We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Intelligent Knowledge Retrieval from Industrial Repositories

Antonio Martin, Mauricio Burbano and Carlos León

Additional information is available at the end of the chapter

Abstract

Actually, a large amount of information is stored in the industrial repositories. Accessing this information is complicated, and the techniques currently used in metadata and the material chosen by the user do not scale efficiently in large collections. The semantic Web provides a frame of reference that allows sharing and reusing knowledge efficiently. In our work, we present a focus for discovering information in digital repositories based on the application of expert system technologies, and we show a conceptual architecture for a semantic search engine. We used case-based reasoning methodology to create a prototype that supports efficient retrieval knowledge from digital repositories. OntoEnter is a collaborative effort that proposes a new form of interaction between users and digital enterprise repositories, where the latter are adapted to users and their surroundings.

Keywords: artificial intelligence, ontology, semantic web, experts systems, CBR, interoperability

1. Introduction

Nowadays, industrial information provides effective knowledge of existing resources through databases and repositories, which provide details on the hosted equipment, including information on their capacity, performance, start/stop dates, turbine and generators models, etc. All of these properties and information are stored in digital repositories, digital files, and business websites. To collect, contribute, and share the knowledge about the resources installed in the industrial area, online databases named digital industry repository (DIR) is used. Therefore, the way in which information and knowledge stored in digital repositories is retrieved is of vital importance. DIRs provide centralized hosting and access to content, establish permissions, and

controls for access to content, the ability to share digital objects or files, to protect the intellectual, and integrity property rights of content owners and creators, etc.

So far, traditional search engines treat the information as an ordinary database that manages the content inefficiently. Current search engines retrieve the information by comparing the contents of the database with searched patterns. The generated result is a list of data that contain this patter. Although search engines are becoming more effectively, information overload hinders search and accurate knowledge retrieval. Consequently, it is necessary to develop new semantic and intelligent models that contribute new possibilities. The presented work offers a new approach to the information retrieval based on semantics and intelligent models. For this, the case-based reasoning (CBR) technique is applied contributing to the goal of improving knowledge recovery in the industrial field.

A significant number of researchers have already investigated the application of intelligence and semantic techniques, but just a few from the point of view of full integrated of both technologies an industrial environment. There are researchers and related field works that include ontology retrieval methods such as [1] present a system that uses an ontology query model to analyze the usefulness of ontologies in effectively performing document searches and proposes an algorithm to refine ontologies for information retrieval tasks with preliminary positive results. In this paper [2], real-time image capture was achieved by using digital camera technology and image processing technology. By extracting the glue line curve from image, thinning glue curve by morphological method, and extracting the frame information, the closure and quality of the glue curve can be detected. Results of test show that the effect is satisfactory and the method is effective. The major contribution of [3] is a novel semantic query expansion technique that combines association rules with ontologies and Natural Language Processing techniques, which utilizes the explicit semantics as well as other linguistic properties of unstructured text corpus. It makes use of contextual properties of important terms discovered by association rules, and ontology entries are added to the query by disambiguating word senses.

Semantic Web utilizes concepts, taxonomic relations, and nontaxonomic relations in a given domain ontology to capture knowledge efficiently. For example, [4] describes one component of a knowledge management platform with a multiagent search module (MASH), which employs domain ontology to search for Web pages that contain relevant information to each concept in the domain of interest. The search is then constrained to a specific domain to avoid as much as possible the analysis of irrelevant information. Ref. [5] expounds the function of each layer and analyses the implementation of this system from the knowledge organization and expression and knowledge retrieval and proposes a framework of knowledge management system based on ontology. This management system establishes a sharable ontology that can be understood both by human and computer, which people can found more relations of different concepts through a better circumstance of knowledge retrieval interface. The work of [6] proposes an ontology-based user model, called user ontology, for providing personalized information service in the Semantic Web which utilizes concepts, taxonomic relations, and nontaxonomic relations in a given domain ontology to capture the users' interests. The research of [7] presents a semantics-based digital project which provides faceted search and represents a novel

approach to Digital Libraries, integrating social Web and multimedia elements in a semantically annotated repository. In other investigation, Ref. [8] describes the architecture of the dynamic retrieval analysis and semantic metadata management system (DREAM) designed to automatically and intelligently index huge repositories of special effects video clips, based on their semantic content, using a network of scalable ontologies to enable intelligent retrieval. Ref. [9] presents an information search and retrieval framework based on the semantically annotated multifacet product family ontology. The major contribution of [10] is an innovative comprehensive semantic search model, which extends the classic information retrieval (IR) model, addresses the challenges of the massive and heterogeneous Web environment.

There are a lot of researchers on applying these new technologies into current information retrieval systems, but no research addresses artificial intelligence (AI) and semantic issues from the whole life cycle and architecture point of view [11]. This article analyses the search methods efficiency in a distributed data space such as industrial information repositories. The paper presents an intelligent proposal to optimize search engines in a specific industrial domain and to focus our discussion on the indexing and retrieval strategies of cases and provides the application technical aspects. This paper describes the current problems of semantic interoperability and proposes an intelligent method to address them. To do these, technologies based on metadata and intelligent techniques are used. The main proposal goal is the intelligent search management in decentralized industrial repositories, where no global information scheme exists. The most important novel introduced by this proposal is that contextual user profiles are built based on ontologies and metadata facilitating the ontological search using expert systems technologies. The objective has focused on creating technologically complex environments industrial domain and incorporates Semantic Web and AI technologies to enable precise location of industrial resources [12]. For this reason, we are improving representation by incorporating more metadata from within the information.

We propose a new paradigm to achieve efficient knowledge retrieval from digital repositories. This paper presents an intelligent search engine for industrial process, especially for resources repositories, and proposes an intelligent agent–based personalized model. One major research area is intelligent systems, with the general intention to replace human operators with intelligent agents. We have used CBR methodology to develop a prototype for supporting efficient retrieval knowledge from DIR.

In the following sections, we review the CBR framework and its features for implementing the reasoning process over ontologies. Section 2 presents a general overview about the industrial domain and technology infrastructure, analyzing its failures and discovering the needs that push us toward new intelligent paradigms. Section 3 analyses ontology requirements and proposes the design criteria to guide the development of ontologies for knowledge-sharing purposes. Then, we show the methodology followed to conduct this research, and we describe the semantic-based management system in DIRs environment. Next section concerns the design of a prototype system for semantic search framework, in order to verify that our proposed approach is an applicable solution. Moreover, the functional requirements of the engine and the knowledge base are described in detail. Finally, we present the results of our work on the adaptation of the framework, and we outline the future works.

2. Importance of applying ontology in industrial sector

Our objective here is thus to contribute to a better knowledge retrieval in the industry repositories field. Ontologies are being developed to facilitate knowledge sharing and reuse. In this section, we explain more formally what ontologies are and what problems can arise from knowledge sharing in industrial area. We have proposed a method to efficiently search the target information on a digital repository network with multiple independent information sources [13]. The use of AI and ontologies as a knowledge representation formalism offers many advantages in information retrieval [14]. In this chapter we analyzed the relationship between both factors ontologies and expert systems.

Currently, the electronic search is based mainly on matching keywords introduced by the users with sought data Web pages containing those keywords. The ambiguity of word blends and phrases and the poor linguistic features of Web content indexing mechanisms greatly affect the results obtained from Web resource searches. The efficiency of the results search obtained can vary depending on the quality of the search query from a limited set of results to a too large number of irrelevant results. For certain cases specifying a couple of keywords can be enough, if they are really specific and no ambiguity is possible. In another way, many Web users search for information that cannot be described easily by a set of keywords, and this is due to the wideness of expected results, which cannot be retrieved from existing search engines just with one search query [15].

Industrial repositories contain a large volume of digital information, generally focusing on making their knowledge resources to improve associate decision-support systems. Within a pool of heterogeneous and distributed information resources, users take site-by-site searching. Thus, considerable effort is required in creating meaningful metadata, organizing and annotating digital documents, and making them accessible. The presentation of semantic-enabled resources introduces some benefits of the Semantic Web technology as a possibility to perform a semantic search, integration of heterogeneous data, and use of semantically annotated search results by software. This work concerns applications of the Semantic Web technology for improving existing information search systems by adding semantically enabled extensions that enhance information retrieval from information systems.

