Abstract: This paper emphasize the role of ontologies in the e-Health domain, namely in Occupational Health. It describes (by means of a case study) the representation and integration of the semantic and modelling layers of a system based on ontologies for the prevention of occupational risks, as an application of ontologies in the occupational health domain. The case study enumerates basic concepts and relationships in the proposed reference ontology and model, as well as the basic reasoning for them. The personalized training on the health, safety, security and environmental issues is related to the work activities. The training is based on the automatic and semantics-based discovery of the prevention documents, actions, methods, etc that fit the user's request.

Key-Words: occupational health; risk prevention; formal ontologies; ontology-based modelling; e-Training.

1 Introduction
According to the Joint ILO/WHO committee in Occupational Health (1950), the goal of the occupational health is the promotion and maintenance of the highest degree of physical, mental and social well-being of the workers in all domains, their protection from risks caused by their working conditions, negligence, lack of training, etc [1]. Hence, the occupational risks are a special type of risks that appear in the work environments with a high probability of harming people or machines. Despite its importance in the risk management (RM) inside and across organizations, the occupational risk prevention is still a manual process, specific to domains like health, construction, transportation, industry, biology, etc. A proactive approach in RM relies on the risk early recognition and prevention.

The training in occupational risk prevention should advise the operator/worker on the health, safety, security and environmental issues related to his work. He can ask for training before or during the execution of an activity or before the use of a certain machine.

The system described in this paper relies on the standard terminology proposed with ISO CD31000 [2], [3], combined with the terminology common to several upper-level ontologies and process models. In Europe, the risk prevention is subject of two directives Seveso I and Seveso II [4] that establish the domain terminology, the obligations and normative documents regarding the large scale industrial hazards.

Ontology in philosophy is a field that studies what exists in the world or human being. In computer science (especially in artificial intelligence), in information science, in bioinformatics and biomedical informatics, it is a sharable, reusable, machine-readable data structure, emphasizing the practical usage [5]. The most typical kind of ontology for the Web has a taxonomy and a set of inference rules [6].

Ontology is practically, “a vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words” [7]. The relations between the concepts in ontology allow inferences for the information interpretation and for the derivation of new information/knowledge. The explicit axioms allow the approximation of the term meaning and the validation of ontology specification [8].

Ontologies play an important role in medical informatics, contributing to the interoperability between systems (possibly distributed on web), to the access to heterogeneous information sources, to the reuse of voluminous and complex information involved in health care activities. The ontologies provide a common language for a domain, a vocabulary that can be used for data description and analysis [9].

There is no system based on knowledge and semantics for risk prevention and for training and
dynamic discovery of prevention information, documents and actions. The ontologies are also used mainly for security management (of assets, networks, information systems, databases, etc). Some examples are in [10]-[15]. The training is given by consulting companies and it is not continuous and mandatory before the execution of the dangerous activities.

However, the risk prevention and training with a similar idea but a simpler reference ontology and model were proposed in [16] for the construction domain.

2 The General Architecture of the System for Risk Prevention

Fig. 1 gives the general architecture of the system for risk prevention.

The intended system will have components distributed on two platforms: platform for the risk design, i.e., for the risk identification, description and analysis (Fig. 1, left); and platform for the risk evaluation and decision-making on the training query results (Figure 1, right). The two platforms share the repository composed of ontologies, rules, queries and documents. The platform for the risk design contains: (1) a Model Editor to build and instantiate domain or application specific models (e.g., the model for risk prevention in Fig. 2); (2) an Ontology Editor to build and instantiate (specialization, composition or list-like) ontologies involved in models; (3) a Rule Editor to define or customize domain or application specific rules in an organization (e.g., for risk prevention); (4) a Query Template Editor to predefine or customize templates for application specific queries on a model (for example, queries for training in a certain domain).

A specific feature of this platform is that the rule and query editors are tightly integrated with the model and ontology editors (as exemplified for queries in Fig. 4, where the concepts and attributes are directly seen and selected during the query composition).

The platform for the risk evaluation and decision-making contains: (1) a Tool for query editing and execution to dynamically compose the training queries and ask for their execution; (2) a Tool for the ontology and model navigation in order to compose a rule or query; (3) an Inference Engine, automatically called after the submission of the query, in order to perform the automatic discovery of the query results (e.g., training documents and the appropriate prevention actions, previously registered in the system and referred to in the domain model). Besides the conditions and constraints in the query, the discovery will also rely on rules previously defined by the risk designer using the rule editor; (4) a Generator of the query results, called by the inference engine, after the result discovery, in order to arrange the query results, according to the model and user's preferences.

