

1. Record Nr.	UNINA9910337563903321
Autore	Sidorov Grigori
Titolo	Syntactic n-grams in Computational Linguistics // by Grigori Sidorov
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2019
ISBN	3-030-14771-1
Edizione	[1st ed. 2019.]
Descrizione fisica	1 online resource (94 pages)
Collana	SpringerBriefs in Computer Science, , 2191-5768
Disciplina	410.285
Soggetti	Natural language processing (Computer science) Computational linguistics Natural Language Processing (NLP) Computational Linguistics
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Preface -- Introduction -- PART I. VECTOR SPACE MODEL IN THE ANALYSIS OF SIMILARITY BETWEEN TEXTS -- Chapter 1. Formalization in computational linguistics -- Chapter 2. Vector space model -- Chapter 3. Vector space model for texts and the tf-idf measure -- Chapter 4. Latent Semantic Analysis (LSA): reduction of dimensions -- Chapter 5. Design of experiments in computational linguistics -- Chapter 6. Example of application of n-grams: authorship attribution using n-grams of syllables -- PART II. NON-LINEAR CONSTRUCTION OF N-GRAMS -- Chapter 7. Syntactic n-grams: the concept -- Chapter 8. Types of syntactic n-grams according to their components -- Chapter 9. Continuous and non-continuous syntactic n-grams -- Chapter 10. Metalanguage of syntactic n-grams representation -- Chapter 11. Examples of construction of non-continuous syntactic n-grams -- Chapter 12. Automatic analysis of authorship using syntactic n-grams -- Chapter 13. Filtered n-grams -- Chapter 14. Generalized n-grams.
Sommario/riassunto	This book is about a new approach in the field of computational linguistics related to the idea of constructing n-grams in non-linear manner, while the traditional approach consists in using the data from the surface structure of texts, i.e., the linear structure. In this book, we propose and systematize the concept of syntactic n-grams, which

allows using syntactic information within the automatic text processing methods related to classification or clustering. It is a very interesting example of application of linguistic information in the automatic (computational) methods. Roughly speaking, the suggestion is to follow syntactic trees and construct n-grams based on paths in these trees. There are several types of non-linear n-grams; future work should determine, which types of n-grams are more useful in which natural language processing (NLP) tasks. This book is intended for specialists in the field of computational linguistics. However, we made an effort to explain in a clear manner how to use n-grams; we provide a large number of examples, and therefore we believe that the book is also useful for graduate students who already have some previous background in the field.

---