A recent comprehensive document covering the main aspects of ontologies in AI research is the technical roadmap of the ontology field in Europe and worldwide produced by the OntoWeb project [16]. In this chapter, we want to emphasize that the first step toward real portability between systems is ontologies. Ontologies can be effectively used to address the problem of global and general models construction between similar domains. Furthermore, it is possible to instantiate and adapt ontology with a specific configuration to automatically build and validate new models [17]. With respect to the research involved in this study, ontologies can provide:

- Share and common understanding of the knowledge domain that can be communicated among agents and application systems.

- Explicit conceptualization that describes the semantics of the data.

The proposal is based on the principle that information items are abstracted to a characterization by introducing metadata, which is used and processed by search engines. This principle is based on a vocabulary/ontology that is shared in order to access the relevant sources of information. This creates new challenges for the research community and motivates scientists to look a recovery approach based on ontologies and intelligent information that automatically search and filter information based on a higher level of understanding required. In this sense, in the present work, we make an effort of investigating techniques that use ontologies to improve the effectiveness in knowledge retrieval. Thus, ontologies are key elements for the definition of the semantic Web [18].

To achieve these objectives, we must consider the interoperability of information. In other words, the ability of different information systems, platforms, and services to share, communicate, and exchange data, information and knowledge effectively and accurately, as well as integrate with other systems, applications, and services in order to deliver new electronic services and products.

Initiatives in this sense such as interoperability between different industrial domains require the establishment of collaborative semantic repositories between private and public sector organizations. Especially, semantic interoperability is necessary, which has a special relevance within the program to support the implementation of distributed services.

3. Challenging the interoperability between systems

Industry and companies are seeking to gain maximum business value from their investments in information and communications technologies. The industry has recognized the ever-increasing importance of systems and software interoperability to enable business process/government service development and the integration of systems and business processes. In the business case, it expands to include the ability of two or more business processes, or services, to easily or automatically work together [19]. In order to reduce costs of industrial integration and inefficiencies, increase business agility, and allow the adoption of new and emerging technologies, the ability to interoperate between systems is key issue. For two systems to be interoperable, they must be able to exchange data and subsequently present that data such that a user can understand it. Interoperability describes the extent to which systems and devices can exchange data and interpret that shared data. Connectivity and interoperation among computers and entities and among software components can increase the flexibility and agility of industrial systems, thus reducing administrative and software costs for industry.

In June 2002, European heads of state adopted the Europe Action Plan 2005 at the Seville summit. It calls on the European Commission to issue an agreed interoperability framework to support the delivery of European digital services to enterprises [20].

This document recommends policies and technical specifications for linking public administration information systems across the EU. This research is based on open standards and the use of open source software. These are the pillars to support the European provision of digital services in the recently adopted European Framework of Interoperability (EIF) [21] and its Spanish equivalent [22]. This document is the reference for the interoperability

of the new Interoperable Program of Digital PanEuropean Services Provision to Public Administrations, Companies, and Citizens (IDAbc). European institutions and bodies should use the EIF for their operations with each other and with the citizens, businesses, and administrations of the respective EU Member States (IEF, 2014). Member States' administrations should use the guidance provided by the EIF to complement its national interoperability frameworks with a pan-European dimension and thus enable pan-European interoperability.

In this context, interoperability is the ability of information and communication technology systems and the business processes they support to exchange data and enable the exchange of information and knowledge. The ISO/IEC 2382 Information Technology Vocabulary defines interoperability as the ability to communicate, execute programs, or transfer data between several functional units in a way that requires the user to have little or no knowledge of the unique characteristics of those units. Interoperability can be considered on very different abstraction levels, and the distinctions to be made in this respect cut across all the other matrix dimensions. An interoperability framework can be described as a set of standards and guidelines, which describe the way in which organizations have agreed, or should agree, to interact with each other.

At the level of technical infrastructure, the industry is approaching interoperability through standards and in many cases conceptualizes those standards through stacks of technology. Technology stacks are conceptual layers of software and software functionality that interoperate between layers within stacks and between stacks in the same conceptual layer [23]. Within a continuum rank from a very concrete to a very abstract perspective, it is possible to distinguish three layers as shown in next **Figure 1**.

The main semantic interoperability objective is to improve communication on industrial knowledge related both between machines and between humans. To achieve this, a two-fold approach is necessary to achieve a unified ontology and tackle specific and clearly delineated issues. Inside semantic interoperability, various dimensions, such as medial/administrative or human/machines levels, can be distinguished. Organizational interoperability is defined as the state in which the organizational components of the industrial system are able to function perfectly together. The goal is an integrated industrial system, which provides efficient, effective, and holistic. The functional objective is to allow data to be exchanged between different platform in various corporations using different software, hardware, equipment, etc. from multiple manufacturers. Technical interoperability allows communication and iteration between systems from different manufacturers. Technical dimension of interoperability include uniform movement of industrial data, uniform presentation of data, uniform user controls, uniform safeguarding data security and integrity, uniform protection of industrial confidentiality, and uniform assurance of a common degree of service quality.

Numerous efforts are being leveraged by many standard efforts to address semantic and organizational interoperability and are proving to be a model for addressing semantic and organizational interoperability such as ebXML, RosettaNet, the new CEFACT/UN work to align their global work process standards with Web services, etc.
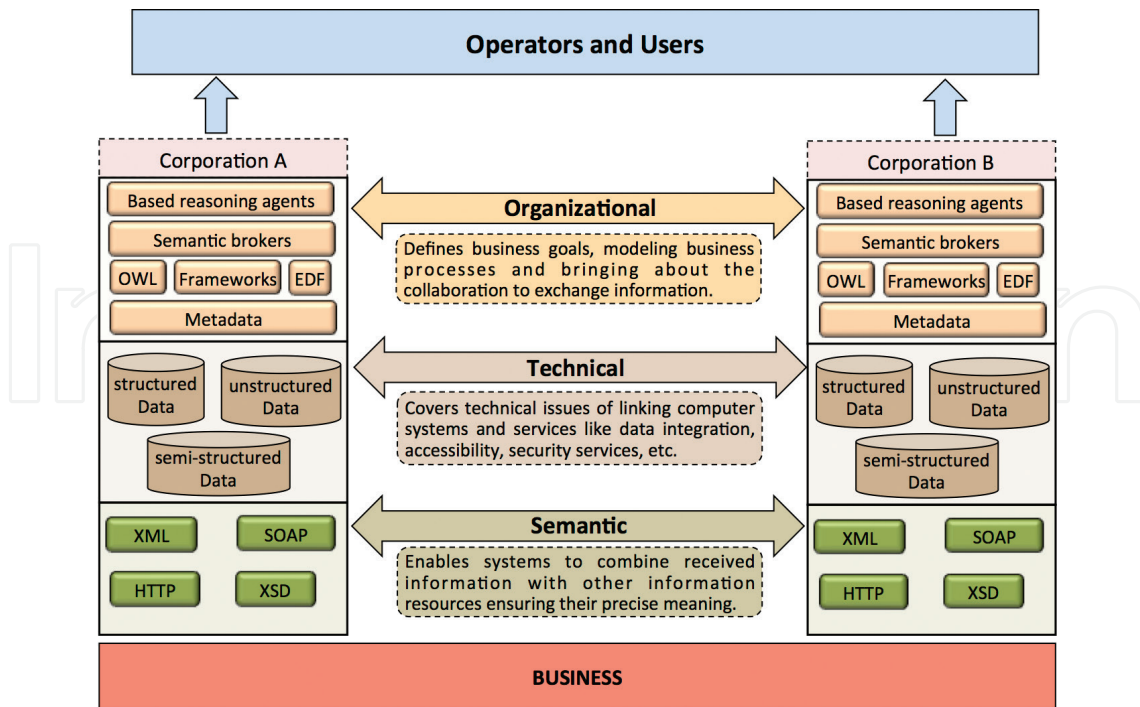
Figure 1. Conceptual interoperability layers.

