Information Management Resource Kit

Module on Management of Electronic Documents

UNIT 2. FORMATS FOR ELECTRONIC DOCUMENTS AND IMAGES

LESSON 2. PROCEDURAL MARK-UP

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.

mark

© FAO, 2003

Objectives

At the end of this lesson, you will able to:

• understand the advantages and limitations of **Microsoft Word**; and

• understand the function of **Rich Text Format**.

Word processors

The most common form of procedural mark-up is that used by **word processing software**.

Word processors are designed primarily to create documents that are to be delivered in **print**, although many people exchange word processing documents by email.

By far the most popular word processor is **Microsoft Word**, although there are other products available, such as **Corel's WordPerfect**.
In recent years, Sun Microsystems have launched the **Star Office** suite as a freely available alternative to Microsoft Word (this is available as open source software, named Open Office, from www.openoffice.org).

Word processors

The mark-up in a word processor serves **two main purposes**. **The first** is to specify how the document should be rendered (**laid out**) when printed.

Some of the rendering **features available** in a word processor like **Microsoft Word** include:

What is clear
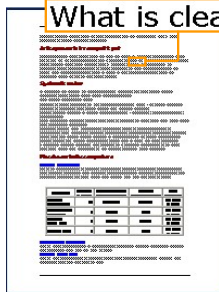
Formatting of text: font family (Times New Roman, Arial, etc), size (10pt, 12pt, etc), style (bold, italic, etc).
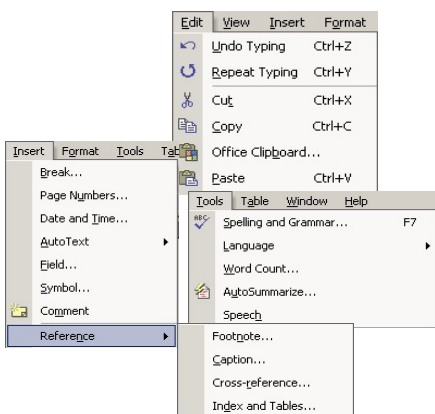
Page Layout: document headers and footers, page size, page orientation (landscape, portrait).

Paragraph level layout: justification, line spacing.

Lists: bulleted or numbered, with indents and nesting (lists within lists).

Tables: with rows and columns, with formatting for table cells, such as spacing, borders, justification (horizontal and vertical).


Word processors

The second purpose of mark-up is to control the **functions** of the word processing application.

Some of the **available functions** include:

- • inserting and deleting text and images;

- • cutting, copying and pasting text and images;

- • tracking (and undoing) changes made to the document;

- • spell checking and grammar checking; and

- • generating indices and tables of contents.

Edit	View	Insert	Format
Undo Typing	Ctrl+Z		
Repeat Typing	Ctrl+Y		
Cut	Ctrl+X		
Copy	Ctrl+C		
Office Clipboard...			
Paste	Ctrl+V		

Insert	Format	Tools	Tab
Break...			
Page Numbers...			
Date and Time...			
AutoText	▶		
Field...			
Symbol...			
Comment			
Reference	▶		

Tools	Table	Window	Help
Spelling and Grammar...			F7
Language			▶
Word Count...			
AutoSummarize...			
Speech			

Footnote...
Caption...
Cross-reference...
Index and Tables...

Stylesheets

Procedural mark-up contains codes that specify how the document is to be displayed on screen or in print.


TITLE

SUBTITLES
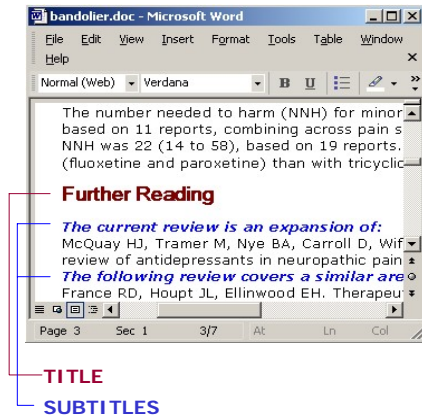
To create a document with a consistent **look and feel**, you could apply the same style to each section header by selecting the correct font, size and layout for each header.

