

Lenguajes de Marcas y Sistemas de Gestión de Información

Tema 1.- Características de los lenguajes de marcas

1	Introducción y conceptos previos	2
2	Mapa conceptual	3
3	Introducción a los lenguajes de marcas	3
4	Breve repaso histórico de los lenguajes de marcas	4
5	Clasificación y características comunes de los lenguajes de marcas	6
5.1	Características comunes	6
5.1.1	Texto plano (plain text)	6
5.1.2	Interoperabilidad o independencia del dispositivo final	7
5.1.3	Flexibles y fáciles de crear	7
5.1.4	Especialización	7
5.2	Clasificación de los lenguajes de marcas	8
6	Organizaciones y estándares	9
7	Ámbitos de aplicación de los lenguajes de marcas	10
8	XML: estructura y sintaxis	10
9	Herramientas de edición	12
9.1	Bloc de notas Windows	12
9.2	Gedit	12
9.3	Notepad++	13

1 Introducción y conceptos previos

Las personas utilizamos en nuestro lenguaje hablado ciertos gestos, diferente entonación, pausas, etc. que acompañan al contenido que queremos transmitir. De modo parecido en el lenguaje escrito utilizamos tamaños de letra, negrita, numeración, viñetas, tablas, colores, etc...que también acompañan a la información para que sea más fácil de entender. En definitiva estamos aplicando un determinado formato a la información que transmitimos.

En el caso de los documentos que intercambiamos a través de Internet, como las páginas web, son los lenguajes de marcas los que nos permiten aplicar dicho formato.

Un documento que contenga exclusivamente texto es perfectamente legible por nosotros, aunque evidentemente, tedioso e inapropiado para publicar páginas web. Si le aplicamos formato mediante un lenguaje de marcas como por ejemplo HTML, obtenemos un archivo también legible pero más difícil de interpretar. De esta tarea se encargará el navegador o mejor dicho el agente de usuario, quien interpreta las marcas de formato y las aplica convenientemente al texto para dar lugar a una página web, que será mucho más agradable de leer que el texto original.

Una “marca” es una señal colocada dentro de un texto, con el fin de delimitar una parte del mismo y en muchos casos, aplicarle un determinado formato (aunque existen marcas con otros propósitos).