The achievement of semantic and organizational interoperability requires strictly agreeing on the meaning of information and aligning business processes between companies/governments. At one level, general interprofessional frameworks and software infrastructure approaches can and are being developed for semantics and business processes. For example, the general semantics of major business transactions, such as purchase orders and invoices, is described through standards such as Universal Business Language (UBL), CEFACT Core Components, and Open Applications Group Integration Standard.

4. Intelligent and semantic architecture

The goal is achieved from a search perspective, with possible intelligent infrastructures to construct decentralized industrial repositories, where no global schema exists. This goal implies the application of CBR technique [24]. The prototype is the main tool to verify that the proposed architecture is an applicable solution, and this work attempts to achieve such verification by documenting in the proposal solution. In order to support the semantic of retrieval knowledge in industrial repositories, we develop a prototype named OntoEnter based on ontologies and expert system technologies. Obviously, our system is a prototype; nevertheless, it gives a good picture of the on-going activities in this new and important field. The architecture of our system is shown in **Figure 2**, which mainly includes three parts: the search engine, ontology knowledge base, and intelligent user interface.

The proposed architecture relies on the approach to efficiently retrieve information through metadata characterization and the inclusion of domain ontology. It involves using ontology

as a vocabulary to define complex and multi-relational case structures to support CBR processes. Our system works by comparing objects that can be retrieved through heterogeneous repositories and capturing a semantic view of the independent world of data representation.

4.1. The case-based reasoning engine

Keeping in mind that our final goal is to reformulate queries in the ontology to queries in another with least loss of semantics, we come to a process for addressing complex relations between ontologies. CBR is widely discussed in the literature as a technology for building information systems to support knowledge management, where metadata descriptions are used to characterize knowledge elements. CBR is a paradigm of problem solving that solves a new problem, in our case a new search, remembering a similar previous situation and reusing information and knowledge of that situation. Recovering one solves a new problem or more previously experienced cases, reusing the case, reviewing, and retaining. This approach when a description of the current problem is input to the system, the reasoning cycle may be described by the following processes **Figure 3**.

The system retrieves the closest matching cases stored in a case base. Reuse a complete design where case-based and slot-based adaptation can be hooked, provided. If appropriate, the validated solution is added to the case for use in solving future problems. Review the proposed solution, if necessary. Given that the proposed result may be inadequate, this process can correct the first proposed solution. Keep the new solution as part of a new case. This process allows CBR to learn and create a new solution. The solution is validated by comments from the user or the environment.

In our CBR application, problems are described by metadata concerning desired characteristics of an industry resource, and the solution to the problem is a pointer to a resource described
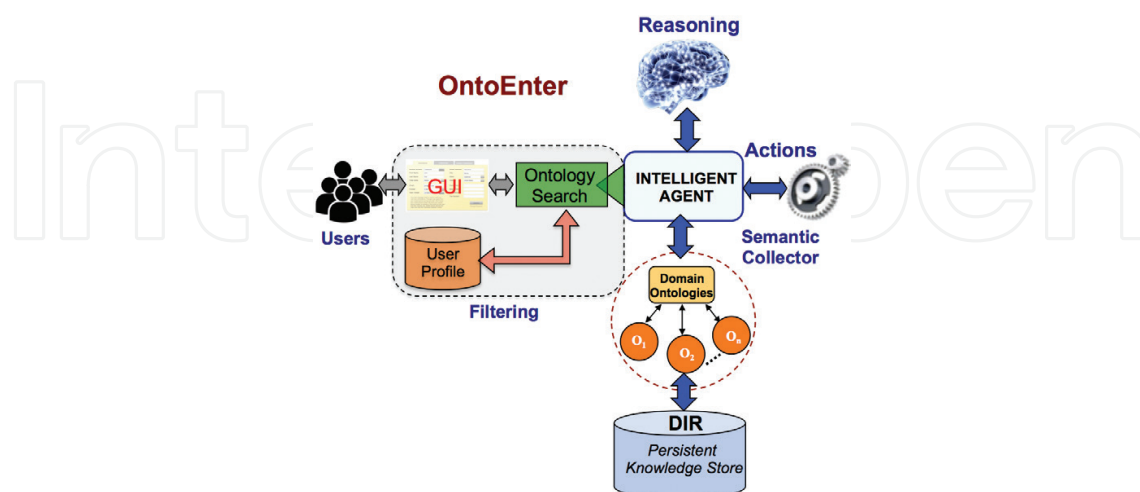

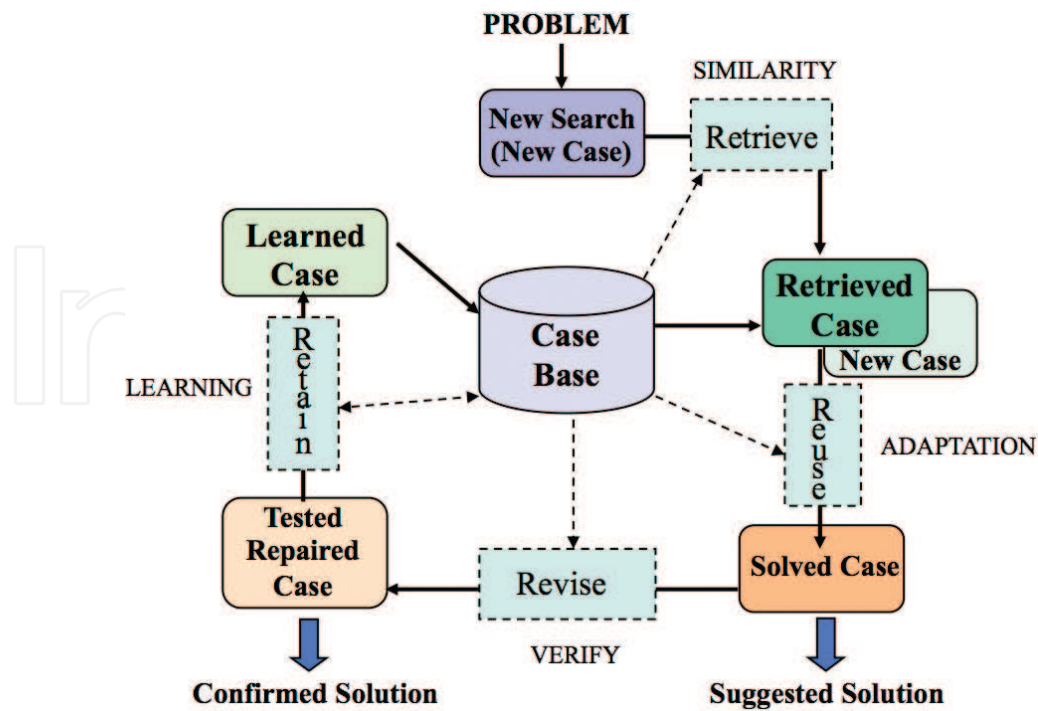**Figure 2.** System architecture of OntoEnter.

Figure 3. Case-based reasoning cycle in OntoEnter.

by metadata [25]. The development of a quite simple CBR application already involves a number of sttif, such as collecting case and background knowledge, modeling a suitable case representation, defining an accurate similarity measure, implementing retrieval functionality, and implementing user interfaces. CBR case data could be considered as a portion of the knowledge (metadata) about an OntoEnter object [26]. Every case contains both description problem and the associated solution.

Compared to other AI approaches, CBR reduces the effort required to acquire knowledge and representation significantly, which is undoubtedly one of the main reasons for the commercial success of CBR applications. However, implementing a CBR application from scratch remains a time-consuming software engineering process and requires a lot of specific experience beyond pure programming skills [27]. In this work, we have chosen framework jColibri to develop the intelligent search.

JColibri is a java-based configuration that supports the development of knowledge intensive CBR applications and helps in the integration of ontology in them [28]. The metadata descriptions of the resources and objects (cases) are abstracted from the details of their physical representation and are stored in the case base. This way the same methods can operate over different types of information repositories. The mapping between the two layers is done by connectors that read the values of the columns and the ontology of the database and return them to the application. That is to say, assign these values to different attributes of the case. Based on the same idea, the case base implements a common interface for similarity methods

to evaluate cases. This includes the generation of case representations, the definition of similarity measures, the testing of retrieval and use of explanation functionality, and finally, the implementation of stand-alone applications. The main focus of methods in this category is to find similarity between cases.

