言語研究のための プログラミング入門

Pythonを活用したテキスト処理

浅尾仁彦・李 在鎬

開拓社

言語研究のための プログラミング入門

Pythonを活用したテキスト処理

淺尾仁彦・李 在鎬

開拓社

Python

はじめに

言語研究にとって，コンピュータの重要性はますます増しています．本書を手に取られる方も，文書作成やメール，ウェブ閲覧などで日常的にコンピュータを利用しているのではないかと思います．また，大規模コーパスの整備やインターネットの発達によって，これまで直感に頼るしかなかったような事柄でもデータに基づいて統計的に調べられるようになってきました．

しかしながら，コンピュータを前にして，単調なコピー・アンド・ペーストの作業を繰り返すなど，人間の側が機械的な作業をしていることがないでしょうか．コンピュータは，人間の代わりに機械的な作業をするためにあります．

時にはソフトウェアが初めから用意されていて，メニューから機能を選ぶだけでコンピュータに機械的な作業をさせることができる場合もあります．しかし，研究の目的は常にユニークなものです．ソフトウェアは，多くの利用者が必要とするであろうという最大公約数的な機能しか用意してくれませんから，既存のソフトウェアでうまく作業を自動化できるとは限りません．このような時に新しい「機能を作る」作業がプログラミングです．

しかしながら，言語研究者がプログラミングを学ぼうとして書店で本を手に取ってみても，ちょうど良い本を見つけるのはなかなか難しいのが実情です．プログラミングのすぐれた入門書は多数出版されていますが，理工系の読者を想定していたり，ウェブサービスの構築など特定の目的を念頭においていたりで，多くの場合，内容が言語研究者の背景知識や関心に合っていないためです．

本書は，言語学研究者に向け，言語研究のための処理を解説しています．コンピュータの操作に慣れていない文系の大学院生や研究者でも実践できるように，丁寧な解説を心がけています．全体の流れにおいても，ごく簡単な内容から出発して，一歩一歩複雑なプログラムが書けるようになるよう，

工夫しています．また同時に，テキスト処理での実用例を多く盛り込み，各機能がどのように役立つのか伝わるように心がけました．その代わりに，一般の入門書であれば書いてある事柄であっても，本書の目的に直接関係しない点については大胆に省略しており，機能をくまなく紹介するということは行っていません．大事なことは，機能の１つ１つを覚えることより，どうやってコンピュータに作業の流れを教え込むのかについての考え方に親しむことだと思うからです．実際に手を動かしながら読み進めていただければ幸いです．本書を通じてプログラミング処理が持つ威力を実感していただき，ご自分の研究にもぜひ応用していただきたいと思います．

本書の執筆のきっかけの１つになったPython勉強会に参加していただいた京都大学文学研究科，人間・環境学研究科の皆様に感謝いたします．また，本書の草稿に目を通して多くの有益なコメントを下さった小松原哲太氏，遠藤智子氏，中川奈津子氏，小木曽智信氏に感謝いたします．そして，Python自体をはじめとして，本書で紹介したソフトウェアや言語資源を公開していただいている開発者やボランティアの方々に感謝いたします．最後に，企画および編集を通じてさまざまな便宜をはかってくださった開拓社の川田賢氏にお礼を申し上げます．

2013年5月

浅尾仁彦・李在鎬

目　次