# Comparison of Collection Data Types in Java & Python

Jamie Kemman

*Department of Computer Sciences, Mathematics, and Engineering*
*Shepherd University*
*P.O. Box 3210*
*Shepherdstown, WV 25443*
*jkemma01@rams.shepherd.edu*

## 1. INTRODUCTION

In the realm of programming and computer science, a collection is a data structure, or object, which exists to hold zero or greater other objects [1]. Python implements four basic variants of a collection under the names List, Tuple, Set, and Dictionary, where each data type contains different properties and access methods. Though Java is structurally and syntactically dissimilar to Python, it also implements collection type data structures which are not unlike the most common examples found in Python.

## 2. LISTS

A python list is an ordered, indexed data structure which can contain multiple items of differing data types [2]. Due to the indexed linear order of the list, multiple native methods exist for accessing entries. As with any mutable collection, items can be inserted to the end of the list, or retrieved from a specific index. Indexing, of course, also allows for removal of specific items in the list. In Java, Arrays mirror the basic functionality of a Python list, but LinkedList objects are most like to the Python List structure. LinkedLists are an abstracted implementation of the default Java Array functionality in which each index of an LinkedList is linked, sequentially, to its neighbors by memeory references which do not require data to be stored in contiguous blocks. LinkedLists are the most direct comparison to a Python list because Java Arrays alone cannot contain data of multiple

types, and they must be a fixed length which is declared when the Array is created. By relying on references to link list elements, LinkedList does not have a fixed length requirement, and each index may contain any data type. In addition, the dynamic nature of LinkedLists allows programmers to insert and remove data from any index of the list with great efficiency [3].

## 3. TUPLES

In Python, a Tuple is an ordered, indexed collection with stucture and access methods nearly equivalent to a List. The primary exception, however, is that Tuples are immutable [2]. Tuples can be cloned or concatenated into new tuples, but once declared a tuple cannot gain or lose elements. A review of Java data structures shows Java has no native analogue of a Python tuple [3]. Java's native collections are mutable after declaration and do not emulate tuple functionality without additional work. Basic functionality to mimic a Python Tuple could be achieved by declaring a Java Array and assigning its element values to constant variables declared outside of the structure. Additionally, programmers could write a class to emulate tuple functionality on a deeper level by cloning a suitable native data structure and removing any setter methods, essentially forcing declaration as the only viable option to assign values to the Tuple.

## 4. SETS

A Set in Python is a mutable, unordered collecion similar to a list, but without indexing, so duplicate values are not possible [2]. Access to set elements occurrs through Hashing, or the practice of mapping data elements to storage addresses instead of containing all data within the structure itself. Hashing provides for data retrieval in constant time, or O(1), on average, given that the target variable name, or key in some instances, is known [2]. In Java, an equivalent structure to a Python Set exists through the collection HashSet [3].

## 5. DICTIONARIES

Dictionaries, like sets, are unordered, mutable hash tables with a constant time curve [2]. Unlike sets, Dictionaries are ordered and contain key-value pairs for every entry instead of single elements—this allows, for example, complicated data entries to be retrieved directly via relaively simple keys. Like to other mutable collection structures in Python, elements can be added and removed from a dictionary at will. However, elements must be initalized as a pair with a unique key corresponding to a unique value. Similar to the Set-HashSet correspondence, Java also implements dictionaries through a type called a HashTable, where unique key value pairs are created and stored in a structure functionally equivalent to a Python Dictionary.

## REFERENCES


[1] Lambert, Kenneth A., *Fundamentals of Python Data Structures*, 2nd ed., Cengage, 2016.

[2] "Python Data Structures," geeksforgeeks.org, 8/16/2024, https://www.geeksforgeeks.org/python-data-structures/ (accessed 9/18/2024).

[3] Parvez, Faizan, "Data Structures in Java – A Beginners Guide 2024," Great Learning, 9/10/2024, https://www.mygreatlearning.com/blog/data-structures-using-java/ (accessed 9/18/2024).