Expert-Based Ontology Construction: a Case-Study in Horticulture

Nicole J.J.P. Koenderink*, Jan L. Top*,†, Lucas J. van Vliet§
*Wageningen University, P.O. Box 17, 6700 AA Wageningen, The Netherlands,
Email: Nicole.Koenderink@wur.nl
†Vrije Universiteit, De Boelenlaan 1083a, 1081 HV Amsterdam, The Netherlands
§Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands

***Abstract*— Experts are capable of performing complex tasks in their specific field of expertise. To do this, they use a vast amount of explicit and tacit domain knowledge. For various applications it may be interesting to represent such detailed domain knowledge in a formal way. We show here the process of elicitating expert knowledge and constructing a domain ontology for a case-study in which experts assess the quality of young greenhouse plants. We have interviewed sorting experts from different plant breeders, created individual ontologies, merged these ontologies, added relevant relations from an observer's point of view and checked the results in a teach-back session. We draw two main conclusions from this work. The first conclusion is that the tacit part of an expert's knowledge is often explicit knowledge for another expert. The resulting merged ontology is richer than the individual ontologies. The second conclusion is that it is essential to involve an objective observer in the creation of the ontology for adding relations to the ontology that are relevant for the final purpose of the ontology, but that are part of the tacit knowledge of the experts.**

I. Introduction

When a knowledge intensive task has to be executed by human experts, they can draw on a range of tacit and explicit knowledge to perform the task successfully. Unfortunately, due to the experts' unique expertise, it happens that when two experts perform a task independently, their view of the situation varies and two different results are obtained. Due to this aspect, the result of the task might be less objective than desired. In our case study, experts assess the quality of young greenhouse vegetable plants. This process is important, since the selection of young plants on their potential productivity increases the yield of the eventual crop. Because of the above mentioned issue, the quality grade of the seedlings could vary up to 10% between two experts.

The purpose of the research presented in this paper is to create an ontology of the expert knowledge concerning the plant morphology and sorting parameters. The thus created domain model will be used as input for a computer vision system that automatically performs the expert's task as proposed in [1]. The overall purpose of the computer vision system is to guarantee objective and consistent use of the domain knowledge and objective quality assessment.

In our case study, we spoke with experts from different plant breeders, each having a unique view on the sorting task. We decided to create for each expert an individual ontology that contains those aspects of the plant structure and plant parameters that he or she deems relevant for the quality assessment task. In the second phase of the process, we have merged these individual ontologies into one all-encompassing knowledge model. We then enriched this merged ontology with relations that were inferred by direct observations of the 3D-images of young plants.

For the plant and seed growing industry, it is important that all seedlings are assessed with one method. At the same time, this method has to be based on the unique quality assessment knowledge of each expert. To ensure this, the final knowledge model should be agreed upon

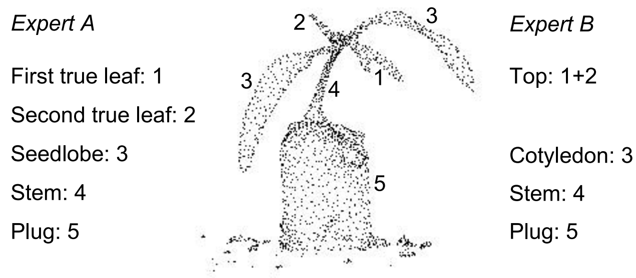

Fig. 1. The plant structure as seen by two different sorting experts.

by all experts. Therefore, we organised a teach-back session in which consensus and commitment of the experts was obtained. A second objective of the teach-back session was to resolve any differences between the expert's knowledge and our interpretation of their knowledge.

In this paper, we present the process of creating one ontology that represents the domain knowledge of a group of experts as accurately as necessary for the execution of their task by describing it for a specific, professional application.

II. Related work

Many papers exist on knowledge elicitation methods, both from the early days of expert system development, and from the present work on ontology construction. We mention here the paper by Lau and Sure [2] who describe the process of creating a skills ontology from three groups of domain experts. Our work differs from theirs in the sense that we involve an objective observer to identify relevant relations with respect to the final purpose of the ontology.

