

# Preface

Grammatical inference, the main topic of this book, is a scientific area that lies at the intersection of multiple fields. Researchers from computational linguistics, pattern recognition, machine learning, computational biology, formal learning theory, and many others have their own contribution. Therefore, it is not surprising that the topic has also a few other names such as grammar learning, automata inference, grammar identification, or grammar induction. To simplify the location of present contribution, we can divide all books relevant to grammatical inference into three groups: theoretical, practical, and applicable. In greater part this book is practical, though one can also find the elements of learning theory, combinatorics on words, the theory of automata and formal languages, plus some reference to real-life problems.

The purpose of this book is to present old and modern methods of grammatical inference from the perspective of practitioners. To this end, the Python programming language has been chosen as the way of presenting all the methods. Included listings can be directly used by the paste-and-copy manner to other programs, thus students, academic researchers, and programmers should find this book as the valuable source of ready recipes and as an inspiration for their further development.

A few issues should be mentioned regarding this book: an inspiration to write it, a key for the selection of described methods, arguments for selecting Python as an implementation language, typographical notions, and where the reader can send any critical remarks about the content of the book (subject-matter, listings etc.).

There is a treasured book entitled “Numerical recipes in C”, in which along with the description of selected numerical methods, listings in C language are provided. The reader can copy and paste the fragments of the electronic version of the book in order to produce executable programs. Such an approach is very useful. We can find an idea that lies behind a method and immediately put it into practice. It is a guiding principle that accompanied writing the present book.

For the selection of methods, we try to keep balance between importance and complexity. It means that we introduced concepts and algorithms which are essential to the GI practice and theory, but omitted that are too complicated or too

long to present them as a ready-to-use code. Thanks to that, the longest program included in the book is no more than a few pages long.

As far as the implementation language is concerned, the following requirements had to be taken into account: simplicity, availability, the property of being firmly established, and allowing the use of wide range of libraries. Python and FSharp programming languages were good candidates. We decided to choose IronPython (an implementation of Python) mainly due to its integration with the optimization modeling language. We use a monospaced (fixed-pitch) font for the listings of programs, while the main text is written using a proportional font. In listings, Python keywords are in bold.

The following persons have helped the author in preparing the final version of this book by giving valuable advice. I would like to thank (in alphabetical order): Prof. Z.J. Czech (Silesian University of Technology), Dr. P. Juszczuk, Ph.D. student A. Nowakowski, Dr. R. Skinderowicz, and Ph.D. student L. Strak (University of Silesia).

Sosnowiec, Poland  
2016

Wojciech Wieczorek

Grammatical Inference  
Algorithms, Routines and Applications  
Wieczorek, W.  
2017, XI, 145 p. 18 illus., Hardcover  
ISBN: 978-3-319-46800-6