2 Mapa conceptual

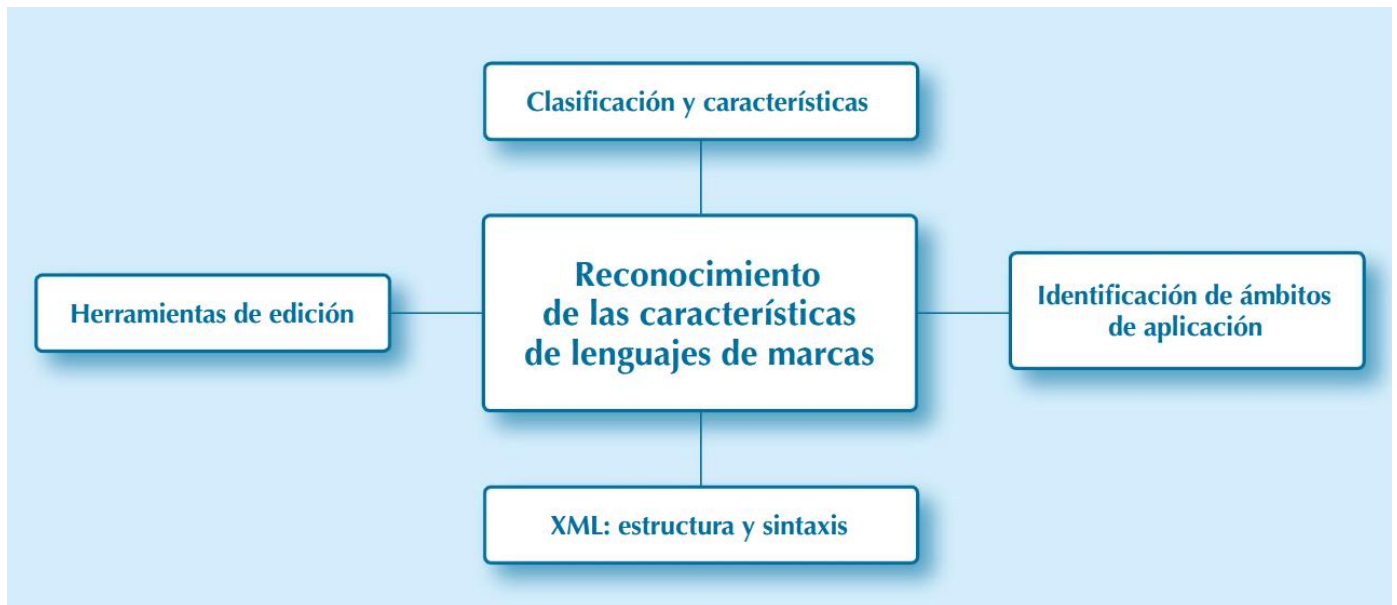


Figura 1 - Mapa conceptual del tema 1

3 Introducción a los lenguajes de marcas

El uso de internet por parte de los usuarios está creciendo en los últimos años. La gran mayoría de los internautas navegan e intercambian información a través de la red, manejando las páginas web que internamente hacen uso de los lenguajes de marcas y permitiendo que todo · uya con normalidad, proceso que es transparente para los usuarios.

Algunos de los programas instalados en el ordenador personal usan lenguajes de marcas, ya sea para almacenar información, configuración de la aplicación, etc.

Las marcas pueden representarse con distintos caracteres, como en TeX, que se usa la barra invertida para el inicio de una marca. Las más usadas son las que se expresan entre el signo menor que (<) y mayor que (>).

En los ejemplos siguientes, se aprecian dos etiquetas de marcas con valores contenidos en ellas: XML (extensible markup language) y HTML (hypertext markup language).

Etiqueta de marca XML

<nombre> con el valor “Contenido”: <nombre>Contenido</nombre>

Etiqueta de marca HTML de cabecera <h1> con el valor “Hola Mundo”:

<h1>Hola Mundo</h1>

No puede compararse un lenguaje de marcas con un lenguaje de programación como C#, Java o Pascal, pero sí permite combinar dentro de un documento lenguajes como PHP (hypertext preprocessor) o JavaScript, aportando funcionalidad, tal y como hace HTML.

4 Breve repaso histórico de los lenguajes de marcas

El origen de estos lenguajes radica en el lenguaje de marcado generalizado estándar SGML (Standard Generalized Markup Language), un estándar que permite definir lenguajes de marcado, considerándose un metalenguaje. Es descendiente de GML (Generalized Markup Language) de IBM. En el año 1986, el GML se convirtió en SGML (Standard Generalized Markup Language), lo que dio lugar a la norma ISO 8879.

Un documento SGML consta de 2 partes:

- El prólogo: contiene la estructura.
- La instancia de documento: contiene los datos.

Ejemplo 1: Los módulos del Ciclo ASIR

Vocabulario: asir, módulo, título, contenido, unidad.

Reglas: asir contiene varios módulos, un módulo tiene un elemento simple título y un elemento contenido, contenido tiene varias unidades y toda unidad debe estar en un contenido, las unidades son texto simple, detrás de cada unidad solo puede ir otra unidad o fin de contenido, detrás de un módulo solo puede ir otro módulo o fin de asir.

```
<asir>
  <módulo><título>Lenguajes de Marcas</título></módulo>
    <contenido>
      <unidad>Introducción</unidad>
      <unidad>HTML</unidad>
      <unidad>CSS</unidad>
      <unidad>XML</unidad>
      <unidad>XSLT</unidad>
      <unidad>Sindicación</unidad>
      <unidad>ERP</unidad>
    </contenido>
  </módulo>
  ...
</asir>
```

Figura 2 - Ejemplo de código SGML

Es importante destacar que SGML no es estrictamente un lenguaje sino un metalenguaje, es decir, un conjunto de normas que permiten crear otros lenguajes de marcas.

