We’ve seen in previous tutorials that parsing a 10-K for the desired information can be a daunting task. As there are multiple roadblocks that get in the way of getting the data. For example, making sure the formatting of the HTML code is consistent or having to convert data to the proper data type after pulling it from the file. In some cases, we will have to resort to this strategy, especially for other financial forms like financial disclosures, insider purchases, and other more specific forms.

However, for the 10K & 10Q form we actually have an easier way to get the data. Inside each filing folder for these forms you’ll notice a multiple documents. Let’s break down what these different forms are:

1. The Index-Headers-HTML**:** This form is the header for the directory, if you open it you’ll notice it’s XML structured data that lists each document, the type of document it is, a link to the document, and some general information about the company. This information is very hard to parse, so I do not recommend parsing it, as there is an easier file to parse it from.
2. The Index-Headers: This file directs you to the EDGAR Query results for that company’s given filing. It would be as if you had gone to the EDGAR Search Tools page, typed in that company’s CIK number, and filing number. This data is definitely more structured than the Index-Headers-HTML file, but it will not contain all the individual file links. However, if I had to parse either one, I would parse this one before that the Headers-HTML one.
3. The .TXT File: This file contains all the information from the 10K or 10Q in a text file. Now it’s the raw HTML code, so in order to get the information out of this file you would have to parse it. It is structured but it does require a good amount of exploration to make sure you are parsing the right items. In previous videos I showed you how to parse, certain sections of the 10K.
4. The XBRL.Zip File: This file is all the files in the directory zipped up in a single file for convenient download, this will come in handy if you don’t want to request each file individually. All we would need to do is download this file.
5. **FilingSummary.xml:** This file is very important! I mentioned up above that the Index-Headers-HTML file is a pain to parse. Parse this one instead, the data is structured in a more organized fashion. Also there are more descriptive features that we can leverage down the road. It contains all the files for this particular filing.
6. Financial-Report.xslx: This is an excel version of the file. Definitely, useful but sometimes hard to navigate. A good macro could make this file much more workable.

The next group of files are named with the following structure:

* 1. **tradingSymbol\_endDateOfSubmission\_FileType.xml**

1. TradingSymbol\_EndDate.xml: This is the file we will work with heavily; it contains all the information in the 10K/Q but in XML format that will make parsing it way easier.
2. TradingSymbol\_EndDate.xsd: This contains all the links to the investor webpage presentations, keep in mind that these links will expire as new filings come out. However, if you want to immediately reference portions of the presentation after they file this would be the file to parse.
3. TradingSymbol\_EndDate\_cal.xml: This contains links to different calculations in the file this will be useful when trying to reference how certain calculations were made.
4. TradingSymbol\_EndDate\_def.xml: This contains links to different definitions used in the 10K/Q and also external resources that define those definitions, this would be entities like FASB.
5. TradingSymbol\_EndDate\_lab.xml: This contains a section for each label in the document along with their corresponding ID used in the XML structuring. Additionally each one these labels also has a locator that will provide a hyperlink to that label in the file.
6. TradingSymbol\_EndDate\_pre.xml: This provides more links to different presentations that contain information regarding different sections of the filing.
7. Files with an R: These are the different reports in the filing, and can range from balance sheets to debt repayment schedules. This will vary on the company but the information can be found in the TradingSymbol\_EndDate.xml file.
8. FILETYPE\_YEAR\_ex\_NUM\_COMPANY: These are specific exhibits in the filing, and are highlighted in the directory.

# Parsing the Filing Summary:

I do encourage, that anyone who is planning to parse a 10K or 10Q that they start with the filing summary, this way you have a good overview of the files available and more importantly it’s a good way to build a dictionary that will store all the information in the different files. Here is the structure of the Filing Summary:

<?xml version='1.0' encoding='utf-8'?>

<FilingSummary>

    <Version>3.19.2</Version>

    <ProcessingTime/>

    <ReportFormat>html</ReportFormat>

    <ContextCount>122</ContextCount>

    <ElementCount>262</ElementCount>

    <EntityCount>1</EntityCount>

    <FootnotesReported>false</FootnotesReported>

    <SegmentCount>19</SegmentCount>

    <ScenarioCount>0</ScenarioCount>

    <TuplesReported>false</TuplesReported>

    <UnitCount>4</UnitCount>

    <MyReports>

        <Report instance="bdr-20190630.xml">

            <IsDefault>false</IsDefault>

            <HasEmbeddedReports>false</HasEmbeddedReports>

            <HtmlFileName>R1.htm</HtmlFileName>

            <LongName>Document - Document and Entity Information</LongName>

            <ReportType>Sheet</ReportType>

            <Role>URL/role/DocumentAndEntityInformation</Role>

            <ShortName>Document and Entity Information</ShortName>

            <MenuCategory>Cover</MenuCategory>

            <Position>1</Position>

        </Report>

</MyReports>

    <InputFiles>

        <File>bdr-20190630.xml</File>

    </InputFiles>

    <SupplementalFiles/>

    <BaseTaxonomies>

        <BaseTaxonomy>http://fasb.org/us-gaap/2019-01-31</BaseTaxonomy>

    </BaseTaxonomies>

    <HasPresentationLinkbase>true</HasPresentationLinkbase>

    <HasCalculationLinkbase>true</HasCalculationLinkbase>

</FilingSummary>

Looking at the data up above, we can see that the main entry point into this XML tree is the <FilingSummary> tag. Immediately, we get some information about the document, regarding the report format, segmentation count, and entity count. This is good information if we need to parse the additional information later because it’ll make sure we know how many values we are expecting and things along that nature.

