mathsf{Team\_Paraguay}) \\ \mathsf{Referee}(\mathsf{Benito\_Archundia}) \\ \ldots \end{array}}}$$

Although contexts are possibly defined on top of any logical language, we focus on RDF and OWL and hence we will build contexts on top of description logics (DL) [17]. This is due to the fact that OWL itself is built on top of the DL $\mathcal{SROIQ}$ [18]. Furthermore we will assume that all the symbols that appear either in the contexts or in dimensional information come from some shared vocabulary Σ.

Definition 1 (Context). *A context is a triple* $\langle \mathcal{C}, \dim(\mathcal{C}), \mathrm{K}(\mathcal{C}) \rangle$ *such that:*

1. $\dim(\mathcal{C})$ *is a set of assertions of the form* $A(\mathcal{C}, v)$ *where* A *is called dimensional attribute and* v *dimensional value;*
2. $\mathrm{K}(\mathcal{C})$ *is a DL knowledge base in* $\mathcal{SROIQ}$ *or some of its sublanguages.*

The set of dimensional values $\dim(\mathcal{C})$ of a given context $\mathcal{C}$ determines its contextual boundaries. Later on, a formal model of a dimensional space will be introduced and we will learn that these values indicate a location of a context in this space. Such a space is composed of several dimensions which provide possible values for all dimensional attributes of a given context. For instance the location dimension provides all possible locations with which contexts can be associated. Similarly for dimensions such as time, topic and possibly others.

This determination may be precise, when all these values are constants. This is the case of the context above, in which location is set to World, time is set to 2010-06-14, and topic is set to FIFA_WC_Match_11. Such contexts are called primitive contexts. They contain information which is tightly bound with a particular set of dimensional values.

Besides for primitive contexts we will also make use of context classes. These are more generic contexts that allow us to specify some information which is valid for multiple sets of dimensional values. This is done by using a concept in place of one or more of the dimensional values. The semantics will then take care that this information is associated with all primitive contexts whose dimensional vectors "instantiate" the dimensional vector of the class context. Let us take for example a generic context class representing a football match:

$$\Gamma = \overset{\mathsf{time}(\Gamma,\top),\ \mathsf{location}(\Gamma,\top),\ \mathsf{topic}(\Gamma,\mathsf{Football\_Match})}{\boxed{\begin{array}{l} \mathsf{TeamA} \sqsubseteq \mathsf{Football\_Team} \\ \mathsf{TeamB} \sqsubseteq \mathsf{Football\_Team} \\ \mathsf{TeamA} \sqsubseteq \neg\mathsf{TeamB} \\ \ldots \end{array}}}$$

Provided that the value FIFA_WC_Match_11 is an instance of the concept Football_Match, the semantics will associate this information with the context representing match 11 listed above, and similarly also with all other contexts representing particular FIFA matches that instantiate the context class.

In addition to context classes CKR offers another more selective option to propagate knowledge across contexts. With qualified concepts and roles, one can selectively query knowledge recorded in some other context. Syntactically this is done by adding the dimensional vector of the queried context into the subscript. For instance, if in the context of FIFA_Match we wish to import information from the context FIFA_WC representing the FIFA World Cup, we may write an axiom of the form:

$$\begin{array}{l} \exists \mathsf{has\_Nationality}_{\mathsf{FIFA\_WC}}{}^{-}.(\mathsf{Referee} \sqcup \mathsf{Assistant\_Referee}) \ \sqcap \\ \qquad\qquad \exists \mathsf{is\_National\_Team\_Of}_{\mathsf{FIFA\_WC}}{}^{-}.(\mathsf{TeamA} \sqcup \mathsf{TeamB}) \sqsubseteq \bot \end{array}$$

By this axiom it is required in the context of a match that referees and assistant referees must not share nationality with any of the teams in the match. Since the information about the nationalities is part of the context FIFA_WC we have used qualified role names like $\mathsf{has\_Nationality}_{\mathsf{FIFA\_WC}}$ in order to access this information. The semantics will take care of that the interpretation of $\mathsf{has\_Nationality}_{\mathsf{FIFA\_WC}}$ is the same as the interpretation of has_Nationality in the context FIFA_WC.

This kind of knowledge transcendence is enabled by the fact that the topics of FIFA World Cup and the one of FIFA match are related (i.e., the latter is a subtopic of the former). Therefore the semantics of qualified concepts and roles is closely related to the hierarchy of contexts, which in turn reflects the hierarchy of dimensions. Therefore a CKR consists of a collection of contexts $\mathfrak{C}$ and of meta knowledge about the dimensions $\mathfrak{D}$, as we formally define below.

