Building Temporally Aware Knowledge Bases in nodegoat

Pim van Bree[1], *Geert Kessels*[1]
(1) LAB1100, Spoorlaan 5D 2495AN Den Haag, The Netherlands
pim@lab1100.com, geert@lab1100.com

Abstract

At LAB1100 we aim to provide scholars in the humanities with powerful tools that allow them to make use of their skills and craftsmanship in a digital environment. Next to data mining and algorithmic approaches, digital applications can also fuel traditional workflows. We equip scholars in the humanities with tools that help them to gather, process and analyse complex and fuzzy primary sources. The main outcome of this philosophy is the tool nodegoat (http://nodegoat.net). We will first introduce this tool followed by the main focus of our proposal.

nodegoat is a web-based data management, analysis and visualisation environment. Using nodegoat, scholars define, create, update, query, and manage any number of data sets by use of a graphic user interface. A custom data model autoconfigures the backbone of notegoat's core functionalities. Within nodegoat scholars are able to instantly analyse and visualise datasets. nodegoat allows scholars to enrich data with relational, geographical and temporal attributes. Therefore, the modes of analysis are inherently diachronic and ready-to-use for interactive maps and extensive trailblazing.

nodegoat follows an object-oriented approach throughout its core functionalities. Borrowing from actor-network theory this means that people, events, artefacts, and sources are treated as equal: objects, and hierarchy depends solely on the composition of the network: relations (Latour, 2005). This object-oriented approach advocates the self-identification of individual objects and maps the correlation of objects within the collective.

Since 2012, nodegoat has been used by the University of Amsterdam to produce diachronic mappings of correspondence networks of nineteenth century intellectuals. The goal of the project is to ‘map the dissemination of cultural nationalism across Europe by charting cultural patterns and networks as they evolve over time’ (http://spinnet.eu/spintimemappings/). The final products show how networks of the intellectuals, all trying to establish clearly defined national communities, transcended existing or newly created national/cultural boundaries.

Together with Dutch research institute Huygens ING, LAB1100 uses nodegoat to map artist networks in 17th century Rome based on existing datasets of ECARTICO (UvA, http://www.vondel.humanities.uva.nl/ecartico/) and HADRIANVS (KNIR, http://hadrianus.it/). nodegoat will be used to connect these databases and to analyse and visualise co-occurances of objects in time and space (http://cdh.uva.nl/projects-2013-2014/knaw---mapping-notes-and-nodes-in-networks.html).

The NIOD Institute for War, Holocaust and Genocide Studies uses nodegoat to map memory landscapes based on interviews of survivors of state and non-state violence. nodegoat allows them to analyse individual accounts and visualise contradictory or conflicting connections within the collective of survivors and perpetrators. This process results in a rhizomatic analysis of knots

of memory (Rothberg, 2010), not seeking for pre-established coherent or socially constructed commonplaces.

Based on our experiences in these and other projects, we have noticed that debate on semantic questions is often focused on the definition of labels and not so much on the underlying content. A well known example is of course ‘nationality’, but also a (historical) occupation/capacity and even seemingly unproblematic concepts like ‘the nineteenth century’ pose several problems.

In order to cope with fuzzy and time dependent concepts, we have developed a methodology that properly handles these concepts. This methodology is implemented in nodegoat and operates over three levels. Firstly, it integrates complex filtering functionalities in the definition of classifications, establishing a dynamic process of segmentation of objects. Secondly, objects can be accessed, analysed and visualised based on the classifications they have been included in. Thirdly, the dynamic interaction between the ‘reversed classifications’ and the objects produces a semantic model that is temporally aware.

In the next paragraphs, we will set out the framework of the methodology of ‘reversed classification’. This will form the focus of our presentation and paper for the workshop on semantic technologies for research in the humanities and social sciences.

In general, classifications emphasise a convention of value and vocabulary. The direction of a classification is outward, relating to the convention unidirectionally. In effect, the classification is unable to communicate/negotiate with the network it classifies. The reversal of classification opens up the convention by disclosing its parameters. Reversal allows the classification to be scrutinised, reconfigured and re-evaluate the objects it classifies.

Simply put: instead of identifying classifications and assigning these to objects in a dataset (like ‘sculptor’ or ‘German’), a scholar defines a multi-faceted filter spanning multiple datasets in which they define any number of parameters that are associated with a classification. This will reverse the classifying process as the definition of the classification is identified by the exchange between parameters of the classification and the attributes of the object.

For the classification of ‘Artist’ this would allow for an inclusion of persons who yield positive on the filter (‘pupil of person who produced paintings’ AND ‘has had paintings in an exhibition’) OR (‘person who studied at School of Visual Arts’ AND ‘has had video installations in an exhibition’). No external labeling or self-labeling of the classification of ‘Artist’ is needed.

By making use of circumstances depending on time and location, a reversed classification is able to accompany variating configurations relating to place and time. Reversed classifications can be employed to cluster people and organisations who are subject to considerable changes over time and in affiliation like activists/freedom fighters/terrorists. Correspondingly, the location and date of a circumstance can also be reversely classified and retrieved from the configuration of the classification. Objects that match ('artifact excavated at a depth of X meters' AND 'in the region of Susa'), could for example be classified with Achaemenid Empire (depending on X) and use the date configured in that classification as its own. The same goes for locations. The domicile of a person can be assessed using titles a person may have had. A location specific title like 'Gouverneur-generaal’ would be classified as ‘Place of residence: Dutch East Indies’.

This reversal works well for concepts and periodisation. When dealing with concepts bound to change over time and space, a reversed classification can be configured to match 17th century objects differently than objects in the 18th century. To facilitate discussions on the definition of periodisations, instead of retagging objects, only the classification has to be reconfigured to match the latest consensus.

This semantic model is as flexible as the understanding of a historian working on a dataset that spans multiple centuries. Just like the historian knows that the twentieth century occupation ‘physician’ means something completely different than the 17th century capacity of ‘physician’, nodegoat is able to handle these concepts accordingly.

References

Latour, B. (2005). Reassembling the Social. Oxford: Oxford University Press.
Rothberg, M. (2010). Introduction: Between Memory and Memory: From Lieux de mémoire to Noeuds de mémoire. Yale French Studies, 118/119, 3-12.

KEYWORDS: Actor-Network Theory, Object-Oriented, Reversed Classification, Visualisation, Research Environment.