National Technical University of Ukraine
«Kyiv Polytechnic Institute»
High Performance Computing Center (hpcc.kpi.ua)

Pospishnyi Oleksandr
pospishniy@kpi.in.ua

Extending Datatype Support for Tractable Reasoning with OWL 2 EL Ontologies

ORE2013
OWL Reasoner Evaluation Workshop

Ulm 2013

Introduction and Motivation

In Description Logics, *datatypes* (also known as concrete domains) can be used to define new concepts by referring to particular values, such as integers or strings.

$$\mathbf{DualCoreCPU} \equiv \mathbf{CPU} \sqcap \exists \mathbf{hasCores}.(=, \mathbf{2})$$
$$\mathbf{QuadCoreCPU} \equiv \mathbf{CPU} \sqcap \exists \mathbf{hasCores}.(=, \mathbf{4})$$
$$\mathbf{MultiCoreCPU} \equiv \mathbf{CPU} \sqcap \exists \mathbf{hasCores}.(>, \mathbf{1})$$

...

$$\mathbf{Xeon} \sqsubseteq \mathbf{CPU} \sqcap \exists \mathbf{hasInstructionSet}.\mathbf{x86\_64} \sqcap \exists \mathbf{manufacturedBy}.\mathbf{Intel}$$
$$\mathbf{E30\text{–}1280} \sqsubseteq \mathbf{Xeon} \sqcap \exists \mathbf{hasCores}.(=, \mathbf{4})$$

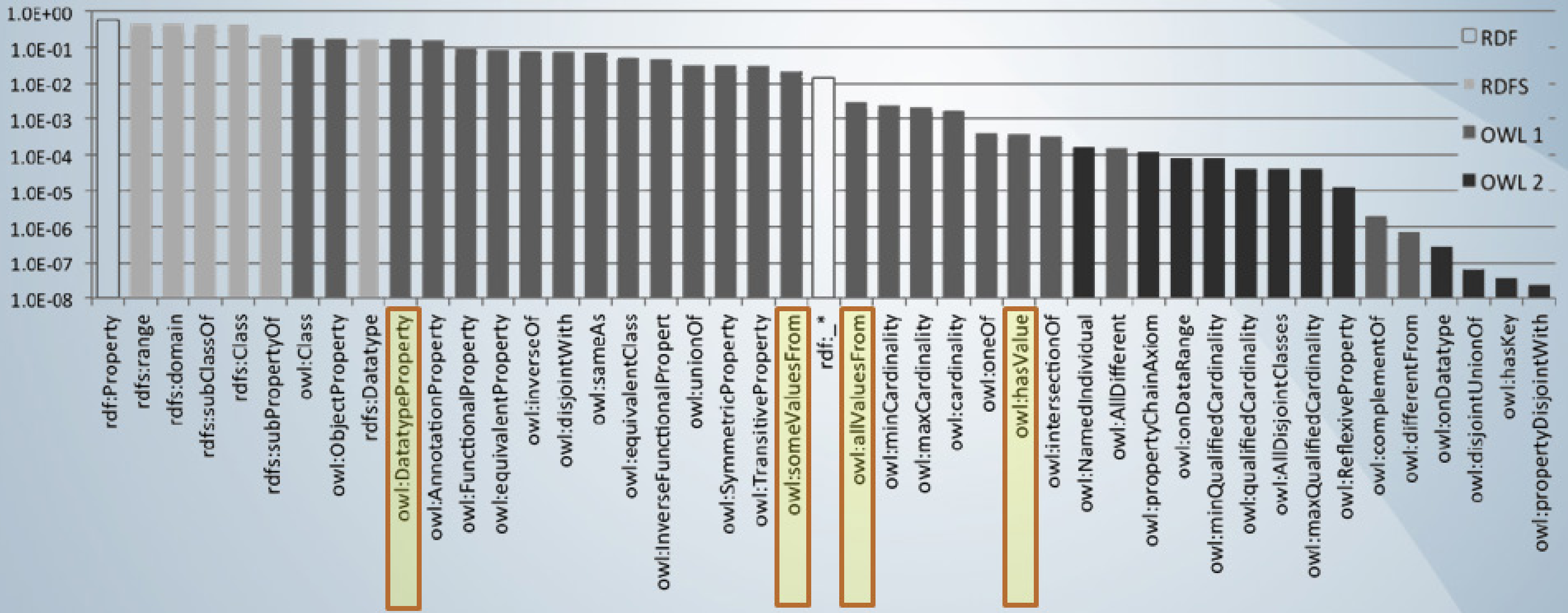
Any ontology reasoner, with a support of datatype expressions should be able to derive that:

$$\mathbf{DualCoreCPU} \sqsubseteq \mathbf{MultiCoreCPU}$$
$$\mathbf{QuadCoreCPU} \sqsubseteq \mathbf{MultiCoreCPU}$$