Definition 2 (Contextualized Knowledge Repository). *A* contextualized knowledge repository (CKR) *is a pair* $\mathfrak{K} = \langle \mathfrak{D}, \mathfrak{C} \rangle$ *where* $\mathfrak{C}$ *is a set of contexts, and* $\mathfrak{D}$ *is a DL knowledge base such that:*

1. $\mathfrak{D}$ *contains* n *distinct roles* $\mathbf{A} = \{A_1, \ldots, A_n\}$ *called dimensions (or dimensional attributes);*

2. *for every dimension $A \in \mathbf{A}$, $\mathfrak{D}$ contains a finite set D_A called the dimension values of A such that each $v \in D_A$ is either a constant symbol or a concept in $\mathfrak{D}$;*
3. *for every dimension $A \in \mathbf{A}$, $\mathfrak{D}$ contains a role covers_A whose domain and range are the constants of D_A;*
4. *the transitive closure of the relation $\{\langle d, d' \rangle \mid \mathfrak{D} \models \mathsf{covers}(d, d')\}$, denoted by $\prec_A$, is a partial order on D_A.*

Due to the hierarchy of dimensions, the organization of contexts in each CKR is hierarchical as well. Given a CKR $\mathfrak{K}$ and two contexts $\mathcal{C}_1$ and $\mathcal{C}_2$ with $\dim(\mathcal{C}_1) = \boldsymbol{d}$ and $\dim(\mathcal{C}_2) = \boldsymbol{e}$. We say that $\mathcal{C}_1$ is covered by $\mathcal{C}_2$ if $\boldsymbol{d}_A \prec_A \boldsymbol{e}_A$ for each $A \in \mathbf{A}$. This is denoted by $\mathcal{C}_1 \prec \mathcal{C}_2$ (also $\boldsymbol{d} \prec \boldsymbol{e}$).

Context classes do not directly participate in the context hierarchy, as some of their dimensional values are not constants. This corresponds with our intuition that a context class is a special collection of knowledge that belongs into multiple contexts. These contexts are called instances of a given context class. A context $\mathcal{C}$ is an instance of a context class Γ, if for each $A \in \mathbf{A}$ either $\dim(\mathcal{C})_A = \dim(\Gamma)_A$ or $\mathfrak{D} \models \dim(\Gamma)_A(\dim(\mathcal{C})_A)$[6]. This is denoted by $\Gamma(\mathcal{C})$ (also $\Gamma(\dim(\mathcal{C}))$).

The semantics of a CKR is an extension of the model-theoretic semantics of DL [17]. A model of a CKR $\mathfrak{K} = \langle \mathfrak{D}, \mathfrak{C} \rangle$ is a class of interpretations $\mathfrak{I} = \{\mathcal{I}_{\boldsymbol{d}}\}_{\boldsymbol{d} \in \mathfrak{D}}$, $\mathcal{I}_{\boldsymbol{d}} = \left\langle \Delta_{\boldsymbol{d}}, \cdot^{\mathcal{I}_{\boldsymbol{d}}} \right\rangle$, that satisfies certain additional constraints. The most important of these constraints are:

1. $\mathcal{I}_{\boldsymbol{d}}$ is a model of the information associated with the dimensional vector $\boldsymbol{d}$ (i.e., $\mathcal{I}_{\boldsymbol{d}} \models \mathrm{K}(\mathcal{C})$ for every $\boldsymbol{d} \in \mathfrak{D}$, and $\mathcal{I}_{\boldsymbol{d}} \models \mathrm{K}(\Gamma)$ for every context class Γ such that $\Gamma(\boldsymbol{d})$);
2. the hierarchy of contexts is reflected by the hierarchy of interpretation domains (i.e., $\Delta_{\boldsymbol{d}} \subseteq \Delta_{\boldsymbol{e}}$ if $\boldsymbol{d} \prec \boldsymbol{e}$);
3. interpretation of constants is shared by all contexts (i.e., $a^{\mathcal{I}_{\boldsymbol{d}}} = a^{\mathcal{I}_{\boldsymbol{e}}}$ for every constant a);
4. interpretation of qualified symbols is based on their home context (i.e., $(C_{\boldsymbol{d}})^{\mathcal{I}_{\boldsymbol{e}}} = C^{\mathcal{I}_{\boldsymbol{d}}} \cap \Delta_{\boldsymbol{e}}$ and $(R_{\boldsymbol{d}})^{\mathcal{I}_{\boldsymbol{e}}} = R^{\mathcal{I}_{\boldsymbol{d}}} \cap \Delta_{\boldsymbol{e}} \times \Delta_{\boldsymbol{e}}$ for any concept C and any role R if $\boldsymbol{d} \prec \boldsymbol{e}$ or $\boldsymbol{e} \prec \boldsymbol{d}$).