The use of structured representations of cases requires approaches for the evaluation of similarities that allow to compare two objects structured in different ways, in particular, objects belonging to different classes of objects. An important advantage of the similarity box recovery is that if there is no case that exactly matches the user's requirements, this may show cases that are more similar to your query. The use of structured representations of cases requires approaches for similarity assessment that allow to compares two differently structured objects, in particular, objects belonging to different object classes. The retrieval strategy used in our system is correlation technique. Correlation is a bivariate analysis that measures the strengths of association between two variables. A line that runs through all the data points and has a positive slope represents a perfect correlation between the two objects. In statistics, the value of the correlation coefficient varies between +1 and −1. When the value of the correlation coefficient is about ±1, then it is said to be a perfect degree of association between the two variables [29]. As the value of the correlation coefficient goes toward 0, the relationship between the two variables will be weaker, **Figure 4**.

We measure correlations with the Pearson correlation method. The Pearson coefficient is a more complex and sophisticated approach to finding similarity. This best fit line is generated by the Pearson coefficient, which is the similarity score. The following formula is used to calculate the Pearson r correlation:

$$\rho_{x,y} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{E\left[(X - \mu_x)(Y - \mu_y)\right]}{\sigma_x \sigma_y}$$

Where:

r = Pearson r correlation coefficient.

N = number of value in each data set.

$\sum xy$ = sum of products of paired scores.

$\sum x$ = sum of x scores.

$\sum y$ = sum of y scores.

$\sum x^2$ = sum of squared x scores.

$\sum y^2$ = sum of squared y scores.

Pearson r correlation is widely used in statistics to measure the degree of the relationship between linear related variables. For example, in the industrial stocks and storage, if we want to measure how two commodities are related to each other, Pearson r correlation is used to measure the degree of relationship between the two commodities. The coefficient is found by dividing the covariance by the product of the standard deviations of the
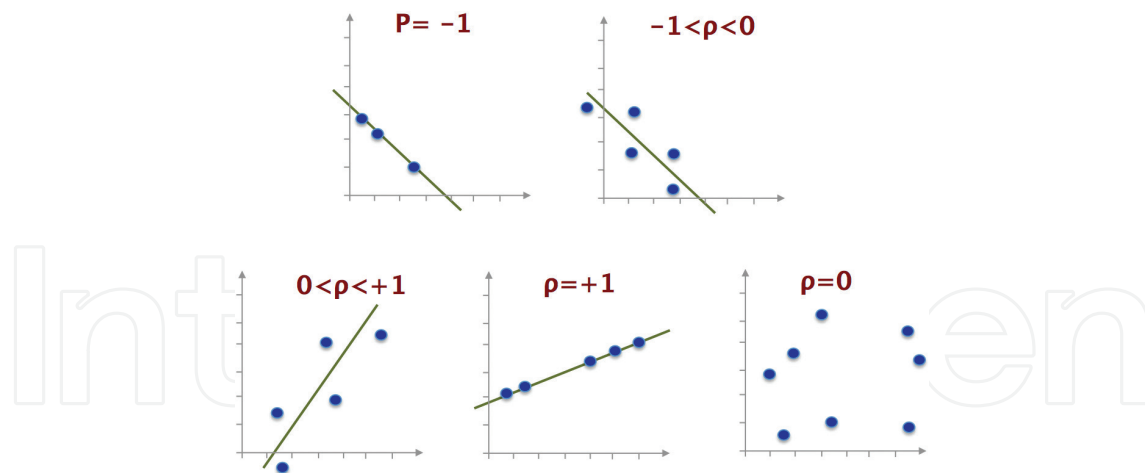
Figure 4. The method generates a best-fit line between attributes in two data objects.

attributes of two data objects. The advantage of the Pearson Coefficient over other techniques is that it is more robust against data that is not standardized. For example, if one person gives the rank for movies "a", "b", and "c" with scores of 1, 2, and 3, respectively, has a perfect correlation to someone who ranked the same movies with 4, 5, and 6. The following python code implements the Pearson coefficient for the same data described previously.

```python
#Return the Pearson Coefficient for person1 and person2
 def similarity_pearson(preferences, person1, person2):

    #Get the list of mutually rated items
    shared_items={}
    for item in preferences[person1]:
        if item in preferences[person2]: shared_items[item]=1

    #Find the number of elements
    num_shared = len(shared_items)

    #If there are no ratings in common, return 0
    if num_shared == 0: return 0

    #Add up all the preferences
    sum1 = sum([preferences[person1][item] for item in shared_items])
    sum2 = sum([preferences[person2][item] for item in shared_items])

    #Sum up the squares
    sum1_squares = sum([pow(preferences[person1][item],2) for item in shared_items])
    sum2_squares = sum([pow(preferences[person2][item],2) for item in shared_items])

    #Sum up the products
    prod_sum = sum([preferences[person1][item]*preferences[person2][item] for item in shared_items])

    #Calculate the Pearson score
    num = prod_sum - (sum1*sum2/num_shared)
    den = sqrt((sum1_squares-pow(sum1,2)/num_shared)*(sum2_squares - pow(sum2,2)/num_shared))
    if den == 0: return 0
    r = num/den

    return r
```

The open source jColibri system provides a framework for building CBR systems based on state-of-the-art software engineering techniques. The reason for choosing the jColibri framework is based on a comparative analysis between it and other frameworks, designed to facilitate the development of CBR applications. jColibri enriches the other shells CBR: CATCBR, CBR * Tools, IUCBRF, Orenge, in several aspects:

- Availability: open source framework.

- Implementation: Java implementation is one of our main requirements with respect to easy integration into the OntoEnter system, which is implemented in the J2EE environment.

Another decision criterion for our choice is related to the fact that jColibri offers the opportunity to incorporate the ontology in the CBR application to use it for the representation of cases and methods of reasoning based on content to evaluate the similarity between them. Providing easy-to-use model generation, data import, similarity modeling, explanation, and test functionality along with convenient graphical user interfaces, the tool allows even CBR beginners to quickly create their first CBR applications. However, at the same time, it ensures sufficient flexibility to enable expert users to implement advanced CBR applications.

4.2. Ontology knowledge base

Semantic modeling can help define the data and relationships between these entities. An information model provides the ability to abstract different types of data and provides an understanding of how data elements are related. A semantic model is a type of information model that supports the modeling of entities and their relationships. The total set of entities in our semantic model comprises the class taxonomy that we use in our model to represent the real world. Together, these ideas are represented by an ontology—the semantic model vocabulary that provides the basis upon which user-defined model queries are formed. The model supports the representation of entities and their relationships and can withstand the constraints on those relationships and entities. This provides the semantic composition of the information model [30].

Semantic models allow users to ask questions about what is happening in a modeled system in a more natural way. As an example, an oil production company could consist of five geographic regions, each region containing three to five drilling rigs and each drilling rig controlled by various control systems, each with a different purpose. One of these control systems could control the temperature of the extracted oil, while another could control the vibration in a pump. A semantic model will allow the user to ask a question like "What is the temperature of the oil extracted in Platform 3?" Without having to understand details such as, what specific control system supervises that information or what physical sensor is reporting temperature of oil on that platform.

The understanding provided through semantic models is fundamental to be able to correctly drive the correct ideas of supervised instrumentation that can ultimately lead to optimize business processes or, in this case, city services. As a result, semantic models can greatly improve the usefulness of the information obtained through operations integration solutions.

Ontology models can be used to relate the physical world, to the real world, in the line-of-business and decision makers. In the physical world, a control point such a valve or temperature sensor is known by its identifier in a particular control system, possibly through a tag name like 14-WW13. This could be one of several thousand identifiers within any given control system, and there could be many similar control systems across an enterprise. To further complicate the problem of information referencing and aggregation, other data points of interest could be managed through databases, files, applications, or component services with each having its own interface method and naming conventions for data accessing. A key value of the semantic model then is to provide access of information in the context of the real world in a consistent way. Within a semantic model implementation, this information is identified using "triples" of the "subject-predicate-object" form; for example:

Tank1 <has temperature> Sensor 7
Tank 1 <is part of> Platform 4
Platform 4 <is part of> Plant1

These triples, taken together, constitute the Plant1 ontology and can be stored on a model server, as is described in more detail later in this article. This information, then, can be easily traversed using the model query language to answer questions such as "What is the temperature of tank 1 on Platform 4," more easily than the case without a semantic model.

