*Research Article*

# A Knowledge Management and Decision Support Model for Enterprises

**Patrizia Ribino, Agnese Augello, Giuseppe Lo Re, and Salvatore Gaglio**

*DINFO Department of Computer Engineering, University of Palermo, Viale delle Scienze, Edificio 6, 90128 Palermo, Italy*

Correspondence should be addressed to Patrizia Ribino, ribino@pa.icar.cnr.it

We propose a novel knowledge management system (KMS) for enterprises. Our system exploits two different approaches for knowledge representation and reasoning: a document-based approach based on data-driven creation of a semantic space and an ontology-based model. Furthermore, we provide an expert system capable of supporting the enterprise decisional processes and a semantic engine which performs intelligent search on the enterprise knowledge bases. The decision support process exploits the Bayesian networks model to improve business planning process when performed under uncertainty.

## 1. Introduction

Knowledge management (KM) is a field of computer science, which proposes methods for making knowledge explicit, by representing and sharing information resources. Knowledge management represents, hence, a crucial activity, especially in medium and large enterprises. Such organizations produce great amounts of documents and information during their daily activities. Knowledge management allows companies to avoid losses of relevant knowledge and to increase their productivity, by means of information contents sharing and reusing. Moreover, a correct representation of enterprise knowledge is important for enabling decisional processes and, more generally, for the problem solving. The great attention paid to such issues has lead to the creation of ad hoc frameworks, known as knowledge management systems (KMS), which enable access, coordination, and processing of knowledge assets [1–3].

The amount of knowledge characterizing an enterprise is composed either of structured knowledge, but also, or of several unstructured documents, typically, written in natural

language. In this scenario, it is relevant to design appropriate representation forms, storage methods, and querying procedures of both types of informative contents. Usually, management of structured contents is carried out by relational databases. However, the representation and queries with traditional DB present some limitations, since a fixed and invariable definition of data set correlations and finite sets of keywords can be exploited to make a query. Unstructured informative contents, on the other hand, are composed by natural language texts, typically characterized by semantic ambiguity. For this reason, it results more difficulty to organize, represent, and extract such kind of information, as the process should involve an automatic comprehension of natural language [4].

The proposed model integrates several artificial intelligence (AI) methodologies for the analysis and representation of knowledge in order to deal with both structured and unstructured contents. Structured knowledge has been modeled by means of ontologies. This choice is due both to the flexibility of ontological representation and to the possibility to infer new knowledge from the already formalized one, using reasoning engines. Such a solution allows companies to enrich their knowledge capital and intellectual assets. Furthermore, the system exploits Bayesian decision networks in order to represent uncertainty, and it also provides a module for document management. This last module is based on an unsupervised methodology of analysis and representation of documents which captures latent relations between words.

## 2. Theoretical Background

### 2.1. Knowledge Management Systems

Knowledge management (KM), according to [3, 5], is the discipline which studies methods and techniques that make knowledge explicit and allow the sharing of the professional expertises and informative resources available in the organizations. In such a view, KM uses information technology tools in order to increase the knowledge diffusion and to improve the efficiency of work teams. Knowledge management systems (KMSs) make data, information, and knowledge, coming from different sources, readily available, managing both explicit and tacit knowledge [6, 7]. In order to accomplish such goals, KMSs may exploit different technologies, taken from different fields such as

(i) *document based* for the creation, administration, and sharing of different documents, managing the explicit knowledge;

(ii) *ontology/taxonomy based* using ontologies and classification for representation of knowledge. Knowledge concepts are structured in hierarchical structures describing also relationships between them. Such methodologies act on explicit and tacit knowledge;

(iii) *AI based* using particular inference engines to resolve specific domain problems.

All these techniques generally manipulate tacit knowledge (e.g., Knowledge-base system).

In this work, we propose different approaches for knowledge representation and reasoning. In particular, we consider a document-based approach, relied on a data-driven creation of semantic spaces, and an ontology-based approach in order to formalize the domain of interest. Moreover, we exploit the ability of Bayesian Networks to represent the knowledge affected by uncertainty. The Bayesian networks provide also decision support for a wide

range of problems involving uncertainty and probabilistic reasoning. These methodologies can be integrated in order to obtain hybrid systems which can benefit from the main advantages of the different approaches.

### 2.1.1. Semantic Spaces

Semantic spaces are subsymbolic models, extensively used in information retrieval, which represent the semantic of large textual corpora in high-dimensional vector spaces. The main feature of these models is that they are generally built by means of unsupervised procedures, exploiting the statistical analysis of documents. Semantic spaces rely on the so-called "geometric metaphor of meaning" [8]; it is possible to get information about a word analyzing the context in which it occurs [9].

The space is created analyzing the distribution of words in the corpus, placing those with similar distributions in close space areas. Natural language elements, such as sentences, sections, and documents, are represented into these models and are related to each other on the strength of the proximity of their corresponding vectors. In this context, documents are seen as "topical units" and words as "topic indicators." In particular, latent semantic analysis (LSA) is a technique that is generally adopted in semantic space representation. The strength of LSA is an induction-dimension optimization obtained through the truncated singular value decomposition (TSVD) applied on a words-documents co-occurrences matrix. The TSVD allows the reduction of the matrix sparseness obtaining a new representation which captures indirect, higher-order associations between words [10].

