

# Chapter 1

## How to Read the Book “Foundations of Biomedical Knowledge Representation”

Peter J.F. Lucas and Arjen Hommersom

### 1.1 On the Nature of Things

Biology and medicine are very rich knowledge domains in which already at an early stage in their scientific development it was realised that without a proper way to organise this knowledge they would inevitably turn into chaos. Early examples of organisation attempts are for example “*De Rerum Natura* (On the Nature of Things)” by Titus Lucretius Carus (99–55 BC), which explains the natural and physical world as known at the time, and of course the work “*Systema Naturae*” by Carl Linnaeus published in 1735. The latter book can be seen as the clear recognition of the need of using *systematic methods*, here principles of taxonomic organisation, to classify nature. As soon as one considers using systematic methods, computer-based representations and algorithms come to mind.

Today, the size and complexity of medical-biological knowledge has risen to such a dazzling height that one cannot even imagine *not* to use computer-based methods. However, so far this has been especially the case for the representation and storage of basic biological knowledge—not the main focus of the present book—rather than for medical and clinical knowledge. The amount of detailed biological knowledge available nowadays is so large that even people specialised in particular biological areas would not be able to remember this specialised part in toto. Thus, the application of formalisms such as description logics to represent knowledge about genetic mechanisms and the proteins involved was in the end unavoidable. Access to these knowledge bases, such as KEGG<sup>1</sup> (Kyoto Encyclopedia of Genes and Genomes), GO<sup>2</sup> (Gene Ontology), and the RCSB PDB<sup>3</sup> (Protein Databank) is essential for the present-day working biologist and biochemist to make scientific progress. These knowledge bases are standard tools and part of the computational environment used in these research areas.

---

<sup>1</sup>[www.genome.jp/kegg](http://www.genome.jp/kegg).

<sup>2</sup>[geneontology.org](http://geneontology.org).

<sup>3</sup>[www.rcsb.org/pdb](http://www.rcsb.org/pdb).

The characteristics of the medical area, however, are different from those of research in biology, even though the former area is firmly grounded on biological knowledge. First, knowledge is not mainly used as part of research but primarily for the management of disease, such as the establishment of the diagnosis, treatment, and prognosis in patients. Second, medical doctors are trained to memorise quite a lot of the knowledge involved in decision making, and this knowledge is often simplified to make the memorisation feasible. As a consequence, the need for computer-based methods is not felt as strongly as in biology, where there is not such a clear rationale for simplifying knowledge.

Simplification of knowledge with the aim of keeping the complexity of the decision-making process manageable to humans has a long tradition in medicine. This is, for example, reflected by the frequent use of acronyms, even for procedures (e.g. CABG, pronounced as ‘cabbage’, i.e. the Coronary Artery Bypass Graft procedure). Yet, with the substantial progress made in biomedical, i.e. both human-biological and clinical, research there are good reasons to consider the biomedical area afresh and wonder whether there may be better, more scientific ways to manage disease in patients. This in itself is not a new idea, and similar ambitions were expressed before in the 1980s by the medical decision making movement [7]. At the same time, there was great belief in the potential of artificial intelligence in medicine with programs such as MYCIN, INTERNIST-I, and CASNET [2, 4]. Reality in biomedicine appeared to be more resistant to change than thought by many people at the time and not much happened.

However, the current circumstances are not the same as those in the 1980s. There is now a stronger tendency to take errors and mistakes in clinical medicine seriously and researchers are identifying ways to prevent them [3]. One also realises that computer-based methods may contribute to a reduction in the avoidable clinical errors and mistakes. At the same time, developments in computing-science methods and tools have continued, which has made it easier to cross the boundary between informal biomedicine and computer-based formalisation.

## 1.2 Towards Biomedical Knowledge Representation

We have now definitely arrived in the digital age and even healthcare workers have entered this era, mainly because they just followed the rest of society. There is some irony in this part of the evolution in healthcare, since complex information systems have been in use in healthcare at least since the 1970s, and so healthcare was in a perfect position to take the lead in digitisation. Despite several attempts, this never happened at the time, at least not on a global scale, mainly because healthcare workers were not convinced that it would contribute to better and more convenient patient care. Nevertheless, the current situation of almost full digitisation has created new opportunities for using computer-based methods for the representation and reasoning with medical knowledge, and this is where this book is about.

Partly because of the growth in basic biological knowledge and partly because of new clinical insights obtained by clinical and epidemiological research, biomedicine

remains one of the most knowledge-intensive areas. Even though basic biological, in particular genetic, knowledge is an important ingredient in clinical decision making nowadays, which is likely to increase more in the near future because of the trends towards personal medicine, there still is this typical practical tendency of medical doctors to control the complexity of the knowledge using its clinical relevancy as the main guiding principle. This is for example reflected in the increasing importance of clinical guidelines and protocols in medical decision making, because clinical guidelines are the result of a process that results in documents that only include what is clinically relevant. Even in this area it is recognised that the impact of the evidence-based medicine movement that is associated with clinical guidelines will have its limitations, because many medical doctors believe that medicine cannot be practised in a systematic way (they call it “cookbook medicine”).

The modern research in biomedical computing takes these developments into account, and this explains for example the work on computer-based guideline representation and execution [5]. Rather than starting with new ways to formalise medical knowledge, researchers take existing ‘representations’, although informal, as a starting point. As researchers working in the computer-based guideline area can only acknowledge, transforming an informal clinical guideline into an executable representation that integrates well with clinical workflow is already a sufficiently big challenge.

