A knowledge-based system for generating interaction networks from ecological data

Willem Coetzer[1,2,5*], Deshendran Moodley[2,4], Aurona Gerber[2,3]

[1] SAIAB: South African Institute for Aquatic Biodiversity, Private Bag 1015, Grahamstown 6140, South Africa

[2] CAIR: Centre for Artificial Intelligence Research, CSIR Meraka, PO Box 395, Pretoria, 0001, South Africa

[3] Department of Informatics, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

[4] Department of Computer Science, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa

[5] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Private Bag X54001, Durban 4000, South Africa

[*] Corresponding author. E-mail: w.coetzer@saiab.ac.za

Abstract

Semantic heterogeneity hampers efforts to find, integrate, analyse and interpret ecological data. An application case-study is described, in which the objective was to automate the integration and interpretation of heterogeneous, flower-visiting ecological data. A prototype knowledge-based system is described and evaluated. The system's semantic architecture uses a combination of ontologies and a Bayesian network to represent and reason with qualitative, uncertain ecological data and knowledge. This allows the high-level context and causal knowledge of behavioural interactions between individual plants and insects, and consequent ecological interactions between plant and insect populations, to be discovered. The system automatically assembles ecological interactions into a semantically consistent interaction network (a new design of a useful, traditional domain model). We discuss the contribution of probabilistic reasoning to knowledge discovery, the limitations of knowledge discovery in the application case-study, the impact of the work and the potential to apply the system design to the study of ecological interaction networks in general.

Keywords: Semantic heterogeneity, ontologies, Bayesian network, knowledge discovery, semantic architecture, interaction network, ecological interactions

1. Introduction

Studies of the behavioural and community ecology of flower-visiting insects, which can be inferred to pollinate flowers, are relevant to theoretical ecology and have important applications in agriculture [1–3] and conservation [4]. Flower-visiting observations are field notes about living insects recorded in nature by ecologists. Many flower-visiting observations are associated with specimen-records of insects now preserved in natural history collections. The data include the names of the plant species whose flowers the insects had visited immediately before the insects were killed and preserved. Such data documenting the relationships between individual organisms (i.e. plants and insects) in a natural setting are considered to be unusually rich ecological data, and the fact that they are 'vouchered' by evidence preserved in museums also means that the identities of the insect organisms can be verified in future. Ecological data (let alone data with rich annotations) are infrequently supported by such physical evidence.

Much progress has been reported in initiatives to advance eScience techniques broadly in the field of biodiversity and ecosystem informatics (BDEI) [5–7]. Due to semantic heterogeneity, however, analysts still face significant challenges when attempting to find, integrate and analyse specific ecological data, including flower-visiting data, among ever larger and more-fragmented datasets and heterogeneous data. Semantic heterogeneity also hampers the interpretation of data. Ecological data typically are incomplete and exhibit uncertainty, and therefore usually require experts, who have implicit knowledge, to analyse or interpret their meaning.

In this work our overall objective was to formalise the specific context of behavioural and community flower-visiting ecology, to infer from the data a network of ecological interactions between plant and arthropod populations, analogous to an ecological community. We describe an application case-study, constrained within a scope and understood by adopting a conceptual stance. In the application case-study, we combined ontologies and a probabilistic knowledge model (a Bayesian network)—a semantic architecture—in a prototype implementation of a knowledge-based system, the purpose of which was to standardise and automate the interpretation of (discovery of knowledge in) flower-visiting data. Standardisation in the construction of interaction networks is required because ecologists need to compare networks from different years or places (e.g. to study the effects of global change). Automation will allow non-experts to construct ecological interaction networks from raw data, or interpret the data without further input from expert ecologists. Modelling knowledge of the diverse array of concepts about interactions between plants and arthropods (of which pollination is but one) will therefore be useful to experts and non-experts alike.

In Section 2 we introduce the background to the problem of integrating and interpreting heterogeneous flower-visiting data. Here we also describe related work in ontology modelling of behaviour, and the application of ontologies to discovery and integration of biodiversity and ecological data. In Section 3 the application case-study is described, including the scope, conceptual stance and modelling approach. Section 4 describes the knowledge models and system architecture, and Section 5 is a system evaluation, including a description of the prototype implementation and results obtained. In section 6 we discuss the extent to which the system was able to infer an interaction network, and the potential impact of this work in the area of automated interpretation of ecological interaction networks in general.

2. Literature review and background

The small size of many flowers and pollinating insects means that the transfer of pollen between flowers, or deposition of pollen on small and inaccessible stigmas, are not readily

observed. This explains why observations of visits by insects to flowers typically are used to infer that pollination occurred. In addition, there are other insect behaviours and consequent ecological interactions between insects and plants, such as foraging for nectar and pollen, which are important in the study of flower-visiting community ecology, but are not readily observed.

The interaction network is a generic modelling construct that is commonly used in the broader domain of ecology to visualise various kinds of relationships between interacting species. Different ways of inferring specific plant-animal interaction networks from data appear in the literature, including mathematical techniques using symbolic computation and algebraic combinatorics [8], statistical techniques, including correlation analysis [9], hierarchical Bayesian models [10] and Bayesian networks [11,12], and computational methods, including machine learning [13] and network theory [14].

Field experiments in flower-visiting ecology typically include a diversity of techniques, and produce data which differ from observations collected in structured field surveys or accumulated over time by natural history museums. There has been some reflection, within the field of flower-visiting ecology, on the heterogeneity of concepts and terminology (e.g. 'pollinator' and 'visitor') [15], and consequent representation of specific pollination networks as opposed to more general flower-visiting networks [16,17]. Interaction networks compiled in different investigations are typically assembled using different methods of data analysis and interpretation, and network nodes and arcs can represent different concepts from study to study. All of these kinds of heterogeneity and uncertainty mean that interaction networks cannot easily be compared, yet explicitly making such comparisons is an important objective in community ecology, especially considering the current focus on global change and the importance of pollination in food security [18–23]. Analysts therefore need a standardised protocol to address semantic heterogeneity in the analysis of flower-visiting data, and standardised techniques to interpret data and automatically and consistently assemble comparable flower-visiting interaction networks.

2.1 Ecological data and knowledge

Traditional approaches to ecological modelling using mathematical equations are hampered by the qualitative nature of ecological knowledge [24]. Typical ecological data and knowledge are 'incomplete, qualitative and fuzzy, often expressed verbally and diagrammatically' [25], and are not easily represented in discrete classes nor subjected to discrete reasoning. The sources of uncertainty in ecology and conservation biology have been summarised in a taxonomy of uncertainty, which lists the various kinds of epistemic uncertainty (e.g. measurement error and model uncertainty) and linguistic uncertainty (e.g. ambiguity and underspecificity), which tend to be compounded in ecological data [26]. An ecologist needs to be comfortable commuting a vast hierarchy of spatio-temporal granularity, from the gene to the ecosystem, into a practical conceptual model. Moreover, the traditions of natural science research do not discourage discursive presentation of knowledge, and an ecologist's conceptual model of causal knowledge may even be largely implicit. However, if ecological knowledge can be made 'explicit, well-organized and computer processable, great predictive power could be harnessed through the integration of quantitative and qualitative knowledge' [22]. This could result in 'more efficient ways of organizing, processing and analysing ecological knowledge to emphasize and facilitate the process of ecological reasoning rather than data reduction' [25,27].

Flower-visiting data are incomplete and uncertain ecological observations that are used to make assertions and express concepts that are not semantically consistent within or between datasets, yet ecologists are able to use implicit knowledge manually to assemble such data into interaction networks to interpret the data (i.e. to interpret data means to assemble interaction networks). The objective of the work described below, therefore, was to standardise and automate the analysis and interpretation of data, i.e. to formalise,

standardise and automate the assembly of flower-visiting interaction networks.

2.2 Ontologies for behavioural ecology

Ontologies have successfully been used to standardise metadata terminology for the discovery, integration (by semantic mediation), and re-use of biodiversity and ecological data [28]. Limited modelling of behavioural concepts has been undertaken e.g. in the context of human neurobiology [29,30]. In biodiversity and ecosystem informatics (BDEI) aspects of behavioural ecology have been modelled in male jumping-spider courtship behaviour, sea-turtle nesting behaviour [31] and the behaviour of social insects [32].

The class 'multi-organism behaviour' was originally defined in the Gene Ontology (GO) [33] (Fig. 1) and is now imported into 8 other ontologies including the Population and Community Ontology [28]. This class includes 'any process in which an organism has a behavioural effect on another organism of the same or a different species' i.e. a behavioural interaction between organisms. The class 'feeding on or from other organism' is also defined in the Population and Community Ontology and the Neuro-behaviour Ontology. These classes are both subsumed by the class BFO:Behavior, a subclass of the class BFO:Biological_Process, which is a kind of BFO:Occurrent [34] (Fig. 1).

A subsumption hierarchy of specific flower-visiting object properties appears in the Relations Ontology (RO) [35] *viz.* the object properties *RO:visits_flowers_of* and *RO:has_flowers_visited_by*, which are ultimately subsumed by *RO:biotically_interacts_with*, which is subsumed by *RO:ecologically_related_to*. These object properties were modelled specifically to facilitate vertical integration of ecological data in broadly defined classes [36]. In contrast, the purpose of the knowledge models described below is to preserve the specific, original context of the data as far as is possible, and enrich the data, so as to discover specific ecological knowledge in the data.

Fig. 1. The definition of the class 'multi-organism behavior', originally defined in the Gene Ontology [37].

Thing
- + entity
    - + occurrent
        - + process
            - + biological_process
                - + response to stimulus
                    - + behavior
                        - + reproductive behavior
                        - + single-organism behavior
                        - - multi-organism behavior

2.3 Ontologies for community ecology

Biodiversity and ecology are complex domains, partly due to the challenge, in knowledge representation and reasoning, of adequately representing spatio-structural granularity, or the hierarchy of levels of organization or complexity observed in biological systems (e.g. cell < tissue < organ in the biomedical domain, and individual < population < community in ecology). A more detailed explanation of the terms 'population' and 'community' and ecological complexity appears below.

Ontologies have been created for ecological informatics, including an 'ecology ontology' as well as ontologies for ecological models, analysis methods and ecological networks (used specifically for food webs) [38]. In these ontologies there is necessarily much emphasis on

representation of discrete knowledge and discrete reasoning e.g. that a herbivore can be inferred to have eaten plants even if what it actually ingested remained unknown, which is useful when generating a food web (food webs are discussed in more detail below).

In ecology a useful ontology model and system architecture will need to represent the complexity of relationships between organisms, as well as between organisms and the environment, at the various scales at which these relationships are thought to be significant. It has been noted that 'there are complementary ways to conceptualise ecological systems,' e.g. as individuals, populations or communities, or as a flow—of information, a substance such as a pollutant or nutrient, or energy [39,40]. Depending on the scale of an observation, processes can be modelled as entities or entities modelled as processes e.g. a population can be modelled as an entity unless it is seen as being composed of individuals, in which case it is a changing process [40].

2.4 Application of ontologies in biodiversity and ecosystem informatics

Many ontologies in the field of BDEI describe low-level concepts about the data record itself (e.g. data provenance), rather than high-level context or causality. For example, ontologies have been used to create semantic annotations of individual records or data-processing steps in scientific workflow systems [41–44]. The Extensible Observation Ontology (OBOE) captures the semantics of generic scientific observation and measurement, and can be used 'to characterize the context of an observation (e.g. space and time), and clarify inter-observational relationships such as dependency hierarchies (e.g. nested experimental observations) and meaningful dimensions within the data' [45]. The Biological Collections Ontology (BCO) has a similar purpose, with classes that describe the methods employed by scientists to collect specimens or observations of individual organisms, or in structured ecological surveys or environmental samples (e.g. a bucket of seawater containing plankton). BCO and related ontologies have been used to link semantically annotated data across sub-disciplines of biodiversity science using the approach of Linked Data [46]. For example, a comprehensive inventory of the non-microbial life on the Pacific island of Moorea has been created [28]. A subset of the data was annotated with classes from BCO and related ontologies, demonstrating how such datasets can be queried easily despite the diversity of methods and sampling situations, which would ordinarily restrict data to discipline-specific silos (e.g. Genetics or Ecology).

Brilhante [47] defined metadata classes for quantitative ecological data by drawing on the EngMath ontology. The resultant ontology was used to synthesize new conceptual ecological models from metadata in datasets by matching an existing model with input metadata concepts constrained by the ontology.

