

# Automated Knowledge Base Management: A Survey

Jorge Martinez-Gil

Software Competence Center Hagenberg (Austria)

email: jorge.martinez-gil@scch.at, phone number: +43 7236 3343 838

**Keywords:** Information Systems, Knowledge Management, Knowledge-based Technology

## Abstract

A fundamental challenge in the intersection of Artificial Intelligence and Databases consists of developing methods to automatically manage Knowledge Bases which can serve as a knowledge source for computer systems trying to replicate the decision-making ability of human experts. Despite of most of tasks involved in the building, exploitation and maintenance of KBs are far from being trivial, significant progress has been made during the last years. However, there are still a number of challenges that remain open. In fact, there are some issues to be addressed in order to empirically prove the technology for systems of this kind to be mature and reliable.

## 1 Introduction

Knowledge may be a critical and strategic asset and the key to competitiveness and success in highly dynamic environments, as it facilitates capacities essential for solving problems. For instance, expert systems, i.e. systems exploiting knowledge for automation of complex or tedious tasks, have been proven to be very successful when analyzing a set of one or more complex and interacting goals in order to determine a set of actions to achieve those goals, and provide a detailed temporal ordering of those actions, taking into account personnel, materiel, and other constraints [9].

However, the ever increasing demand of more intelligent systems makes knowledge has to be captured, processed, reused, and communicated in order to complete even more difficult tasks. Nevertheless, achieving these new goals has proven to be a formidable challenge since knowledge itself is difficult to

explicate and capture. Moreover, these tasks become even more difficult in fields where data and models are found in a large variety of formats and scales or in systems in which adding new knowledge at a later point is not an easy task.

But maybe the major bottleneck that is making very difficult the proliferation of expert systems is that knowledge is currently often stored and managed using Knowledge Bases (KBs) that have been manually built [11]. In this context, KBs are the organized collections of structured and unstructured information used by expert systems. This means that developing a system of this kind is very expensive in terms of cost and time. Therefore, most current expert systems are small and have been designed for very specific environments. Within this overview, we aim to focus on the current state-of-the-art, problems that remain open and future research challenges for automatic building, exploiting and maintaining KBs so that more sophisticated expert systems can be automatically developed and practically used.

The rest of this work is structured as follows: Section 2 presents the state-of-the-art concerning automated knowledge-base management. Section 3 identifies the problems that remain open. Section 4 propose those challenges that should be addressed and explain how their solution can help in the advancement of this field. Finally, we remark the conclusions.

## **2 State-of-the-art**

Although the challenge for dealing with knowledge is an old problem, it is perhaps more relevant today than ever before. The reason is that the joint history of Artificial Intelligence and Databases shows that knowledge is critical for the good performance of intelligent systems. In many cases, better knowledge can be more important for solving a task than better algorithms [38].

It is widely accepted that the complete life cycle for building systems of this kind can be represented as a three-stage process: creation, exploitation and maintenance [14]. These stages in turn are divided into other disciplines. In Table 1 we can see a summary of the major disciplines in which the complete cycle of knowledge (a.k.a. Knowledge Management) is divided<sup>1</sup>.

---

<sup>1</sup>In general, there is no agreement about the nomenclature used in the literature, but we will try to explain these discrepancies. In general we will use the expression a.k.a. (also knows as) for the same discipline receiving different names

Knowledge Creation	Knowledge Exploitation	Knowledge Maintenance
Knowledge Acquisition	Knowledge Reasoning	Knowledge Meta-Modeling
Knowledge Representation	Knowledge Retrieval	Knowledge Integration
Knowledge Storage and Manipulation	Knowledge Sharing	Knowledge Validation

Table 1: Summary of concepts in the Knowledge Management field

Concerning the automatic creation of KBs (a.k.a. knowledge learning, knowledge extraction or knowledge generation), there are three major steps that should be fulfilled: automatic acquisition of the knowledge, appropriate representation of that knowledge, and storage and manipulation of the knowledge into the KB. These major steps are summarized below:

- The process of automatic **knowledge acquisition** starts by extracting concepts and relations among the concepts from texts or document libraries using some kind of methods for terminology extraction [7]. Then, concrete instances for these concepts should be also extracted [47]. This usually involves the use of natural language processing techniques [23]. Then statistical or symbolic techniques are applied to extract relations between the terms and concepts [17]. The intentional aspects of domain are formalized by means of a schema or ontology. Meanwhile, the extensional part is based on instances of concepts and relations on the basis of the given schema or ontology.
- **Knowledge representation** phase consists of providing a formal specification of a knowledge domain using some kind of logical notation to represent the concepts, properties for these concepts, relations among these concepts, and the underlying rules of that domain [5]. The conditions and constraints of knowledge formation and organization have to be formally specified [2]. A notation of this kind follows a logical specification using expressions and symbolical structures, such as taxonomies, classes, and axioms [43].
- Another important aspect consists of **storing and manipulating large KBs**. This means the design of a physical and logical support, on which applications and users can rely in order to store and share the knowledge [4]. This involves using standard ways to communicate knowledge units and retrieve them [15]. Metadata and annotations should be properly taken into account. Ignoring

the inherent inferential capability given by KBs each KB is also a database in the sense that there is a schema, i.e. the concepts and roles, and a set of instances. Therefore, adopting database technology as key method to address this issue is an idea adopted by most of the solutions.

