Representing Situational Knowledge for Disease Outbreaks in Agriculture

Markus Stocker[1], Jussi Nikander[2], Hanna Huitu[3], Marja Jalli[4], Markku Koistinen[5], Mauno Rönkkö[6], Mikko Kolehmainen[7]

I N F O

A B S T R A C T

We present a software system for automated projection of situational knowledge for disease outbreak in agriculture. The system supports farmers and agricultural advisers in obtaining and maintaining awareness of present and future disease outbreaks in crops grown at agricultural parcels. It models objects such as plant pathogens and agricultural parcel crops, and their relations, as entities in situations observed by an environmental monitoring system. It utilizes a mechanistic disease pressure model to obtain knowledge about observed situations from forecast data for various weather parameters. It represents obtained situational knowledge explicitly and manages represented knowledge in a knowledge base. We evaluate the system for 3 fungal plant pathogens, 2 cereal crops, and 17 agricultural parcels located in Finland, for a growing season. We underscore how the explicit representation of situational knowledge is useful toward various purposes, including reasoning, query and visualization, and is, thus, vastly superior to having situational knowledge only implicit in high-level data products such as maps.

1. Introduction

Farmers have arguably been relying on timely information about factors that influence crop growth and quality since the dawn of agriculture, and modern Precision Agriculture is "intrinsically information intensive" (Fountas et al., 2006). Plant disease is a factor that influences crop growth and quality, and the continuous assessment of plant disease pressure is important as it supports farmers in their decision for when and how to protect crops (Roberts et al., 2006). Ideally, the assessment should forecast disease pressure (Roberts et al., 2006) and include necessary information. The requirement of forecast is important because it gives farmers time to plan and take action, and reduce losses (Bourke, 1970).

During a growing season, disease pressure can rapidly increase and various models have been proposed to describe disease progress (van Maanen and Xu, 2003). It is therefore necessary to assess disease pressure regularly. Already with relatively few agricultural parcels it becomes quickly evident that disease pressure assessment is laborious. In fact, agricultural parcels are often spatially scattered over large areas. Visiting them requires time and material resources. Both visual and laboratory assessments are time consuming and costly. Lacking direct observation, disease pressure assessment is overly dependent on educated guesses by farmers. Automated assessment using computer systems may address this problem. Indeed, computer systems have served farmers and agricultural advisers for many decades and purposes, including as decision support systems for plant protection. Bouma (2007) recently reviewed the historical development of computer-based support for plant protection since the 1980s in various European countries.

[1] Markus Stocker, markus.stocker@uef.fi, University of Eastern Finland
[2] Jussi Nikander, jussi.nikander@luke.fi, Natural Resources Institute Finland
[3] Hanna Huitu, hanna.huitu@luke.fi, Natural Resources Institute Finland
[4] Marja Jalli, marja.jalli@luke.fi, Natural Resources Institute Finland
[5] Markku Koistinen, markku.koistinen@luke.fi, Natural Resources Institute Finland
[6] Mauno Rönkkö, mauno.ronkko@uef.fi, University of Eastern Finland
[7] Mikko Kolehmainen, mikko.kolehmainen@uef.fi, University of Eastern Finland

Disease outbreak caused by a plant pathogen in a crop of an agricultural parcel at a particular point in time can be modelled as a situation, following the formalization of the concept proposed in Situation Theory (Barwise and Perry, 1981; Devlin, 1991). The situation involves five objects, namely the pathogen, crop, agricultural parcel, spatial location, and temporal location. The outbreak relates the pathogen and the crop in the situation. Disease outbreak assessment can, thus, be understood as a particular type of situation assessment, which is the process used to achieve situation awareness. Endsley (1995) defined situation awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." Situation awareness is the state of knowledge held, for instance, by farmers.

Our objective is to discuss the implementation – as well as the execution, for a growing season – of a software system that performs automated daily situation assessment to support the maintenance of situation awareness for disease outbreak in agricultural crops. The result is a plant disease predictive system (Gent et al., 2013), in particular a weather-related situation awareness system that automatically predicts – or, in Endsley's terms, projects – situations for disease outbreak. Situation assessment is performed daily during a growing season and encompasses 3 fungal plant pathogens, 2 cereal crops, and 17 agricultural parcels, located in Finland.

We implement the system by extending a software framework for situation awareness in environmental monitoring with knowledge and program logic required for the domain and problem of interest here. Specifically, extensions include (1) knowledge about the agricultural parcels, such as crop susceptibility to disease and tillage method, and (2) program logic necessary to obtain forecast data for various weather parameters as well as to compute accumulated risk from forecast data using an implementation for a disease pressure model. Situational knowledge for (acute) outbreak of pathogens in crops is obtained from accumulated risk, is represented explicitly in machine readable form, and is persisted in a knowledge base. The knowledge base serves as a repository for situational knowledge. We present how such knowledge can be visualized in time and space to provide experts, such as farmers and agricultural advisers, with an intuitive overview of (projected) outbreak situations at agricultural parcels and, thus, contribute to expert situation awareness.

