Ontologies for Probabilistic Networks

Eveline M. Helsper

Linda C. van der Gaag

institute of information and computing sciences, utrecht university

technical report UU-CS-2003-042

www.cs.uu.nl

Ontologies for Probabilistic Networks

Eveline M. Helsper and Linda C. van der Gaag
Institute of Information and Computing Sciences, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
{eveline,linda}@cs.uu.nl

Abstract. Building a probabilistic network for a real-life domain of application is a hard and time-consuming process, which is generally performed with the help of domain experts. As the scope and, hence, the size and complexity of networks are increasing, the need for proper documentation of the elicited domain knowledge becomes apparent. To study the usefulness of ontologies for this purpose, we constructed an ontology for the domain of oesophageal cancer, based upon a real-life probabilistic network for the staging of cancer of the oesophagus and the knowledge elicited for its construction. In this paper, we describe the various components of our ontology and outline the benefits of using ontologies in engineering probabilistic networks.

1 Introduction

More and more knowledge-based systems build upon the formalism of *probabilistic networks* for their knowledge representation. A probabilistic network consists of a graphical structure, encoding the important statistical variables from the domain of application along with the influential relationships between them, and an associated numerical part, encoding a joint probability distribution over the represented variables [1]. The suitability of the formalism for capturing complex problem domains and their uncertainties is demonstrated by an increasing number of successful applications in for example medical diagnosis and prognosis, information retrieval, and weather forecasting.

Building a probabilistic network for a real-life domain of application is generally considered a hard and time-consuming process. The process involves three basic tasks [2]. The first of these is to identify the statistical variables that are of importance in the domain, along with their possible values. Once the important variables have been identified, the second task is to identify the relationships between them. The variables and relations are then expressed in a graphical structure. The last task is to obtain the probabilities that are required for the network's numerical part. The various tasks are typically performed with the help of domain experts, from whom the knowledge required is elicited in a series of interviews.

In building a probabilistic network, numerous modelling and design decisions are taken [3]. Many of these decisions originate from a trade-off between the desire for a rich model on the one hand and the costs of construction and maintenance on the other hand. Generally, a rich model is preferred, that properly reflects the intricacies in the domain of application and as a consequence is likely to be accepted by the domain's experts. A richer model further is often expected to have a better performance. Representing additional knowledge and adding more details, however, will often increase the time and, hence, the costs of construction and maintenance. A highly detailed network, for example, tends to require many probabilities that may be hard to come by and as a consequence may prove to be a source of unacceptable inaccuracy. Other design decisions are enforced by the formalism of probabilistic networks itself. For example, since a probabilistic network in essence is a model of a joint probability distribution, multi-valued domain concepts must be modelled as statistical variables, which are single-valued by definition. As a result of these decisions, the knowledge from the domain experts may not always be recognizable from the resulting probabilistic network. Moreover, a considerable amount of background knowledge may not have been included in the network at all. For a small-scaled network that is developed in a laboratory setting by a single knowledge engineer, the elicited domain knowledge is readily shared between the experts and engineer, and the various modelling and design decisions can easily be reconstructed. We have experienced, however, that larger probabilistic networks that are being developed over various years involving different engineers, may become inaccessible to those who have not been involved in their construction from the very beginning. Construction and maintenance in fact can be seriously hampered if the elicited domain knowledge itself is not made explicit by proper documentation.

As the scope and, hence, the complexity and size of real-life probabilistic networks are increasing, researchers are beginning to look into methodologies for engineering network-based systems. K.B. Laskey and S.M. Mahoney, for example, have advocated an overall systems-engineering approach to building probabilistic networks [4, 5]. More specifically, they propose the spiral life-cycle model as the most appropriate approach. According to this model, systems engineering is a repeating cycle of system design, development, operation, and evaluation. The alternative waterfall model resembles the spiral model, but has a more linear nature. The cyclic approach is preferred, since engineering a probabilistic network is not a process during which an already existing model of knowledge is extracted from the domain expert; it is basically a process of discovering a model of knowledge, during which the understanding of both the knowledge engineer and the domain expert evolves over time.

Within the overall systems-engineering approach, often the use of techniques from the field of knowledge engineering is suggested for actually constructing a probabilistic network for the domain at hand. For engineering knowledge-based systems in general, various sophisticated methodologies are available. A well-known example is the CommonKADS methodology [6], that has been designed to fit in with a life-cycle approach. Knowledge-engineering methodologies generally advise to first capture the knowledge from a domain in a knowledge model, before actually developing the system. Such a model then contains the knowledge that is to be incorporated in the system and typically also includes the additional background knowledge that is necessary for understanding the domain under study. The model is used in acquiring (additional) domain knowledge, in specifying the design of the system, and for maintenance. The knowledge in the model is represented independently from any tool or implementation language, to avoid biases. Developing the system then includes selecting a representation formalism and expressing the knowledge from the model in the selected formalism. In building probabilistic networks, however, it is common practice to express the elicited domain knowledge into the network formalism directly, thereby side-stepping the construction of an explicit knowledge model. Various approaches to expressing elicited knowledge into the network formalism have been proposed, which include the application of network fragments [7] and of the concept of object orientation [8].

As we have argued above, we feel that explicit documentation of elicited domain knowledge is necessary to facilitate a shared understanding of the knowledge among the experts and engineers involved in building and maintaining a probabilistic network. We therefore advocate the construction of an explicit knowledge model. In this paper, we investigate whether ontologies can be used for this purpose. In the field of knowledge-based systems, the term *ontology* is used to denote an explicit specification of shared domain knowledge. An ontology typically describes the domain knowledge and its structure in terms of concepts and the various different types of relation between them. The content of the ontology is understood to be agreed upon by all agents involved. These agents can be software agents that operate on the same knowledge [9], or the people involved in the construction of a knowledge-based system [10]. The ontology serves to allow the agents to communicate about the domain knowledge without any misconceptions. Ontologies have been used in knowledge engineering before. In fact, our idea of using an ontology for explicit documentation of elicited domain knowledge has been motivated by G. van Heijst et al. [11], who exploit ontologies as (part of) a knowledge model.

An ontology that is constructed for our purpose of knowledge sharing and documentation contains a body of knowledge that is relevant in a domain of application and can also be used as a knowledge model as outlined above. More specifically, the ontology can be used to derive the graphical structure of a probabilistic network. In [12] we present an initial study to this end. Some of the approaches that are currently in use for constructing a network directly from elicited domain knowledge can in fact also be exploited in deriving a network from the ontology. The explicit separation of the process of knowledge acquisition from the design of the network enables the knowledge engineer to focus on the acquisition issues and the design issues separately.

To study the usefulness of ontologies in engineering probabilistic networks, we constructed an ontology for the domain of oesophageal cancer. We had developed before, over a period of more than five years, a real-life network for the staging of cancer of the oesophagus. We now created an ontology for the same domain through reverse engineering of this network and building upon the knowledge that had been elicited for its construction. In constructing the ontology, we noticed that a considerable amount of background knowledge underlies the network, which rendered it practically inaccessible to anyone other than the domain experts and the knowledge engineer involved. We found that our ontology serves to make this background knowledge explicit.

In this paper, we describe the oesophagus ontology and its construction. It is not our intention to introduce a new methodology for developing ontologies or for specifying their components. Our main goal is to propose the use of ontologies in engineering probabilistic networks. The paper is organised as follows. Section 2 briefly introduces the oesophagus network. In Section 3, we use ontologies for making the knowledge in our domain of application explicit, both at different levels of abstraction and from different perspectives. In Section 4, we elaborate on the benefits that can be expected from constructing and maintaining ontologies in engineering probabilistic networks in general. The paper ends with our concluding observations and directions for further research in Section 5.