Concerning the automatic exploitation of KBs (a.k.a. knowledge exploitation or knowledge application) can be divided in two subgroups: knowledge utilization and knowledge transfer. At the same time, the utilization of knowledge can be used for knowledge reasoning or for knowledge retrieval (in the way the Question and Answering (Q & A) systems work [44]). Meanwhile, the purpose of knowledge sharing (a.k.a. knowledge exchange) is the process through which explicit or tacit knowledge is communicated to others.

- **Knowledge reasoning** consists of inferring logical consequences from a set of asserted facts or axioms [27]. The notion of a reasoner generalizes that of an inference engine, by providing a richer set of mechanisms to work with [46]. Formal specification is required in order to be able to process ontologies and reasoning on ontologies automatically. By reasoning, it is possible to derive facts that are not expressed in the KB explicitly. Some of the facts that can be automatically derived could be:
  - Consistency of ABox with respect to TBox, determine whether individuals in ABox do not violate descriptions and axioms described by TBox
  - Satisfiability of a concept, determine if a description of the concept is not contradictory
  - Subsumption of concepts, determine whether concept A subsumes concept B
  - Retrieval of individuals, find all individuals that are instances of a concept
  - Realization of an individual, find all concepts which the individual belongs to, especially the most specific ones
- **Knowledge retrieval** aims to help users or software applications to find knowledge that they need from a KB through querying, browsing, navigating and/or exploring [36]. The goal is to return information in a structured form, consistent with human cognitive processes as opposed to plain

lists of items [41]. It is important to remark that traditional information retrieval organize information by indexing. However, knowledge retrieval aims to organize information by indicating connections between different elements [40].

- **Knowledge sharing** consists of exchanging knowledge units between entities so that each entity gets access to more than the knowledge it has been able to build up [28]. Obviously, each entity is then more prepared to make the correct choices in their field. In this way, unprecedented situations can be resolved satisfactorily. However, knowledge is currently exchanged inefficiently. This means that exchange mechanisms are restricted to very specific domains. This fact reduces knowledge propagation in space and time. To address this problem, the Knowledge Interchange Format (KIF) was designed [16]. KIF is a language designed to be used for exchange of knowledge between different expert systems by representing arbitrary knowledge units using the first order predicate logic.

Concerning the automatic maintenance of KBs (a.k.a. knowledge maintenance or knowledge retention), there are three important phases: knowledge meta-modeling, i.e. modeling knowledge about the KB, knowledge integration which consists on merging past and new knowledge, and knowledge validation to assure the correctness of the new knowledge added to the KB.

- **Knowledge meta-modeling** can be considered as a process for adding explicit descriptions (constructs and rules) of how a domain-specific KB is built [24]. In particular, this comprises a formalized specification of the domain-specific notations, a centralized repository about data such as meaning, relationships to other data, origin, usage, and format. This repository is mainly accessed by the various software modules of the KB itself, such as query optimizer, transaction processor or report generators.
- **Knowledge integration** is considered to be the process of incorporating new information into a body of existing knowledge with an interdisciplinary approach [18]. A possible technique which can be used is semantic matching [29]. This process involves determining how the new information and the existing knowledge interact, how existing knowledge should be modified to ac-

commodate the new information, and how the new information should be modified in light of the existing knowledge [32]. These techniques can be used for going beyond the literal lexical match of words and operate at the conceptual level when comparing specific labels for concepts (e.g., Finance) also yields matches on related terms (e.g., Economics, Economic Affairs, Financial Affairs, etc.). As another example, in the healthcare field, an expert on the treatment of cancer could also be considered as an expert on oncology, lymphoma or tumor treatment, etc.

- **Knowledge validation** is a critical process in the maintenance of the KBs. Validation consists of ensuring that something is correct or conforms to a certain standard. A knowledge engineer is required to carry out data collection and data entry, but they must use validation in order to ensure that the data they collect, and then enter into their systems, fall within the accepted boundaries of the application collecting the data [1]. Therefore, the ultimate goal of this process is to make the KB satisfy all test cases given by human experts [20]. This is further complicated by factors such as temporal validity, uncertainty and incompleteness. Most of current expert systems incorporate simple validation procedures within the program code. After the expert system is constructed, it is usually maintained by a domain expert.