The contribution of this work is three fold. First, we show that disease outbreak in agriculture can be modelled as situations grounded in formal Situation Theory (Barwise and Perry, 1981; Devlin, 1991). Second, we show that forecasting disease outbreak from weather forecast data can be understood as a situation assessment process. Hence, we ground the problem of disease outbreak forecast in situation awareness theory (Endsley, 1995). Third, we demonstrate the application of a generic software framework for situation awareness in environmental monitoring to a software system that performs automated daily situation assessment to support the maintenance of situation awareness for disease outbreak in agricultural crops.

The most interesting feature of the presented system is arguably the explicit and machine readable representation of situational knowledge. Information is not implicit in maps or reports designed for experts. Rather, information is explicit, readable and interpretable by computer systems. The system thus aims at automated interpretation of data using terminology and semantics shared with experts. Automating this process is interesting in particular because the amount of data, in particular environmental monitoring data, can be overwhelming (Wang et al., 2006; Gaire et al., 2013).

2. Materials and methods

We first describe the materials and then the methods relevant to this study. Materials consist primarily of software, data, and a disease pressure model. Methods are for data processing, knowledge extraction, and knowledge representation.

2.1 Materials

Situation assessment was implemented using the Wavellite software framework for situation awareness in environmental monitoring. Wavellite supports data acquisition and processing, the representation of situational knowledge extracted from processed data, and situational knowledge processing. Situations are structured parts of reality (Devlin, 1991) and in environmental monitoring

Markus Stocker, Jussi Nikander, Hanna Huitu, Marja Jalli, Markku Koistinen, Mauno Rönkkö, Mikko Kolehmainen:
Representing Situational Knowledge for Disease Outbreaks in Agriculture

(Wiersma, 2004) they are 'observed' by an environmental monitoring system consisting of hardware, software, and experts, as well as, often, an environmental sensor network (Martinez et al., 2004; Hart and Martinez, 2006). Generally, environmental sensor networks do not observe situations directly. Rather, they monitor certain physical properties of environmental phenomena that occur in observed parts of reality.

Wavellite supports two abstractions for data: sensor observation and dataset observation. Sensor observation is the abstraction for sensor data while dataset observation is the abstraction for processed data. Situation is the abstraction for knowledge. Situations are grounded in Situation Theory (Barwise and Perry, 1981; Devlin, 1991). Following the notation by Devlin (1991), situations are formalized by means of the expression $s \vDash \sigma$ (read s supports σ) meaning that the infon σ is "made factual" by the situation s. The object ‹‹R, a_1, …, a_m, i›› is a well-defined infon if R is an n-place relation and a_1, …, a_m (m ≤ n) are objects appropriate for the argument places i_1, …, i_m of R, and if the filling of argument places i_1, …, i_m is sufficient to satisfy the minimality conditions for R, and i = 0,1 is the polarity. Minimality conditions "determine which particular groups of argument roles need to be filled in order to produce an infon" (Devlin, 1991). The polarity is the 'truth value' of the infon. If i = 1 then the objects a_1, …, a_m stand in the relation R; else the objects do not stand in the relation R. Parameters, denoted as ȧ, make reference to arbitrary objects of a given type. For instance, İ and ṫ typically denote parameters for arbitrary objects of type spatial location and temporal location, respectively. Anchors are a mechanism to assign values to parameters. The parameter ṫ may anchor the value for the current time.

Sensor observation, dataset observation, and situation are concepts defined in upper ontologies. Sensor observation is aligned with the concept SSN Observation of the Semantic Sensor Network (SSN) ontology (Compton et al., 2012); dataset observation is aligned with the concept QB Observation of the RDF Data Cube Vocabulary (QB) (Cyganiak et al., 2014); situation is defined in the Situation Theory Ontology (STO) (Kokar et al., 2009), as STO Situation. Sensor observations, dataset observations, and situations are persisted and managed by a knowledge base.

The study was for 17 agricultural parcels, which are part of the Vakola precision agriculture research station located in Vihti, Finland, operated by the Natural Resources Institute Finland (LUKE). For each agricultural parcel, we required its preceding crop (wheat or barley), current crop (wheat or barley), current crop susceptibility (sensitive, normal, or resistant), tillage method (reduced or intensive), and seeding date. This data was for a growing season. In addition, we had the polygon data for the spatial extent of each agricultural parcel. This data was available encoded as ArcGIS shapefile. Agricultural parcel data was made available by LUKE.

We used 48-hours hourly forecast data for the four properties temperature [C], wind speed [m/s], relative humidity [%], and precipitation amount [mm] and for six grid points that delimited a geographical area that spatially contained the 17 agricultural parcels. Forecast data was computed by the Finnish Meteorological Institute (FMI) and was made available via the FMI Open Data Web service interface. NetCDF (Rew and Davis, 1990) was the encoding of the forecast data obtained by FMI. We used forecast data in order to run the presented system ahead of current time so that it predicts disease pressure, i.e. projects situations.

