Ontology Matching for Big Data Applications in the Smart Dairy Farming Domain

Jack P.C. Verhoosel, Michael van Bekkum and Frits K. van Evert

TNO Connected Business, Soesterberg, The Netherlands
{jack.verhoosel,michael.vanbekkum}@tno.nl
Wageningen UR, Wageningen, The Netherlands
frits.vanevert@wur.nl

Abstract. This paper addresses the use of ontologies for combining different sensor data sources to enable big data analysis in the dairy farming domain. We have made existing data sources accessible via linked data RDF mechanisms using OWL ontologies on Virtuoso and D2RQ triple stores. In addition, we have created a common ontology for the domain and mapped it to the existing ontologies of the different data sources. Furthermore, we verified this mapping using the ontology matching tools HerTUDA, AML, LogMap and YAM++. Finally, we have enabled the querying of the combined set of data sources using SPARQL on the common ontology.

1 Background and context

Dairy farmers are currently in an era of precision livestock farming in which information provisioning for decision support is becoming crucial to maintain a competitive advantage. Therefore, getting access to a variety of data sources on and off the farm that contain static and dynamic individual cow data is necessary in order to provide improved answers on daily questions around feeding, insemination, calving and milk production processes.

In our SmartDairyFarming project, we have installed sensor equipment to monitor around 300 cows each at 7 dairy farms in The Netherlands. These cows have been monitored during the year 2014 which has generated a huge amount of sensor data on grazing activity, feed intake, weight, temperature and milk production of individual cows stored in databases at each of the dairy farms. The amount of data recorded per cow is at least 1MB of sensor values per month, which adds up to 3.6GB of data per dairy farm per year. In addition, static cow data is available in a data warehouse at the national milk registration organization, including date of birth, ancestors and current farm. Finally, another existing data source contains satellite information on the amount of biomass in grasslands in the country that is important for measuring the feed intake of cows during grazing.

We focused on decision support for the dairy farmer on feed efficiency in relation to milk production. Thus, the big data analysis question is: “How much feed did an individual cow consume in a certain time period at a specific grassland parcel and how does this relate to the milk production in that period?”.

2 Ontology matching approach

We selected one of the dairy farms (DairyCampus) and created with TopBraid composer a small ontology with 12 concepts that covers among others the grasslands of a farm and grazing periods of cows. This ontology contains the concept “perceel” which is Dutch for parcel. In addition, we selected the data source with satellite information about biomass in grasslands (AkkerWeb, www.akkerweb.nl). This data source already had an ontology defined with 15 concepts that contains the concept “plot” which is similar to parcel but with different properties. Furthermore, we created with TopBraid composer a common ontology for the domain with 28 concepts on feed efficiency (see **Fig. 1**).

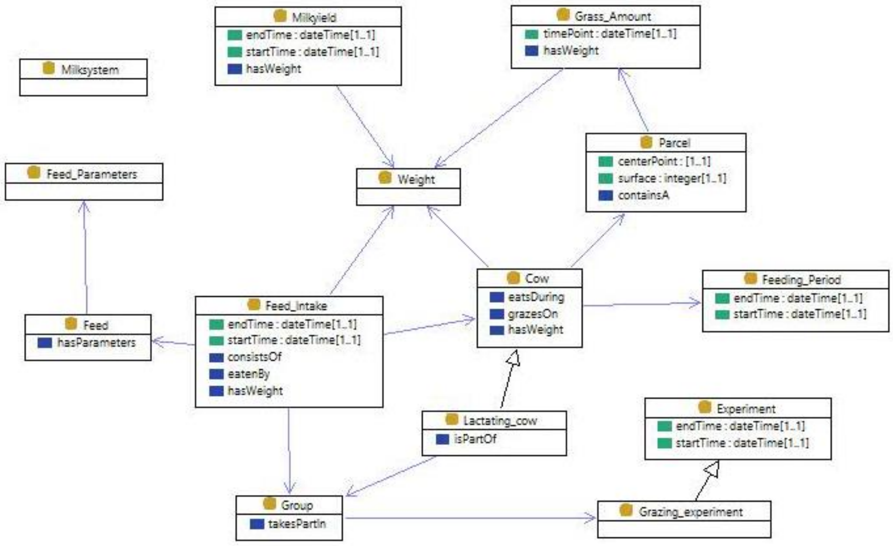

Fig. 1. Common ontology excerpt for feed efficiency in dairy farming.

The challenge was to find a match between the concepts and properties in the common ontology and both specific DairyCampus and Akkerweb ontologies, especially regarding the concepts “parcel”, “perceel” and “plot”.

We have initially created manual mappings between classes and properties in TopBraid using rdfs:subClassOf and owl:equivalentProperty relations. Based on relatively few and simple matches we created initial alignments between properties and classes (see **Fig. 2**).

Use of a matching tool or system however, provides us with opportunities to verify our current findings and better support our efforts in finding alignments between the other concepts in our ontologies. We used a literature survey of matching techniques and supporting matching systems in [1] to identify both a suitable matching technique and find tools supporting that technique. We consider language-based matching as the appropriate type of matching since it focuses on syntactic element-level natural language processing of words.