There are similar developments in other areas. For example, in the clinical setting of diagnosis most of the work is now focussing on assisting medical doctors in a particular diagnostic task, for example to help in the interpretation of radiological images. There are still people who pursue the old idea, initially investigated with the development of the INTERNIST-I system, of a diagnostic computer-based system that covers the whole area of medicine, but now they do this using modern methods offered by Bayesian networks<sup>4</sup>.

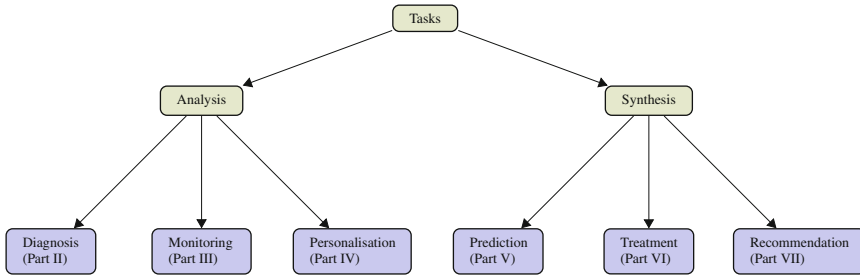
Similar developments in techniques has made it possible to assist and give insight to clinicians in treating diseases in patients and in making predictions of the outcome of treatment. Finally, the book also covers modern developments in representation techniques for personal medicine, recommender systems and monitoring of disease, where time is an important aspect of the representation formalism. Disease monitoring is an important topic in the context of eHealth [6].

### 1.3 Organisation of the Book

Knowledge representation methods [1] have been used for many different types of knowledge-intensive tasks in biomedicine. The concept of ‘task’ has been used as a way to capture the generic aspects of particular procedures, such as diagnostic problem solving. The same ideas can also be applied to other domains; in that sense only part of the tasks described in this book are domain specific. Nevertheless, in

---

<sup>4</sup>[www.symptomate.com](http://www.symptomate.com).



**Fig. 1.1** Overview of the book.

this book we distinguish particular tasks and give them the names that they have in medicine even though there may be similar tasks in other domains. The fact that we deal with the domain of biomedicine often has implications for the way we represent the domain, for example, often we use causality as a way to structure the domain knowledge; *abduction*, i.e. explaining observations in terms of active causes is then one possible *method* to implement the task of diagnosis. Many of these issues will hopefully have become clear after reading the introductory chapter (Chap. 2) of the book.

We have made a distinction between tasks that can be seen as a form of *analysis*, and other tasks that put more emphasis on *synthesis*, as summarised in Fig. 1.1. The book consists of 7 parts, where the first part introduces the book and the various techniques used in the book; the other 6 parts are concerned with individual tasks. Each part is started with an introductory chapter that is followed by one or more specialised chapters. The following parts are distinguished:

**Part I Introduction.** A general description of knowledge representation methods that are relevant for the different chapters included in this book.

**Part II Diagnosis of Disease** includes a general overview of diagnostic methods and a chapter that describes the use of these methods for medical image interpretation.

**Part III Monitoring of Health and Disease and Conformance** concerns a description of general characteristics of the monitoring task and applications in the context of clinical guidelines and the individual patient.

**Part IV Assessment of Health and Personalisation** puts a focus on the use of graphical knowledge-representation formalisms, such as Bayesian networks and chain graphs, to capture the features of disease at the level of the individual. Genetic information and the modelling of the relationship between genetic information and disease is key here.

**Part V Prediction and Prognosis of Health and Disease** is concerned with statements of what is going to happen in the future, and uncertainty is something one has to take into account. This explains the use of probabilistic methods in this context.

**Part VI Treatment of Disease.** Treatment is concerned with following a sequence of actions, taking into account the uncertainty in the diagnosis and the uncertainty

in the expected outcome of the treatment. Both general principles of knowledge representation of the treatment task and actual applications are described in three chapters.

**Part VII Recommendation.** This last part of the book deals with supporting physicians through recommendations based on the best available evidence. Knowledge representation techniques have been used to develop computer-interpretable clinical guidelines that can be used for various reasoning tasks. Furthermore, this part also presents an alternative to guidelines by aggregation of clinical evidence through argumentation theory.

With the early work as briefly summarised at the beginning of this chapter in mind it becomes clear that modern knowledge representation and reasoning methods cover a much broader area of biomedicine than the earlier methods, which often only dealt with a specific clinical diagnostic problem. The modern methods now also have a sound mathematical foundation in terms of logic, probability theory and decision theory. This explains the title of the present book “Foundations of Biomedical Knowledge Representation”. As the applications described in the book already make clear, we are now finally on the edge that principles of knowledge representation are creating impact in the biomedical field.

## References

1. van Harmelen, F., Lifschitz, V., Porter, B. (eds.): Handbook of Knowledge Representation. Elsevier, Amsterdam (2008)
2. Lucas, P.J.F., van der Gaag, L.C.: Principles of Expert Systems. Addison-Wesley, Wokingham (1991)
3. Rylander, M., Guerrasio, J.: Heuristic errors in clinical reasoning. In: Clinical Teaching (2015). doi:[10.1111/tct.12444](https://doi.org/10.1111/tct.12444)
4. Szolovits, P.: Artificial intelligence in medicine. In: AAAS Selected Symposium, vol. 51. Westview Press, Boulder, Colorado (1982)
5. ten Teije, A., Miksch, S., Lucas, P.J.F. (eds.): Computer-Based Clinical Guidelines and Protocols: a Primer and Current Trends. IOS Press, Amsterdam (2008)
6. Velikova, M., Lucas, P.J.F., van der Heijden, M.: Intelligent disease self-management with mobile technology. *IEEE Comput.* **48**(2), 32–39 (2015)
7. Weinstein, M.C., Fineberg, H.: Clinical Decision Analysis. Saunders, Philadelphia (1980)