

УДК 004.432

**КОМПАРАТИВНЫЙ АНАЛИЗ ПОПУЛЯРНЫХ ЯЗЫКОВ
ПРОГРАММИРОВАНИЯ****Матушкова Анастасия Алексеевна,**

Магистрант 2 курса,

кафедры прикладной информатики и математических методов в экономике (ПИММЭ)

Института экономики и управления (ИУЭ) Волгоградского государственного университета
(ВолГУ),

Россия, Волгоград,

E-mail: piem-201_625542@volsu.ru

Седова Алена Игоревна

Магистрант 2 курса,

кафедры ПИММЭ ИЭУ ВолГУ,

Россия, Волгоград,

E-mail: piem-201_524818@volsu.ru

Аннотация

В статье рассматриваются языки программирования Python, Java и C++. Приведена сравнительная таблица для этих языков. Был сделан вывод, какой из языков является наиболее подходящим для предобработки и анализа данных.

Ключевые слова: язык программирования, компаративный анализ, Python, Java, C++**COMPARATIVE ANALYSIS OF POPULAR PROGRAMMING LANGUAGES****Anastasia A. Matushkova**

2nd year master's student,

IEM VolSU,

Russia, Volgograd,

E-mail: piem-201_625542@volsu.ru

Alena I. Sedova

2nd year master's student,

IEM VolSU,

Russia, Volgograd,

E-mail: piem-201_524818@volsu.ru

ABSTRACT

In the article considered Python, Java and C++ programming languages. The article also contains a comparative table these languages. It concluded which of these languages is more suitable for data preprocessing and analysis.

Keywords: programming language, comparative analysis, Python, Java, C++

Язык программирования – это искусственный язык, созданный для обеспечения общения между человеком и компьютером. В мире существует более двух тысяч языков программирования, что не упрощает выбор.

Рассмотрим Python, Java и C++ поподробнее, чтобы понять, какой из них является наиболее подходящим для эффективного анализа больших массивов данных.

Согласно рейтингам TIOBE[6] и IEEE Spectrum[3] за 2021 год наиболее популярными были Python, Java, C и C++.

Согласно руководству «Python 3. Самое необходимое»: «Python – это интерпретируемый, объектно-ориентированный, тьюринг-полный язык программирования высокого уровня, предназначенный для решения самого широкого круга задач» [9, с. 2]. Python был разработан в начале 1990-х годов Гвидо ван Россумом [5].

Язык программирования Python используется для разработки web-сайтов и приложений, корпоративных, настольных и научных приложений, а также для программирования контроллеров устройств [3].

Python имеет ряд характеристик, выгодно выделяющих его на фоне других языков программирования:

- динамическая типизация (переменная связывается с типом не в момент ее объявления, а в момент присваивания ей значения; в разных частях программы переменная может принимать значения разных типов);
- автоматическое управление памятью (устраняет необходимость самостоятельно освобождать память, выделенную для объекта, и попытки получить доступ к памяти уже удаленного объекта);
- полная интроспекция (можно получить всю информацию о структуре любого объекта);
- встроенный механизм обработки исключений (производится с помощью операторов try, except, else, finally и raise, которые образуют блок исключений);
- поддержка многопоточных вычислений (возможность выполнять несколько потоков одновременно; реализуется модулями thread и threading);
- высокоуровневые структуры данных (наличие смысловых конструкций, описывающих такие структуры данных и операции, проводимые над ними, для которых описания на машинном коде сложные и громоздкие);
- модульность (программа организуется в виде отдельных блоков – модулей).

Для проведения анализа данных с помощью Python есть несколько свободно распространяемых библиотек:

- NumPy – это библиотека Python, которая предоставляет многомерные массивы, различные производные объекты и набор подпрограмм для быстрых операций с массивами, включая математические, логические манипуляции, сортировку, выборку и т.д.

- Pandas – библиотека, предоставляющая высокопроизводительные, простые в использовании структуры данных и инструменты анализа данных.
- Matplotlib – библиотека для создания статичных, анимированных и интерактивных визуализаций в Python.
- SciPy – библиотека, представляющая алгоритмы для оптимизации, интегрирования, интерполяции, алгебраических уравнений, дифференциальных уравнений, статистики и многих других классов задач.
- Scikit-learn – библиотека, предоставляющая простые и эффективные инструменты для предиктивной аналитики данных, классификации, регрессии, кластеризации, уменьшения размерности, выбора модели, предварительной обработки и т.д.
- Seaborn – библиотека визуализации данных Python, основанная на Matplotlib и тесно интегрированная со структурами данных pandas. Она предоставляет возможность создавать статистические графики.

Недостаток Python состоит в том, что код, написанный на нём, за счёт его интерпретируемости, в общем случае, будет выполняться медленнее, чем эквивалентный ему написанный на компилируемом языке [7, с.24], а также в том, что из-за гибкости типов данных повышается расход памяти.

