Simon Kendal
Malcolm Creen

An Introduction to Knowledge Engineering

Springer

An Introduction to Knowledge Engineering

S.L. Kendal and M. Creen

An Introduction to Knowledge Engineering

With 33 figures

Springer

S.L. Kendal
School of Computing & Technology
University of Sunderland
Tyne and Wear
UK

M. Creen
Learning Development Services
University of Sunderland
Tyne and Wear
UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2006925857

ISBN 10: 1-84628-475-9 Printed on acid-free paper
ISBN 13: 978-1-84628-475-5

Printed in the United States of America (TB/MVY)

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

To my wife Janice, who is a better partner than I could wish for, and my daughter Cara, a gift from God.

—Simon Kendal

To Lillian and Sholto—with love.

—Malcolm Creen

Foreword

An *Introduction to Knowledge Engineering* presents a simple but detailed exploration of current and established work in the field of knowledge-based systems and related technologies. Its treatment of the increasing variety of such systems is designed to provide the reader with a substantial grounding in such technologies as expert systems, neural networks, genetic algorithms, case-based reasoning systems, data mining, intelligent agents and the associated techniques and methodologies.

The material is reinforced by the inclusion of numerous activities that provide opportunities for the reader to engage in their own research and reflection as they progress through the book. In addition, self-assessment questions allow the student to check their own understanding of the concepts covered.

The book will be suitable for both undergraduate and postgraduate students in computing science and related disciplines such as knowledge engineering, artificial intelligence, intelligent systems, cognitive neuroscience, robotics and cybernetics.

Contents

1

An Introduction to Knowledge Engineering

Introduction

This chapter introduces some of the key concepts in knowledge engineering. Almost all of the topics are covered in summary form, and they will be explained in more detail in subsequent chapters.

The chapter consists of three sections:

1. Data, information and knowledge
2. Skills of a knowledge engineer
3. An introduction to knowledge-based systems (KBSs).

Objectives

By the end of this chapter, you will be able to:

- define knowledge and explain its relationship to data and information
- distinguish between knowledge management and knowledge engineering
- explain the skills required of a knowledge engineer
- comment on the professionalism, methods and standards required of a knowledge engineer
- explain the difference between knowledge engineering and artificial intelligence
- define KBSs
- explain what a KBS can do
- explain the differences between human and computer processing
- state a brief definition of expert systems, neural networks, case-based reasoning, genetic algorithms, intelligent agents and data mining.

SECTION 1: DATA, INFORMATION AND KNOWLEDGE

Introduction

This section defines knowledge and explains its relationship to data and information.

Objectives

By the end of this section you will be able to:

- develop a working definition of knowledge and describe its relationship to data and information.

What Is Knowledge Engineering?

'Knowledge engineering is the process of developing knowledge based systems in any field, whether it be in the public or private sector, in commerce or in industry' (Debenham, 1988).

But what, precisely, is knowledge?

What Is Knowledge?

Knowledge is 'The **explicit** functional associations between items of information and/or data' (Debenham, 1988).

Data, Information and Knowledge

What is data? Is it the same as information? Before we can attempt to understand what knowledge is, we should at least attempt to come closer to establishing exactly what data and information are.

Activity 1

The following activity introduces you to the concepts of data and information:

1. Read the following descriptions and definitions of 'data' drawn from a variety of sources:

   Data (the plural of *datum*) are just raw facts (Long and Long, 1998).

   Data . . . are streams of raw facts representing events . . . before they have been arranged into a form that people can understand and use (Laudon and Laudon, 1998).

   Data is comprised of facts (Hayes, 1992).

   Recorded symbols (McNurlin and Sprague, 1998).

2. Make a note of any factors common to two or more of the descriptions.

Feedback 1

You will have noticed that data is often spoken of as the same as 'facts'—often 'raw' and, in the first quotation, considered to move in a 'stream'. The final quotation from Hayes appears to look deeper in defining data more fundamentally as recorded symbols.

Hayes actually goes on to insist that data are not facts and that treating them as such can produce 'innumerable perversions' for example, in the form of propaganda or lies—which are still 'data'.

You do not need to accept or reject any of the definitions you encounter—simply be aware that there are no universally accepted definitions of data.

Similarly, in connection to the meaning of the term 'information', we find that there are many attempts at definitions in the textbooks on information systems and information technology. In many ways the meanings of the words 'data' and 'information' only become clearer when we approach the differences between them. The following activity will help you to appreciate this.