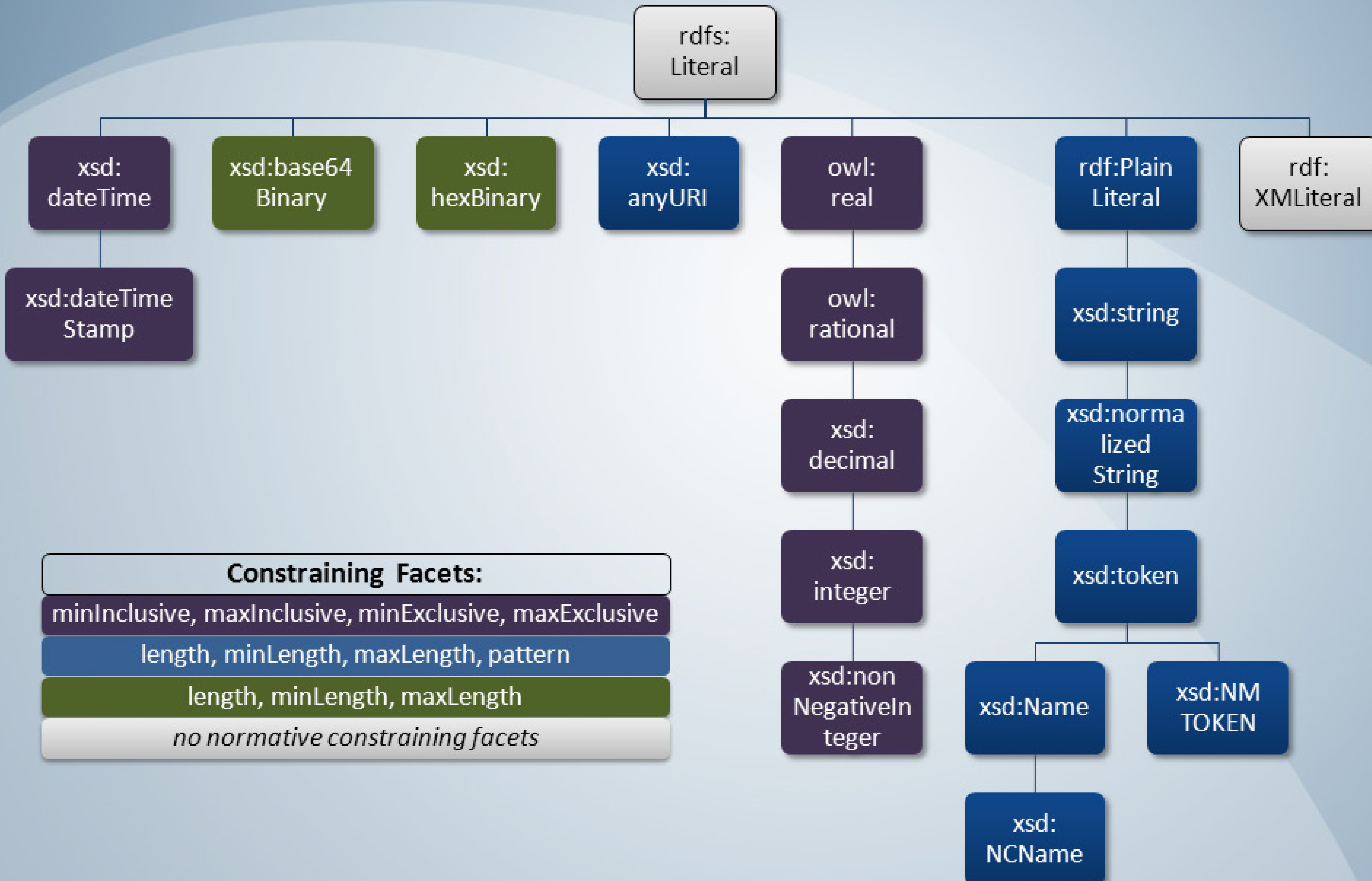
$$\mathbf{E30\text{–}1280} \sqsubseteq \mathbf{QuadCoreCPU}$$
$$\mathbf{E30\text{–}1280} \sqsubseteq \mathbf{MultiCoreCPU}$$

Datatype Expressions in the wild

Birte Glimm, Aidan Hogan, Markus Krötzsch, Axel Polleres. "OWL: Yet to arrive on the Web of Data?." arXiv preprint arXiv:1202.0984 (2012).