Java – это строго типизированный, объектно-ориентированный, высокоуровневый, тьюринг-полный язык программирования общего назначения. Он был разработан в 1995 году компанией Sun Microsystems[2, с. 34].

Этот язык программирования также является кроссплатформенным, это свойство было достигнуто за счет создания виртуальной машины Java (Java Virtual Machine или JVM).

Виртуальная машина Java – это компонент технологии, отвечающий за независимость от аппаратного обеспечения и операционной системы, небольшой размер скомпилированного кода и способность защищать пользователей от вредоносных программ [4].

Java поддерживает автоматическое управление хранилищем и имеет встроенный механизм обработки исключений.

Язык программирования Java используется для разработки web-сайтов и приложений, корпоративных, настольных и научных приложений, а также приложений для мобильных устройств [3].

В проведении анализа данных могут помочь следующие инструменты Java:

- Паттерн RegEx (regular expressions, регулярные выражения) – шаблон для поиска строки в тексте.
- Univocity parsers – это набор чрезвычайно быстрых и надежных парсеров на основе Java для файлов CSV, TSV и файлов с фиксированной шириной (fixed-width files).
- JDOM предоставляет полное решение на основе Java для доступа, обработки и вывода XML-данных из кода Java.
- JSON.simple – это простая библиотека Java для обработки JSON (JavaScript Object Notation (нотация объектов JavaScript) – это облегченный, текстовый, независимый от языка формат обмена данными), чтения и записи данных JSON и полного соответствия спецификации JSON (RFC 7159).

- jsoup – это библиотека Java для работы с HTML, которая позволяет очищать и анализировать HTML из URL, файла или строки, находить и извлекать данные, используя обход DOM или селекторы CSS, управлять HTML-элементами, атрибутами и текстом. Эта библиотека предназначена для работы со всеми существующими разновидностями HTML.
- Tablesaw – это библиотека данных и визуализации, которая поддерживает загрузку, очистку, преобразование, фильтрацию и суммирование данных. Она также поддерживает описательную статистику и может использоваться для подготовки данных для работы с библиотеками машинного обучения.
- Smile (Statistical Machine Intelligence and Learning Engine) – это быстрая и комплексная система машинного обучения, НЛП, линейной алгебры, графов, интерполяции и визуализации на Java и Scala.

Язык программирования Java также имеет свои недостатки: из-за использования виртуальной машины страдает производительность, отсутствие инструмента для создания дизайна пользователя, сложность и громоздкость кода

C++ – это универсальный, мультипарадигменный, тьюринг-полный, компилируемый, статически типизированный язык программирования, разработанный в 1983 году Бьёрном Страуструпом.

Язык программирования C++ используется для разработки корпоративных, настольных и научных приложений, программирования контроллеров устройств, а также приложений для мобильных устройств [3].

В C++ также реализован механизм обработки исключений.

Для проведения анализа данных с C++ можно использовать следующие библиотеки:

- Shark – это быстрая, модульная, multifunctionальная C++ библиотека для машинного обучения с открытым исходным кодом. Она предоставляет методы линейной и нелинейной оптимизации, нейронные сети и другие методы машинного обучения. Является мощным инструментом как для приложений, так и для исследований.
- stats++ (Statistical software in C++, статистическое ПО (программное обеспечение) на C++) – это передовое комплексное статистическое ПО. Предоставляет инструменты для сбора и предварительной обработки данных, статистики, машинного обучения, постобработки данных.
- Plotutils – пакет, содержащий программное обеспечение как для программистов, так и для технических пользователей.
- Matplotlib++ – это графическая библиотека для визуализации данных, которая обеспечивает интерактивное построение графиков, предоставляет средства для экспорта графиков в высоком качестве, компактный синтаксис, совместимый с аналогичными библиотеками, а также десятки категорий графиков, несколько стилей кодирования и т.д.

Основными недостатками C++ являются сложность в изучении и отсутствие автоматического управления памятью. Также к его недостаткам относят наличие указателей, дружественных функций и глобальных переменных, которые могут привести к проблемам с безопасностью.

Сравнение основных характеристик вышерассмотренных языков программирования приведено в таблице 1.