The system is developed using Microsoft Visual Studio 2008 and about Suite product for ASP.Net [19]. It integrates the expression evaluator given in [15], adapted to the use of concepts and attributes in ontologies; and, the interface for the rule and query editors is inspired from [16].
3 Layers for the Representation of Occupational Risk Prevention

The e-Training system will integrate three layers based on semantics for the representation of the occupational risks and of the context for their occurrence and prevention:

The **semantic layer** is composed of the reference ontology (the generic framework for the identification, analysis, evaluation, classification and description of the risk factors, of the relationships between them, of the consequences and actions needed for their prevention) and of domain ontologies (the domain-specific implementation of the reference ontology) that give the basic vocabularies for domains with potential risks. The domain ontologies are populated by domain experts (risk designers) using the ontology editor. They are represented by: (1) domain-specific taxonomies, i.e., hierarchies composed of concepts connected, in this system, by relationships like: specialization or synonymy or composition (part-of) or list-like relationships; and (2) attributes of and constraints upon the concepts and relationships in ontologies.

The **modelling layer** is needed in addition to the semantic layer in order to represent ontology-based models for applications. Hence, a model is seen here as a union of relationships between concepts in different ontologies, along with their attributes and the constraints on them. In this system, the *inter-ontology relationships* are defined according to the reference model represented in Fig. 2.

The **execution (technological) layer** represents the ontologies and models, the documents, rules, constraints and queries in a format interpretable by the software.

Fig. 2 shows how the semantic and modelling layers for risk prevention are integrated from the conceptual point of view. The modeling layer represents the relationships between the ontologies defined on the semantic layer. These relationships have been selected depending on the needed reasoning on them and on the context for the risk identification, analysis, evaluation and prevention, identified at this moment.

In the system described in this paper, the modelling layer is composed of:

- A reference model *(application modelling schema)*, built by the application designer;
- A domain model *(application modelling content)* that instantiate the reference model. Domain models can be populated by the application designer and by another domain expert.

![Fig. 2 Basic types of ontologies and relationships in the reference model for occupational risk prevention](image_url)

The model is seen as an *ontology-based extension of an ER (Entity-Relationship) model* and graphs. The model in our system replaces the entity types with ontologies. This modification introduces an additional problem caused by the *semantic relationships* inside ontologies, in addition to the *modelling relationships* between entity types in the ER model.

The model graph definition in this system is summarized below.

**ModelGraph** = *(ModelNode, ModelEdge)*, where

**ModelNode** = ConceptNode ∪ ConceptInstanceNode

Nodes can be concepts or concept instances.

**ConceptNode** = \{C(O), \{Subtree(C(O))\}\};
- \{CAtr(C(O), \{Subtree(C(O))\)}

**ConceptInstanceNode** = \{C(O), \{Subtree(C(O))\}\};
- \{CAtr(C(O), \{Subtree(C(O))\)}

**ModelEdge** = Set of all the relations between the concepts and their instances.
The generic concepts, their attributes, relationships and constraints in the reference model are specialized and instantiated by the risk designers, resulting in domain models (e.g., for biological or industrial risks). For any domain concept, the designer instantiates the concept attributes defined in the ontology. They can be reference attributes, (for the concept unique identification) or inherited attributes (that can be ontology or concept specific attributes). Also, by their instantiation, the inter-ontology relationships in the reference model become domain-specific relationships between concrete concepts in ontologies or between concept instances.

\[
\{\nu_{ijk}\}_{i=1:I, j=1:J, k=1:K}
\]

where

- \(O_i\) is an ontology in the model;
- \(C_i(O_i)\) is a concept in \(O_i\);
- \(\text{Subtree}(C_i(O_i))\) is the set of the semantic subtrees for the concept \(C_i(O_i)\) (composed of the concept (specialization) subtypes or (composition) parts and of its synonyms). It is NULL, if the concept belongs to a list-like ontology and does not have synonyms.
- \(\text{CAtr}_k\) is the set of attributes for a concept \(C_i(O_i)\) or for a concept in one of sub-trees;
- \(\nu_{ijk}\) is the value of an attribute in the set \(\text{CAtr}_k\).

For each domain-specific concept, regardless of the taxonomy, its author should instantiate its specific attributes that help for its unique identification and compose the concept descriptor. Also, the generic constraints in the reference ontology will be adapted/ instantiated in order to become domain-specific constraints; or, new constraints and rules might be imposed to certain domain-specific concepts or relationships.

### 4 Building Domain-Specific Ontologies and Models

The generic concepts, relationships and constraints in the reference ontology and model are specialized in domain ontologies and models (for instance, for the biological and industrial risks) by the risk designers/ managers in each workplace/ organization.