A principios de los años noventa, Tim Berners-Lee desarrolla HTML, que permite visualizar la información de las páginas web en los distintos navegadores. Dos de los lenguajes de marcas que se verán en próximos capítulos son HTML y XML.

Una herramienta fundamental que se usa para visualizar documentos XML, HTML, XHTML, etc., es el navegador web o browser en inglés. Son aplicaciones instaladas en distintos dispositivos cuya funcionalidad es dar acceso a la web, mostrando a los usuarios la interpretación de los lenguajes de marcas, incorporando multitud de nuevas funciones.

Hoy día, se usan para consultar el correo electrónico y las páginas web, realizar búsquedas, acceder a la configuración de un router, etc...

En el siguiente ejemplo se observa la estructura de un documento HTML que, al abrirlo en el navegador, muestra el mensaje “Hola Mundo”. Dicho documento puede crearse con el Bloc de notas y ha de tener la extensión .html

```
<html>
<head>
  <title>Página Hola Mundo</title>
</head>
<body>
  <p>
    <b>Hola Mundo</b>
  </p>
</body>
</html>
```

Figura 3 - Código HTML

La ejecución del ejemplo da como resultado una página con el texto “Hola Mundo”, donde se aprecia que ninguna de las marcas que se han creado en el documento se muestra en el navegador. Para poder visualizar el contenido, simplemente hay que abrir el documento HTML con un navegador que se tenga instalado en el ordenador y, automáticamente, se visualizará el resultado.

5 Clasificación y características comunes de los lenguajes de marcas

5.1 Características comunes

No todos los lenguajes de marcas comparten las mismas características debido a la gran cantidad que puede encontrarse en el mercado actual. En este apartado, se tratarán las características comunes más importantes de estos lenguajes como puede ser el uso de texto plano, flexibilidad, etc...

5.1.1 Texto plano (plain text)

Este tipo de documentos, también llamado *texto sin formato* o *texto simple*, no permite

almacenar información con formato (color, tipo de letra, negrita, tamaño, subrayados, etc...) tal y como puede hacerlo un procesador de textos avanzado. Están formados exclusivamente por caracteres (letras, números, caracteres especiales y de control).

Uno de los editores que puede encontrarse en Windows para crear un documento de texto plano es el Bloc de notas (Notepad), cuya extensión es .txt, mientras que, si se usa Linux, tenemos Vi o Gedit.

Tal y como se ha mencionado anteriormente, al estar formado por caracteres, se emplea una codificación específica que suele ser ASCII, ISO 8859-1, Unicode etc...

Estos estándares de codificación permiten representar los distintos caracteres en todos los idiomas.

5.1.2 Interoperabilidad o independencia del dispositivo final

Se considera el texto plano como formato universal, ya que puede abrirse y editarse desde cualquier máquina, aunque ha de tenerse en cuenta la codificación empleada. Son independientes a la plataforma usada y a cualquier sistema de representación.

De esta forma, el mismo documento puede ser interpretado de diferentes formas dependiendo del dispositivo final. Así tendremos diferentes resultados si se usa un dispositivo móvil, un ordenador de sobremesa, etc...

5.1.3 Flexibles y fáciles de crear

Simplemente, se necesita un editor de texto para poder crearlos y guardarlos en la extensión que se desee. Algunos de ellos permiten combinarse con otro lenguaje para darle mayor funcionalidad.

5.1.4 Especialización

Inicialmente los lenguajes de marcas se idearon para visualizar documentos de texto, pero progresivamente se han empezado a utilizar en muchas otras áreas como gráficos vectoriales, notación científica, GPS, etc...

5.2 Clasificación de los lenguajes de marcas

El uso de los lenguajes de marcas es muy diverso. Se utilizan para mensajería instantánea XMPP, servicios web WSDL, páginas web HTML, sindicación de contenidos RSS, gráficos vectoriales, documentación electrónica RTF, coordenadas GPS en formato GPX, etc...