For the knowledge acquisition process, we used the interview technique described by Scott *et al* [3]. Part of each interview session consisted of the expert performing his or her task in our presence, while explaining his or her actions. This technique was inspired by the 'think aloud method' as described by Van Someren *et al* in [4].

We use the CommonKADS method specified by Schreiber *et al* [5] to guide the knowledge engineering process. The creation of the individual ontologies and the merged ontology is in agreement with the proposed modelling method in Noy and McGuinness [6]. We used the Web Ontology Language as specified in [7] to express the ontology.

Within the Semantic Web community, many papers exist on automated merging of ontologies. Most papers on automated merging or matching domain knowledge need a large set of instances or a considerable number of natural language documents as starting material. In our relatively small individual ontologies we have neither. Therefore, the (semi-)automatic merging methods were not applicable to our case study, and as a consequence, we have merged the ontologies by hand.

III. Case study

A. Interviewing the Experts

We have interviewed sorting experts from thirteen different seed growing or plant breeding companies. We consider a person to be an expert when he or she performs the sorting task on a regular basis. For the interviews, the open interviewing technique [3] was used. The experts had to explain the relevant plant part parameters for the sorting task. They had access to a tray of seedlings, which they used to illustrate their explanations. We photographed the interesting plants of these trays to create a data set to obtain objective material to supplement the interviews.

B. Ontology Construction

We analysed the thirteen interviews and constructed for each interview a separate ontology. To create these individual plant ontologies, we first extracted all plant concepts from each interview and used these to form the rudimentary structure for the ontology. In Figure 1, the plant part models as sketched by two of the thirteen experts

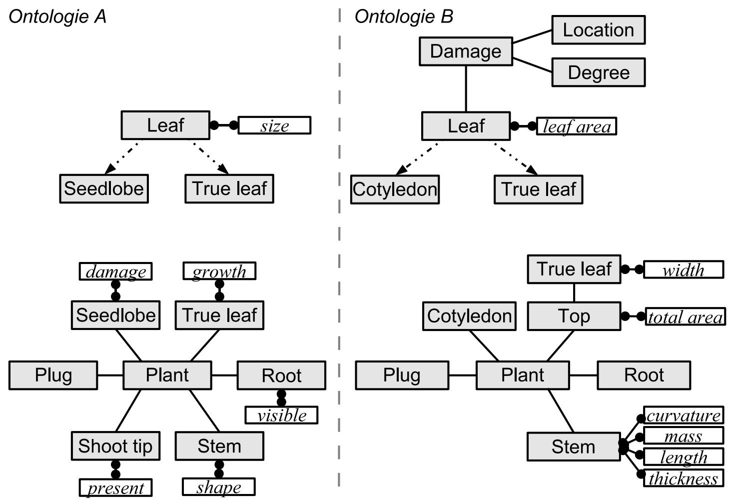

Fig. 2. Schematic representation of two individual plant ontologies.

are depicted. The differences between these two sets of concepts illustrate the two main issues that we encountered in the interviews. Firstly, we noticed that two experts may have a different set of concepts to describe a plant. One expert considers for example the two top leaves of a seedling as separate entities *First true leaf* and *Second true leaf* respectively, whereas a second expert describes them as one entity, the *Top* of the plant. Secondly, the used terminology differs between any two experts. For example, the leaves that are indicated with a '3' in Figure 1 are referred to as *Cotyledons* by Expert A but as *Seedlobes* by Expert B.

The second step in creating the individual ontologies consisted of analysing the interviews. We added attributes and relevant relations between concepts to the initial ontology. In Figure 2, we see a representation of part of the ontologies from Figure 1. This time, the relations between and the attributes of the concepts are displayed. We see that the structure of the two ontologies differs, since not all concepts of ontology A are present in ontology B and vice versa. Another difference stems from the variation in emphasis on the relevance of the parameters: where in ontology *A* damage may only occur at the level of the *Seedlobe*, in ontology B damage to the true leaves is also taken into account and the location and degree of the damage are also important. Finally, we notice that the concepts in the ontologies may contain different attributes. Sometimes, the attributes are synonyms of each other. This holds for the attributes *size* and *leaf area* of the concept *Leaf*. In other cases, attributes exist in only one of the two ontologies.

