

## From science to practice: Bringing innovations to agronomy and forestry

Don J.M. Willems<sup>1</sup>, Nicole J.J.P. Koenderink<sup>2</sup>, Jan L. Top<sup>3</sup>

### INFO

Received

Accepted

Available on-line

Responsible Editor:

### Keywords:

dialogue, ontology, annotation,  
innovation, query and answer  
system

### ABSTRACT

The challenge of the work presented here is to make innovative research output in the agronomy and forestry domain accessible to end-users, so that it can be practically applied. We have developed an approach that consists of three key-elements: an ontology with domain knowledge, a set of documents that have been annotated and meta-annotated, and a system (ask-Valerie) that is based on a dialogue to represent the interaction between end user and system.

We show that the dialogue-metaphor is a good way of modelling the interaction between user and system. The system helps the user in formulating his question and in answering it in a useful way. Meta-annotations of key-paragraphs in the document-base turn out to be relevant in assessing in one glance what the content of a document is.

End-users are very enthusiastic about the possibilities that ask-Valerie offers them in translating scientific results to their own situation.

## 1. Introduction

There is a compelling need for agricultural and forestry research to play a significant role in the future to meet the challenges of increased demand for food and fiber. This should be balanced against the need to deliver other vital services provided by land management, such as regulating the flow of water and maintaining valuable habitats. If this role is to be fulfilled, innovative research outputs need to be made accessible to end-users so that they can be practically applied. Many EU and nationally funded research projects in the fields of agriculture and forestry provide excellent scientific results. However, outreach and interpretation of these results into 'ready-to-use' farming and forestry formats is limited. The challenge here is to boost innovation by facilitating knowledge exchange between practitioners (farmers, foresters, advisers etc.) and researchers to improve the uptake of research outputs.

The EU-funded Valerie-project (<http://www.valerie.eu>) aims to address this challenge<sup>4</sup>. Ideally, experts relate scientific findings to applications by discussing possible implementations with practitioners in the field. In these discussions, the practitioner's background, the precise problem context, possible solutions and further actions are clarified. However, the experts are not available 24/7, their time is expensive, and they do not have immediate access to results of others. This is why the digital expert-assistant comes into play. Currently, computer aided advisory systems have become accessible and cheap, can access large repositories with relevant reports and publications.

In the standard 'Google' approach to this kind of advisory systems, the user enters few search terms. The system returns a ranked list of documents and the user refines the search terms if needed. Although this is nowadays common for a generic information space as the web, it is a very poor reflection of the

<sup>1</sup> Don Willems, Wageningen UR Food & Biobased Research, the Netherlands, [don.willems@wur.nl](mailto:don.willems@wur.nl)

<sup>2</sup> Nicole J.J.P. Koenderink, Wageningen UR Food & Biobased Research, the Netherlands, [nicole.koenderink@wur.nl](mailto:nicole.koenderink@wur.nl)

<sup>3</sup> Jan L. Top, Wageningen UR Food & Biobased Research, the Netherlands and Vrije Universiteit Amsterdam, the Netherlands, [jtop@cs.vu.nl](mailto:jtop@cs.vu.nl)

<sup>4</sup> Cited from (Valerie brochure 2015)

expert-practitioner interaction one would expect in a specific domain, in our case agriculture and forestry innovation. We submit that this is a consequence of the fact that in standard solutions, background knowledge on the domain is missing in the interaction between practitioner and system. In this paper we present how a digital but knowledgeable ‘assistant-expert’ can serve as an intermediate between experts and practitioners, optimizing the effectiveness of the interaction between them. An important challenge is to use the modelled domain knowledge in such a way that the digital assistant-expert is able to have an effective dialogue with the user. In such a dialogue, the initial question of a farmer may be vague and broad, or very specific. The assistant-expert helps to translate his question into terms in which solutions are formulated, independent of a particular language, and helps the farmer to narrow down or broaden the questions, or suggest alternative lines of thinking.

The ideal expert has access to a large set of knowledge sources that he has read and understood using unambiguous terminology. He knows which parts of which documents are relevant to find particular answers. He also knows that a text that is about ‘Rusty Spot’ may be applicable even if the question was about ‘potato disease’. In short, the digital expert should be understanding, ask further questions, give useful answers, but also inspire and know its limitations.

Our assistant expert is called ask-Valerie. ask-Valerie is to assist the user in asking the right question, give answers that are informed and to-the-point and, be fast. Our approach is based on (1) setting up an ontology on innovations in agriculture and forestry to remove the ambiguities and vagueness of natural language in this domain (2) select and interpret useful digital material and (3) performing a dialogue. We submit that by combining these three elements we can progress from standard search solutions to approaching the advisory capabilities of a human advisor.

The paper is set up as follows. In Section 2, an overview of related research is given to set the stage for our contribution. Section 3 describes the method used. In Section 4, we show the results of the first development cycle. We finish with conclusions, a discussion and recommendations for future work.