We concentrate on the critical issue of metadata/ontology-based search and expert system technology. The main objective of the system is to improve the modeling of a semantic coherence to allow the interoperability of different modules of environments dedicated to the industrial area. We have proposed to use ontology together with CBR in acquiring expert knowledge in the specific domain. The primary information managed in the OntoEnter domain is metadata about industrial resources, such as guides, digital services, alarms, information, etc. We need a vocabulary of concepts, resources, and services for our information system described in the scenario that requires definitions about the relationships between objects of discourse and their attributes [31]. OntoEnter project contains a collection of codes, visualization tools, computing resources, and data sets distributed across the grids, for which we have developed a well-defined ontology using RDF language. Our ontology can be regarded as quaternion OntoEnter:={caller, resources, properties, relation), where profiles represent the user kinds, collection contains all the services and resources of the institutional repository, the matter cover the different information sources: electronic services, web pages, BB.DD., guides, etc., and a set of relationships intended primarily for standardization across ontologies.

We integrate three essential sources for the system: electronic resources, the catalog of documents, and the personal database. The W3C defines standards that can be used to design an ontology [32]. We write the description of these classes and properties in RDF semantic markup language. We chose Protégé as our ontology editor, which supports the acquisition of knowledge and the development of knowledge bases. Protégé provides an environment for the creation and development of semantic knowledge structures-ontologies and semantically annotated Web services. Protégé organizes these elements as a dynamic workflow [33]. For the construction of the ontology of our system, we follow the sttif detailed below.

- Determine the domain and scope of the ontology. This should provide the location of different online resources. These are included from different sources: Catalog of Publications, Websites, Electronic Resources, etc.

- Enumerate important terms in the ontology. It is useful to write down a list of all terms we would like either to make statements about or to explain to a user.

- Define the classes and the class hierarchy. When designing the ontology, we need first group together related resources of the institutional repositories. There are three major groups of resources: users, services, and resources. IA detailed picture of our effort in designing this ontology is available in **Figure 5**. This shows the high level classification of classes to group together OntoEnter resources as well as things that are related with these resources.

- Generating the ontology instances with SW languages. To provide a conversational CBR system to retrieve the requested metadata satisfying a user query, we need to add enough initial instances and item instances to the knowledge base.

After designing the ontology, we write the description of these classes and properties in RDF semantic markup language. Then the domain expert, in this case, the administrative staff fills the blank units of the instance according to the knowledge of the domain. A total
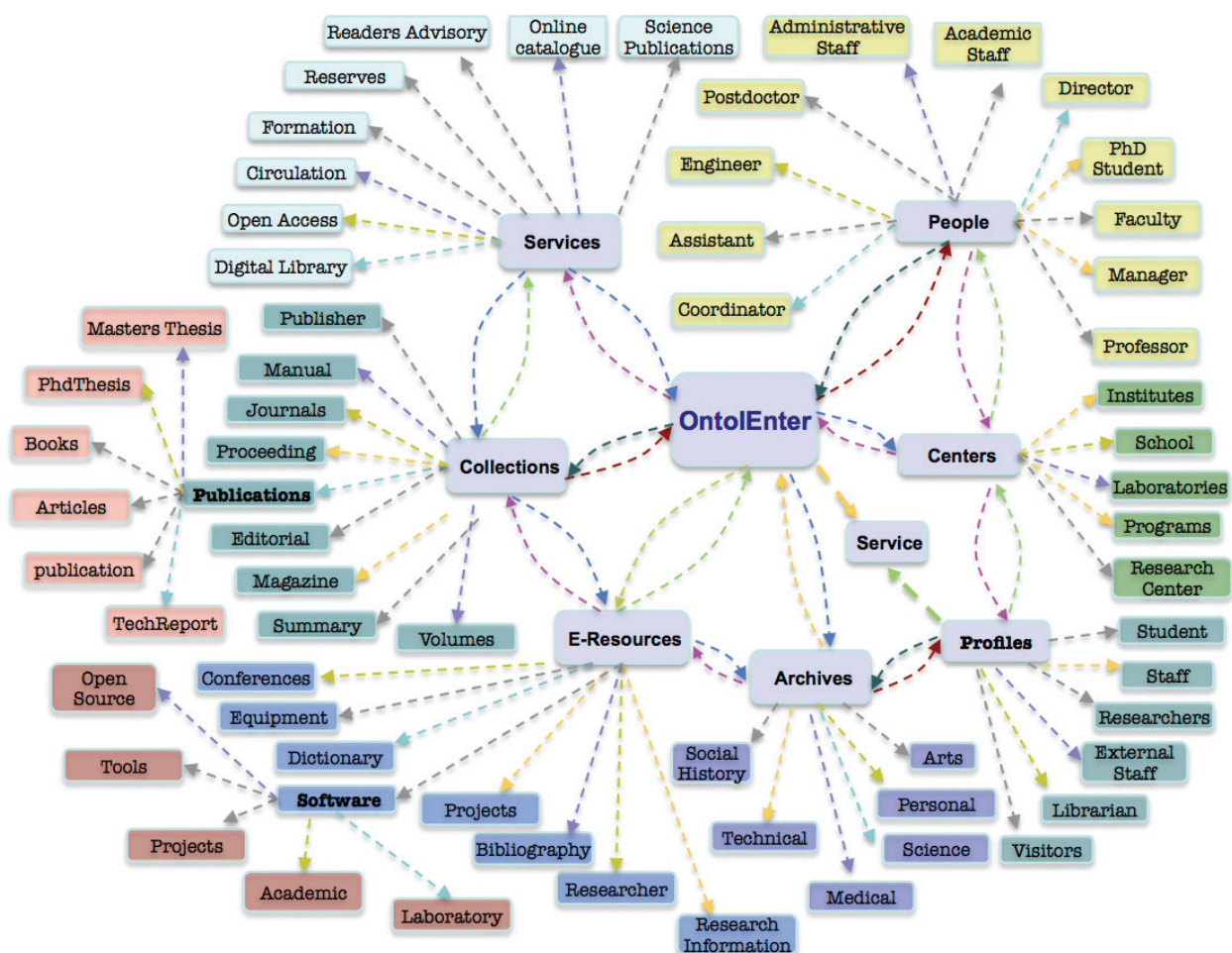

Figure 5. Class hierarchy for the OntoEnter ontology.

of 13,000 cases were collected for user profiles and their different resources and services. Each case contains a set of attributes related to both metadata and knowledge. As a plus, domain specific rules defined by domain experts can infer more complex high-level semantic descriptions, for example, by combining low-level features in local repositories. Considering that our final objective is to reformulate queries in the ontology to queries in another with less loss of semantics, we arrive at a process to approach complex relations between two ontologies. As mentioned in previous sections, the relationships between ontologies can be compounded as a form of declarative rules, which can be handled in inference engines.

In addition, our system combines ontologies with rules into the ontology-based search due to the description limitations in current ontology languages. As mentioned in previous sections, relations among ontologies can be composed as a form of declarative rules, which can be further handled by inference engines.

4.3. Intelligent user Interface

OntoEnter is software, which is an intermediate link between users and search engine. By using OntoEnter user can tune the query in accordance with his needs. Advanced conversational user interface interacts with users to solve a query, defined as the set of questions selected and answered by the user during a conversation. The real way to get an individualized interaction between a user and a website is to present the user with a variety of options and allow the user to choose what is of interest at that specific time. In our system, the user interacts with the system to fill the gaps to retrieve the correct cases.

The algorithms used allow to give intelligent advice on improving the search query to obtain more relevant results to a narrow number of documents obtained or, conversely, to extend it. But the main task of OntoEnter is to specify in which exact word, the word is used and to formulate the "question" to the search engine, excluding answers from an inappropriate domain and adding semantically similar results [34].