We employed a disease pressure model developed by LUKE and Rural Advisory Centre Nyland Finland (NSL) (Nikander et al., 2015). It is a mechanistic model in which the state of disease pressure is updated once per day. Disease pressure stands for the progress of a disease infection in the field and tells the probability for the need of additional plant protection measures. Disease pressure is computed as the cumulative value $AR_t = AR_{t-1} + DR$, where AR_t is the accumulated disease pressure value on day t, and DR is the change on a given day. The magnitude of AR tells the current probability of a disease outbreak in a field. Typically, AR is a monotonic never decreasing variable. However, plant protection measures, such as chemical spraying, can decrease the value and prevent the progress of disease for a defined amount of time. DR is constructed from a base risk value (BR) modified by daily modifiers (DM), namely DR = BR x DM. The base risk BR depends on the susceptibility of the selected crop and farming history. The daily modifiers DM are related to the weather for the given day, specifically average temperature, humidity, wind, and the amount of rainfall. All diseases

Markus Stocker, Jussi Nikander, Hanna Huitu, Marja Jalli, Markku Koistinen, Mauno Rönkkö, Mikko Kolehmainen: Representing Situational Knowledge for Disease Outbreaks in Agriculture

included in the model follow this general disease pressure model. How the base risk or the daily modifiers are used, however, depends on the disease, as all diseases react to changes in the environmental variables in a unique manner.

Accumulated disease pressure is computed for the pair (c_{ap}, p) consisting of the crop c_{ap} of an agricultural parcel ap and a pathogen p. The fungal pathogens currently supported by the model are *Pyrenophora teres*, *Pyrenophora tritici-repentis*, and *Stagonospora nodorum. Pyrenophora teres* affects barley while the other two pathogens affect wheat. Thus, valid pairs (c_{ap}, p) are those in which the pathogen p affects the crop c_{ap}. The disease pressure model served as mechanistic (physically-based) model in the computation of accumulated risk. Situations were for outbreak and acute outbreak of pathogens in the crop of an agricultural parcel. The disease pressure model was available as textual description in Finnish language.

All software was developed using the Java programming language. The knowledge base was implemented by the Stardog RDF database. The Resource Description Framework (RDF) (Manola et al., 2004) is a data model. RDF is used to represent sensor observations, dataset observations, and situations. RDF Schema (RDFS) (Brickley and Guha, 2004) is a semantic extension of RDF (Hayes and McBride, 2004) and provides the basic constructs to build ontologies. Ontology is a means to represent knowledge of a domain, i.e. the concepts of interest and relations that hold among them as well as concept and role assertions. Among other features, RDFS supports the building of subclass hierarchies. The Web Ontology Language (OWL 2) (Hitzler et al., 2012) builds on RDF and RDFS and introduces further constructs that allow for the building of ontologies with richer semantics. RDFS and OWL ontologies are machine readable and interpretable. The Simple Protocol and RDF Query Language (SPARQL) (Prud'hommeaux and Seaborne, 2008) is a query language for RDF, and together with RDFS and OWL they are technologies of the Semantic Web (Berners-Lee et al., 2001).

2.2 Methods

We created a domain ontology to define the class AgriculturalParcel as sub class of the class Feature as defined by GeoSPARQL (Perry and Herring, 2012). Each agricultural parcel considered in this study was an individual instance of the class AgriculturalParcel. As described in Section 2.1, for each agricultural parcel we had data for the growing season in textual form and ArcGIS spatial data. This data was converted into RDF. In addition to agricultural parcel data for the growing season, we also processed the ArcGIS shapefile to extract the polygons that located agricultural parcels.

The textual description of the disease pressure model needed to be first translated into software. In the presented system, this software served as an independent library and was responsible for computing the daily accumulated risk for crop and pathogen pairs, given the model input parameters for pathogen, preceding crop, current crop susceptibility, tillage, daily mean temperature, daily mean precipitation amount, daily mean wind speed, daily duration leaves are wet [hours], daily flag for whether or not leaves are wet [true/false], and the current total accumulated risk. The library was created for the particular purpose of this study.

Each day between April 30 and August 27 we collected NetCDF encoded forecast data via the FMI Open Data Web service interface. The forecast data was processed to a daily dataset consisting of dataset observations with component properties for temporal location, spatial location, and the forecast data for the four properties. The forecast data collection and processing was automated. The data collection process was responsible to fetch the NetCDF encoded forecast data and persist the file on local disk. The data processing process was responsible to read the NetCDF file from disk and parse the data to dataset observations, elements of the corresponding daily dataset for forecast data.