### 2.1.2. Ontologies

In computer science, the use of the term "ontology" means the study of the "beings," the fundamental categories of which all is composed and their relationships. They are adopted to formulate an exhaustive and rigorous conceptual scheme of a particular application domain [11]. The noteworthy entities, the existing relationships between them, the rules, the axioms, and the specific domain constraints are formalized in a hierarchical structure. The use of ontologies allows for information reusing and sharing by means of a common vocabulary; as a consequence, a reduction of conceptual and terminological ambiguities and the integration of different domains into a coherent framework are obtained. Moreover, ontologies can be used to perform an intelligent, knowledge-enhanced retrieval.

### 2.1.3. Bayesian Networks

A Bayesian network (BN) [12], also called belief network, is a graphical model used to describe probabilistic dependencies among the variables of a given domain. In recent years, BNs have been successfully used in many fields such as data mining [13] and decision support system [14]. A BN is a directed acyclic graph, where the nodes represent a set of stochastic variables, and the directed edges connect pairs of nodes representing cause/effect relationships. Each node is enriched with a probability table which specifies the conditional probability of each state given its parents. The relationships between the variables follow the Bayes' rule:

$$p(y \mid x) = \frac{(p(x \mid y) * p(y))}{p(x)}. \tag{2.1}$$

A Decision network is a BN where the vertices represent decision and utility nodes. A decision node defines a finite set of the different possible choices to follow in order to achieve the desired goal. Utility nodes represent a preference level associated with the possible choices.

Let $C = c_1, \ldots, c_n$ be a set of mutually exclusive choices. The utility of a choice depends on the state of some variables. Let m be the number of associated random variables. The goal is achieved by optimizing the expected utility function (EU) that estimates the preferences among the possible states of the world. The expected utility bound to the $i$th choice $c_i$ is given by

$$EU(c_i) = \sum_{j=1}^{m} U(c_i, V_j) P(V_j \mid c_i), \tag{2.2}$$

where $U(c_i, V_j)$ is the utility value for the $i$th choice associated to the state of the $j$th stochastic variable, and $P(V_j \mid c_i)$ represents the probability of $V_j$ conditioned by the $i$th choice.

## 3. State of the Art

During the last decade, several research activities on knowledge management and decision support have faced issues regarding new strategies, tools, and systems in order to organize, store, and share data and know-how; at the same time, the biggest ICT (information and communication technology) companies have exhibited increasing interest in the development of internal knowledge management instruments, using novel data representation models and involving modern AI techniques.

A web and ontology-based KMS, called WAICENT (World Agriculture Information Center), has been adopted by the United Nations Food and Agriculture Organization, in order to improve food security through information exploitation [15]. Sui [16] and Daneva et al. [17] analyzed benefits of a decision support system within the business management. In the field of the decisional systems, Fenton et al. [18] and Noothong and Sutivong [19] investigated the usage of Bayesian networks in order to take decisions on software projects.

Jung et al. [20] propose an intelligent retrieval system for enterprises. They use a Bayesian network to constantly update information stored in the ontology repository.

Several sources propose methodologies of information analysis and extraction from enterprise documents. For instance, in the musing project, an ontology-based information extraction is proposed [21]. In particular, a specific module annotates documents detecting information specified in a domain ontology; the annotations are then used by an ontology population module. The extracted information is used in financial and ICT operational risk management applications.

Cumby et al. [22] propose an information extraction module, based on a vectorial representation of semistructured documents, capable of extracting and ranking a multiplicity of enterprises semantic entities, such as employees, companies, clients, and projects. Other architectures for document analysis and understanding propose the use of Prolog sets of rules as in [23, 24].

A preliminary version of our system is described in [25], where firstly the ontological approach was proposed.

In this paper, we propose a knowledge management system for decision support. Improving the results obtained in [25], we implement a novel fusion of artificial intelligent techniques in order to improve the reliability of decisional processes and the precision of

information retrieval. Differently from the above mentioned works, the goal is achieved by means of a different knowledge representation which allows more complex reasonings.

The proposed system and the relative application context are thoroughly described in the following sections.

## 4. Application Context

A big challenge for the business environment is constituted by the management of the organizational knowledge (the knowledge owned by an organization, it is a fusion of individual expertise, know-how, documents, etc.) in a coherent and productive way. Some useful and important knowledge is usually owned unconsciously by organizations. It can be found in the experience of organization members, into organizational processes, within files or documents.

Such a knowledge represents a value and may be used in several different ways, for instance, to estimate costs and necessary resources for new products or services, to find similarities in different application contexts, or to increase the reuse of previous projects.

In this work, the application scenario for the proposed KMS is represented by an ICT society whose main goal is the development of software systems for the process automation of government agencies. In such a company, as well as in many others operating in the same areas, employees can assume different organizational roles. Among those, the most important are the *program* and *project manager* roles. A program manager maintains a global vision of the entire organization and of its products, while a project manager leads a group of people knowing only the work of his team.