For the distributed retrieval of learning resources, we use profile users, which are used for personalized searches according to user specifications. The methodology was based on incremental user profiling, which assumes mapping of a user's keywords to the concepts of the domain ontology according to the presented transformation rules. Transformation algorithm was implemented in the research prototype as the combined capability of the query transformation agent and the ontology agent of the intelligent multi-agent information retrieval mediator. The user interface helps user to build a particular profile that contains his interest search areas in the industry repositories domain. In an intelligence profile setting, people are surrounded by intelligent interfaces merged, thus creating a computing-capable environment with intelligent communication and processing available to the user by means of a simple, natural, and effortless human-system interaction. The objective of profile intelligence has focused on creating of user profiles: Plan Managers, Assistants, Operators, and Engineers. If the information space is well designed, then this option is easy, and the user achieves optimal information through the use of natural intelligence, that is, the options are easy to understand so users know what they will see if they click in a link, and what they negate by not following other links, **Figure 6**.

Figure 6. User profiles, graphical user interface.

The profile agent is an environment, in which software agents can be executed to retrieve E-learning resources and which is wrapped by a Web service. This configuration contains the user requirements that typically described the relative needs, tasks, and goals of the user for an individual search. Profile agents help students with the search, according to the specifications they made. Search parameters in a profile, the initiation of a search, or access to the list of retrieved learning objects can be controlled by invoking appropriate search operations that extract learning resource metadata. Ideally, profile agents learn from their experiments, communicate and cooperate with other agents, around in DL. A profile agent uses a registry to locate learning searches. The agent compares the metadata and the search keywords for possible matches and presents the search results to the user. For this, a statistical analysis has been done to determine the importance values and establishing specified user requirements.

5. Experimental evaluation

In order to validate the approach, we have developed intelligent control platform in an electrical power system. This system integrates management knowledge into network resource specifications. We study an example of alarm detection and intelligent resolution of incidents related to a private network. We have used a telecommunications network belonging to a company in the electricity sector Sevillana-Endesa (SE), a Spanish electricity company. OntoEnter is used to optimize the operation of hundreds of connected sensors currently installed. Many

of these sensors are wireless because they can be installed more quickly and at less cost than their wired equivalents, often with no required downtime. These low-cost wireless sensors and accompanying analytics can dramatically improve plant performance, increase safety, and pay for themselves within months. The Spanish electricity grid has a wireless network in the regional high voltage grid. Part of the long distance traffic in this network is controlled by a wireless intelligent system distributed through this private network. The use of knowledge integration in agents can help the system administrator to use the maximum capabilities of the intelligent network management platform without having to use another specification language to customize the application [35].

The intelligent development of the system must meet the following requirements: it must be robust, and the management activity should not interfere with the normal operations of the network and should only intervene when necessary. We will use the SCADA system due to management limitations of the network communication equipment. SCADA consists of the following subsystems **Figure 7**.

- Remote terminal units (RTUs) are connected to sensors in the process, the conversion of sensor signals to digital data and the sending of digital data to the monitoring system.

- Communication infrastructure that connects the monitoring system to RTUs.

- A monitoring system (computer), collecting (acquiring) data about the process and sending commands (control) to the process, which is our IA.

SCADA systems are configured around standard basic functions such as data acquisition, monitoring and event processing, archiving and data storage analysis, etc. The RTU encodes sensor inputs in protocol format and sends them to the SCADA master. The fundamental role of an RTU is the acquisition of various types of power process data, accumulation, packaging,
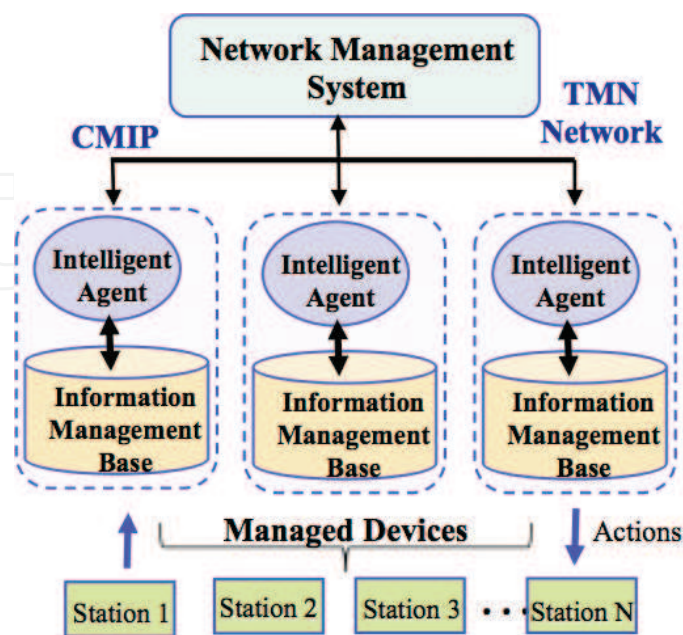

Figure 7. Elements of the prototype.

and data conversion in a form that can be communicated back to the master, interpretation and output of the commands received from the master, local filtering performance, calculation, and processes to allow specific functions to be performed locally. The supervision below and RTU includes all network devices and substation and feeder levels like circuit breakers, reclosers, autosectionalizers, the local automation distributed at these devices, and the communications infrastructure [36].

OntoEnter can monitor, in real time, the network's main parameters, making use of the information supplied by the SCADA, placed on the main company building, and the RTUs are installed at different stations. From the information provided, the operator can take action to solve any errors that may arise or send a technician to repair the station equipment. OntoEnter allows the operator to search for information, alarms, or digital and analog parameters of measurement, registered in each IA or RTU. The system has the ability to select the IA that is best suited to satisfy the client's requirements, without the client being aware of the details about the agent. In addition, the AI is able to communicate and negotiate with the other IAs. Collaborative IAs are useful, especially when a task involves several systems in the network.

6. Evaluation and proofs

When we perform a search on a search engine, we are looking to find the most relevant material, while minimizing the junk that is retrieved. This is the basic objective of any search engine. Get important information while avoiding junk is difficult, if not impossible to accomplish. The experiments carried out, in order to evaluate the effectiveness of the assignment of runtime ontology. The main objective has been to verify if the agent-assisted query formulation mechanism provides a suitable tool to increase the number of significant documents extracted from the DIRs to be stored in the CBR. For our experiments, we included about 50 users with different profiles. It set a context for users, they were asked to at least start their essay before issuing any query to the system. They were also asked to look through all results returned by OntoEnter before clicking on any result [37].

We compared the top 10 search results for each keyword phrase per search engine. Our application recorded the results they clicked on, which we used as a form of implicit user relevance in our analysis. We must consider that the relevance of recovered documents is subjective. That's different people can assign different relevance values to the same document. In our study, we have agreed different values to measure the quality of recovered documents, excellent, good, acceptable and poor, as can be seen in **Table 1**. After the data were collected, we had a record of queries with an average of 5 queries per user. From these queries, some of

	Excellent	Good	Acceptable	Poor
OntoEnter	7.5%	42.3%	35.1%	14.4%
Traditional SE	1.4%	25.7%	31.5%	21.3%

Table 1. Analysis of retrieved document relevance for select queries.

them had to be deleted, either because multiple results were clicked, no results were clicked or no information was available for that particular query.

In each experiment, we report the average rank of the user-clicked result for our baseline system, another search engine, and for our search engine OntoEnter. Thus basically, we can define two set-based measures: precision and recall.

$$precision = \frac{\left| \left\{ relevan\, t_{document} \right\} \cap \left\{ retrieve\, d_{documents} \right\} \right|}{\left| \left\{ retrieve\, d_{documents} \right\} \right|}$$

$$recall = \frac{\left| \left\{ relevan\, t_{documents} \right\} \cap \left\{ retrieve\, d_{documents} \right\} \right|}{\left| \left\{ relevan\, t_{documents} \right\} \right|}$$

It is possible to measure how well a search performed with respect to these two parameters. For each such set, precision and recall values can be plotted to give a precision-recall curve. We need these measures, if we are to evaluate the ranked retrieval results for search engines. These measures are computed using unordered sets of documents. The remaining queries were analyzed and evaluated **Table 2**.

It is easy to compare several classifiers in the precision graph. Curves near the perfect precision-recall curve have a better performance level than those closest to the baseline. In other words, a curve above the other curve has a better performance level (**Figure 8**).

