

# ANÁLISIS DE RELACIONES SEMANTICAS DEL LEXICO DISPONIBLE EN MATEMATICAS EN UN HIPERMEDIO ADAPTATIVO

**Dr. Pedro Salcedo Lagos**  
Universidad de Concepción  
Chile  
psalcedo@udec.cl

**Dr. Oscar Nail Kroyer**  
Universidad de Concepción  
Chile  
onail@udec.cl

**(c) Mg. Carla Arzola Zapata**  
IP Virginio Gómez  
Chile  
carlaarzola@gmail.com

## ABSTRACT

The paper presents the module structure of semantic analysis of an Adaptive hypermedia via the Internet, which allows to describe and analyze the semantic relationships that occur in the lexicon available in mathematics of students accessing the system. The analysis is done by means of graphs and through the main uses statisticians lexicostatistics.

## RESUMEN

El trabajo presenta la estructura del módulo de análisis de relaciones semánticas de un Hipermedio Adaptativo vía Internet, que permite describir y analizar las relaciones semánticas que se presentan en el léxico disponible en matemáticas de los alumnos que acceden al sistema. El análisis es realizado por medio de grafos y a través de los principales estadígrafos que utiliza el léxico-estadística.

## KEYWORDS

Relaciones semánticas, Léxico disponible, Disponibilidad léxica, Léxico Latente, Léxico y Matemáticas.

## INTRODUCCIÓN

El léxico es un “elemento clave en la comprensión y producción del conocimiento” [2]. Los estudiantes manejan cierto léxico que les permite comunicarse, pero no es suficiente para comprender de forma adecuada textos con temáticas específicas como los del ámbito escolar (por ejemplo un libro de matemáticas). Esto se debe al deficiente vocabulario de este tipo de textos y a su incidencia en los procesos de lectura y escritura. El léxico es parte esencial del conocimiento lingüístico y su “manejo instrumental pleno”, resulta fundamental en los niveles de aprendizaje [20]. Desde un marco psicológico más amplio [21], [9] y [14], el léxico es fundamental en el proceso de apropiación del saber, puesto que mejora el desempeño curricular de los estudiantes y es “la herramienta cognitiva” que les permite entrar en diferentes áreas del conocimiento.

Pastora Herrero [12], citando a Ausubel, afirma que el aprendizaje del vocabulario, está inserto en el aprendizaje de las representaciones. Para utilizar correctamente un vocablo, se debe comprender su concepto y en qué contexto está. Los vocablos pueden tener varios significados según la combinación morfosintáctica de la proposición u oración donde se encuentre. Hay una estrecha relación entre los tres estadios de aprendizaje que se inicia con las representaciones, a lo que sigue la significación de ideas y termina con los conceptos [13]. Para el alumno será una ventaja conocer más

palabras, porque podrá utilizarlas en los marcos en que sea adecuada. Es importante contextualizar el léxico para obtener su significado correcto en el texto.

El mundo aparece representado en el lenguaje por el léxico, el léxico es “el módulo lingüístico que relaciona más directamente el conocimiento del mundo y el saber lingüístico” [8]. Un concepto es “una configuración de conocimientos que puede activarse o recuperarse de una manera más o menos consistente y homogénea” [1]. El significado del léxico es la suma de sus usos posibles.

La léxico-estadística es la ciencia que se encarga de contabilizar y dar a conocer el uso real del lenguaje (tanto oral como escrito) en ciertas temáticas, en un grupo común de hablantes, o bien, de los hablantes pertenecientes a una región geográfica determinada [5]. Entonces, por medio de esta ciencia es posible saber, medir, conocer el léxico de cierto grupo de personas. Al principio la léxico-estadística obtenía la frecuencia de las palabras extraídas de textos con no más de mil lexias, luego se diferenciaba entre palabra y vocablo (palabras diferentes), y las listas de frecuencia obtenidas daban a conocer los vocablos que se utilizaban, cuál era el que más se ocupaba y cuál era el menos usado. Se pueden distinguir dos tipos de léxico, que juntos forman el léxico fundamental de una lengua; el léxico básico: formado por las palabras que más se utilizan cotidianamente, y el léxico disponible: formado por las palabras, que aunque no se utilicen con frecuencia, se recuerdan y utilizan de acuerdo al tema específico que se esté tratando.

