

# Herramienta para la visualización y análisis de las normas de asociación libre

Trabajo Terminal No. 2016-1311

Alumno: Peña López Kevin Irvin

Directores: M. en C. José David Ortega Pacheco, Dra. Natalia Arias Trejo

**Resumen** – El presente Trabajo Terminal tiene como objetivo el desarrollo de una herramienta que permita la visualización de grafos, que permite analizar la relación existente entre palabras, a través de algoritmos de reconocimiento de patrones definidos en las normas lingüísticas, con el fin de proporcionar un reporte de los resultados obtenidos. Finalmente, una vez la herramienta se encuentre implementada se trabajará en un caso de estudio para ejemplificar su uso.

**Palabras clave** – Semántica, red, grafos, ingeniería de software, desarrollo web, reconocimiento de patrones, inteligencia artificial

## 1. Introducción

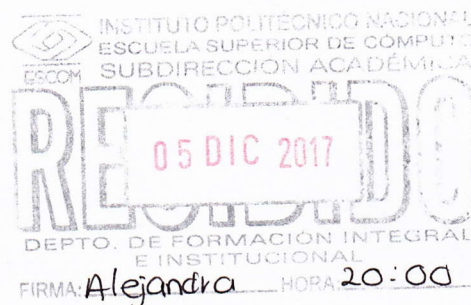
Las Normas de Asociación de Palabras (NAP) reflejan la relación existente entre dos palabras. En este tipo de estudios normativos la recolección de datos se hace a través de técnicas de asociación libre o asociación múltiple.

La técnica comúnmente empleada para la recolección de datos es la de asociación libre la cual consiste en pedirle a los sujetos que externen deliberadamente ya sea de manera verbal o escrita la respuesta que les produce el estar expuestos a ciertas palabras estímulo. La asociación de palabras ha sido de sumo interés para la psicolingüística como una herramienta para investigar los mecanismos que subyacen a la memoria semántica, ofreciendo a los investigadores una medida sobre la información semántica que normalmente se accede cuando una palabra se lee o escucha. Específicamente las NAP se han empleado en la psicolingüística para documentar las relaciones más frecuentes entre palabras (p. ej., perro-gato) por parte de hablantes de una misma lengua. Estas relaciones también son empleadas en investigaciones sobre memoria semántica y procesos de facilitación e inhibición lingüística (Nelson, Bennett & Leibert, 1997).

Además de los estudios de NAP, existen otros estudios normativos que autores como Proctor y Vu, (1999) han clasificado de acuerdo al tipo de variable y material que se estandariza y técnica empleada en 1) bases de índices psicolingüísticos, 2) asociación de palabras, 3) diccionarios de frecuencia léxica, 4) imaginabilidad, concretud, familiaridad y valencia afectiva, 5) homógrafos, homófonos y homónimos, 6) cadenas de letras y sílabas, 7) categorías conceptuales.

Algunos Artículos relacionados con *semantic prime* son:

Aplicación	ÁREA	APLICACIONES ESPECÍFICAS
Gephi	Análisis de la Lengua	Visualizador de palabras en 3D
Knowledge Structures	Análisis de la Lengua	Visualizador en 3D, y analizador de palabras en el idioma Alemán
Semantic Networks Analysis: Applications for Social Sciences	Análisis de la Lengua	Visualizador en 2D, para el conjunto de técnicas para estudiar las creencias, significados, identidades, el conocimiento, el discurso y el lenguaje en una amplia variedad de dominios sociales



Algunos Artículos relacionados con *semantic prime* son:

Artículo	Autores	Publicación	Objetivo
<b>Longitudinal Analysis of Early Semantic Networks</b>	Thomas T. Hills, Mounir Maouene, Josita Maouene, Adam Sheya, and Linda Smith	IEEE	Analizar las redes semánticas a través de mecanismo de involucramiento preferencial, para uso en el aprendizaje en las personas. Con el apoyo de grafos para identificar los puntos de inflexión de aprendizaje y la asociación de conceptos.
<b>The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth</b>	Mark Steyversa, Joshua B. Tenenbaum	IEEE	Describir el comportamiento de redes complejas, a través de parámetros y algoritmos de reconocimiento de patrones para la inclusión de nuevos nodos y su simulación, con el fin de aprender sobre la relación de conceptualización semántica abstracta, teniendo la posibilidad de formar un modelo matemático para su análisis.
<b>Latent Semantic Analysis Model as a Representation of Free-Association Word Norms</b>	Natalia Arias Trejo, Julia B. Barrón Martínez y David Ortega Pacheco	IEEE	Tiene como objetivo validar, por medio de un modelo computacional, una base de datos empíricos de las normas de asociación palabra- gratuitas de español de México. Para detectar el peso asociado de pares de palabras palabra-, y para proporcionar una comprensión de una red léxica.
<b>NORMAS DE ASOCIACIÓN DE PALABRAS EN UNA MUESTRA MEXICANA DE ADULTOS JÓVENES</b>	Julia B. Barrón Martínez	Tesis de la UNAM	Creación de Normas de Asociación de Palabras (NAP) para el español de México, documentar las relaciones más frecuentemente producidas por los hablantes de esta variante del español, que sirva como recurso para estudios en el área experimental.