Los lenguajes de marcas suelen clasificarse en tres tipos:

- Tipo 1. De procedimiento. Suele emplearse para la presentación del texto, siendo visible para el usuario. Pueden usarse etiquetas para poner un título centrado, reducir el tamaño de letra poner el texto cursivo, etc... Ejemplos de este tipo de marcado son TeX, LaTeX.
- Tipo 2. De presentación. Se refiere al que define el formato del texto. Aunque fácil de crear, resulta complicado de mantener o modificar. Las etiquetas de marcado, por lo general, no son visibles a los usuarios. Uno de sus usos puede ser la maquetación de documentos para que el lector pueda leerlos correctamente. Por ejemplo: Microsoft Word y las marcas o caracteres ocultos.
- Tipo 3. Descriptivos o semánticos. Es un marcado flexible que usa etiquetas sin especificar la manera de representarlas ni el orden. Las marcas dan información sobre su estructura y una descripción del contenido. Algunos de estos lenguajes son SGML, HTML y XML.

En el ejemplo se observa una serie de etiquetas creadas para un documento XML que almacena información relacionada con los exámenes de distintas materias de un ciclo. En este caso concreto, se han creado dos materias de las que se almacenan el nombre, el número de alumnos aprobados y suspensos y las unidades de cada una.


```
<?xml version="1.0" encoding="iso-8859-1"?>
<exámenes>
  <materia>
    <nombre>Lenguaje de marcas</nombre>
    <aprobados>15</aprobados>
    <suspensos>10</suspensos>
    <unidades>4</unidades>
  </materia>
  <materia>
    <nombre>sistemas operativos</nombre>
    <aprobados>5</aprobados>
    <suspensos>20</suspensos>
    <unidades>5</unidades>
  </materia>
</exámenes>
```

Figura 4 - Ejemplo de código XML

6 Organizaciones y estándares

Podemos definir la normalización o estandarización como el proceso de especificación de normas, para garantizar el correcto funcionamiento de elementos contruidos de forma independiente.

Aplicado al contexto de los lenguajes de marcas, sería por ejemplo el desarrollo de páginas web atendiendo a las especificaciones del lenguaje utilizado.

Para la definición de estas normas existen organismos internacionales, nacionales e incluso organizaciones privadas,

Las organizaciones más importantes en materia de software son W3C (www.w3.org) , ISO (www.iso.org) ...

Toda organización de estándares pretende desarrollar normas que sean de amplio seguimiento

por parte de la comunidad, para lo cual es imprescindible el consenso con las empresas involucradas como los navegadores, buscadores, desarrolladores web y fabricantes de dispositivos.

Entre sus miembros se encuentran las principales empresas del sector como Microsoft, Apple o Google entre otras.

Su web es

7 Ámbitos de aplicación de los lenguajes de marcas

El ámbito de aplicación de los lenguajes de marcas es muy diverso, ya que permite el intercambio de datos entre distintas aplicaciones, independientemente de la plataforma usada y la tecnología en la que estén creadas. Desde un fichero XML, pueden generarse vistas como HTML, WML o PDF. Java EE usa ficheros XML para poder especificar datos de configuración. Visual Studio, a la hora de crear servicios web, genera varios documentos con estructura XML. Tanto Windows Phone como Android Studio usan esta estructura para guardar las vistas. En general, existe una gran cantidad de software que usa datos en formato XML para configuración o guardar información. En el mundo real, puede usarse para bases de datos, frameworks de desarrollo, sistemas de publicación de contenidos, definición de interfaces gráficas, etc.

8 XML: estructura y sintaxis

XML, sigla de *eXtensible Markup Language* (lenguaje de marcas extensibles), se considera un metalenguaje y surge para resolver los problemas que plantea HTML, al mostrar las etiquetas el significado de sus datos. Una de las ventajas es que no requieren conocimientos de programación para poder crear o modificar un documento sencillo. La importancia radica en el intercambio de información de manera segura entre distintos programas, lo que permite la reutilización de contenido, crear etiquetas propias y representar una estructura y diseño independientes.

Al ser XML un metalenguaje, sirve de base para crear otros lenguajes.

Aunque este apartado se estudiará con más profundidad más adelante, se mostrará una introducción a su estructura y sintaxis para ir asimilando la terminología empleada.

