Computers and Electronics in Agriculture 152 (2018) 82–89

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Agricultural recommendation system for crop protection

Javier Lacasta*, F. Javier Lopez-Pellicer, Borja Espejo-García, Javier Nogueras-Iso, F. Javier Zarazaga-Soria

Aragon Institute of Engineering Research (I3A), Universidad de Zaragoza, Spain

ARTICLE INFO

Keywords:
Ontology creation
Ontology population
Data integration
Intelligent systems
Pest control

ABSTRACT

Pests in crops produce important economic loses all around the world. To deal with them without damaging people or the environment, governments have established strict legislation and norms describing the products and procedures of use. However, since these norms frequently change to reflect scientific and technological advances, it is needed to perform a frequent review of affected norms in order to update pest related information systems. This is not an easy task because they are usually human-oriented, so intensive manual labour is required. To facilitate the use of this information, this work proposes the construction of a recommendation system that facilitates the identification of pests and the selection of suitable treatments. The core of this system is an ontology that models the interactions between crops, pests and treatments.

1. Introduction

Agriculture is a vital sector in the economy of any country, but depending on the crop between 26% and 80% of the agricultural production is lost because of pests (Oerke, 2006). Crop protection is vital but also challenging due to the multiple pests that affect them, such as insects, plant pathogens and weeds, and the toxic effects of most of the existing solutions (Alavanja, 2009). Because of these effects, most countries have established strict regulations for their use and promote non-chemical solutions (European Parliament, 2009).

In general, the norms about pest control are published in heterogeneous and human oriented formats, so intensive manual labour is required to identify the most suitable solution for a given pest. An example of this heterogeneity can be found in the data collections provided by the Spanish Ministry of Agriculture[1] where the description of how to control each type of pest is distributed among multiple heterogeneous textual sources. For example, each document has a layout slightly different from the rest and the names of the pests in the document title are variants of those used in the pest description. This lack of interoperability affects critically tasks requiring some degree of data integration such as identifying the different crops affected by a single organism, finding similitude in the treatment of different species, and comparing the approved pesticides in different countries. Additionally, as new products and techniques are frequently approved, a continuous review is required (Ricci et al., 2010). This happens not only in Spain, but also in many other countries such as United Kingdom,[2] United States[3] and Canada.[4]

To facilitate the usability of this information, we need systems able to provide it in an integrated and harmonized way. For this task, in this paper, we propose the "Pests in Crops and their Treatments" Ontology (PCT-O). To populate it, we suggest a conversion process for the transformation of non-ontological heterogeneous resources into ontological ones. As use case, this process is applied to transform content from selected Spanish data sources into instances according to PCT-O model. Finally, we describe the structure of the information retrieval (IR) system and the recommendation process that simplifies the identification of a pest and the selection of a suitable treatment.

2. State of the art

The use of ontologies is a classical solution to deal with heterogeneity and interoperability problems. In the biology area, Walls et al. (2012a) remark how semantic models facilitate the creation of intelligent applications that manage living species information. The inference capability of ontologies are especially relevant in the biology area, because it can be used in the taxonomic structures used for

* Corresponding author.
E-mail address: jlacasta@unizar.es (J. Lacasta).

[1] http://www.mapama.gob.es/.
[2] https://secure.pesticides.gov.uk/pestreg/.
[3] https://www.epa.gov/pesticide-registration.
[4] https://www.canada.ca/en/health-canada/services/consumer-product-safety.html.

https://doi.org/10.1016/j.compag.2018.06.049
Received 7 December 2017; Received in revised form 4 June 2018; Accepted 29 June 2018

classification to simplify conceptual interoperability, data integration and search. However, the creation of ontologies is difficult. The main challenges are the modelling of the information for the desired task, the availability of data for population, and the data transformation complexity. Data modelling is difficult due to different interpretations of the selected knowledge area. With respect to data availability, the availability of data sources conditions the extension and depth of a semantic model. Something similar happens with data transformation. Too complex or too heterogeneous data collections may not be added to the model due to transformation costs.

Several works in the literature categorize living species, the interactions between them or the effects produced by chemical substances. This section describes the main works in these fields, remarks the parts of these models that can be used to describe pest control information, and indicates the shortcomings solved by the proposed PCT-O.

With respect to living being descriptions, the Integrated Information Taxonomic System (ITIS) (Integrated Taxonomic Information System, 2010) contains taxonomic information of aquatic and terrestrial flora and fauna, the Catalogue of Life model (Jones et al., 2000) describes 2 million of species, and the NCBI taxonomy (Gene Ontology Consortium, 2004; Federhen, 2012) stores the organism names and taxonomic lineages in the INSDC database. All these models provide a comprehensive collection of species but they do not provide very detailed information about their features and behaviour. The search capabilities of the portals providing them are limited to the use of names or database codes.

