

WP8

Case Study: Cultural Heritage

Aarne Ranta (UGOT)

Summary

Data collection: CRM ontology, Gothenburg City Museum database
(Dana Dannélls, Milen Chechev)

Prototype with natural language generation for English, Finnish, French,
Italian, Swedish (Dana Dannélls, Ramona Enache, Aarne Ranta)

Issues

Human resources for this WP: we suggest extending it to Month 36.

Data collection

D8.1, September 2011

Gothenburg City Museum objects to a database

CIDOC-CRM ontology concept to GF grammars

Convert OWL to GF abstract syntax

Museum query language based on MOLTO KRI

Multilingual text generation

D8.2, March 2012

Resource grammar + text patterns

CRM painting concepts

Future plan: at least 15 languages

With RGL + native informant, a new language takes ca. one hour to implement (150 LOC).

Text pattern implementation

Abstract syntax

```
fun MkGenText :  
  Painting -> Painter -> PaintingType ->  
  OptColour -> OptSize -> OptMaterial ->  
  OptYear -> OptMuseum -> GenText ;
```

English (passive past tense, anaphoric reference):

Painting was painted by Painter on Material in Year. It is of Size and it is painted in Colour. This PaintingType is displayed in Museum.

Finnish (OSV topicalization, use of cases):

Maalauksen Painting maalasi Painter Materialille vuonna Year. Se on kokoa Koko ja se on maalattu Colourilla. Tämä PaintingType on esillä Museumissa.

French (composed past passive, gender agreement):

Le tableau Painting a été paint sur Material par Painter en Year. Il est de taille Size et il est paint en Colour. Ce PaintingType est exposé dans Museum.

Italian (composed past passive, gender agreement, pro-drop, *misurare*):

Il quadro Painting è stato dipinto su Material da Painter nel Year. Misura Size ed è dipinto in Colour. Questo PaintingType è esposto in Museum.

Swedish (s-passive, gender agreement):

Painting målades på Material av Painter år Year. Den är av storlek Size och den är målad i Colour. Den här Painting Type är utställd på Museum.

Soundness of text w.r.t. database

Dependent types of GF

cat

VerifiedText ;

CompletePainting

Painting PaintingType Painter OptYear

OptMuseum OptColour OptSize OptMaterial ;

fun

MkVerifiedText :

(pg : Painting) -> (pr : Painter) -> (pt : PaintingType) ->

(cr : OptColour) -> (se : OptSize) -> (ml : OptMaterial) ->

(yr : OptYear) -> (mm : OptMuseum) ->

CompletePainting pg pt pr yr mm cr se ml -> VerifiedText ;

Future plan: lexicon mining

TermFactory WP3

Lemon WP12?

Future plan: information retrieval

MOLTO-KRI query language (Ontotext) specialized to museum queries
and lexicon

When was "Guernica" painted?

Support for the same languages as text generation

Future plan: multilingual Wiki

Generate Wikipedia articles for museum objects

Use technology from WP11 (UZH)

Integrate

- multilingual text generation
- multilingual search and queries
- collaborative authoring

Future plan: use case

System usable by visitors in the Gothenburg City Museum

- descriptions in 15 languages
- queries and answers in (controlled) natural language

Issues

Human resources for this WP: we have used only 6 pm of 29

Part-time work by Aarne Ranta, Ramona Enache, Olga Caprotti, Dana Dannélls (UGOT), Inari Listenmaa (UHEL), Milen Chechev (Ontotext)

Key person: Dana Dannélls (UGOT) will join 100% in November 2012 (after PhD defence)

We suggest extending WP8 from Month 30 to Month 36 (end 28 February 2012).

Move 3 Ontotext PM's from WP6 (change with UPC)