Entailment in a CKR is defined with respect to a particular context.

Definition 3. *A formula ϕ (i.e., a subsumption $X \sqsubseteq Y$ or an assertion $A(a)$) is entailed by the CKR $\mathfrak{K}$ in $\boldsymbol{d}$, denoted by $\mathfrak{K} \models \boldsymbol{d} : \phi$, if for each model $\mathfrak{I}$ of $\mathfrak{K}$, $\mathcal{I}_{\boldsymbol{d}} \models \phi$[7].*

All basic decision problems such as concept satisfialbility checking and entailment with respect to a CKR knowledge base are decidable, and the complexity of reasoning is not increased when compared to classical DL [9].

---

[6] The somewhat clumsy notation $\mathfrak{D} \models \dim(\Gamma)_A(\dim(\mathcal{C})_A)$ refers to the standard DL-related reasoning task of instance checking, i.e., it is to be read: the knowledge base $\mathfrak{D}$ implies that the constant $\dim(\mathcal{C})_A$ is an instance of the concept $\dim(\Gamma)_A$ [17].

[7] The definition of $\mathcal{I} \models X \sqsubseteq Y$ and $\mathcal{I} \models A(x)$ is the standard one for description logics [17].

The advantages of explicit tracking of knowledge provenance by attaching contextual meta information are apparent. In addition, our system provides means for efficient manipulation of generic information that is valid in multiple contexts using the construct of context class. Also, knowledge propagation in the system is encoded in the semantics of qualified concepts and roles and the complexity is hidden from the user which leads to practical and efficient modeling. We will demonstrate this in the next section.

4 Modeling with CKR

In this section we describe how we model the domain of football, in particular the domain of FIFA Would Cup 2010, within a contextualized knowledge repository. The modeling is composed of three basic components: The first component is the dimension hierarchies and the meta knowledge $\mathfrak{D}_{\mathsf{Football}}$. In this knowledge base we represent the structure of the contexts in which we organize all the information about football. The structure of $\mathfrak{D}_{\mathsf{Football}}$ will be inspired by the structure of the FIFA World Cup 2010 web site. The second component of the modeling consists of the context classes which describe types of contexts that we will deal with, such as for instance matches, team, groups, etc. Each context class contains all the axioms that should hold in all of its instance-contexts. Finally, the third component of the model are the contexts, with all the specific knowledge.

4.1 Context dimensions

The knowledge base $\mathfrak{D}_{\mathsf{Football}}$ contains the meta knowledge and formalizes the structure of the contexts in the repository in a logical form. In the proposed implementation we consider the following three dimensions:

time: values are time intervals of the form $\langle \mathsf{start\text{-}time}, \mathsf{end\text{-}time} \rangle$.[8] Coverage relation is the standard containment relation between intervals. It does not require explicit representation because the containment relation between two intervals can be evaluated on the fly;

location: values determine the geographical region in which the set of statements in a context is true. This structure is represented by geographic ontology encoded in OWL and constructed from Geonames[9], the resource of geographical places. The ontology defines generic concepts, such as for instance Geographic_Area or Country as well as specific individuals for geographical places, such as World, Italy, or Florence. The concepts are then used as dimension values of context classes of the repository, while the individuals are used as dimension values of particular contexts. The hierarchical

---

[8] For the sake of notation simplicity, to refer to the common temporal intervals, e.g., the whole year or day, we will just use the year or date instead of start-end tuple.