In semantic environmental modelling, ontologies can be used to declare a semantically enriched model by specifying [48] :

    a) the modelled entities, by identifying the relevant concepts and properties;

    b) the underlying relationships among these entities, to capture the structure of causality in the system as understood by the modeller.

There may be limitations to the application of ontologies to nuanced biodiversity and ecological data and knowledge, including that ontologies do not explicitly support causal modelling or uncertainty.

2.5 Bayesian networks

Differential equations are often used to represent causal knowledge in environmental and ecological modelling. Causal knowledge has also been represented using Bayesian networks. A Bayesian network (BN) is a graphical model that probabilistically represents causal (or correlative) relationships among variables [49,50]. Nodes in the graph represent event variables which are connected by arcs representing causal influences between events. A BN node is implicitly understood to be an event which can be in one of a number of states at a given time. To specify the probability distribution of a BN, one 'must give the prior probabilities of all root nodes (nodes with no predecessors) and the conditional probabilities of all nonroot nodes given all possible combinations of their direct predecessors' [49]. Bayesian networks have been used widely in ecology and natural resource management, e.g. to evaluate the potential effects of alternative forest management decisions, and represent uncertainty and variability of costs and benefits assigned to model outcomes [51]. Bayesian networks have also been used specifically to infer ecological interaction networks using only species and habitat abundance [11,13], but not from the observation of, or knowledge about, behavioural interactions between individual organisms, such as the work described below.

While ontologies and BN models have been applied in the geospatial domain, no study could be found, in the domain of BDEI, which combined both of these formalisms. In the field of the Sensor Web (distributed instruments and data for Earth observation) an approach to knowledge discovery involved combining ontologies and Bayesian networks in a probabilistic reasoning system. Bayesian networks were used to represent uncertainty and causal relations between environmental variables. This 'eases conceptual modelling and allows for more flexible reasoning' [52]. Specifications of scientific theories and system modelling were integrated into the Sensor Web Agent Platform (SWAP), a 'comprehensive framework for representing all aspects of geospatial data (space, time, theme and uncertainty) and the knowledge (theories and models) to interpret and analyse the data', as well as software agents to manage and dynamically apply knowledge to the data [52]. A SWAP Bayesian Network has two types of nodes: observation or measurement nodes, which represent sensor observations, and inferred nodes, which represent natural phenomena. One of the novel aspects was a mapping mechanism between observations captured in ontologies and event (observation) variables in the Bayesian Network [52,53]. The combination of formalisms and semantic architecture of SWAP [52] may therefore be a promising approach to automated knowledge discovery in biodiversity and ecological data.

3. Application case-study

The scope, conceptual stance and modelling approach, which are described below, served to constrain the work to a specific real-world application case-study. The work was an exploration of the potential to combine ontologies and a Bayesian network to model the context of flower-visiting community ecology, automatically to infer an ecological interaction network from semantically heterogeneous data. Specific knowledge was elicited from experts to create knowledge models, which were used to design a knowledge-based system. The system output was evaluated by experts.

Three data-stores of flower-visiting observations were used, namely those of the Albany Museum (AM) in Grahamstown, Iziko South African Museum (SAM) in Cape Town and Plant Protection Research Institute (SANC) in Pretoria. In previous work we had used an application ontology to enrich the meaning of raw data and integrate the data by semantic mediation [54]. In further work we combined semantically enriched records of plant-arthropod associations with expert knowledge of the species' behavioural ecology in a semantic Bayesian network [55] to detect meaningful situations or behaviours e.g. that an organism was probably 'foraging for nectar' on a flower. Previously our work has been limited to the transformation of raw data into high-level, knowledge-rich abstractions of individual organisms.

Below we describe a continuation of the previous application case-study [54]. We incorporate probabilistic reasoning into the system architecture and extend the knowledge modelling to a higher level of abstraction to aggregate, further analyse and automatically interpret enriched records by assembling an interaction network, a commonly used domain modelling construct. Rather than attempting to produce a universal or comprehensive model of plant-arthropod interactions, we aimed to test a specific combination of knowledge models and formalisms to discover knowledge of these interactions within the constraints of the scope and conceptual stance, and by using the expert knowledge elicited in previous work [55].

3.1 Scope

The behaviour of anthophilous (flower-visiting) arthropod species occurring outside of Africa was excluded from the scope (e.g. orchid bees, which collect fragrances from flowers to attract mates). We modelled three specific behaviours that distinguish the more specialised anthophilous African insect species, which typically pollinate flowers, from arthropod species that can be found on flowers but are not typical flower-visitors (i.e. they are either opportunistic or incidental flower-visitors). These specialised behaviours are foraging for nectar, foraging for pollen and foraging for oil (or foraging for a floral product or 'reward'), and they typify, but are not restricted to, the bees (Anthophila) and the wasp subfamily Masarinae (pollen wasps). We also modelled the passive transfer of pollen (a pre-requisite of pollination), which is an incidental consequence of these specialised foraging behaviours. This ultimately explains the evolution of the plant-pollinator mutualism. Pollination is the benefit received by the plant organism in return for offering the floral 'reward' to pollinators. Whereas pollination can sometimes be caused by bees collecting floral resin and nectar, not for ingestion but for nest construction (behaviour that is exhibited by many species in the family Megachilidae), this behaviour was not explicitly included in the scope because it is not a foraging behaviour.

We excluded arthropod observations that are not linked to preserved museum specimens because we planned to enumerate and aggregate organisms of the same species (i.e. an instance of 'at least two organisms'), and therefore needed to be certain that different database records represent different individual arthropod organisms, each labelled with a unique museum catalogue number. Knowledge modelling was limited to arthropod specimens collected on seedplants (gymnosperms and angiosperms) (i.e. for aggregation into the class *ArthropodPopulation*). Preservation of plant specimens, however, is not routinely practiced as part of arthropod field surveys, so the modelling of plant populations (i.e. the class *PlantPopulation*) was not limited to preserved specimens.

3.2 Conceptual stance

The conceptual stance was informed by ecological theory as well as more recent philosophical work in ecology, which was the source [56] of the following practical definitions of the most widely used ecological units. The first two concept definitions were used (whereas the high-level concept of an ecosystem is beyond the scope but is provided for the sake of completeness):

Population: a group of individual organisms of the same species in space and time;
Community: an assemblage of organisms of different types (species, life forms) in space and time;
Ecosystem: an assemblage of organisms of different types (species, life forms) together with their abiotic environment in space and time.

Specifically, we mean that the behaviour of individual organisms can be observed, and the repeated occurrence of a particular behaviour by many individual organisms (aggregated behaviour) within a spatio-temporal context is meaningful at the population level—*e.g. the*

foraging behaviour exhibited by a population of bees of a particular species in an apple orchard during the summer of 2015-2016. This population of bees can be said to interact with other, co-existing organisms of a different population (and different species e.g. the population of apple trees), and this gives rise to a phenomenon that may be termed a *foraging ecological interaction* between organisms of the population of bees and organisms of the population of apple trees. The ecological interaction cannot be defined intensionally at any level of ecological organisation but rather emerges as a consequence of the ensemble of this bee population's individual members behaving (e.g. specifically foraging for nectar), at this time, in a particular way towards individuals of the population of apple trees. It will be seen below, however, that if space and time are removed, the picture is not lost but rather looks different and has a different meaning. Importantly, therefore, this conceptual stance allowed us to aggregate individual organisms as well as their behaviour, in instances that exist at a higher level of organisation (aggregated organisms and aggregated behaviour). The result of abstracting the salient behavioural and ecological phenomena is the abstraction hierarchy shown in Fig. 2.

This conceptual stance is commensurate with that of individual-based computational modelling as applied to ecology, in which the individual, population and community levels of ecological organisation are recognised. Processes occurring at the individual level produce patterns at higher levels of organisation, and small individual differences can lead to significant effects at the population or community levels [57].

Fig. 2. The abstraction hierarchy of behavioural and ecological phenomena.
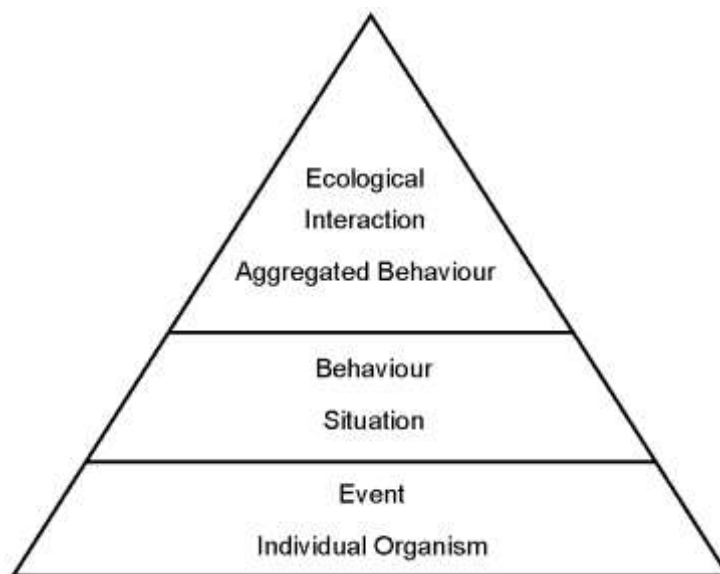
3.3 Modelling approach

The knowledge that was modelled included knowledge of (the organisation of) ecological phenomena, generally available knowledge of plant and arthropod species, and their occurrences (observations, i.e. data), as well as expert knowledge of the behaviour of flower-visiting arthropods [55]. The ontology classes formalised:

1) the required context of the behavioural ecology of individual organisms now preserved as natural history specimens;

2) the context of plant-arthropod (mutualistic) ecological interactions (at a higher level of abstraction) inferred from records of these individual organisms.

While the modelled classes were not specifically integrated with existing domain or foundational ontologies, their concept definitions nevertheless are aligned with concepts encoded in the Basic Formal Ontology (BFO) and, at the domain level, the Darwin Semantic Web Ontology (DSW; e.g. the class `DSW:Occurrence`) [58]. A specific knowledge engineering methodology was not followed. Rather, the modelling of ontology classes was informed by interviews with flower-visiting ecologists and through reading relevant literature (top-down approach), as well as analysing flower-visiting data (bottom-up approach). Modelling in OWL was executed using the Protégé tool [59] and in accordance with the middle-out ontology construction approach [60].

In accordance with the conceptual stance and modelling approach, concepts that have instances at the individual level of organisation (i.e. the behaviour of individual organisms, or the study of behavioural ecology) were separated from concepts that have instances at the community level of organisation (i.e. ecological interactions between populations, or the study of community ecology). Two ontologies were therefore developed, named respectively, the Individual Plant-Arthropod Associations Ontology (IPA) and the Interaction Network Ontology (IN). A notional whole population, represented by the *PopulationSample* class, was included in the community level because we did not model concepts used in the study of population ecology, such as population size or the rate of population growth.

Instead of using differential equations, a Bayesian network was used to capture causal knowledge of behavioural ecology. The combination of ontology and Bayesian network therefore follow an approach demonstrated in Earth Observation [52]. The causal knowledge model was of central importance because it was used to reason about the behaviours of individual organisms, and it was these behaviours that were aggregated at higher levels of organisation to represent the higher-level context (i.e. community ecology).

4. System description

The purpose of the knowledge-based system is to transform typical natural history occurrence data into a flower-visiting interaction network by combining the data with relevant (if qualitative and uncertain), generally available knowledge and expert knowledge. The semantic architecture (Fig. 3) consists of three layers which reflect the abstraction hierarchy of behavioural and ecological phenomena introduced above.

4.1 Overview of the system architecture

An overview of the three layers of the system architecture is given below, and each layer is described in more detail in the following sections.

Layer 1: The Semantic Enrichment and Mediation Layer

This layer enriches data and performs semantic mediation (see [54]) to integrate instances of the `IPA:PlantAssociationEvent` class and its subclasses. Each processed record is passed to Layer 2 via a mapping which sets the states of nodes in Layer 2.

Layer 1 contains two ontologies. The IPA Mapping Ontology maps records from the data-stores to the main ontology, *viz.* the *Individual Plant-Arthropod Associations Ontology* (IPA). The IPA Ontology captures knowledge of preserved specimens of plant and arthropod organisms and the low-level associations (events) between them in nature, as well as knowledge of plant and arthropod species that is relevant to flower-visiting behaviour.

