**Metadata Tensor Arrangement**

Each position in the tensor corresponds to a specific metadata attribute of a cell. The arrangement is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Pos** | **Attribute** | **Description** | **Values** |
| 0 | type | Data type of the cell here each type is mapped to an integer | 0 - 14 |
| 1 | is\_filled | Denotes whether cell is filled or not with 0,1 | 0,1 |
| 2 | halign | Horizontal alignment keys | 0 - 7 |
| 3 | valign | Vertical alignment keys | 0 - 4 |
| 4 | cell\_fam | Family of the font key | 0 - 5 |
| 5 | cell\_size | Size of the font in pt | Float |
| 6 | bold | Denotes whether font is bold or not | 0,1 |
| 7 | italic | Denotes whether font is italic or not | 0,1 |
| 8 | underline | Denotes what underline type font has | 0,1, 2 |
| 9 | script | Denotes what scripting font has | 0,1, 2 |
| 10 | case | Denotes the casing of cell when alphabets are present as inapplicable, lower,upper, mixed case | 0,1,2,3 |
| 11 | is\_col | Denotes whether font is colored or not | 0,1 |
| 12 | btop | Denotes style of top border | 0 - 13 |
| 13 | bbot | Denotes style of bottom border | 0 - 13 |
| 14 | bleft | Denotes style of left border | 0 - 13 |
| 15 | bright | Denotes style of right border | 0 - 13 |
| 16 | hmerge | Holds horizontal merge tag showing if part of merged range or not | 0,1,2 |
| 17 | vmerge | Holds vertical merge tag showing if part of merged range or not | 0,1,2 |

**Metadata Key Mappings**

This section stores the key mappings for each metadata and what each integer key means for each type of metadata in the tensor

**1. Datatype**

The **xlrd** library provides integer keys for the datatype for **XLS** files, now these are exclusive to xls and need to streamlined across **CSV** and **XSLX** files also. From keys 0 to 6 are default types returned by the func and the rest are custom ones

|  |  |  |
| --- | --- | --- |
| **Key** | **Data Type** | **Description** |
| 0 | empty | Empty type which is only retrieved for xls files, has no formatting. This needs to be discarded as in set to nothing for this cell if encountered |
| 1 | text | Text data |
| 2 | number | Numeric type has 4 subclasses |
| 3 | int | Integer subclass of number |
| 4 | float | Decimal subclass of number |
| 5 | pct | Percentage subclass of number |
| 6 | curr | Currency subclass of number |
| 7 | scien | Scientific subclass of number |
| 8 | date | Date type has 2 subclasses |
| 9 | time | Time subclass of date |
| 10 | datetime | Datetime subclass of date |
| 11 | boolean | Boolean type with T or F value |
| 12 | error | Error type corresponding to #REF!, #VALUE! etc. in excel |
| 13 | blank | Empty type with formatting, needs to be checked before discarding |
| 14 | unk | Unknown type not in our keys |

**2. Is Filled (Highlight)**

This key holds 0 if the cell is not highlighted or in excel terms, filled with a color. If the cell has default white background also its considered 0, if any other color besides None or white is encountered then the cell is considered filled with 1

**3. Alignment**

The alignment key in the metadata tensor is mapped the following way

|  |  |  |  |
| --- | --- | --- | --- |
| **Horizontal** | | **Vertical** | |
| **Key** | **Description** | **Key** | **Description** |
| 0 | General | 0 | Top |
| 1 | Left | 1 | Center |
| 2 | Center | 2 | Bottom |
| 3 | Right | 3 | Justify |
| 4 | Fill | 4 | Distributed |
| 5 | Justify |  |  |
| 6 | Center Continuous |  |  |
| 7 | Distributed |  |  |

**4. Font Family**

The ft\_fam key in the metadata dictionary represents the macro-style font family used in a cell. This value is stored as an integer, corresponding to the specific style of the font. The mapping of the integer values to font family styles is as follows:

**Key (Integer) Font Family Style**

0 Not Applicable

1 Roman (e.g., Times New Roman)

2 Swiss (e.g., Arial)

3 Modern (e.g., Courier New)

4 Script (e.g., Brush Script)

5 Decorative (e.g., Old English)