Although the individual ontologies contain different descriptions of a seedling, each description represents a plant model with which the expert could successfully fulfill the sorting task. Consequently, the thirteen individual ontologies correspond to thirteen valid expert's views on the seedling domain.

C. Ontology Merging

The first step in merging the thirteen individual ontologies was to identify the concepts of the new merged ontology, choose appropriate names for them and indicate the other name(s) as synonym(s). The concept *Cotyledon*, for example, had six different synonyms in the interviews[1]: Cotyledon, Seedlobe, Lobe, Ear, *et cetera*. Adding the synonyms to the merged ontology is a necessary step to be able to use the ontology to interpret in a later stage sorting rules that have been expressed by the experts in their own vocabulary.

We incorporated the different plant models from all individual ontologies by introducing hierarchical subdivisions of plant concepts. The concept *Top* for example

[1]In Dutch: cotyl, kiemlob, kiemlobbe, lob, lobblad, oor, zaadlob

consists of all *True leaves*. By adding a concept *Top* in the ontology of Expert A, we did not change the perception of the plant fundamentally, but just added an additional layer in the plant structure. We noticed that most differences between the concepts described in individual ontologies could be resolved by adding such layers.

When all concepts had been added to the merged ontology, for each concept the attributes and relations had to be defined. For this, we had to decide on the domain of the relations. Due to the different views of the experts on the plant structure, the domain of the relations differs. As an example, we consider the situation described in Figure 2. The relation *has-true-leaf* has as domain either *Plant* or *Top*. Another modelling decision needed to be made on the range of the attributes. Some attributes had been defined in some individual ontologies qualitatively and in other ontologies quantitatively. This was the case for the attribute *length* of the concept *Stem*. The value of this attribute can be expressed quantitatively as a number (in cm), or qualitatively in comparison to the average stem length. Deciding on the range of an attribute is important, since it may influence the handling of the data in a later stage. When *e.g.* the *length* attribute is expressed as a qualitative value (*e.g.* 'large', 'small'), a new attribute *average stem length* should be added to *Tray*, to obtain a comparative value.

With the merging of the concepts and relations, a rough outline of the final merged ontology had been obtained. The last step in the modelling process dealt with refining the merged ontology to accommodate certain implicit nuances that were used by some of the experts. The concept *Damage*, for example, was replaced by the notions of *Localised Damage* and *Global Damage*. The experts indicated that for certain defects the location of the defect is important in the quality assessment, where this was not true for other types of defects.

Using this procedure, we have created a merged ontology that contains all information and as much of the structure of the individual ontologies as possible. This way a detailed description of the required plant structure knowledge for the quality assessment task has been obtained.

D. Observer Perspective

After completing the merging process, we had obtained objective understanding of the relevant issues for the assessment task. Next, we studied the data set, consisting of 366 two-dimensional and 46 three-dimensional plant images, through the eyes of an unprejudiced observer. With the final use of the ontology in mind, *i.e.* an automated sorting system, we searched the data set for relations that were too trivial for the experts to mention, but that were essential to the ultimate goal. Relations that we found were *e.g.* the relation that a *Cotyledon* is connected to a *Stem*, and the fact that a *Plant* can have only one *Stem*. Since these additional relations were deduced from observation, we had to confirm those relations in the teach-back session.

We studied the data set after the creation of the merged ontology, since it was essential to have a rough idea of the relevant aspects of the expert's task, before any meaningful observations could be made.

E. Teach-Back Session

In the teach-back session, we showed the experts the merged ontology enriched with the knowledge from the data analysis. The goal of this presentation was twofold. On the one hand we wanted to obtain feedback on the ontology to ensure that the knowledge had been interpreted correctly from the interviews and data. On the other hand, we wanted to use this opportunity to reach consensus on the relevant domain knowledge for the sorting task.