Our main concern will be the <MyReports> list, the <InputFiles> list, and the <BaseTaxonomies> list. The <MyReports> tag contains a list of all the reports created for this particular filing, you’ll notice that each report has the following information:

<Report instance="bdr-20190630.xml">

<IsDefault>false</IsDefault>

<HasEmbeddedReports>false</HasEmbeddedReports>

<HtmlFileName>R1.htm</HtmlFileName>

<LongName>Document - Document and Entity Information</LongName>

<ReportType>Sheet</ReportType>

<Role>URL/role/DocumentAndEntityInformation</Role>

<ShortName>Document and Entity Information</ShortName>

<MenuCategory>Cover</MenuCategory>

<Position>1</Position>

</Report>

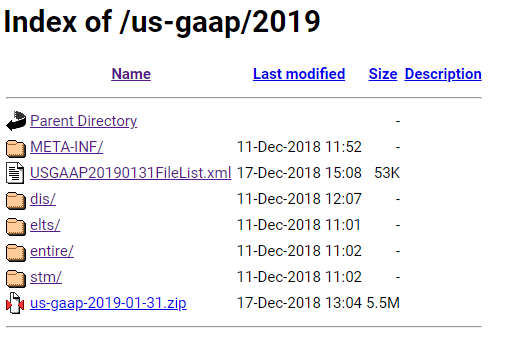
When I was exploring this information, I found that the position information was useful, the name both long and short, and the instance attribute. What I found was to store this information in a python dictionary, that way I could keep a reference of all the documents in a given filing. Eventually, I would take that information and store it in a SQL database, that would be my master filing list. Here is the code for converting the Filing Summary to a python dictionary.

Here are some common abbreviations you’ll encounter while exploring the different directories.

**Directories For Entry Points Schema:**

|  |  |
| --- | --- |
| Abbreviation | Description |
| dis | Disclosures |
| elts | Elements |
| stm | Statements |
| entire | Entry point for entire Taxonomy |
| META-INF | Manifest file to identify entry points automatically |

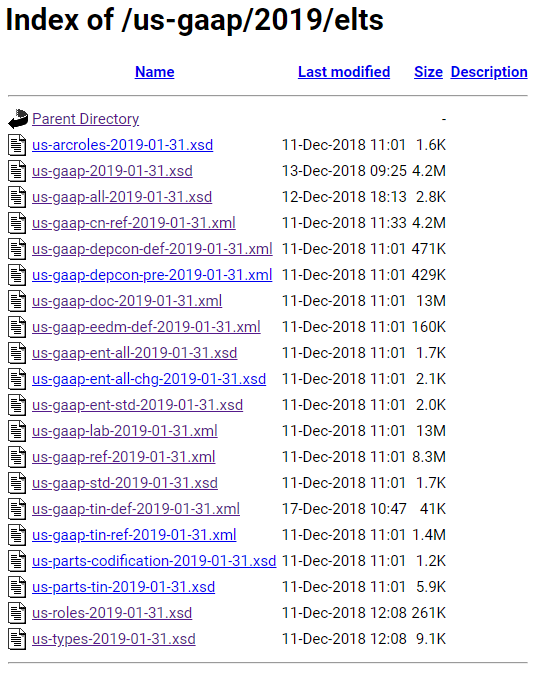
Here is an example of how the directory will look, along with abbreviations.



**Abbreviations Used in File Names:**

|  |  |
| --- | --- |
| Abbreviation | Description |
| -all- | Contains labels, relationships with information about deprecation, and documentation and references for concepts. |
| -std- | Loads the Taxonomy with labels but no documentation or references. |
| -dep- | Contains labels and relationships with information about deprecation |
| -chg- | Contains descriptions and relationships with information about Taxonomy changes and taxonomy implementation notes. |
| -eedm- | Contains domain of members for use with concepts of type enum:enumerationsItemType. |
| -\*tmp- | Contains taxonomy disclosure templates |

Here is an example of a directory of files that contain different abbreviations.



**Abbreviations Entry Point Types:**

|  |  |
| --- | --- |
| Abbreviation | Description |
| -dis- | A disclosure schema or linkbase |
| -ent- | A document schema entry point |
| -stm- | A statement schema or linkbase |
| -entryPoint- | The root of the entire taxonomy |
| -tmp- | Taxonomy disclosure templates |
| -wotmp- | Excludes taxonomy disclosure templates |

**Abbreviations Statement Types:**

|  |  |
| --- | --- |
| Abbreviation | Description |
| -com- | Common, contains definitions and other relationships whose only purpose is to be copied by users into other links |
| -scf- | Statement of Cash Flows |
| -scp- | Statement of Partner Capital |
| -sfp- | Statement of Financial Position (Balance Sheet) |
| -she- | Statement of Shareholder Equity |
| -soc- | Statement of Comprehensive Income |
| -soi- | Statement of Income |

**Prefixes For the Main File Groups:**

|  |  |
| --- | --- |
| Abbreviation | Description |
| us-gaap- | U.S. GAAP Taxonomy prefix |
| srt- | SEC Reporting Taxonomy prefix |

**Abbrevations Link Base Groups:**

|  |  |
| --- | --- |
| Abbreviation | Description |
| -cal- | calculation |
| -def- | definition |
| -doc- | Documentation |
| -lab- | labels (contains labels having standard role "label" and others) |
| -pre- | presentation |
| -ref- | reference |
| -dep- | deprecation (contains relationships among deprecated and normal concepts) |
| -tin-ref- | Taxonomy implementation notes using reference syntax |
| -tin-def- | Taxonomy implementation note relationships in definition linkbase to associate an alternative concept for a superseded concept for period of and after adoption; identify which particular element in the financial statements includes the value for that concept when the element representing the concept is not separately disclosed in the financial statements. |
| -cn-ref- | Taxonomy change notes using reference syntax |
| -\*tmp\*- | Linkbase used with a Taxonomy Disclosure Template |