However, to make sure that the section headers are displayed consistently, and to reduce the amount of work needed to format the document, you can **define a style** for a header and **then apply** it to each piece of header text.

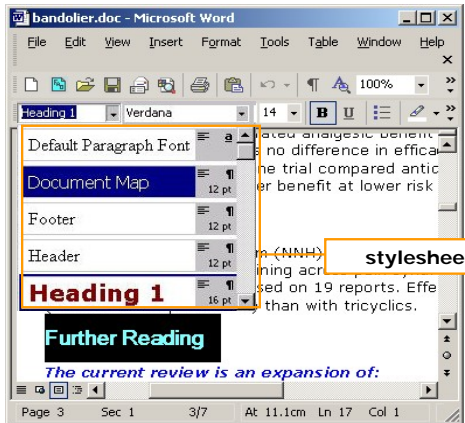Let's see how this is done...

Stylesheets
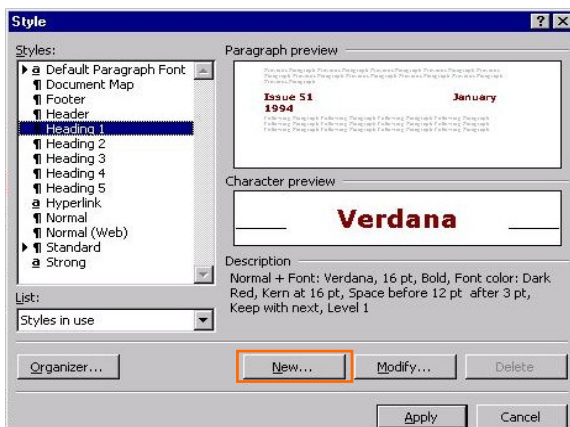
To **define** a style in Microsoft Word, you can:

• select **New Style** from **Format/Style** menu;

• set the features of your style; and

• click on the OK button.

In this case the style will display the header text in Verdana font, 16pt bold, Dark Red with 12pt space before and 3pt space after.

Stylesheets

stylesheet

To **apply** a style to a heading, you can:

• highlight the heading; and

• from **Format/Style** menu, select the style you have chosen for this heading.

The **set of styles** used to format a particular type of document is commonly referred to as a **stylesheet**.

Templates

When you have to create documents of the **same type** or class, like memos, letters, technical reports, research articles and invoices, document **templates** can help you.

This is the last report for today... If I didn't have the template I wouldn't have made it on time!
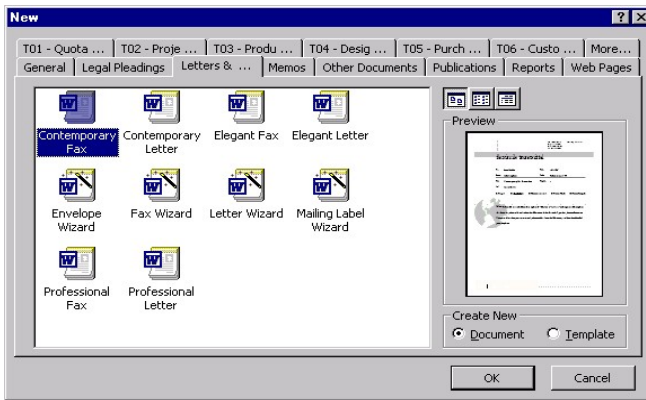

The template will typically **contain the stylesheet** that will be used to format this type of document.

In addition, it will also contain a **framework** with elements such as a standard front page, headers and footers, a standard set of sections and headings, etc.

The framework can be modified by the user to create individual documents of that class.

Templates

How do you apply a template?

When creating a new document you can select from a **library of templates** on which to base your new document.

Some software applications will allow you to apply a different template to a document you have **already created**, so that all the styles and standard text elements are changed instantly.
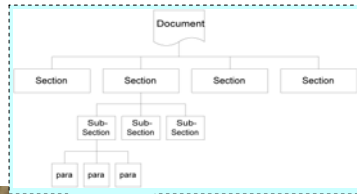

Templates

To summarize, what is the relationship between the template, the stylesheet and the style?

