Analizador Léxico

FLEX

Preparado por

Alexis Boza

Silvia Delgado

Instituto Tecnológico de Costa Rica

01/04/2013

# Descripción del Problema

Para el curso de “Compiladores e Intérpretes” de la Escuela de Computación del Instituto Tecnológico de Costa Rica es necesario implementar un pequeño analizador léxico o scanner para el lenguaje de etiquetas XHTML, en cual se deberán de tomar en cuenta las diferentes etiquetas, palabras reservadas, palabras o comandos no reconocidos por el lenguaje, etc., esto con el fin de crear los tokens que serán necesarios para un futuro analizador sintáctico o parser.

Además de generar los tokens necesarios el analizador deberá retornar al STDERR aquellas palabras que lexicalmente no se encuentran declaradas dentro de la especificación de XHTML. Según especificaciones brindadas con anterioridad el scanner deberá funcionar solamente utilizando argumentos de línea de comandos.

El control del proyecto deberá ser llevado haciendo uso de la herramienta GITHUB, con el fin de introducir el conocimiento y el uso de sistemas de control de versiones.

# Diseño del Programa

## FLEX

En la etapa de planificación del proyecto decidimos utilizar Flex debido a que decidimos como grupo que existía más documentación y material de ayuda para esta herramienta, además de que hace uso del lenguaje C/C++, en cual creemos necesitábamos más practica a diferencia de Jflex que hace uso de Java.

## Tokens a utilizar

### Palabras Reservadas

* base
* body
* head
* html
* meta
* title
* abbr
* acronym
* address
* b
* bdo
* big
* blockquote
* br
* cite
* code
* del
* dfn
* div
* em
* h1
* h2
* h3
* h4
* h5
* h6
* hr
* i
* ins
* kbd
* p
* pre
* q
* samp
* small
* span
* strong
* sub
* sup
* tt
* var
* dd
* dl
* dt
* li
* ol
* ul
* a
* link
* area
* img
* map
* object
* param
* caption
* col
* colgroup
* table
* tbody
* td
* tfoot
* th
* thead
* tr
* button
* fieldset
* form
* input
* label
* legend
* optgroup
* option
* select
* legend
* noscript

### Atributos Css

* href
* class
* id
* dir
* style
* xml:lang
* profile
* xmlns
* content
* http-equiv
* name
* scheme
* datetime
* accesskey
* charset
* hreflang
* rel
* rev
* tabindex
* type
* media
* alt
* coords
* shape
* nohref:nohref
* height
* longdesc
* src
* usemap
* width
* classic
* codebase
* codetype
* data
* declare:declare
* archive
* standby
* value
* align
* char
* charoff
* valign
* borders
* frame
* rules
* summary
* axis
* colspan
* headers
* rowspan
* scope
* disable:disable
* accept
* charset
* action
* enctype
* method
* checked:checked
* maxlegth
* readonly:readonly
* size

### Caracteres de Estructura y Etiquetado XHTML

* <
* >
* <
* <--
* -->
* =
* ( )

### Caracteres no reconocidos por XHTML/HTML

* á
* à
* é
* è
* í
* ì
* ó
* ò
* ú
* ù
* @
* ñ/ñ

# Conclusión

Al realizar este proyecto y el analizador en general, dimos con varios puntos interesantes dentro del desarrollo de un scanner y del lenguaje XHTML:

* Lexicamente XHTML no posee errores (al menos no los encontramos), pero existen caracteres que son irreconocibles para este lenguaje.
* Flex como su nombre lo indica es una forma rápida de crear analizadores léxicos bastantes eficientes y complejos.
* La mayoría de errores durante el proceso de compilación serán capturados en las etapas de análisis sintáctico y semántico (XHTML).
* El uso de GitHub permite un mayor control dentro de la realización de un proyecto, en especial en aquellos relacionados con desarrollo de software.