2 The oesophagus network

As a consequence of a lesion of the oesophageal wall or associated with smoking and drinking habits, a tumour may develop in the oesophagus. A primary tumour of the oesophagus has various characteristics that influence its prospective growth. These characteristics include the tumour's location in the oesophagus, its length, and its macroscopic shape. The tumour typically invades the oesophageal wall and upon further growth may affect such neighbouring organs as the trachea and bronchi or the diaphragm, dependent upon its location in the oesophagus. In time, the tumour may result in secondary tumours, or metastases, in lymph nodes and in other organs, such as the liver and the lungs. A distinction is made between lymphatic metastases and haematogenous metastases that result from transference of cancer cells via the lymph vessels and via the blood vessels, respectively. The depth of invasion and the extent of metastasis, summarised in the cancer's stage, largely influence a patient's life expectancy and are indicative of the effects and complications to be expected from the different available therapeutic alternatives.

Every year some eighty patients receive treatment for oesophageal cancer at the Antoni van Leeuwenhoekhuis of the Netherlands Cancer Institute. These patients are assigned to a therapy by means of a standard protocol that includes a small number of prognostic factors. Based upon this protocol, 75% of the patients show a favourable response to the therapy provided; one out of every four patients, however, develops more or less serious complications as a result of the therapy. To arrive at a more fine-grained protocol with a more favourable response rate, a knowledge-based system is being developed for patient-specific therapy selection. The kernel of our system is a probabilistic network that captures the state-of-the-art knowledge about cancer of the oesophagus. The graphical structure of the part of the oesophagus network that pertains to the staging of a patient's cancer is depicted in Fig. 1; the figure also shows the prior probability distribution for each of the variables.

The oesophagus network was constructed and refined with the help of two experts in gastrointestinal oncology from the Netherlands Cancer Institute. The network currently includes over 80 statistical variables, for which some 4000 probabilities have been specified. In a sequence of eleven interviews of two to four hours each over a period of two years, the experts identified the relevant diagnostic and prognostic factors to be captured as statistical variables in the network and the relationships between them. The elicitation of the probabilities took five interviews of approximately two hours each over a period of fifteen months [13].

3 An ontology for oesophageal cancer

To study the value of ontologies in engineering probabilistic networks, we constructed an ontology for the domain of oesophageal cancer. In this section, we describe our ontology in some detail, elaborating on the various methodological issues that we encountered.

We based our oesophagus ontology on the probabilistic network described in the previous section and on the knowledge that had been elicited during the network's construction. We decided not to arrange additional knowledge-acquisition sessions to construct the ontology together with our domain experts. The experts were already quite familiar with the graphical structure of the network and might be biased by it when describing the knowledge of their domain for the purpose of the ontology. Moreover, since knowledge-acquisition sessions generally are very demanding of experts, we preferred not to ask them to re-perform a task they had, roughly speaking, performed before. For similar reasons, we decided not to validate the ontology against our experts.

We would like to note that the construction of the oesophagus ontology has been far from straightforward, even given the available knowledge in the domain of application. Although we could build upon a wealth of literature on ontologies, we still had to resolve numerous issues. Also it quickly became apparent that a considerable amount of background knowledge underlies the oesophagus network. Since one of the authors had engineered the network and the other author, who had no prior knowledge of the domain, constructed the ontology, all background knowledge had to be shared between the two authors.

In Section 3.1, we discuss a number of general criteria for ontologies that are constructed for the purpose of knowledge sharing between knowledge engineers and domain experts. In Section 3.2, the implications of these criteria for the oesophagus ontology are discussed. Sections 3.3 through 3.6 describe the different components of our ontology. The various properties of an ontology depend on the goal for which the ontology is to be used and on the domain of application. The goal of the oesophagus ontology as well as the nature of the domain are therefore reflected in our ontology. In Section 3.7 we briefly review some other approaches to specifying ontologies.

3.1 Ontologies for knowledge sharing and documentation

We recall that our main purpose of developing an ontology for a probabilistic network under construction is to help the knowledge engineers involved in understanding the intricacies of the domain and to facilitate communication between the engineers and the domain experts.

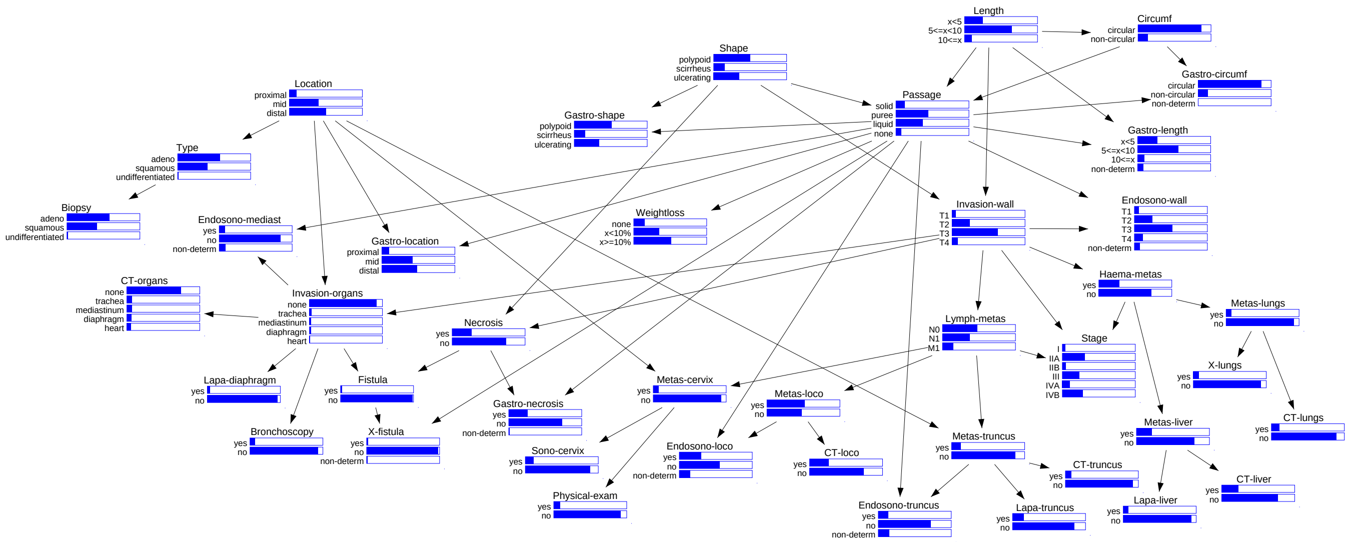

Figure 1: The oesophagus network

Th.R. Gruber identifies five criteria for an ontology that has knowledge sharing for its main purpose [9]:

- *Clarity*: the meanings of the concepts and relations that are specified in the ontology must be clear.
- *Extendibility*: the ontology should be easy to extend and maintain, without inducing the need to change major parts of the ontology.
- *Coherence*: the ontology should be internally coherent.
- *Minimal encoding bias*: the representation language used for expressing the domain knowledge should introduce as little bias as possible.
- *Minimal ontological commitment*: the ontology should be based on as few assumptions as possible about how the ontology, and the knowledge it contains, will be used.

We adopt these criteria for our oesophagus ontology and describe the implications of three of these criteria for ontologies that are built for the purpose of knowledge sharing and documentation in general. The more detailed implications of the five criteria depend on the domain of application and on the people involved. We elaborate on these implications for our oesophagus ontology in Section 3.2.

According to the *clarity* criterion, the meanings of the concepts and relations in the ontology must be clear to both knowledge engineers and domain experts. To meet this criterion, the level of detail in which the domain knowledge is specified, needs to be chosen with care. It needs to be tuned to the goal of the ontology and to the engineers and experts involved. On the one hand, the domain knowledge should be represented in sufficient detail to avoid multiple interpretations and lack of understanding; all relevant intricacies should be captured. On the other hand, representing the knowledge in too much detail may hamper the understandability and transparency.

The criterion of *minimal encoding bias* states that the language used in the ontology for representing the domain knowledge should introduce as little bias as possible in the contents and structure of the knowledge. If the criterion is not met, the ontology may capture a biased view of the domain. Since the ontology is used for the construction of a knowledge-based system, the resulting system may be biased as well, possibly even in an unforeseen way. The development of an independent knowledge model, recommended by most knowledge-engineering methodologies, in fact has its origin in this observation.