Concerning explanation delivery, the purpose is that expert systems may be able to give the user clear explanations of what it is doing and what it has deduced. The most sophisticated expert systems are able to detect contradictions [3] in user information or in the knowledge and can explain them clearly, revealing at the same time the expert's knowledge and way of thinking, what makes the process much more interpretable.

### 3 Open problems

From the state-of-the-art, we can deduce that a lot of successful work have been done in the field of automated knowledge-base management during the last years. However, despite of these great advancements, there are still some problems that remain open. These problems should be addressed to support a more effective and efficient knowledge-base management. Therefore, the gist of these problems is to

support the complete life cycle for large KBs so that computer systems can exploit them to reflect the way human experts take decisions in their domains of expertise. These tasks are often pervasive because large KBs must be developed incrementally, this means that segments of knowledge are added separately to a growing body of knowledge [6]. Satisfactory results in this field can have a great impact in the advancement of many important and heterogeneous disciplines and fields of application. However, there are a number of challenging questions that should be successfully addressed in advance. These problems which are summarized as follows:

- The first problem concerns the **automatic generation of large KBs**. Every expert system has a major flaw: knowledge collection and its interpretation into rules is quite expensive in terms of effort and time [10]. Most expert systems have no automated methods to perform this task. Instead it is necessary to work manually, increasing the likelihood of errors and the costs in terms of money and time. In order to develop new methods for automatic knowledge learning, it is important to have a strong methodology for their evaluation and comparison. This problem is even more critical in environments working with large KBs, as it is not viable to manually evaluate the inclusion of new knowledge.
- The second problem concerns the efficiency of **methods for exploiting KBs**. These methods include: knowledge reasoning, knowledge sharing and knowledge retrieval (e.g. Question & Answering tools [48]). Beside quality, the efficiency of this kind of methods is of prime importance in dynamic applications, especially, when it is not possible to wait too long for the system to respond or when memory is limited. Current expert systems are mostly design-time tools which are usually not optimized, this means that many useful systems cannot be practically used mainly due to the lack of scalability.
- The third problem concerns automatic selection, combination and/or tuning of **methods for KB maintenance**. These methods include knowledge integration, meta-modeling or new knowledge validation. For example, the vital task of knowledge integration (inclusion of external knowledge in the KBs) requires complex methods for identifying semantic correspondences in order to

proceed with the merging of past and new knowledge [22]. For the detection of semantic correspondences, it is necessary to perform combination and self-tuning of algorithms that identify those semantic correspondences at run time [30]. This means that efficiency of the configuration of different search strategies becomes critical. As the number of available methods for KB maintenance as well the knowledge stored in the KB increases, the problem of their selection will become even more critical.

- The fourth problem concerns **explanation delivery** in order to improve the expert systems, thereby providing feedback to the system, users need to understand them. It is often not sufficient that a computational algorithm performs a task for users to understand it immediately. In order for expert systems to gain a wider acceptance and to be trusted by users, it will be necessary that they provide explanations of their results to users or to other software programs that exploit them. This information should be delivered in a clear and concise way so that it cannot be any place for misunderstanding.

## 4 Future challenges

In view of the state of the art and the open problems that need to be investigated, it is possible to identify four major future research challenges that should be addressed:

### 4.1 Challenge 1: methodology for the comparison and evaluation of KBs which have been automatically built.

We know that **evaluation of KBs** refers to the correct building of the content of a KB, that is, ensuring that its definitions correctly implement requirements or perform correctly in the real world. The goal is to prove compliance of the world model (if it exists and is known) with the world modeled formally. From the literature, we have found that the problem of evaluating an automatically-built KB involves six criteria:



- Accuracy which consists of determining the precision of the extracted knowledge and its level of confidence.
- Usefulness which consists of determining the relevancy of the knowledge for target tasks, its level of redundancy, and its level of granularity.
- Augmentation which consists of determining if the new knowledge added something new to the past knowledge.
- Explanation which consists of determining the provenance of the knowledge [12], and if there is something contradictory.
- Adaption which consists of determining if current knowledge could be adapted to new languages and domains and how much effort should be made to do that.
- Temporal qualification which consists of determining the temporal validity of the knowledge.