Fue Michea [10] el primero en separar palabras frecuentes (ó atemáticas) de palabras disponibles (ó temáticas). Las palabras frecuentes se pueden encontrar en cualquier texto con un número moderado de páginas y sin importar su contenido, como adjetivos, verbos y sustantivos comunes (ó nombres muy generales); en cambio las palabras disponibles se relacionan con cierto tema y son, en su mayoría, palabras concretas.

El léxico disponible se obtiene a través de encuestas, donde por un estímulo se intenta que el informante actualice su lexicón mental, que según Emmorey y Fromkin [4] es el “componente de la gramática que contiene información de las palabras necesarias para el hablante”, esta información de las palabras se ha obtenido mediante información fonológica, morfológica, sintáctica y semántica (significado o estructura conceptual, [7]). Según Hall [6] las palabras en el lexicón mental se adquieren y/o retienen en base a su pronunciación, ortografía, marco sintáctico y concepto ó significado de la palabra.

## LECTURA Y APRENDIZAJE DE MATEMATICAS

La matemática como toda ciencia tiene su lenguaje particular, donde cada vocablo debe corresponderse con un acto comunicativo para evitar errores en la comunicación interpersonal. Diversos autores [11], [19], [15], [3], [17] se han referido a la importancia que tiene, para profesores y estudiantes, dominar la terminología para construir significados matemáticos. En Matemáticas se confunden los significados de las palabras, lo que impide comprender conceptos, se necesitan analizar significados e interpretaciones de las palabras. Según Reyna y Roque [16], hay tres categorías de palabras usadas en la enseñanza de las Matemáticas:

Categoría 1: palabras técnicas (o específicas del área matemática)

Categoría 2: palabras del área matemática, que también están en el lenguaje cotidiano, pero que no tienen siempre los mismos significados.

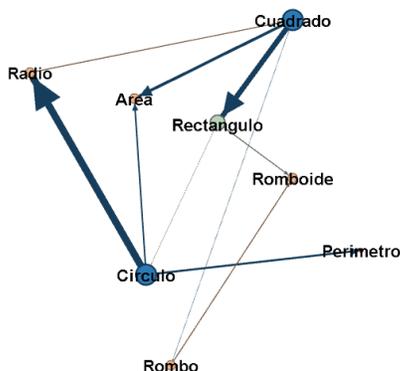
Categoría 3: palabras que tienen significados iguales o muy próximos en ambos contextos.

## GRAFOS

Un grafo es una representación gráfica de un conjunto de objetos o puntos, que se conocen como nodos o vértices, los cuales se encuentran unidos a través de líneas que reciben el nombre de enlaces o aristas.

Un grafo es un par  $G = (V, E)$ , donde  $V$  es un conjunto de objetos, llamados nodos o vértices, y  $E$  es un conjunto de pares de nodos, llamados enlaces o aristas. En la Figura 1 se presenta un ejemplo, con 8 nodos y 10 enlaces, donde se pueden observar los vértices con más enlaces en un tamaño mayor, y las aristas con más grosor representando a los enlaces entre vértices que tienen más frecuencia de ser citados.

Los Grafos se pueden usar para estudiar las interrelaciones entre unidades que interactúan unas con otras, así modelar, estudiar y optimizar muchos tipos de redes y sistemas, por ejemplo: redes de routers en internet, carreteras que conectan ciudades, redes y circuitos eléctricos, redes de alcantarillados, manejo de proyectos complejos, etc.



**Figura 1.** Ejemplo de un grafo de palabras etiquetado con 8 nodos o vértices y 10 enlaces o aristas

En un grafo podemos encontrar vecindades (neighbors) y agrupaciones (clusters), las cuales expresan valores

semánticos tanto de unidades (palabras) como de conjuntos (categorías).

En este trabajo el módulo semántico busca proporcionar suficiente información al usuario, para realizar un análisis lo más exacto posible de la información semántica que se puede estar dando en una población de sujetos. Es así como el módulo permite presentar conceptos (agrupaciones) que son formadas a través de los vértices con la mayor cantidad de enlaces, para lo cual esconde las aristas irrelevantes. O bien al posicionar el puntero sobre los nodos con más peso (mayor tamaño).

## MODELO GENERAL DEL HA AL LEXICO MATEMATICO

Este trabajo describe la estructura del módulo de análisis de relaciones semánticas de un Hipermedio Adaptativo vía Internet [18], que permite describir y analizar las relaciones semánticas que se presentan en el léxico disponible en matemáticas de los alumnos que acceden al sistema. El análisis es realizado por medio de grafos y a través de los principales estadígrafos que utiliza la léxico-estadística.