The company is therefore organized in a collection of independent groups, each one with a project manager. Different teams may replicate the same activities and develop the same semifinished products already done by another team. The reuse is minimal, and there is not any effort for knowledge management and sharing of experiences. This situation produces knowledge as islands that could not be shared among teams, with a limited capability of learning from previous knowledge and without any unified vision.

However, each ICT project is described and documented by several technical writings and documents. Usually each document contains explanation of the used technology, description of the system, financing sources, responsibility information, processes to be automated, and so on. The number of pages that compose a generic technical writing ranges from tens to thousand pages for each project.

In such a scenario, it was noted that

(i) the lack of an intelligent system, able to manage the great amount of knowledge (such as documents, technical reports, and mails) produced during organizational activities, causes a great waste of time to find useful information;

(ii) the lack of communication and knowledge sharing among workers causes a restricted view of corporate know-how;

(iii) a program manager operates without any kind of support in decision-making processes.

For these reasons, the adoption of an integrate system that effectively supports the organization workers in their business activities, and the use of smart engines for decision making, can improve organizational performance.

Making the organizational knowledge accessible to all those who can benefit from its application is unfortunately a difficult task due to the different forms as knowledge is stored.

The main goal of the proposed organizational framework is to manage different kinds of knowledge in order to provide a useful tool for business decision support and to improve the organizational productivity. In order to achieve this result, we propose a novel fusion of artificial intelligence techniques capable of dealing with tacit and explicit knowledge.

## 5. System Architecture

The system we implemented is a KMS for enterprises and specifically for ICT organizations. It is composed by the following components, as shown in Figure 1: a *knowledge base* (KB), a *knowledge management component* (KMC), an *ontology editor* (OE), a *document repository* (DR), and some graphical user interfaces (GUIs). The core of the system is the KMC, which is further constituted by two subsystems: an inference engine and a semantic engine.

The system provides two main functionalities: intelligent search and decision support. The search exploits the potential of two different but related knowledge representation models, namely, the ontologies and the semantic spaces. On one hand, search is performed by means of an inference engine used to interact with the ontology and a retrieval mechanism based on semantic spaces. On the other hand, the decision support exploits Bayesian networks to improve enterprise planning process under uncertainty.

### 5.1. Knowledge Base

The knowledge base (KB) is the knowledge repository, where the relations and concepts are described using an ontological framework of instances in order to collect and manage data [26, 27]. The KB is built using the Protégé system, a free and open source platform developed at Stanford University [28].

The considered knowledge concerns topics relevant to the practice of software designing and the definition of the government offices structure. Such a knowledge is modeled by means on two kinds of different but correlated ontologies that we called *domain* and *project* ontology.

The *domain ontology* is a formal representation of the structure and activities of government's agencies. It is used to characterize the environment in which the system works, and it is organized as a set of concepts and relations; it replicates the logical architecture of government offices arranged in layers, each depending on underlaying ones in a hierarchical fashion.

The *project ontology* is used to describe ICT company projects and their main features.

In this paper, we define a project as a set of formal descriptions that explain how to solve a given problem. Basically, each project is formed by a set of documents that specify its content from different points of view (such as cost analysis, resources employment, requirements, and implementation specification). The project ontology involves the specifications of ICT projects by means of a set of concepts (such as project manager, allocated funded, and software and hardware requirements), and relationships among these concepts (such as *a project has a project manager and a project satisfies a software requirement*).

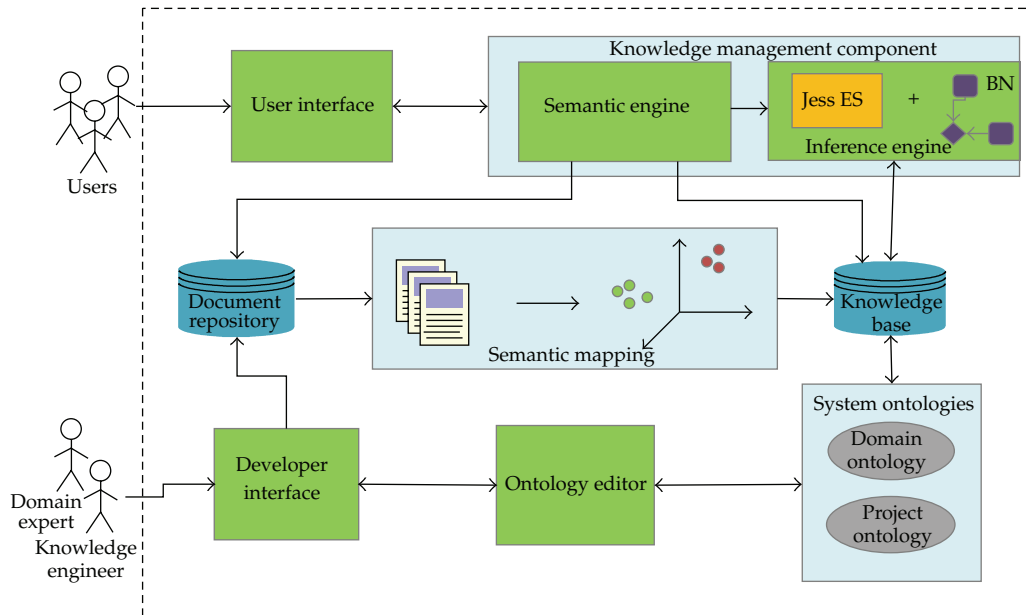Furthermore, a vectorial coding resulting from a semantic mapping of its documents is associated to each project.