Activity 2

This activity introduces you to some definitions of information and its relationship to data.

1. Read the following definitions and descriptions of information. As in the last activity look for common denominators.

   That property of data which represents and measures effects of processing them (Hayes, 1992).

   By information we mean data that have been shaped into a form that is meaningful and useful to human beings (Laudon and Laudon, 1998).

Information is data that have been collected and processed into a meaningful form. Simply, information is the meaning we give to accumulated facts (data) (Long and Long, 1998).

Information is the emergent property which comes from processing data so that it is transformed into a structured whole (Harry, 1994).

Information is data presented in a form that is meaningful to the recipient (Senn, 1990).

Information is data in context (McNurlin and Sprague, 1998).

Information is data endowed with relevance and purpose (Drucker, 1988).

2. Make a note of any similarities between the different descriptions.

Feedback 2

You should have noted that information is commonly thought to be data, processed or transformed into a form or structure suitable for use by human beings. Such words as 'meaning', 'meaningful', 'useful' and 'purpose' are in evidence here.

You may also have noted that information is considered a property of data. This implies that the former cannot exist without the latter.

In the definitions of information you will have seen how the meaning of the word becomes clearer when the differences between it and data are considered. For example, whereas the 'rawness' of data was emphasised earlier, information is considered to be some refinement of data for the purposes of human use.

In addition, the words 'knowledge' and 'communication' have emerged as having a relationship to data and information. What is also worth emphasising at this point is that the interface between data and a human being's interpretation of it is where information—determined by 'meaning'—really emerges.

The two terms are still often used interchangeably and no definition of either will apply in all the situations you might encounter.

Knowledge

In common language, the word knowledge is obviously related to information, but it is clear that they are not the same thing. So, how can we define knowledge in the same flexible way in which we have arrived at working definitions of information and data?

Activity 3

This activity extends your understanding of data and information.

Look at the seven topics described briefly below. Which of them would you consider yourself as ‘knowing’, and which would you consider yourself as having information about?

(a) A second language in which you are fluent.
(b) The content of a television news programme.
(c) A close friend.
(d) A company’s annual report.
(e) Your close friend’s partner whom you have yet to meet.
(f) The weather on the other side of the world.
(g) The weather where you are now.

Feedback 3

It is probable—but by no means certain—that you will have been inclined to consider items (a), (c) and (g) as things you can know about and the others as things for which you may have information. Note that the items that you would not describe yourself as possessing knowledge of could actually become known if circumstances were different, e.g. you might come to know your close friend’s partner.

It is also worth noting that all of this depends on individual perceptions rather than measurable facts. You may only think you know your close friend. Similarly, your fluency in the second language will always be relatively poorer than that of a native speaker.

Activity 4

This activity brings you closer to a definition by helping you highlight the differences between having information and possessing knowledge.

What would you suggest is the primary characteristic that distinguishes the ‘having information’ situations from the ‘knowing’ situations you categorised in the previous activity? You will need to make sure that your description does not simply describe information or data, but must particularly take account of the former.

Feedback 4

You should have been able to identify specific characteristics of knowledge that distinguish it from information similar to those highlighted in the following quotations. According to experts in the field, knowledge is:

the result of the understanding of information (Hayes, 1992)
the result of internalising information (Hayes, 1992)
collected information about an area of concern (Senn, 1990)
information with direction or intent—it facilitates a decision or an action (Zachman, 1987).

Here it has become clear that knowledge is what someone has after understanding information. Often this understanding follows the development of a detailed or long-term relationship with the known person or thing. Such a process can often be accelerated when the need to use the information for a critical decision arises. This application of information to a decision or area of concern is particularly relevant in an organisational situation.

However, it should be clear that data, information and knowledge are not static things in themselves but stages in the process of using data and transforming it into knowledge. On this basis they can be considered points along a continuum, moving from less to more usefulness to a human being, in much the same way as we all move along a continuum from young to old, but at no point can we be defined as either.

Activity 5

Temperature and humidity readings are taken from various locations around one city. These readings are taken four times each day, and the results collated in a central location.

The city is 12 miles in diameter. Readings taken on the periphery of the city can show, over time, how rain or adverse weather conditions start at one side of the city and move across to the other side.

Details of adverse weather can be used to warn weather-sensitive activities such as cricket or tennis matches when to expect a break in play.

Explain how a series of temperature and humidity readings can be transformed from data into knowledge.

