Let's now go ahead and read the document WExercise.docx file. The

following code will help us in getting the object that points to the

WExercise.docx file:

import docx

doc = docx.Document('WExercise.docx')

print "Document Object:", doc

The output of the preceding code is shown in the following screenshot. Reading a

Word document is conceptually pretty similar to reading a file in Python. Just like

how we created a file handle using the open() method, we create a document

handle in this code snippet:

3. Now, if we want to get the basic information about a document, we can use the

document object, doc, from the preceding code. For instance, if we want to

retrieve the title of a document, we can do that with the following code. If you

look at the code carefully, we use the paragraphs object to get the text.

Paragraphs are the lines present in the document. Assuming that the title of the

document is the first line of the document, we get the 0 index for the paragraphs

in the document and call the text attribute to get the text of the title:

import docx

doc = docx.Document('WExercise.docx')

print "Document Object:", doc

print "Title of the document:"

print doc.paragraphs[0].text

Notice in the following screenshot how we print the title of our exercise

document:

4. Wow, that's cool! Let's go ahead and read other attributes of the Word document

that we care about. Let's use the same doc object:

print "Attributes of the document"

print "Author:", doc.core\_properties.author

print "Date Created:", doc.core\_properties.created

print "Document Revision:", doc.core\_properties.revision