One possible way to evaluate these criteria could consist of treating the KB as a set of assertions, and use set-oriented measures such as precision and recall to determine the accuracy of the recently built KB. Treating each assertion as atomic avoids the need to perform alignment between the expert system output and ground truth. Comparing the expert system and ground truth KB should require encoding the assertions in compatible or mappable ontologies. Identifying the differences should take into account the logical dependencies between assertions for not over-penalizing an expert system for missing assertions from which many others are derivable [13]. Evaluation of temporal qualification can be partially handled by treating the KB as a sequence of fixed sets of assertions over time. Augmentation can also be examined by performing ablation studies over the assertions in the KB.

The TAC KBP 2013 Cold Start Track<sup>2</sup> could serve as a base for this research. The idea behind this workshop is to test the ability of proposed methods to extract specific knowledge from text and other sources and place it into a KB. The schema for the target KB is specified a priori, but the KB is otherwise empty to start. Expert systems should be able to process some sources, extracting information

---

<sup>2</sup><http://www.nist.gov/tac/2013/KBP/ColdStart>

about entities mentioned in the collection, adding the information to a new KB, and indicating how entry point entities mentioned in the collection correspond to nodes in the KB [33]. The sources consists of tens of thousands of news and web documents that contain entities that are not included in existing well-known KBs.

## **4.2 Challenge 2: improving the efficiency of the knowledge exploitation methods.**

The second challenge should lay on the development of strategies for improving the efficiency of tasks exploiting the KB. Moreover, these strategies should not alter the capability of current methods to produce desired results by comparing them with task requirements. These methods are those concerning to knowledge reasoning, knowledge retrieval, and knowledge sharing. It is necessary to focus in many different aspects and requirements brought by these exploitation methods. Some of them may concern on efficiency, e.g., time and space complexity of the algorithms developed, and the rest will concern the effectiveness in relation to efficiency, e.g. correctness, completeness, and so on. Therefore, the problem needs to be addressed from a point of view involving multi-decision criteria.

The ultimate goal is to measure and improve the extent to which time, effort or cost is well used for the intended KB exploitation methods. According the literature, efficiency issues are currently tackled through a number of computational strategies. This strategies could be organized as follows:

- Parallelization of exploitation methods.
- Distribution of exploitation methods over computers with available computational resources.
- Approximation of results, which over time become better (more complete).
- Modularization of the KB, yielding smaller more targeted exploitation tasks.
- Optimization of existing exploitation methods.

To the best of our knowledge the first two items above remain largely unaddressed so far. Maybe the reason is that researchers thought that more computing power does not necessarily improve effectiveness of exploitation methods. However, it is possible to think that, at least at the beginning, it would

accelerate the first run and the analysis of the bottlenecks [42]. In order that parallelization and distribution become really useful it is possible that a traditional approach of divide-and-conquer with an iterative procedure whose result converges towards completeness over time will be necessary. Concerning the last three items, some solutions have been proposed in the past but only in relation to the TBox [42]. Therefore, there are still many interesting insights on potential further developments of the themes of approximation, modularization and optimization in the ABox.

### 4.3 Challenge 3: automatically selection, combination and tuning of algorithms for the maintenance of a KB.

The research challenge 3 should focus primarily in the development of methods for the maintenance of the KBs. These methods include the maintenance of the meta-knowledge, the knowledge integration tasks, and the knowledge validation using test cases from users. It is necessary to design solutions avoiding to choose these methods arbitrarily, but in automatic way, by applying machine learning techniques on an ensemble of methods. Sequential and parallel composition of methods should be also studied. Hereby it is necessary to learn rules for the correctness on the output of different methods and additional information about the nature of the elements to be operated. The final goal is to leverage the strengths of each individual method.

Let us focus in **semantic matching** which is a well established field of research in the field of **Knowledge Integration**. Two entities in a KB are assigned a score based on the likeness of their meaning [37]. Automatically performing semantic matching is considered to be one of the pillars for many computer related fields since a wide variety of techniques rely on a good performance when determining the meaning of data they work with.

More formally, we can define semantic matching as a function  $\mu_1 \times \mu_2 \rightarrow \mathbb{R}$  that associates the degree of correspondence for the entities  $\mu_1$  and  $\mu_2$  to a score  $s \in \mathbb{R}$  in the range  $[0, 1]$ , where a score of 0 states for not correspondence at all, and 1 for total correspondence of the entities  $\mu_1$  and  $\mu_2$ .

Traditionally, the way to compute the degree of correspondence between entities has been addressed

from two different perspectives: using semantic similarity measures and semantic relatedness measures. Fortunately, recent works have clearly defined the scope of each of them [39]. Firstly, semantic similarity is used when determining the taxonomic proximity between objects. For example, automobile and car are similar because the relation between both terms can be defined by means of a taxonomic relation. Secondly, the more general concept of semantic relatedness considers taxonomic and relational proximity. For example, blood and hospital are not completely similar, but there is still possible to define a naive relation between them because both belong to the world of healthcare.