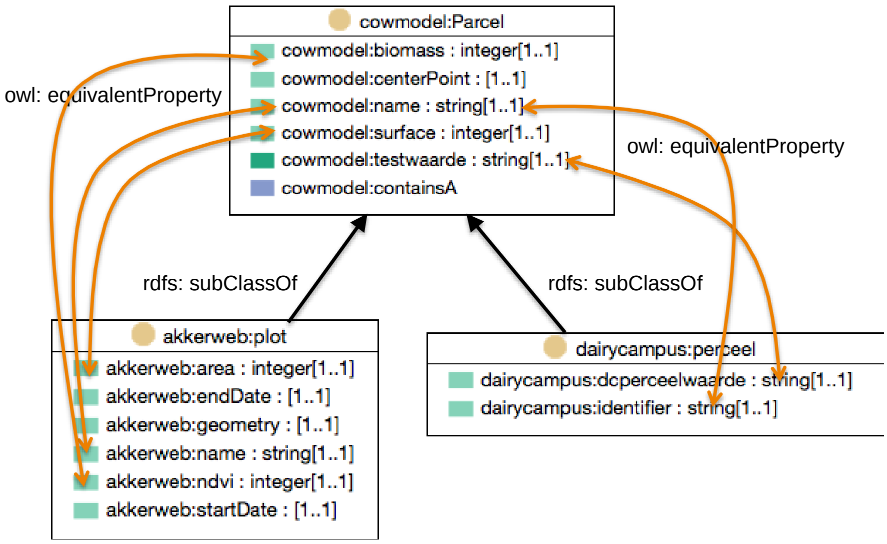

Fig. 2. Mapping of classes and properties based on the matching result.

There are numerous tools available that support this specific matching technology, mostly from academic efforts. Some however are no longer in active use, either being outdated or not maintained anymore [2].

We have selected several matching systems that support our requirement of language-based matching: HerTUDA [3,4], AgreementMaker Light (AML) [5], LogMap [6], and YAM++ [7]. We have started to investigate the possibilities of these tools to find alignments of concepts and properties in our ontologies. Initial efforts with the concepts shown in **Fig. 2** have not led to successful matches and alignments yet, however. The HerTUDA, LogMap and YAM++ tools were difficult to install and execute. The AML worked fine, but could not entirely find the relation between “parcel”, “perceel” and “plot”. Further analysis is required to find out whether this is due to inappropriate matching techniques or to the specific ontologies that we offered to the tool.

3 SPARQL queries and triple stores

In order to show that the mapping of the common ontology to the specific ontologies works properly, we generated in Topbraid a few instances of an Akkerweb plot and a DairyCampus perceel. In addition, we build a simple select query using the common ontology to retrieve all parcels and for each parcel the properties name, biomass, surface and test.

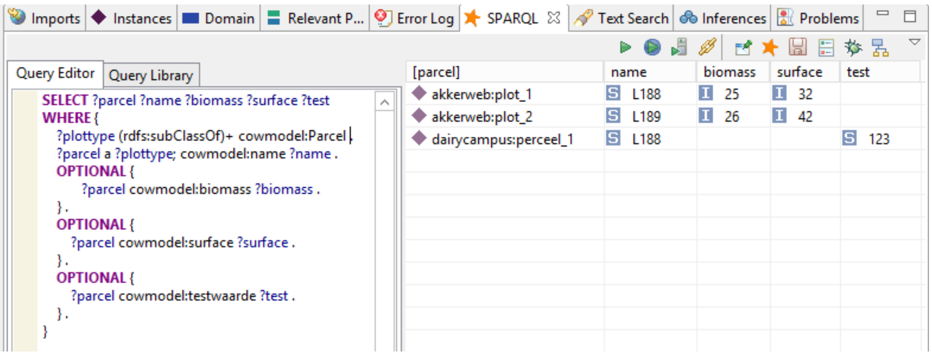

Fig. 3. Select query on common ontology to retrieve all parcels.

The query and its results are shown in **Fig. 3**. As can be seen, the query retrieves both Akkerweb plots and DairyCampus percelen. In addition, Akkerweb contains data about a plot with name "L188" and DairyCampus contains data on a perceel with an identifier "L188". This means that both databases contain the same parcel and the properties can be combined.

The specific ontologies for DairyCampus and Akkerweb formed the basis to generate triples from the relational data sources of DairyCampus and Akkerweb. The triples have been made available via Virtuoso as well as directly from the D2RQ tool (www.d2rq.org). A system that is based on the common ontology can take the big data question to create federated SPARQL queries on the DairyCampus and Akkerweb triple stores using the matched ontologies. As a result, farmers can pose questions in terms of the concepts in the common ontology instead of the detailed and specific concepts of the DairyCampus and Akkerweb data sources.

The farmer can use such a system for decision support purposes on various daily operations, such as which amount of feed to provide to which cow in which period, when to inseminate a specific cow and how to deal with the transition of a cow towards calving.

4 Future work

The approach that is describe in this paper is currently in an experimental phase. We have reached a set-up by filling the triple stores for 3 farms with cow-data of 1 month which adds up to a total of 7 million triples. This needs to be upgraded to all farms with all data from 2014. Thereby, we can test the scalability of our system. In addition, we need to do more detailed analysis of the matching tools that we used and the reasons for not adequately solving the simple matching problem that we proposed.

References

1. Otero-Cerdeira, L., Rodriguez-Martinez, F.J., Gomez-Rodriguez, A.: Ontology matching: A literature review. Journal on Expert Systems with Applications, 949-971 (2015)
2. Ontology matchings tool overview: www.mkbergman.com/1769/50-ontology-mapping-and-alignment-tools/

3. Hertling, S.: Hertuda results for OAEI 2012. In Ontology Matching 2012 workshop proceedings, 141-144 (2012)
4. HerTUDA download: www.ke.tu-darmstadt.de/resources/ontology-matching/hertuda
5. AgreementMakerLight website: somer.fc.ul.pt/aml.php
6. LogMap website: www.cs.ox.ac.uk/isg/tools/LogMap/
7. YAM++ website: www.lirmm.fr/yam-plus-plus