



# XML Files Types and Their Differences and Denominators with Plain TXT File

Zahir AlHashami\*

CS Instructor, Oman-Muscat

## Abstract

Easy and simplicity of transformation, sharing and storing files among programs, platforms and machines. From these formats are XML formats and plain and flat TXT file formats. Complex formats of the files will cause difficulty in treating with, for instance files transformation while the simple files could treat with them with insurance of no need for compatibility requirements and no virus effect. The method of this research are content size and comparison between XML file format and TXT flat file. The XML file formats are very simple to treat in structure in such as transforming and in converting among various applications. This study got that XML files are structured programmed content and took some Kilobytes size, such as XML Spreadsheets formats which are taken for a text and a picture content just 3KB, whilst in XML PowerPoint Presentation took 97KB. TXT plain txt file is simpler than XML file which are both text based content because that, the same content after converting to TXT format took just 1KB. The plain text file format is fairly not organized and unstructured text-based content except the head section TXT file.

**Keywords:** Simplicity; XML; Flat TXT File; Transformation; Size; Structure

## Introduction

XML stands for an Extensible Markup Language (XML) file format used to create, share and transform information formats and share both the format and the data on the World Wide Web (internet) by use standard ASCII text.

Usage and the aims of XML are to simplify, generalize the data across the internet. It is a textual data with strong support by coding the data for different human languages. It widely used for the representation of data such as those used in web services.

XML is similar to Hypertext Markup language (HTML). Both XML and HTML contain markup symbols to describe the contents of a page or file.

The identification of XML based on: Norman Walsh (1998), said: that, it is a markup language for documents containing structured information. This structured information contains the words and the pictures. In the content of XML document the section heading has a different meaning from content in a footnote, which means something different than content in a figure caption or content in a database table. Also he said the all documents have some structure [1].

## XML files sample

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
<to>Tovee</to>
<from>Jani</from>
<heading>Reminder</heading>
<body> Don't forget me this weekend!</body>
</note>
```

## XML file formats transformation with office programs

In the 1990s and after advent of XML, based on Frank Rice (2006), corporation customers realize the valuable of open formats files and basic standardization in computer-regard products and applications.

As example in Microsoft Corporation the IT specialists get benefit from the files formats are possible open with XML format, because of two reasons: first the capacity and files size and second is the easily to read by many applications, and many operating system as well as internet browsers.

Historically transformation transition begun with developers to see the need to transform from binary files format to XML format. Office programs are able to do many functions but can face such challenges included transporting and easily moving data between Operating Systems and various applications.

Frank Rice also mentioned that, MS 2007 in Microsoft Corporation has the ability to transform such as: Doc, Ppt and Xls to XML files format. The new file format called Office Open XML formats. With Open office XML format can be accessed and supported with any application, Frank Rice [2].

## TXT file format

TXT file identification is a file extension for a text file, used by a variety of text editors like Note. Text is a human-readable sequence of the characters and the words they form that can be encoded into computer-readable format. There is no standard definition if a text file, though there are several common formats, including ASCII and ANSI.

According to Varun Agarwal (2008), Flat file is generally a delimited file; with some specific delimiter such as semi-colon, comma. It shell may specify the delimiter in map properties tab in Syndicator. Whereas in case of XML file should need to specify the schema which will define the structure of XML file [3].

\*Corresponding author: Zahir AlHashami, CS Instructor, Oman, Tel: 96896166363; Fax: 96896166363; E-mail: [ac.res@hotmail.com](mailto:ac.res@hotmail.com)

Received July 16, 2015; Accepted August 05, 2015; Published August 12, 2015

Citation: AlHashami Z (2015) XML Files Types and Their Differences and Denominators with Plain TXT File. J Theor Comput Sci 2: 130. doi:10.4172/2376-130X.1000130

Copyright: © 2015 AlHashami Z, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### TXT files content sample

```

===Plugin Name ===
Contributors: (this should be a list of wordpress.org user ID's)
Donate link: http://example.com
Tags: comments, spam
Requires at least: 3.0.1
Tested up to: 3.4
Stable tag: 4.3
License: GPL v2 or later
License URL: http://www.gnu.org/licenses/gpl-2.0.html
    
