

ToTalf: Tutorial online para Teoría de Autómatas y Lenguajes Formales *

Román Sosa, Francisco de Sande

Dpto. de Estadística, I.O. y Computación

Universidad de La Laguna

38271 La Laguna

e-mail: {alu1970,sande}@etsii.ull.es

Resumen

El Tutorial Online para Teoría de Autómatas y Lenguajes Formales (*ToTalf*) ha sido diseñado como una herramienta de apoyo a la realización de prácticas en la asignatura de *Teoría de Autómatas y Lenguajes Formales*. En los planes de estudios de la Universidad de La Laguna, esta asignatura aparece con una fuerte componente práctica (la mitad de sus créditos) de modo que las prácticas de laboratorio constituyen una parte esencial de la misma.

ToTalf permite a los alumnos experimentar con los conceptos estudiados en las clases de teoría, a través de la ejecución de distintos algoritmos que se aplican sobre los diferentes tipos de modelos de cómputo que se estudian en las clases teóricas. La herramienta ha sido desarrollada en Java, y está disponible tanto en forma de aplicación como de applet, de modo que los alumnos pueden instalarla localmente en sus ordenadores o bien utilizarla a través de cualquier navegador. Dispone de un sistema de ayuda que facilita el uso de la interfaz de usuario y para cada uno de los diferentes modelos de cómputo contemplados se ofrecen al usuario ejemplos predefinidos que se pueden cargar directamente y que también ilustran el modo de operación de la herramienta.

Uno de los objetivos esenciales en el diseño de la herramienta ha sido que ésta fuera una aplicación abierta de modo que en el futuro sea muy fácil incorporar nuevas funcionalidades,

ejemplos o completar y/o modificar el sistema de ayuda.

1. Introducción

El proyecto *ToTalf* [11] es un sistema de ayuda para el aprendizaje de conceptos básicos en la materia de autómatas y lenguajes formales. Concretamente, ha sido desarrollado para los alumnos que cursan la asignatura de *Teoría de Autómatas y Lenguajes Formales* (TALF) [5] en la Universidad de La Laguna, aunque algunas de sus capacidades son también aprovechables en la asignatura de *Compiladores*. El desarrollo del proyecto ha tenido una duración aproximada de seis meses y el primer prototipo ha sido implantado a lo largo del curso académico 2003-2004.

ToTalf aborda el estudio de la asignatura mediante un enfoque práctico, dividiendo el temario de la asignatura en cuatro módulos principales.

1. Autómatas finitos y expresiones regulares
2. Gramáticas independientes del contexto
3. Máquinas de Turing
4. Autómatas a pila

En cada uno de los módulos, el usuario puede introducir un ejemplo del concepto correspondiente, y aplicarle los distintos algoritmos que se encuentran implementados, y que se imparten en la asignatura. De esta forma, el alumno puede aplicar directamente los conocimientos y algoritmos estudiados en las clases de teoría, en vez de tener que resolver cualquier problema usando lápiz y papel, lo que en muchos

*El proyecto ToTalf ha sido financiado por el Vicerrectorado de Calidad Docente y Nuevos Estudios de la Universidad de La Laguna en la II Convocatoria de Proyectos de Innovación Docente

casos conlleva demasiado tiempo, es propenso a los errores, y sumamente tedioso.

En la actualidad es posible encontrar herramientas de características similares a las de *ToTalf*. Es el gran auge de Internet quien ha permitido la proliferación de programas de este tipo. Cada uno suele tener alguna propiedad interesante, pero la carencia de otras los hacen ser herramientas deficientes para nuestros propósitos. En general, la mayoría se centra en un único aspecto: la simulación de autómatas finitos. En el caso de las aplicaciones desarrolladas fuera de España, el hecho de que todos los menús así como el sistema de ayuda esté en inglés, desanima a algunos alumnos a utilizarlas. Revisemos a continuación las aplicaciones más relevantes:

- JFLAP [2] es un paquete de herramientas gráficas diseñado como asistente en la enseñanza de conceptos básicos en TALF. Se trata posiblemente de la herramienta más sofisticada y evolucionada de este tipo. Con respecto a *ToTalf*, la característica más notable es que dispone de una interfaz gráfica a través de la cual el usuario puede introducir autómatas así como visualizar gráficamente la salida de los diferentes algoritmos. La versión actual incorpora diferentes mejoras que han sido añadidas a las sucesivas versiones de este conocido paquete.
- ABCEZ (Applet Based Computability Enlightenment Zone) [3] es otra de las herramientas basadas en applets con mayor impacto visual. Igual que JFLAP también posee una cuidada interfaz gráfica, pero en este caso centra su atención exclusivamente en la simulación de autómatas finitos. Una de las características más destacables de este programa es su sistema de ayudas.
- El applet diseñado por J. Calera [4] para la asignatura *Lenguajes, Gramáticas y Autómatas* en la Universidad de Alicante es uno de los pocos ejemplos disponibles de herramientas similares a *ToTalf* en el contexto español. Se centra también en algoritmos orientados a autómatas finitos, y al igual que en nuestro trabajo, en lugar

de invertir esfuerzo en el diseño de una interfaz gráfica avanzada, se ha optado por implementar un amplio conjunto de algoritmos.

- También mencionamos [8], [9] y [10], trabajos sobre simulación en autómatas finitos, y que no aportan grandes diferencias respecto a los ya comentados en los puntos anteriores.

2. Contexto y Objetivos

En la *Escuela Técnica Superior de Ingeniería Informática (ETSII)* de la Universidad de La Laguna se imparten los estudios correspondientes a las Ingenierías Técnicas en Informática de Sistemas y de Gestión, así como los de Ingeniero en Informática.

La asignatura de TALF figura como troncal para las dos Ingenierías Técnicas, y tiene asignados un total de nueve créditos, repartidos en partes iguales entre teoría y prácticas. La asignatura se imparte a lo largo del primer cuatrimestre del segundo curso.

TALF es una asignatura con unos contenidos eminentemente abstractos que se han desarrollado históricamente a partir de los problemas concretos a los que se ha ido enfrentando la comunidad científica en su intento de utilizar lenguajes de alto nivel en la programación de ordenadores, de traducir automáticamente o de matematizar el lenguaje humano para procesarlo automáticamente.

La repercusión más directa de los contenidos de TALF, quizás se produce en las asignaturas del plan de estudios relacionadas con compiladores: *Compiladores* en las ingenierías técnicas, así como *Procesadores de Lenguajes* en la Ingeniería en Informática. Aparte de esta repercusión directa, los contenidos de TALF también son fundamentales en muchos otros campos y asignaturas estudiados en la carrera.

2.1. Objetivos generales de TALF

Con esta asignatura se pretende transmitir al alumno los conceptos fundamentales de la *Teoría de Autómatas y Lenguajes Formales*, pre-

sentándole una visión global de la misma y profundizando principalmente en los aspectos más aplicados como son las herramientas básicas que se utilizan en el diseño de analizadores léxicos y sintácticos a partir de expresiones regulares y gramáticas independientes del contexto, respectivamente. Los objetivos principales de la asignatura son:

- Presentar los distintos métodos formales de representación de lenguajes, estudiando métodos generativos (gramáticas) y métodos por aceptación (autómatas).
- Conocer las herramientas que se utilizan para describir, reconocer y caracterizar dichos lenguajes.
- Desarrollar la capacidad de abstracción y análisis teórico en relación con la teoría de lenguajes.

2.2. El programa de teoría

El temario de teoría impartido en la asignatura es el siguiente:

1. Introducción al lenguaje C++.
2. Conceptos básicos
3. Lenguajes regulares y autómatas finitos.
4. Gramáticas. Lenguajes independientes del contexto. Autómatas a pila.
5. Máquinas de Turing. Lenguajes recursivos y recursivamente enumerables.
6. Resolubilidad.

2.3. El programa de prácticas

Sin desdeñar los contenidos teóricos, indispensables para el desarrollo de cualquier técnica, consideramos que los estudios de Informática son eminentemente prácticos y es por ello que en las asignaturas que impartimos, las prácticas con ordenador constituyen un aspecto fundamental de las mismas como complemento insustituible a las clases de teoría. Semanalmente los alumnos presentan en una sesión de corrección de prácticas una de las prácticas cuyo enunciado se les ha propuesto con antelación. Los contenidos de las diferentes prácticas pueden variar de un año a otro, pero como es natural, siempre están íntimamente relacionadas con los temas que se estudian en teoría.