Layer 2: The Situation Detection Layer

This layer uses a Bayesian network knowledge model, *viz.* the *Individual Flower-Visiting Behaviour Bayesian Network* (IFBN)*.* Using knowledge represented by the IPA, the IPA-IFBN mapping sets the nodes of the IFBN to the required evidence states. The IFBN is executed to detect the most probable high-level situation that occurred. It infers each arthropod organism's behaviour on the flower while it was alive, given the semantically enriched behavioural ecology observations and prior knowledge of the plant and arthropod species (received from Layer 1).

Layer 3: The Interpretation Layer

A mapping between IFBN and IN aggregates instances of individual arthropod organism behaviours received from Layer 2 into aggregated behaviour instances, which are then assembled into a generalised flower-visiting network or a specific ecological interaction network (according to the spatio-temporal parameters input by the user).

This layer uses a knowledge model, *viz.* the *Interaction Network Ontology* (IN), to represent aggregated behaviour (and specialised ecological interactions), aggregations of individuals (and specialised population samples), and structural classes of interaction networks and their constituent nodes.

4.2 System input and output

A system input query allows the user to specify spatio-temporal parameters (indicated as **s** and **t** in Fig. 3), and uses these to limit the spatio-temporal extent of the interaction network produced by the system. The user is further required to specify whether or not an ecological community of contemporaneously interacting populations (indicated as **c** in Fig. 3) can be expected to occur within the supplied spatio-temporal envelope.

The system output is a semantically consistent interaction network which has been enriched with the general context of plant-arthropod mutualistic interactions. Each interaction network is either enriched with the specific context of community ecology or with the more generalised context of evolutionary history. This will be further explained below.

Fig. 3 The system architecture.

Interaction network (s_1, t_1)

INTERPRETATION LAYER

Interaction Network Ontology **IN** (s,t)

IFBN <-> IN Mapping

SITUATION DETECTION LAYER

Individual Flower-Visiting Behaviour Bayesian Network **IFBN**

IPA<->IFBN Evidence Mapping

SEMANTIC ENRICHMENT AND MEDIATION LAYER

Individual Plant-Arthropod Associations Ontology **IPA**

IPA Mapping Ontology

io

SANC Mapping

io

SAM Mapping

io

AM Mapping

Jena API

BNJ Inference Engine

HermiT OWL Reasoner

Spatio-temporal Query (s,t,(c))

SANC

SAM

AM

Heterogeneous flower-visiting observations

4.3 Layer 1: The Semantic Enrichment and Mediation Layer

This layer receives occurrence records from the data-stores via the data-store mappings and enriches these by creating object properties, thereby creating associated events and associated species properties. The layer's output is enriched event instances (of the class `IPA:PlantAssociationEvent`) which are input into the situation detection layer.

IPA Mapping Ontology

Records received from a data-store are classified by an instance, unique to the data-store, of the IPA Mapping Ontology (Table 1), which has been created by an expert who has classified descriptions of arthropod behaviour in the Behaviour field of the data-store into one of the subclasses of the `IPA:PlantAssociationEvent` class (described below). The IPA Mapping Ontology also contains the `IPA:ForagingBehaviour` subsumption hierarchy (described below) because an expert is capable of asserting that a specific foraging behaviour was directly observed.

Table 1. Examples of data from SAM and AM, instantiated using the IPA Mapping Ontology.

Data-store	Value in Behaviour field in data-store	IPA Class
sam-m	Collecting pollen on yellow flowers	*PollenForagingBehaviour*
sam -m	Feeding on Brunia laevis pollen	*PollenForagingBehaviour*
sam -m	Foraging on nectar of Euphorbia flowers	*NectarForagingBehaviour*
sam -m	Visiting extra-floral nectaries	*PlantUtilizingEvent*
am-m	On foliage	*PlantUtilizingEvent*
am -m	On stem of plant	*PlantUtilizingEvent*
am -m	Visiting flowers	*FlowerVisitingEvent*
am -m	In flowers	*FlowerUtilizingEvent*
am -m	On flowers	*FlowerUtilizingEvent*

Individual Plant-Arthropod Ontology

The purpose of the IPA Ontology is two-fold:

1) to identify instances of the important class of events, *viz*. `IPA:PlantAssociationEvent`, in which plant and arthropod organisms are associated with each other, by filtering data-store records 'from the bottom up' via the data-store mappings;
2) to enrich these event instances with the necessary background knowledge or context (behavioural ecology and species knowledge) to allow the events to be interpreted in higher system layers.

The IPA Ontology therefore encodes two kinds of classes:
i) Knowledge of occurrences of species (now preserved as specimens in museum collections), which originated in the data-stores and was expressed with rich semantics;
ii) knowledge of species, or the 'species knowledgebase', consisting of generally available ecological knowledge of plant and arthropod species, which was not included in the data-

stores but compiled separately. This knowledge was expressed with a minimum ontological component.

i) Knowledge of occurrences of species

The definition of the `IPA:PlantAssociationEvent` class is shown in Fig. 4. The subsumption hierarchy specialising the `IPA:PlantAssociationEvent` class is of central importance in semantic mediation because heterogeneous data-store records are enriched with these classes, and the events are further interpreted and enriched in higher system layers. In this subsumption hierarchy the word 'association' means that there is no observational evidence with which to assert that an arthropod visited (e.g. landed on) a plant or flower, whereas the word 'visiting' means that such evidence does appear on the arthropod specimen label.

Fig. 4. Definition of the `IPA:PlantAssociationEvent` class.

Definitions of the `IPA:PlantOccurrence` and `IPA:ArthropodOccurrence` classes are shown in Fig. 5a and 5b. The latter are subclasses of the `DSW:Occurrence` class, defined as 'an organism at a time and place' [58]. The class definitions employ the classes `IPA:PlantOrganism` and `IPA:ArthropodOrganism`, and the property restriction:

```
(occursAt some TimeAndPlace).
```

Fig. 5a. Restriction of the IPA:PlantOccurrence class.

Fig. 5b. Restriction of the IPA:ArthropodOccurrence class.

ii) The species knowledgebase

The species knowledgebase (Table 2) in the IPA Ontology encodes classes and properties defining species knowledge, the choice of which was informed by knowledge representation and reasoning requirements elicited from expert flower-visiting ecologists [55]. The decision to use a probabilistic model to interpret arthropod organisms' behaviour, and link this to the IPA Ontology, therefore dictated which classes were needed in the IPA Ontology. In other words, the states of BN variables would need to be set from corresponding instances of equivalent classes in the IPA Ontology, and this informed the choice of classes needed in the IPA Ontology. Like the IPA Mapping Ontology, the species knowledgebase is static and was created prior to using the system to interpret ecological data.

Table 2a. Examples of knowledge in the plant species knowledgebase in the IPA Ontology.

Plant species	Floral product combination	Pollination system	Sexual system	Earliest flowering month	Latest flowering month
Pterygodium hallii	Oil_And_Pollinaria	Entomophily	Hermaphroditic	8	10

Table 2b. Examples of knowledge in the arthropod species knowledgebase in the IPA Ontology.

Arthropod species	Flower visitor type	Family	Subfamily (Required only when the family is Masaridae)
Rediviva macgregori	Oil_And_Nectar_And_ Pollen_Forager	Mellitidae	

4.4 Layer 2: The Situation Detection Layer

The specific behaviour of an arthropod organism was modelled using a Bayesian network because this allowed the inferencing of an expert flower-visiting ecologist (i.e. using causal knowledge) to be simulated. The situation detection layer receives input from the IPA Ontology, via the IPA-IFBN evidence mapping (described below), in the form of the requisite states in which to set the BN nodes. It reasons with this contextual expert knowledge of events, occurrences and species to interpret the most probable behaviour of an individual flower-visiting arthropod, and sends this behaviour instance to the layer above it in the semantic architecture (the Interpretation Layer).

Individual Flower-Visiting Behaviour Bayesian Network

The knowledge represented by the IFBN was elicited from 5 experts in the field of flower-visiting insect behavioural ecology [55]. Specifically, the structure of the IFBN and its conditional probability tables were designed using the input of these experts. The implemented IFBN is shown in Fig. 6 and selected variables of the IFBN are explained below.

Fig. 6. The Individual Flower-Visiting Behaviour Bayesian Network (IFBN). Variables representing generally available knowledge have bold outlines.
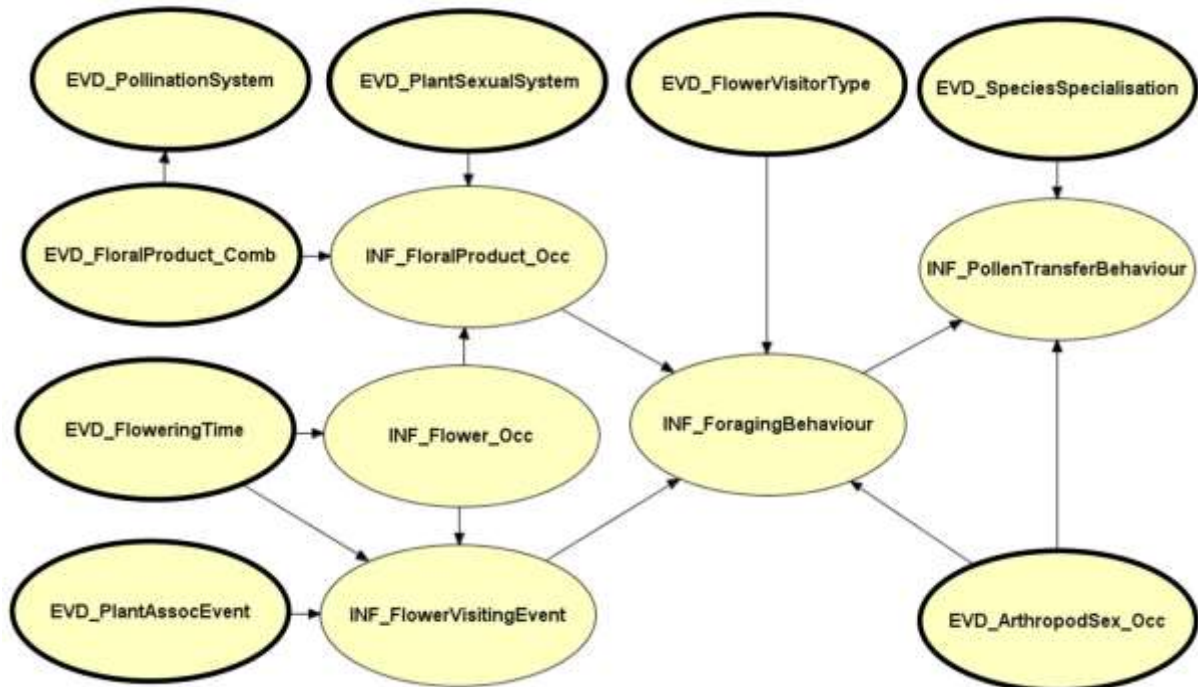
The posterior probability distributions over the states of the `INF_ForagingBehaviour` node and the `INF_PollenTransferBehaviour` node are the targets. The states of three nodes influence the belief of an expert that pollen will be transferred: the `INF_ForagingBehaviour` node, the `EVD_SpeciesSpecialisation` node (which is set by the IPA-IFBN evidence mapping using an inference made by the IPA Ontology reasoner) and the `EVD_ArthropodSex_Occ` node (the sex of the arthropod occurrence). The reason for this is that the fewer plant species an arthropod species is known to visit (specifically to forage for nectar i.e. the condition of oligophagy), the higher the chance that an organism of this species will visit the flowers of a different plant organism of the same plant species, and therefore transfer conspecific pollen between flowers. Female arthropods fly from flower to flower to collect nectar and pollen to provision their nests, and therefore have a higher chance than males of transferring loose pollen accidentally adhering to their bodies.

Table 3 shows the prior probabilities of the states of the node `EVD_FloralProduct_Comb`, or the combinations of floral products presented by an African seedplant species. Grass species, for example, do not secrete nectar because they are wind-pollinated. Many orchid species, which have pollen sacs, or pollinaria, rather than granular pollen, do not secrete nectar or oil (though some orchid flowers do secrete oil).