In the remainder of this paper we use the role of the *practitioner* as a common denominator for ‘farmer’, ‘forestry worker’, ‘extension worker’, i.e., any member of the Valerie’s target audience, as opposed to the *expert*, who can be any knowledge provider<sup>5</sup>.

## 2. Related work

The huge amount of information on the Internet has made the task of finding the best possible available information for a specific query very challenging. From the analysis of user search logs (e.g. (Morrison 2008)), it seems that users pose short queries and do mostly not look beyond the first ten search results. Therefore, it is relevant to find as many well-fitting results as possible in the top 10 of automatically generated answers, to optimally serve the user with an answer.

Keyword-based search is not well-suited for finding the best results (Anderson 2013). The Information Retrieval (IR) community has searched for better methods. A frequently used approach is to expand the query that users enter (Abdelali et al. 2007; Bhogal et al. 2007). Query expansion helps to prevent the problem of word mismatch (Xu and Croft 1996). Three commonly used methods (Bhogal et al. 2007) for query expansion are (i) relevance feedback (Brillakis 2005), (ii) corpus dependent knowledge models (Khan and Khor 2004; Oakes and Taylor 1998) and (iii) corpus independent knowledge models. Corpus independent knowledge models usually use thesauri or ontologies to enrich the user-defined query (Dey et al. 2005; Díaz-Galiano et al. 2009; Thenmalar and Geetha 2014; Uma-Devi and Meera Ghandi 2015; Yadav 2009). These sources show that ontology-based search helps in improving the finding of relevant information. Two bottlenecks can be identified, though. First, the development of the domain ontology and its regular update to incrementally improve the search process usually takes a significant amount of time and is therefore a bottleneck in the process (Thenmalar and Geetha 2014). Moreover, usually people are not interested in the complete document, but need a

---

<sup>5</sup> Both roles can of course be played by any gender; for simplicity we use ‘he’ throughout the text.

fragment that fits their needs. Finding relevant fragments in documents is another bottleneck in the process (Yadav 2009).

The objective of ask-Valerie is similar to many Question and Answering (QandA) Systems (Voorhees, 1999), i.e. providing an answer or solution to a question that the user has. The use of semantic vocabularies to (re)formulate the question has been used previously in QandA systems, such as the systems created by Athira et al. (2013) and Guo and Zhang (2008). Semantic technology becomes especially pertinent when the answers to the users' questions are also stored in semantic databases or on the semantic web (Bordes et al. 2015).

Typically, in QandA systems, the question can be formulated by the user using natural language. Athira et al. (2013) compare QandA systems with Information Retrieval (IR) and Information Extraction (IE) systems in which a relevant subset of documents are found using IR by matching keywords provided by the user with an indexed set of words taken from the full set of documents. In this sense, ask-Valerie has more in common with IR systems than with QandA systems. But unlike IR systems, the ask-Valerie system, does aim to provide the user with an exact answer, while IR systems typically provide the user with documents that are related to the answer (Athira et al. 2013).

Core to ask-Valerie is the use of semantic annotations to identify the relevant parts of a document. Work on semantically annotating documents or information and using those annotations in Information Retrieval (IR) is not new (Oshaiba et al. 2015). A lot of research has been targeted towards automatic annotation systems, such as the User Oriented Semantic Indexer (USI, Fiorini et al. 2015) or GeneView (Thomas et al. 2012), which is one of many tools that uses automatic annotation to search PubMed.

Manual annotation in computer science is mostly used to create datasets used in the evaluation of automatic or semiautomatic annotation (e.g. Kiyavitskaya et al. 2006). An overview of manual annotation tools is given in Hinze et al. (2012). The issue with manually annotating documents is that it is very laborious for non-trivial documents sets. Semi-automatic annotation is used to pre-annotate documents to make the process of manual annotation much less laborious (Kiyavitskaya et al. 2006, Névéol et al. 2011) while still retaining the advantages of manually annotating documents over automatic annotation, i.e. it is less error-prone (see for instance; Lin et al. 2015).

### 3. Method

Since our goal is to approach a discussion between the expert and the practitioner, we will relate the different steps that our method takes to the elements of such a discussion and the preparation of this discussion. Where steps can be automated we refer to the assistant-expert as a metaphor for the supporting application.

Firstly, the expert and practitioner define a common language to express questions and answers. While the practitioner and the expert will have some common background, as they work in the same domain, differences will exist in the level of expertise and jargon used. This means that they not only have to choose a shared natural language, but also a corresponding set of words and grammar, understandable for the assistant-expert.

Secondly, the expert selects a collection of potentially relevant reports, papers, etc. We assume that the knowledge that is needed to provide answers is contained in a set of digital documents. The expert 'translates' the documents into the shared terminology and identifies relevant sections. He also makes the documents available for the system to access.