Los objetivos docentes a alcanzar a través de las prácticas de la asignatura los podemos resumir en:

1. Introducir a los estudiantes en un lenguaje de programación orientado a objetos, que es nuevo para ellos (durante el primer curso de la carrera los alumnos desarrollan sus prácticas usando Pascal).
2. La consolidación a través de las experiencias de laboratorio de los conceptos adquiridos en las clases teóricas.
3. Mejorar el dominio de la programación adquirido en el curso anterior en las asignaturas de *Metodología y Tecnología de la Programación I y II*.

2.4. Objetivos del Proyecto ToTalf

El proyecto *ToTalf* surge como una necesidad del profesor para mejorar la enseñanza de la asignatura. Para este fin, *ToTalf* implementa la totalidad de los algoritmos que se programan en las prácticas (hay que tener en cuenta que las prácticas de la asignatura cambian en cada uno de los cursos), y muchos otros algoritmos interesantes estudiados en las clases de teoría, incluyendo algunos correspondientes a la asignatura de *Compiladores*. Entre los objetivos que perseguía *ToTalf* en su planteamiento inicial, cabe destacar los siguientes:

- Como ya se comentó en la introducción, trata de ser una herramienta para ayudar al alumno en la comprensión de la teoría y los conceptos relacionados con la asignatura. El poder experimentar con dichos conceptos mediante la aplicación de algoritmos debe ser una actividad que facilite esa comprensión. Gracias a *ToTalf*, el alumno puede comprobar directamente si ha realizado de manera correcta un ejercicio teórico. Además, en el caso de problemas de mayor tamaño, el alumno invierte su tiempo en el análisis de los resultados de los algoritmos, en vez de utilizarlo en realizar trazas sobre papel.
- Ayuda en la realización de prácticas. Los alumnos pueden experimentar con los algoritmos antes de acometer las prácticas que deben realizar. Pero fundamen-

talmente, es un marco de comprobación. En general, los alumnos dedican mucho tiempo a la implementación de las prácticas, pero poco a la comprobación de los resultados. Se suelen conformar con comprobar su código con los ejemplos que se le han suministrado, lo cual a veces resulta insuficiente. Con *ToTalf* se pretende que el alumno tenga de una manera sencilla una forma de examinar los resultados que devuelve su programa.

- Así mismo, es útil en la realización de prácticas de cierta envergadura, en las que haya que calcular ciertos resultados intermedios, los cuales no son objetivo de estudio de la práctica.
- Puede ser también una herramienta interesante para aquellos alumnos curiosos que gusten de experimentar por su cuenta.

3. ToTalf: la herramienta

3.1. Implementación

ToTalf está programado en Java. La elección de Java como lenguaje de programación responde a la gran ventaja de su ejecución multiplataforma, que permite su uso en multitud de entornos, con la ventaja de poder ejecutarse de manera sencilla y directa en un navegador.

Concretamente, *ToTalf* está programado en Java 1.4, y se proporciona tanto en versión applet como en versión aplicación.

Para la ejecución del applet (es decir, la ejecución a través del navegador) es necesario un navegador con soporte para la máquina virtual Java de Sun [7]. Por lo tanto, se necesita disponer del paquete J2RE (entorno de ejecución de Java) o el paquete J2SDK (entorno de desarrollo de Java), disponibles ambos en la sección de descargas.

Sin embargo, para un uso cotidiano, se recomienda la versión aplicación, que no está afectada por las restricciones de seguridad impuestas a los applets. Con la versión aplicación podremos cargar y guardar archivos, y copiar y pegar entre aplicaciones y el applet, o viceversa. Para la ejecución de la aplicación, es ne-

cesario, al igual que antes, J2RE o J2SDK, y también descargar el fichero jar del proyecto.

Los requisitos principales de la aplicación a desarrollar fueron los siguientes:

- Disponer de un sistema de ayuda para que el usuario pudiera consultar cómo operar con la herramienta. La ayuda debe, además, poder modificarse con sencillez.
- Facilidad de introducción de nuevos algoritmos.
- Posibilidad de introducción de nuevos ejemplos predefinidos para cada tipo de concepto.
- Desarrollo en Linux.