Algunos Trabajos Terminales que se relacionan con el tema de *face recognition* son:

Trabajo Terminal	Desarrolladores	Resumen
<b>Herramienta para la obtención de bases de datos de normas de la asociación libre y análisis latente semántico</b>	Iván David Luqueño Jiménez	Obtención de información de las normas de asociación libre y aplicación de análisis semántico que permite obtener información mediante un proceso experimentación.



## 2. Objetivo

### General

Desarrollar una herramienta de análisis de normas de asociación libre del español en México, que permita una fácil visualización de una red semántica y con esto permita la aplicación de algoritmos estadísticos y de comportamiento en una red semántica para su análisis, además de concentrar la información en un reporte para su estudio.

### Específicos:

- visualización: tener una visualización en un grafo de las palabras del español en México, además de ver las conexiones que tienen con otras palabras

- métricas de análisis en grafos: obtener la información detallada de cómo se construye una red de palabras con sus respectivos datos

- reportes: Conjunto de resultados de las métricas de una red de palabras dada para su análisis

## 3. Justificación

En la última década en diferentes países como Alemania, Estados Unidos y Rusia, y centro de investigación de universidades tal como University of Base, Department of Psychological at Indiana University, CGES (Centre for German and European Studies), se han centrado en la investigación de la lingüística de su idioma así como las reglas de asociación libres de cada lenguaje, esto se debe que en cada país las reglas son diferentes, como lo son su semántica, la forma de relacionarse y la conceptualización de las palabras mediante su cultura, esto tiene una importancia cultural y de aprendizaje para la cada país ya que alguno estudios han favorecido para facilitar el aprendizaje de niños mediante técnicas experimentales de enseñanza. A largo plazo esto podría ser la base de una nueva pedagogía en la enseñanza así como tener definido las reglas de asociación de los idiomas más usados en el mundo, por ejemplo el español es la tercera lengua más usada a nivel mundial según la ONU, y la segunda con hablantes nativos, ya que en más de treinta países se habla el español.

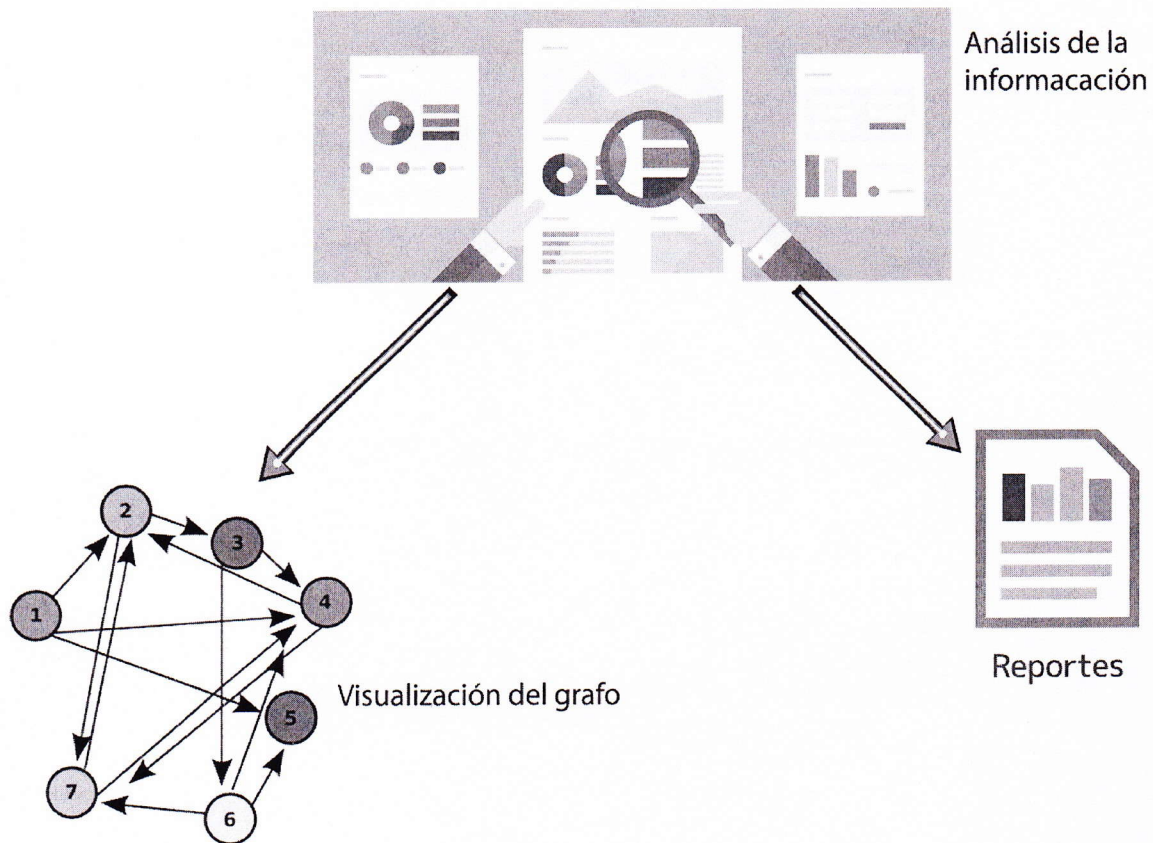
Aún así no existen las normas de asociación libres para este idioma. En México en la facultad de psicología de la UNAM, el departamento de psicolingüística tiene más de cinco años estudiando estas normas, por ejemplo existe una tesis que habla de estas normas para adultos jóvenes.