On April 30, the current total accumulated risk was zero for all valid pairs (c_{ap}, p) of agricultural parcel crop c_{ap} and pathogen p. We determined April 30 to be the first day following 5 consecutive days with temperature > 5°C and, thus, the start of the growing season in Vihti, Finland. Starting April 30, for each day and applicable pair (c_{ap}, p), the system computed the total accumulated risk for the next day using the forecast data, the agricultural parcel data, and current total accumulated risk for (c_{ap}, p). A pair (c_{ap}, p) was applicable if the current date was greater than the seeding date of the crop c_{ap} and c_{ap} was not protected (e.g. by a chemical agent). The first day for which total accumulated risk

was computed was, thus, May 1. The result of such daily computation were dataset observations consisting of temporal location, spatial location, pathogen, and total accumulated risk. Such dataset observations were elements of the disease pressure dataset. This computation made use of the library that implemented the disease pressure model.

Situational knowledge was extracted from disease pressure dataset observations. Knowledge was for situations of outbreak or situations of acute outbreak. Using the notation by Devlin (1991), a situation s of outbreak $s \vDash$ <<outbreak, $\dot{p}$, $\dot{t}$, l, 1>> supports an outbreak relation infon with parameters for pathogen, $\dot{p}$, temporal location, $\dot{t}$, and spatial location, $\dot{l}$, which refer to the arbitrary objects that stand in the outbreak relation. An outbreak situation is represented for each disease pressure dataset observation with total accumulated risk r such that $50 < r \leq 75$. A situation s of acute outbreak $s \vDash$ <<acute-outbreak, $\dot{p}$, $\dot{t}$, l, 1>> supports an acute-outbreak relation infon with parameters $\dot{p}$, $\dot{t}$, and $\dot{l}$, which refer to the arbitrary objects that stand in the acute-outbreak relation. An acute outbreak situation is represented for each disease pressure dataset observation with total accumulated risk r such that $r > 75$.

In addition to outbreak and acute outbreak situations, we also model chemical plant protection situations. Plant protection is the result of applying a chemical agent, specifically fungicides, to the crop, with the aim of protecting the crop from infection. Aside killing pathogens, for a certain time period chemical agents also prevent pathogens from developing. Experts determined this time period to be two weeks.

To represent situations lasting for periods longer than the next projected day, e.g. a one-week acute outbreak, we implemented temporal reasoning for situations covering 11 of the 13 relations of Allen's interval algebra (Allen, 1983). Temporal reasoning inferred continuous periods of unchanged situations. Temporal reasoning on situations was implemented as a post-processing step on situational knowledge.

3. Results and discussion

For the growing season, the system represented 112 situations, considering plant protection situations and temporal reasoning. For agricultural parcels at which wheat was grown the system represented outbreaks and acute outbreaks for both *Stagonospora nodorum* and *Pyrenophora tritici-repentis*. In contrast, for agricultural parcels at which barley was grown the system represented outbreaks and acute outbreaks for *Pyrenophora teres*. This is expected. The greatest number of situations was 11 at the only agricultural parcel that was sprayed with fungicides twice, the first time in the second half of May shortly after the system had projected *Stagonospora nodorum* outbreak. All other outbreak situations turned into acute outbreak situations before agricultural parcels were sprayed with fungicides. The least number of situations was 1.

Notable is the considerable difference in number of represented situations involving *Stagonospora nodorum* or *Pyrenophora tritici-repentis* and situations involving *Pyrenophora teres*. This difference is explained by the current crop and its susceptibility grown at agricultural parcels. In fact, the current crop at agricultural parcels with 9 or more situations was wheat and its susceptibility was either normal or sensitive. In contrast, the current crop at agricultural parcels with one or two situations was barley and its susceptibility was resistant. The exception is an agricultural parcel growing wheat of susceptibility normal with a total of 4 situations. The parcel was never sprayed with fungicides; the few situations started hence in the second half of May and lasted throughout the growing season.

With current crop barley and not having been sprayed with fungicides, two agricultural parcels developed outbreaks for *Pyrenophora teres*. However, such situations occurred during the first half of July, which is almost two months after the first onset of outbreaks at agricultural parcels with wheat as current crop. Three agricultural parcels growing resistant barley were sprayed with fungicides during the second half of June and no situations developed at these parcels as the remainder of the season (July and August) was not sufficiently long for development of disease outbreaks in resistant barely, according to our system.

This suggests that resistant barley is less prone to outbreaks than sensitive wheat, which is expected by definition of resistant. The system represented 4 outbreaks of *Pyrenophora teres* (which affects barley) and 93 outbreaks for *Stagonospora nodorum* or *Pyrenophora tritici-repentis* (which affect wheat). *Stagonospora nodorum* outbreaks occurred generally before *Pyrenophora tritici-repentis* and in one instance the agricultural parcel was protected before the *Stagonospora nodorum* outbreak turned into an acute outbreak.

The mean and standard deviation duration of outbreaks is 4.6 ± 4.5 days while acute outbreaks last on average 33.7 ± 22.2 days. Outbreaks relatively quickly turn into acute outbreaks. Given that agricultural parcels have typically been sprayed with fungicides only once, the duration of acute outbreaks is relatively long. With 23 ± 4 days mean duration, *Pyrenophora teres* outbreaks lasted considerably longer than *Stagonospora nodorum* (4.7 ± 2.9) and *Pyrenophora tritici-repentis* (3.4 ± 1.3) outbreaks. Arguably, this can be ascribed to different crops and crop susceptibility.