Para cumplir el primer requisito, se optó por realizar la ayuda de *ToTalf* en formato *HTML*. Se trata de una solución estándar y ampliamente utilizada, ya que los ficheros de ayuda se puede crear y modificar fácilmente, más aún teniendo en cuenta la multitud de herramientas disponibles en este campo.

Se puso un especial énfasis en el diseño de los objetos que representan los diferentes tipos de conceptos (o modelos de cómputo) con el fin de permitir la escalabilidad del programa. Cada objeto tiene abundantes métodos para modificar y acceder a las estructuras del mismo, de tal forma que implementar un nuevo algoritmo sea una tarea sencilla.

Se contempla el añadido de nuevos ejemplos predefinidos mediante la redefinición de una tabla de conceptos predefinidos en el código. Cada ejemplo se codifica al igual que se hace en la interfaz de usuario, con lo que es fácil la modificación de dicha tabla.

El último de los requisitos del proyecto fue que se desarrollara sobre Linux, en un entorno de desarrollo libre, ya sea en modo texto o gráfico. A la vez, la herramienta o herramientas de desarrollo debían ser de fácil uso, y que permitieran la modificación del código fuera de él, y de forma sencilla. La herramienta de desarrollo JBuilder8 Personal [6] se ajusta a estos parámetros y está disponible gratuitamente para usos no comerciales.

3.2. Utilizando ToTalf

El tutorial contiene cuatro grandes módulos que abarcan los diferentes aspectos abordados

en la asignatura: autómatas finitos y expresiones regulares, gramáticas independientes del contexto, máquinas de Turing y autómatas a pila.

En cada módulo el usuario puede introducir una descripción textual del concepto correspondiente. Dicha descripción es igual a la que se utiliza en los enunciados de las prácticas, para facilitar el uso del programa a los alumnos. Cada descripción puede guardarse en fichero para su uso posterior, o también puede cargarse una descripción ya salvada. Se dispone también de la opción de cargar unos ejemplos predefinidos, los cuales son representativos del concepto en cuestión.

A la descripción introducida, se le pueden aplicar los algoritmos implementados, que se corresponden con la totalidad de los algoritmos programados en las prácticas de la asignatura más algunos otros interesantes estudiados en clase.

Describiremos a continuación cada uno de los módulos que componen *ToTalf*.

- Autómatas finitos y expresiones regulares. En este módulo, se contemplan autómatas finitos deterministas (AFD) y no deterministas (AFN), al igual que un subconjunto de las expresiones regulares extendidas. Se encuentran disponibles los siguientes algoritmos:
 - Transformación de AFN a AFD.
 - Eliminación de ϵ -transiciones.
 - Minimización de un AFD.
 - Obtención de un AFN a partir de una expresión regular (Construcción de Thompson).
 - Obtención de una gramática regular a partir de un AFD.
 - Obtención de los AFNs que reconocen $L(M_1) \cup L(M_2)$ y $L(M_1)L(M_2)$ a partir de los autómatas M_1 y M_2 .

En las expresiones regulares, se reconocen los operadores básicos (alternativas, concatenaciones, asociaciones y cierre), junto a varias características de las expresiones regulares extendidas: conjuntos de caracteres (por ej. `[abc]`, `[a-c]`), el cierre positivo, el operador de interrogación y algu-

nas clases de conjuntos de caracteres (por ej. `[:alpha:]` y `[:digit:]`).

- Gramáticas independientes del contexto (GIC). En este módulo, los algoritmos implementados son:
 - Cálculo de los conjuntos first y follow.
 - Algoritmo de Cocke-Younger-Kasami para el análisis sintáctico de una frase.
 - Cálculo de la forma Normal de Chomsky.
 - Simplificación de una GIC.
 - Eliminación de la recursividad por la izquierda de una GIC.
 - Transformación de una gramática regular lineal por la derecha a un AFN.
- Máquinas de Turing. En el módulo de máquinas de Turing, el único algoritmo implementado es la simulación de una máquina de Turing con una cinta infinita en ambos sentidos.
- Autómatas a pila deterministas. El único algoritmo disponible es el de la simulación de un autómata a pila determinista.