Feedback 5
Data. Individual temperature and humidity readings, by themselves, are simply numbers, and therefore represent data.

Information. Information on where the readings have been taken (e.g. at which point in the city) and at what time provides a trend to show how the temperature is currently changing. This information can be used by someone to make a decision.

Knowledge. Knowing how the temperature and humidity are changing AND, knowing about how the weather can affect people living or working in the city will allow decisions to be made concerning the use of umbrellas, warm clothing, running a cricket or tennis match, etc. In this situation, two or more sets of information are related and can be processed to reach a decision.

The movement from data to knowledge implies a shift from facts and figures to more abstract concepts, as shown in Figure 1.1.

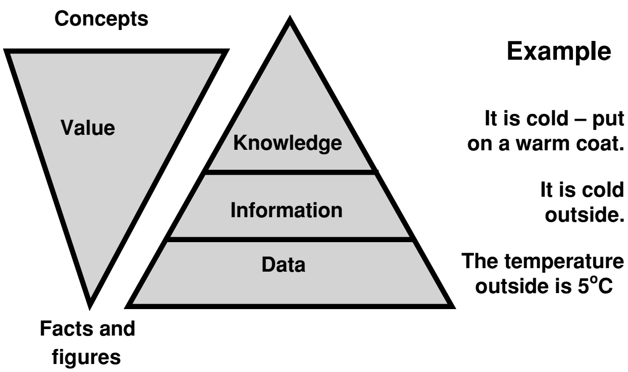

FIGURE 1.1. Data, information and knowledge.

In other words:

It is 5°C—data.
It is cold—information.
It is cold outside AND if it is cold you should wear a warm coat—knowledge.

From a knowledge engineering perspective, it is useful to consider knowledge as something that can be expressed as a rule or useful to assist a decision, i.e.,

IF it is cold outside THEN wear a warm coat.

The perceived value of data increases as it is transferred into knowledge, because the latter enables useful decisions to be made.

Activity 6

Knowledge engineering normally involves five distinct steps (listed below) in transferring human knowledge into some form of knowledge based system (KBS).

Explain what you think should be involved in each of these activities.

1. Knowledge acquisition
2. Knowledge validation
3. Knowledge representation
4. Inferencing
5. Explanation and justification.

Feedback 6

Knowledge acquisition involves obtaining knowledge from various sources including human experts, books, videos and existing computer sources of data such as databases and the Internet.

In *knowledge validation***,** knowledge is checked using test cases for adequate quality.

Knowledge representation involves producing a map of the knowledge and then encoding this knowledge into the knowledge base.

Inferencing means forming links (or inferences) in the knowledge in the computer software so that the KBS can make a decision or provide advice to the user.

Explanation and justification involves additional computer program design, primarily to help the computer answer questions posed by the user and also to show how a conclusion was reached using knowledge in the knowledge base.

Knowledge Engineering and Knowledge Management

The terms ‘knowledge management’ and ‘knowledge engineering’ seem to be used as interchangeably as the terms data and information used to be. The term ‘manage’ relates to exercising executive, administrative and supervisory direction, whereas, to engineer is to lay out, construct or contrive or plan out, usually with more or less subtle skill and craft.

The main difference seems to be that the (knowledge) manager establishes the direction the process should take, where as the (knowledge) engineer develops the means to accomplish that direction.

We should therefore find knowledge managers concerned with the knowledge needs of the enterprise, e.g. discovering what knowledge is needed to make decisions and enable actions. They should be taking a key role in the design of the enterprise and from the needs of the enterprise they should be establishing the enterprise level knowledge management policies.

On the other hand, if we were to look in on the knowledge engineers we should find them concerned with data and information *representation and encoding methodologies*, data repositories, etc. The knowledge engineers would be interested in what *technologies* are needed to meet the enterprise's knowledge management needs.

The knowledge engineer is most likely a computer scientist specialising the development of knowledge bases but a knowledge manager may be the chief information officer or the person in charge of the information resource management.

Summary

This section has introduced the concept of knowledge engineering and the relationship between data, information and knowledge.

Self-Assessment Question

Try and think of other systems within your immediate environment that result in data being transferred into information and then knowledge.