1

| | CONTAINS | → | | CONTAINS | → | |

A

| STYLE | TEMPLATE | STYLESHEET |

Click each option, drag it and drop it in the corresponding box.
When you have finished, click on the confirm button.

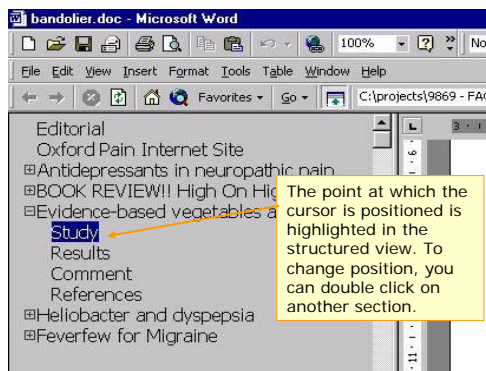**Structured Documents and Procedural Mark-up**

If you want to organize a document with a set of sections, containing sub-sections and paragraphs, then you need to create a **structured document**.

A structured document is one which consists of a set of elements that are organized in a **hierarchical structure**.

We can create such a structured document using our word processor by creating **styles that define** section and sub-section **headers** as well as the **text** in the paragraphs.

Structured Documents and Procedural Mark-up

In MS Word, if you select the option View/Document Map you can view an **expandable tree** that represents the structure of your document.

The point at which the cursor is positioned is highlighted in the structured view. To change position, you can double click on another section.

But remember: in a word processor the header styles merely indicate **where sections or sub-sections start**, rather than defining elements in the hierarchical structure.
For example, you can put a Header2 (indicating the start of sub-section) without first entering Header1 (used for a section).

The procedural mark-up doesn't support the concept nor enforce the rules of a structured document, it just gives **visual clues** as to what the underlying structure is.

Standardization problems

We have seen that Microsoft Word offers several benefits. However, there is a problem **when you open a Microsoft Word file in a plain text editor**.

Is this image distinguishable?
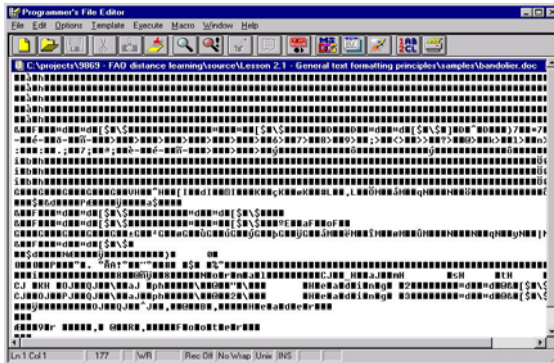
As you can see, most of the mark-up codes cannot be displayed in a plain text editor.

This is because Microsoft Word files use a **proprietary, binary format** for their procedural mark-up.

But what is a proprietary binary format?

Standardization problems

Proprietary, binary formats cause some problems in terms of standardization and open information exchange:

• Users are obligated to use only applications from the vendor that owns the format.

• The vendor may (and usually does) change the format without notice or consultation with users – older versions slowly become obsolete, since they do not support the features available in the latest releases of the word processing application.

• It is difficult for other vendors or individuals to create applications that process the same mark-up format.

A **proprietary** format is one that is owned and controlled by a particular commercial organization and (generally) the details of the mark-up coding scheme are **not available to other** organizations or individuals.
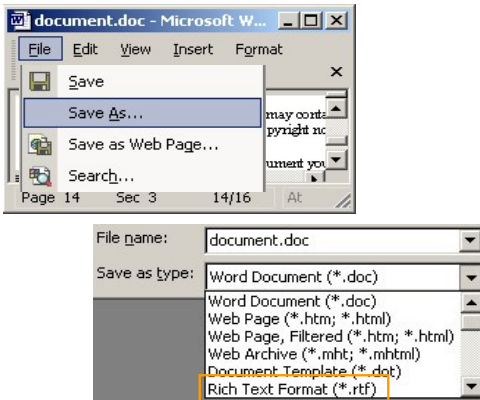
A **binary** format is one which represents characters using codes that lie outside the normal range (1 to 128) that can be displayed by a plain text editor.

Proprietary, binary formats cause some problems in terms of **standardization** and open information exchange.