Allowed Datatypes in OWL 2 EL



Tractable Datatype Reasoning using ELK

In order to support reasoning with abovementioned allowed datatypes, a set of new inference rules was implemented for the ELK reasoner:

$$\mathbf{R}^{\mathcal{D}}_{\sqsubseteq}\ \frac{C \sqsubseteq \exists R.\, r_{+}}{C \sqsubseteq D} \quad \exists R.\, r_{-} \sqsubseteq D \in \mathcal{O}, \quad r_{+} \subseteq r_{-}$$

$$\mathbf{R}^{\mathcal{D}}_{\bot}\ \frac{}{C \sqsubseteq \bot} \quad C \sqsubseteq \exists R.\, r_{+} \in \mathcal{O}, \quad r_{+} = \emptyset$$

$$\mathbf{R}_{\sqsubseteq}\ \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\mathbf{R}^{+}_{\sqcap}\ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} : D_1 \sqcap D_2 \in \mathcal{O}$$

$$\mathbf{R}^{-}_{\sqcap}\ \frac{C \sqsubseteq D_1 \sqcap D_2}{\begin{matrix} C \sqsubseteq D_1 \\ C \sqsubseteq D_2 \end{matrix}}$$

$$\mathbf{R}^{+}_{\exists}\ \frac{C \sqsubseteq D}{\exists S.\, C \rightarrow \exists S.\, D} : \exists S.\, D \in \mathcal{O}$$

$$\mathbf{R}^{-}_{\exists}\ \frac{C \sqsubseteq \exists R.\, D}{D \sqsubseteq D}$$

$$\mathbf{R}_{\mathcal{H}}\ \frac{D \sqsubseteq \exists R.\, C \quad \exists S.\, C \rightarrow E}{D \sqsubseteq E} : R \sqsubseteq^{*}_{\mathcal{O}} S$$

$$\mathbf{R}^{+}_{\top}\ \frac{C \sqsubseteq C}{C \sqsubseteq \top}$$

$$\mathbf{R}_{\mathcal{T}}\ \frac{D \sqsubseteq \exists R.\, C \quad \exists S.\, C \rightarrow E}{\exists T.\, D \rightarrow E} : \begin{matrix} R \sqsubseteq^{*}_{\mathcal{O}} T \sqsubseteq^{*}_{\mathcal{O}} S \\ \mathrm{Trans}(T) \in \mathcal{O} \end{matrix}$$

Rule $\mathbf{R}_{\perp}^{\mathcal{D}}$

$$\mathbf{R}_{\perp}^{\mathcal{D}} \frac{}{\mathrm{C} \sqsubseteq \perp} \qquad \mathrm{C} \sqsubseteq \exists \mathrm{R}.\mathrm{r}_{+} \in \mathcal{O}, \qquad \mathrm{r}_{+} = \emptyset$$

where $\mathcal{O}$ is an ontology, C and D are the concepts, R is a datatype property and $\exists \mathrm{R}.\mathrm{r}_{+}$ denotes an existential datatype expression created with constraining facets.

With $\mathrm{r}_{+} = \emptyset$ we represent an empty value space produced by the datatype restriction r_{+}.

Example:

$$\mathbf{A} \equiv \exists \mathrm{R}.\boldsymbol{integer}(> \mathbf{10}, < \mathbf{5})$$
$$\mathbf{B} \sqsubseteq \exists \boldsymbol{R}.\boldsymbol{integer}(> \mathbf{4}, < \mathbf{5})$$
$$\mathbf{C} \sqsubseteq \exists \boldsymbol{R}.\boldsymbol{nonNegativeInteger}(\leq \mathbf{0})$$

Rule $\mathbf{R}^{\mathcal{D}}_{\sqsubseteq}$

$$\mathbf{R}^{\mathcal{D}}_{\sqsubseteq}\ \frac{\mathrm{C} \sqsubseteq \exists \mathrm{R}.\mathrm{r}_{+}}{\mathrm{C} \sqsubseteq \mathrm{D}} \quad \exists \mathrm{R}.\mathrm{r}_{-} \sqsubseteq \mathrm{D} \in \mathcal{O}, \qquad \mathrm{r}_{+} \subseteq \mathrm{r}_{-}$$

where $\mathcal{O}$ is an ontology, C and D are the concepts, R is a datatype property and $\exists \mathrm{R}.\mathrm{r}_{\pm}$ denotes an existential datatype expression created with constraining facets. Symbols $_{+}$ and $_{-}$ indicate that an expression is occurring right or left of $\sqsubseteq$ respectively.