After the expert has made these preparations, the digital assistant-expert can talk to the practitioner. The assistant helps the practitioner to use the shared language for articulating his question. The system suggests alternative directions to explore, and helps the user to zoom in or out of certain details. It also translates selected terms into queries to consult the document repository. Moreover, it uses narrower terms for a given search term to find more documents. It selects fragments from relevant documents and ranks them. The user takes action with this information or reformulates his question.

In this way, ask-Valerie aims to help farmers to find the most relevant information based on the knowledge of many experts in the domains of agronomy and forestry. In the following paragraphs we describe in more detail how we extract and apply the underlying knowledge given the three steps describe above, focusing on the (technical) details of the implementation in ask-Valerie.

### 3.1. Creating a shared vocabulary

An ontology defines a set of representational primitives (concepts) and relations between these primitives (Gruber 2009). These concepts include information about their meaning in the form of the name(s) or possibly an extended textual description of the concept. Multiple alternative names (labels) can be attached to one concept, allowing for the definition of synonyms and translations (e.g., *soil* has synonyms *ground* and *earth* and has translations *aarde* in Dutch and *suelo* in Spanish). Different types of relations can be defined, such as *is-a* relations providing for the creation of hierarchies (e.g., *sandy clay* 'is-a' *soil*). Concepts are often identified by unique identifiers allowing for the disambiguation between them (e.g. *earth* as in soil and *Earth*, the planet). If a question for an information source can be expressed in these concepts, the question will be unambiguous. Any other relation between concepts can be defined, such as soil is-part-of a farm. The network of relations determines the formal semantics of the associated concepts, allowing applications to act in an informed and intelligent way.

The VALERIE domain ontology is created using the ROC<sup>+</sup> tool (Koenderink et al. 2008). It provides users who are unfamiliar with ontologies a simple tool to create the domain ontology that formalises their expert knowledge. The ROC<sup>+</sup>-method consists of five steps: (i) entering of concepts into the ontology, (ii) identifying synonyms, (iii) use existing ontologies to suggest per concept in the ontology other possibly relevant concepts, (iv) create the hierarchy (*skos:narrower/skos:broader*) between the concepts, (v) indicate *skos:related* relations between concepts.

In workshop sessions, domain experts and potential users of the ask-Valerie system are asked to provide relevant concepts in the agronomics and forestry domains, place these concepts in a 'is-a' hierarchy using ROC<sup>+</sup> and identify concepts that are related in another way. This ontology is expressed in the SKOS language (W3C 2012), which provides relations such as *skos:narrower*, and *skos:broader* for *is-a* relations, and *skos:related*. Literal properties such as *skos:prefLabel* (to specify the preferred label to be used for each language considered) and *skos:altLabel* can be used to specify alternative names for concepts, together with possible synonyms.

### 3.2. Document selection and annotation

Meanwhile, domain experts select a set of documents. The documents in this set are selected for their potential to help farmers to find the innovations relevant to their problems and questions. These documents are then automatically annotated using the domain ontology. This means that the documents are parsed and their text is extracted. The text is then scanned for occurrences of the concepts from the domain ontology (i.e., for all labels of all concepts), resulting in links from words in the text to ontological concepts. It is important to realise that these automatic annotations are on the level of individual words, or short sequences of words, and not on paragraph or higher levels.

These automatically annotated documents are then manually checked by the same domain experts as those who selected the set of documents. Manual annotation involves the checking the correctness of the automatic annotations (add annotations that were missed or delete erroneous annotations). It may also involve adding annotations at the paragraph level, e.g., specifying that a paragraph covers a specific concept.

Manual annotation is supported by the Annotative Authoring (AnnA) tool developed within the VALERIE<sup>6</sup> and COMMIT/eFoodLab<sup>7</sup> projects. It allows the expert to view existing (automatically or

---

<sup>6</sup> <http://www.valerie.eu>

<sup>7</sup> <http://commit-nl.nl/projects/e-foodlab>

manually created) annotations within a document as highlighted words, and to add and remove manual annotations. The experts can use AnnA to select a piece of text in the document and then type the relevant concept from the domain ontology in a separate text field (based on text-completion using the labels in the ontology). They can select single words but also complete or multiple paragraphs. The manual annotation is then added automatically to the selected text.

The steps outlined above (creation of the domain ontology, selection of the document set, automatic and manual annotation, and creation of the semantic index) are all pre-processing steps conducted by domain experts. These steps formalise the domain knowledge in terms of the ontology. This will allow ask-Valerie to answer the farmer's question.

### 3.3. Iterative question-and-answer

One important feature of ask-Valerie is to help the user (practitioner) to formulate his question as a formal query to the ask-Valerie knowledge base. The query should capture the meaning of the question. For the user to be able to express his question the following functionality is used.

