Identifying explicit and tacit knowledge in a life science knowledge base

Johanna Saoud[1], Alain Gutierrez[1], Marianne Huchard[1], Pierre Silvie[2,3] and Pierre Martin[2]

[1] LIRMM, Univ Montpellier, CNRS, Montpellier, France
[2] CIRAD, UPR AIDA, F-34398 Montpellier, France
AIDA, Univ Montpellier, CIRAD, Montpellier, France
[3] PHIM Plant Health Institute, Montpellier University,
IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

Corresponding author: johanna.saoud@etu.umontpellier.fr

An alternative to the use of synthetic pesticides and antibiotics in agriculture is to spray local plants extracts, in aqueous or essential oil form. To this end, the Knomana knowledge base [1] compiles various knowledge sets on plant use such as the 42000 descriptions of pesticidal plant uses for plant, animal, and public health presented in the literature. As the One Health approach dictates to be aware of the additional uses of these pesticidal plants to prevent their unintended effects on the animal, the human, and their environment, the challenge for the domain experts (e.g. entomologist, pathologist) is thus to identify the pesticidal plants in Knomana considering the One Health approach.

With the aim to present knowledge to the expert using a compact and comprehensive formalism, in [2], we computed the Duquenne-Guigues basis (DGB) of implications on an excerpt of Knomana, in which each plant is described using its taxonomy (i.e. species, genus, and family), to be consumed as food, and to be used in medical care. The DGB method is based on Formal Concept Analysis (FCA) and provides a cardinality-minimal set of non-redundant implications. By considering a reduced knowledge set, this work identified 3 types of knowledge elements in the implications: knowledge on plant use at diverse taxonomy levels (e.g. *Plants from Meliaceae family are not consumed as food*), plant taxonomy (e.g. *A plant from Salvia genus is from Lamiaceae family*), and side effect of the knowledge set, e.g. *a plant from the Piperaceae family is from the genus Piper*. This latter illustration is not in accordance with taxonomic referential and thus informs on the extend of knowledge inserted in Knomana. Moreover, as plant taxonomy is known by the experts, removing it from the implications eases their reading but makes it tacit knowledge.

Implementing this method to select pesticidal plants requires to consider Knomana as a multidimensional (ternary) dataset, and thus to use the extension of FCA devoted to this kind of knowledge discovery, i.e. Relational Concept Analysis (RCA). Therefore, computing the DGB of implications based on RCA provides linked set of implications which includes the existential quantifier. Converting this formulation as practical expression is a need for the domain experts.

This poster describes the product line that formulates Knomana knowledge on pesticidal plants as implications, from which the implicit knowledge elements were removed and the side effects are highlighted to alert the expert. This product line was developed using the library fca4j from Cogui software (http://www.lirmm.fr/cogui/), that provides the RCA based DGB of implications, and using a post-process which differentiates the 3 types of knowledge elements within the implications. As an illustration, this poster presents the implications on *Spodoptera frugiperda*, a highly polyphagous insect that is close to invade South of Europe. The perspective of this work is to identify pesticidal European plants species that share chemical components similarities with plants used to control this pest in its native area.

References

[1] Pierre J. Silvie, Pierre Martin, Marianne Huchard, Priscilla Keip, Alain Gutierrez, and Samira Sarter. Prototyping a knowledge-based system to identify botanical extracts for plant health in sub-saharan africa. *Plants*, 10(5), 2021.

[2] Johanna Saoud, Alain Gutierrez, Marianne Huchard, Pascal Marnotte, Martin Silvie, and Pierre Martin. Explicit versus Tacit Knowledge in Duquenne-Guigues Basis of Implications: Preliminary Results. Montpellier, France, May 2021. Submitted to Analyzing Real Data with Formal Concept Analysis, RealDataFCA'2021.