Precision and retrieve are inversely related, i.e., as precision increases recall falls and vice-versa. When a relevant document is not retrieved at all, the precision value in the above equation is taken to be 0. A balance between these two needs to be achieved by the search engine that to achieve this and to compare performance, the precision-recall curves come in the practice.

Recall	OntoEnter Precision	Traditional SE Precision
0.10	0.90	0.60
0.20	0.85	0.56
0.30	0.79	0.43
0.40	0.68	0.35
0.50	0.61	0.25
0.60	0.52	0.05
0.70	0.39	—
0.80	0.20	—
0.90	0.05	—
1.00	—	—

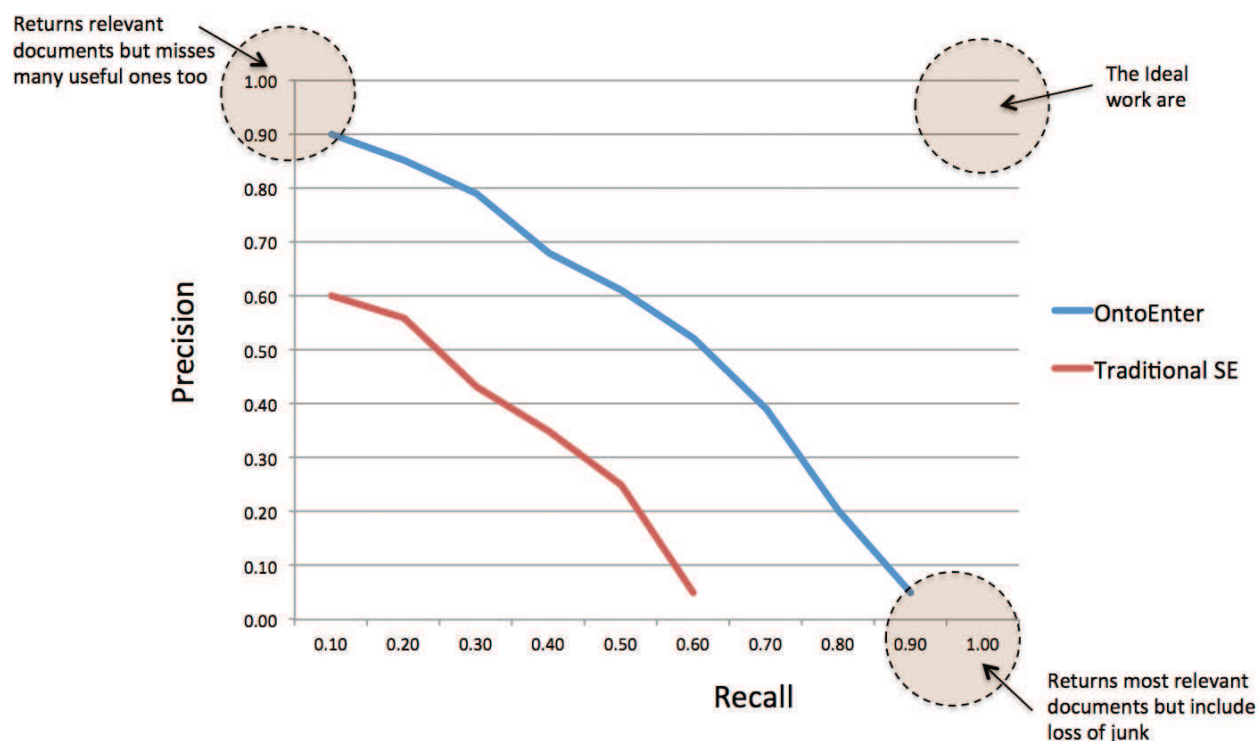**Table 2.** Precision and recall values.

Figure 8. Performance OntoEnter & Traditional Search Engine (TSE).

This trade-off between precision and recall can be observed using the precision-recall curve and an appropriate balance between the two obtained. The precision-recall curves for two algorithms are shown. Depending on the requirement of high precision at the cost of the recall, or high recall with lower precision, an appropriate algorithm can be chosen. In our case, we choose the appropriate system depending on the high precision and data with false positives allowed. Two precision-recall curves represent the performance levels of the two search engines. The search engine OntoEnter clearly outperforms TSE in this domain example. Our system performs satisfactorily with about a 95.2% rate of success in real cases.

Another important aspect of the design and implementation of OntoEnter is the determination of the degree of speed in the answer that the system provides. During experimentation, heuristics and measures that are commonly adopted in information retrieval have been used. A statistical analysis was performed to determine the importance values in the results. While users were performing these searches, an application continued to run in the background on the server and captured the content of the written queries and search results. We can establish that OntoEnter speed in our domain improves the answer time and the average of the traditional search engine. **Figure 9** shows a graphic of these parameters that was collected as a part of the experiment.

We can establish that the speed in the system improves the procedure time and the average of the traditional search engine. Results for OntoEnter are 15.1% better than procedure time and 19.5% better than running time/sec searches on traditional search engines.
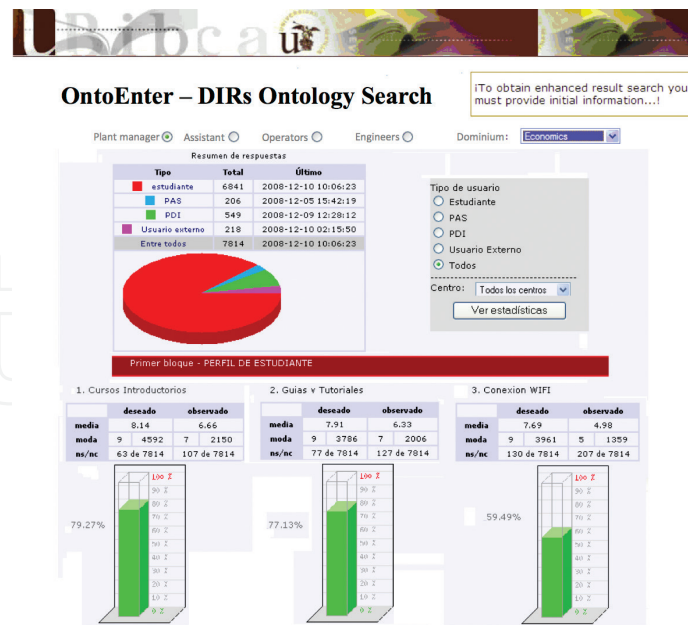
Figure 9. OntoEnter search analysis report.

7. Conclusion and future works

In this chapter, we investigated how the semantic technologies can be used to provide additional semantics from existing resources in industrial repositories. For this purpose, we presented a system based on ontology and AI architecture for knowledge management in industrial repositories. We describe an effort to design and develop a prototype to manage resources in a repository such as the OntoEnter project and exploit them to help users as they select resources. Our study addresses the main aspects of a Semantic Web Information Retrieval System architecture attempting to respond to the requirements of the next generation of Semantic Web users. This scheme is based on the following principle: knowledge elements are abstracted from a characterization by a metadata description that is used for further processing.

In this chapter, we offer different possibilities, which the semantic Web opens for the industry. An important goal is to study appropriate industrial cases, compile arguments, launch industrial projects, and develop prototypes for industrial companies that not only create with us but also benefit from the semantic Web.

As described here, semantic models play a key role in the evolving solution architectures that support the business goal of obtaining the complete view of "what is happening" within operations and then deriving business insights from that view. Semantic models based on industry standards take that one step further, especially as application vendors adopt those standards (which, as always, will happen more rapidly through pressure from the user community). This study addresses the main aspects of a semantic and intelligent information retrieval system architecture trying to answer the requirements of the next-generation semantic search engine. We have investigated how the semantic technologies can be used to provide additional semantics from existing resources in institutional repositories.

This scheme is based on the principle of knowledge elements that are abstracted from a metadata description characterization that is used for further processing. We have proposed to use ontology together with CBR in acquiring expert knowledge in the industry specific domain. We have developed the domain ontology, and we have studied how the content-based similarity between the concepts typed attributes could be assessed in CBR system. The study analyses the implementation results and evaluates the viability of our approaches in enabling search in intelligent-based digital repositories. It introduced a prototype Web-based CBR retrieval system, which operates on an RDF file store. With this, characteristic of the model ability of an individual will be increased to learn through collective searches experience. Furthermore, an IA was illustrated for assisting the user by suggesting improved ways to query the system on the ground of the resources in industry repositories according to his own preferences, which come to represent his interests. We have used all the profile agents effectively to generate relevant and recommended personalized profile for the different users.