Con este proyecto se podrá ayudar a continuar los estudios de estas normativas, así como facilitar la extracción de datos entre las palabras que donde ya esté definida una asociación, esto aportará una forma de visualización que permita el análisis de más palabras de forma paralela a los investigadores, se espera que con este proyecto se pueda continuar el análisis de la semántica del idioma español de México, además de ayudar a los investigadores a crear un banco de información con un modelo matemático y estadísticas sobre la relación de las palabras, ayudando a su trabajo de investigación y de forma que tengan una herramienta desarrollada especialmente para el español, ya que hoy en día existen muy pocas y se concentran en otros idiomas donde se realiza un trabajo similar, en la industria no se enfoca a hacer el software que permita el avance de investigación.

De acuerdo a la anterior, se pretende ayudar a la obtención no presencial de datos experimentales, mediante un desarrollo web que permita visualizar las normas de asociación libres con apoyo de un grafos, además de resultados de del análisis a través de algoritmos estadísticos que aportarán información concentrada para una mejor conclusión en la investigación de las normativas de asociación en el idioma español de México.

#### 4. Productos o Resultados esperados

A continuación se muestra el diagrama inicial de la solución propuesta para la implementación de la herramienta:



Los módulos que conforman el sistema se especifican como sigue:

- Análisis de Información es un módulo que permitirá al usuario aplicar un tratamiento a la información pre-procesada y será una de las opciones de algoritmos estadísticos de los que ofrecerá la herramienta, así como métricas que se utilizan en la NAP, como son la media, el peso de conexión, el tipo de nodo, la probabilidad de palabra. Se analizará la opción de utilizar una biblioteca ya creada para cubrir la mayor cantidad de algoritmos posibles, o implementarlos de manera directa.
- El módulo de la visualización, permitirá ver las conexiones entre palabras de forma gráfica, para su análisis.
- Finalmente, el módulo de Reportes entrega al usuario un reporte del análisis completo, producto del análisis de información, con el cual se obtendrá una representación gráfica de los resultados que lo ameriten.

Respecto al caso práctico, de manera inicial y no siendo éste el motivo del trabajo terminal, se tiene la idea de trabajar con datos médicos relacionados con diabetes, pues el caso práctico sólo tiene como objetivo ilustrar el uso de la herramienta y no resolver dicho caso.



Como consecuencia del trabajo realizado se esperan obtener:

- Herramienta implementada.
- Documentación técnica.
- Manual de usuario y de instalación.

## 5. Metodología

Para poder llegar a la construcción final de un producto existe una gran variedad de modelos definidos por la ingeniería de software, los cuales se aplican dependiendo a las características del proyecto a desarrollar, así como cada uno optimiza el desarrollo del mismo dependiendo de su definición.

Las metodologías ágiles permiten aplicar modelos en los que se puede tener un cambio en los requerimientos que varía conforme se desarrolla el proyecto, uno de estos modelos es el modelo de desarrollo evolutivo (también denominado prototipado evolutivo) el cual se basa en la elaboración de una versión inicial del sistema, la cual es expuesta a comentarios del cliente y es refinada a lo largo de diferentes versiones hasta llegar al sistema final. Este modelo considera las actividades de especificación, desarrollo y validación, las cuales se relacionan para poder generar un prototipo a partir de especificaciones generales e ir modificando hasta llegar al sistema final como se muestra en la siguiente figura.