```

Here is a short description of the plugin. This should be no more than 150 characters. No markup here.

== Description ==

This is the long description. No limit and can use Markdown (as well as in the following sections).

For backward compatibility, if this section is missing, the full length of the short description will be used, and markdown parsed.

A few notes about the sections above:

“Contributors” is a comma separated list of wp.org/wp-plugin.org username

“Tags” is a comma separated list of tags that apply to the plugin

“Requires at least is the lowest version that the plugin will work on

“Tested up to” is the highest version that you’ve successfully used to test the plugin\*. Not that it might work on higher version... just this the highest one you’ve verified.

Stable tag should indicate the subversion “tag” of the latest stable version, or “trunk,” if you use /trunk/ for stable.

### Application case of XML files use

Analyzing data in the DrugBank aspect is included detailed drugs like chemical and pharmacological compounds by using XML documents. Each drug described with than 150 data fields. There is an application name KNIME, the XML files loaded into it. The XML processing nodes extract several information from XML file like compound name and compound description by using XPath node. The row filter is responsible to extract the information from approved drugs. In the Meta node in the XML file may extract further textual features like the organisms and manufacturer, which are including other free text fields like description and pharmacology. These are analyzed in text mining part of the XML document, Koetter, 2012 [4].

## Results

### Content size

The XML files of Word document, Excel book and Power Point presentation contain the same content, which is text identification of a computer and a figure of the computer as appear in the below Figure 1.

Converting the Office programs to XML files appeared various sizes like in the Table 1.

### Size of XML word document and TXT plain text file

The same content in XML word and TXT plain text size will be like the Table 2.

### TXT and word XML document

**TXT file:** When convert and open XML word document with same content which is containing text and one picture, all the file appears just text file like the below Figure 2.

### XML word document

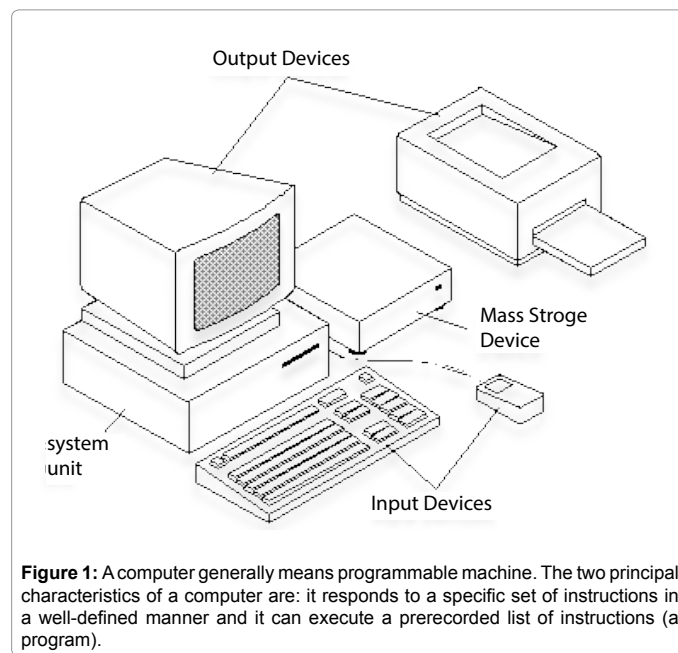
When converting the same file that containing text and one picture as XML file, the document will appear in structured and organized XML file like the below Figure 3.

**XML document file in note:** Opening XML file document by the Note will change the structure of organized and hierarchical structure of the XML file and appear like flat TXT file Figure 4.

## Discussion

XML files used for analyzing data in the DrugBank aspect which is included XML data with XML formats. These uses are like, medicine description, names of medicine and chemical and pharmacological compounds by using XML documents [4].

In the Stackoverflow.com website, Pascal Martin (2015) in his article titled: «XML file and text file» said that, XML file is structured document and standard validation consists of DTD, schema and standardized way to parsing the structure of XML. In the other hand



**Figure 1:** A computer generally means programmable machine. The two principal characteristics of a computer are: it responds to a specific set of instructions in a well-defined manner and it can execute a prerecorded list of instructions (a program).

XML Spreadsheet	XML word document	XML Power point presentation
3KB	57KB	97KB

**Table 1:** Converting the Office programs to XML files appeared various sizes like in this table.

TXT file	XML word document
1KB	57KB

**Table 2:** The same content in XML word and TXT plain text size will be like the table below.

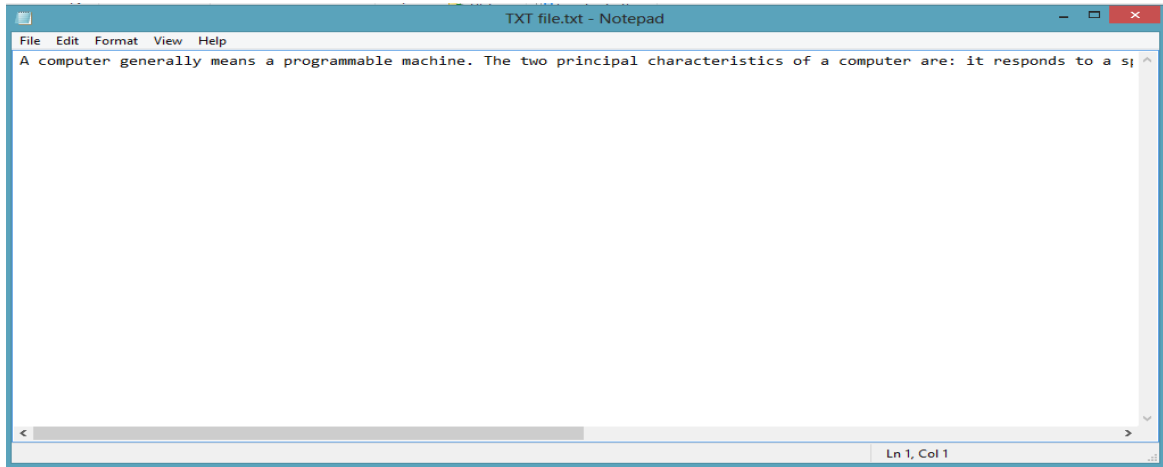


Figure 2: Conversion of open XML word document.

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
<?mso-application progid="Word.Document"?>
- <pkg:package xmlns:pkg="http://schemas.microsoft.com/office/2006/xmlPackage">
  - <pkg:part pkg:padding="512" pkg:contentType="application/vnd.openxmlformats-package.relationships+xml" pkg:name="/_rels/.rels">
    - <pkg:xmlData>
      - <Relationships xmlns="http://schemas.openxmlformats.org/package/2006/relationships">
        <Relationship Target="docProps/app.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/extended-properties" Id="rId3"/>
        <Relationship Target="docProps/core.xml" Type="http://schemas.openxmlformats.org/package/2006/relationships/metadata/core-properties" Id="rId2"/>
        <Relationship Target="word/document.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/officeDocument" Id="rId1"/>
      </Relationships>
    </pkg:xmlData>
  </pkg:part>
  - <pkg:part pkg:padding="256" pkg:contentType="application/vnd.openxmlformats-package.relationships+xml" pkg:name="/word/_rels/document.xml.rels">
    - <pkg:xmlData>
      - <Relationships xmlns="http://schemas.openxmlformats.org/package/2006/relationships">
        <Relationship Target="webSettings.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/webSettings" Id="rId3"/>
        <Relationship Target="settings.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/settings" Id="rId2"/>
        <Relationship Target="styles.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/styles" Id="rId1"/>
        <Relationship Target="theme/theme1.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/theme" Id="rId6"/>
        <Relationship Target="fontTable.xml" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/fontTable" Id="rId5"/>
        <Relationship Target="media/image1.gif" Type="http://schemas.openxmlformats.org/officeDocument/2006/relationships/image" Id="rId4"/>
      </Relationships>
    </pkg:xmlData>
  </pkg:part>
  - <pkg:part pkg:contentType="application/vnd.openxmlformats-officedocument.wordprocessingml.document.main+xml" pkg:name="/word/document.xml">
    - <pkg:xmlData>
      - <w:document xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape"
        xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml" xmlns:wpi="http://schemas.microsoft.com/office/word/2010/wordprocessingInk"
        xmlns:wpg="http://schemas.microsoft.com/office/word/2010/wordprocessingGroup" xmlns:w15="http://schemas.microsoft.com/office/word/2012/wordml"
        xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml" xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
        xmlns:w10="urn:schemas-microsoft-com:office:word" xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing"
        xmlns:wp14="http://schemas.microsoft.com/office/word/2010/wordprocessingDrawing" xmlns:v="urn:schemas-microsoft-com:vmml"
        xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math">
```