Figure 1 illustrates the temporal visualization of situations. We show the time line of one agricultural parcel. Outbreak, acute outbreak, and plant protection situations extend over time intervals. The growing season started in May and, according to the system, at the agricultural parcel the first outbreaks started around mid-May, *Stagonospora nodorum* first and *Pyrenophora tritici-repentis* later. The first outbreak of *Stagonospora nodorum* lasted four times longer than the first outbreak of *Pyrenophora tritici-repentis*. On June 7, the parcel was plant protected and the protection lasted for two weeks (expert assessment). After the protection came to an end, it took about 10 days for the system to model new situations. Notably, the second *Stagonospora nodorum* outbreak was four times shorter than the first. The vertical line represents current time. As we can see, system modelling of situations is one day ahead of current time.
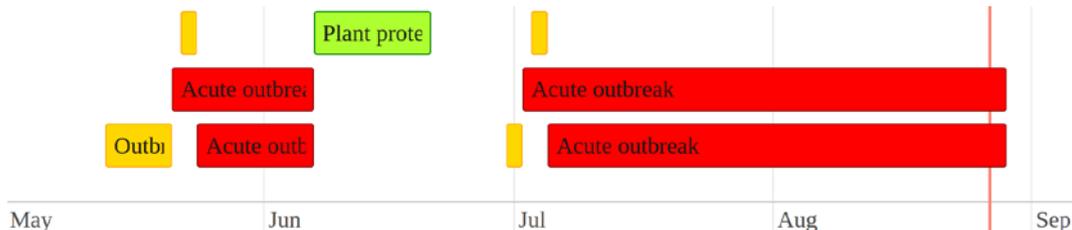
Figure 1. Temporal visualization of situations for an agricultural parcel. Situations of relatively short duration (orange) are for outbreaks while situations of longer duration (red) are for acute outbreaks. The additional situation (green) is for plant protection.

We have described the application of an open source software framework, called Wavellite, for situation awareness in environmental monitoring to the problem of disease outbreak forecast in agriculture. The framework supports the modelling of disease outbreak as situations grounded in formal Situation Theory (Devlin, 1991) and the modelling of disease outbreak assessment as situation assessment (Endsley, 1995). The state of knowledge, achieved and maintained by the system and shared with experts (Stanton et al., 2006, 2010), for the projected situations of outbreak and acute outbreak of pathogens in agricultural parcel crops in time is situation awareness (Endsley, 1995).

The software system presented here evolved from an earlier prototype for the computation and visualization of disease pressure in agriculture (Thessler et al., 2011). Earlier, the computation was on weather observation data by the SoilWeather WSN (Kotamäki et al., 2009) and the result of computation was a static color-coded GIS map with coarse information for disease pressure in the region. There are significant differences between the earlier prototype and the system presented here. First, the system presented here executes the disease pressure model on (FMI) forecast weather data while the earlier prototype was based on (SoilWeather WSN) observation weather data. Second, visualizations presented here (e.g. Figure 1) are interactive, meaning that temporal and spatial features can be selected to obtain information for the situations that occurred at particular locations. Third, the granularity of information is at agricultural parcels, rather than larger regions, typically including many agricultural parcels. Fourth, there are major technological differences between the two systems. Most importantly, the system presented here builds on Semantic Web technologies and, thus, holds

Markus Stocker, Jussi Nikander, Hanna Huitu, Marja Jalli, Markku Koistinen, Mauno Rönkkö, Mikko Kolehmainen: Representing Situational Knowledge for Disease Outbreaks in Agriculture

semantically rich representations of data and metadata. Fifth, and perhaps most importantly, here the information is explicit. The software system 'knows' about agricultural parcels, crops, pathogens, and situations. Because information is represented explicitly, in symbolic form, software can manipulate it to produce many more results beyond a map. For instance, we can employ inference techniques to obtain new information, compute summary statistics for situations, provide access to situational knowledge to other systems, or integrate situational knowledge provided by other systems to enrich shared expert-machine situation awareness. In contrast, in an image the information is implicit in the colouring of particular regions. While experts can interpret the visualization to obtain relevant information, computer systems can hardly access and manipulate such implicit information.

The use of weather forecasts in the prediction of plant disease has a long history dating back to at least the early second half of the last century (Bourke, 1970), with modern computer-based weather-related decision support systems surfacing towards the end of the century (Bouma, 2007). Jensen et al. (2000) describe a Web-based decision support system for crop management which, among other information, included graphical maps for risk of disease and pests. As for the earlier LUKE prototype briefly discussed above, the key difference with the system presented here is arguably the explicit representation of the information content in maps, abstracted as situational knowledge.

