

## *o książce*

---

*Przetwarzanie języka naturalnego w akcji* to praktyczny przewodnik po przetwarzaniu i generowaniu tekstu w języku naturalnym w realnym świecie. W książce tej dajemy wam wszystkie narzędzia i techniki potrzebne do zbudowania backendowych systemów NLP do obsługi wirtualnego asystenta (chatbota/agenta dialogowego), filtra spamu, moderatora forum, systemu do analizy wydźwięku, generatora baz wiedzy, analizatora tekstów języka naturalnego i każdego innego zastosowania NLP, jakie możecie sobie wyobrazić.

Książka ta jest skierowana do deweloperów Pythona, od średnio zaawansowanych do zaawansowanych. Dla czytelników, którzy już umieją projektować i tworzyć złożone systemy, książka także będzie przydatna, gdyż daje wiele przykładów dobrych praktyk oraz wgląd w możliwości najnowszych algorytmów NLP. Wiedza na temat programowania obiektowego w Pythonie będzie przydatna do budowy lepszych systemów, ale nie jest wymagana przy nauce z tej książki.

W przypadku specjalnych tematów podajemy odpowiednią ilość materiału pomocniczego i cytujemy zasoby (tekstowe i online) dla tych, którzy chcą szczegółowo się z nimi zapoznać.

### **Plan działania**

Osoby nieznające Pythona i przetwarzania języka naturalnego powinny najpierw przeczytać część 1, a potem dowolne rozdziały z części 3, które związane są z ich zainteresowaniami lub wyzwaniem w pracy. Jeśli chcecie przyspieszyć zdobywanie nowych możliwości NLP, na które pozwala deep learning, będziecie chcieli przeczytać po kolei część 2. Buduje ona zrozumienie sieci neuronowych, przyrostowo zwiększając złożoność i możliwości sieci neuronowych.

Gdy znajdziecie rozdział lub część z fragmentem programu, który możecie „uruchomić w głowie”, powinniście wykonać go na swoim komputerze. A jeśli jakieś przykłady

wyglądają, jakby można je było zastosować w swoich dokumentach, powinniście wstawić ten tekst do CSV lub do pliku tekstowego (jeden dokument w wierszu) w katalogu `nlpia/src/nlpia/data/`. Wtedy możecie użyć funkcji `nlpia.data.loaders.get_data()` do uzyskania tych danych i wykonania przykładów na swoich danych.

## O książce

Rozdziały z części 1 dotyczą logistyki pracy z językiem naturalnym i przekształcania go w liczby, które można wyszukiwać i obliczać. To „tworzenie bloków i chwywanie” słów daje nagrodę w postaci zadziwiająco użytecznych aplikacji, jak wyszukiwanie informacji i analiza wydźwięku. Gdy już dobrze opanujecie podstawy, okaże się, że już przy użyciu prostej arytmetyki wykonywanej w pętli można rozwiązać niektóre dość skomplikowane problemy, jak filtrowanie spamu. Filtry spamu, które zbudujecie w rozdziałach od 2 do 4, są tym, co ocaliło globalny system e-mailowy przed anarchią i stagnacją. Dowiedziecie się, jak tworzyć filtry spamu o dokładności ponad 90% za pomocą technologii z ery lat dziewięćdziesiątych – obliczając jedynie liczbę słów i proste średnie tych liczb.

Cała ta matematyka dotycząca słów może wydawać się uciążliwa, ale w istocie to dość dobra zabawa. Bardzo szybko będziecie budować algorytmy, które będą mogły podejmować decyzje dotyczące języka naturalnego tak samo dobrze lub lepiej niż my sami (a na pewno dużo szybciej). Może to być pierwszy raz w waszym życiu, gdy będziecie mieli perspektywę, aby ocenić w pełni sposób, w jaki słowa odzwierciedlają i wzmacniają wasze myślenie. Spojrzenie na słowa i myśli przez perspektywę wielowymiarowej przestrzeni wektorowej pozostawi nasz mózg w ruchu w rekurencyjnych pętlach samo-odkrywania.