In most of cases, the problem to face is more complex since it does not involve the matching of two individual entities only, but two complete KBs. This can be achieved by computing a set of semantic correspondences between individual entities belonging to each of the two KBs. A set of semantic correspondences between entities is often called an alignment. It is possible to define formally an alignment  $A$  as a set of tuples in the form  $\{(id, \mu_1, \mu_2, r, s)\}$ , where  $id$  is a unique identifier for the correspondence,  $\mu_1$  and  $\mu_2$  are the entities to be compared,  $r$  is the kind of relation between them, and  $s$  the score in the range  $[0, 1]$  stating the degree of correspondence for the relation  $r$ .

Therefore, when matching two KBs, the challenge that scientists try to address consists of finding an appropriate semantic matching function leading to a high quality alignment between these two KBs. Quality here is measured by means of a function  $A \times A_{ideal} \rightarrow R \times R$  that associates an alignment  $A$  and an ideal alignment  $A_{ideal}$  to two real numbers  $\in [0, 1]$  stating the precision and recall of  $A$  in relation to  $A_{ideal}$ .

Precision represents the notion of accuracy, that it is to say, states the fraction of retrieved correspondences that are relevant for the matching task (0 stands for no relevant correspondences, and 1 for all correspondences are relevant). Meanwhile, Recall represents the notion of completeness, thus, the fraction of relevant correspondences that were retrieved (0 stands for not retrieved correspondences, and 1 for all relevant correspondences were retrieved).

#### **4.4 Challenge 4: methods which can explain what happens inside a KB in a clear and concise way.**

The fourth challenge should find a way to provide explanations in a simple, clear and precise way to the users or software applications in order to facilitate informed decision making. In particular, most of techniques used by expert systems do not yield simple or symbolic explanations. It is necessary to take into account that different types of explanations may be needed. For example, if negotiating agents trust each others information sources, explanations should focus on the manipulations. If on the other hand, the sources may be suspect, explanations should focus on meta information about sources. If a user wants an explanation of the reasoning engine used by the expert system, a more complex explanation may be required.

There are some preliminary works which try to address the problem, i.e. laying the foundations about how an expert system should deliver explanations. According the literature, there are a set of requirements that are intended to act as criteria for the evaluation of explanations given by expert systems. For instance, Moore [34] states that explanations given by an expert system should have the characteristics listed below:

- **Naturalness.** Explanations should appear natural to the user. Explanations that are not structured according to standard pattern of human discourse often obscure critical elements of an explanation.
- **Responsiveness.** An expert system should have the ability to accept feedback from the user and to answer follow-up questions.
- **Flexibility.** An explanation should be able to offer an explanation in more than one way in order to accommodate differences knowledge and abilities of users.
- **Sensitivity.** An explanation should take into account the user's goals, the problem solving situation and the previous explanatory dialogue.
- **Fidelity.** An explanation should accurately reflect the knowledge from the KB and reasoning from the engine.

- **Sufficiency.** An explanation should be able to answer a range of questions users want to ask and not to be limited to those questions predicted by developers.
- **Extensibility.** An expert system should be easy to extend in order to accommodate questions not conceived at design time.

To the best of our knowledge, current expert systems are not able to cover all of these characteristics. Some expert systems try to use pre-written text attached to knowledge units or apply simple transformations to produce explanations from program code. This type of explanation is very simple to generate and easy to understand. However, this way to proceed is very far from the guidelines we have reviewed above. It is necessary to follow these guidelines for designing and evaluating new explanation delivery mechanisms so that they can meet this set of desirable characteristics.

## **4.5 Impact of future challenges**

As already described before, the importance of satisfactorily addressing these four research challenges is given by contributions in several areas.

- Concerning the systematic development of a methodology for the comparison and evaluation of recently built KBs; the essence of the creation stage is that no one fully understands the idea or emerging body of knowledge, not even those creating it [45]. The process of creation is messy by nature and does not respond well to formal methodologies or rigid time lines [25]. The design of an evaluation methodology can definitely help to develop automatic solutions for addressing this problem since it is possible to learn the impact of each decision in the final quality of the KB.
- The importance of creating strategies for improving the efficiency of the knowledge exploitation methods is out of doubt. The impact of this approach is given by the fact that current methods for exploiting KBs are developed without taking into account its efficiency. For example, reasoning is a very complex process which needs a lot of time and memory space to be performed. As the KB grows, this issue become more critical, so it is quite important to build methods that meet their goals but these methods should also meet them efficiently.

