

## Обзор библиотек Python для анализа геопространственных данных

Научный руководитель – Коваленко Сергей Михайлович

*Ямашкина Екатерина Олеговна*

*Аспирант*

МИРЭА - Российский технологический университет, Институт информационных технологий, Кафедра вычислительной техники, Москва, Россия

*E-mail: eoladanova@yandex.ru*

На сегодняшний день в рамках развития цифровой экономики России инфраструктуры геопространственных данных играют важную роль в решении проблемы поиска средств эффективного анализа, систематизации и использования больших объемов данных, их визуализации и распространения для оптимизации функционирования территориальных систем. Однако возможности существующих программных систем не всегда достаточны для решения данных задач. Расширение функционала системы можно осуществлять с помощью подключения к ней библиотек. В данной статье будут рассмотрены решения для языка Python, предназначенные для анализа и визуализации геопространственных данных.

Библиотека Pandas обладает высокой производительностью и является одной из самых известных библиотек для обработки пространственных данных [1]. Эффективность Pandas заключается в ее структуре данных, оптимизированной для работы с большими данными. Geopandas, как и Pandas, предназначена для работы с геопространственными данными. Но вместо простого табличного анализа Geopandas использует географические компоненты. Для операций наложения используются Fiona и Shapely - собственные библиотеки Python.

Библиотека ArcPy поддерживает обработку и преобразование геопространственных данных, а также составление карт с помощью Esri ArcGIS; GDAL/OGR используется для перевода между GIS-форматами и поддерживает более 250 форматов; RSGISLib представляет набор инструментов для дистанционного зондирования, классифицирует изображения и собирает статистику.

PyProj работает с системами пространственных ссылок, может преобразовывать координаты систем отсчета, а также выполнять геодезические расчеты для любых заданных значений. NumPy получает таблицу атрибутов и помещает ее в структурированный массив, что существенно увеличивает скорость обработки [2]. NumPy может работать с другими библиотеками Python, такими как SciPy, которая предназначена для тяжелых статистических операций.

Для обработки больших массивов данных и их последующего визуального отображения существует библиотека Matplotlib. Она позволяет строить графики, диаграммы, карты. Библиотека Scikit комбинирует NumPy, SciPy и Matplotlib и подходит для анализа данных, классификации или прогнозирования. В работе был представлен обзор некоторых библиотек Python для анализа геопространственных данных. Результаты исследования будут использованы в дальнейшем для решения практических задач.

Благодарности. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-70055.

### Источники и литература

- 1) 0 Minutes to Pandas – Pandas 0.18.0 documentation: <http://pandas.pydata.org/pandas-docs/stable/10min.html>

- 2) Ямашкин С.А. Разработка образовательного курса "Цифровая инфраструктура пространственных данных" // Образовательные технологии и общество. 2019. Т. 22. № 3. С. 113-120.