**Figure 1:** System architecture.

A more detailed description of knowledge retrieval process will be described in the following subsections.

## 5.2. Document Semantic Mapping

### 5.2.1. Semantic Space Creation

A semantic representation of the enterprise documents was built by the mapping of document contents in a semantic space. The semantic space has been designed according to the latent-semantic-analysis- (LSA-) based approach described in [29, 30].

Authors of these papers extended the theoretical model of LSA suggesting a statistical interpretation. According to the discussion reported in [30], the truncated singular value decomposition (TSVD) can be considered as an estimator that maps a word-document frequency sample in a probability distribution.

The aim is to obtain a sufficient estimator, capable of capturing only the information relevant to the determination of the semantic relationships among language elements, removing the information that are only closely related to the specific sample. It can be shown that such an estimator can be obtained finding the best approximation of the co-occurrences matrix with respect to the Hellinger distance measure [30]. It can be also proved that this can be achieved performing the TSVD after a process of substitution of each matrix element with its square root. According to this definition, we created a semantic space applying TSVD on an $m \times n$ co-occurrences matrix obtained analyzing a specific texts corpus, where each $(i, j)$th entry of the matrix represents square root of the number of times the $i$th word appears in the $j$th document.

The matrix has been created analyzing a large meaningful text corpus coherent with the active topics of the enterprise. A better representation has been obtained using the corpus

of enterprise documents extended with a set of documents related to the enterprise thematic areas. The choice of these documents represents a critical phase, since the quality of the corpus determines the effectiveness of the space. The Wikipedia repository [31] can be used for the retrieval of documents related to enterprise themes; the topics are used as keywords, and a relevance threshold, experimentally fixed, is used.

After the decomposition, we can obtain a representation of words and documents in the reduced semantic space. To evaluate the similarity between two vectors $v_i$ and $v_j$ belonging to this space, it is possible to use several measures, such as the cosine or the scalar product between the vectors. If we normalize the obtained vectors, we can consider them as probability distributions. Therefore, to remain consistent with the statistical interpretation of the semantic space, according to Agostaro [30], we normalize the similarity measure in order to obtain a probability. The similarity measure is defined as follows [30]:

$$
\text{sim}(d_i, d_j) = \begin{cases} \cos^2(d_i, d_j) & \text{if } \cos(d_i, d_j) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}
$$

The closer the above value to 1, the higher is the similarity degree.

### 5.2.2. Project Features Extraction and Semantic Mapping

Each project $P$ is characterized by a set of features, $F_1, F_2, \ldots F_n$. Examples of these information features are the project code, its title, the project manager, the subject, its goals and functionalities, a summary, and so on. A project can also be described by several documents. As mentioned in the previous sections, projects are formalized in the ontology along with their properties.

In this work, we automatically associate to each project a vectorial representation in the semantic space, inferring also its thematic area. Each project is described by different documents, where each document is composed by a set of words, having a correspondent vector encoding into the semantic space. Therefore, a vectorial representation of a document can be obtained evaluating the sum of the vectors associated to the words composing it, and similarly, a representation of a project is given by the sum of vectors associated to the documents describing it.

Moreover, documents and projects can be classified with respect to an enterprise thematic area, by means of the nearest neighbors algorithm. The classification is obtained comparing the document or project vectorial representation to a set of classified samples, given by the documents used to build the space. Given a project to classify, its vectorial representation $p$ is compared with the set of documents $d_j$ with $j = 1, \ldots, n$ used to build the space. The comparison is made by means of the similarity measure reported in (5.1), cutting off all results which do not exceed an experimentally fixed threshold. A set of $n'$ documents is therefore obtained. The thematic area related to the project will be the same of the closest document $d_x$, where $(\text{sim}(p, d_x > \text{sim}(p, d_j), \text{ for all } j = 1, \ldots, n')$. The calculated thematic area will be associated to the project, inserting an ad hoc relationship into the ontology. If any value does not exceed the threshold, then it can be inferred that the training set is not complete enough, or the project belongs to a new thematic area. In this case, the thematic area is manually assigned.

### 5.3. Semantic Engine

The semantic engine allows users to search projects or documents into the semantic space. Namely, the document closer to a particular query, will be returned through the graphic user interface. If the search is performed in order to retrieve a project, the comparison can be limited by considering only document representations associated to the projects. Comparison is performed evaluating the semantic similarity measure, defined in (5.1), between an element such as a document, a project, or a query, and the basic elements of the space. The semantic search carries out the similarity metric computation between the vectors identifying the elements to compare; such distance measures how much semantically related are the involved entities. For instance, users could search the following

    (i) projects or documents related to a specific thematic area;

    (ii) projects or documents closer to a specific query expressed in natural language;

    (iii) projects or documents similar to a given project.