Figure 3: Structured and organized XML file.

text file is easier to write which is no need to respect and write any structured tag and that text file not necessary to have a well-defined structure [5].

Also Jacek Konieczny (2015) about the same article in Stackoverflow.com commented that the XML file is a text file, but with a well-structured framework which provides ways to respect complicated data structures.

There are attributes that XML file has like verbosity and higher processing costs. The XML file is good choice when use different applications or systems to being data exchanged, while these data is preferred to be well-structured [6].

Coombs, James and Allen (1987), said that files which contain markup or other meta-data can consider them as flat or plain text while

remain text readable [6].

The flat or the plain text file is use to be recovered and to be saved for viruses trouble and loss of data. To save it in an immune disk in computer or any media and save it in case of incompatibility. According to Unicode Standard that the plain text is a pure sequence of character codes. Styled text, also known as rich text, is any text representation contains plain text with font size, color and hyperlink [7].

Vangie Beal (2015), also said about text (TXT) file that it is in each byte of the text represents one character based on ASCII Standard (American Standard Code for Information Interchange) code. He mentioned that the files that are formatted with word processor to save the formats must be as binary file to be stored.

The ASCII files are textual information or TXT data. The plain text



## References

1. Norman W (1998) A Technical Introduction to XML.
2. Frank R (2006) Introducing the Office (2007) Open XML File Formats.
3. Varun A (2008) Difference between XML & Flat files.
4. Koetter T (2012) DrugBank.
5. Pascal M, Jacek K (2015) XML file and text file.
6. James CH, Renear AH, DeRose SJ (1987) Plain text.
7. Vangie B (2015) ASCII file-Plain Text.
8. Paul M (2015) Plain text files-Flat Files.