In our opinion, the language of probabilistic networks fails to meet the criterion of minimal encoding bias. Using this language may cause loss of information because of the necessity to capture all domain knowledge in terms of statistical variables and independence relations. The language of probabilistic networks, for example, does not allow for explicitly distinguishing between the different natures of the relations that exist between the various concepts. In our domain, for example, the concepts of *secondary tumour* and *tumour* are connected by a hierarchical *is-a* relation. The process of metastasis via the blood vessels and the metastases in the liver, on the other hand, are connected by a relation that involves time and causality. Both relations are represented in a probabilistic network by dependences that are captured by means of unlabeled arcs. Also, multi-valued domain concepts should be modelled as statistical variables which are single-valued by definition. Moreover, for the purpose of knowledge sharing, the clarity criterion implies that the ontology should be represented in a language that is readable and understandable for both the knowledge engineers and the domain experts involved. The formalism of probabilistic networks is often considered to be suitable for this purpose: since the (qualitative) knowledge is represented by a graphical structure, the knowledge is intuitively clear to the readers of the network, the domain experts included [1]. We experienced, however, that experts who are not familiar with the formalism nor with the concept of probabilistic independence, often tend to misinterpret the graphical structure. For example, a medical diagnostic probabilistic network will typically contain arcs from a disease to the various symptoms of the disease. A physician may be inclined to reverse these arcs, since she often reasons from symptoms to disease when establishing a diagnosis for a patient. A network containing the reversed arcs, however, may not correctly represent the independences that hold in the domain of application. Based upon these observations, we feel that an ontology that is expressed as a probabilistic network cannot in general serve as a means of communication between domain experts and knowledge engineers, mainly because the conceptual distance between the ontology and the way the experts think and talk about their domain would be too large.

For expressing the knowledge from a domain of application in an ontology, therefore, a more suitable representation language should be chosen. The issue of selecting an appropriate representation language has been addressed by many researchers. Some suggest that domain knowledge should be represented by a language that is highly informal, semi-informal, or semi-formal [10]; others argue that ontologies should be specified in a rigorously formal language and, in fact, should be machine readable [14]. The level of formality is best tuned to the properties of the domain knowledge, to the people involved, and to the goal of the ontology. If the use of a formal language is uncommon in a domain of application, for example, then a formal language may be less suitable for the purpose of knowledge sharing between the knowledge engineers and the domain experts in the domain at hand.

The criterion of *minimal ontological commitment*, to conclude, states that the ontology should be developed independently of the projected use of the ontology and its contents. Any commitment to the problem-solving method that will be applied to the domain knowledge, for example, will influence and thereby bias the way the knowledge is captured in the ontology [15]. Such commitments may therefore hamper the extendibility and reuse of the ontology. Commitments should not be avoided, however, as they serve to delimit the domain knowledge to be captured. When made explicit, commitments in fact enhance the clarity of the ontology and thereby facilitate the knowledge engineers to keep the ontology consistent upon maintenance.

3.2 An overview of the oesophagus ontology

For the oesophagus ontology, we selected a language that we considered suitable for representing the knowledge in our domain of application, and that would also be understandable to both the knowledge engineers and the experts in our domain, thereby allowing them to share the domain's knowledge. We chose a semi-formal representation language, including tables, graphs, depictions and natural language. We feel that a more formal representation language would be less suitable for our goals, because the use of formal languages is quite uncommon in our domain of application.

As discussed before, ontological commitments may bias the represented knowledge and its structure, yet also serve to restrict the scope of an ontology. For our oesophagus ontology we decided to explicitly commit to our application in the sense that a patient is assumed to actually have oesophageal cancer. As the Netherlands Cancer Institute is a specialised centre for cancer treatment, this assumption seems reasonable. In fact, the same assumption also underlies the oesophagus network. As a consequence of our ontological commitment, all knowledge involved in determining the stage of a patient's cancer and in establishing a prognosis, is included; knowledge about the etiology, that is, about the conception of the cancer, is not included, since this knowledge is not needed in the projected use of the ontology.

To meet the clarity criterion, the agreed-upon interpretation of the domain's terminology is explicitly captured in the oesophagus ontology. For this purpose the ontology contains a *glossary*, listing the names of the relevant concepts along with their meaning. This glossary basically serves to avoid confusion and ambiguity of terms; it is discussed in section 3.3. To provide for extendibility, the oesophagus ontology contains separate components specifying different types of knowledge from different perspectives. We distinguish between two perspectives. The *static perspective* on the domain knowledge mainly addresses the organisation of concepts in hierarchies. The *dynamic perspective* focuses on the relations between the concepts in which time plays an important role. Our main purpose of specifying the knowledge from different perspectives is to enable a knowledge engineer to concentrate on a particular component and to modify it independently of another component. Sections 3.4 and 3.5 discuss the static perspective and the dynamic perspective, respectively. To ensure that no internal inconsistencies will arise in the ontology upon maintenance, it includes various *coherence con-*

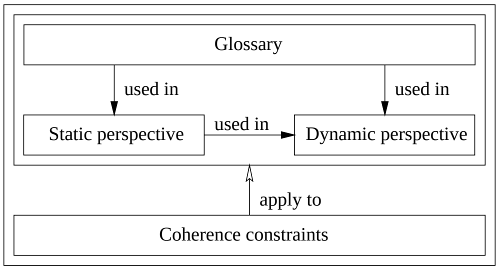

Figure 2: The components of the oesophagus ontology

straints. These constraints specify requirements that the represented knowledge should adhere to, and thus are instrumental in meeting the coherence criterion. Section 3.6 discusses the coherence constraints in more detail. The glossary, the static and dynamic perspectives, and the coherence constraints constitute the four components of the oesophagus ontology, which is schematically depicted in Fig. 2.

To conclude our overview of the oesophagus ontology, we observe that the knowledge for dealing with an instance in a domain of application generally comprises both declarative knowledge, including causal relations and hierarchical relations between concepts, and procedural knowledge that pertains to performing a specific task, such as establishing a diagnosis or a prognosis. An ontology typically captures the declarative knowledge and therefore is composed of concepts and relations between concepts. In this paper, we thus focus on the declarative knowledge in the domain of oesophageal cancer.

3.3 The glossary

The *glossary* of the oesophagus ontology specifies the names of the concepts in the domain of application along with a description of their meaning. All relevant concepts are included, irrespective of whether it refers to a concrete entity, a set, a process, or a more abstract item. The purpose of the glossary is to list the terms used in the domain in such a way that their meaning is unambiguous and agreed upon by the domain experts and knowledge engineers. When a specific term is used in the domain to refer to two different concepts, new terms are introduced to reflect the different meanings, to avoid confusion and ambiguity. An example is the term *metastasis* which is commonly taken to refer to the process of transference of cancer cells as well as to the secondary tumours resulting from the process. Our glossary includes the separate terms *metastasis* and *metastatic tumour* to denote and distinguish between the two meanings. In the glossary, moreover, the terms are chosen to be independent of the projected use of the knowledge, to avoid too many ontological commitments. Table 1 shows a part of the glossary of our oesophagus ontology.

Term	**Meaning**
lamina propria	the first inner layer of the oesophageal wall
lymphatic metastasis	a metastatic tumour in a lymph node
metastasis	the transference of cancer cells from the primary tumour via blood or lymph vessels
metastatic tumour	a secondary tumour at another site than the primary tumour
site	location in the human body

Table 1: Part of the glossary of the oesophagus ontology

3.4 The static perspective

The oesophagus ontology captures the static relations between the various concepts in the domain, that is, the relations in which time does not play a role, in a separate component. Within this component, two types of relation are distinguished: *structural* relations and *definitional* relations.

3.4.1 Structural relations

The concepts in a domain are typically related to one another in many different ways. A concept may, for example, be a generalisation or a superset of another concept. It may also be a property of another concept. These relations in essence describe the structure of the domain knowledge and will be termed *structural relations*. Studying the structural relations in the domain of oesophageal cancer resulted in a number of *hierarchies*. As an example, Fig. 3 shows the hierarchy of pathological entities; the hierarchy is simplified for ease of presentation. Pathological entities, where the term pathological indicates that an entity deviates from what is considered normal, are of primary importance in our domain.

