Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cosrev

Survey

Automated knowledge base management: A survey

*Jorge Martinez-Gil**

Software Competence Center Hagenberg, Austria

HIGHLIGHTS

- We present the state-of-the-art concerning knowledge base management systems.
- We review the open problems that remain open in this field.
- We identify the future research challenges for automatically building, exploiting and maintaining knowledge base management systems.

ARTICLE INFO

Article history:
Received 20 March 2014
Received in revised form
16 September 2015
Accepted 30 September 2015
Published online 30 October 2015

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 the tasks involved in the building, exploitation and maintenance of KBs are far from being trivial, and 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.

Contents

* Tel.: +43 7236 3343 838.
E-mail address: jorge.martinez-gil@scch.at.

http://dx.doi.org/10.1016/j.cosrev.2015.09.001
1574-0137/

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, material, and other constraints [1].

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 [2]. 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 proposes those challenges that should be addressed and explains 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 [3].

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 [4]. 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.[1]

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 [5]. Then, concrete instances for these concepts should be also extracted [6]. This usually involves the use of natural language processing techniques [7]. Then statistical or symbolic techniques are applied to extract relations between the terms and concepts [8]. 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 [9]. The conditions and constraints of knowledge formation and organization have to be formally specified [10]. A notation of this kind follows a logical specification using expressions and symbolical structures, such as taxonomies, classes, and axioms [11].
- 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 [12]. This involves using standard ways to communicate knowledge units and retrieve them [13]. 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

[1] 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.