If the elements to be compared are already coded in the semantic space, the comparison can be performed straightforwardly. Otherwise, if the item represents a user query in natural language, or a new element, it should be first coded in the space. The procedure will retrieve all those vectors with similarity distance from the query vector exceeding a given threshold. The user can inspect the results ordered by relevance to the query, and select those that he deems well matching his request.

### 5.4. Inference Engine

The goal of an expert system (ES) supporting decisions mainly consists of assisting workers during their activities in order to find solutions, that usually need the intervention of specifically skilled people.

The proposed ES is implemented through the joint adoption of two technologies, rule-based systems, and decision networks in order to face two kinds of problems concerning ICT enterprises: software reuse and project selection.

*Software Reuse* is not only a technology problem, but basically a knowledge management problem. Reuse can be defined as further use or repeated use of an asset (e.g., software components). A new product can be created choosing applicable assets from an "asset base" (a repository of all enterprise assets). A correct knowledge management allows for finding best assets candidate for reuse.

*Project selection* concerns the choice of the best proposal on the basis of a cost benefit analysis. In order to decide which of the proposed projects should be selected, a number of factors must be considered. Namely, each project is characterized by its own complexity, environmental advantages and disadvantages, tangible and intangible benefits, costs, allocation of human and hardware resources, and many others features.

The following subsections illustrate how we use a rule-based system and a decision network in order to address the above issues.

#### 5.4.1. Ontology Query for Knowledge Reuse

In a rule-based expert system, the knowledge is coded into a set of rules, each representing a small portion of expert's skills. Each rule is, generally, an "if-then" statement. The ES

prototype implemented for our framework uses a rule-based system developed in Jess (Java Expert System Shell) [32].

Such system component is built to assist project managers during their activities and to find solutions that can improve software component reuse. More specifically, our ES can offer support for evaluation of previous projects' functionalities reuse. The problem addressed is essentially to carry out complex ontology queries in order to find software requirements matching the user demands.

An ICT organization usually develops products capable of automatizing workers activities. A process can be seen as a set of interrelated activities, grouped in phases. Therefore, each ICT product (defined into a project) is composed by a set of components. Each component generally supports a singular phase of an entire process. In our study case, the ICT organization develops products in order to automatize process performed by government agencies.

There are several government agencies which perform common activities within different process. Software reuse can be optimized using components previously developed.

The ontology query module may perform not only simple queries (such as the name of the responsible of a given project and the features of a given office) but also complex queries that allow to infer new knowledge exploring direct and indirect relationships between concepts of the ontological model. For instance, it is possible to investigate which offices have common tasks or to search which software components could be used to automatize a particular activity, and so forth.

The JESS code shown below is a portion of a rules set used by our ES (see Algorithm 1).

The first part expresses a portion of the ontological concept of a Sicilian government office with its attributes and relationships. Namely, a Sicilian government office is identified by a name; it belongs to a specific agency; it can provide several services; it has several related projects and so on.

The second part of the reported code illustrates how the ES performs the query, exploring the chain of relationships that directly or indirectly are related to the "office concept."


### 5.4.2. A Bayesian Network for Decision Support

A project-planning process is a typical decision problem. In complex and dynamical environments, such as the marketing context, it is difficult to take decisions since there are several factors to be considered. Incautious decisions could produce negative effects on future incomes for the enterprise. Making the right decisions is a fundamental task for enterprise survival. Moreover, the greatest part of the factors influencing the project planning process (such as time, costs, and resources) are sources of uncertainty that must be opportunely evaluated in order to optimize the decisional task.

In the last years, BNs have become a popular representation for encoding uncertain knowledge into expert systems [33].

In this section, we present a model of Bayesian network in order to establish which product (i.e., software system) for the ICT enterprise is more convenient to develop by means of a costs/benefits analysis. The development of the network has been divided in three phases: domain analysis, relationships discovery among the variables of interest, and probability tables estimation. The obtained decision network is shown in Figure 2.

```
   (DEFCLASS office
//represent a general company offices
(multislot linked_projects
(comment "projects of the department"
(type INSTANCE))
(single-slot name
(type STRING)
(cardinality 0 1))
(single-slot council_of_competence
(type INSTANCE)
(allowed-classes COUNCIL)
(cardinality 0 1))
(multi-slot comprehend_services
(type INSTANCE)
(allowed-classes SERVICE)
. . .
(defrule office_processes ?instance
<-(object (is-a OFFICE)
(name ?n&:(call ? *ric* equals
(slot-get "+ off+" name)
(lowcase ?n))))=>(bind $?area
(slot-get (instance-name ?instance)
comprehend_areas
(foreach ?j $?area
(if (call ?*ric* different
(slot-get (instance-name ?j) name) empty)
then . . .
```

**Algorithm** 1

*Domain Analysis*

A set of stochastic variables characterizing the domain of interest and their possible states are defined as follows.

(i) *Project complexity (PC)* = measures how the project is complex. States values: high, medium, and low.

(ii) *Founds (Fs)* = available financial support in million of Euro (MEUR): States values: less than 1 MEUR, between 1 and 3 MEUR, and more than 3 MEUR.

(iii) *Human resources (HRs)* = availability of employees. States values: full time and part time.

(iv) *Specific skills (SSs)* = availability of human resources with specific technical know-how. States values: high, medium, and low.