[9] www.geonames.org

```
sport
  Football
    Football_domestic_competitions
    Football_international_competitions
      Continental_League
      FIFA_WC
        FIFA_WC_qualification
        FIFA_WC_final_tournament
          FIFA_WC_group_stage
              FIFA_WC_group_A
                 FIFA_WC_Match_1
              ...
              FIFA_WC_group_H
                 FIFA_WC_Match_48
          FIFA_WC_knockout
            FIFA_WC_round_of_16
              FIFA_WC_Match_49
              ...
            FIFA_WC_semi-final
              FIFA_WC_Match_61
              FIFA_WC_Match_62
            FIFA_WC_third_place_match
            FIFA_WC_final_match
```

Fig. 1. Excerpt of a topic dimension hierarchy

coverage relation $\mathsf{covers}_{\mathsf{location}}$ corresponds to actual geopolitical relations between the regions as defined by Geonames (i.e., the division of the world into continents, the continents into countries, and countries into administrative regions, etc.);

topic: values of this dimension determine the subject topic that a context pertains to. Similarly to locations, the structure of the topic dimension is encoded by an OWL ontology defining generic concepts and specific individuals, such as the concept Football_Match and the individual FIFA_WC_final_match. The hierarchical coverage relation between individuals is asserted by evaluating topic generalization/specification. Figure 1 shows an excerpt of topic hierarchy, manually created following the organizational structure of the FIFA World Cup event.

Using these three dimensions, it is possible to specify detailed contexts, some examples are listed below:

Context description	time	location	topic
European Champions League 2010	2010	Europe	Continental_league
FIFA World Cup 2010	2007–2010	World	FIFA_WC
The final tournaments of the FIFA World Cup of the last 20 years	1990–2010	World	FIFA_WC_final_tournament
The group A of FIFA World Cup 2006	2006	World	FIFA_WC_group_A
The final match of the latest FIFA World Cup	2010	World	FIFA_WC_final_match

4.2 Context classes

In the CKR architecture, context classes can be viewed as a tool which makes the representation of knowledge more effective. Axioms that are valid in many contexts are asserted in a context class and they are automatically imported by the semantics into all contexts that instantiate the class. For example, in every context that describes a football match, certain general axioms should hold: e.g., there are two teams, one playing against the other, the number of players in the match is eleven or less then per team, there is a goalkeeper for each team, etc. Instead of explicitly including all these axioms in each single context associated with a football match, one can create the context class Match containing the axioms and impose that all context that describe a particular match are instances of this context class.

Example 1. As an explanatory example we show the definition of the context class Match associated to every match

$$\mathsf{Match} = \overset{\begin{array}{l}\mathsf{time}(\mathsf{Match}, \mathsf{Time\_Interval} \sqcap \exists \mathsf{start}.(\geq 1900)),\\ \mathsf{location}(\mathsf{Match}, \mathsf{World}),\\ \mathsf{topic}(\mathsf{Match}, \mathsf{Football\_Match})\end{array}}{\boxed{\begin{array}{l}\mathsf{Team} \equiv \mathsf{TeamA} \sqcup \mathsf{TeamB}\\ \mathsf{TeamA} \sqsubseteq \neg \mathsf{TeamB}\\ \mathsf{Team} \sqsubseteq {=}11.\mathsf{in\_Lineup}^{-}.\top\\ \mathsf{Team} \sqsubseteq \exists \mathsf{has\_Captain}.\top\\ \mathsf{has\_captain} \sqsubseteq \mathsf{in\_Lineup}\\ \ldots\end{array}}}$$

The context class described above matches with all the contexts which describe any football match which is taken in any part of the world after 1900. When a context is declared to be an instance of this class it inherits the basic structure of a football match and as well as all the constraints defined in the context class in terms of axioms.

Example 2. Another example is a context class that specifies the composition of a group during the FIFA World Cup and basic relations between the teams the group, for instance, there are exactly four teams in the group, and at the end there is one winner and one runner-up winner who are among these four teams:

$$\mathsf{FIFA\_Group} = \overset{\mathsf{topic}(\mathsf{FIFA\_Group}, \mathsf{FIFA\_Group})}{\boxed{\begin{array}{l}\mathsf{Team} \equiv \mathsf{Team1} \sqcup \mathsf{Team2} \sqcup \mathsf{Team3} \sqcup \mathsf{Team4}\\ \mathsf{Winner} \sqsubseteq \mathsf{Team}\\ \mathsf{Runner\text{-}up} \sqsubseteq \mathsf{Team}\\ \ldots\end{array}}}$$

The context class FIFA_Group will match any of the eight contexts representing the actual groups of the FIFA World Cup, that is, groups A, B, C, D, E, F, G, H, and as a result it will define the notions of the winner and the runner-up in each of the groups.

4.3 Contexts

The declaration of the actual contexts in the repository concludes the modeling process. It is important to stress that when a context is loaded into the CKR, first all matching context classes are identified and then axioms contained in them are copied into the context. Let us now see some examples.