Table 3. The table of prior probabilities of the `EVD_FloralProduct_Comb` IFBN node

EVD_FloralProduct_Comb	Prior Probability
Oil_And_Pollinaria	0.05
Nectar_And_Pollinaria	0.1
Oil_And_Pollen	0.05
Nectar_And_Pollen	0.5
Secretion_Absent_Pollen_Present	0.15
Secretion_Absent_Pollinaria_Present	0.15

Table 4 is the table of prior probabilities of the `EVD_FlowerVisitorType` node. Females of most anthophilous insect species forage for nectar and pollen, whereas males forage only for nectar. A small group of bees (i.e. the genus *Rediviva*) are unique in that the females forage for floral oil on specific oil-producing plant species, though they also forage for nectar and granular pollen on nectar-producing plants (insects are not said to 'forage' for pollinaria because these adhere passively to insects). Again, the males of these species only require nectar for energy, whereas females actively collect oil to provision their nests.

Table 4. Table of prior probabilities of the `EVD_FlowerVisitorType` BN node.

EVD_FlowerVisitorType	Prior Probability
Nectar_Forager	0.1
Pollen_Forager	0.1
Nectar_And_Pollen_Forager	0.7
Oil_And_Nectar_And_Pollen_Forager	0.1

IPA-IFBN Evidence Mapping

A mechanism was needed to set the states of BN nodes from corresponding instances of equivalent classes in the IPA Ontology. This can be seen as linking the ontology (see Figure 4 and Table 2) to the BN (Figure 6) for each processed record (i.e. for each associated plant and arthropod specimen-record). Table 5 shows how instances of four IPA classes are mapped to four corresponding states of IFBN target nodes, through matching the object property names with node names, and instance names with state names (in bold type). The rest of the BN nodes' states are set in the same way from the IPA Ontology.

Table 5. An extract from the IPA-IFBN evidence mapping.

IPA Ontology (source class and instance name)	IFBN (*target node and state name*)	Description
hasPlantParticipant hasPlantSpecies hasFloralProduct Combination **Oil_And_Pollen**	*EVD_hasPlantParticipant_ hasPlantSpecies_ hasFloralProduct_Comb* ***Oil_And_Pollen***	Which floral product combination characterises the plant species
hasArthropodParticipant hasArthropodSpecies hasFlowerVisitorType **Oil_And_Nectar_And_Pollen _Forager**	*EVD_hasArthropodParticipant _ hasArthropodSpecies_ hasFlowerVisitorType* ***Oil_And_Nectar_And_Pollen _Forager***	The type of flower-visiting behaviour exhibited by the arthropod species
hasArthropodParticipant hasArthropodSpecies hasSpeciesSpecialisation **HighSpecialisation**	*EVD_hasArthropodParticipant _ hasArthropodSpecies_ hasSpeciesSpecialisation* ***HighSpecialisation***	Whether an arthropod species specialises (high, medium or low specialisation) in foraging for nectar from a small number of plant species.
hasArthropodParticipant hasForagingBehaviour **Oil_Foraging_Behaviour Pollen_Foraging_Behaviour**	*INF_hasArthropodParticipant_ hasForagingBehaviour* ***Oil_Foraging_Behaviour Pollen_Foraging_Behaviour***	The inferred foraging behaviour of the arthropod occurrence

4.5 Layer 3: The Interpretation Layer

IFBN-IN Mapping

In Layer 2 the IFBN is executed for each instance (associated plant and arthropod specimen-record) received from the IPA, but Layer 3 aggregates these instances. The IFBN-IN mapping performs this aggregation through a simple code routine which gives effect to the following logic. If, for a given pair of associated arthropod and plant species in the spatio-temporal envelope, the number of IFBN *ForagingBehaviour* nodes in the TRUE state is greater than n, these will be aggregated (e.g. n can be set to 1). The same applies to IFBN *PollenTransferBehaviour* nodes. The result will be either an instance of the IN:AggregatedBehaviour class or an instance of the IN:EcologicalInteraction class, depending on whether the user set the 'ecological community' criterion to False or True, respectively.

The function of the Interpretation Layer is to receive records of aggregated *ForagingBehaviour* and *PollenTransferBehaviour* from the IFBN-IN mapping, and create instances of classes in the Interaction Network Ontology representing aggregated behaviour exhibited by aggregations of organisms.

Interaction Network Ontology

The Interaction Network Ontology creates the high-level domain context and also contains classes and properties for assembling the network infrastructure. The more generalised classes of aggregated individual organisms and aggregated behaviours will be described first, followed by the more specific classes of population samples and ecological interactions, and finally the network infrastructure classes.

Aggregations of Individual Organisms

Parallel subsumption hierarchies specialise the classes `In:AggregationOfIndividualsBySpecies` and `IN:PopulationSample` (Fig. 7), for both the arthropod aggregation and the plant aggregation. The class `IN:AggregationOfIndividualsBySpecies` is defined as 'More than one individual organism of the same species', and is specialised into the class `IN:AggregationOfArthropodIndividualsBySpecies`. For example, the class `IN:ArthropodPopulationSample` is a subclass of the class `IN:AggregationOfArthropodIndividualsBySpecies`.

Fig. 7. The `IN:AggregationOfIndividualsBySpecies` and `IN:PopulationSample` classes.

Aggregated Behaviours

The names of classes have been simplified for easier reading by omitting the words 'Aggregated' and 'Behaviour' in specialised subclasses. The class `IN:AggregatedBehaviour` is specialised into a subsumption hierarchy (Fig. 8) which mirrors that of the `IPA:Behaviour` class. The classes are further restricted by a subsumption hierarchy of object properties (Fig. 9).

**Fig. 8.** The class restriction of the `IN:PlantAssociationBehaviour` class.
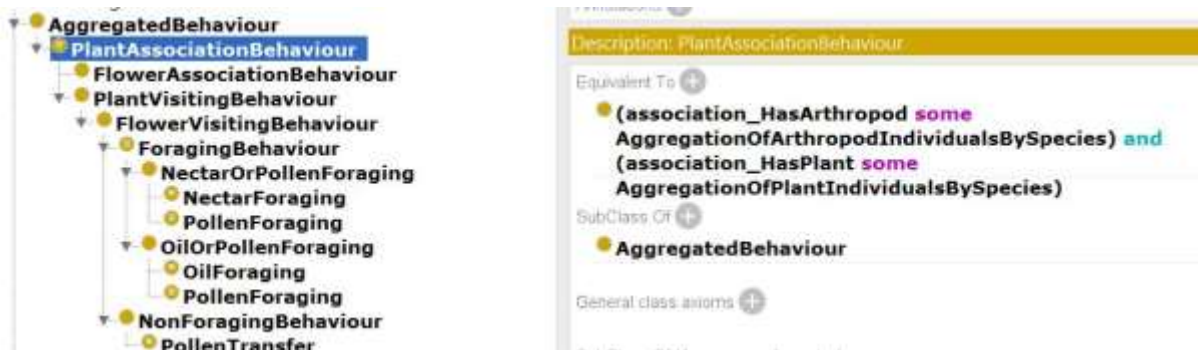
Fig. 9. The subsumption hierarchy of object properties in the IN Ontology.
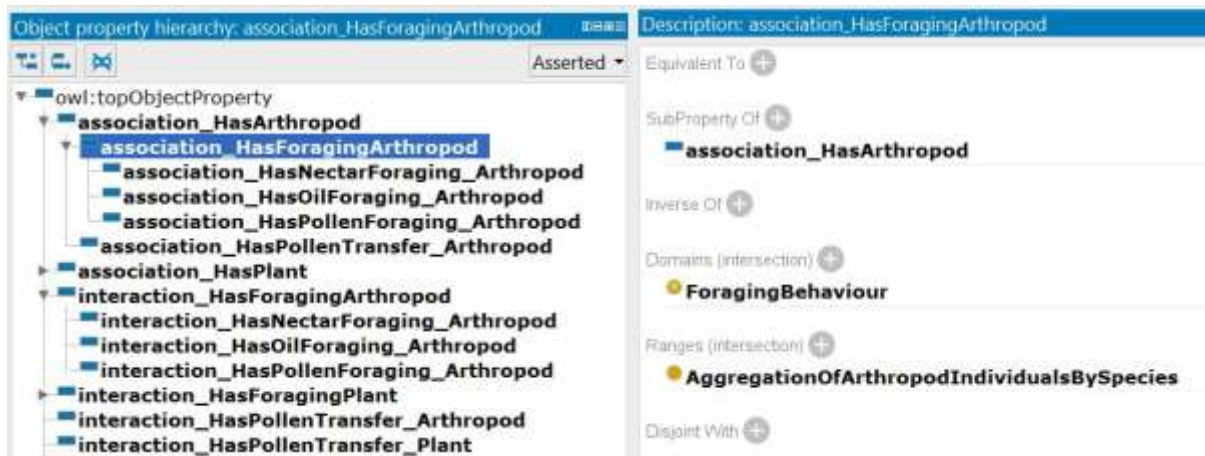
The class `IN:ForagingBehaviour` (Fig. 10) is restricted by the object property `IN:association_HasForagingArthropod`, the value of which is an instance of the class `IN:AggregationOfArthropodIndividualsBySpecies`. Similarly, Asserting the `IN:ForagingBehaviour` instance's object property `IN:association_HasForagingPlant` with a plant species as its value serves to link the `IN:ForagingBehaviour` instance to an instance of `IN:AggregationOfPlantIndividualsBySpecies` (Fig. 10)

Fig. 10. The class restriction of the `IN:ForagingBehaviour` class.
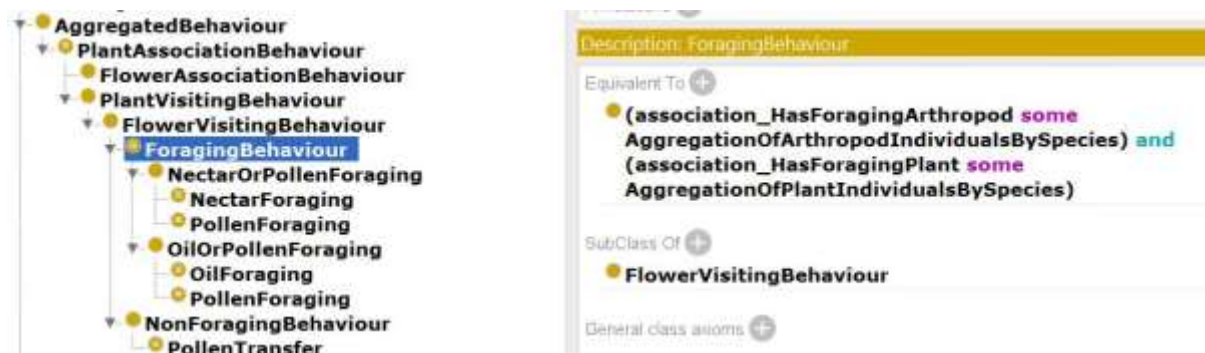
The three classes representing aggregated foraging behaviour for specific floral products are similarly restricted by object properties specific to each floral product, e.g. the class `IN:NectarForaging` and the object property `IN:association_HasNectarForaging_Arthropod`. The `IN:PollenTransfer` class (Fig. 11) is restricted by object properties which are not subsumed by a foraging object property, because pollen transfer is incidental to foraging, i.e. the properties `association_HasPollenTransfer_Arthropod` and `association_HasPollenTransfer_Plant`.

Fig. 11. The class restriction of the `IN:PollenTransfer` class.

Ecological Interactions

The class `IN:EcologicalInteraction` is specialised into a subsumption hierarchy beneath, and reflecting that of, the `IN:AggregatedBehaviour` class (Fig 12).

Fig. 12. The `IN:EcologicalInteraction` subsumption hierarchy.

Specific object properties are used to restrict the specific foraging ecological interactions, e.g. `IN:interaction_HasNectarForaging_Arthropod`.

The class restriction of the class `IN:PlantArthropodMutualisticInteraction` requires a foraging ecological interaction (which is assumed to benefit the arthropod) as well as a pollen transfer ecological interaction (which is assumed to benefit the plant) (Fig. 13).

Fig. 13. The class restriction of the `IN:PlantArthropodMutualisticEcologicalInteraction` class.

