This is what I did to get from the input document to the output document.

1285550 –

1. Go to the sheet “Note\_2\_Summary\_of\_Significant\_”
2. Each line that was italicized appears to be one of the accounting policies that I’m after.
3. Copy and paste the text into the “Text” column of the output sheet.
4. I’m not sure how to handle something like the “Revenue Recognition” policies as it has several sub-policies. Ideally I would like all the text of the individual sub-items just rolled up into the “Revenue Recognition” row but I don’t see anything in the xlsx sheet that I could systematically use as a distinction there. If you have ideas here that would be great.
5. Important. I do NOT want/need the tables. The text is what I’m interested in.

1259515 –

1. Go to the sheet “Description\_of\_Business\_and\_Su”
2. It appears that there is significant variation in the naming of the sheet that I want extracted. It is likely that I’ll have to iterate on the Regex that extracts the sheet. Besides the name, another distinction in the sheet that I want is that the sheet is always mostly text and there is more than 1000 words of text.
3. Each line that was italicized and bolded was one of the accounting policies.
4. Copy and paste the text into the “Text” column

351789 –

1. Go to the sheet “Basis\_of\_Presentation\_Policy”
2. Again, funky name. Most of the target sheets will be called something like “Summary of Significant”. There is a cell in the sheet with the text “Nature of Business and Significant Accounting Policies [Abstract]”. It appears to be unique to the sheet that I want. So maybe we’ll have to look within sheets for the phrase “Significant Accounting Policies” rather than the name of the sheet.
3. Each line that was bolded was one of the accounting policies.
4. I’ve noticed that even copy/pasting manually is cumbersome because of the blank rows that they leave in between blocks of texts within a given accounting policy.
5. Copy and paste the text in to the “Text” column.

13372 –

1. Go to the sheet “SUMMARY\_OF\_SIGNIFICANT\_ACCOUNT”.
2. Finally, a normally named sheet. As I’ve been doing this, it wouldn’t be the end of the world if I had to manually help the script choose correct sheet.
3. Each line that was bolded was one of the accounting policies.
4. Seems like the accounting policies being italicized or bolded is going to help significantly in identifying the policies.
5. Each line that was bolded was one of the accounting policies. In addition, this file numbers the policies alphabetically.
6. Copy and paste the text into the “Text” Column.
7. Potential issue. This file italicizes a lot within the text fields. I doubt that this is prevalent and may just be an example of a failure.
8. Observation. Seems like the text that I want is always coming from the “B” Column.