Firstly, the search field tries to automatically complete the text entered by the user. The system uses the labels for the different concepts in the ontology. When the user starts typing 'soi', the text completion functionality proposes terms such as 'soil' and 'soil management'. The user selects one of the completions in a drop down box. Not only are terms suggested that start with the string typed by the user, but also terms in which the string appears in another position are being suggested, though at the end of the drop down list. Using the suggested terms and therefore the concepts associated with these terms, the user is assisted in using ontological concepts in his question, which facilitates the translation of the query to the language understood by ask-Valerie.

Secondly, to help the user in his query formulation even further, the system presents suggestions for alternative search directions by showing slightly different concepts. These concepts link to the originally selected concepts, either by *is-a* relations (broader or parent concepts) or by *related-to* relations. The user can use these suggestions as inspiration for modifying his query.

Finally, the system expands the search query entered by the user with concepts that are implied by the search terms. For example, if the user enters the string 'ground', the ask-Valerie systems recognises the string as a synonym of the concept *soil*. Because we use a semantic index, also documents connected to other synonyms of *soil* show up as results. It does not matter which synonym (or translation) the user uses, as long as it is known in the ontology. This prevents problems related to different languages and jargons used by the practitioner and the assistant-expert. This type of query expansion is not only limited to synonyms but it also uses narrower concepts. *Mite* is, for instance, a narrower concept of *pest insect*. The system will find documents or paragraphs containing '*mite*' even if only '*pest insect*' has been mentioned in the search query.

### 3.4. Meta-annotation

Whereas the domain specific annotation as described above enables a dialogue between the assistant-expert and the practitioner at the content level, it does not guide the problem-solving process as such. We therefore have introduced the notion of *meta-level annotation* in our approach. The idea is to have experts identify paragraphs in the documents that play a specific role in the different phases of the dialogue. As a first step to exploit the feasibility of this approach and its effect on the results found, we have defined the meta-level concepts '*innovation challenge*' and '*innovation*'. They are closely linked to the *question* asked by the practitioner respectively the *answer* given by the assistant-expert, which are basic elements of any dialogue. We ask experts to manually annotate sections in the documents with these terms. These meta-annotations then identify the most crucial parts of a document with regard to possible questions and solutions for the practitioner's problem.

We assume that by using these meta-level annotations, both the selection of fragments and the ranking of documents will be improved using this kind of expert judgment. We rank documents with domain-specific search terms inside '*innovation (challenge)*' fragments highest. Next follow those

documents that have domain terms *and* one or more ‘innovation (challenge)’ fragment (but with domain terms occurring outside of these fragments). The lowest rank is given to documents without ‘innovation (challenge)’ fragments. Otherwise, results are ranked using the occurrence of the domain terms.

Fragments that are meta-level annotated also show up differently in the result list than those that are not. When the meta-annotation is not used, a predefined number of sentences before and after the search term in the text is shown. This means that the cut-off may be rather arbitrary in the eye of the user. However, when the meta-annotations are used, the entire fragments annotated with ‘innovation’ and ‘innovation challenge’ are presented to the user. These fragments will in general be more self-explaining and complete than the machine-generated fragments.