Ecological or evolutionary context

An aggregation of instances of the class `IPA:ForagingBehaviour` either becomes an instance of the `IN:AggregatedBehaviour` class or the `IN:EcologicalInteraction` class. The specialised class `IN:EcologicalInteraction` is created if the user has specified that an ecological community can be expected to occur in the specified spatio-temporal envelope. In this case an instance of the class `IN:PopulationSample` will be created (i.e. for both the plant and arthropod species), and an instance of the class `IN:EcologicalInteraction` will link the population sample instances. If the user is not modelling an ecological community, an instance of the class `IN:AggregationOfIndividualsBySpecies` will be created for the plant and arthropod species, and an instance of the class `IN:AggregatedBehaviour` will link them. As discussed below, this is the context of the evolution of flower-visiting interactions.

Network structure

The classes `IN:InteractionNetwork` and `IN:InteractionNetworkNode` are defined (Fig. 14) to associate each node with an instance of a specific network,

e.g. the population sample:

```
node3 IN:represents C_deflexum;
```

or the aggregated foraging behaviour:

```
node2 IN:represents AggNecForBehaviour10;
```

and the interaction network:

```
IN1 IN:hasNode node3.
```

This allows different instances of interaction networks to be created so that a user can visualise more than one network, to compare networks assembled within different spatio-temporal envelopes (e.g. different places at the same time or different times at the same place).

Nodes represent the class `IN:AggregationOfIndividualsBySpecies` or its subclass, `IN:PopulationSample` (of either plants or arthropods), as well as the class `IN:AggregatedBehaviour` or its subclass, `IN:EcologicalInteraction`.

Creating a network

An instance of the class `IN:AggregatedBehaviourNode` links exactly two nodes in the same interaction network, through its object properties `IN:from` and `IN:to` (Fig. 14).

Fig. 14. Instances of the class `IN:InteractionNetworkNode` are related to an instance representing the whole interaction network (class `IN:InteractionNetwork`).
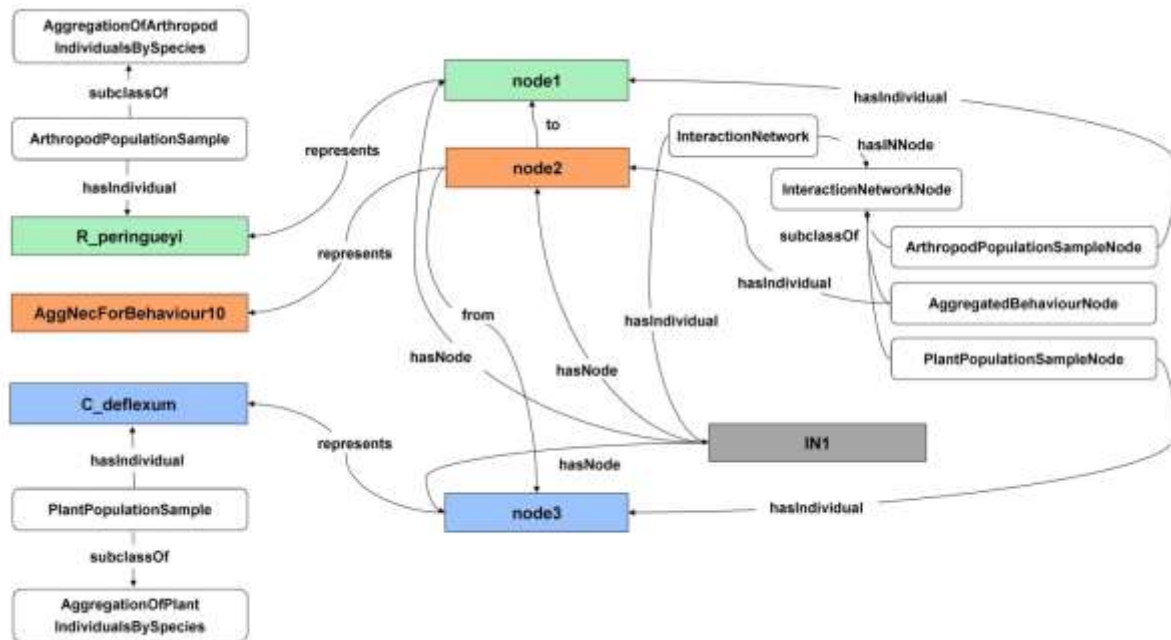
5. Implementation and evaluation

A prototype of each system component was implemented and executed to obtain system output (results) for expert verification. The Jena API was used to instantiate the ontology classes and execute the reasoner (HermiT), and the BN tools in Java (BNJ) suite was used to execute the Bayesian network (see the implementation of SWAP [52,53]). The knowledge models and parts of the code can be downloaded from the following URL: https://github.com/wcoetzer/flowervisiting.

5.1 Execution

Below we describe how the system functions to enrich, transform and aggregate flower-visiting records through each system layer.

Semantic Enrichment and Mediation Layer

This layer receives instances of the class `IPA:PlantAssociationEvent` from the data-stores' IPA Ontology mappings [54], creates instances of the classes `IPA:PlantOccurrence` and `IPA:ArthropodOccurrence`, and asserts two object properties of the `IPA:PlantAssociationEvent` instance, *viz.* `IPA:hasArthropodParticipant` and `IPA:hasPlantParticipant` (Fig. 15).

If a specific foraging behaviour was directly interpreted by the observer the record will be sent to the Situation Detection Layer but will bypass probabilistic reasoning because the IPA-IFBN mapping will set the probability of the corresponding state of the `INF_ForagingBehaviour` node to 1.

Figure 15. Event, occurrence and species instances enriched with object properties.
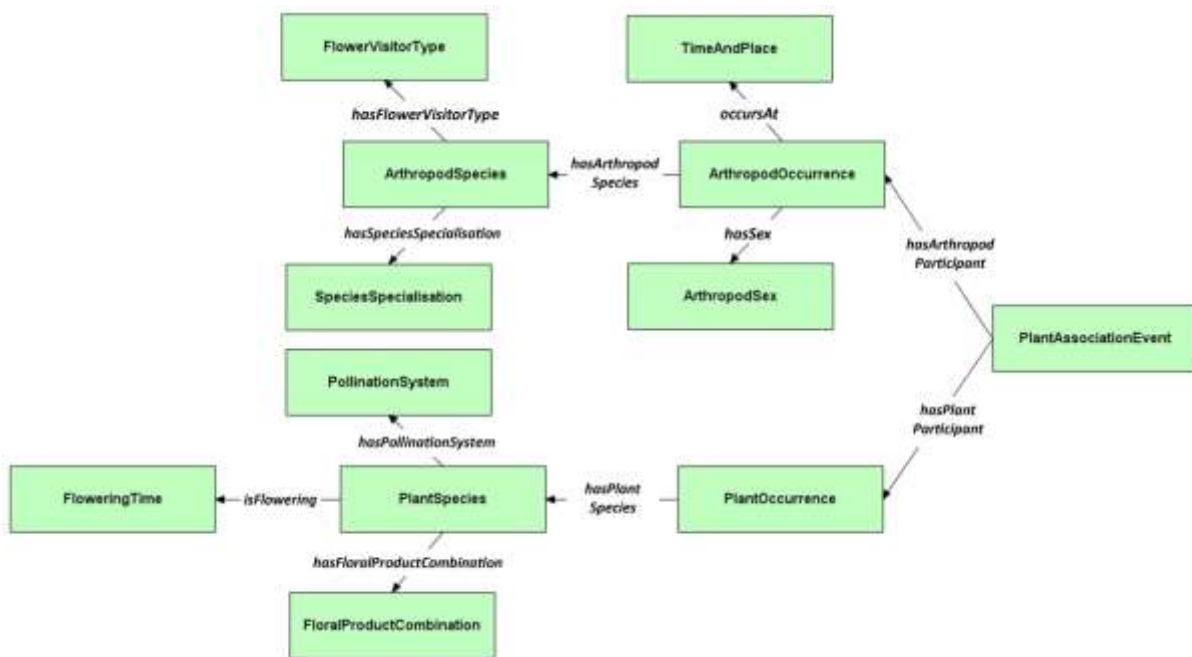
The `IPA:PlantOccurrence` instance becomes enriched with an object property (Fig. 16a) which assigns it a species name. Similarly the `IPA:ArthropodOccurrence` instance becomes enriched with object properties (Fig. 16b) which locate the instance in space and time, and assign it a species name, sex and catalogue number.

Figure 16a. Object property of the `IPA:PlantOccurrence` instance.

Figure 16b. Object properties of the `IPA:ArthropodOccurrence` instance.

Instances of the `IPA:PlantSpecies` and `IPA:ArthropodSpecies` classes are then enriched with object properties from the IPA species knowledgebase (Fig. 17a and Fig. 17b), which results in the enriched species instances shown in Fig. 15.

Figure 17a. Object properties of the `IPA:PlantSpecies` instance.

Figure 17b. Object properties of the `IPA:ArthropodSpecies` instance.

The appropriate `IPA:SpeciesSpecialisation` subclass (e.g. `HighSpecialisation`) is inferred from the family name (or in the case of the family Masaridae, the subfamily name) i.e. the value of the object property `IPA:hasFamily` or `IPA:hasSubfamily`. Second, since the `IPA:PlantSpecies` object properties `IPA:hasEarliestFloweringMonth` and `IPA:hasLatestFloweringMonth` have been asserted, the values of these can be used to infer the value of the property `IPA:isFloweringTime`.

Situation Detection Layer

Corresponding object properties in the IPA ontology are identified and their values used to set the states of the BN evidence nodes before executing the BN to calculate the posterior probabilities of states of the `INF_ForagingBehaviour` and `INF_PollenTransferBehaviour` nodes.

The IPA-IFBN evidence mapping uses the value of the `IPA:isFloweringTime` property to set the state of the `EVD_FloweringTime` IFBN node, i.e. to `True` or `False`, to allow the IFBN to infer whether or not the `IPA:PlantOccurrence` was probably flowering when the `IPA:ArthropodOccurrence` participated in the `IPA:PlantAssociationEvent`. Similarly, discrete reasoning by the IPA Ontology allows the plant sexual system and the flower visitor type to be inferred, and the corresponding IFBN nodes to be set accordingly.

Interpretation layer

The IFBN-IN mapping aggregates records received from the IFBN. For example, in the case that there is more than one record of nectar-foraging behaviour (interpreted by the IFBN) an instance of the class `IN:AggregatedNectarForagingBehaviour` is asserted, as shown in Fig 18a. In the specific case of an ecological interaction network, an instance of the class `IN:EcologicalInteraction` is asserted, as shown in Fig. 18b.

Figure 18a. An instance of the class `IN:NectarForaging` is asserted.

Figure 18b. An instance of the class `IN:NectarForagingEcologicalInteraction` is asserted.

The 'Ecological Community' criterion, received at the data input stage, determines whether the user has indicated that populations of plants and arthropods can co-exist in the specified spatio-temporal envelope. If this criterion is `True`, the specific case of a community of interacting populations applies, and the `IN:PopulationSample` instances therefore can be represented as an ecological interaction network of plant and arthropod populations (i.e. linked by instances of the class `IN:EcologicalInteraction`), or else the general case applies and the `IN:AggregationOfIndividualsBySpecies` instances must be represented as a generalised flower-visiting interaction network (i.e. the class `IN:AggregatedBehaviour` instead of the specialised `IN:EcologicalInteraction`).

An instance of the class `IN:EcologicalInteraction` is classified as an `IPA:PlantArthropodMutualisticInteraction` if both foraging and pollen transfer occurred (Fig. 19). While the inferred class `IN:PollenTransferInteraction` is not displayed in Fig. 19 due to an artefact of the Protégé application's interface design, a DL query for this class does return this instance.

Figure 19. An instance of the class

`IN:PlantArthropodMutualisticInteraction` is inferred.

5.2 Results and evaluation

Specimen-records (n=126) were spatially filtered from the database of the Iziko South African Museum. The filtered records had been created as part of a survey conducted in February and December 1986. These records were used as input into the prototype system implementation. Table 6 shows the resultant ecological interaction network.

The records were deemed to represent ecologically interacting populations of plants and arthropods, and the 'ecological community' criterion was therefore set to TRUE. The IFBN was therefore executed 126 times (for each specimen-record). The minimum number of plant or arthropod specimens/individuals in a population was 2 (IFBN-IN mapping; refer to Section 4.5).

