

Chapter 6

Language Processing and Python

Belsini Glad Shiya V.

Agurchand Manmull Jain College, India

Sharmila K.

VISTAS, India

ABSTRACT

Natural language processing is the communication between the humans and the computers. It is the field of computer science which incorporates artificial intelligence and linguistics where machine learning algorithms are used to analyze and process the enormous variety of data. This chapter delivers the fundamental concepts of language processing in Python such as text and word operations. It also gives the details about the preference of Python language for language processing and its advantages. It specifies the basic concept of variables, list, operators, looping statements in Python and explains how it can be implemented in language processing. It also specifies how a structured program can be written using Python, categorizing and tagging of words, how an information can be extracted from a text, syntactic and semantic analysis, and NLP applications. It also concentrates some of the research applications where NLP is applied and the challenges of NLP processing in the real-time area of applications.

INTRODUCTION

Nowadays data gathering and analysis of data are essential in every field of business for understanding the needs and passions of the user so that the organizations and companies can satisfy their customer's essentials and expectations. For data analytics different fields of computer Science, technologies, statistics, algorithms, analytical tools are used according to the types of data to be processed and depending on the field that data belongs. Natural Language Processing is one of the data analytical fields of computer science which comprises linguistics and artificial intelligence which mainly concentrates the communications between computers and humans (Vismaya & Reynald, 2017). It processes and analyze huge natural language data which inference the computers to understand the documents and gives in forma-

DOI: 10.4018/978-1-7998-7728-8.ch006

tions about that documents. NLP supports computers to interact with humans in their own language and process other language related methods to help computers to read text data, hear voice data and interpret it. NLP comprise utilization of algorithms to recognize and draw out the rules of natural language which is in the form of unstructured data format in turn processed and changed into structured data in which the computer can easily understand.

NLP process functions based on rules which can take more time and effort of people. Some other performs with plenty of data using Statistical methods and use machine learning algorithms to obtain the inferences according to the data. The set of data is trained by the machine learning algorithms. With the help of the trained data model the data can be tested to predict the outcome or the result(Thanaki 2017)

NLP plays an important role in research. The research includes speech recognition, text classification, machine translation, question answering. The researches work on Natural Language Processing (NLP) collect the data related to the behavior of human being to process the language by understanding and hence to use the suitable tools and techniques to put together the computers to recognize and process the natural language(Vismaya, Reynald,2017). They converts the linguistic knowledge of data into a rule based implementation by means of Machine learning and Deep learning algorithms for simple manipulation and distribution of data in language processing.

In certain cases the machine learning program in python will be implemented by three step processing to look with the key words relating the event. The first step the NLP cleans the data as the initial stage. In the second step another form of cleaning such as vectorization or tokenization will be performed where the text is converted into tokens and in the third step the data is split to train and test for further processing.(Szlosek, Ferrett, 2016)

This chapter delivers the fundamental concepts of Language processing in Python such as text and word operations. It also gives the details about the preference of Python language for Language Processing and its advantages. It specifies the basic concept of variables, list, operators, looping statements in Python and explains how it can be implemented in language processing. It outlines the knowledge attainment of basic applications of NLP in Python, preparing the data set for NLP applications, Context free grammar, Stepping into NLTK, Raw text processing . It also specify how a structured program can be written using python, Categorizing and tagging of words, how an information can be extracted from a text, Syntactic and Semantic analysis and NLP applications. It also concentrates some of the research applications where NLP is applied and the challenges of NLP processing in the real time area of applications.

NATURAL LANGUAGE PROCESSING WITH PYTHON

Python is an interactive, interpreted, object oriented programming language which is easy to learn and develop applications and algorithms to analyze and process huge natural language data. (Menzenski, 2015). It was developed by Guido Van Rossum in the year 1991. Python allows easy encapsulation of methods and data and hence the code can be repeatedly used. The variables can be used as dynamic in our program to improve fast development of program codes. It also contains standard library which contains a collection of build-in modules for performing specific functions which includes a core part of the language. It also contains components for Graphical Programming, Web Connectivity and processing numerical data. The Syntax and Semantics are well defined in Python which gives special meaning to the language. Python is a open source software and hence it is easily downloaded according to the operating systems used by the user. It furnish the programmers with various NLP tools and libraries

25 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/language-processing-and-python/284205

Related Content

Recent Trends in Deepfake Detection

Kerenalli Sudarshanaand Mylarareddy C. (2021). *Deep Natural Language Processing and AI Applications for Industry 5.0* (pp. 1-28).

www.irma-international.org/chapter/recent-trends-in-deepfake-detection/284200

Author Profiling Using Texts in Social Networks

Iqra Ameerand Grigori Sidorov (2021). *Handbook of Research on Natural Language Processing and Smart Service Systems* (pp. 245-265).

www.irma-international.org/chapter/author-profiling-using-texts-in-social-networks/263105

Research Journey of Hate Content Detection From Cyberspace

Sayani Ghosaland Amita Jain (2021). *Natural Language Processing for Global and Local Business* (pp. 200-225).

www.irma-international.org/chapter/research-journey-of-hate-content-detection-from-cyberspace/259790

MorseEx: A Communication Application for the Deaf-Blind

Suresh Kumar Nagarajan, Geetha N., Raghav Talwarand Shivoma Ahuja (2023). *Deep Learning Research Applications for Natural Language Processing* (pp. 218-228).

www.irma-international.org/chapter/morseex/314146

Deep Learning for Sentiment Analysis: An Overview and Perspectives

Vincent Karasand Björn W. Schuller (2021). *Natural Language Processing for Global and Local Business* (pp. 97-132).

www.irma-international.org/chapter/deep-learning-for-sentiment-analysis/259785