More recently, semantic technologies also gained traction in agriculture informatics. Gaire et al. (2013) provide a brief overview of the Kirby smart farm in Armidale, New South Wales, Australia, and its ICT infrastructure, which builds on Semantic Web technologies and Complex Event Processing to support event notification. Contrary to Gaire et al. (2013), who adopt the concept of event, we use the concept of situation in knowledge modelling. Combining the notion of situation and event may be an interesting future direction. Dutta et al. (2014) employ semantic technologies in an agricultural decision support system for water resource estimation. Of particular interest here is that their system seems to explicitly represent knowledge obtained from data using models, data-driven models in particular (Morshed et al., 2013). Pokharel et al. (2014) present the integration of data from heterogeneous sources using Semantic Web technologies in the domain of agriculture, with the aim of improving the effectiveness of farming, specifically in Nepal. The authors make use of AGROVOC8 (Caracciolo et al., 2013), a thesaurus covering Food and Agriculture Organization of the United Nations (FAO) terminology, also available in RDF. AGROVOC could be used in the work presented here. For instance, the thesaurus provides a term for Pathogenic fungi (Term Code 29233) which could serve in the classification of the fungal pathogens of interest here, e.g. *Pyrenophora teres* (Term Code 34565).

The framework for situation awareness in environmental monitoring has been used in various applications. Stocker et al. (2014b) describe the application of an early version of the framework to intelligent transportation systems. The objects in situations are vehicles, which were detected and classified using observation data by a sensor network that monitored road-pavement vibration, by means of machine learning (Mitchell, 1997). Stocker et al. (2014a) describe the application of the framework to the representation of situational knowledge for atmospheric phenomena. The objects in situations are new particle formation and (rainy) cloud events, which were detected and characterized in observation data by an environmental sensor network that monitored aerosols, visibility and precipitation. Machine learning was used to detect and characterize new particle formation in daily observation data of particle size distribution for polydisperse aerosols. Cloud events were detected and characterized using complex event processing (Luckham, 2002).

There are three important differences between the applications developed so far and the application described here. First, so far the applications were on historical data. In contrast, the application presented here occurs in real time. Second, in previous applications, situations were represented as they occurred (in historical data). In contrast, here situations are projected into the near future. Third, in previous applications, situational knowledge acquisition was by means of data-driven techniques, in particular machine learning. In contrast, the application presented here relies on a mechanistic ecological model, i.e. the disease pressure model. The application shows that the framework can support extensions which include mechanistic ecological that use processed data (of environmental sensor networks). In the framework, situational knowledge obtained via models, including mechanistic

Markus Stocker, Jussi Nikander, Hanna Huitu, Marja Jalli, Markku Koistinen, Mauno Rönkkö, Mikko Kolehmainen:
Representing Situational Knowledge for Disease Outbreaks in Agriculture

ecological models, is represented explicitly, in symbolic form, as situations, i.e. information objects grounded in formal Situation Theory (Devlin, 1991). Situations of a knowledge base can be queried, shared, integrated, manipulated, visualized. To explicitly represent situational knowledge is a mechanism that addresses the problem of having such knowledge only implicitly in, for instance, maps, plots, or tables.

We highlight an important limitation of this study. Lacking data for ground truth during the growing season we cannot evaluate the agreement of situations modelled by the system with actual situations on the ground. Experts did perform a quantitative analysis for the percentage of infected samples on July 7 for several agricultural parcels. We summarize here the results for two agricultural parcels. For the first parcel, the analysis shows that on July 7 44% of the samples were infected with *Stagonospora nodorum* and *Pyrenophora tritici-repentis*. On July 7, the system modelled acute outbreak situations at this parcel for *Stagonospora nodorum* since May 29 and for *Pyrenophora tritici-repentis* since June 1, i.e. for over one month. For the second parcel, the analysis shows that on July 7 10% of the samples were infected with *Pyrenophora teres*. On July 7, the system modelled outbreak situation at this parcel for *Pyrenophora teres* for the second day. Experts estimate the acute outbreak threshold at 20% infection. At 44%, infection at the first parcel is thus well beyond the threshold. This may suggest agreement with system modelling for the acute outbreaks that had been lasting for over one month. At 10%, infection at the second parcel is below the threshold and, according to expert opinion, in line with system modelling for outbreak, in particular if the crop is dense and keeps the night humidity for some hours, enough for spores to infect. While these results are promising, the currently available data is unfortunately insufficient to systematically validate system performance. We consider a comprehensive system performance evaluation using laboratory analysis and/or surveys to be an opportunity for future work.

4. Conclusion

We have presented and evaluated the application of an open source software framework for situation awareness in environmental monitoring to the problem of disease outbreak situation awareness in agriculture. The resulting system was capable of real time and automated explicit representation of knowledge for outbreak situations of plant pathogens in agricultural crops. Situational knowledge was obtained using forecast data for various weather parameters and a disease pressure model that acted as mechanistic ecological model for the computation of accumulated risk using weather and agricultural parcel data. The system was evaluated on 3 fungal pathogens, 2 crops, and 17 agricultural parcels located in Finland during a growing season. We also briefly compared the system presented here to an earlier prototype for the computation and visualization of disease pressure in agriculture.