3.3. Descripciones

En cada uno de los módulos de *ToTalf*, la entrada a los diferentes algoritmos viene dada por una descripción del modelo de cómputo correspondiente.

```
// gramática que genera expresiones
// regulares simples
ER -> ER + ALT
ER -> ALT
ALT -> ALT PIEZA
ALT -> PIEZA
PIEZA -> ATOMO
PIEZA -> ATOMO *
ATOMO -> t
ATOMO -> ( ER )
```

Listado 1: Fichero de representación de una gramática

Las descripciones tratan de ser lo más general y flexible posible. Así, por ejemplo, el número de caracteres prohibidos intenta ser mínimo, se permite el uso de comentarios en las descripciones, o en el caso de las gramáticas, los

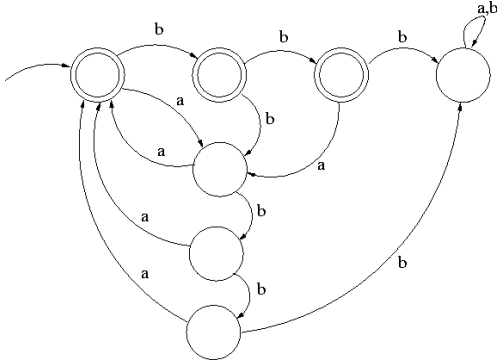


Figura 1: Autómata que reconoce cadenas con un n° par de 'a' y que no contengan la subcadena 'bbb'

símbolos pueden constar de más de un carácter.

Como ejemplo, el listado 1 muestra una descripción de una gramática que genera expresiones regulares.

3.4. Un ejemplo de uso con autómatas

Supongamos que el alumno desea resolver el siguiente problema: *Dado el alfabeto $\Sigma = \{a, b\}$ hallar un autómata finito con un número mínimo de estados que reconozca las cadenas de Σ^* que contengan un número par de símbolos 'a' y no contengan la subcadena 'bbb'.*

Cuando se les plantea este problema a los estudiantes, la mayoría obtienen directamente la solución, que se muestra en la figura 1.

Sin embargo muchos otros alumnos tratan de obtener la solución siguiendo un proceso alternativo, en el que *ToTalf* es una herramienta muy útil a la hora de simplificar algunos de los cálculos que han de realizarse.

Para resolver este problema, puede resultar cómodo partir de dos autómatas que reconozcan los siguientes lenguajes:

$$L_1 = \{w \in \Sigma^* | w \text{ contiene un número par de símbolos 'a'}\}$$

$$L_2 = \{w \in \Sigma^* | w \text{ no contiene la subcadena 'bbb'}\}$$

Los AFD que reconocen L_1 y L_2 son los que aparecen en la figura 2.

El autómata que se nos pide puede obtenerse teniendo en cuenta la siguiente relación

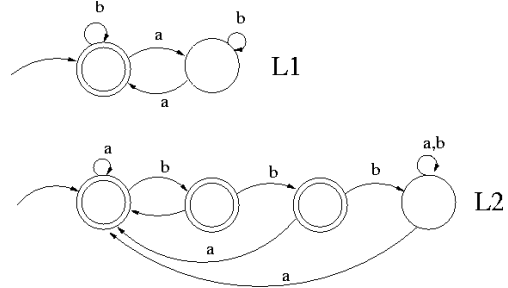


Figura 2: Autómatas que reconocen L_1 y L_2

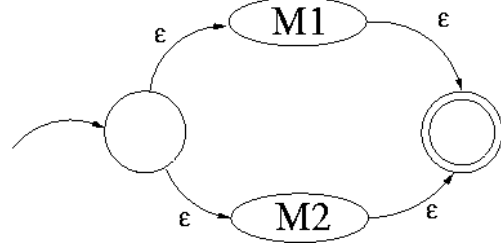


Figura 3: Autómata que reconoce $L(M_1) \cup L(M_2)$

entre lenguajes, consecuencia de la aplicación de una de las leyes de Morgan:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

Dado un AFD M que reconoce un lenguaje L , para obtener el AFD que reconoce \overline{L} basta con cambiar en M los estados que son de aceptación por estados que no sean de aceptación y convertir en estados de aceptación aquellos que en M no lo eran.