- The idea of developing methods to automatically select, combine and/or tune algorithms for the maintenance of a KB is of vital importance [31]. The design of novel approaches that attempt to tune and adapt automatically current solutions to the settings in which an user or application operates are vital for a real automatic maintenance of large expert systems become real. This may involve the run time reconfiguration of the methods by finding their most appropriate parameters, such as thresholds, weights, and coefficients. In this way, tasks than currently are performed by humans can be automated.
- The novelty of the research concerning explanation delivery in a simple, clear and precise way to the users or software applications can have a better understanding of the knowledge provided by the expert systems. The idea to standardize explanations or proofs of tasks inside the KB in order to facilitate the interaction of expert systems with people or other software programs will have a positive impact in the development of this field and widespread of expert systems.

#### **4.6 Fields of application that could get benefit**

The spectrum of potential application domains that could be benefited from these advances is really wide. Let us summarize some application fields which can be benefited from satisfactorily addressing the aforementioned research challenges:

**Financial decision support.** The financial services industry has been a traditional user of expert systems [8]. Some systems have been created to assist bankers in determining whether to make loans to businesses and individuals, insurance firms have used expert systems to assess the risk presented by a given customer or software applications has been built for foreign exchange trading. Therefore, advances in this field could be beneficial for improving the traditional systems, by aggregating new knowledge sources, improving the real time performance, explaining the rationale behind financial decisions, and so on.

**Manufacturing industry.** Configuration, whereby a solution to a problem is synthesized from a given set of elements related by a set of constraints, is one of the most important of expert system ap-

plications [19]. Configuration applications were pioneered by computer companies as a means of facilitating the manufacture of semi-custom minicomputers. Nowadays, expert systems have found its way into use in a wide range of different industries, from textile industry where fabrics must be optimally cut, to failure detection in factories which consists of deducing faults and suggest corrective actions for malfunctioning devices or processes [26].

**Question & Answering systems.** Expert systems in this field are able to deliver knowledge that is relevant to the user's problem, in the context of the user's problem [35]. In case ne improvements may be proposed, very interesting Q & A systems would be built. For example, a computational assistant which may give some hints to a user on appropriate grammatical usage in a text, or a tax advisor that accompanies a tax preparation program and advises the user on individual tax policy. Therefore, advances in this field could help to the popularization of this kind of systems in many additional fields like education, eTourism, personal finance, and so on.

**Scientific research.** Scientists need to be able to easily gain access to all information about chemical compounds, biological systems, diseases, and the interactions between these kinds of entities, and this requires data to be effectively integrated in order to provide a greater level view to the user, for instance, a complete view of biological activity [21]. Therefore, advances on the automatic building, exploitation and maintenance of large KBs will certainly help scientists to more easily work with all knowledge of their interest. More specifically, the benefits include the aggregation of heterogeneous sources using explicit semantics, and the expression of rich and well-defined models for working with knowledge.

## 5 Conclusions

In this work, we have presented the current state-of-the-art, problems that are still open and future research challenges for automated knowledge-base management. Our aim is to overview the past, present and future of this discipline so that complex expert systems exploiting knowledge from knowledge bases can be automatically developed and practically used.



Concerning the state-of-the art, we have surveyed the current methods and techniques covering the complete life cycle for automated knowledge management, including automatic building, exploitation and maintenance of KBs, and all their associated tasks. That is to say, knowledge acquisition, representation, storage and manipulation for automatic building of KBs. Knowledge reasoning, retrieval and sharing for exploitation of KBs, and knowledge meta-modeling, integration and validation for the automatic maintenance.

From the current state-of-the-art, we have identified some problems that remain open and represent a bottleneck that is avoiding the rapid proliferation of systems of this kind. In fact, we have identified flaws in some areas including: a) automatic generation of large KB, b) lack of efficiency in methods for exploiting KBs, c) lack of automatic methods for smartly configuring maintenance tasks, and d) need of improving explanation delivery mechanisms.

Finally, in view of the state-of-the-art and open problems, we have identified four research challenges that should be addressed in the future: a) methodology for the comparison and evaluation of KBs which have been automatically built, b) improving the efficiency of the knowledge exploitation methods, c) automatic selection, combination and tuning of algorithms for the maintenance of a KB, and d) design of methods which can explain what happens inside a KB in a clear and concise way. We think that satisfactorily addressing these challenges could have a positive impact not only in basic research but in a lot of application domains too.

The final goal of automated knowledge-base management should be to empower people and intelligent software applications with the knowledge that they need to make well-informed decisions in an increasingly complex and changing world.

## **Acknowledgments**

We would like to thank in advance the reviewers for their time and consideration. This work has been partially funded by Vertical Model Integration within Regionale Wettbewerbsfähigkeit OO 2007-2013 by the European Fund for Regional Development and the State of Upper Austria.