None of the records included directly observed arthropod behaviour, and were therefore all inferred by the IFBN to represent a particular behaviour. Whereas 15 plant-arthropod species pairs were present in the data, only 10 species pairs were eventually represented by instances of the `IN:EcologicalInteraction` class. There were two reasons why such an instance was not created: either there was not more than one of either the ForagingBehaviour or PollenTransfer IFBN node, or the nodes were not interpreted to be in a foraging behaviour state - in this case male bees visiting the flowers of an oil-producing plant rather than a nectar-producing plant (males cannot use floral oil). Seven species pairs were ultimately linked through mutualistic ecological interactions (i.e. because there was both foraging and pollen transfer, and each of these benefited either the plant or arthropod population).

The interaction network model (system output) was visualised to enable experts to validate the network semantics and structure. The interaction network may be a specific ecological interaction network (e.g. Fig. 20) or a more generalised flower-visiting interaction network (with a broad spatio-temporal extent or without space and time constraints i.e. the context of evolution). In the case of a generalised flower-visiting interaction network, instances of the classes `IN:AggregationOfArthropodIndividualsBySpecies` and `IN:AggregationOfPlantIndividualsBySpecies` are asserted instead of their subclasses, viz. `IN:ArthropodPopulationSample` and `IN:PlantPopulationSample`.

Examples of ecological interaction networks were submitted to five independent experts for verification of the network semantics and structure. All five experts agreed that the networks produced by the system were semantically and structurally the same as manually constructed interaction networks, and that the semantic consistency and objectivity of the automatically interpreted networks could be useful in research. Therefore the knowledge-based system was able to infer interaction networks from flower-visiting data by formalising the specific contexts of behavioural and community flower-visiting ecology, and the evolutionary history of flower-visiting relationships.

Table 6. Results obtained from analysing 126 specimen records by using the prototype system. Each combination of plant and arthropod species is a potential ecological interaction. The sex of the arthropod is an important determinant of the kind of potential ecological interaction.

Layers 1 and 2 (Enrichment, Mediation and Interpretation)					Layer 3 (Aggregation)	
Plant Species	**Arthropod Species**	**Sex**	**ForagingBehaviour**	**Count**	**Ecological Interaction**	**Mutualistic Ecol. Int.**
Bowkeria verticillata	Rediviva rufocincta	female	OilOrPollen (47%)	12	OilOrPollenForaging / PT (88.6%)	True
Bowkeria verticillata	Rediviva rufocincta	male	(NotForagingForAFloralProduct)	16		
Buchenroedera lotonoides	Rediviva brunnea	female	NectarOrPollen (47%)	1		
Diascia anastrepta	Rediviva neliana	female	OilOrPollen (47%)	3	OilOrPollenForaging / PT (88.6%)	True
Diascia anastrepta	Rediviva pallidula	male	(NotForagingForAFloralProduct)	1		
Diascia barberae	Rediviva pallidula	female	OilOrPollen (47%)	25	OilOrPollenForaging / PT (88.6%)	True
Diascia cordata	Rediviva neliana	female	OilOrPollen (47%)	6	OilOrPollenForaging / PT (88.6%)	True
Diascia cordata	Rediviva neliana	male	(NotForagingForAFloralProduct)	2		
Diascia cordata	Rediviva pallidula	female	OilOrPollen (47%)	1		
Diascia intergerrima	Rediviva neliana	female	OilOrPollen (47%)	17	OilOrPollenForaging / PT (88.6%)	True
Diascia intergerrima	Rediviva neliana	male	(NotForagingForAFloralProduct)	5		
Diascia intergerrima	Rediviva pallidula	female	OilOrPollen (47%)	7	OilOrPollenForaging / PT (88.6%)	True
Diascia megathura	Rediviva neliana	female	OilOrPollen (47%)	8	OilOrPollenForaging / PT (88.6%)	True
Diascia megathura	Rediviva neliana	male	(NotForagingForAFloralProduct)	8		
Geranium drakensbergensis	Rediviva neliana	female	NectarOrPollen (47%)	1		
Ortholobium polystictum	Rediviva rufocincta	female	NectarOrPollen (47%)	1		
Ortholobium polystictum	Rediviva rufocincta	male	Nectar (92%)	3	NectarForaging	False
Sandersonia aurantiaca	Rediviva colorata	female	NectarOrPollen (47%)	1		
Wahlenbergia undulata	Rediviva neliana	female	NectarOrPollen (47%)	1		
Wahlenbergia undulata	Rediviva neliana	male	Nectar (92%)	4	NectarForaging	False
Wahlenbergia zeyheri	Rediviva neliana	male	Nectar (92%)	3	NectarForaging	False

Fig. 20. An ecological interaction network that visualises the results obtained in the system test described above. Green rectangles represent population samples of oil-collecting bees, blue rectangles represent population samples of oil-producing plants, and orange diamonds represent different kinds of ecological interactions. N = NectarForagingEcologicalInteraction, OP = OilOrPollenForagingEcologicalInteraction. A double diamond represents a PlantArthropodMutualisticEcologicalInteraction (meaning that a PollenTransferEcologicalInteraction also occurred).
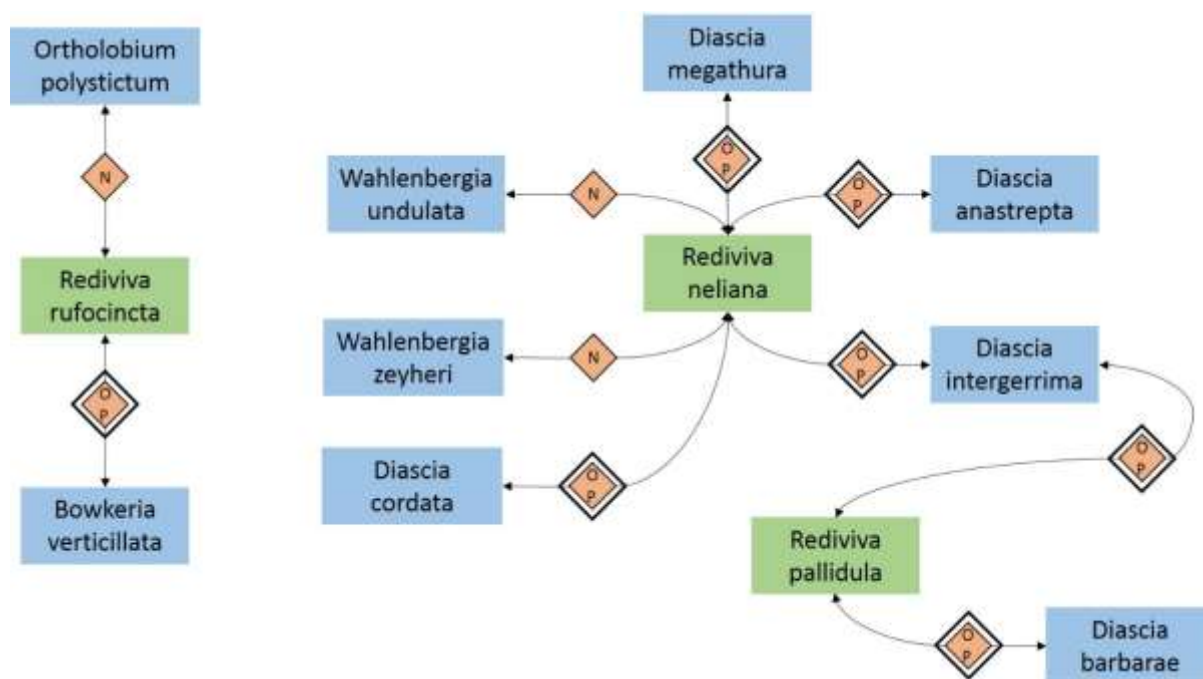
6. Discussion

We reflect on the development of the knowledge-based system and discuss the extent to which the described implementation can infer ecological interactions and interpret these as an interaction network. Limitations of knowledge representation and reasoning with ecological data are considered and emphasis is placed on the system's use of probabilistic reasoning. As an outcome of this reflection we ask: How might ecological reasoning be demonstrated differently? The potential to infer more-generalised interaction networks, and the potential impact of the work in ecology, are discussed. We also refer to related work in this under-researched area.

6.1 Development of the knowledge-based system

Ultimately the high-level context that was formalised was that of community ecology of plant-animal mutualistic interactions, and the way that this context was generated was through a specific semantic architecture that combined ontologies and a Bayesian network. The Bayesian network was not included in a previous version of the semantic architecture [54], in which records from data-stores were mapped to ontology classes which already had been interpreted by an expert (e.g. `FV:FlowerNectarIngestingEvent`). In contrast, the IPA Ontology in the described system uses a subsumption hierarchy of uninterpreted, low-

level events (i.e. the class `IPA:PlantAssociationEvent`) for semantic enrichment and mediation. This is more objective because the specific foraging behaviours of anthophilous arthropods are mostly difficult to observe directly, either because the arthropods are too small or because they fly and forage too quickly. If, however, an expert has observed a high-level behaviour directly, the system can accommodate this evidence. Otherwise, the process of high-level inferencing begins with how the (now semantically consistent) low-level events are further enriched with specific, expert causal knowledge of flower-visiting behavioural ecology. The most probable specific behaviours of individual arthropod organisms are then inferred by the Individual Flower-Visiting Behaviour Bayesian Network (IFBN). This causal knowledge and probabilistic reasoning are exemplified by two variables—the combinations of floral products offered by plant species (e.g. some plants offer pollen and oil) and the types of arthropods visiting flowers (e.g. some bees specialise in foraging for floral oil). The system was therefore able to interpret behavioural ecology data automatically, as they were found in the context of a natural history museum, and at a high level of behavioural and ecological abstraction.

The system architecture, and combination of knowledge models and formalisms of the semantic architecture, represent a first attempt to automate the interpretation of flower-visiting ecological data. The two ontologies represented relevant knowledge of plant and arthropod species, and aggregated behaviour and ecological interactions. The IFBN addressed the uncertainty inherent in all ecological data by representing qualitative, causal knowledge, and therefore may represent an advance in ecological reasoning at the individual level of organisation. There is much potential to apply probabilistic reasoning to the automated interpretation of natural history data, especially when causal behavioural knowledge about specialised classes of organisms (such as different kinds of anthophilous arthropods) can be modelled e.g. in pest control (and biological control), freshwater biomonitoring, intertidal ecology, food webs (isotope analysis) or animal movement studies [55]. Further research is needed to develop a more generalised model of behavioural ecology.

The conceptual stance of the described work did not require modelling the properties that characterise behaviour as an unfolding process. This approach may not be appropriate when adopting a different conceptual stance e.g. when population dynamics, *per se,* are important. A qualitative approach to population and community dynamics, named qualitative reasoning [24], has been used to develop a rich vocabulary describing objects, situations, relations and mechanisms of change as well as causal interpretations of system behaviour. Qualitative reasoning has been demonstrated in a causal model of predation (including consequent increases and decreases in population sizes) and in a model of the succession of a community of cerrado vegetation in Brazil. Where population dynamics of flower-visiting arthropods and plants are important, qualitative reasoning could therefore potentially be integrated into the system design.

6.2 Limitations of knowledge modelling

The important role of the IFBN leads to the question of whether the knowledge elicited from experts [55] was comprehensive enough to make the requisite inferences. An example is given to further characterise the particular combination of quality, complexity and uncertainty that is typical of ecological data, and to illustrate why the Bayesian network and probabilistic reasoning are well suited to ecological knowledge discovery. The sexual system of a plant species is generally available knowledge [61], and 90% of plants are hermaphroditic, with bisexual flowers containing both pollen and a floral reward (though in some species male and female parts mature asynchronously so as to prevent self-fertilisation). When a plant species has gender dimorphic populations (or is monoecious or dioecious) [62], however, a given plant organism, or specific flower, may either be male (and have pollen) or female (and have a floral reward e.g. oil or nectar), and unless this is recorded in data it cannot be

known. In the described application case-study the effect of this lack of knowledge probably was minimal.