Example 3. To represent the FIFA World Cup group A 2010, one gives the following definitions:

time(GA_2010, 2010), location(GA_2010, World), topic(GA_2010, FIFA_WC_group_A)
GA_2010 =
Team1(Team_Uruguay)
Team2(Team_Mexico)
. . .
Winner(Team_Uruguay)
Runner-up(Team_Mexico)

Note that the axioms about the teams and the winners are imported due to the fact that the dimension value FIFA_WC_group_A is an instance of the concept FIFA_Group and hence the whole context is an instance of the eponymous context class.

Another powerful modeling mechanism proposed in the CKR is the notion of qualified concepts and roles. This mechanism allows to refer from one context to concepts and roles of another context. For example, according to the format of the FIFA World Cup, after the group stage matches only the winner and the runner-up team from each group pass into the next tournament stage, called "round of sixteen", here, group A is one context and a match in the round of sixteen is another. With help of qualified of qualified concepts and roles we are able to express complex axioms in the latter context reusing knowledge of the former.

Example 4. To represent the constraint that during the match 46 in the round of sixteen the teams Uruguay and Republic Korea are actually the winner and the runner-up of the groups A and B accordingly, we insert the following axioms in the context of the match:

time(Match_49, 2010), location(Match_49, World), topic(Match_49, FIFA_WC_Match_49)
Match_49 =
TeamA(Team_Uruguay)
$\mathsf{TeamA} \sqsubseteq \mathsf{Winner}_{\mathsf{2010,World,FIFA\_WC\_group\_A}}$
. . .

The semantics of qualified symbols will take care that the qualified concept $\mathsf{Winner}_{\mathsf{2010,World,FIFA\_WC\_group\_A}}$ actually refers to Winner in the context with the dimensions $\langle \mathsf{2010, World, FIFA\_WC\_group\_A} \rangle$, that is, the one that we above named GA_2010.

Example 5. To represent the context of the final match of the FIFA World Cup 2010, the following declarations can be used:

time(Final_2010, 2010), location(Final_2010, World), topic(Final_2010, FIFA_WC_final_match)
Final_2010 =
TeamA(Team_Spain)
TeamB(Team_Netherlands)
in_Lineup(Team_Netherlands, Maarten_Stekelenburg)
in_Lineup(Team_Netherlands, Giovanni_van_Bronckhorst)
has_Captain(Team_Netherlands, Giovanni_van_Bronckhorst)
. . .

We also assert in the meta-knowledge that FIFA_WC_final_match is an instance of the concept Football_Match and hence this context will inherit all the axiom from the context class Match. In addition, please note the role of each player and the shirt number of the player do not change between the matches in one tournament, thus this information is not specified here but it can be imported from the more general context of the FIFA World Cup 2010 whenever need. This approach is also reflected at the actual web site of the FIFA tournament.

5 Conjunctive Query Answering over CKR

In this section we will take a look on conjunctive query answering in a CKR. We show how the notion of conjunctive query has to be generalized in order to be useful in a situation involving multiple contexts and we will explain how answers for the conjunctive queries are defined. We will then show several examples of conjunctive queries and the answers building on our FIFA World Cup scenario introduced above.

In the classical setting a conjunctive query (CQ) is an expression of the form $Q(\boldsymbol{x}) \leftarrow \exists \boldsymbol{y} \bigwedge_{i=1}^{n} \phi_i(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are tuples of variables, and each $\phi_i(\boldsymbol{x}, \boldsymbol{y})$ is either unary or binary predicate taking variables from $\boldsymbol{x}$ and $\boldsymbol{y}$ [19].

In the multi-contextual setting, conjunctive queries need to be parametrized also with respect to a context.

Definition 4 (Contextual Conjunctive Query). *A* contextual conjunctive query (CCQ) *over a CKR is an expression of the form* $Q(\boldsymbol{x}) \leftarrow \exists \boldsymbol{y} \bigwedge_{i=1}^{n} \boldsymbol{d}_i : \phi_i(\boldsymbol{x}, \boldsymbol{y})$ *where for each* i, $\boldsymbol{d}_i$ *is a possibly partial dimensional vector, and* $\phi_i(\boldsymbol{x}, \boldsymbol{y})$ *is a conjunction of unary and binary predicates taking variables from* $\boldsymbol{x}$ *and* $\boldsymbol{y}$*. If* $\boldsymbol{x}$ *is empty then the query is called* boolean*. If for each* i, $\boldsymbol{d}_i$ *is a total dimensional vector then the query is said to be* fully contextualized.