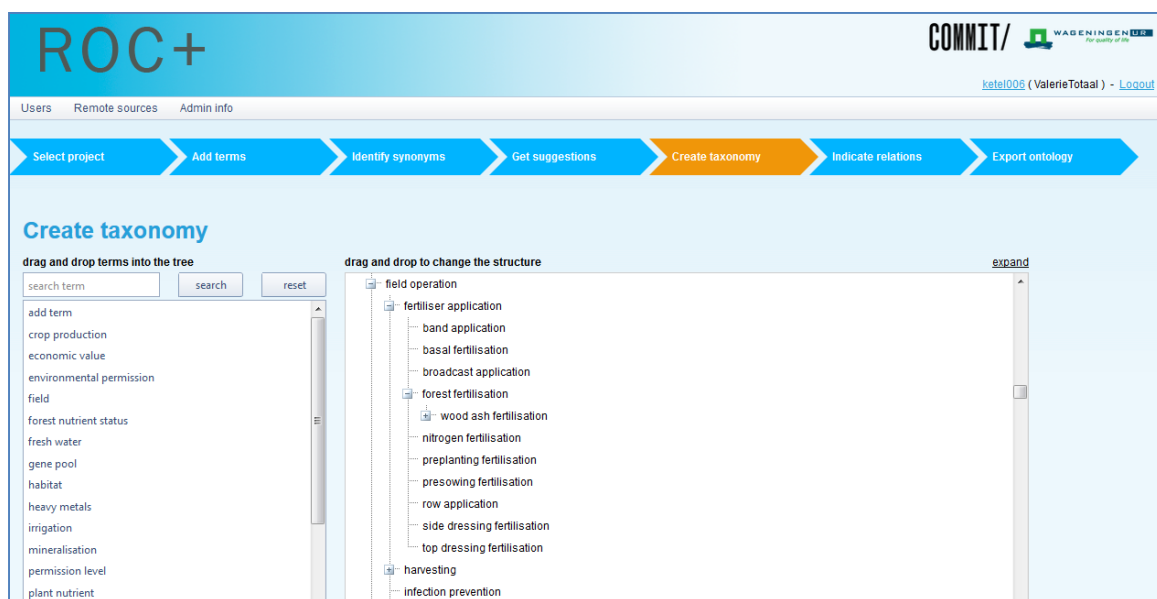
## 4. Results

### 4.1. The creation of the ontology

The Valerie-ontology was successfully created by domain experts using the ROC<sup>+</sup> ontology creation tool [3]. For the considered domain, AGROVOC (Agrovoc 2015) and Eurovoc (Eurovoc 2011) are relevant existing ontologies, so they were used as sources of inspiration in step (iii).

Sixteen teams of two experts each have systematically charted their own domain knowledge using ROC<sup>+</sup> in a workshop session of four hours (see Figure 1). In this way, expertise on the themes ‘crop rotation’, ‘integrated pest management’, ‘eco-system services’, ‘soil management’, ‘water management’, ‘sustainable integrated supply chain services’, ‘innovative farm management’ and ‘recycling of biomass and food waste’ has been covered, complemented with knowledge from ten agronomy and forestry case studies distributed over Europe (Italy, UK, Spain, Finland, Poland, the Netherlands, and France).

The result of this modelling activity is a detailed ontology consisting of 1726 terms (in April 2015), which can be found as linked open data on FoodVoc (Valerie-ontology 2015). In the ontology, four types of relations have been added: 524 synonyms (using `skos:altLabel`), hierarchical terms using `skos:narrower`, `skos:broader` (max. depth is 11 levels, 761 top level concepts exist), and 3589 other types of relations using `skos:related`. The time taken by the experts to create this ontology was approximately 1 day per person in which the ROC<sup>+</sup> workshop was given (1 hour per person) and the ontology was constructed. Additionally, two experts have spent time to improve the structure of the created ontology, each approximately two days. This amounts to a total time of approximately 150 hours.



**Figure 1.** A screenshot of the ROC<sup>+</sup> tool in which part of the Valerie ontology is shown.

## 4.2. The selection and annotation of documents

The experts who created the ontology have identified a start set of 728 PDF documents in which answers to user questions are expected to be present. These documents include guideline documents, fact sheets and scientific papers.

The first application of the Valerie-ontology is to automatically annotate this set of potentially relevant documents using the annotation tool AnnA (see Figure 2). The annotation relates individual terms and paragraphs in these documents to domain concepts. The automatic annotation process was able to annotate 710 documents in the set. The remaining 18 documents contained images, often scans of text. As the PDF parser does not support text extraction (OCR), these documents were not annotated and are not included in the search index.

In 566 documents manual annotations were found, created by the experts. Of the 1726 concepts in the domain ontology, 174 were used in manual annotations, most only once. The domain concept that was added most often was *Subsurface irrigation* with 22 annotations in 5 documents. The concept *Soil organic matter* was used with 17 annotations in 16 documents.

With respect to meta-annotation, the concept *Innovation* has been added most often (1,490 annotations in 525 files), followed by *Innovation challenge* (766 annotations in 460 documents).

The manual and meta-annotation process combined took an average of 13 minutes per document, which totals to approximately 123 hours of annotation time.

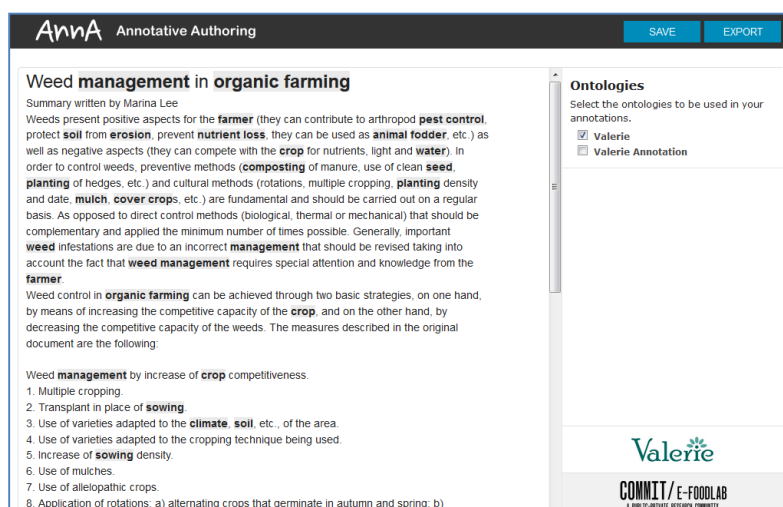
## 4.3. Ontology enrichment as result of the annotation process

The experts who annotated the 728 PDF documents noticed that terms occur in the documents that are missing from the ontology. By adding these terms to the ontology, the content of the documents will be covered better. Therefore in parallel to the manual annotation process, the experts created a list of concepts that are to be added to the ontology.