## References

- [1] P. Ardimento, M. T. Baldassarre, M. Cimitile, and G. Visaggio. Empirical validation of knowledge packages as facilitators for knowledge transfer. *JIKM*, 8(3):229–240, 2009.
- [2] M. Arevalillo-Herráez, D. Arnau, and L. Marco-Giménez. Domain-specific knowledge representation and inference engine for an intelligent tutoring system. *Knowl.-Based Syst.*, 49:97–105, 2013.
- [3] N. Arman. Fault detection in dynamic rule bases using spanning trees and disjoint sets. *Int. Arab J. Inf. Technol.*, 4(1):67–72, 2007.
- [4] R. Balch, S. Schrader, and T. Ruan. Collection, storage and application of human knowledge in expert system development. *Expert Systems*, 24(5):346–355, 2007.
- [5] R. Balzer. Automated enhancement of knowledge representations. In *IJCAI*, pages 203–207, 1985.
- [6] R. Bareiss, B. W. Porter, and K. S. Murray. Supporting start-to-finish development of knowledge bases. *Machine Learning*, 4:259–283, 1989.
- [7] K. Barker, J. Blythe, G. C. Borchardt, V. K. Chaudhri, P. Clark, P. R. Cohen, J. Fitzgerald, K. D. Forbus, Y. Gil, B. Katz, J. Kim, G. W. King, S. Mishra, C. T. Morrison, K. S. Murray, C. Otstott, B. W. Porter, R. Schrag, T. E. Uribe, J. M. Usher, and P. Z. Yeh. A knowledge acquisition tool for course of action analysis. In *IAAI*, pages 43–50, 2003.
- [8] O. Ben-Assuli. Assessing the perception of information components in financial decision support systems. *Decision Support Systems*, 54(1):795–802, 2012.
- [9] Y. Chen, L.-J. Zhang, and Q. Wang. Intelligent scheduling algorithm and application in modernizing manufacturing services. In *IEEE SCC*, pages 568–575, 2011.
- [10] P. Cimiano, A. Hotho, and S. Staab. Comparing conceptual, divide and agglomerative clustering for learning taxonomies from text. In *ECAI*, pages 435–439, 2004.

- [11] J. de Bruijn, D. Pearce, A. Polleres, and A. Valverde. A semantical framework for hybrid knowledge bases. *Knowl. Inf. Syst.*, 25(1):81–104, 2010.
- [12] R. Q. Dividino, S. Schenk, S. Sizov, and S. Staab. Provenance, trust, explanations - and all that other meta knowledge. *KI*, 23(2):24–30, 2009.
- [13] M. Dredze, P. McNamee, D. Rao, A. Gerber, and T. Finin. Entity disambiguation for knowledge base population. In *COLING*, pages 277–285, 2010.
- [14] A. Felfernig and F. Wotawa. Intelligent engineering techniques for knowledge bases. *AI Commun.*, 26(1):1–2, 2013.
- [15] I. Filali, F. Bongiovanni, F. Huet, and F. Baude. A survey of structured p2p systems for rdf data storage and retrieval. *T. Large-Scale Data- and Knowledge-Centered Systems*, 3:20–55, 2011.
- [16] M. L. Ginsberg. Knowledge interchange format: the kif of death. *AI Magazine*, 12(3):57–63, 1991.
- [17] F. Gomez and C. Segami. Semantic interpretation and knowledge extraction. *Knowl.-Based Syst.*, 20(1):51–60, 2007.
- [18] H.-F. Hung, H.-P. Kao, and Y.-Y. Chu. An empirical study on knowledge integration, technology innovation and experimental practice. *Expert Syst. Appl.*, 35(1-2):177–186, 2008.
- [19] A. Jones, R. H. Weston, B. Grabot, and B. Hon. Decision making in support of manufacturing enterprise transformation. *ADS*, 2013, 2013.
- [20] E. S. Jr. and H. T. Dinh. On automatic knowledge validation for bayesian knowledge bases. *Data Knowl. Eng.*, 64(1):218–241, 2008.
- [21] P. D. Karp. Development of large scientific knowledge bases. In *ICAART (I)*, page 23, 2010.
- [22] J. L. Kenney and S. P. Gudergan. Knowledge integration in organizations: an empirical assessment. *J. Knowledge Management*, 10(4):43–58, 2006.

