

1. Record Nr.	UNINA9910337952203321
Autore	Couto Francisco M
Titolo	Data and Text Processing for Health and Life Sciences [[electronic resource] /] / by Francisco M. Couto
Pubbl/distr/stampa	Cham, : Springer Nature, 2019 Cham : , : Springer International Publishing : , : Imprint : Springer, , 2019
ISBN	3-030-13845-3
Edizione	[1st ed. 2019.]
Descrizione fisica	1 online resource (XV, 98 p. 483 illus., 74 illus. in color.)
Collana	Advances in Experimental Medicine and Biology, , 0065-2598 ; ; 1137
Disciplina	614.5999
Soggetti	Cancer research Human genetics Database management Data mining Cancer Research Human Genetics Database Management Data Mining and Knowledge Discovery
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Preface -- Introduction -- Resources -- Data Retrieval -- Text Processing -- Semantic processing -- Index.
Sommario/riassunto	This open access book is a step-by-step introduction on how shell scripting can help solve many of the data processing tasks that Health and Life specialists face everyday with minimal software dependencies. The examples presented in the book show how simple command line tools can be used and combined to retrieve data and text from web resources, to filter and mine literature, and to explore the semantics encoded in biomedical ontologies. To store data this book relies on open standard text file formats, such as TSV, CSV, XML, and OWL, that can be open by any text editor or spreadsheet application. The first two chapters, Introduction and Resources, provide a brief introduction to the shell scripting and describe popular data resources in Health and Life Sciences. The third chapter, Data Retrieval, starts by introducing a

common data processing task that involves multiple data resources. Then, this chapter explains how to automate each step of that task by introducing the required commands line tools one by one. The fourth chapter, Text Processing, shows how to filter and analyze text by using simple string matching techniques and regular expressions. The last chapter, Semantic Processing, shows how XPath queries and shell scripting is able to process complex data, such as the graphs used to specify ontologies. Besides being almost immutable for more than four decades and being available in most of our personal computers, shell scripting is relatively easy to learn by Health and Life specialists as a sequence of independent commands. Comprehending them is like conducting a new laboratory protocol by testing and understanding its procedural steps and variables, and combining their intermediate results. Thus, this book is particularly relevant to Health and Life specialists or students that want to easily learn how to process data and text, and which in return may facilitate and inspire them to acquire deeper bioinformatics skills in the future.
