Building a Semantic Virtual Museum: from Wiki to Semantic Wiki using Named Entity Recognition

Alain Plantec, Vincent Ribaud, Vasudeva Varma

To cite this version:


HAL Id: hal-02912811
https://hal.univ-brest.fr/hal-02912811
Submitted on 24 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A semantic wiki aims to add meaning to the values and the links embedded in wiki pages. Shared semantics in a community relies on the use of a common ontology of universals: the definition of the domain concepts (classes) and of the properties used to represent and associate concepts. When we are building a semantic wiki to share knowledge about a domain, it requires to identify the particulars (instances) about which we have knowledge, then to instantiate classes and properties in order to represent these particulars. Traditional wiki pages contain a list of unstructured knowledge and our research works aims at providing the end-users with methods and tools that might help in extracting the semantic knowledge from regular wiki pages.

A typical wiki page about an ammeter is represented below. Knowledge about this ammeter may be represented with a list of couples: (property, object) such as (has dimension, 50) or (shows features of, an Ammeter). NER helps also to produce information regarding the type of the particulars (attribute or class), e.g. Number or Instrument.

Overview of the translation process

This translation process is performed in three steps: named entity recognition (NER), CIDOC-CRM class recognition, and CIDOC-CRM property disambiguation. The two former steps are automatic but require human validation to ensure that named entities or CRM classes were correctly recognized. The last step is a computer-aided process.

Excerpts of the translation process

An "ammeter" is a [[measuring instrument]] used to measure the [[electric current]] in [ampere(s) [A]], hence the name.

The earliest design is the [[Jacques-Arsène d'Arsonval]] ([[galvanometer]]) or "moving coil" ammeter.

An "instrument=ammeter" is a [[measuring instrument]] used to measure the [[electric current]] in a [[Electrical circuit]].

The earliest design is the [[Person=Jacques-Arsène d'Arsonval] Person=Person[[Person=D'Arsonval]]] or "moving coil" ammeter.

The earliest design is the [[Person=Jacques-Arsène d'Arsonval Person=Person]] ([[galvanometer]]) or "moving coil" ammeter.

The earliest design is the [[E101 Instrument=ammeter]] ([[E101 Instrument]]) or "moving coil" ammeter.

Another ammeter is a [[measuring instrument]] used to measure the [[electric current]] in a [[Electrical circuit]].

The earliest design is the [[E101 Instrument=ammeter]] ([[E101 Instrument]]) or "moving coil" ammeter.

"Amperemeter" was measured in [[Measure=ampere]] and "ammeter".

Another ammeter is a [[measuring instrument]] used to measure the [[electric current]] in a [[Electrical circuit]].

Electric currents are measured in [[Measure=ampere]] and "moving coil" ammeter.

Electric currents are measured in [[Measure=ampere]] and "moving coil" ammeter.

The earliest design is the [[E101 Instrument=ammeter]] ([[E101 Instrument]]) or "moving coil" ammeter.

"[[Category:E101 Instrument]]"

[[Category:E101 Instrument]]

Wiki pages are contents

Yalta Conference

This graph represents the decomposition of the composite "Yalta Conference". If we assume that Actors – such as Roosevelt or Churchill - Things – such as the document Yalta agreements, Image - The Spa and Places are particulars that exist on their own in the wiki, the whole Event called "Yalta Conference" is composed of 2 Activities (the creation of the conference itself and the creation of the Yalta agreements). Hence, any of the components of the composite may be the subject or the object of the RDF triples that can be extracted from the whole graph.

RDF provides constructs to deal with groups of things: containers and images that can be edited (and edited) and images that can be edited (and created) independently. The page can also be manipulated, for example, refereced or printed, as a single entity.

Our approach assumes that all the knowledge embedded in a wiki page is related to the same particular - acting as the subject of all generated RDF triples. But, most pages are composites that combines several entities into a collective entity that can be referenced as if it were atomic.

A typical use of composites in a wiki is to see a page as a collection of subsections and images that can be edited (and edited) and images that can be edited (and created) independently. The page can also be manipulated, for example, refereced or printed, as a single entity.

References