The difficulty of representing high-level ecological units (i.e. the concept of a population and that of a community) was a general limitation on knowledge representation and reasoning. This is a philosophical problem: "Although there may be something that is a 'real' supraorganismal entity of nature, there is no way for us to know these entities in their reality and totality. The essence of any ecological unit thus has to be defined and cannot just be 'found' " [56]. Ecological units can be defined either by drawing discontinuities in space (topographical boundaries) or by extension of functional relationships between elements of the unit (functional boundaries), i.e. 'by what means does something become an element of a unit? Is it by virtue of its presence in a particular area, or by virtue of functional relationships with other elements of the unit?' [56]. A dual approach to the use of ecological units has been proposed [56]: 'Generic meanings of 'population', 'community', and 'ecosystem' can be retained only as heuristically useful perspectives, while specific and 'operational' definitions of the concepts as units should be developed, depending on specific purposes of their use'. This view supports our decision to delegate the delimitation (by functional boundaries) of the populations and community of interest to the user. The classes *AggregatedBehaviour* and *AggregationOfIndividualsBySpecies*, the salient classes in the system, were not encapsulated entirely within any single knowledge model. Rather, these classes were defined as the aggregation of instances of classes defined at lower levels of abstraction, and creating instances of these high-level classes therefore required traversing the whole semantic architecture. The semantic architecture was therefore a kind of knowledge model of aggregated organisms and aggregated behaviours, which represented the most important knowledge within the scope and case-study. Traversing the levels of ecological organisation, represented by the whole semantic architecture, was therefore considered a kind of ecological reasoning.

6.3 The choice of the combination of formalisms

Hunter and Liu [63] surveyed the formalisms used for representing and reasoning with scientific knowledge, including description logics, logic programming, argumentation systems, uncertainty formalisms, and systems for combining knowledge. While Bayesian networks were considered to be useful, other uncertainty formalisms, such as probabilistic logic programming, also showed potential (to combine probabilistic and logical reasoning) in the case of making statistical assertions e.g. when conducting experimental trials. In the presented work, however, the behaviour of each individual organism needed to be interpreted, so a possible-worlds approach was appropriate.

Ontologies and description logics offer a valuable approach for capturing meta-knowledge on the provenance and quality of (data as well as) knowledge in any area of science, and reasoning with this knowledge [63]. This knowledge is an important aspect of justifying a model, i.e. 'to know where the original information comes from, how it was formalized, and what conflicts and uncertainties were flagged' [5,63].

The combination of formalisms was therefore chosen because it allowed both context and causality to be modelled. In addition to describing the context of data, the Individual Plant-Arthropod Associations Ontology (IPA) was used to perform discrete reasoning in three instances, *viz.* to decide: the specialisation of an arthropod species, the sexual system of a plant species, and whether a plant species was probably flowering. For reasoning the semantic architecture relied more on the causal model (and probabilistic reasoning) than the IPA Ontology (and discrete reasoning). Ultimately this was due to the degree to which uncertainty pervaded the data and knowledge. It was also due to the facility of modelling causal knowledge using the Bayesian network formalism, compared to modelling uncertain ecological concepts and ecological causality using discrete ontology classes. A discrete model of ecological concepts would have been considerably more complex, and contained

more classes, than the Bayesian network. This facility came at no expense in knowledge representation when considering the semantic architecture as a whole, and the specific purpose of enrichment and knowledge discovery in the application case-study.

6.4 Potential for broader application in ecology

The described system was designed to discover ecological interactions between co-existing plant and arthropod populations, specifically foraging by arthropods for floral products and consequent pollen transfer. The design of the semantic architecture, however, will allow it to be used to model behavioural ecology in more general terms, and address the different scales of ecological organisation inherent in all ecological data and knowledge. The meaning of the system output potentially can be broadened to interpret other kinds of ecological interactions, e.g. parasitism and predation, from heterogeneous data (but this would require further, specific modelling, design and implementation work).

In environmental science, beyond semantic annotation of data and automatic integration of datasets, models and analytical pipelines [42,44], semantic modelling has been applied in a knowledge-driven approach [48], where 'knowledge is the key to overcoming scale and paradigm differences and to novel potential for model design and automated knowledge discovery.' In the context of distributed databases semantic modelling allows new techniques to be developed, such as model-driven query [48,64], in which a generic version of a model can be used as a constraint over a distributed knowledge base to discover new knowledge in an automated way. For example, the concept of a species-area relationship can be modelled and the model applied to distributed data to identify other potential instances of species-area relationships by finding patterns that match the model. Similarly a model of an ecological community, such as the model in the present work, could be used to discover among distributed, heterogeneous data other instances of ecological communities.

Flower-visiting interaction networks over broader spatio-temporal scales

The implemented system is already capable of distinguishing between two different contexts, i.e. an ecological interaction network and a more-generalised flower-visiting interaction network defined at a broader spatio-temporal scale. Both of these contexts have been verified through consultation with experts and reading the literature.

In community ecology, flower-visiting interaction networks belong to a class of interaction networks (including food webs) which are explicitly constrained in space and time, so that the network nodes represent real-world populations which are said to interact with each other through emergent ecological interactions. We assigned the user two input decisions relevant to the spatio-temporal scale of the interaction network under construction:

1) limiting the continuous variables of space and time that set the limits of the produced interaction network, and
2) deciding whether the supplied space and time period are small and short enough not to preclude co-existing populations.

Therefore if the user has specifically limited the input data to relate to co-existing, potentially interacting populations of plants and arthropods (e.g. occurring in a particular forest during this summer) then the interaction network produced by the system is a specific network of interacting populations (an ecological interaction network), which is analogous to an ecological community. Decisions limiting space and time 'are based on habitat borders as perceived by the researcher and knowledge about the extent of the flowering season. The method most often used is to choose a study plot of a type of vegetation and then score interactions between all flowering plant and flower-visitor species through, most often, a season' [65].

From consulting with experts and reading the literature we found that a generalised flower-visiting interaction network assembled from data collected through a broad spatial extent (including globally) and long period of time (or excluding time), which would preclude the existence of interacting populations or communities, is valid and has a different meaning. Such a flower-visiting interaction network is used to represent the evolutionary relationships of the flower-visiting mutualism between plants and insects, abstracted from ecology and studied in the light of evolutionary history.

Including other ecological interactions

Interaction networks are used in two ecological disciplines, *viz*. the study of mutualisms (including pollination and seed dispersal) and the study of food webs. Food webs are important tools in community and ecosystem ecology [36,67,68], e.g. it has been noted that 'several of the most ambitious theories in ecology describe food webs that document the structure of strong and weak trophic links, which are responsible for ecological dynamics among diverse assemblages of species' [68].

Within a community, a 'food chain' (or feeding interaction network) links species being eaten (e.g. insects) with species eating them (e.g. frogs) and species eating these (e.g. large birds) and so on. If the taxonomic species are grouped into 'trophic species' e.g. including fish as well as frogs, which 'typically eat insects' and are 'typically eaten by large birds and otters', the interaction network is a 'food web'. A food web links all the discrete food chains in a community and broadly depicts 'who eats whom'. With the exception of pollen-transfer, the ecological interactions included in the presented knowledge models are trophic interactions i.e. relevant to food or feeding, but the context of the resultant network (i.e. pollination) is not the same as that of a food web (i.e. the flow of energy). There is, however, much potential to apply the described system design to the analysis of heterogeneous data to construct more consistent food webs, or to integrate non-feeding interactions such as pollen transfer into food webs [69].

7. Conclusion

We demonstrated that by combining ontologies and a Bayesian network in a semantic architecture, expert knowledge could be represented and the manual inferences made by ecologists using implicit knowledge could be replicated and automated. Further, the results of automated interpretation were accepted by domain experts. Interpreting semantically heterogeneous flower-visiting data specifically meant inferring a standardised and consistent interaction network (a modelling construct already used in the domain), and further distinguishing between the ecological and evolutionary context of flower-visiting by arthropods. The combination of discrete and probabilistic reasoning was the key to knowledge discovery because this allowed important causal knowledge to be modelled and used in reasoning, functionality which may have been difficult to achieve otherwise. This causal knowledge was used to generate the higher-level context of community ecology.

The approach can therefore be recommended for knowledge discovery in other kinds of ecological and biodiversity data, especially when there is potential to replicate existing domain models as a way to automate data interpretation (perform automated knowledge discovery). In future work the semantic architecture could be extended to accommodate unvouchered observations, including a way to aggregate records of individuals without the risk of counting the same individual more than once. Data from flower-visiting field experiments could also be included to allow the strength of interactions (e.g. frequency of visits) or pollinator effectiveness [15] to be estimated. In an evolutionary context, network properties may be useful additions.

Interaction networks are used as tools to detect ecological and evolutionary patterns, and standardising and automating these tools could bring significant benefits to ecological research. Extension and refinement in the areas mentioned above could lead to new insights to develop techniques for ecological reasoning or ecological knowledge discovery.

Acknowledgements

References

[1]   A.-M. Klein, B.E. Vaissière, J.H. Cane, I. Steffan-Dewenter, S.A. Cunningham, C. Kremen, T. Tscharntke, Importance of pollinators in changing landscapes for world crops, Proc. R. Soc. B Biol. Sci. 274 (2007) 303–313. http://www.ncbi.nlm.nih.gov/pubmed/17164193.

[2]   P.G. Kevan, T.P. Phillips, The economic impacts of pollinator declines: An approach to assessing the consequences, Ecol. Soc. 5 (2001).

[3]   N. Gallai, J. Salles, J. Settele, B. Vaissiere, Economic valuation of the vulnerability of world agriculture confronted with pollinator decline, Ecol. Econ. 68 (2009) 810–821. doi:10.1016/j.ecolecon.2008.06.014.

[4]   A. Pauw, Collapse of a pollination web in small conservation areas., Ecology. 88 (2007) 1759–1769. http://www.ncbi.nlm.nih.gov/pubmed/17645022.

[5]   L. Vogt, eScience and the need for data standards in the life sciences: in pursuit of objectivity rather than truth, Syst. Biodivers. (2013) 1–14. doi:10.1080/14772000.2013.818588.

[6]   W.M. Hochachka, R. Caruana, D. Fink, A. Munson, M. Riedewald, D. Sorokina, S. Kelling, Data-Mining Discovery of Pattern and Process in Ecological Systems, J. Wildl. Manage. 71 (2007) 2427. doi:10.2193/2006-503.

[7]   S. Kelling, W.M. Hochachka, D. Fink, M. Riedewald, R. Caruana, G. Ballard, G. Hooker, Data-intensive Science: A New Paradigm for Biodiversity Studies, Bioscience. 59 (2009) 613–620. doi:10.1525/bio.2009.59.7.12.

[8]   P. Vera-Licona, R. Laubenbacher, Inference of Ecological Interaction Networks, Ann. Zool. Fennici. 45 (2008) 459–464. doi:10.5735/086.045.0509.

[9]   W. Zhang, Constructing ecological interaction networks by correlation analysis: hints from community sampling, Netw. Biol. 1 (2011) 81–98.

[10]   A. Aderhold, D. Husmeier, J.J. Lennon, C.M. Beale, V.A. Smith, Hierarchical Bayesian models in ecology: Reconstructing species interaction networks from non-homogeneous species abundance data, Ecol. Inform. 11 (2012) 55–64. doi:10.1016/j.ecoinf.2012.05.002.

[11]   I. Milns, C.M. Beale, V.A. Smith, Revealing ecological networks using Bayesian network inference algorithms., Ecology. 91 (2010) 1892–1899. doi:10.1890/09-0731.1.

[12]   N. Trifonova, D. Duplisea, A. Kenny, A. Tucker, A Spatio-temporal Bayesian Network Approach for Revealing Functional Ecological Networks in Fisheries, in: H. Blockeel, M. van Leeuwen, V. Vinciotti (Eds.), Adv. Intell. Data Anal. XIII 13th Int. Symp. IDA 2014, Leuven, Belgium, 30 Oct.

– 1 Novemb. 2014, Springer, 2014.

[13]   A. Faisal, F. Dondelinger, D. Husmeier, C.M. Beale, Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods, Ecol. Inform. 5 (2010) 451–464. doi:10.1016/j.ecoinf.2010.06.005.

[14]   C. Campbell, S. Yang, R. Albert, K. Shea, A network model for plant–pollinator community assembly, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 197–202. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3017189&tool=pmcentrez&rendertype=abstract.

[15]   G. Ne'eman, A. Jürgens, L. Newstrom-Lloyd, S.G. Potts, A. Dafni, A framework for comparing pollinator performance: effectiveness and efficiency, Biol. Rev. Camb. Philos. Soc. 85 (2009) 435–51. http://centaur.reading.ac.uk/16350/.

[16]   G. Ballantyne, K.C.R. Baldock, P.G. Willmer, Constructing more informative plant– pollinator networks: visitation and pollen deposition networks in a heathland plant community, Proc. R. Soc. B Biol. Sci. 282 (2015) 20151130. doi:http://dx.doi.org/10.1098/rspb.2015.1130.

