

Техническое задание на составление таблицы индексов и алгоритм генерации формул

Таблица с индексами

Требуется создать excel таблицу с информацией о **ВСЕХ ПОЛЕЗНЫХ** нам BR и индексах (далее индексы).

Полный список индексов и полезность каждого нужно определить в разговоре с Дмитрием Владимировичем Макаровым.

Индексы нужно брать из:

- <https://www.indexdatabase.de/>
- Статей, которые читал Макаров Дмитрий Владимирович
- Статей, которые читал Унаров Айталь Айалович
- Собственных разработок Унарова Айтала Айаловича
- Из списка использованных нами на Чукотке (файл Чукотка.xlsx)

В результате работы должна появиться excel таблица со следующими колонками

Название колонки	Обязательная	Описание колонки	Пример
Название	Да	Название индекса на английском языке	Adjusted transformed soil-adjusted VI
Название на русском языке	Да	Название индекса на русском языке	Нормализованный вегетационный индекс
Аббревиатура	Да	Краткая аббревиатура индекса. Будет использоваться в названии файлов, поэтому не должна содержать спецсимволы. Должна удовлетворять регулярному выражению ^[\w_]+\$.	GNDLI
Общая формула	Да	Общая формула индекса, в которой участвуют только длины волн или диапазоны длин волн. Конкретная грамматика для записи формул на усмотрение исполнителя.	([2145:2185]*[2235:2365]) / ([2185:2225]*[2185:2225]) Или (850-710)/(850-680) 550/1650

		<p>Формулы должны легко читаться человеком.</p> <p>Формулы должны использоваться в алгоритме генерации формул под конкретный сенсор.</p>	
Формула для Landsat-1/2/3/4/5 MSS	Да, если возможна	Формула должна использовать корректные названия каналов (см. в таблице Приложения/спутники.xlsx) и быть корректным python выражением без побочных эффектов.	(B4-B4)/(B3+B4), B02*B8A / B04**2
Формула для Landsat-1/2/3 RBV	Да, если возможна		
Формула для Landsat-4/5 TM	Да, если возможна		
Формула для Landsat-7 ETM+	Да, если возможна	Гарантируется, что все названия каналов удовлетворяют регулярному выражению ^B\w+\\$	
Формула для Landsat-8/9 OLI/TIRS	Да, если возможна		
Формула для Sentinel-2 MSI	Да, если возможна	P.S. Оно НЕ будет использовать в функциях eval, exec, compile и т.д.	
Формула для ASTER	Да, если возможна	Будет написана своя реализация парсера выражений с использованием библиотеки lark.	
Формула для EMIT	Да, если возможна	Реализовывать это не требуется.	
Формула для EnMAP	Да, если возможна		
Описание	Да	Описание индекса на русском языке. Будет вставляться в autogenerated documentation. Должно быть понятен для геолога. Должно быть заверено у Дмитрия Владимировича Макарова и Дмитрия Александровича Грицая одновременно.	Индекс показывает то-то, то-то. Его можно использовать для того-то, для того-то. Хорошо работает, когда мы ищем такой-то тип месторождения.
Статья	Нет	<p>Статья, из которой был взят индекс. Должно быть названием файла. Сам файл должен быть приложен в папке articles.</p> <p>Если индекс был взят не из статьи или журнала, например, разработан самостоятельно, то нужно</p>	some_article.pdf, vestnik_chukotki_oct_2025.doc

		оставить пустым. Это обязательно должно быть указано в описании.	
Источник статьи	Нет	Ссылка на статью, название журнала или что-то похожее	https://example.com/article.pdf , Журнал «Молодой геолог» июнь 1933
DOI	Нет	DOI статьи, если есть. Должен быть в виде ссылки	https://doi.org/10.5066/F7WD3Z10
Настройка легенды для qgis	Да	<p>Название xml файла с настройками стилей в qgis. Можно получить так: в qgis ПКМ по слою с BR -> Стили -> Копировать стиль -> Все категории. Файл должен быть приложен к таблице в папке styles.</p> <p>Настройки легенды будут применены к автосгенерированному раству с индексом и в автосгенерированном qgis проекте к слою с этим индексом, поэтому должны работать в любых условиях.</p>	В прикреплённом файле Приложения/legend_example.xml
Комментарий	Нет	Дополнительные исполнителя комментарии по поводу индекса	

Все поля, кроме «Комментарий», будут использоваться в автогенерации документации и qgis проекта, поэтому должны быть максимально точными и проверенными.

Шапка таблице есть в файле Приложения/indices_template.xlsx

Алгоритм генерации формулы

В дополнении к таблице требуется разработать алгоритм генерации формулы для расчета по спецификации инструмента съемки.