Each node in the hierarchy of pathological entities captures an object in the domain, that represents a set of individual entities. Each link separately describes a subset-of relation; the link is directed from the subset to the superset. All links pointing to a single superset are assumed to originate from mutually exclusive and collectively exhaustive subsets, called *blocks*; these links thus capture a *partition* of the superset. For example, the set of *tumours* is partitioned into the set of *primary tumours* and the set of *secondary tumours*. Although in our ontology, we have chosen to direct the links from a subset to a superset, thereby following the direction of the *block-of* relation, we could also have directed the links from the superset to the subset, following the direction of the *partitions-into* relation.

Associated with the various objects in the hierarchy are *attributes*. We distinguish between element attributes and set attributes. An *element attribute* describes a property

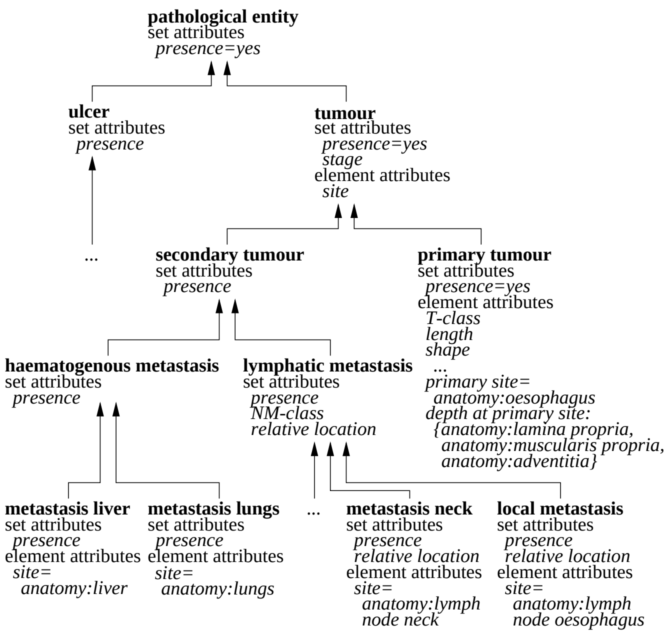

Figure 3: The hierarchy of pathological entities

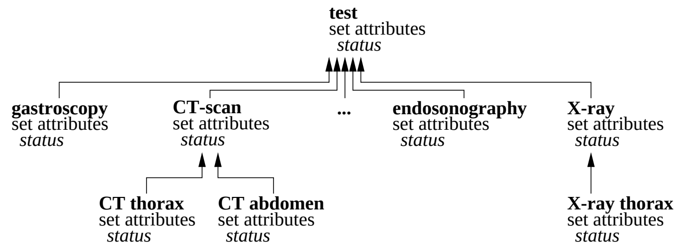

Figure 4: The hierarchy of tests

of each individual, or element, in the associated set. For example, each individual in the set of *tumours* has a *site*. Since the links in our hierarchy capture the partition relation, all element attributes that are associated with a particular node are also associated with the nodes that describe its subsets. All individuals in the set of *primary tumours*, which is a subset of the set of *tumours*, will thus have associated a *site*. Similarly, concrete values that are specified for the element attributes of a node are associated with the attributes of the nodes that describe its subsets. We thus do not allow (implicit) exceptions.

In addition to element attributes an object can have associated *set attributes* that specify properties of the represented set as a whole. For example, the set attribute *presence* of the object *tumour* represents the presence of at least one tumour in a patient. If this attribute has the value *yes*, the patient may have a primary tumour, but no secondary tumours. The set attribute *presence* of the object *secondary tumour* therefore does not necessarily have the value *yes*. The knowledge that is needed for establishing the value of a set attribute can be domain independent. For example, the attribute *presence* of the set of *haematogenous metastasis* has the value *yes* if and only if the attribute *presence* of at least one of its subsets has the value *yes*. This type of knowledge is considered to be common-sense knowledge and is not specified explicitly in our ontology for reasons of clarity. Set attributes, however, can also require domain-dependent knowledge for establishing their values. An example is the attribute *stage* of the node *tumour*. This attribute basically captures the stage of the set of tumours and, hence, of the cancer. The domain knowledge required to establish its value is specified explicitly in our ontology using *definitional relations*. Definitional relations constitute a second type of static relation, and will be discussed in more detail below.

An attribute, whether an element or a set attribute, can specify not just a value but also a reference to another object. Such a reference is depicted in the hierarchy by the name of the hierarchy in which the other object is located as a node, a colon, and the name of the referenced object. For example, the attribute *site* of the object *metastasis liver* specifies a reference to the object *liver* in the hierarchy of anatomical knowledge. Attributes may further specify domain declarations. For example, the depth of invasion of an oesophageal tumour at the primary site is specified to be one of the three layers, *lamina propria*, *muscularis propria* and *adventitia*, of the wall of the oesophagus. The set of possible values of an attribute that is specified in a node is equal to or a restriction of the set of values that is specified in the parent of the node.

To conclude, the hierarchy can be seen as a place-holder for data of a single patient. The element and set attributes that are specified in both leaf nodes and internal nodes can be assigned a patient-specific value. In the hierarchy of pathological entities, it is assumed that the sets that are represented by a leaf node contain one individual only, that is, we assume for example that a patient can have at most one *primary tumour*.

Upon constructing the oesophagus ontology, we noticed that organising the structural relations between the concepts in our domain was not as easy as it seemed at first sight. Often, there are several different ways of organising concepts in hierarchies [16], each having its specific advantages and disadvantages. We illustrate this observation by the hierarchies of tests and test results. These hierarchies are shown in Fig. 4 and 5; again, the hierarchies are simplified for ease of presentation.

The hierarchy of tests, shown in Fig. 4, contains nodes that represent the tests that are used to gain information about (often hidden) pathological entities. For example, a CT-scan of the thorax can be used to gather information about the presence or absence of metastases in the lungs. In organising the different tests in use in a hierarchy, we