Answer to Self-Assessment Question

You might have thought of the following example:
50 litres (*Data*)—e.g. the amount of petrol your car can hold.
Having filled the tank, this can implicitly indicate that you can now travel 320 miles. (*Information*)
With the information above, a map and the addresses of several friends, you can now decide who you can visit within a 160 mile radius (assuming that the next refuelling will take place back at home). (*Knowledge*)

SECTION 2: SKILLS OF A KNOWLEDGE ENGINEER

Introduction

This section introduces one of the most important people in knowledge engineering; namely the knowledge engineer. The knowledge engineer is responsible for obtaining knowledge from human experts and then entering this knowledge into some form of KBS. To undertake these activities, specific skills are required.

Objectives

By the end of this section you will be able to:

- explain the skills and knowledge required of a knowledge engineer
- comment on the professionalism, methods and standards required of a knowledge engineer.

Knowledge Required of a Knowledge Engineer

To begin with, a knowledge engineer must extract knowledge from people (human experts) that can be placed into knowledge based systems (KBSs).

This knowledge must then be represented in some format that is understandable both to the knowledge engineer, the human expert and the programmer of the KBS.

A computer program, which processes that knowledge or makes inferences, must be developed, and the software system that is being produced must be validated. The knowledge engineer may be involved in the development of the program, or this may be delegated to another person.

In developing these systems the knowledge engineer must apply methods, use tools, apply quality control and standards.

To undertake these activities, the knowledge engineer must plan and manage projects, and take into account human, financial and environmental constraints.

Overview of Knowledge Engineers Work

To summarise the above points, knowledge engineering includes the process of knowledge acquisition, knowledge representation, software design and implementation.

To meet the objective of designing a KBS, the knowledge engineer will have to:

- acquire the knowledge from the expert to be used in the system
- use an appropriate method for representing knowledge in a symbolic, processable form.

This means that to deserve the title of knowledge engineer we must really apply professional and rigorous approaches to the development of a product. The engineer will also use various techniques to ensure quality and work to standards.

Knowledge engineering is a multi-stage process, and traditionally a business being tackled by a range of professionals. These include psychologists, computer scientists, software engineers, project managers, systems analysts, domain (or subject) experts and knowledge specialists.

Types of Knowledge

The knowledge engineer will normally be dealing with three types of knowledge:

- *Declarative knowledge* tells us facts about things. For example, the statement 'A light bulb requires electricity to shine' is factually correct.
- *Procedural knowledge* provides alternative actions based on the use of facts to obtain knowledge. For example, an individual will normally check the amount of water in a kettle before turning it on; if there is insufficient water in the kettle, then more will be added.
- *Meta-knowledge* is knowledge about knowledge. It helps us understand how experts use knowledge to make decisions. For example, knowledge about planes and trains might be useful when planning a long journey and knowledge about footpaths and bicycles might be useful when planning a short journey.

A knowledge engineer must be able to distinguish between these three types of knowledge and understand how to codify different knowledge types into some form of KBS.

Activity 7

A knowledge engineer will be involved in the following tasks:

- Advising the expert on the knowledge required for a system
- Acquiring knowledge from the expert
- Encoding the knowledge in some form ready for inclusion in the knowledge base
- Entering the knowledge into a knowledge base on a computer system
- Validating the knowledge in that knowledge base to ensure that it is accurate
- Training users to access and use the knowledge in the knowledge base.

Knowledge engineers are trained in techniques to extract knowledge from experts, in the same way that systems analysts and other specialists are trained to obtain user requirements.

Think of a situation where you have either had to provide knowledge to someone or even had to obtain knowledge from a third party—this will help you answer the following question:

What tools or techniques are available to assist the knowledge engineer in carrying out these activities?

Feedback 7
Advising and obtaining knowledge from the expert can be supporting by some formal elicitation techniques, or use of interviews, questionnaires and similar fact-finding methods.

In addition to standard techniques, software including text editors and specialised knowledge representation languages such as KARL (Fensel, 1996) can assist in the encoding of knowledge for inclusion in a knowledge base.

Specialised programs such as TEIRESIAS (Davis, 1993), help to validate knowledge and check for errors within a knowledge base.

Professionalism, Methods and Standards

Apart from the skills required to place knowledge into a KBS, a knowledge engineer will also normally be expected to:

- be bound by a professional code of conduct
- update their knowledge and skills on a regular basis
- adhere to appropriate rules, regulations and legal requirements.

The following managerial and interpersonal skills are also expected from knowledge engineers. The most important skills are identified at the top of the list.

- Knowledge representation
- Fact finding
- Human skills
- Visualisation skills
- Analysis
- Creativity
- Managerial.