Todo documento XML tiene estructura jerárquica arborescente y está compuesto por dos partes fundamentales: *prólogo*, que será la primera parte del documento y estará compuesto por una o varias líneas, y el *cuerpo*.

En este ejemplo el prólogo estaría compuesto por la primera línea donde se declara la versión de XML, así como la codificación empleada para representar los caracteres, mientras que el resto del documento, el cuerpo, estaría compuesto por elementos anidados. Dentro de la etiqueta <cine>, se incluyen las distintas películas de la cartelera. En este caso, se ha introducido una película y se guarda información relativa al director, la duración y el título.

```
<?xml version="1.0" encoding="iso-8859-1"?>
<cine>
  <película>
    <director>Señor X</director>
    <duración>60</duración>
    <título> Las aventuras de los SGML</título>
  </película>
</cine>
```

Figura 5 - Código XML

Las etiquetas van encerradas entre el signo menor que (<) y mayor que (>) y suelen ir emparejadas, una de apertura y otra de cierre. También existen etiquetas vacías, tal y como se aprecia en el ejemplo siguiente donde existe en la primera línea una etiqueta de apertura y cierre para el elemento email y un etiqueta vacía para el elemento casado con un atributo v con valor "no". Los atributos añaden propiedades a los elementos y van dentro de la etiqueta de apertura.

```
<email>666777888@prueba.com</email>  
<casado v="no" />
```

Figura 6 - Código XML

9 Herramientas de edición

En este apartado, se tratan las herramientas básicas relacionadas con la edición de páginas web, así como otras para la creación de documentos XML, estudiando algunas de mayores prestaciones en capítulos posteriores. AL igual que para otro tipo de software, pueden encontrarse tanto herramientas *offline*, que tienen que descargarse e instalarse en el ordenador personal para poder usarlas, y *online*, donde el único requisito es disponer de una conexión a Internet y el uso de un navegador para acceder.

9.1 Bloc de notas Windows

El Bloc de notas (Notepad) es un editor de texto que viene con los sistemas operativos Windows que permite crear documentos de texto plano. La extensión de dichos documentos es .txt y permite realizar la codificación de caracteres en ANSI, Unicode, UTF-8,...