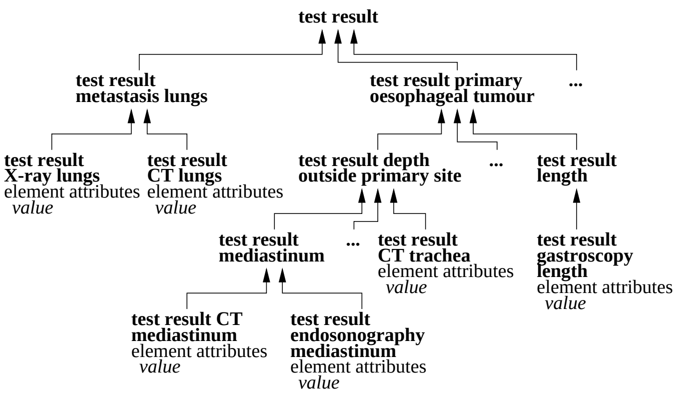

Figure 5: The hierarchy of test results

took the technology and the site it can be applied to for our criterion. For example, the CT-scan and the X-ray are two different technologies, represented by two nodes at the same level in the hierarchy. A CT-scan can be applied to the thorax, represented by the node *CT-scan thorax*, or to the lower abdomen where the liver is located, represented by the node *CT-scan abdomen*. The properties of a test such as the costs involved, can be included in the hierarchy as element attributes. Also the reliability characteristics of a test can be included, thereby following common practice in the field of medical decision making. In essence, we could have organised the results, or outcomes, of the various tests in the same hierarchy, for example by adding appropriate attributes to the objects that represent the tests. We chose, however, to capture the test results in a separate hierarchy, which is shown in Fig. 5. In this hierarchy, the results of the tests are organised according to the entities they provide information about. For example, a CT-scan of the thorax and an X-ray of the thorax both serve to yield information about the presence or absence of a secondary tumour in the lungs. The results of the two sets are represented by two separate nodes that are linked to the node *test result metastasis lungs*. If we would have merged the knowledge about tests and about their results into a single hierarchy, the information with respect to a single pathological entity would have been distributed over the hierarchy. For similar reasons, we decided not to extend the hierarchy of test results to include knowledge about the tests that are performed to obtain them. Knowledge about a test, for example, to which sites it can be applied, would have been distributed over the hierarchy of test results. For example, a CT-scan can be applied to the thorax and to the lower abdomen, yielding test results with respect to different pathological entities. In addition, a single CT-scan can have multiple results, providing information about different pathological entities. For example, a single CT-scan of the thorax may provide information about the presence or absence of metastases in the lungs and about the depth of invasion of the primary tumour into other organs than just the oesophagus. Test results with respect to different entities that are obtained using a single technology may thus be obtained from applying a single test that has multiple results, or from applying the technology to multiple sites. Including knowledge about the tests in the hierarchy of test results may obscure the knowledge with respect to the sites a test can be applied to. By representing the structural relations with respect to tests and their results in separate hierarchies, we feel that we captured the structure of the knowledge more faithfully than we would have had by merging the knowledge into a single hierarchy. To conclude, we would like to note that we explicitly distinguished between test results and the values of their attributes on the one hand, and pathological entities and their attribute values on the other hand, since test results do not always perfectly reflect the true value of an entity's attribute.

In addition to the three hierarchies discussed above, the oesophagus ontology contains a hierarchy of *pathological processes*. These processes underlie the progression of disease and play an important role in our domain of application. Knowledge about manifestations of diseases, such as weight loss and swallowing problems, is captured by the hierarchy of *manifestations*. Manifestations of a disease are explicitly distinguished from test results as such manifestations are typically noticed by the patient and therefore do not require a test for their observation. The ontology further captures anatomical knowledge, that is, knowledge about the structure and organisation of the human body. Anatomical knowledge is often referred to in the remainder of the ontology. We decided to represent the anatomical knowledge by a hierarchy and by pictures. The hierarchy captures the anatomical units and the way they are related by set inclusion. The locations of the various units in the human body relative to one another are represented by the pictures. Capturing this knowledge in a more formal way, for example using the attributes associated with the objects in the hierarchy, would result in a less compact representation, that would be more difficult to access for the domain experts. We feel that the rather informal representation in pictures is closer to the way the experts think about their domain.

In the hierarchies of our ontology, we have made little commitment to the ontology's projected use. One of the few commitments that we have made, is the value *yes* for the attribute *presence* that is associated with the object *primary tumour*. This value reflects our previous assumption that a patient is known to be suffering from oesophageal cancer. The ontology further is committed as little as possible to the task at hand. We have decided, however, to keep the hierarchies restricted in scope, which in essence is a commitment to the application aimed for. To conclude, we would like to note that the hierarchies of the oesophagus ontology and the glossary exhibit some overlap in the knowledge they represent. For example, the description of the *lamina propria* in the glossary specifies some knowledge that is also captured in the representation of anatomical knowledge. For reasons of clarity, we decided to maintain both representations. To ensure coherence upon maintenance, the relationship between the two representations is captured by the coherence constraints. We will return to these constraints in Section 3.6.

3.4.2 Definitional relations

The second type of static relation that we distinguished in our oesophagus ontology, is the *definitional relation.* A definitional relation is a relation between attribute values: it basically defines the value of an attribute in terms of the values of some other attributes. As an example, Table 2 shows how the value of the attribute *stage* of the object *tumour* is defined in terms of the attribute *presence* of the object *haematogenous metastasis*, the attribute *NM-class* of the object *lymphatic metastasis*, and the attribute *T-class* of the object *primary tumour.* The values of the latter attributes are, in turn, defined in terms of values of yet other attributes.

In addition to the definitional relations themselves, the ontology captures their overall structure. This structure is represented by means of a graph that has the various attributes involved in the definitional relations for its nodes. The incoming arcs for an attribute originate from the attributes whose values serve to define its value. The graph thus abstracts from specific values and shows whether or not two or more attributes are related at the value level; it basically constitutes an additional *description level.* The definition graph for our domain of application is depicted partially in Fig. 6. It shows, for example, that the value of the attribute *relative location* of the object *lymphatic metastasis* is defined in terms of the values of the attributes *relative location* of the objects *metastasis neck*, *metastasis truncus*, and *local metastasis.*

To conclude, we recall that our ontology, in addition to the definitional relations, includes a glossary. This glossary contains an informal, yet explicit and unambiguous description of the meaning of a term; such a description

haematogenous metastasis *presence*	lymphatic metastasis *NM-class*	primary tumour *T-class*	tumour *stage*
no	N0	T1	I
no	N0	T2 or T3	IIA
no	N1	T1 or T2	IIB
...	...	...	...
yes	-	-	IVB

Table 2: Definitional relation for the stage of an oesophageal cancer

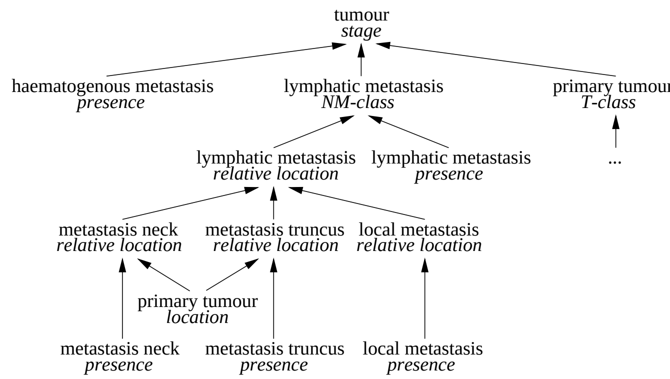

Figure 6: Part of the attribute-level graph of the definitional relations

may refer to other terms. The main purpose of the glossary is to provide for communication and comprehension. The definitional relations in contrast, are represented in a much more formal way and always describe terms using other terms. The main purpose of this part of the ontology is to capture the exact relations between the attribute values in the hierarchies.

3.5 The dynamic perspective

A domain of application may involve processes that have important effects over time. In the domain of oesophageal cancer, for example, the pathological process of metastasis via the blood vessels may result in a secondary tumour in the liver. The process of metastasis precedes, in time, the presence of metastatic cancer in the liver. We use the term *dynamic relation* to refer to such time-involving relations.

Our domain of oesophageal cancer includes various dynamic relations between attribute values. These relations are represented in the ontology by means of tables. The ontology contains, for example, a table that captures the dynamic relation that expresses that the attribute *presence* of the process of *metastasis via blood vessels* having the value *yes* may result in the attribute *presence* of *metastasis liver* adopting the value *yes.* Dynamic relations, like static relations, have associated a type. The relation that is described above is an example of a *resulting* relation, since it relates a process to its result.

The tables of the dynamic perspective of our ontology describe the dynamic relations between the various attributes at the level of their values. As with the definitional relations in the static perspective, our ontology includes a graph showing whether or not two or more attributes are related at the value level. This graph thus makes the overall structure of the dynamic relations explicit. Fig. 7 depicts a part of this attribute-level graph. It shows, for example, that the pathological process of invasion may affect the depth of invasion of the primary

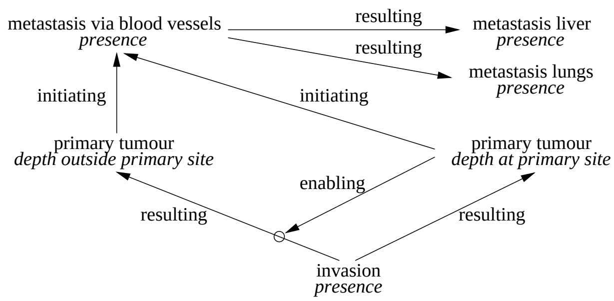

Figure 7: Part of the attribute-level graph of the dynamic relations

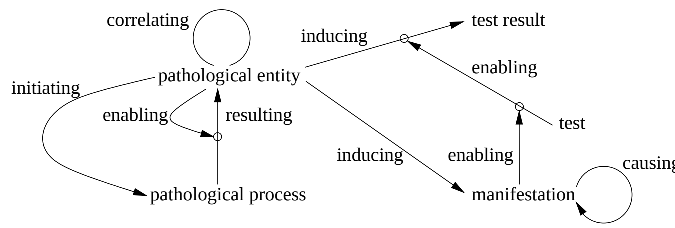

Figure 8: The high-level graph of dynamic relations

tumour into the oesophageal wall. The primary tumour may also invade organs outside the oesophagus, *provided* it has grown through the entire oesophageal wall. We say that the depth of invasion of the tumour at the primary site *enables* the invasion outside the oesophagus. Furthermore, a tumour that has invaded the oesophageal wall may *initiate* a process of metastasis via the blood vessels, which in turn may *result* in secondary tumours in the liver and in the lungs.