Tempo nauki może osiągnąć najwyższy punkt w połowie książki. Rdzeniem książki w części 2 jest poznawanie skomplikowanej sieci obliczeń i komunikacji z sieciami neuronowymi. Sieciowy efekt małych jednostek logicznych w interakcji z siecią „myślenia” dał maszynom moc rozwiązywania problemów, którymi w przeszłości zajmowali się tylko ludzie, obejmując takie elementy jak pytania o analogie, streszczanie tekstu i tłumaczenia między językami naturalnymi.

W miarę nauki poznacie wektory słów, ale bez obaw, o wiele więcej. Będziecie w stanie dokonać wizualizacji słów, dokumentów i zdań w chmurze powiązanych pojęć, rozciągającej się znacznie poza trzy wymiary, które jesteśmy gotowi ogarnąć. Zaczniecie myśleć o dokumentach i słowach jak o postaciach w grze *Dungeons and Dragons*, z miriadą losowo wybranych charakterystyk i zdolności, które wyewoluowały i urosły z biegiem czasu, lecz tylko w naszych głowach.

Docenienie tego wielotematycznego świata słów i ich znaczenia stanie się podstawą do *coup-de-grace* z części 3, gdzie nauczycie się, jak budować maszyny, które prowadzą konwersację i odpowiadają na pytania tak samo dobrze jak ludzie.

## O kodzie

Książka zawiera wiele przykładów kodu źródłowego zarówno w ponumerowanych listingach, jak i w wierszach normalnego tekstu. W obu przypadkach kod jest zapisany czcionką o stałej szerokości (jak ta) i oddzielony od reszty tekstu. Czasami kod jest także

**pogrubiony**, aby podkreślić zmiany w stosunku do poprzednich kroków w rozdziale, jak w przypadku dodania nowej funkcji w istniejącym wierszu kodu.

W wielu przypadkach oryginalny kod źródłowy został przeformatowany. Przełamaliśmy wiersze i dodaliśmy wcięcia, aby dostosować do dostępnej szerokości kolumny. W rzadkich przypadkach nawet to nie wystarczyło i listingi zawierają znaczki kontynuacji kodu (➡). Ponadto w przypadkach, gdy kod jest opisany w tekście, komentarze z kodu źródłowego zostały usunięte z listingów. Wielu listingom towarzyszą adnotacje kodu, aby wypuklić ważne pojęcia.

Kod źródłowy wszystkich listingów z tej książki jest dostępny do pobrania z witryny wydawnictwa Manning pod adresem <https://www.manning.com/books/natural-language-processing-in-action> oraz na GitHub pod adresem <https://github.com/totalgood/nlpia>.

## Forum dyskusyjne liveBook

Zakupienie tej książki obejmuje bezpłatny dostęp do prywatnego forum sieciowego prowadzonego przez Manning Publications, gdzie możecie komentować książkę, zadawać pytania techniczne i otrzymać pomoc od autorów i innych użytkowników. Aby mieć dostęp do forum, trzeba wejść na stronę <https://livebook.manning.com/#!/book/natural-languageprocessing-in-action/discussion>. Więcej o forum Manninga i obowiązujących tam zasadach można się dowiedzieć na stronie <https://livebook.manning.com/#!/discussion>.

Manning zobowiązuje się zapewnić czytelnikom forum, na którym może mieć miejsce sensowny dialog między czytelnikami oraz między czytelnikami a autorami. Nie jest to zobowiązanie do konkretnego stopnia zaangażowania ze strony autorów, których udział w forum jest dobrowolny (i bezpłatny). Sugerujemy, aby zadawać autorom trudne pytania, co przyciągnie ich uwagę! Forum i archiwa zawierające wcześniejsze dyskusje są dostępne na witrynie wydawcy od chwili wydrukowania tej książki.