[17]   C. King, G. Ballantyne, P.G. Willmer, Why flower visitation is a poor proxy for pollination: Measuring single-visit pollen deposition, with implications for pollination networks and conservation, Methods Ecol. Evol. 4 (2013) 811–818. doi:10.1111/2041-210X.12074.

[18]   J. Memmott, P.G. Craze, N.M. Waser, M. V Price, Global warming and the disruption of plant-pollinator interactions., Ecol. Lett. 10 (2007) 710–7. doi:10.1111/j.1461-0248.2007.01061.x.

[19]   J. Memmott, The structure of a plant-pollinator food web, Ecol. Lett. 2 (1999) 276–280. doi:10.1046/j.1461-0248.1999.00087.x.

[20]   T. Petanidou, A.S. Kallimanis, J. Tzanopoulos, S.P. Sgardelis, J.D. Pantis, Long-term observation of a pollination network: fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization., Ecol. Lett. 11 (2008) 564–575. http://www.ncbi.nlm.nih.gov/pubmed/18363716.

[21]   L.A. Burkle, R. Alarcón, The future of plant-pollinator diversity: understanding interaction networks across time, space, and global change., Am. J. Bot. 98 (2011) 528–38. doi:10.3732/ajb.1000391.

[22]   L.A. Burkle, J.C. Marlin, T.M. Knight, Plant-pollinator interactions over 120 years: loss of species, co-occurrence and function, Science. 339 (2013) 1611–1615.

[23]   Y.L. Dupont, B. Padrón, J.M. Olesen, T. Petanidou, Spatio-temporal variation in the structure of pollination networks, Oikos. 118 (2009) 1261–1269. doi:10.1111/j.1600-0706.2009.17594.x.

[24]   P. Salles, B. Bredeweg, Qualitative reasoning about population and community ecology, Ai Mag. 24 (2004) 77–90.

[25]   E.J. Rykiel, Artificial intelligence and expert systems in ecology and natural resource management, Ecol. Modell. 46 (1989) 3–8.

[26]   H.M. Regan, M. Colyvan, M.A. Burgman, A taxonomy and treatment of uncertainty for ecology and conservation biology, Ecol. Appl. 12 (2002) 618–628. doi:10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2.

[27]   T. Nuttle, B. Bredeweg, P. Salles, M. Neumann, Representing and managing uncertainty in qualitative ecological models, Ecol. Inform. 4 (2009) 358–366.

doi:10.1016/j.ecoinf.2009.09.004.

[28] R.L. Walls, J. Deck, R. Guralnick, S. Baskauf, R. Beaman, S. Blum, S. Bowers, P.L. Buttigieg, N. Davies, D. Endresen, M.A. Gandolfo, R. Hanner, A. Janning, L. Krishtalka, A. Matsunaga, P. Midford, N. Morrison, E. O´ Tuama, M. Schildhauer, B. Smith, B.J. Stucky, A. Thomer, J. Wieczorek, J. Whitacre, J. Wooley, Semantics in support of biodiversity knowledge discovery: An introduction to the Biological Collections Ontology and related ontologies, PLoS One. 9 (2014) e89606. doi:10.1371/journal.pone.0089606.

[29] G. V. Gkoutos, P.N. Schofield, R. Hoehndorf, The Neurobehavior Ontology: An ontology for annotation and integration of behavior and behavioral phenotypes, Int. Rev. Neurobiol. 103 (2012) 69–87.

[30] G. V. Gkoutos, R. Hoehndorf, L. Tsaprouni, P.N. Schofield, Best behaviour? Ontologies and the formal description of animal behaviour, Mamm. Genome. 26 (2015) 540–547. doi:10.1007/s00335-015-9590-y.

[31] P.E. Midford, Ontologies for behavior, Bioinformatics. 20 (2004) 3700–3701. doi:10.1093/bioinformatics/bth433.

[32] C. Mungall, Social Insect Behavior Ontology, (n.d.). http://www.obofoundry.org/ontology/sibo.html.

[33] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium., Nat. Genet. 25 (2000) 25–29. http://www.ncbi.nlm.nih.gov/pubmed/10802651.

[34] R. Arp, B. Smith, Function, role, and disposition in Basic Formal Ontology, Nat. Preced. 1941.1 (2008) 1–4. doi:10.1038/npre.2008.1941.1.

[35] P.L. Buttigieg, E. Pafilis, S.E. Lewis, M.P. Schildhauer, R.L. Walls, C.J. Mungall, The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation, J. Biomed. Semantics. 7 (2016) 57. doi:10.1186/s13326-016-0097-6.

[36] J.H. Poelen, J.D. Simons, C.J. Mungall, Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets, Ecol. Inform. 24 (2014) 148–159. doi:10.1016/j.ecoinf.2014.08.005.

[37] Definition of "multi-organism behaviour" class, (n.d.). http://www.ontobee.org/ontology/GO?iri=http://purl.obolibrary.org/obo/GO_0051705.

[38] R.J. Williams, N.D. Martinez, J. Golbeck, Ontologies for ecoinformatics, Web Semant. Sci. Serv. Agents World Wide Web. 4 (2006) 237–276.

[39] M. Keet, Factors affecting ontology development in ecology, in: Data Integr. Life Sci. Second Int. Work. DILS 2005, San Diego, CA, USA, July 20-22, 2005: pp. 46–62.

[40] D.D. Pennington, I.N. Athanasiadis, S. Bowers, S. Krivov, J. Madin, M. Schildhauer, F. Villa, Indirectly driven knowledge modelling in ecology, Int. J. Metadata, Semant. Ontol. 3 (2008) 210–225. doi:10.1504/IJMSO.2008.023569.

[41] W. Michener, J.H. Beach, M.B. Jones, B. Ludäscher, D.D. Pennington, R.S. Pereira, A. Rajasekar, M. Schildhauer, A knowledge environment for the biodiversity and ecological sciences, J. Intell. Inf. Syst. 29 (2007) 111–126. doi:10.1007/s10844-006-0034-8.

[42]   W. Michener, M.B. Jones, Ecoinformatics: supporting ecology as a data-intensive science, Trends Ecol. Evol. 27 (2012) 85–93. doi:10.1016/j.tree.2011.11.016.

[43]   J.S. Madin, S. Bowers, M.P. Schildhauer, M.B. Jones, Advancing ecological research with ontologies., Trends Ecol. Evol. 23 (2008) 159–68. doi:10.1016/j.tree.2007.11.007.

[44]   B. Leinfelder, S. Bowers, M. O'Brien, M.B. Jones, M. Schildhauer, Using semantic metadata for discovery and integration of heterogeneous ecological data, Proc. Environ. Inf. Manag. Conf. EIM 2011. (2011) 1–6. https://semtools.ecoinformatics.org/repository/docs/pubs/EIM-2011/main.pdf (accessed October 10, 2014).

[45]   J.S. Madin, S. Bowers, M. Schildhauer, S. Krivov, D. Pennington, F. Villa, An ontology for describing and synthesizing ecological observation data, Ecol. Inform. 2 (2007) 279–296. doi:10.1016/j.ecoinf.2007.05.004.

[46]   C. Bizer, T. Heath, T. Berners-Lee, Linked Data - The Story So Far, Int. J. Semant. Web Inf. Syst. 5 (2009) 1–22. doi:10.4018/jswis.2009081901.

[47]   V. Brilhante, An ontology for quantities in ecology, in: Proc. Brazilian Symp. Artif. Intell. Lect. Notes Artif. Intell. 3171, Springer Berlin / Heidelberg, 2004: pp. 144–153.

[48]   F. Villa, I. Athanasiadis, A. Rizzoli, Modelling with knowledge: A review of emerging semantic approaches to environmental modelling, Environ. Model. Softw. 24 (2009) 577–587. doi:10.1016/j.envsoft.2008.09.009.

[49]   E. Charniak, Bayesian Networks without Tears, AI Mag. 12 (1991) 50. doi:10.1609/aimag.v12i4.918.

[50]   R.E. Neapolitan, Learning Bayesian Networks, Mol. Biol. 6 (2003) 674. http://www.amazon.com/Learning-Bayesian-Networks-Richard-Neapolitan/dp/0130125342.

[51]   R.K. McCann, B.G. Marcot, R. Ellis, Bayesian belief networks: applications in ecology and natural resource management, Can. J. For. Res. 36 (2006) 3053–3062. doi:10.1139/x06-238.

[52]   D. Moodley, I. Simonis, J. Tapamo, An architecture for managing knowledge and system dynamism in the worldwide Sensor Web. International Journal of Semantic Web and Information Systems: Special issue on Semantics-enhanced Sensor Networks, Int. J. Semant. Web Inf. Syst. 8 (2012) 64–88.

[53]   D. Moodley, Ontology Driven Multi-agent Systems: An Architecture for Sensor Web Applications, University of KwaZulu-Natal, 2009.

[54]   W. Coetzer, D. Moodley, A. Gerber, A knowledge-based system for discovering ecological interactions in biodiversity data-stores of heterogeneous specimen-records: A case-study of flower-visiting ecology, Ecol. Inform. 24 (2014) 47–59. doi:http://dx.doi.org/10.1016/j.ecoinf.2014.06.008.

[55]   W. Coetzer, D. Moodley, A. Gerber, Eliciting and Representing High-Level Knowledge Requirements to Discover Ecological Knowledge in Flower-Visiting Data, PLoS One. 11 (2016) e0166559. doi:10.1371/journal.pone.0166559.

[56]   K. Jax, Ecological units: definitions and application., Q. Rev. Biol. 81 (2006) 237–258. doi:10.1086/506237.

[57]   M. Huston, D. DeAngelis, W. Post, New computer models unify ecological theory, Bioscience. 38 (1988) 682–691. http://www.jstor.org/stable/1310870.

[58] S. Baskauf, C. Webb, Darwin-SW: Darwin Core-based terms for expressing biodiversity data as RDF, Semant. Web – Interoperability, Usability, Appl. 1213–2425 (2014). doi:10.3233/SW-150203.

[59] M. Horridge, A practical guide to building OWL ontologies using Protege 4 and CO-ODE Tools, Edition 1.3, (2011).

[60] M. Uschold, M. Gruninger, Ontologies: principles, methods and applications, Knowl. Eng. Rev. 11 (1996) 1–63.

[61] S.S. Renner, The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database, Am. J. Bot. 101 (2014) 1588 – 1596.

[62] S.C.H. Barrett, The evolution of plant sexual diversity., Nat. Rev. Genet. 3 (2002) 274–284. doi:10.1038/nrg776.

[63] A. Hunter, W. Liu, A survey of formalisms for representing and reasoning with scientific knowledge, Knowl. Eng. Rev. 25 (2010) 199–222. doi:10.1017/S0269888910000019.

[64] F. Villa, A semantic framework and software design to enable the transparent integration, reorganization and discovery of natural systems knowledge, J. Intell. Inf. Syst. 29 (2007) 79–96. doi:10.1007/s10844-006-0032-x.

[65] J.M. Olesen, Y.L. Dupont, M. Hagen, C. Rasmussen, K. Trojelsgaard, Structure and dynamics of pollination networks: the past, present, and future, in: S. Patiny (Ed.), Evol. Plant-Pollinator Relationships, Cambridge University Press, London, 2012: pp. 374–391.

[66] J. Bascompte, P. Jordano, J.M.J. Olesen, Asymmetric coevolutionary networks facilitate biodiversity maintenance, Science. 312 (2006) 431–433. doi:10.1126/science.1123412.

[67] C.S. Parr, B. Lee, B.B. Bederson, EcoLens: Integration and interactive visualization of ecological datasets, Ecol. Inform. 2 (2007) 61–69. doi:10.1016/j.ecoinf.2007.03.005.

[68] R.J. Williams, N.D. Martinez, Simple rules yield complex food webs., Nature. 404 (2000) 180–183. doi:10.1038/35004572.

[69] S. Kéfi, E.L. Berlow, E.A. Wieters, S.A. Navarrete, O.L. Petchey, S.A. Wood, A. Boit, L.N. Joppa, K.D. Lafferty, R.J. Williams, N.D. Martinez, B.A. Menge, C.A. Blanchette, A.C. Iles, U. Brose, More than a meal... integrating non-feeding interactions into food webs, Ecol. Lett. 15 (2012) 291–300. doi:10.1111/j.1461-0248.2011.01732.x.