A detailed investigation of the dynamic relations in the domain of oesophageal cancer revealed several regularities. These regularities are not easily recognised from the tables capturing the relations, however, as they are implicitly present; nor are they readily seen from the attribute-level graph of the relations. Since the ontology serves to represent the relevant knowledge of the domain explicitly, we decided to make the regularities that we found explicit by using additional description levels. The dynamic relations are not just represented at the level of attributes and at the level of their values, as discussed above, but also at the level of the objects involved. We recall that these objects are captured as nodes in the various hierarchies in our ontology. By taking the graphical representation of the structure of the dynamic relations at the attribute level as a starting-point and abstracting from the attributes, a graph of the object-level regularities is obtained. This graph shows whether or not objects are related at the attribute level. Further abstraction, using the *block-of* relations from the hierarchies, then results in a graph that explicitly represents the high-level regularities among the dynamic relations in the domain. Fig. 8 depicts this graph of dynamic relations of our oesophagus ontology. The left-hand part of this high-level graph is an abstraction of the attribute-level graph of Fig. 7. In constructing the high-level graph, *metastasis via blood vessels* and *metastasis liver*, for example, have been generalised to *pathological process* and *pathological entity*, respectively, using knowledge from the hierarchies. The high-level graph thus shows that pathological processes may result in pathological entities. It further shows that manifestations may cause other manifestations. For example, a patient's difficulties with swallowing food may cause weight loss. The high-level graph of Fig. 8 further reflects that a pathological entity, or, at a less abstract level, a property of a pathological entity may be observed. In essence, the pathological entity *induces* the test result. The result can only become available, however, if an appropriate test is performed to this end. In our domain of oesophageal cancer, for example, a gastroscopic examination, that is, letting a camera into the oesophagus, may reveal the length of the primary tumour. The test basically *enables* the observation. If the oesophagus is obstructed by the primary tumour, however, the camera cannot pass the obstruction and may not give the results aimed for. The observation then is disabled, or negatively enabled, by the manifestation.

In distinguishing between the various types of dynamic relation we build upon the concept of *state*. A *state* reflects a specific situation, in time, in which a property of an entity has a certain value; such an entity can be any object in our ontology, excepting pathological processes. A *causing* relation now asserts that a particular state may cause some other state to occur. In such a relation, the latter state can never precede the former in time. Generally, two types of causation are distinguished [17]. *Continuous causation* denotes that the cause needs to be present continuously for the effect to persist. An example of a continuous causation is the relation between a patient's difficulties with swallowing food and weight loss. In *one-shot causation*, on the other hand, the cause is required to be present only momentarily for the effect to occur and continue. For example, once metastases in the liver occur, their continuous existence is independent of the presence of their cause. In our ontology, we adopted this distinction. We attach to each *causing* relation a label denoting it as capturing one-shot or continuous causation. In addition, we specify for each *causing* relation whether or not the occurrence of the effect may be delayed from the onset of the cause. For a *causing* relation that is labeled as capturing continuous causation, moreover, we indicate whether or not a delay may occur in the termination of the effect after the cause has halted to be present. We decided not to specify time-intervals for the possible delays, since the exact delays are not relevant for our application.

For ease of presentation, the labels that are attached to the various *causing* relations are not shown in the figures in this section. As mentioned above, our ontology also includes *initiating* and *resulting* relations. The former relation asserts that a state may initiate a process to happen, whereas the latter relation asserts that this process may then result in another state. The distinction between *initiating*, *resulting*, and *causing* relations that we made in our ontology is closely related to the distinction that has been made in [17]. As the *causing* relations, each *initiating* or *resulting* relation has attached a label denoting whether it captures a *one-shot* or a *continuous* mechanism and whether or not it involves a delay.

An *inducing* relation asserts that a particular state may induce another state. An *inducing* relation, like a *causing* relation, involves time in the sense that the induced state cannot occur before the inducing state has occurred. Different types of knowledge may underly an *inducing* relation, however. We abstracted from this knowledge, since it does not pertain to the primary processes that are involved in the progression of the cancer; this is, in essence, an ontological commitment to our application. An *enabling* relation differs from *causing* and *inducing* relations in that it relates a state to a relation rather than to another state. It asserts that a particular state is a prerequisite for some dynamic relation to be active. To conclude our discussion of the various different types of dynamic relation, we observe that the high-level graph of Fig. 8 includes *correlating* relations between pathological entities. If two pathological entities are correlated, they typically have a common cause that has not been explicitly captured in the ontology. The correlations in our oesophagus ontology originate from processes that play a role in the etiology of the cancer; as argued in Section 3.2, we decided not to model the etiology in our ontology. Since the two correlated entities have a common cause, the relation between them involves time and therefore is dynamic in nature. As the common cause remains implicit, however, a *correlating* relation is undirected with respect to time.

We would like to note that there are several different ways of modelling the *causing* relations in our ontology [17]. We consider, as an example, the knowledge that a pathological entity may initiate a pathological process, which in turn may result in a pathological entity. This knowledge is modelled by means of two dynamic relations. We could also have modelled this knowledge by a single *causing* relation between two pathological entities, leaving the pathological process underlying this dynamic relation implicit. Making the process of *metastasis via blood vessels* implicit, Fig. 9 shows the *causing* relations between the attributes *depth outside primary site* and *depth at primary site* of *primary tumour* on the one hand and the attributes *presence* of *metastasis lungs* and *presence* of *metastasis liver* on the other hand. By making the process implicit, the knowledge is represented in a more abstract way. We decided, however, to represent explicitly the pathological processes that play an important role in the domain of oesophageal cancer and in the communication with our domain experts. As the processes underlying the *causing* relation from a *manifestation* to another *manifestation*, on the other hand, are of less importance to our application in the making, we decided not to specify them explicitly for reasons of clarity. Capturing the *causing* relations between manifestations in less detail than the *causing* relations between pathological entities is, in essence, an ontological commitment to our application.

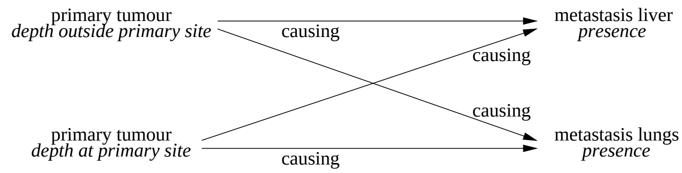

Figure 9: *Causing* relations, replacing *resulting* and *initiating* relations

3.6 Coherence constraints

To meet the coherence criterion, our ontology contains several coherence constraints specifying requirements that the knowledge represented in the glossary, the static component and the dynamic component should adhere to. These constraints typically refer to the overlapping parts in the ontology and are instrumental in ensuring that no incoherences will arise when maintaining the ontology. An example coherence constraint pertains to the domain declaration of the attribute *depth at primary site* of the object *primary tumour*. It specifies that a particular value can be adopted by the attribute if and only if the value is known to be part of the site of the primary tumour. The coherence constraint thereby makes explicit the knowledge underlying the overlap between the hierarchy of pathological entities and the representation of anatomical knowledge.