In CKR, queries can span over multiple contexts with different conceptualizations and hence the result can be seen as a mash-up of knowledge from different contexts. We will make use of substitution and completion. By $\phi[\boldsymbol{x}/\boldsymbol{a}]$ we will understand the expression derived from ϕ in which each element of $\boldsymbol{x}$ is replaced by the respective element of $\boldsymbol{a}$. Given two dimensional vectors $\boldsymbol{d}$ and $\boldsymbol{e}$, by $\boldsymbol{d} + \boldsymbol{e}$ we will understand a completed vector which contains all elements of $\boldsymbol{d}$ plus the elements of $\boldsymbol{e}$ for those dimensions which $\boldsymbol{d}$ is missing. The semantics of CCQ is defined as follows.

Definition 5. *A fully contextualized boolean CCQ* $Q() \leftarrow \exists \boldsymbol{y} \bigwedge_{i=1}^{n} \boldsymbol{d}_i : \phi_i(\boldsymbol{y})$ *is said to be entailed by a CKR* $\mathfrak{K}$*, if for each model* $\mathfrak{I}$ *of* $\mathfrak{K}$ *there is a substitution* $\boldsymbol{u}$ *such that* $\mathcal{I}_{\boldsymbol{d}_i} \models \phi(\boldsymbol{y})[\boldsymbol{y}/\boldsymbol{u}]$*. This is denoted by* $\mathfrak{K} \models Q()$*.*

The expression $\boldsymbol{e} : \boldsymbol{a}$ *formed by a dimensional tuple* $\boldsymbol{e}$ *and a tuple of constants* $\boldsymbol{a}$ *is an answer for a CCQ* $Q(\boldsymbol{x}) \leftarrow \exists \boldsymbol{y} \bigwedge_{i=1}^{k} \boldsymbol{d}_i : \phi_i(\boldsymbol{x}, \boldsymbol{y})$ *with respect to a CKR* $\mathfrak{K}$*, if* $\mathfrak{K} \models \bigwedge_{i=1}^{k} \boldsymbol{d}_i + \boldsymbol{e} : \phi_i(\boldsymbol{x}, \boldsymbol{y})[\boldsymbol{x}/\boldsymbol{a}]$*.*

Let us now see some practical examples of CCQ.

Example 6. As a simple example, consider the following query which retrieves the list of all goalkeepers and the teams they play for in the current World Cup:

$$Q(x, y) \leftarrow \langle \mathsf{2010}, \mathsf{World}, \mathsf{FIFA\_WC} \rangle : \mathsf{Goalkeeper}(x) \wedge \mathsf{in\_Squad}(x, y)$$

The answer set will look like:

x	y
Gianluigi_Buffon	Team_Italy
Federico_Marchetti	Team_Italy
Tim_Howard	Team_USA
. . .	

Example 7. The previous query is concerned with a sole context of the 2010 World Cup, which is fully specified in the query. We may of course query over multiple contexts. The next query retrieves the list of goalkeepers who played for the team Italy in any FIFA World Cup:

$$Q(x) \leftarrow \langle \mathsf{World}, \mathsf{FIFA\_WC} \rangle : \mathsf{Goalkeeper}(x) \wedge \mathsf{in\_Squad}(x, \mathsf{Team\_Italy})$$

As we can see, by omitting one of the dimensions in the dimensional vector within the query, this query is evaluated by substituting all possible dimensional values of for this dimension and querying in every resulting context. This yields the answer set:

Context			
time	location	topic	x
2010	World	FIFA_WC	Gianluigi_Buffon
2010	World	FIFA_WC	Federico_Marchetti
2006	World	FIFA_WC	Gianluigi_Buffon
2006	World	FIFA_WC	Angelo_Peruzzi
. . .	World	FIFA_WC	. . .

Note that the answer set is contextualized, that is, for each answer we get also the context in which it answers the query. We can also see that some individuals are listed more than once, each time in a different context.

Example 8. Of course, we have the possibility to explicitly address more than one context. In the following query, we ask about the list of players who did in fact play in some match. For this we have to consider the context FIFA_WC, which contains the list of players and then the contexts of particular matches.

$$Q(x, y) \leftarrow \exists z \langle \mathsf{2010}, \mathsf{World}, \mathsf{FIFA\_WC} \rangle : \mathsf{in\_Squad}(x, y) \bigwedge \langle \mathsf{2010}, \mathsf{World} \rangle : \mathsf{in\_Lineup}(z, x)$$

The first context we have specified fully and for the second one we again use the same feature as above: we omit the value for the topic dimension, thus all possible values are substituted here. Due to the fact that the predicate in_Lineup is inherent to the contexts that represent matches, the query is evaluated once per each match. The answer set will look as follows:

Context				
time	location	topic	x	y
2010	World	FIFA_WC_Match_11	Gianluigi_Buffon	Team_Italy
2010	World	FIFA_WC_Match_11	Federico_Marchetti	Team_Italy
2010	World	FIFA_WC_Match_28	Federico_Marchetti	Team_Italy
2010	World	FIFA_WC_Match_41	Federico_Marchetti	Team_Italy
2010	World	. . .	. . .	. . .