The study shows that disease outbreak in agriculture--in particular the objects involved, such as pathogens and agricultural crops--can be modelled as situations. Situation Theory is a mathematical ontology for the modelling of information about situations. Based on various applications in environmental monitoring, including the one presented here, we think that situation is a powerful abstraction for information about the structured parts of reality monitored by environmental monitoring systems, specifically also those involving environmental sensor networks.

The explicit representation of (situational) knowledge obtained in environmental monitoring systems from processed data is arguably an interesting and important step. In fact, symbolic knowledge managed by a knowledge base can be queried, shared, integrated, and reasoned on. Generic tools can be developed to process symbolic knowledge toward various aims, e.g. visualization. Together with our previous studies, we think that the application presented here further supports the claim that the explicit representation of (situational) knowledge obtained using data-driven or mechanistic ecological models from processed data in environmental monitoring systems is a powerful mechanism to avoid that such knowledge only remains implicit in maps, plots, or text.

References

Allen, J.F., 1983. Maintaining Knowledge About Temporal Intervals. Commun. ACM 26, 832-843. doi: 10.1145/182.358434

Barwise, J., Perry, J., 1981. Situations and Attitudes. The Journal of Philosophy 78, 668-691. doi: 10.2307/2026578

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The Semantic Web. Scientific American 284, 29-37. doi: 10.1111/j.1365-2338.2007.01119.x

Bouma, E., 2007. Computer aids for plant protection, historical perspective and future developments. EPPO Bulletin 37, 247-254. doi: 10.1111/j.1365-2338.2007.01119.x

Bourke, P.M.A., 1970. Use of Weather Information in the Prediction of Plant Disease Epiphytotics. Annual Review of Phytopathology 8, 345-370. doi: 10.1146/annurev.py.08.090170.002021

Brickley, D., Guha, R., 2004. RDF Vocabulary Description Language 1.0: RDF Schema. Recommendation. W3C. URL: http://www.w3.org/TR/2004/REC-rdf-schema-20040210/.

Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., Keizer, J., 2013. The AGROVOC Linked Dataset. Semantic Web 4, 341-348.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W.D., Phuoc, D.L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K., 2012. The SSN ontology of the W3C semantic sensor network incubator group. Web Semantics: Science, Services and Agents on the World Wide Web 17, 25-32. doi: 10.1016/j.websem.2012.05.003

Cyganiak, R., Reynolds, D., Tennison, J., 2014. The RDF Data Cube Vocabulary. Recommendation. W3C. URL: http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/.

Devlin, K., 1991. Logic and Information. Cambridge University Press.

Dutta, R., Morshed, A., Aryal, J., D'Este, C., Das, A., 2014. Development of an intelligent environmental knowledge system for sustainable agricultural decision support. Environmental Modelling & Software 52, 264-272. doi: 10.1016/j.envsoft.2013.10.004

Endsley, M.R., 1995. Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors 37, 32-64. doi: 10.1518/001872095779049543

Fountas, S., Wulfsohn, D., Blackmore, B., Jacobsen, H., Pedersen, S., 2006. A model of decision-making and information flows for information-intensive agriculture. Agricultural Systems 87, 192-210. doi: 10.1016/j.agsy.2004.12.003

Gaire, R., Lefort, L., Compton, M., Falzon, G., Lamb, D., Taylor, K., 2013. Semantic Web Enabled Smart Farming, in: Proceedings of the 1st International Workshop on Semantic Machine Learning and Linked Open Data (SML2OD) for Agricultural and Environmental Informatics, Sydney, Australia.

Gent, D.H., Mahaffee, W.F., McRoberts, N., Pfender, W.F., 2013. The Use and Role of Predictive Systems in Disease Management. Annual Review of Phytopathology 51, 267-289. doi: 10.1146/annurev-phyto-082712-102356. pMID: 23682914.

Hart, J.K., Martinez, K., 2006. Environmental Sensor Networks: A revolution in the earth system science? Earth-Science Reviews 78, 177-191. doi: 10.1016/j.earscirev.2006.05.001

Hayes, P., McBride, B., 2004. RDF Semantics. Recommendation. W3C.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F., Rudolph, S., 2012. OWL 2 Web Ontology Language Primer (Second Edition). Recommendation. W3C. URL: https://www.w3.org/TR/owl2-primer/.