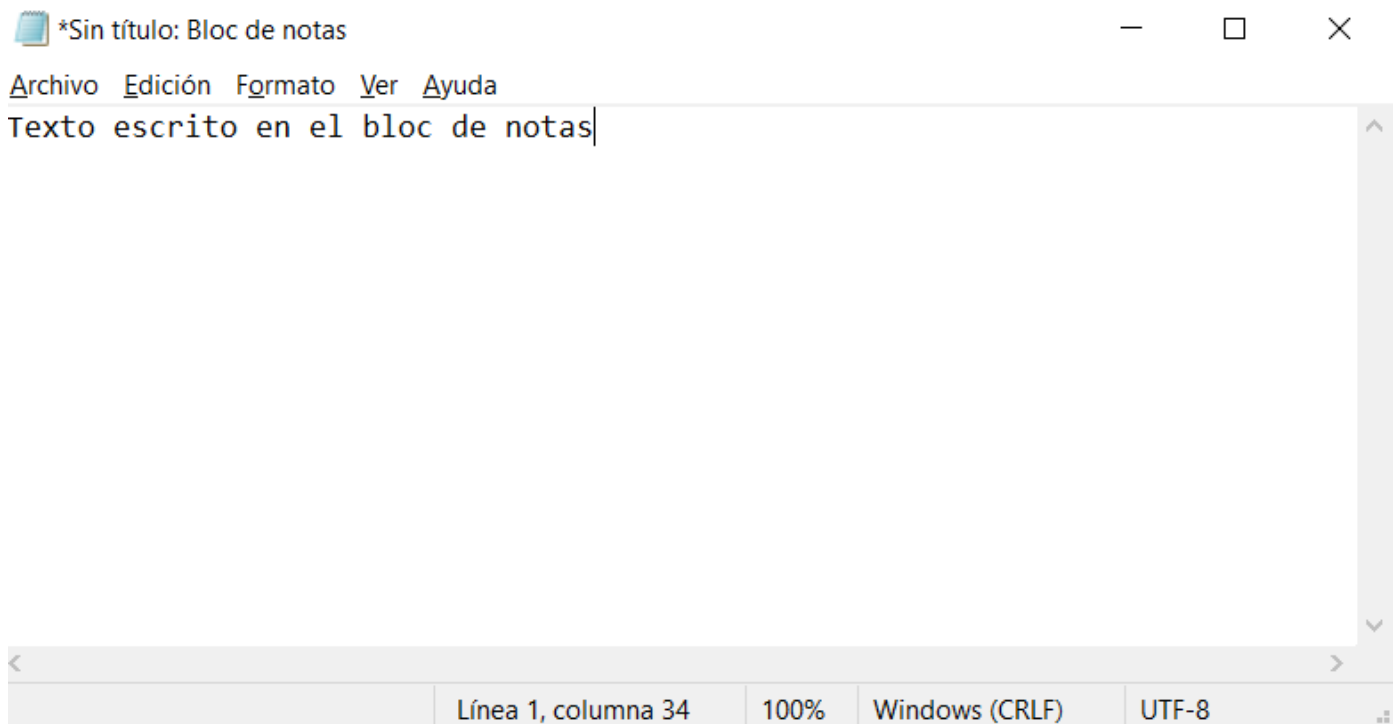


Figura 7 - Bloc de notas

9.2 Gedit

Es un editor de texto de fácil uso para el entorno de escritorio GNOME, aunque también está disponible para Windows.

Suele emplearse como simple bloc de notas, utilizar las características avanzadas y usarlo como entorno de desarrollo.