We can see that same goalkeepers played in several matches, but in addition there are also matches in which more than one goalkeeper played.

From the examples, we conclude that CCQ are a natural and particularly versatile extension of CQ that provides a flexible mechanism for data retrieval over a CKR that allows us to retrieve and combine data from multiple contexts and in addition it allows predicating over contextual meta data in order to refine the query.

6 Related Work

In this section we outline some notable approaches and systems for modeling contextualized knowledge with semantic web technologies, particularly focusing on the football domain selected for the use case in the present work. Though the football domain is characterized by well structuredness of the available information (e.g., tournaments, teams, players, matches, etc.), which simplifies its representation with RDF/OWL semantic web standards, the challenging problem is to reflect and consistently represent context dependency of the most of knowledge with RDF/OWL (e.g., line ups within a certain match, authors of goal shots, etc.).

The 2006 FIFA World Cup was one of the main application scenarios investigated within the SmartWeb[10] research project. The corresponding knowledge on the world cup has been encoded in the SWIntO Sport Event RDFS ontology [20]. The modeling choice pursued to represent context-dependency of knowledge was through a specialization of concepts and properties. For example, depending on the context, the notion of team can refer to the football club, complete squad composition in a certain tournament or the actual lineup in a certain match. To distinguish between these different granularities the concept `Team` is specialized into `FootballClubTeam`, `Squad` or `FootballMatchTeam` respectively, and instantiated accordingly into different individuals `germany`, `germany_fifa_2006` and `germany_14_june_2006`, which afterwards are linked together using specific relation `personatedBy`. The similar modeling approach is used to represent a notion of player, having different roles in a club, in a team or a specific match; different officials in different matches, etc. In our approach there is no need to proliferate creation of specialized concepts and individuals for distinguishing contextual qualifications, because contexts allow us to treat differences of concepts in different contexts and consistently use the same name for individuals through out different contexts.

Another notable example extensively modeling football domain is Freebase. Freebase[11] is a massive collaboratively-edited RDF-exportable knowledge base of facts about people, organizations, events, etc. The knowledge base is organized into domains (e.g., sport disciplines, politics, etc.), grouping relevant types (e.g., sport championships, clubs and players, politicians and parties, etc.).

---

[10] smartweb.dfki.de

[11] freebase.com

Types have properties (e.g., `Date_of_birth` for type `Person`), and can be organized in inheritance hierarchies (e.g., `Football_Player` type extends generic type `Person`) allowing for property inheritance. For example, for representation of facts about 2010 FIFA World Cup Freebase contains a dedicated type `FIFA_World_Cup_2010`[12]. One of the distinguishing characteristics of Freebase is extensive use of reification in order to support compound multi-dimensional properties, allowing to assign contextually bounded values. An example of such a compound property is `Football_Player_Match_Participation` allowing to assert for a given match a player and a team he plays.

7 Conclusion

Contextualized Knowledge Repository (CKR) constitutes a novel architecture for the Semantic Web that has been lately proposed and implemented [8, 9]. It is completely embedded within the current Semantic Web standards represented by RDF and OWL. It builds on top of these standard formalisms and enhances them in the following aspects: (1) knowledge is organized in contexts which are hierarchically sorted according to the coverage relation defined with respect to the contextual metadata; (2) the coverage relation is itself formalized in an RDF/OWL ontology, which introduces flexibility on the structure of contexts and it allows to reason, not only inside the contexts but also on the contextual organization; (3) with context classes generic knowledge that is valid in multiple contexts can be asserted effectively and with minimal redundancy; (4) so called qualified concepts and roles allow for fully automated knowledge "importing" between the contexts that relies on their hierarchical structure, it is intuitive to use, and whose technicalities are hidden from the user; (5) contextual conjunctive queries provide a flexible data retrieval mechanism in which also contextual metadata are returned with the answers and furthermore the querying can be also refined by such metadata.