(v) *Develop functionalities (DFs)* = project functionality that can be implemented. States values: many, on average, and few.

(vi) *Complete description (CDs)* = if the customer has provided an exhaustive description of the product. States values: true and false.

(vii) *Customer changes (CCs)* = if the customer can bring changes to the requirements of the product during the working progress. States values: true and false.
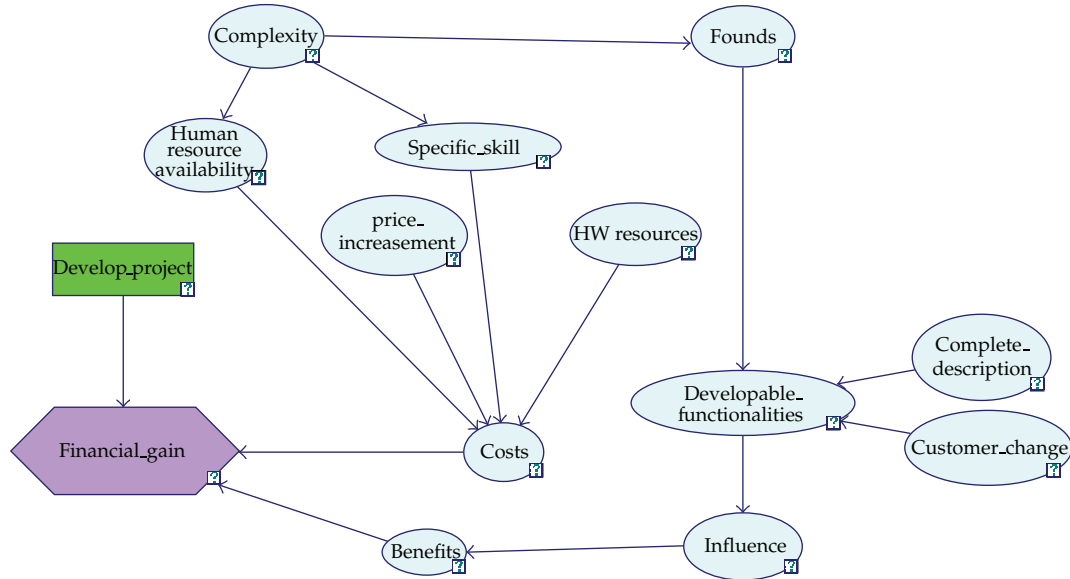
**Figure 2:** Projects decision network.

(viii) *Influence (I)* = how much the project is important. States values: many, on average, and few.

(ix) *Costs (Cs)* = total costs for the development of the project. States values: high, medium and low.

(x) *Price increment (PI)* = prices growing (raw material, renewal employment contract etc.). States values: true and false.

(xi) *HW resources (HW)* = measures the how much hardware is necessary to project development. States values: high, medium, and low.

(xii) *Benefits (Bs)* = improvement of human process, of time and costs. States values: high and low.

(xiii) *Financial gain (FG)* = It contains information about the decision maker's goals.

(xiv) *Development project (DP)* = it models decision maker's options.

*Relationships Discovery*

The relationships discovery phase allows for the discovering of causal relationships between the variables of our observation. In our modeled domain, the following set of dependence and independence conditions were discovered:

$$PC, PI, CD, CC, HW \mid \emptyset; \quad F, HR, SS \mid PC;$$

$$C \mid HW, PI, HR, SS; \quad DF \mid SS, CD, CC; \quad B \mid I; \quad I \mid DF. \tag{5.2}$$

The symbol | indicates the dependence of the set of variables on the right side by the set of variables on the left side.
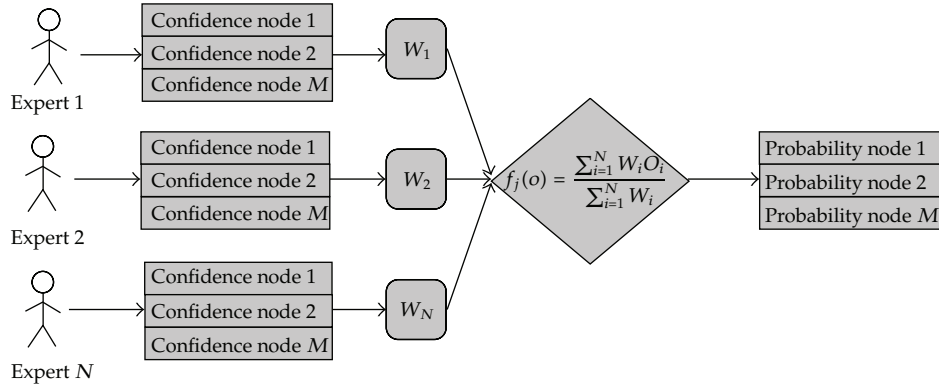
**Figure 3:** Fusion schema of experts' opinion.

*Probability Tables Definition*

Building the tables of the conditional probability distributions for each node of the BN was one of the difficult tasks to carry out. Making a correct evaluation of the confidence of an event is a "human sensitive" task. Namely, a decision maker generally maintains a personal opinion about a problem to solve, rather than an objective judgment about the evaluation of the criteria influencing the particular problem. Moreover, since the human opinions are based on individual capabilities and experiences, several different opinions are possible on a particular event occurrence.