3.7 Other approaches to specifying ontologies

In building the oesophagus ontology, we have taken a number of design decisions pertaining, among others, to the representation language used and to the decomposition of the ontology into components. While the decisions that we have taken, are appropriate for our domain of application, they may not be the most suitable for other domains. Many of our decisions in fact depend on the properties of the domain knowledge, on the goal of the ontology, and on the people involved in the construction and projected use of the application. Representing the knowledge in a highly structured and more formal domain of application, for example, may require a more formal representation language than the one that we decided to use. More in general, different applications pose different requirements to the ontology and to the representation language to be used for its specification.

Two well-known languages for representing ontologies are *KIF* [18] and *Ontolingua* [19, 20]. KIF is a rather formal language, that is based on first-order predicate calculus. It provides for the definition of objects, relations and functions; the interpretation of these constructs is constrained by logical axioms. Ontolingua is based upon KIF in the sense that the constructs offered by KIF are also offered by Ontolingua. Ontolingua extends KIF, however, by offering higher-order constructs that are composed of basic KIF building blocks. Examples are explicit constructs for specifying hierarchies and *object-attribute-value* relations. Although such relations can be specified in KIF as well, it does not offer an explicit construct for this purpose. Ontolingua, moreover, offers the additional facility of including natural-language annotations with each specification in an ontology. Ontolingua has been used to describe ontologies in various domains of application, including the medical domain of acute radiation syndrome [11], enterprise modelling [21], and chemistry [22].

KIF and Ontolingua do not allow, for example, for specifying a separate, informal glossary of terms as we do in the oesophagus ontology. Building upon the constructs offered, however, the knowledge in a glossary can be captured by specifying, for each term, a relation between the term and a description of its meaning. In Ontolingua, the facility of adding natural-language annotations can be used for this purpose as well. Both languages also do not provide constructs for including pictures, such as we use in our ontology for representing the human anatomy. Nor do the languages provide graphical constructs for representing relations, for example, for representing the higher-level relations in a graph. G. van Heijst et al., however, developed a tool for visualising relations in an Ontolingua ontology [11].

For constructing ontologies in a formal language such as Ontolingua, the *Methontology* methodology [22] proposes the use of intermediate representations that are less formal. The knowledge to be included in the ontology then is captured in diagrams, tables, trees and natural language before the formal ontology is actually constructed. The intermediate representations serve to bridge the gap between the way domain experts think and talk about their domain, and the language in which the ontologies are formalised. The use of the intermediate representations proposed by Methontology has been illustrated with an ontology in the domain of chemistry [22]. We note that the aim of facilitating communication between the domain experts and the knowledge engineers also underlies the use of a semi-formal language for our oesophagus ontology.

In addition to the special-purpose methodologies that are tailored to the construction of ontologies, various more general knowledge-engineering methodologies exist. These methodologies support the representation of domain knowledge in a knowledge model, that serves a similar purpose as our ontology. The well-known *CommonKADS* methodology [6], for example, offers a semi-formal language for specifying a knowledge model that integrates several representations. It provides graphical constructs such as diagrams, mainly based on UML, and textual constructs. From a more methodological perspective, CommonKADS recommends to compose a list of domain terms with their explanations, and advises to include the resulting glossary in a document that contains background documentation of the application that is being constructed. It further recommends to specify hierarchies and relations much in the way we have used them in our oesophagus ontology. The design of our oesophagus ontology has in fact been motivated by the general, domain-independent CommonKADS methodology. We further refined its approach, however, to meet the requirements of the knowledge in our domain of application, and of the goal of our ontology. For example, CommonKADS does not propose decompositions to distinguish between a static and a dynamic perspective, and to explicitly capture the overlap between the various parts. Nor does it explicitly support the representation of domain knowledge at the attribute-level.

The overview in this section is not exhaustive. A variety of languages for representing ontologies exist, ranging from natural language to first-order predicate logic. Also, a wide range of approaches for developing ontologies exist. [16] and [22] provide lists of references to ontology-related work. An overview and analysis of methodologies for developing ontologies is given in [23], whereas [24] provides a framework for comparing ontologies, and, in addition, contains a discussion of a number of existing ontologies. In choosing a suitable representation language and approach for an application at hand, the properties of the domain knowledge, the goal of the ontology, and the people involved play a major role.

4 Benefits of ontologies for probabilistic networks

The common practice of engineering probabilistic networks is to model knowledge directly into a network. In our experience, however, domain knowledge often cannot be represented straightforwardly into a probabilistic network. In fact, modelling knowledge in the network directly may result in considerable loss of information, as we have argued in Section 3.1. Moreover, background knowledge is often not explicitly represented in a probabilistic network, nor are the underlying structure of domain concepts and the regularities in the domain. These issues can hamper the communication between the domain experts and knowledge engineers; moreover, it may hamper a shared understanding of the domain of application among all people involved in building and maintaining the network, thus impeding further development and maintenance of the network. To forestall these problems, we feel that constructing an ontology should be integrated in a systems-engineering approach to developing probabilistic networks.

Constructing an ontology that captures domain knowledge has many benefits. A major benefit is that it renders background knowledge explicitly available, including knowledge about the structure and the regularities in the domain. Moreover, the ontology enables explicit representation of all elicited concepts and relations from the domain of application, as opposed to the language of probabilistic networks. Since we have used a rather informal language for specifying our ontology, we feel that the knowledge contained is readily recognizable for domain experts and for knowledge engineers. This will facilitate a shared understanding of the domain knowledge among the domain experts and the knowledge engineers involved in building and in maintaining the system. The ontology can serve as a means of communication between knowledge engineers and experts during knowledge acquisition and validation, but also as documentation of the domain knowledge during system design and maintenance.

As an additional benefit, we observe that, since an ontology explicitly represents the structure and the regularities of the domain knowledge, it can be used to guide further acquisition efforts [11]. For example, if the domain expert states that a specific pathological process plays an important role, the knowledge engineer may focus her questions on discovering what pathological entities may result from this process. Also, the represented knowledge can be validated more easily against completeness and consistency, upon further knowledge acquisition or maintenance. For example, if a new test is to be represented in our ontology, then the knowledge engineer should establish whether or not this test will always yield results, irrespective of the presence of certain manifestations. Moreover, if a dynamic relation is to be added to an initial collection of relations, the knowledge engineer should verify that it meets the regularities in the domain. An irregularity, for example an indication that a test affects a pathological entity, then serves as a warning to further investigate the newly acquired relation. If it appears to be correct, the high-level graph should be adjusted accordingly.

Once the ontology is considered a faithful representation of the relevant domain knowledge, it can be exploited for actually building the graphical structure of the network in the making. Supported by carefully designed guidelines, the structure can be derived from the ontology in a number of steps. A knowledge engineer can focus on the modelling and design decisions that are relevant in a particular step, in isolation of the issues pertaining to other steps. We refer the reader to [12] for more details about the specific steps that can be taken in deriving the graphical structure of a network from an ontology.

Since the ontology explicitly represents all elicited domain knowledge, it can also be used to facilitate the application of other existing approaches for constructing a probabilistic network from elicited domain knowledge. For example, M. Neil, N. Fenton, and L. Nielsen propose the use of *idioms*, which are reusable network structures that represent generic patterns [25]. They can be used as building blocks in the construction of probabilistic networks. The idioms are typically instantiated by matching them against pieces of the real world problem at hand. Using an ontology will facilitate this process of instantiation: the idioms can easily be matched against segments of the documented domain knowledge. Another example pertains to the application of *network fragments*, that can be combined to yield a probabilistic network [7]. Within our ontology-based approach, the network fragments can be created using the explicitly represented domain knowledge. A similar observation can be made with respect to, for example, approaches that build upon the concept of object orientation [4, 5, 8, 26, 27]: the objects and the relations involved can be derived from the documented knowledge in the ontology. The various approaches that are currently in use for constructing a network directly from elicited domain knowledge can thus often be facilitated by using an ontology. Since the ontology explicitly represents the knowledge in the domain, including its structure and regularities, it provides a solid foundation for the development of a probabilistic network using the various existing approaches.