Tras analizar este método de trabajo se concluyó que es el adecuado para trabajar en el desarrollo del sistema, ya que permite la elaboración de prototipos que se pueden mejorar a través de pruebas y es posible agregar funcionalidad respecto al avance que se realice. Esto /representa un beneficio ya que permite/ trabajar en el sistema tanto práctica como teóricamente, de esta forma se podrán realizar pruebas que permitirán mejorar el prototipo que se tenga e ir agregando la funcionalidad necesaria.

A continuación se describen de manera general los prototipos definidos de manera inicial para la Herramienta de Minería de Datos:

**Prototipo 1.** Descripción del sistema, en este prototipo estará centrado en las consideraciones generales de arquitectura, análisis y diseño.

**Prototipo 2.** Versión inicial (análisis, implementación y pruebas) de las etapas.

**Prototipo 3.** Versión intermedia (análisis, implementación y pruebas) de las etapas.

**Prototipo 4.** Versión final (análisis, implementación y pruebas) de las etapas.

Se tiene contemplado que al inicio del trabajo terminal se realice un análisis de los prototipos propuestos de manera inicial, para determinar si es necesario realizar un ajuste a la presente propuesta.

## 6. Cronograma

Nombre del alumno(a): Peña López Kevin Irvin

TT No.:

Título del TT:

Actividad	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV
Investigación de Trabajo previo											
Prototipo 1											
*Prototipo 2											
Presentación de TT1											
Proto 3											
Proto 4*											
Presentación TT2											
Documentación											

## 7. Referencias

- [1] K. Kendall, Análisis y diseño de sistemas, Sexta ed., Pearson, 2005.
- [2] Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 53A, 741-764.
- [3] Ferrand, L., & New, B. (2003). Semantic and associative priming in the mental lexicon. In P. Bonin (Ed.), Mental Lexicon: Some words to talk about words (pp. 25-43). Hauppauge, NY: Nova Science Publisher.
- [4] Vivas, J. (2010). Modelos de memoria semántica. Evaluación de Redes Semánticas. Instrumentos y Aplicaciones. Capítulo del libro.
- [5] Macizo, P., Gómez-Ariza, C., & Bajo, M. T. (2000). Associative norms of 58 spanish for children for 8 to 13 years old Psicológica, 21, 287-300.
- [6] NORMAS DE ASOCIACIÓN DE PALABRAS EN UNA MUESTRA MEXICANA DE ADULTOS JÓVENES (2013) Tesis de JULIA BEATRIZ BARRÓN MARTÍNEZ, UNAM



## 8. Alumnos y Directores

*Peña López Kevin Irvin* .- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2012630339 , Tel. 5559064038 , email kevinpe.07@gmail.com

Firma: \_\_\_\_\_



*M. en C. José David Ortega Pacheco*.- Ingeniero en electrónica con especialidad en sistemas digitales y computadoras egresado de la UAM Azcapotzalco en 2006. Obtuvo el grado de Maestro en Ciencias de la Computación con especialidad en Inteligencia Artificial en el CIC-IPN en 2008. Actualmente se encuentra trabajando como profesor en la ESCOM del IPN. Áreas de interés: Procesamiento de Lenguaje Natural, Reconocimiento de Patrones, Cómputo Reconfigurable y áreas relacionadas. Teléfono: 57296000 Ext. 52032, correo electrónico: david82d@gmail.com.

Firma: \_\_\_\_\_



Natalia Arias Trejo PhD Experimental Psychology, University of Oxford, UKMA Applied Linguistics, University of Sussex, UKBA Lengua y Literaturas Hispánicas, Universidad Nacional Autónoma de México, México. Sus intereses se centran en el procesamiento lingüístico tanto en etapas tempranas del desarrollo como en la etapa adulta. Actualmente dirige dos proyectos de investigación en la Facultad de Psicología de la Universidad Nacional Autónoma de México (UNAM): CONACYT 167900 Mecanismos en la Formación y Modulación de Redes Semánticas durante la Infancia y la Etapa Adulta.

Firma: \_\_\_\_\_



CARÁCTER: Confidencial  
FUNDAMENTO LEGAL: Art. 3, fracc. II, Art. 18, fracc. II y  
Art. 21, lineamiento 32, fracc. XVII de la L.F.T.A.I.P.G.  
PARTES CONFIDENCIALES: No. de boleta y Teléfono.