This list consists of 1095 newly identified terms. Of these, 139 terms are synonyms of existing concepts in the ontology. In total, 956 new concepts need to be added to the Valerie-ontology. The annotation process allows experts to look beyond their top-of-mind knowledge and to recognise domain terms that are relevant for the problem context.

## 4.4. Effect of meta-annotation on the quality of the search results

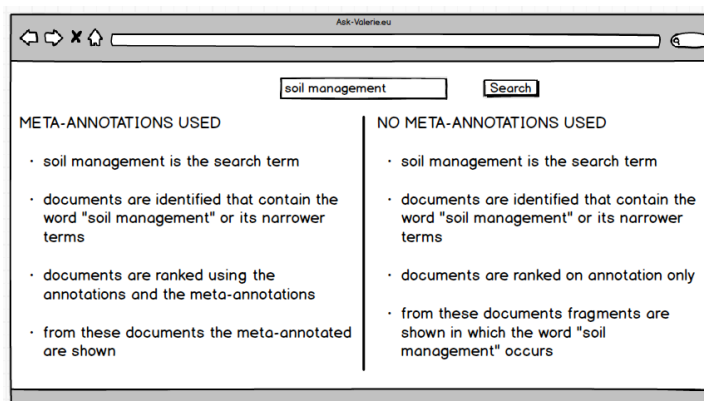
The meta-annotations of the document, where fragments of the document are indicated as “*innovation challenge*” or “*innovation*” turned out to be very relevant for improving the quality of the search results for the user.



**Figure 2:** A screenshot of AnnA, in which a document from the document base is being annotated. Highlighted terms have automatically been found.

We have setup a test portal in ask-Valerie, in which we compared search results with and without meta-annotations. In Figure 3, a schematic overview of this test portal has been displayed. The meta-annotations were used for ranking and for presenting the results.

In a feedback session with 20 users (divided in couples of 2 people), the ranking and presentation of the documents with meta-annotations was unanimously preferred over the presentation without meta-annotation.



**Figure 3:** A schematic overview of the test portal. On the left-hand side results are ranked and presented using information from the meta-annotations. On the right-hand side, no meta-annotations are used.

## 5. Conclusion and Discussion

The ask-Valerie portal provides the functionality to easily search the set of documents selected by domain experts to be relevant for the domain. Three key elements have been mentioned in this paper: (i) the creation of the Valerie-ontology, (ii) the selection and annotation of the documents in the document-base and (iii) the search functionality formed after the dialogue metaphor. Our work results in the following six conclusions:

- **Conclusion 1:** A dialogue is a useful metaphor for modelling the interaction between practitioner and assistant-expert. The assistant-expert helps the practitioner in formulating the query. It helps answering the question in interaction with the expert.
- **Conclusion 2:** ROC<sup>+</sup> is a good method to allow experts to build an ontology by themselves catered to a specific domain and task.
- **Conclusion 3:** The automatic annotation of documents with concepts from an ontology is an important step in both getting better search results than with text-based search, but also in identifying and filling knowledge gaps in the ontology.
- **Conclusion 4:** Meta-annotation is a good way to present better results to the user.
- **Conclusion 5:** The automatic annotation process performed by AnnA is a good method to perform domain annotation; manual annotation to correct the automatic annotation is superfluous.
- **Conclusion 6:** The meta-annotation process is a process where the specific capability of experts to comprehend texts comes into play. By restricting the meta-annotation-ontology to few terms, the process of meta-annotation is feasible, although in future work we will look for options of supporting this process automatically.

In the next paragraphs, we discuss the observations and future areas of research for the three mentioned key-elements.

The first pillar of our approach is the domain ontology. The creation of the ontology has been described in this paper. The created ontology, though, is not a static entity. To stay relevant for answering practitioners' questions, it will continuously be expanded and enriched. To make this process as user-friendly as possible, we will allow users to suggest new ontological terms at any time using ask-Valerie. These terms will in the future be processed by a 'knowledge gardener' who will add the terms in the appropriate place in the ontology. Moreover, AnnA will be equipped with the possibility for experts to suggest and add new terms to the ontology as well. In the near future, we will encapsulate the 956 concepts already found in the annotation process in the ontology and publish a new version on <http://www.foodvoc.org>.



The second pillar of the approach is the selection of documents for the document-base followed by the automatic annotation and the manual annotation and meta-annotation. Up to now, the selection of documents has been a manual process. With this initial set of relevant documents for the domain and the Valerie-ontology, we will for the future develop a method to automatically find documents that have a high probability of being relevant as well.