More information about standardization problems

Rich Text Format

To alleviate some of the standardization problems, Microsoft have invented the procedural format **Rich Text Format** (**RTF**).

RTF is a **plain text** format, meaning that you can open and read RTF files in **any standard text editor**.

It is easy to create an RTF file from most word processing applications:

• select **File/Save As** from the main menu, and
• select **RTF** from the list of format options.

Rich Text Format

You can take an RTF file that has been created in this way and open it using a **plain text editor**. As you can see, all the text and mark-up codes can be displayed (and edited), although it is not very easy to distinguish the text content of the document from the RTF mark-up codes.
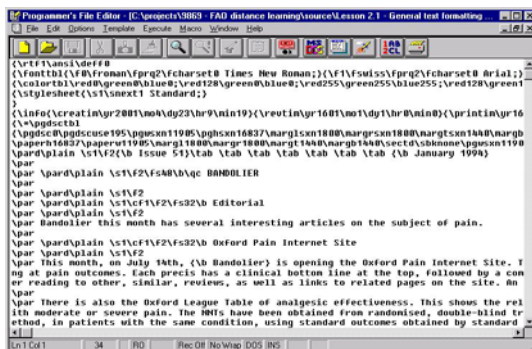
RTF mark-up contains all the information necessary to support the layout and functionality of the Microsoft Word processing application.

Because of these features, RTF is often used as the **exchange format** between word processing applications from different vendors.

Summary

• **Word processing** software uses the most common form of procedural mark-up.

• Word processing format, such as Word, is useful when you have to **create or edit** a document.

• The mark-up in a word processor serves to specify how the document should be **laid out** when printed, and to control the **functions** of the word processing application.

• Using a word processor such as Microsoft Word, you can set the **stylesheet**, apply **templates** and create a **visual structure** for your document.

• Microsoft Word use a proprietary, binary format: this causes problems in terms of **standardization**.

• To resolve these problems, Microsoft have created another procedural format, RTF, that is a plain text format used as the **exchange format** between word processing applications.

Exercises

The following four exercises will allow you to test your understanding of the concepts covered in the lesson and provide you with feedback.

Good luck!

Exercise 1

Why is it appropriate to use a Word Processor format such as Microsoft Word as the source format of a document?

• Because it can be used as exchange format among different applications.

• Because Word Processors are designed for creating and editing documents.
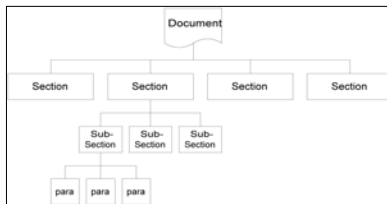
Click on the answer of your choice

Exercise 2

What is a stylesheet?

• The set of styles used to format a particular type of document.

• The set of features of a particular style (e.g. type of font, size, layout).

• A framework with elements such as a standard front page, headers and footers, etc.

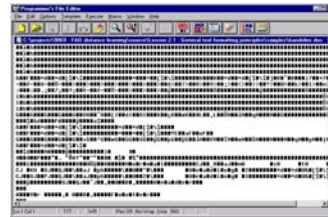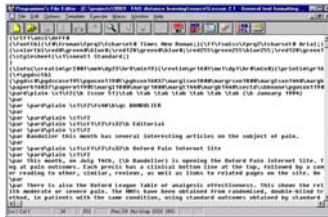*Click on the answer of your choice*

Exercise 3

Which of these images shows a document template?

○

○

Click on the answer of your choice

Exercise 4

Which of the images shows an RTF file opened in a plain text editor?

○

○

Click on the answer of your choice

Webopedia (www.webopedia.com). Online dictionary and search engine for computer and Internet technology. Try searching there for "word processor", "RTF", etc

Microsoft Word (www.microsoft.com) the most widely used word processor

Corel (www.corel.com) a leading supplier of word processing and office applications software, including WordPerfect.

Adobe (www.adobe.com) supplier of PDF applications software and desktop publishing packages

Open Office (www.openoffice.org) the leading open source (freely available) office application suite, including a word processor

StarOffice (www.sun.com) commercial version of Open Office from Sun Micorsystems

AbiWord (www.abisource.com) a free word processor, similar to MS Word