

INSTITUTO TECNOLÓGICO DE COSTA RICA

Analizador Léxico

Del lenguaje XHTML, realizado en Flex

Andres Ramirez 201013880
Larissa Rivas Carvajal 201022184

08/04/2013

Contenido

Contenido	2
Descripción del programa	3
Diseño de programa	3
Decisión de diseño:	3
Lista de tokens a usar:	4
Razón de escogencia de FLEX:	4
Librerías externas a utiliza	5
Conclusión personal	5
Referencias	6

Descripción del programa

El programa consiste en crear un analizador léxico del lenguaje de xhtml en Flex que identifique cada uno de sus componentes dividiendo en token. Para ello se deberá de utilizar las expresiones regulares que asocien acciones para que luego el analizador sintáctico lo pueda realizar en su función con base en lo que este compilador léxico le envíe.

Este debe de tener una salida donde indique si reconoció cada carácter del texto de entrada. Además deberá de identificar si se ha introducido un carácter no valido para el lenguaje de xhtml, de ser así se debería de indicarlo por el stderr. El programa debe de recibir la entrada de la información por el stdin, y no hacer ningún manejo de archivos adicional, lo que se imprimirá será un archivo con cada token debidamente identificado por el stout.

Diseño de programa

Decisión de diseño:

Por la complejidad que representa el lenguaje de xhtml, que depende del contexto para la identificación de sus componentes, decidimos realizar el programa bajo ese concepto, que cada elemento trabaja bajo su propio contexto, y por tal motivo decidimos utilizar las condiciones iniciales (start Conditionals) para creación de mini escáneres que permitan analizar cada elemento dependiendo de su localización dentro del documento. Además de definir una lista de tags, atributos y eventos que son posibles dentro de este programa.

Lista de tokens a usar:

La lista de tokens se detalla a continuación:

S: espacio en blanco

EQ: signo igual con cualquier cantidad de espacios antes o después.

VN: numero de version.

Name: token que empieza con letra, raya abajo, o dos puntos seguido de caracteres.

CONTENIDO_ATRIBUTO: todo aquello que se encuentre entre comillas después de un atributo y un símbolo de igual.

SCRIPT: función con () en su contenido.

PubldLit: identificador de archivo, propiedad de SYSTEM o PUBLIC.

ENCODNAME: nombre de codificación del archivo.

Contenido cDATA: todo carácter aceptado dentro de los cDATA.

ATRIBUTO: atributos que pertenecen a a los elementos del TAG. Lista definida en la sección de definiciones.

TAG: nombre de elemento de xhtml que cumple con una serie de requisitos. Lista definida en la sección de definiciones.

Eventos: eventos que pertenecen a los elementos del TAG. Lista definida en la sección de definiciones.

COMMENT: Sección que ignora los caracteres dentro de esta sección.

?xml: inicia una condición inicial para que se evalúen sus componentes.

"<!DOCTYPE": inicia una condición inicial para que se evalúen sus componentes.

"</": inicia un tag final, este inicia una condición inicial.

"<": inicia el mini scanner de tag, ya sea un tag inicial o un tag corto.

<CONTENIDO>: es iniciada únicamente por los tags, y es el contenido que ese encuentra dentro de un TAG inicial y TAG Final.

Razón de escogencia de FLEX:

Durante el tiempo de carrera hemos estado en contacto con mucho lenguaje orientado a objetos, por lo que hemos optado por aprender mas acerca de C. A parte de esto, logramos encontrar información suficiente para animarnos a probar algo nuevo, puesto que anteriormente hemos trabajado en java. Por otra parte, en clase se ha venido trabajando sobre este programa, por lo cual nos pareció una buena opción ya que se ha visto materia introductoria sobre el software el cual nos ha parecido bueno para la realización del proyecto. Incluso, consideramos el hecho de que el asistente como así otros compañeros que han llevado este curso, lo cual sería un recurso útil a consultar o pedir asesoría.

Librerías externas a utiliza

No se utilizaron librerías externas para la realización de este programa.

Conclusión personal

Este proyecto nos ha ayudado a entender como un compilador realiza su primer etapa de análisis, más específicamente el análisis léxico el cual es la parte donde se analizan los token de un archivo. A partir de este proyecto podemos concluir:

- Para la realización de un analizador léxico, se debe de echar mano de expresiones regulares que permitan definir patrones que generen tokens validos por el lenguaje.
- Flex es una gran herramienta para la generación de analizadores léxicos gracias a su facilidad y sencillez a la hora de programar. Además que posee una gran integración con el lenguaje de programación C que permite realizar programas más robustos con inclusión de librerías y y creación funciones.
- La implementación de un analizador léxico para xhtml es más que definir unas pocas palabras reservadas o distintos patrones, ya que muchos de ellos dependen del contexto con el que se esta trabajando.

No cabe duda, que la implementación de este proyecto, reflejara un incremento en el conocimiento del funcionamiento del procesador, quien debe de ser construido con reglas que implementen la mejor forma de resolución tomando en cuenta la importancia o peso de los elementos que componen el lenguaje.

Referencias

1. E. ,. K. MIT, «he World Wide Web Consortium,» 2004. [En línea]. Available: <http://www.w3.org/>. [Último acceso: 03 2013].
2. T. R. o. t. U. o. California., «The Flex Project.,» 2001-2007,2012. [En línea]. Available: <http://flex.sourceforge.net/manual/>. [Último acceso: 04 2013].