First, the main types of activities/ tasks with major risks, at both individual and sequential level, should be identified and analyzed. Then, for each identified activity, the followings actions should be performed by the risk designers:

- Selection, classification (possibly, a multi-criteria classification) and analysis of the major risk factors specific to the target workplace and domain (elements/ conditions that substantially increase the occurrence probability of accidents and diseases).
- Instantiation of the predefined descriptors specific to the risk factors; of the hierarchical relationships between the risk factors and with concepts in other taxonomies; of the risk evaluation criteria; of the risk control and prevention measures. Each concept type will have a predefined descriptor, composed of predefined and type-specific attributes that should be instantiated for each concept added in the corresponding taxonomy.

The identification of the workplace and domain-specific major events (types of accidents and diseases) that can occur during each activity/ task execution;

The correlation of the identified workplace and domain-specific activities, risks and events with concepts in other taxonomies in the reference ontology.

For each domain-specific concept, the needed knowledge and reasoning for the risk prevention will include:

- Rules for the composition and analysis (verification of the query semantic completeness and correctness) of the training requests (queries);
- Rules for the risk evaluation and for the search and discovery of the appropriate documents and actions for the training in the requested context;
- Rules for the navigation on taxonomies and models;
- Inheritance rules for the concept attribute values inside each taxonomy. For any taxonomy with a predefined descriptor, the values of some generic description attributes will be inherited by a concept from its parent. Other attribute values can be specific to the particular concept and their values are not inherited (e.g. the ID, the concept author, the concept creation date, risk severity, risk probability, workplace condition, etc).
- Inheritance rules for the concept relationships. E.g. the relationship of a risk with a certain activity/ workplace can be inherited by the subtypes of the respective risk.
• Rules for ordering the query results depending on certain attribute values for the involved concepts.

Two kinds of queries will be allowed for training on risk prevention: queries based on the model for risk prevention or queries based on user-defined concepts and relationships. Also, the requestor will be allowed to select the target repository for search: either the system's repository (including other predefined servers with information provided by risk experts or training companies); or, the Web.

For the model-based query, the requestor will follow the steps below. The steps 5 and 6 will enclose a specific reasoning, briefly described below.

1. **System**: Displays the model (similar to Figure 2) and the request for information about:
   - The concepts types (and taxonomies) involved in the query as search (or source) concepts (generic concepts from the model, for which the requestor will later provide values/instances);
   - The concepts types (and taxonomies) involved in the query as requested (or target) concepts (generic concepts from the model, for which the requestor asks for instances after the query execution);

2. **Requestor**: (optional) Selects the search concepts (e.g. Process, Activity, Workplace, Work_Instrument); and/or (mandatory) Selects the requested concepts (e.g. Risk, Event, Document, Preventive_Action); Submits the information.

3. **System**: Opens the query editor and asks the requestor to compose the query with instances of the search concepts.

4. **Requestor**:
   - *(if he/she has specified search concepts)* Selects from the corresponding search taxonomies the domain-specific concepts he wants to be searched for (e.g. the domain-specific process, activity, workplace and work instrument); Selects the logical/grouping operators between the search concept instances (AND, OR, , ).
   - *(if he/she did not specify search concepts)* Fills the query with his own concepts correlated by logical/grouping operators;
   - Submits the query.

5. **System**: For each search concept: Navigates the corresponding taxonomy and adds in the query all its subtypes (grouped and correlated by OR);

6. **System**: Executes the query by composing the query result based on the model as follows:
   - Includes in the result (correlated by AND) all instances of the requested concepts that have direct relationships with the instances of the search concepts in the query. E.g. the instances of the requested concepts Risk, Event, Document, Preventive_Action that have been associated by the risk designers with the specific process, activity, workplace and work instrument given in the query. In Figure 2, there are direct relationships Activity-Risk, Workplace-Event, Work_Instrument-Event. So, the corresponding events and risks will be inserted in the result.
   - If certain requested concepts do not have direct relationships with the search concepts (e.g. Document and Preventive_Action), the system will search for their direct relationships with the requested concepts having already found instances. For the example above, according to the model, the instances of the Document and Preventive_Action can be searched in relationship with the identified instances of the concept Risk, which is in direct relationship with the two requested concepts.
   - If there is a requested concept without any direct relationship with a search concept or with another requested concept having instances already identified, the system (1) will search in the model for an intermediate concept type directly correlated with a search concept; and, (2) will ask the requestor to provide an instance of the respective intermediate concept to be added to the query. In Figure 2, there is no example for this situation, but it must be considered in the system reasoning.

System groups and orders the results depending on the relationships between the requested concepts and the search concepts.

**4 Conclusion**

The system architecture was adapted to a semantics-based view on the risk prevention. Its interface dedicated to the domain experts moves the work for ontology editing and risk design, from IT experts to the domain experts. The system and
its portal will contribute to a knowledge repository for risk prevention inside and cross organizations. It will be accessible from Web and will gradually replace the periodical training in organizations.

References: