# LENGUAJE DE PROGRAMACIÓN PARA GRAFOS COMPACTOS

# (Programming language for compact graphs)

# Simón Oroño1, Gerardo Pirela1

1 Laboratorio de Lenguajes y Modelos Computacionales, Departamento de Computación, Facultad Experimental de Ciencias, Universidad del Zulia, Maracaibo - Venezuela.

simonorono@protonmail.com

# Resumen

En la actualidad las redes son estudiadas como un medio para analizar complejas estructuras relacionales. Debido a esta tendencia, muchas herramientas se han desarrollado para realizar operaciones sobre redes, sin embargo, estas tienden a consumir de forma desmesurada recursos computacionales. En este trabajo se propone una solución a este problema: un lenguaje de programación que soporte en su sistema de tipos grafos almacenados de forma que se pueda reducir el uso de recursos computacionales requeridos para su manipulación.

**Palabras claves:** lenguaje de programación, grafos, compilador

# Abstract

Currently, networks are been studied as a mean to analyze complex relational structures. It is because of this tendency, that many tools have been developed in order to perform operations on networks; however, these tend to consume lots of computational resources. In this work, a solution to this problem in proposed: a programming language that supports in its type system graph stored in a way that will allow reducing the amount of computational resources required to handle them.

**Keywords:** programming language,graphs, compiler

# Introducción

El diccionario de la lengua española define una red como un conjunto de elementos organizados para un determinado fin. Dada esta definición general, las redes pueden modelar muchos tipos de relaciones y procesos en distintos campos del conocimiento, tales como biología, economía, epidemiología, sociología, entre otros. Según Barabási (2002), la construcción y estructura de las redes es la clave para entender el complejo mundo a nuestro alrededor.

De acuerdo con Mitchell (2009), hasta hace muy poco, el estudio de las redes no era visto como un campo en sí mismo. Epidemiólogos estudiaban la transmisión de enfermedades a través de redes de personas que interactúan, ecólogos estudiaban redes tróficas, sociólogos estudiaban la estructura de redes sociales humanas, economistas estudiaban el comportamiento de redes económicas. Todos estos campos trabajaban de forma independiente los unos de los otros, generalmente inconscientes de los avances de los demás.

Por su parte, las matemáticas y ciencias de la computación estudian las redes en un campo conocido como la teoría de grafos, que vio sus inicios en 1736, cuando el matemático Leonhard Euler demostró que era imposible hallar una solución para el problema de los siete puentes de Königsberg. Gracias a Euler y a la teoría de grafos, se puede asegurar que las redes y los grafos tienen propiedades que limitan o aumentan nuestras capacidades de hacer cosas con ellos (Barabási, 2002).

Durante las últimas décadas, un creciente grupo de matemáticos y físicos aplicados se han interesado en desarrollar un conjunto de principios unificadores que gobiernen las redes de cualquier tipo en la naturaleza, sociedad y tecnología (Mitchell, 2009), debido a esto, existe un creciente número de herramientas para manipular redes representadas como grafos: librerías para lenguajes de programación que añaden soporte para grafos, aplicaciones para la manipulación de grafos, lenguajes para recorrer y hacer consultas a bases de datos basadas en grafos, lenguajes para describir grafos, entre otras.

Sin embargo, las redes estudiadas por los distintos campos del conocimiento son, por lo general, muy extensas (por ejemplo: la Internet, redes neuronales humanas, redes y tramas tróficas, redes de transmisión de enfermedades, entre otras) y las herramientas actuales consumen muchos recursos computacionales para poder representarlas y manipularlas. Con una representación optimizada de grafos en memoria se podrían almacenar grafos muy grandes en muy poco espacio; los grafos representados de esta forma se denominan grafos compactos.

En la actualidad, no existen herramientas que manipulen de forma nativa y computacionalmente eficiente grafos compactos, es por ello que se propone con esta investigación la creación de un lenguaje de programación que dé soporte para grafos compactos de forma nativa en su sistema de tipos y en su librería estándar.

# Grafo

Un grafo consiste en un conjunto finito no vacío , cuyos elementos reciben el nombre de vértices o nodos y un conjunto de pares de elementos de , cuyos elementos se conocen como aristas (Rosen, 2012) y representan una conexión entre nodos. Si las aristas de un grafo no representan dirección, se dice que el grafo es no dirigido; si estás son aristas dirigidas (o arcos) se dice que el grafo es dirigido y los arcos se representan como pares ordenados de nodos que son elementos de (Rosen, 2012).

# Lista de adyacencia

La representación en forma de lista de adyacencia de un grafo consiste en un arreglo de listas, una por cada vértice en . Para cada , la lista de adyacencia contiene todos los vértices tal que existe una arista (Cormen y col., 2010). Esta es la representación idónea para grafos dispersos (dónde es mucho menor que ).

# Matriz de adyacencia

Para la representación en forma de matriz de adyacencia de un grafo , se asume que los vértices se encuentran numerados de una forma arbitraria ; de esta forma, la matriz de adyacencia consiste en una matriz de elementos, tal que para cada elemento perteneciente a se cumple que:

# Grado de un vértice

El grado de un vértice en un grafo no dirigido es el número de aristas incidentes con este, con la excepción de los bucles que se cuentan dos veces. (Rosen, 2012).

En un grafo dirigido se tienen aristas dirigidas, por lo que cada vértice cuenta con dos grados: el in-grado, que es el número de aristas dirigidas hacía , y el ex-grado, que es el número de aristas dirigidas desde (Rosen, 2012).

# Distribución de grados

Si se hiciera un recuento en un grafo del número de nodos por cada grado se tendría una distribución de grados del grafo, que sería entendido igualmente como la distribución de probabilidad de un grado en dicho grafo (Barabási, 2012).

La distribución de nodos de un grafo está definida como la fracción de nodos en el grafo con grado . Si existen nodos en un grafo y de ellos tienen grado , entonces .

Los grafos se han podido clasificar en función de la distribución de grados en diversas topologías, siendo las más conocidas:

* Topología exponencial: se da en redes evolucionistas y ocurre siempre que cada nuevo nodo que se añade posea la misma probabilidad de ser enlazado con el resto.
* Topología libre de escala: la distribución de grados de estas redes sigue una ley de potencias de la forma . En estas redes se cumple el hecho de que existan pocos nodos con alta conectividad, pero el grado de conexión de casi todos los nodos es bastante bajo.

# Compilador

Según Aho y col. (2006), un compilador es un programa que puede leer un programa escrito en un lenguaje (fuente) y traducirlo al programa equivalente en otro lenguaje (destino), reportando cualquier error que consiga durante este proceso.

# Estructura básica de un compilador

El programa fuente que será traducido por el compilador pasa por varias fases y cada una de estas transforma una representación del programa fuente a otra (Aho y col. 2006).

* Análisis léxico: es la primera fase del proceso de compilación, y en ella se lee el flujo de caracteres del programa fuente y se agrupan en secuencias significativas conocidas como lexemas (Aho y col., 2006). Por cada lexema, el analizador léxico genera un *token* de la forma:

Donde “nombre” esun símbolo abstracto que se utiliza durante el análisis sintáctico y “valor” apunta a una entrada en la tabla de símbolos para este *token*.

* Análisis sintáctico: es la segunda fase del compilador. El analizador sintáctico lee el flujo de *tokens* resultante de la primera fase e intenta crear una representación intermedia en forma de árbol para describir la estructura gramatical del programa fuente; este árbol, conocido como árbol sintáctico, será utilizado por las siguientes fases del proceso de compilación (Aho y col., 2006).
* Análisis semántico: utiliza el árbol generado por el analizador sintáctico y la tabla de símbolos para comprobar que el programa fuente es consistente con lo establecido en la definición del lenguaje (Aho y col., 2006).

Durante esta fase, se realiza el chequeo estático de tipos, donde se puede verificar que cada operador tenga operandos cuyos tipos coincidan, que el tipo de la expresión retornada por una función coincida con su tipo de retorno, que las funciones sean llamadas con el número correcto de argumentos y los tipos de estos coincidan con la declaración de la función, entre otras cosas.

* Generación de código intermedio: durante el proceso de compilación un compilador puede construir una o más representaciones intermedias del programa fuente (Aho y col., 2006). Los árboles sintácticos usados durante el análisis semántico son un ejemplo de estas representaciones.

Luego del análisis semántico, algunos compiladores generan una representación intermedia de bajo nivel —similar al código máquina—, que podemos considerar como un programa para una máquina abstracta y que debe tener dos propiedades importantes: debe ser fácil de producir y fácil de traducir en la máquina destino (Aho y col., 2006).

* Optimización de código: la fase de optimización de código independiente de la máquina trata de optimizar el código intermedio generado en la fase anterior en pro de producir mejor código destino (Aho y col., 2006). En este contexto, mejor código puede significarse código más rápido, más corto o un código con menor consumo de recursos.
* Generación de código: el generador de código recibe la representación intermedia del programa fuente como entrada y produce un programa equivalente en el lenguaje destino (Aho y col., 2006). Si el lenguaje destino es código máquina, durante esta fase se seleccionan los registros que utilizará cada variable del programa durante su ejecución.

# Propuesta

Se propone la creación de un lenguaje de programación con soporte nativo para grafos tanto en su sistema de tipos como su librería estándar. Se debe definir la estructura gramatical del lenguaje, al igual que su semántica.

El lenguaje sería lo más pequeño posible, de forma de hacer más sencillo al compilador y que la curva de aprendizaje sea poco inclinada, para este fin, la manipulación de grafos sería parte de la librería estándar en forma de funciones y no parte del lenguaje.

Toda forma de verificación de tipos se realizaría de forma estática por el compilador, reduciendo así el tamaño del entorno de ejecución y haciendo más rápidos a los programas escritos en este lenguaje.

# Bibliografía

Aho, A., Lam, M., Sethi, R. y Ullman, J. (2006): **Compilers: Principles, Techniques, and Tools.** Segunda Edición. Addison Weaslie. 1000pp.

Barabási, A. (2002): **Linked: The New Science of Networks.** Primera edición. Cambridge, Estados Unidos. Perseus Publishing. 280pp.

Cormen, T., Leiserson, C., Rivest, R., y Stein, C. (2010): **Introduction to Algorithms.** Tercera edición. The MIT Press. 1251pp

Mitchell, M. (2009): **Complexity: A Guided Tour.** Primera edición. Nueva York, Estados Unidos. Oxford University Press. 227-288.

Real Academia Española: **Diccionario de la lengua española.** [Disponible en] http://dle.rae.es/. Consultado enero 2016.

Rosen, K. (2012): **Discrete Mathematics and Its Applications.** Séptima edición. McGraw Hill. 1071pp.