During the teach-back session, some concepts had to be refined. The experts agreed for example that the concept *Leaf stem*, that connects a leaf to the main stem was part of the *Leaf* and not of the *Stem*. Some relations and attributes were removed from the merged ontology. The attribute *is beautiful*, *e.g.*, is not accurately definable and was therefore rejected. In two cases, attributes were discussed on which the sorting experts could not reach a consensus. These were the attributes *uniformity* of a *Tray*, which was outside the scope of the specialists' expertise, and the attribute *delicacy* of a *Plant*, which is so complex that it must be defined for each cultivar separately. The approval of the observed relations and attributes of the concepts from the data set was without problems, since these were for the experts of almost trivial nature.

In general, we noticed that the experts easily reached consensus. This was on the one hand caused by the fact that for the most part, they recognised the synonyms and view points of each other. We believe that this was partly the result of an ongoing communication process in the sector. On the other hand, the experts had become so involved in the modelling process, that they were keen on making the project a success and on making compromises to reach an optimal, all-encompassing representation of the sorting domain.

Another observation is that the official guidelines that experts are expected to follow cover only a small part of the situations occurring in practice. Often the experts use additional knowledge, referred to as 'intuition' or 'experience'. The confrontation between several experts has allowed us to uncover part of this knowledge.

IV. Discussion

In this paper we have presented the overall process of formalising expert knowledge in the domain of horticultural quality assessment. Our approach combines the ComonKADS knowledge engineering approach with ontology construction. The latter replaces the rather informal methods that were used before in knowledge acquisition and proofs to be quite efficient. The stepwise process of identifying concepts and properties has enforced rigour and precision to the process.

Interviewing many experts has resulted in two types of benefits. Firstly, by combining several different viewpoints we succeeded in obtaining a relatively rich domain model, which at the same time can easily be extended for future purposes. Secondly, tacit knowledge of one expert could be filled in by explicit knowledge of another.

Another novel element of our knowledge acquisition approach was to include the objective observer's perspective. Given the goal of the assessment task and an elementary understanding of the domain, the objective observer identifies relations that were left implicit by the experts. Possibly, this kind of relations could also have been detected in a semi-automated way. Of course a teach-back session is needed to validate these additional relations.

The teach-back session was critical to validating the merged ontology. However, equally important was the effect of this session on the experts, since their support and commitment increased significantly.

In conclusion, we have shown the process of creating a knowledge model using ontologies in an ambiguous and informal domain. This approach can be applied in other domains with similar results: a detailed representation of expert knowledge that was previously largely implicit and heterogeneous, combined with enthusiasm and willingness from the experts involved.

Acknowledgments

The work was supported by the Dutch Ministry of Economic affairs (IOP Beeldverwerking, IBV02010), the Dutch Ministry of Agriculture, Nature and Food Quality (DLO 391), and Plantum (Glasgroentegroep). We thank the experts for their input, M. van den Berg and L. Bak-Sinnecker for transcribing the interviews, and M. Klein for answering detailed questions on OWL.

References

[1] N. Koenderink, J. Top, L. van Gool, and L. van Vliet, "Knowledge-based method for computer vision problems," *submitted in July 2004*, 2004.

[2] T. Lau and Y. Sure, "Introducing ontology-based skills management at a large insurance conpany," in *Proceedings of Modellierung 2002, Modellierung in der Praxis - Modellierung fr die Praxis*, 2002.

[3] S. Scott, J. Clayton, and E. Gibson, *A practical Guide to Knowledge Acquisition*. Addison-Wesley Publishing Company, 1991.

[4] M. van Someren, Y. Barnard, and J. Sandberg, *The Think Aloud Method: a Practical Guide to Modelling Cognitive Processes*. Academic Press, London, 1994.

[5] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. van de Velde, and B. Wielinga, *Knowledge Engineering and Management, The CommonKADS Methodology*. MIT Press, 1999.

[6] N. Noy and D. McGuinness, "Ontology development 101: A guide to creating your first ontology," Stanford University, Tech. Rep. SMI-2001-0880, 2001.

[7] M. Smith, C. Welty, and D. McGuinness, "Owl web ontology language guide – w3c recommendation 10 february 2004," 2004, http://www.w3.org/TR/owl-guide.