Jensen, A.L., Boll, P.S., Thysen, I., Pathak, B., 2000. Pl@nteInfo--a web-based system for personalised decision support in crop management. Computers and Electronics in Agriculture 25, 271-293. doi: 10.1016/S0168-1699(99)00074-5

Kokar, M.M., Matheus, C.J., Baclawski, K., 2009. Ontology-based situation awareness. Information Fusion 10, 83-98. doi: 10.1016/j.inffus.2007.01.004

Kotamäki, N., Thessler, S., Koskiaho, J., Hannukkala, A.O., Huitu, H., Huttula, T., Havento, J., Järvenpää, M., 2009. Wireless in-situ Sensor Network for Agriculture and Water Monitoring on a River Basin Scale in Southern Finland: Evaluation from a Data User's Perspective. Sensors 9, 2862-2883. doi: 10.3390/s90402862

Luckham, D.C., 2002. The Power of Events. volume 204. Addison-Wesley Reading.

van Maanen, A., Xu, X.M., 2003. Modelling Plant Disease Epidemics. European Journal of Plant Pathology 109, 669-682. doi: 10.1023/A:1026018005613

Manola, F., Miller, E., McBride, B., 2004. RDF Primer. Recommendation. W3C.

Martinez, K., Hart, J.K., Ong, R., 2004. Environmental Sensor Networks. Computer 37, 50-56. doi: 10.1109/MC.2004.91

Mitchell, T.M., 1997. Machine Learning. 1 ed., McGraw-Hill, Inc., New York, NY, USA.

Morshed, A., Dutta, R., Aryal, J., 2013. Recommending Environmental Knowledge As Linked Open Data Cloud Using Semantic Machine Learning, in: Data Engineering Workshops (ICDEW), 2013 IEEE 29th International Conference on, pp. 27-28. doi: 10.1109/ICDEW.2013.6547421

Nikander, J., Jalli, M., Erlund, P., Kivipelto, J., Peltonen, S., Huitu, H., 2015. Model for Cereal Crop Leaf Disease Pressure Calculation. Plant Disease (Manuscript).

Perry, M., Herring, J., 2012. OGC GeoSPARQL - A Geographic Query Language for RDF Data. Technical Report OGC 11-052r4. Open Geospatial Consortium Inc.

Pokharel, S., Sherif, M.A., Lehmann, J., 2014. Ontology Based Data Access and Integration for Improving the Effectiveness of Farming in Nepal, in: Proceedings of the International Conference on Web Intelligence, IEEE, Warsaw, Poland. doi: 10.1109/wi-iat.2014.114

Prud'hommeaux, E., Seaborne, A., 2008. SPARQL Query Language for RDF. Recommendation. W3C. URL: http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/.

Rew, R., Davis, G., 1990. NetCDF: an interface for scientific data access. Computer Graphics and Applications, IEEE 10, 76-82. doi: 10.1109/38.56302

Roberts, M.J., Schimmelpfennig, D.E., Ashley, E., Livingston, M.J., Ash, M.S., Vasavada, U., 2006. The Value of Plant Disease Early-Warning Systems: A Case Study of USDA's Soybean Rust Coordinated Framework. Economic Research Report 7208. United States Department of Agriculture, Economic Research Service. URL: https://ideas.repec.org/p/ags/uersrr/7208.html.

Stanton, N.A., Salmon, P.M., Walker, G.H., Jenkins, D.P., 2010. Is situation awareness all in the mind? Theoretical Issues in Ergonomics Science 11, 29-40. doi:10.1080/14639220903009938.

Stanton, N.A., Stewart, R., Harris, D., Houghton, R.J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M.S., Linsell, M., Dymott, R., Green, D., 2006. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics 49, 1288-1311. doi: 10.1080/00140130600612762. pMID: 17008257.

Stocker, M., Baranizadeh, E., Portin, H., Komppula, M., Rönkkö, M., Hamed, A., Virtanen, A., Lehtinen, K., Laaksonen, A., Kolehmainen, M., 2014a. Representing situational knowledge acquired

from sensor data for atmospheric phenomena. Environmental Modelling & Software 58, 27-47. doi: 10.1016/j.envsoft.2014.04.006

Stocker, M., Rönkkö, M., Kolehmainen, M., 2014b. Situational knowledge representation for traffic observed by a pavement vibration sensor network. IEEE Transactions on Intelligent Transportation Systems 15, 1441-1450. doi: 10.1109/TITS.2013.2296697

Thessler, S., Kooistra, L., Teye, F., Huitu, H., Bregt, A.K., 2011. Geosensors to Support Crop Production: Current Applications and User Requirements. Sensors 11, 6656-6684. URL: http://www.mdpi.com/1424-8220/11/7/6656, doi: 10.3390/s110706656

Wang, N., Zhang, N., Wang, M., 2006. Wireless sensors in agriculture and food industry--Recent development and future perspective. Computers and Electronics in Agriculture 50, 1-14. doi: 10.1016/j.compag.2005.09.003

Wiersma, G., 2004. Environmental Monitoring. CRC Press. doi: 10.1201/9780203495476

Markus Stocker, Jussi Nikander, Hanna Huitu, Marja Jalli, Markku Koistinen, Mauno Rönkkö, Mikko Kolehmainen:
Representing Situational Knowledge for Disease Outbreaks in Agriculture