5 Conclusions and future research

Building a probabilistic network for a real-life domain of application is a hard and time-consuming process. In this process, numerous design decisions are taken. We have further noticed that often a lot of knowledge that is needed to understand the domain, is not represented explicitly in a probabilistic network. Construction and maintenance of a network are seriously hampered if the elicited domain knowledge, including the background knowledge, is not made explicit by proper documentation. In this paper, we have studied the usefulness of ontologies for this purpose by constructing an ontology for the domain of oesophageal cancer, based upon our probabilistic network for the staging of cancer of the oesophagus and the knowledge elicited for its construction.

The oesophagus ontology is composed of various components that represent different types of knowledge from different perspectives at different levels of abstraction. Although the ontology has not been explicitly validated against the domain experts involved in the construction of the oesophagus network, we feel that it represents a rich, well-organised body of knowledge for further reference. By incorporating the ontology into the documentation of the oesophagus network, only the knowledge in the application domain is documented explicitly. However, numerous design decisions have been taken in the network's construction, for example pertaining to the translation of domain concepts into statistical variables. These decisions basically constitute the link between the oesophagus ontology and the network [12]. We feel that the documentation should be extended to include an explicit specification of these design decisions.

Based upon our experience with constructing the oesophagus ontology, we see many benefits from the use of ontologies for explicitly documenting domain knowledge in engineering probabilistic networks. These benefits pertain to the various stages in the engineering process, including the acquisition of domain knowledge, the design of the probabilistic network and system maintenance. Our ontology-based approach can be combined with other approaches that are currently in use for building networks. Moreover, the application of such approaches will often be facilitated by the explicit and structured representation of the domain knowledge in an ontology.

Acknowledgements. This research has been (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organisation for Scientific Research. We are grateful to Babs Taal and Berthe Aleman from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, who provided the domain knowledge for the construction of the oesophagus network. This paper is an improved and extended version of "A case study in ontologies for probabilistic networks", in *Research and Development in Intelligent Systems XVIII*, M. Bramer, F. Coenen, and A. Preece, Eds. London: Springer-Verlag, 2002, pp. 229–242.

References

[1] F.V. Jensen, *Bayesian Networks and Decision Graphs.* New York: Springer-Verlag, 2001.

[2] M.J. Druzdzel and L.C. van der Gaag, "Building probabilistic networks: "Where do the numbers come from?" Guest editors' introduction", *IEEE Transactions on Knowledge and Data Engineering*, vol. 12(4), pp. 481–486, 2000.

[3] L.C. van der Gaag and E.M. Helsper, "Experiences with modelling issues in building probabilistic networks", in *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Proceedings of EKAW 2002*, Lecture Notes in Artificial Intelligence, vol. 2473, A. Gómez-Pérez and V.R. Benjamins, Eds. Berlin Heidelberg: Springer-Verlag, 2002, pp. 21–26.

[4] S.M. Mahoney and K.B. Laskey, "Network engineering for complex belief networks", in *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, E. Horvitz and F. Jensen, Eds. San Francisco: Morgan Kaufmann, 1996, pp. 389–396.

[5] K.B. Laskey and S.M. Mahoney, "Network engineering for agile belief network models", *IEEE Transactions on Knowledge and Data Engineering*, vol. 12(4), pp. 487–497, 2000.

[6] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. Van de Velde, and B. Wielinga, *Knowledge Engineering and Management: The CommonKADS Methodology.* Cambridge, Massachusetts: The MIT Press, 2000.

[7] K.B. Laskey and S.M. Mahoney, "Network fragments: representing knowledge for constructing probabilistic models", in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, D. Geiger and P. Shenoy, Eds. San Francisco: Morgan Kaufmann, 1997, pp. 334–341.

[8] D. Koller and A. Pfeffer, "Object-oriented Bayesian networks", in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, D. Geiger and P. Shenoy, Eds. San Francisco: Morgan Kaufmann, 1997, pp. 302–313.

[9] Th.R. Gruber, "Towards principles for the design of ontologies used for knowledge sharing", *International Journal of Human-Computer Studies*, vol. 43, pp. 907–928, 1995.

[10] M. Uschold and M. Gruninger, "Ontologies: principles, methods and applications", *The Knowledge Engineering Review*, vol. 11(2), pp. 93–136, 1996.

[11] G. van Heijst, A.Th. Schreiber, and B.J. Wielinga, "Using explicit ontologies in KBS development", *International Journal of Human-Computer Studies*, vol. 46(2/3), pp. 183–292, 1997.

[12] E.M. Helsper and L.C. van der Gaag, "Building Bayesian networks through ontologies", in *ECAI 2002: Proceedings of the 15th European Conference on Artificial Intelligence*, F. van Harmelen, Ed. Amsterdam: IOS Press, 2002, pp. 680–684.

[13] L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B.M.P. Aleman, and B.G. Taal, "Probabilities for a probabilistic network: a case-study in oesophageal cancer", *Artificial Intelligence in Medicine*, vol. 25(2), pp. 123–148, 2002.

[14] R. Studer, V.R. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods", *Data & Knowledge Engineering*, vol. 25, pp. 161–197, 1998.

[15] T. Bylander and B. Chandrasekaran, "Generic tasks for knowledge-based reasoning: the "right" level of abstraction for knowledge acquisition", in *Knowledge Acquisition for Knowledge-Based Systems*, vol. 1, B.R. Gaines and J.H. Boose, Eds. London: Academic Press, pp. 65–77, 1988.

[16] B. Chandrasekaran, J.R. Josephson, and V.R. Benjamins, "What are ontologies, and why do we need them?", *IEEE Intelligent Systems & Their Applications*, vol. 14(1), pp. 20–26, 1999.

[17] C. Rieger and M. Grinberg, "The declarative representation and procedural simulation of causality in physical mechanisms", in *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Los Altos, California: William Kaufmann, 1977, pp. 250–256.

[18] M.R. Genesereth and R.E. Fikes, "Knowledge Interchange Format version 3.0", Computer Science Department, Stanford University, Tech. Rep. Logic-92-1, 1992.

[19] Th.R. Gruber, "A translation approach to portable ontology specifications", *Knowledge Acquisition*, vol. 5, pp. 199–220, 1993.

[20] A. Farquhar, R. Fikes, and J. Rice, "The Ontolingua Server: a tool for collaborative ontology construction", *International Journal of Human-Computer Studies*, vol. 46, pp. 707–727, 1997.

[21] M. Uschold, M. King, S. Moralee, and Y. Zorgios, "The Enterprise Ontology", *The Knowledge Engineering Review*, vol. 13(1), pp. 31–89, 1998.

[22] M. Fernández-López, A. Gómez-Pérez, J. Pazos Sierra, and A. Pazos Sierra, "Building a Chemical Ontology using Methontology and the Ontology Design Environment", *IEEE Intelligent Systems & Their Applications*, vol. 14(1), pp. 37–46, 1999.

[23] M. Fernández-López and A. Gómez-Pérez, "Overview and analysis of methodologies for building ontologies", *The Knowledge Engineering Review*, vol. 17(2), pp. 129–156, 2002.

[24] N. Fridman Noy and C.D. Hafner, "The state of the art in ontology design: a survey and comparative review", *AI Magazine*, vol. 18(3), pp. 53–74, 1997.

[25] M. Neil, N. Fenton, and L. Nielsen, "Building large-scale Bayesian networks", *The Knowledge Engineering Review*, vol. 15(3), pp. 257–284, 2000.

[26] D. Koller and A. Pfeffer, "Probabilistic frame-based systems", in *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*, Menlo Park, California: AAAI Press / The MIT Press, 1998, pp. 580–587.

[27] O. Bangsø and P.-H. Wuillemin, "Top-down construction and repetitive structures representation in Bayesian networks", in *Proceedings of the Thirteenth International Florida Artificial Intelligence Research Society Conference*, J. Etheredge and B. Manaris, Eds. Menlo Park, California: AAAI Press, 2000, pp. 282–286.