With $\mathrm{r}_{+} \subseteq \mathrm{r}_{-}$ we represent a fact that a value space constrained by the datatype restriction r_{+} is a subset of a value space constrained by the r_{-} datatype restriction.

During the computation of all conclusions under the inference rules, for each encountered $\exists \mathrm{R}.\mathrm{r}_{+}$ expression a search is conducted for all $\exists \mathrm{R}.\mathrm{r}_{-}$ expressions in the ontology where $\mathrm{r}_{+} \subseteq \mathrm{r}_{-}$.

Example:

$$\begin{array}{l} \mathbf{A} \sqsubseteq \exists \mathbf{R}.(> \mathbf{5}) \\ \exists \mathbf{R}.(\geq \mathbf{3}) \sqsubseteq \mathbf{B} \end{array} \quad \Rightarrow \quad (> \mathbf{5}) \subseteq (\geq \mathbf{3}) \quad \Rightarrow \quad \mathbf{A} \sqsubseteq \boldsymbol{B}$$

Rule $\mathbf{R}^{\mathcal{D}}_{\sqsubseteq}$: Value spaces

All possible value spaces, created by the datatype restrictions, could be conventionally divided into 3 categories:

Values	Binary value
	Date-time value
	Literal value
	Numerical value
Intervals	Numeric interval
	Date-time interval
	Length restriction
Other	Empty value space
	Entire value space
	Patter restriction

Subsumption matrix $B \subseteq A$ $(r_+ \subseteq r_-)$

A \ B	EMPTY VALUESP.	ENTIRE VALUESP.	BINARY VALUE	DATETIME VALUE	LITERAL VALUE	NUMERIC VALUE	DATETIME INTERVAL	LENGTH RESTRICTION	NUMERIC INTERVAL	PATTERN
EMPTY VALUESP	-	-	-	-	-	-	-	-	-	-
ENTIRE VALUESP	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$	$B_D \subseteq A_D$
BINARY VALUE	-	-	$B_D = A_D$ $A = B$	-	-	-	-	-	-	-
DATETIME VALUE	-	-	-	$A =^1 B$	-	-	-	-	-	-
LITERAL VALUE	-	-	-	-	$A =^2 B$	-	-	-	-	-
NUMERIC VALUE	-	-	-	-	-	$A =^3 B$	-	-	-	-
DATETIME INTERVAL	-	-	-	$B_D \subseteq A_D$ $A_{low} \leq B$ $A_{high} \geq B$	-	-	$B_D \subseteq A_D$ $A_{low} \leq B_{low}$ $A_{high} \geq B_{high}$	-	-	-
LENGTH RESTRICT.	-	-	$B_D \subseteq A_D$ $A_{low} \leq B_{len}$ $A_{high} \geq B_{len}$	-	$B_D \subseteq A_D$ $A_{low} \leq B_{len}$ $A_{high} \geq B_{len}$	-	-	$B_D \subseteq A_D$ $A_{low} \leq B_{low}$ $A_{high} \geq B_{high}$	-	$B_D \subseteq A_D$ $B \subseteq^4 A$
NUMERIC INTERVAL	-	-	-	-	-	$B_D \subseteq A_D$ $A_{low} \leq^3 B$ $A_{high} \geq^3 B$	-	-	$B_D \subseteq A_D$ $A_{low} \leq^3 B_{low}$ $A_{high} \geq^3 B_{high}$	-
PATTERN	-	-	-	-	$B_D \subseteq A_D$ $B \xrightarrow{A} \notin \emptyset$	-	-	$B_D \subseteq A_D$ $B \subseteq^4 A$	-	$B_D \subseteq A_D$ $B \subseteq^4 A$

Evaluation

Configuration: Intel Core 2 Duo T9300 @ 2.50GHz, 4 Gb RAM, OpenSUSE 12.1 (Linux 3.1.10), Java 1.7.0_05 (-Xmx3200M -Xms3200M), Pellet 2.2.0, HermiT 1.3.6

Full ontology: 1,087,124 axioms, 65 classes, 33 object properties, 109 datatype properties and 131,637 individual assertions

Truncated ontology: 230,670 axioms 65 classes, 33 object properties, 109 datatype properties and 36,191 individual assertions

Reasoner	Truncated ontology classification time, ms	Full ontology classification time, ms
ELK	7 366	54 912
ELK[1]	5 598	12 567
Pellet	165 419	out of mem
HermiT	timeout	timeout

[1] Simplified literal reasoning algorithm

Pospishnyi Oleksandr
pospishniy@kpi.in.ua

Source Code:

https://code.google.com/p/elk-reasoner/

‘**elk-parent-datatypes**’ branch