Se puede encontrar en <https://wiki.gnome.org/Apps/Gedit>

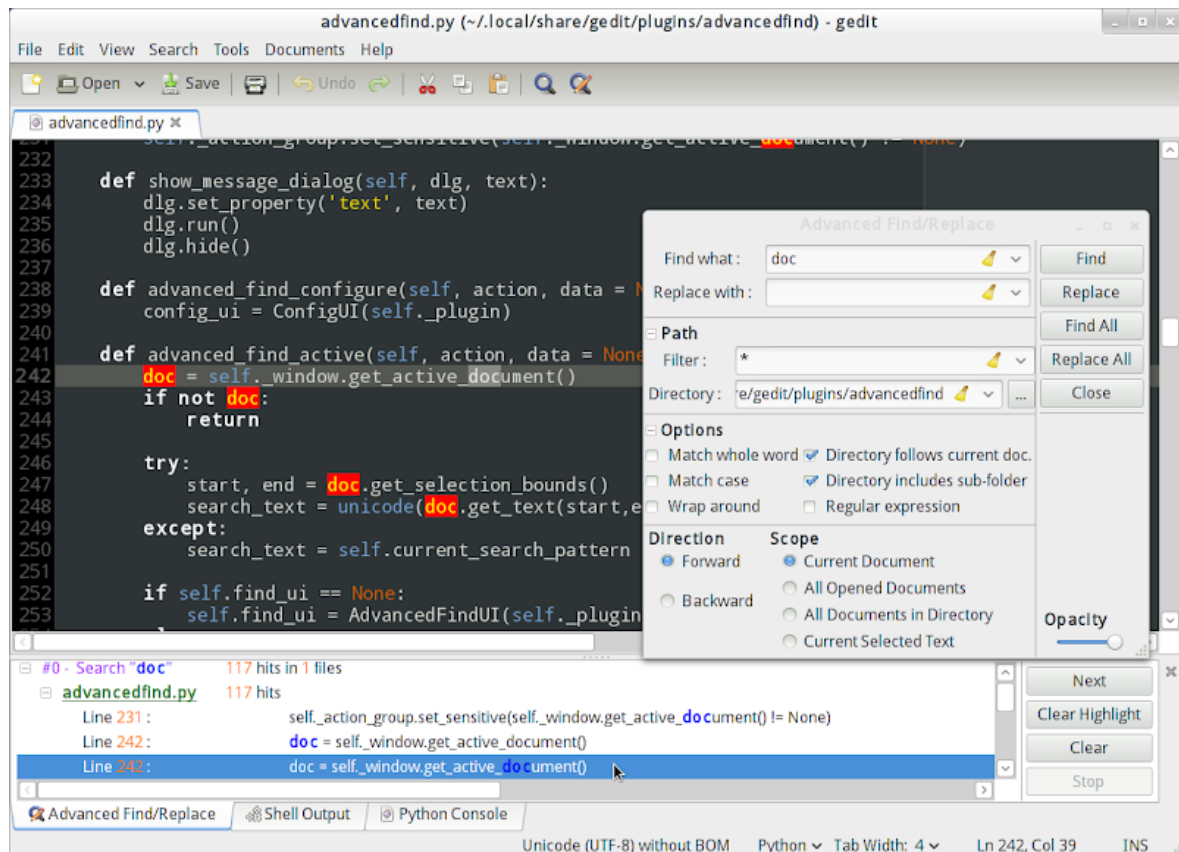
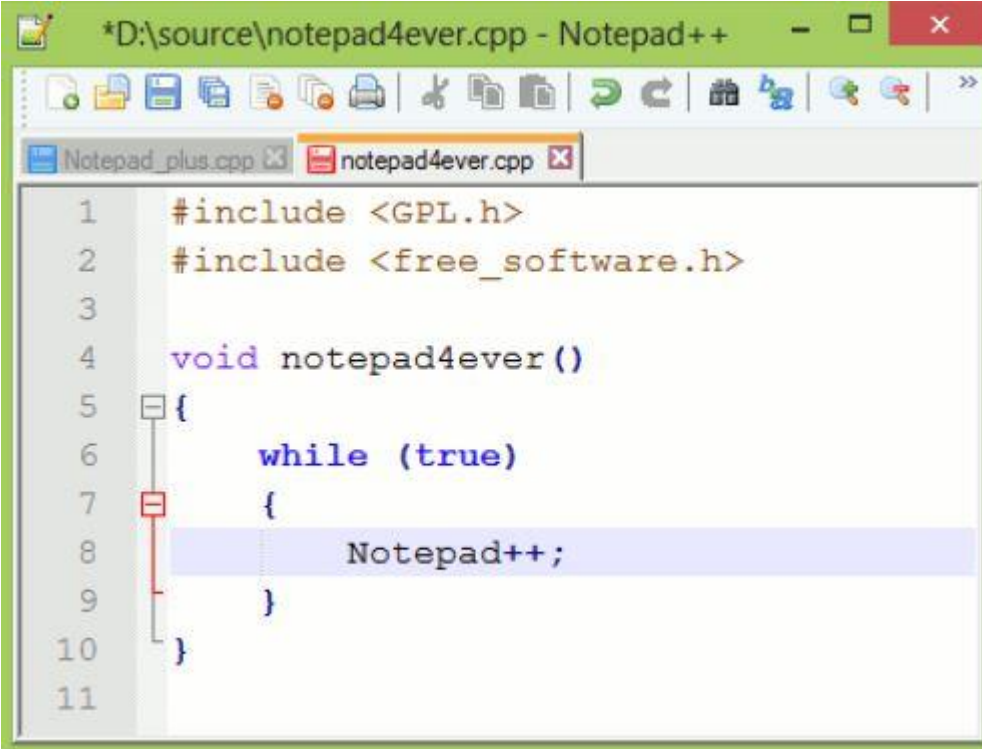


Figura 8 - Editor Gedit

9.3 Notepad++

Es un potente editor de texto gratuito con licencia GPL para diversos sistemas operativos. Permite seleccionar multitud de lenguajes de programación, como C, Java o HTML. Permite ampliar y reducir texto, autocompletar, búsqueda y reemplazo y resaltado de sintaxis.

Se puede descargar en <https://notepad-plus-plus.org>



```
1  #include <GPL.h>
2  #include <free_software.h>
3
4  void notepad4ever()
5  {
6      while (true)
7      {
8          Notepad++;
9      }
10 }
11
```

Figura 9 - Notepad++