Функциональные требования

1. Алгоритм должен возвращать формулы под конкретный инструмент на основании общей формулы и спецификации инструмента
2. Алгоритм должен выдавать формулы в упрощённом виде. Например, должно быть не $(B1+B1)/(B2+B2)$, а $B1/B2$.
3. Алгоритм должен быть чистой функцией на python
4. Если получающаяся формула - константа, то алгоритм должен явно сообщать об этом через возврат специального объекта

5. Должны присутствовать полная документация и типизация

Не функциональные требования

1. Алгоритм должен выдавать такие же формулы, как указаны в таблице, для Landsat-1/2/3/4/5/7/8/9, Sentinel-2, ASTER, Emiss, EnMAP
2. Требуется предоставить файл с unit тестами этой функции, как минимум для Landsat-1/2/3/4/5/7/8/9, Sentinel-2, ASTER, Emiss, EnMAP.
3. Код должен запускаться на python >= 3.12.0
4. Код должен использовать минимум внешних библиотек
5. Все python файлы должны проходить проверку ruff == 0.14.2 без ошибок и предупреждений. Настройки ruff приложены в файле Приложения/pyproject.toml
6. Все python файлы не должны изменяться после выполнения команды ruff format. Настройки ruff приложены в файле Приложения/pyproject.toml
7. Все python файлы должны проходить проверку wemake-python-styleguide == 1.4.0 без ошибок и предупреждений. Настройки wemake-python-styleguide приложены в файле Приложения/setup.cfg (секция flake8)
8. Все python файлы должны проходить проверку mypy == 1.18.2 без ошибок и предупреждений. Настройки mypy приложены в файле Приложения/setup.cfg
9. **Спецификация не должна быть захардкожена!**
10. Желательно добавить формулы для EO-1 Hyperion и EO-1 ALI

Спецификация инструмента предоставляется в виде списка каналов. Для каждого канала известно:

Название. Например, B12,

Диапазон в нм в формате начало-конец. Например, 400-2500.

Функция спектрального отклика в виде таблицы с колонками длина волны и отклик.

Поиск, предобработка, разработка формата, преобразование спецификации для алгоритма лежат на исполнителе. Данные о каналах есть в файле Спутники.xlsx. Часть функций спектрального отклика доступна по ссылкам и в приложении:

Инструмент	Ссылка	Файл
Landsat-8 OLI	https://landsat.gsfc.nasa.gov/satellites/landsat-8/spacecraft-instruments/operational-land-imager/spectral-response-of-the-operational-land-imager-in-band-band-average-relative-spectral-response/	Landsat-8 OLI.xlsx
Landsat-8 TIRS	https://landsat.gsfc.nasa.gov/satellites/landsat-8/spacecraft-instruments/thermal-infrared-sensor/spectral-response-of-the-thermal-infrared-sensor/	Landsat-8 TIRS.xlsx
Landsat-9 OLI-2	https://landsat.gsfc.nasa.gov/satellites/landsat-9/landsat-9-instruments/oli-2-design/oli-2-relative-spectral-response/	Landsat-9 OLI-2.xlsx
Landsat-9 TIRS-2	https://landsat.gsfc.nasa.gov/satellites/landsat-9/landsat-9-instruments/tirs-2-design/tirs-2-relative-spectral-response/	Landsat-9 TIRS-2.xlsx
Sentinel-2 MSI	https://sentinelwiki.copernicus.eu/web/s2-mission#S2Mission-MSIInstrumentS2-Mission-MSI-Instrumenttrue	Sentinel-2 MSI.xlsx

	https://sentiwiki.copernicus.eu/_attachments/1692737/COPE-GSEG-EOPG-TN-15-0007%20-%20Sentinel-2%20Spectral%20Response%20Functions%202024%20-%204.0.xlsx?inst-v=f7ebcaa0-df8b-4852-b778-860b5386404b	
ASTER	Нужно найти	-
EMIT	<p>https://upcommons.upc.edu/server/api/core/bitstreams/fdf01653-1a28-4d1b-8657-0947644e3a70/content</p> <p>https://earth.jpl.nasa.gov/emit/internal_resources/280/</p> <p>Гауссиана, аналогично EnMAP</p>	-
EnMAP	<p>https://www.enmap.org/data/doc/EnMAP_FAQ.pdf</p> <p>Раздел 5.2: All the required information for the estimation of the spectral response functions can be found in the product metadata file. The basic information on band center wavelength and full width at half maximum (FWHM) assuming a Gaussian response function are given in the structure bandCharacterisation in units of nm</p> <p>Данные можно взять из таблицы Приложения/Спутники.xlsx</p>	EnMAP_FAQ.pdf

Ожидаемые результаты

В результате ожидается zip архив со следующей структурой:

- specification - папка с техническим заданием
- articles - папка со статьями
- styles - папка со стилями qgis
- indices.xlsx - таблица с описание индексов
- formula_generator.py - функция генерации формулы
- test_formula_generator.py - код для тестирования функции генерации формулы, должен запускаться через pytest

Критерии приёмки

1. Выполнение всех пунктов технического задания
2. Выполнение всех правок к ТЗ
3. Одобрение Макарова Дмитрия Владимировича (научный консультант)
4. Одобрение Грицая Дмитрия Александровича (научный консультант)
5. Одобрение Шаронова Дмитрия Александровича (заказчик)