In this paper, we describe a modeling scenario from the domain of football, by which we demonstrate the features of the CKR architecture, we show how to model with it in practise, and highlight its particular advantages. The choice of this particular domain is due to its inherent contextual nature and sufficient complexity. Although equivalent modeling can surely be done in any ontology language such as OWL that does not provide any context aware features, it is apparent from our demonstration that with CKR an increased efficiency of the representation and more intuitive modeling are achieved. We believe that the scenario can be remodeled also in any other contextualized Semantic Web framework and thus it may serve in future as a modeling benchmark in this area.

References

1. Ding, L., Finin, T., Peng, Y., da Silva, P.P., McGuinness, D.L.: Tracking rdf graph provenance using rdf molecules. In: Proceedings of the Fourth International

---

[12] www.freebase.com/view/en/2010_fifa_world_cup

Semantic Web Conference. (2005)

2. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named graphs, provenance and trust. In: WWW '05: Proceedings of the 14th international conference on World Wide Web, New York, NY, USA, ACM (2005) 613–622
3. Nickles, M.: Social acquisition of ontologies from communication processes. Appl. Ontol. **2**(3-4) (2007) 373–397
4. Liao, H.C., Tu, C.C.: A rdf and owl-based temporal context reasoning model for smart home. Information Technology Journal **6** (2007) 1130–1138
5. Guha, R., Mccool, R., Fikes, R.: Contexts for the semantic web. In: ISWC, volume 3298 of Lecture Notes in Computer Science, Springer (2004) 32–46
6. H.Stoermer: Introducing Context into Semantic Web Knowledge Bases. In: Proceedings of the CAISE*06 Doctoral Consortium. (June 2006)
7. O.Khriyenko, V.Terziyan: A framework for context sensitive metadata description. Int. J. Metadata Semant. Ontologies **1**(2) (2006) 154–164
8. M.Joseph, L.Serafini, A.Tamilin: Context shifting for effective search over large knowledge bases. In: Workshop on Context, Information And Ontologies (CIAO-2009), collocated with the 6th European Semantic Web Conference (ESWC-2009). (2009)
9. Serafini, L., Homola, M., Joseph, M., Tamilin, A.: Contextualized knowledge repositories for the semantic web. Technical report, Fondazione Bruno Kessler – IRST, Trento, Italy (2010) To appear.
10. Buvac, S., Mason, I.A.: Propositional logic of context. In: In Proceedings of the Eleventh National Conference on Artificial Intelligence. (1993) 412–419
11. Buvac, S.: Quantificational logic of context. In: In Proceedings of the Eleventh National Conference on Artificial Intelligence. (1996) 412–419
12. Giunchiglia, F., Serafini, L.: Multilanguage hierarchical logics, or: how we can do without modal logics. Artif. Intell. **65**(1) (1994) 29–70
13. Ghidini, C., Giunchiglia, F.: Local models semantics, or contextual reasoning = locality + compatibility. Artificial Intelligence **127** (2001)
14. Ghidini, C., Serafini, L.: Distributed First Order Logics. In Gabbay, D., de Rijke, M., eds.: Frontiers Of Combining Systems 2. Studies in Logic and Computation, Research Studies Press (1998) 121–140
15. M.Benerecetti, P.Bouquet, C.Ghidini: Contextual Reasoning Distilled. Experimental and Theoretical AI **12**(3) (2000) 279–305
16. M.Benerecetti, Bouquet, P., C.Ghidini: On the Dimensions of Context Dependence. In P.Bouquet, L.Serafini, R.H.Thomason, eds.: Perspectives on Contexts. CSLI Lecture Notes. Center for the Study of Language and Information/SRI (2007) 1–18
17. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., eds.: The Description Logic Handbook. Cambridge University Press (2003)
18. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible $\mathcal{SROIQ}$. In: Procs. of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR 2006), AAAI Press (2006) 57–67
19. Horrocks, I., Tessaris, S.: A conjunctive query language for description logic aboxes. In: AAAI/IAAI. (2000) 399–404
20. Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., Mougouie, B., Vembu, S., Baumann, S., Romanelli, M., Buitelaar, P., Engel, R., Sonntag, D., Reithinger, N., Loos, B., Porzel, R., Zorn, H., Micelli, V., Schmidt, C., Weiten, M., Burkhardt, F., Zhou, J.: DOLCE ergo SUMO: On foundational and domain models in the SmartWeb Integrated Ontology (SWIntO). Web Semantics **5**(3) (September 2007) 156–174