Other works provide extended taxonomies with additional information such as species descriptions, biology, lifecycle, habitat, and interaction with other species. An example of this type of works is Wikispecies (Wikimedia Foundation, 2017), which contains near half a million of species, although the information provided for each species is limited. Focusing on plants, the U.S. plants database (Natural Resource Conservation Service, 2016) includes a quite detailed textual description of U.S. plant, their distribution, life cycle, and common pests. Another system is the European Nature Information System (EUNIS) (Davies et al., 2004). It includes a large collection of species obtained from other databases and indicates the geographical distribution and the level of extinction threat of those species. A relevant work is the Encyclopedia of Life (Li et al., 2004), which provides more detailed information about a million of species and even a basic description of the interaction between species. However, it does not detail the kind of interaction they have (predator, prey, symbiosis, and so on). Sini (2009) describes the AGROVOC vocabulary, an agriculture thesaurus. A part of it provides a taxonomy of living beings that includes the main used crops and pests in the form of hierarchically related concepts. DBpedia (Auer et al., 2007) also contains a formal structure for the information about living species in Wikipedia and Wikispecies. However, the number of provided species is more limited. Finally, GeoSpecies (DeVries, 2013) relates each concept to the Encyclopedia of Life, Wikipedia, Wikispecies, NCBI, ITIS, and other similar systems. Instead of providing proper information about the stored species, it focuses on providing equivalences between the aligned models. The search capabilities in these systems are more complete, allowing textual search in the data content. In the semantic models, such as AGROVOC, DBpedia and GeoSpecies, arbitrary searches are also possible.

Some works specifically focus on the interactions between species. Rodríguez-Iglesias et al. (2017) propose an ontology that details the pathogens that affect plants. It integrates data related to both plant physiology and plant pathology with the objective of facilitating the interpretation of phenotypic responses and disease processes. Similar to this, Walls et al. (2012b) analyse the infectious diseases of plants and the pathogens that cause them. They reuse vocabularies from other plant, pathogen and disease ontologies such as the Infectious Disease Ontology (IDO) (Cowell and Smith, 2010). Finally, the Plant Ontology Consortium (2002) defines a set of ontologies to describe plants, their genes, diseases and growing process that include the relation between plants and harmful virus and bacteria. All these models, as in the previous cases, provide semantic searches that make possible detailed queries and precise results.

With respect to crop treatments, PubChem model (Fu et al., 2015) describes chemical structures, biological activities and biomedical annotations. This includes pesticides and the environmental effects they produce. However, this information is text-based and it is not linked to any living species model. ChEBI ontology is another model describing chemical substances (Degtyarenko et al., 2008). It contains natural molecular entities and synthetic products that affect living organisms. However, it also lacks a semantic relation with the species affected by each chemical product. Here, depending on the part of the models, textual or semantic searches are possible.

Other works integrate parts of all these and other agricultural aspects together. Damos (2013) proposes the definition of ontologies that allow describing all the characteristics of cultivations. He also indicates the need to link the created models to other related data collections that complement them. Damos et al. (2017) show an ontology to describe pest and the treatments approved by the Greek Ministry of Rural Development and Food. The core of the ontology contains the pests that are related to the affected crops and existent treatments. On a broader context, Athanasiadis et al. (2009) describe several ontologies for data integration in the agricultural field. Especially relevant is their agricultural activities ontology for crop management. Goumopoulos et al. (2009) describe an ontology for precision agriculture. It focuses on describing plants and all the technological and electronic devices that surround them in precision agriculture. Finally, Rehman and Shaikh (2011) describe another precision agriculture ontology whose core includes concepts for describing crops and their pests.

The objective of the ontology proposed in this paper (PCT-O) is to connect crops, pests and treatments into a unified model. The formal description of living species taxonomies can be managed with the previously described ontologies such as NCBI taxon or GeoSpecies, the description of plant pathologies is covered by Rodríguez-Iglesias et al. (2017) illnesses ontology, and PubChem covers the application of chemical substances. However, they do not model all the crop protection aspects. Specifically, they do not cover the relation between crops, pests that affect them, and the solutions approved by each country to deal with them. Only Damos et al. (2017) make a proposal to relate information about pests and treatments to the affected crops. However, they propose a high-level model that does not provide detailed properties about each of the proposed classes. The proposed PCT-O allows describing the conditions required by a pest to produce outbreaks and the restrictions on the treatments.

3. Structure of the PCT-O

This section describes the ontology created for the description of pests, crops and their treatments. The core of the proposed model can be considered as an extension of the disease triangle described in Rodríguez-Iglesias et al. (2017), which consists of a virulent pathogen, a susceptible host, and a propitious environment. It has been extended to include non-pathogen pests and the definition of treatments for the pests. We have also modelled the provenance of the information to allow updates and correction of errors in the sources and in the generation process.

The ontology has been created with the Methontology methodology (Gómez-Pérez et al., 2004). Specifically, the modelling has been guided to answer the following competence questions: Which is the pest that is affecting a given crop? Which treatment do I have to apply to deal with the pest? When do I have to apply the treatment? What are the sanitary/environmental restrictions of the treatment?

In the construction process of the PCT-O, we have put a special emphasis on reusing existing models to improve the ontology interoperability. Specifically, we have analysed widely used models of living species (which include both crops and pest) and chemical substances