The schema shown in **Figure 3** was adopted in order to consider the different opinions on the same event expressed by different domain experts. The weighted average method (see (5.3)) is used to merge these individual contributions in order to obtain a single final opinion expressed as a probability. We defined a metric to quantify the credibility of each expert on the basis of his experience and the work years (see (5.4)).

The final opinion about node $j$, produced by $N$ experts, is

$$
f_j(o) = \frac{\left(\sum_{i=1}^{N} w_i o_i\right)}{\left(\sum_{i=1}^{N} w_i\right)}, \tag{5.3}
$$

where $O_i$ indicates the percentage value of the opinion of the expert $i$, while $w_i$ is the weight (credibility) associated to $i$th expert. $w_i$ is calculated it follows:

$$
w_i = \frac{G_i * A_i}{(G_{\max} * A_{\max})} \quad \text{with } G \in [1, G_{\max}], \ A \in [1, A_{\max}], \tag{5.4}
$$

where $G$ indicates the degree of experience of the expert in the company, and $A$ the number of years he has worked for the company.

### 5.5. Semantic and Inferential Engines Integration

The proposed methodologies complement each other. Their integration allows the system to combine several results, obtained by means of intelligent searches and reasoning processes.

　　The fusion of semantic and inference engines can provide a more efficient and effective knowledge retrieval. This way, it is possible to perform ontology queries driven by search in the semantic space: for instance, the system can first find a project related to a textual query using the semantic engine, and then it can retrieve information associated to such project (as project manager, allocated funded, costs, etc.) by means of the inferential engine. This can be useful if the user does not know the name of the project he is looking for, or if he has to write a new project and wants to explore the enterprise KB in order to find related projects and to reuse parts of the related documentation. Similarly, the system can perform an ontology query, in order to retrieve specific information about a particular project. Then it can expand the research through the semantic engine to obtain other results such as, for instance, a list of related projects or topics. This allows the system to expand the research analyzing features not explicitly modeled in the ontology, but that are automatically inferred by the system.

　　Furthermore, the integration of logical and probabilistic reasoning allows the system to obtain effective results, even when there is not a complete and deterministic knowledge of the domain. As an example, if we want to determine the feasibility of a project within an estimated deadline, we make a query to the system. Then, some information related to the project are set as evidences into decisional Bayesian network. A probabilistic reasoning is performed to evaluate the probability to meet the given deadline. If such probability results are too low, it is possible to obtain further suggestions on possible changes to adopt in order to respect the timing constraints.

## 6. Conclusion

In this work, we proposed a knowledge management and decision support model for ICT enterprises. The main feature is the integration of two well-known techniques in order to improve the ability of decision support systems. The combination of different knowledge representation and reasoning methodologies is a current trend in intelligent systems search and is performed in order to implement different capabilities in a system. Currently, at the best of our knowledge, it is unusual to find the fusion of these methodologies in decision support systems for enterprises. Nevertheless, artificial intelligence methodologies, and moreover their integration, can strongly improve knowledge management and decisional processes in enterprises.

　　The main advantage of the proposed model is its adaptability to different kinds of informative contents. The explicit knowledge, coming from structured information or natural language documents repositories, can be easily managed. Both the tacit knowledge and the reasoning capabilities owned by the domain experts can be represented by means of models such as Bayesian networks and ontologies.

　　Ontologies compared to traditional relational databases allow the interoperability between heterogeneous knowledge sources by means of a common vocabulary. Since the ontological knowledge representation is "concepts oriented" rather than "data oriented," a better usability of information can be obtained. The proposed system also includes an expert system which represents a valid support tool for enterprise decision makers. In our specific context, it allows for searching and reusing components previously developed in other ICT

projects, and it is able to suggest the best strategies for projects planning. In addition, the possibility to query the system trough natural language queries improves its usability. The evolution of this work will involve the development of a module of information extraction from textual documents in order to automatically detect the main components of the business projects.

## Acknowledgments

## References

[1] S. Staab, R. Studer, H. P. Schnurr, and Y. Sure, "Knowledge processes and ontologies," *IEEE Intelligent Systems and Their Applications*, vol. 16, no. 1, pp. 26–34, 2001.

[2] D. E. O'Leary, "Enterprise knowledge management," *IEEE Computer*, vol. 31, no. 3, pp. 54–61, 1998.

[3] M. Alavi and D. E. Leidner, "Review: knowledge management and knowledge management systems: conceptual foundations and research issues," *Management Information Systems Quarterly*, vol. 25, no. 1, pp. 107–136, 2001.

[4] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison Wesley, 1999.

[5] T. H. Davenport, D. W. De Long, and M. C. Beers, "Successful knowledge management projects," in *The Knowledge Management Yearbook 1999-2000*, J. Cortada and J. A. Woods, Eds., pp. 89–107, 1999.

[6] H. Takeuchi and I. Nonaka, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press, New York, NY, USA, 1995.

[7] I. Nonaka, "A dynamic theory of organizational knowledge creation," *Organization Science*, vol. 5, no. 1, pp. 14–37, 1994.