The manual annotation process consists at present of two tasks: (i) the correction of the automated annotation and (ii) the task-specific meta-annotation where key paragraphs are indicated. The results of the first task show that it hardly resulted in changes to the documents. This indicates that the automated annotation process works quite well. In the meta-annotation process, experts tend to select complex fragments -- one sentence or even a paragraph of text. This turned out to give very relevant information for the ranking and presenting of results. For future annotations, we therefore choose to skip the first task and only have the meta-annotation process performed by the experts. Ideally, even this meta-annotation task would be conducted automatically using machine learning techniques, but it is difficult to automatically identify fragments about a complex concept like 'innovation' in texts.

The third key-element in our approach is the dialogue metaphor. Future work lies in allowing practitioners in posing questions in natural language, and in using the meta-annotated text fragments in better understanding the needs of the practitioner. An overview of 'innovation challenges' in documents with similar search terms as the query, may help in formulating the query even more precise. We will make this part of our future research.

We conclude this paper with the remark that the work that we have presented here is a promising step in making innovative research outputs accessible to end-users, so that they can be practically applied. Feedback from those end-users clearly showed that answers to previously unanswered questions are already being found. However, we have only taken the first steps in creating an interactive and helpful assistant-expert. Our future work consists in refining the ontology, expanding the (meta-)annotated document base and adding functionality to ask-Valerie. Moreover, for the long-term existence of ask-Valerie, we will need to work on further minimising the input of the expert's time while at the same time keeping him involved in providing his essential domain knowledge.

## Acknowledgement

This publication was supported by the Dutch national program COMMIT and by the EU-funded project Valerie. The authors would like to thank the experts and case study owners of the Valerie project team for their input.

## 6. References

- Valerie brochure, 2015, Valorising European Research for Innovation in Agriculture and Forestry. Available from <http://www.valerie.eu> [April 6, 2015].
- Gruber, T., 2009, Ontology, In: Encyclopedia of Database Systems, Springer US, pp. 1963-1965.
- Koenderink, N.J.J.P., Assem, M. van, Hulzebos, J.L., Broekstra, J. and Top, J.L. (2008), 'ROC: A Method for Proto-Ontology Construction by Domain Experts', LNCS 5367 (2008), pp. 152 - 166. doi: [10.1007/978-3-540-89704-0\\_11](https://doi.org/10.1007/978-3-540-89704-0_11)
- Agrovoc, Agrovoc-ontology. Available from: <http://aims.fao.org/access-agrovoc>. [April 6, 2015].
- Eurovoc, Eurovoc-ontology. Available from: <http://eurovoc.europa.eu/>. [14 December 2011].
- Valerie-ontology. Valerie-ontology Available from: <http://www.foodvoc.org/page/Valerie/>. [April 6, 2015].
- W3C, 2012, SKOS Simple Knowledge Organization System. Available from <http://www.w3.org/2004/02/skos/>, [May 29, 2015].
- Morrison, P.J., 2008, 'Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web', Information Processing and Management, vol. 44, pp. 1562-1579. doi: [10.1016/j.ipm.2007.12.010](https://doi.org/10.1016/j.ipm.2007.12.010)

- Anderson, N., 2013, 'Context-aware web search using dynamically weighted information fusion', *Concurrency Computation: Practice and Experience*, vol. 25, pp. 672–685. doi: [10.1002/cpe.1805](https://doi.org/10.1002/cpe.1805)
- Abdelali, A., Cowie, J. and Soliman, H.S., 2007, 'Improving query precision using semantic expansion', *Information Processing and Management*, vol. 43, pp. 705–716. doi: [10.1016/j.ipm.2006.06.007](https://doi.org/10.1016/j.ipm.2006.06.007)
- Bhagal, J., Macfarlane, A. and Smith, P., 2007, 'A review of ontology based query expansion', *Information Processing and Management*, vol. 43, pp. 866–886. doi: [10.1016/j.ipm.2006.09.003](https://doi.org/10.1016/j.ipm.2006.09.003)
- Xu, J. and Croft, W.B., 1996, 'Query Expansion Using Local and Global Document Analysis', *SIGIR '96 Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 4–11. doi: [10.1145/243199.243202](https://doi.org/10.1145/243199.243202)
- Khan, M.S. and Khor, S., 2004, 'Enhanced Web Document Retrieval Using Automatic Query Expansion', *Journal of the American Society for Information Science and Technology*, vol. 55, no. 1, pp. 29–40. doi: [10.1002/asi.10341](https://doi.org/10.1002/asi.10341)
- Oakes, M.P. and Taylor, M.J., 1998, 'Automated Assistance in the Formulation of Search Statements for Bibliographic Databases', *Information Processing and Management*, vol. 34, no. 6, pp. 645–668. doi: [10.1016/s0306-4573\(98\)00029-6](https://doi.org/10.1016/s0306-4573(98)00029-6)
- Dey, L., Singh, S., Rai, R. and Gupta, S., 2005, 'Ontology Aided Query Expansion for Retrieving Relevant Texts', *AWIC 2005, LNAI 3528*, pp. 126–132. doi: [10.1007/11495772\\_21](https://doi.org/10.1007/11495772_21)
- Díaz-Galiano, M.C., Martín-Valdivia, M.T. and Ureña-López, L.A., 2009, 'Query expansion with a medical ontology to improve a multi modal information retrieval system', *Computers in Biology and Medicine*, 39, pp. 396–403. doi: [10.1016/j.combiomed.2009.01.012](https://doi.org/10.1016/j.combiomed.2009.01.012)
- Thenmalar, S. and Geetha, T.V., 2014, 'Enhanced ontology-based indexing and searching', *Aslib Journal of Information Management*, vol. 66, no. 6, pp. 678–696. doi: [10.1108/ajim-08-2014-0098](https://doi.org/10.1108/ajim-08-2014-0098)
- Uma Devi, M., and Meera Ghandi, G., 2015, 'Wordnet and Ontology Based Query Expansion for Semantic Information Retrieval in Sports Domain', *Journal of Computer Science*, vol. 11, no. 2, pp. 361–371. doi: [10.3844/jcssp.2015.361.371](https://doi.org/10.3844/jcssp.2015.361.371)
- Yadav, S.B., 2009, 'A conceptual model for user-centered quality information retrieval on the WorldWide Web', *Journal of Intelligent Information Systems*, vol. 35, pp. 91–121. doi: [10.1007/s10844-009-0090-y](https://doi.org/10.1007/s10844-009-0090-y)
- Voorhees, E. M., 1999, 'The TREC-8 Question Answering Track Report', National Institute of Standards and Technology (NIST).
- Athira P, M., Sreeja, M., and Reghuraj, P. C., 2013, 'Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System', *International Journal of Web and Semantic Technology (IJWest)*, vol.4-4, pp 31-39. doi: [10.5121/ijwest.2013.4403](https://doi.org/10.5121/ijwest.2013.4403)
- Guo, Q. and Zhang, M., 2008, 'Question Answering System Based on Ontology and Semantic Web', *RSKT* 652–659. doi: [10.1007/978-3-540-79721-0\\_87](https://doi.org/10.1007/978-3-540-79721-0_87)
- Bordes, A., Usunier, N., Chopra, S. and Weston, J., 2015, Large-scale Simple Question Answering with Memory Networks. [arXiv.org](https://arxiv.org/abs/1506.04838) **1506**, 2075.
- Oshaiba, M. M. Z., Houbay, El, E. M. F. and Salah, 2015, 'A. Semantic annotation for biological information retrieval system.', *Advances in Bioinformatics*, Vol. 2015. doi: [10.1155/2015/597170](https://doi.org/10.1155/2015/597170)
- Fiorini, N., Ranwez, S., Montmain, J. and Ranwez, V., 2015, 'USI: a fast and accurate approach for conceptual document annotation.', *BMC Bioinformatics* Vol. 16, 83. doi: [10.1186/s12859-015-0513-4](https://doi.org/10.1186/s12859-015-0513-4)
- Thomas, P., Starlinger, J., Vowinkel, A., Arzt, S. and Leser, U., 2012, 'GeneView: a comprehensive semantic search engine for PubMed.', *Nucleic Acids Research*, Vol. 40, W585–W591. doi: [10.1093/nar/gks563](https://doi.org/10.1093/nar/gks563)
- Kiyavitskaya, N., Zeni, N., Mich, L., Cordy, J. R. and Mylopoulos, J., 2006, 'Text Mining Through Semi Automatic Semantic Annotation', chapter 13 in Reimer, U. and Karagiannis, D. (eds), 'Practical Aspects of Knowledge Management', pp. 143–154, Springer Berlin Heidelberg. doi: [10.1007/11944935\\_13](https://doi.org/10.1007/11944935_13)
- Hinze, A., Heese, R., Luczak-Rösch, M. and Paschke, 2012, 'A. Semantic Enrichment by Non-experts: Usability of Manual Annotation Tools.' In: *International Semantic Web Conference*, pp. 165–181, doi: [10.1007/978-3-642-35176-1\\_11](https://doi.org/10.1007/978-3-642-35176-1_11)

Lin, C.-H., Wu, N.-Y., Lai, W.-S. and Liou, D.-M., 2015, 'Comparison of a semi-automatic annotation tool and a natural language processing application for the generation of clinical statement entries', Journal of the American Medical Informatics Association : JAMIA, Vol. 22, pp. 132–142. doi: [10.1136/amiajnl-2014-002991](https://doi.org/10.1136/amiajnl-2014-002991)

Névéol, A., Dogan, R. I. and Lu, Z., 2011, 'Semi-automatic semantic annotation of PubMed queries: A study on quality, efficiency, satisfaction.', Journal of Biomedical Informatics, Vol. 44, pp. 310–318. doi: [10.1016/j.jbi.2010.11.001](https://doi.org/10.1016/j.jbi.2010.11.001)