Таблица 1. Сравнение языков программирования

Характеристика	Python	Java	C++
Полнота по Тьюрингу	Да	Да	Да
Кроссплатформенность	Да	Да	Нет
Типизация	Динамическая	Статическая	Статическая
Лаконичность и изящность кода*	Да	Нет	Нет
Автоматическое управление памятью	Да	Да	Нет
Наличие механизма обработки исключений	Да	Да	Да
Наличие библиотек для машинного обучения	Да	Да	Да

Для демонстрации лаконичности и изящности кода, помеченного в таблице 1 звездочкой, приведем в качестве примера программу «Hello, world!»[1] на каждом из языков в таблице 2:

Таблица 2. Демонстрация кода «Hello, world!» на Python, Java и C++

Язык	Python	Java	C++
Программа	<pre>print("Hello, world!")</pre>	<pre>public class HelloClass { public static void main(String[] args) { System.out.println("Hello, world!"); } }</pre>	<pre>#include <iostream> int main() { std::cout << "Hello, world!" << std::endl; return 0; }</pre>

C++ – объектно-ориентированный язык программирования, который отлично подходит для разработки ПО, управляемого GUI (Graphic User Interface, графический интерфейс пользователя) и игр [8].

Java – объектно-ориентированный язык, который создает код, предназначенный для запуска на виртуальной машине. Java является популярным выбором для разработки web-приложений и проведения автоматического тестирования [8].

Python – объектно-ориентированный интерпретируемый язык, который получил большую часть своей популярности за счет большого количества библиотек, предназначенных для машинного обучения и научных вычислений. Python является языком, наиболее часто используемым для обработки данных [8].

Таким образом, наиболее подходящим для предобработки и анализа данных является Python за счет простоты в изучении, удобной, на наш взгляд, динамической типизации, наличия большого количества свободно распространяемых библиотек (см. таблицу 1), лаконичного синтаксиса (см. таблицу 2) и востребованности в области анализа данных.

Список литературы:

1. Bogdanchikov, A. Python to learn programming / A. Bogdanchikov, M. Zhaparov, R. Suliyev // Journal of Physics: Conference Series. – 2013. – № 423.

2. Buyya, R., Selvi, S T., Chu, X. Object-oriented Programming with Java: Essentials and Applications. // R. Buyya, S T. Selvi, X. Chu // India:Tata McGraw-Hill 2009. 649 p.
3. IEEE Spectrum. Top Programming Languages 2021. // IEEE [сайт]. URL: <https://spectrum.ieee.org/top-programming-languages/> (дата обращения: 21.04.2022)
4. Java Virtual Machine Specification. Chapter 1. Introduction // Oracle [сайт]. URL: <https://docs.oracle.com/javase/specs/jvms/se8/html/jvms-1.html> (дата обращения: 22.04.2022)
5. Python. History and License - Python 3.10.4 documentation. // Python : [сайт]. – URL: <https://docs.python.org/3/license.html> (дата обращения: 22.04.2022).
6. TIOBE Index. // TIOBE : [сайт]. – URL: <https://www.tiobe.com/tiobe-index/> (дата обращения: 21.04.2022)
7. Маккинни, У. Python и анализ данных / У. Маккинни; перевод с английского А. А. Слинкина. 2-ое изд., испр. и доп. // Москва : ДМК Пресс, 2020. 540 с.
8. Рейтинг языков программирования 2021. // Хабр : [сайт]. – URL: <https://habr.com/ru/post/543346/> (дата обращения: 23.05.2022)
9. Прохоренок, Н.А. Python 3. Самое необходимое / Н.А. Прохоренок, В.А. Дронов. 2-е изд., перераб. и доп. // СПб.: БХВ-Петербург, 2018. 608 с.

References:

1. Bogdanchikov, A. Python to learn programming / A. Bogdanchikov, M. Zhaparov, R. Suliyev // Journal of Physics: Conference Series. – 2013. – № 423.
2. Buyya, R., Selvi, S T., Chu, X. Object-oriented Programming with Java: Essentials and Applications // R. Buyya, S T. Selvi, X. Chu. // India:Tata McGraw-Hill 2009. 649 p.
3. IEEE Spectrum. Top Programming Languages 2021 // IEEE [site]. URL: <https://spectrum.ieee.org/top-programming-languages/> (date of application: 21/04/2022)
4. Java Virtual Machine Specification. Chapter 1. Introduction // Oracle [site]. URL: <https://docs.oracle.com/javase/specs/jvms/se8/html/jvms-1.html> (date of application: 22/04/2022)
5. Python. History and License - Python 3.10.4 documentation // Python : [site]. – URL: <https://docs.python.org/3/license.html> (date of application: 22/04/2022).
6. TIOBE Index // TIOBE : [site]. – URL: <https://www.tiobe.com/tiobe-index/> (date of application: 21/04/2022)
7. McKinney, W. Python for Data Analysis / W. Mckinney : translated from English A.A. Slinkina, 2nd edition // Moscow: DMK Press, 2020. 540 p.
8. Rating of programming languages 2021. // Habr : [site]. – URL: <https://habr.com/ru/post/543346/> (date of application: 23.05.2022)
9. Prokhorenok, N.A. Python 3. The most nessesary / N.A. Prokhorenok, V.A. Dronov. 2nd edition. // SaintPetersburg : BHV-Petersburg, 2018. 608 p.