- [23] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Extracting large-scale knowledge bases from the web. In M. P. Atkinson, M. E. Orlowska, P. Valduriez, S. B. Zdonik, and M. L. Brodie, editors, *VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases, September 7-10, 1999, Edinburgh, Scotland, UK*, pages 639–650. Morgan Kaufmann, 1999.
- [24] I. Lahoud, D. Monticolo, V. Hilaire, and S. Gomes. A metamodeling and transformation approach for knowledge extraction. In *NDT (2)*, pages 54–68, 2012.
- [25] D. B. Lenat. Problems of scale in building, maintaining and using very large formal ontologies. In *FOIS*, page 3, 2006.
- [26] W. Lepuschitz, V. Jirkovský, P. Kadera, and P. Vrba. A multi-layer approach for failure detection in a manufacturing system based on automation agents. In *ITNG*, pages 1–6, 2012.
- [27] H. J. Levesque. A completeness result for reasoning with incomplete first-order knowledge bases. In *KR*, pages 14–23, 1998.
- [28] W.-B. Lin. The effect of knowledge sharing model. *Expert Syst. Appl.*, 34(2):1508–1521, 2008.
- [29] J. Martínez-Gil, E. Alba, and J. F. Aldana-Montes. Optimizing Ontology Alignments by Using Genetic Algorithms. *NatuReS*, 2008.
- [30] J. Martínez-Gil and J. F. Aldana-Montes. Evaluation of two heuristic approaches to solve the ontology meta-matching problem. *Knowl. Inf. Syst.*, 26(2):225–247, 2011.
- [31] J. Martínez-Gil and J. F. Aldana-Montes. An overview of current ontology meta-matching solutions. *Knowledge Eng. Review*, 27(4):393–412, 2012.
- [32] J. Martínez-Gil. An overview of textual semantic similarity measures based on web intelligence. *Artif. Intell. Rev.*, 42(4): 935–943, 2014.
- [33] P. McNamee, H. T. Dang, H. Simpson, P. Schone, and S. Strassel. An evaluation of technologies for knowledge base population. In *LREC*, 2010.

- [34] J. D. Moore. *Participating in explanatory dialogues : interpreting and responding to questions in context*. ACL-MIT Press series in natural language processing. Cambridge, Mass. MIT Press, 1995. A Bradford book.
- [35] J. Morrissey and R. Zhao. R/request: A question answering system. In *FQAS*, pages 79–90, 2013.
- [36] D. Nevo, I. Benbasat, and Y. Wand. The knowledge demands of expertise seekers in two different contexts: Knowledge allocation versus knowledge retrieval. *Decision Support Systems*, 53(3):482–489, 2012.
- [37] G. Pirrò. A semantic similarity metric combining features and intrinsic information content. *Data Knowl. Eng.*, 68(11):1289–1308, 2009.
- [38] I. J. Rudas, E. Pap, and J. C. Fodor. Information aggregation in intelligent systems: An application oriented approach. *Knowl.-Based Syst.*, 38:3–13, 2013.
- [39] D. Sánchez, M. Batet, and D. Isern. Ontology-based information content computation. *Knowl.-Based Syst.*, 24(2):297–303, 2011.
- [40] A. Sharma and K. D. Forbus. Automatic extraction of efficient axiom sets from large knowledge bases. In *AAAI*, 2013.
- [41] S. Shekarpour. Dc proposal: Automatically transforming keyword queries to sparql on large-scale knowledge bases. In *International Semantic Web Conference (2)*, pages 357–364, 2011.
- [42] P. Shvaiko and J. Euzenat. Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.*, 25(1):158–176, 2013.
- [43] S.-S. Wang and D. Liu. Knowledge representation and reasoning for qualitative spatial change. *Knowl.-Based Syst.*, 30:161–171, 2012.
- [44] S.-J. Yen, Y.-C. Wu, J.-C. Yang, Y.-S. Lee, C.-J. Lee, and J.-J. Liu. A support vector machine-based context-ranking model for question answering. *Inf. Sci.*, 224:77–87, 2013.

- [45] L. Zang, C. Cao, Y. Cao, Y. Wu, and C. Cao. A survey of commonsense knowledge acquisition. *J. Comput. Sci. Technol.*, 28(4):689–719, 2013.
- [46] X. Zhang and Z. Lin. An argumentation framework for description logic ontology reasoning and management. *J. Intell. Inf. Syst.*, 40(3):375–403, 2013.
- [47] L. Zhao and R. Ichise. Instance-based ontological knowledge acquisition. In *ESWC*, pages 155–169, 2013.
- [48] G. Zhou, Y. Liu, F. Liu, D. Zeng, and J. Zhao. Improving question retrieval in community question answering using world knowledge. In *IJCAI*, 2013.

## Biography

Jorge Martinez-Gil is a Spanish-born computer scientist working in the Knowledge Engineering field. He got his PhD in Computer Science from University of Malaga in 2010. He has held a number of research position across some European countries (Austria, Germany, Spain). He currently holds a Team Leader position within the group of Knowledge Representation and Semantics from the Software Competence Center Hagenberg (Austria) where he is involved in several applied and fundamental research projects related to knowledge-based technologies.