OntoEnter can be part of a bigger framework of interacting global information networks including other DIRs, scientific repositories, commercial providers, and relies as much as possible on standards and existing building blocks as well as be based on Web standards. The combination of effective information retrieval techniques and IAs continues to show promising results in improving the performance of the information that is being extracted from the online repositories for users. Our findings suggest that IA is the central manager in the knowledge transfer process. Their mediation is essential to help adapt the knowledge produced by academics and makes it easier to adopt and use by the educational community. We conclude pointing out an important aspect of the obtained integration: improving representation by incorporating more metadata from within the information and intelligent techniques into the retrieval process, the effectiveness of the knowledge retrieval is enhanced. The model has good characteristics in providing preference to the users with a novel approach of finding nearby meaning of query and user can also recommend result pages by their opinion.

Future work will address the exploitation of information from other institutional repositories and digital services and refine the suggested queries, expand the system to provide other support, and refine and evaluate the system through user testing.

Future work will focus on the design of distributed and self-managed services based on the Web and services, which are:

- Able to examine and filter information based on semantic similarity and closeness

- Able to handle heterogeneous data/knowledge /intelligence sources.

- Able to discover, compose, and integrate heterogeneous components automatically.

- Able to create, deploy, and exploit linked data.

- Able to perform automated and user-driven application/service orchestration and choreography, etc.

Author details

Antonio Martin, Mauricio Burbano and Carlos León*

*Address all correspondence to: cleon@us.es

Technology Electronic Department, Seville University, Spain

References

[1] Jimeno-Yepes A, Berlanga-Llavori R, Rebholz-Schuhmann D. Ontology refinement for improved information retrieval. Information Processing & Management. 2010;**46**(4): Semantic Annotations in Information Retrieval

[2] Yang Z, An Y, Sun Y, Zhang J. Research on intelligent glue-coating robot based on visual servo. Physics Procedia. 2012;**24**(Part C):2165-2171

[3] Song M, Song I, Hu X, Allen R. Integration of association rules and ontologies for semantic query expansion. Data & Knowledge Engineering. October 2007;**63**(1):63-75

[4] Bañares-Alcántara R, Jiménez L, Aldea A. Multi-agent systems for ontology-based information retrieval. In: Puigjaner L, Espuña A, editors. Computer Aided Chemical Engineering. Vol. 20. Elsevier; 2005. p. 1549-1554

[5] Zhang J, Zhao W, Xie G, Chen H. Ontology- based knowledge management system and application. Procedia Engineering. 2011;**15**:1021-1029

[6] Jiang X, Tan A. Learning and inferencing in user ontology for personalized semantic web search. Information Sciences. 2009;**179**(16):2794-2808 ISSN

[7] García-Crespo A et al. Digital libraries and web 3.0. The CallimachusDL approach. Computers in Human Behavior. 2011;**27**(4):1424-1430

[8] Badii a et al. Semi-automatic knowledge extraction, representation and context-sensitive intelligent retrieval of video content using collateral context modelling with scalable ontological networks. Signal Processing: Image Communication. 2009;**24**(9):759-773

[9] Lim SCJ, Liu Y, Lee WB. Multi-facet product information search and retrieval using semantically annotated product family ontology. Information Processing & Management. 2010;**46**(4, Semantic Annotations in Information Retrieval):479-493

[10] Fernandez M et al. OnlineSemantically Enhanced Information Retrieval: An Ontology-Based Approach. Web Semantics: Science, Services and Agents on the World Wide Web; 2016 In Press, Corrected Proof

[11] Govedarova, D., Stoyanov, S., and Popchev, I. (2008) "An ontology based CBR architecture for knowledge management in BULCHINO catalogue," in Proc. International Conference on Computer Systems and Technologies (CompSysTech).

[12] Stuckenschmidt, H., and Harmelen, F.V. (2011). "Ontology-based metadata generation from semi-structured information," K-CAP, pp. 163-170, ACM.

[13] Ding H. Towards the metadata integration issues in peer-to-peer based digital libraries. In: Jin GCCH, Pan Y, Xiao N, Sun J, editors. LNCS. Vol. 3251. Berlin, Germany: Springer; 2004

[14] Guha R, McCool R, Miller E. Semantic search. In: Proceedings of WWW2003; 2003

[15] Martín Antonio, Mauricio Burbano, Iñigo Monedero, Joaquín Luque. (2016). "Semantic reasoning method to troubleshoot in the industrial domain", the fifth international conference on Intelligent systems and applications, pp.89.

[16] OntoWeb project. OntoWeb—Technical Roadmap v2.0. Deliv- erable 1.1.2 of IST Project IST-2000-29243 OntoWeb, November 2002

[17] Ceccaroni L, Cortés U. OntoWEDSS: Augmenting environmental decision-support systems with ontologies. Environmental Modelling and Software, pp. 2003;**3**:

[18] Martín A. Intelligent search engine to a semantic knowledge retrieval in the digital repositories. International Journal on Advances in Intelligent Systems. 2015;**8**:67

[19] Segaran T. Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media; 2007a

[20] SEC. Commission Staff Working Paper: linking up Europe, the importance of interoperability for egovernment services [Online 2016]. Available from: http://europa.eu.int/ISPO/ida/export/files/en/1523.pdf, 2016.05.3

[21] EIF. European Interoperability Framework Version 2. [Online 2016]. Available from: http://ec.europa.eu/isa/strategy/doc/annex_ii_eif_en.pdf, 19 April 2015.

[22] MAP. Aplicaciones utilizadas para el ejercicio de potestades. Criterios de Seguridad, Normalización y Conservación. Ministerio de Administraciones Públicas. [Online 2016]. Available from: http://www.csi.map.es/csi/criterios/index.html. 5 March 2016;

[23] CompTIA. European Interoperability – ICT Industry Recommendations; 2004. p. 13

[24] Toussaint J, Cheng K. Web-based CBR (case-based reasoning) as a tool with the application to tooling selection. International Journal of Advanced Manufacturing Technology. 2006;**29**(1):24-34

[25] Sure Y, Studer R. Semantic web technologies for digital libraries. Library Management Journal, Emerald. 2005;**26**:190-195

[26] Bechhofer S, Harmelen FV, Hendler J, Horrocks I. OWL web ontology language reference. W3C recommendation. 2004;**10**(February): Publisher W3C

[27] Martin A, León C, Expert Knowledge Management Based on Ontology in a Digital Library. Departamento de Tecnología Electrónica: Universidad de Sevilla; 2010. p. 4

[28] GAIA—Group for Artificial Intelligence Applications. jCOLIBRI project—Distribution of the development environment [Online 2016]. Available from: http://gaia.fdi.ucm.es/research/colibri/jcolibri/ 25 April 2016

[29] Segaran T, "Programming Collective Intelligence: Building Smart Web 2.0 Applications, Published by O'Reilly Media, August 23rd 2007.

[30] Hanis T, Noller D. The role of semantic models is smart industrial operations", developer works. IBM. 2011;2

[31] Taniar D, Rahayu JW. Web Semantics and Ontology. Hershey, PA: Idea Group; 2006

[32] W3C. RDF Vocabulary Description Language 1.0: RDF Schema. [Online 2016]. Available from: http://www.w3.org/TR/rdf-schema/, 10 August 2016.

[33] PROTÉGÉ. The Protégé Ontology Editor and Knowledge Acquisition System. [Online 2016]. Available from: http://protege.stanford.edu/, 5 July 2016.

[34] Terziyan V. Semantic Search Facilitator: Concept and Current State of Development. Industrial ontologies group; 2004

[35] Martín A, León C. Intelligent management experience on efficient electric power system. ICREPQ. 2014;**12**:842

[36] Warren P. Applying Semantic Technologies to a Digital Library: A Case Study. Library Management Journal: Emerald; 2005

[37] Amerland D. Google Semantic Search: Search Engine Optimization (SEO) Techniques that Get your Company More Traffic, Increase Brand Impact and Amplify your Online Presence. Que Publishing Kindle Edition; 2013