[8] M. Sahlgren, *The Word-Space Model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*, Ph.D. thesis, Department of Linguistics, Stockholm University, 2006.

[9] W. Lowe, "Towards a theory of semantic space," in *Proceedings of the COGSCI*, pp. 576–581, Lawrence Erlbaum Associates, 2001.

[10] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge," *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.

[11] N. Guarino, "Formal ontology in information systems," in *Formal Ontology in Information Systems—Proceedings of the 1st International Conference (FOIS '98)*, Trento, Italy, IOS Press, Amsterdam, The Netherlands, 1st edition, June 1998.

[12] D. Heckerman, "A tutorial on learning with Bayesian networks," in *Learning in Graphical Models*, M. Jordan, Ed., MIT Press, Cambridge, Mass, USA, 1999.

[13] D. Heckerman, "Bayesian networks for data mining," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 79–119, 1997.

[14] S. Z. Zhang, N. H. Yang, and X. K. Wang, "Construction and application of Bayesian networks in flood decision supporting system," in *Proceedings of the 1st International Conference on Machine Learning and Cybernetics*, pp. 718–722, 2002.

[15] D. E. O'Leary, "A multilingual knowledge management system: a case study of FAO and WAICENT," *Decision Support Systems*, vol. 45, no. 3, pp. 641–661, 2008.

[16] L. Sui, "Decision support systems based on knowledge management," in *Proceedings of the International Conference on Services Systems and Services Management (ICSSSM '05)*, vol. 2, pp. 1153–1156, June 2005.

[17] M. Daneva, J. Peneva, R. Rashev, and R. Terzieva, "Knowledge-based decision support system for competitive software audit," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 1974–1979, October 1995.

[18] N. Fenton, W. Marsh, M. Neil, P. Cates, S. Forey, and M. Tailor, "Making resource decisions for software projects," in *Proceedings of the 26th International Conference on Software Engineering (ICSE '04)*, pp. 397–406, May 2004.

[19] T. Noothong and D. Sutivong, "Software project management using decision networks," in *Proceedings of the 6th International Conference on Intelligent Systems Design and Applications (ISDA '06)*, vol. 2, pp. 1124–1129, IEEE Computer Society, Washington, DC, USA, 2006.

[20] M. Jung, H. B. Jun, K. W. Kim, and H. W. Suh, "Ontology mapping-based search with multidimensional similarity and Bayesian network," *The International Journal of Advanced Manufacturing Technology*, vol. 48, pp. 367–382, 2010.

[21] H. Saggion, A. Funk, D. Maynard, and K. Bontcheva, *Ontology-Based Information Extraction for Business Intelligence*, vol. 4825/2007, pp. 843–856 of *Lecture Notes in Computer Science*, 2007.

[22] C. Cumby, K. Probst, and R. Ghani, "Retrieval and ranking of semantic entities for enterprise knowledge management tasks," in *Proceedings of the Workshop on Semantic Search (SemSearch '09)*, 2009.

[23] D. Niyogi and S. N. Srihari, "Using domain knowledge to derive the logical structure of documents," in *Proceedings of the SPIE*, pp. 114–125, 1996.

[24] A. Dengel and G. Barth, "ANASTASIL: a hybrid knowledge-based system for document layout analysis," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '89)*, pp. 1249–1254, 1989.

[25] P. Ribino, A. Oliveri, G. L. Re, and S. Gaglio, "A knowledge management system based on ontologies," in *Proceedings of the IEEE International Conference on New Trends in Information and Service Science (NISS '09)*, pp. 1025–1033, July 2009.

[26] M. Minsky, "A framework for representing knowledge," Tech. Rep., Massachusetts Institute of Technology, Cambridge, Mass, USA, 1974.

[27] V. K. Chaudhri, A. Farquhar, R. Fikes, P. D. Karp, and J. P. Rice, "OKBC: a programmatic foundation for knowledge base interoperability," in *Proceedings of the 15th National/10th Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence (AAAI '98/IAAI '98)*, pp. 600–607, American Association for Artificial Intelligence, Menlo Park, Calif, USA, 1998.

[28] Resources available at, http://protege.stanford.edu/.

[29] F. Agostaro, G. Pilato, G. Vassallo, and S. Gaglio, "A sub-symbolic approach to word modelling for domain specific speech recognition," in *Proceedings of the 7th IEEE International Workshop on Computer Architecture for Machine Perception (CAMP '05)*, pp. 321–326, July 2005.

[30] F. Agostaro, *Metriche per l'Analisi della Semantica Latente finalizzata ai Modelli del Linguaggio*, Ph.D. thesis, Dipartimento di Ingegneria Informatica, Università degli Studi di Palermo, 2006, S. Gaglio, supervisor.

[31] Resources available at, http://www.wikipedia.org/.

[32] E. Hill, *Jess in Action: Java Rule-Based Systems*, Manning Publications, Greenwich, Conn, USA, 2003.

[33] D. Heckerman, A. Mamdani, and M. P. Wellman, "Real-world applications of Bayesian networks," *Communications of the ACM*, vol. 38, pp. 24–26, 1995.