Por otra parte, partiendo de las descripciones de los autómatas M_1 y M_2 que reconocen $\overline{L_1}$ y $\overline{L_2}$ respectivamente, y utilizando uno de los algoritmos implementados en *ToTalf*, podemos obtener el AFN que reconoce $\overline{L_1} \cup \overline{L_2}$. La figura 3 muestra el fundamento teórico de esta construcción. Si aplicamos a este autómata el algoritmo de conversión de un AFN en AFD y al resultado le aplicamos la minimización del número de estados de un AFD, obtendremos el AFD que se muestra en la figura 1, que es la solución del problema planteado.

3.5. Un ejemplo de uso con gramáticas

Supongamos otro caso real de trabajo: construir un analizador sintáctico descendente re-

- Obtención del APND asociado con una gramática independiente del contexto (GIC).
- Cálculo de la tabla LL(1) para una GIC.
- Obtención de la GIC que genera el lenguaje reconocido por APND.
- Simulación de máquinas de Turing con varias cintas.

Pero sin duda, una de los aspectos más interesantes a incluir es una interfaz gráfica para la definición de los autómatas, en vez del actual esquema de texto. La creación de autómatas sería más intuitiva, y el análisis de los autómatas resultado de los algoritmos, mucho más sencillo, lo que conllevaría un mejor entendimiento por parte del alumno.

Tal como ha sido concebido, el prototipo actual de *ToTalf* no es más que el núcleo básico de una herramienta que tiene toda la potencialidad de proseguir su desarrollo. En el caso de herramientas similares, lo único que podemos hacer es recomendar a nuestros alumnos su utilización, mientras que *ToTalf* es un prototipo sobre el que tenemos control total. También estamos considerando la posibilidad de liberar el código fuente de la aplicación, haciéndolo de dominio público, de forma que toda la comunidad académica de habla hispana pueda contribuir al desarrollo de la herramienta.

En resumen, consideramos satisfactorio el esfuerzo invertido en este proyecto de calidad docente. Pensamos que *ToTalf* es en estos momentos una de las herramientas de estas características más completas desarrollada en España, y que se ajusta plenamente a los objetivos con los que fue concebida.

Referencias

- [1] John E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, London, 1979.
- [2] Susan H. Rodger. JFLAP 4.0b10 version. Disponible electrónicamente en <http://www.cs.duke.edu/~rodger/tools/jflap/index.html>.
- [3] MindSpring Software. ABCEZ (applet based computability enlightenment zone). Disponible electrónicamente en <http://www1.cs.columbia.edu/~zeph/ComputabilityApplets/abcez.html>.
- [4] J. Calera. *Applet Java para realizar diversas operaciones con autómatas finitos*. Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. Disponible electrónicamente en <http://www.dlsi.ua.es/assignaturas/lga/applet/Afapplet.html>.
- [5] Francisco de Sande. *Páginas web de Teoría de Autómatas y Lenguajes Formales*. Escuela Técnica Superior de Ingeniería Informática. <http://assignaturas.pcg.uill.es/etsii/talf/>.
- [6] Jbuilder downloads. Disponible electrónicamente en http://www.borland.com/products/downloads/download_jbuilder.html.
- [7] Java. Página principal: <http://java.sun.com>.
- [8] Nicolas Christin. DFAppllet a deterministic finite automata applet simulator. Disponible electrónicamente en <http://www.cs.virginia.edu/~nc2y/dfa/>.
- [9] Dynalab. Finite State Automaton Applet. Disponible electrónicamente en <http://www.cs.montana.edu/~dynalab/fsa/fsa.html>.
- [10] Thomas Dunn. Interactive animation of finite-state automata. Disponible electrónicamente en <http://www.ccs.neu.edu/home/tdunn/COM1350/honors/>.
- [11] Román Sosa González. ToTalf tutorial online de teoría de autómatas y lenguajes formales. Disponible electrónicamente en http://assignaturas.pcg.uill.es/etsii/talf/ToTALF/Applet_ToTalf.html.