El Hipermedio Adaptativo está desarrollado en función del modelo que se presenta en la figura 2, el cual cuenta con 4 componentes bien definidos, los cuales interactúan para presentar al usuario un hipermedio según sus necesidades léxicas.

- Componente Modelo del Alumno: compuesto por las bases de datos encargadas de mantener la disponibilidad léxica de cada alumno, tanto en los centros de interés generales como en los específicos.
- Componente Modelo del Dominio: compuesto por las bases de datos asociadas a los medios y actividades apropiadas, según la propuesta didáctica, para el aumento del léxico disponible en los distintos centros de interés, además del léxico ideal que se obtendrá de la aplicación de la misma encuesta a los docentes.
- Componente del modelo experto: compuesto por todas las reglas necesarias para determinar el léxico general del alumno y el léxico no disponible. Además del motor de inferencia, el cual es el encargado de extraer las actividades apropiadas en función del modelo del alumno y del modelo del dominio.
- Interfaz: a través de este componente se recopila el léxico, se presentan las actividades personalizadas y se emiten informes. Entre los informes se encuentran los estadígrafos (de la léxico estadística) y los modelos de Relaciones Semánticas, que es el módulo que a continuación.

## MODULO DE RELACIONES SEMANTICAS

Utilizando un test de disponibilidad Léxica online, el que junto a algunos antecedentes personales, en dos minutos permite determina el léxico en cuatro centros específicos (álgebra, geometría, números y datos y azar). El módulo de análisis semántico y las relaciones de secuencia de las palabras disponibles, se presentan a través de este módulo el cual genera gráficamente los grafos, representando las palabras por nodos y las relaciones por las aristas. Los grafos se interpretan entonces como redes cuya configuración

expresa las relaciones semánticas subyacentes (como se explicó anteriormente). El software además permite determinar diversos índices que ayudan a realizar un análisis cualitativo más acabado de las relaciones encontradas en las redes de conocimiento.

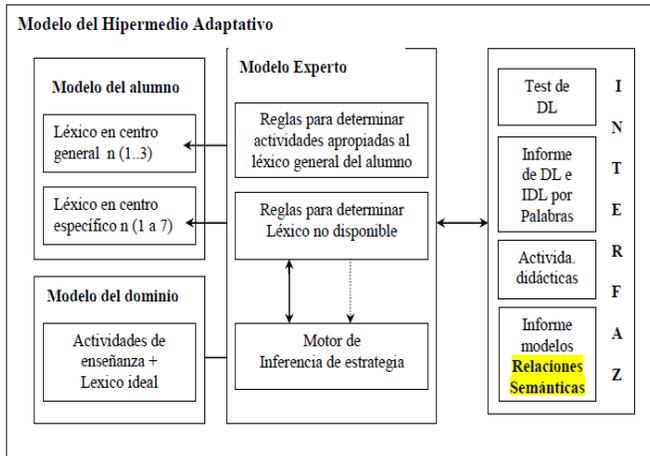


Figura 2. Módulo de relaciones semánticas en el modelo del HA

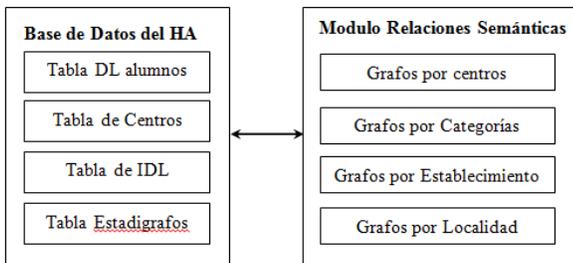


Figura 3. Módulo relaciones semánticas

### Implementación del módulo de Relaciones Semánticas

Para el desarrollo del módulo se ha utilizado el Script Gexf.js desarrollado en el 2011 por Brandon Aaron (<http://brandaaron.net>) y abierto para su utilización con fines académicos, el cual utiliza la tecnología de Gephi para presentar los grafos.

Gephi es una herramienta para la exploración, navegación y análisis de grafos. Permite a los usuarios interactuar con las distintas representaciones, manipular las estructuras, formas y colores que revelan propiedades ocultas. El objetivo es ayudar a los analistas de datos a hacer hipótesis, descubrir patrones, aislar singularidades en las estructuras o encontrar fallas en los datos.

Gephi se destaca por ser una herramienta libre de código abierto y que corre tanto en Windows, Linux como Mac, puede ser descargado desde el sitio <http://gephi.org>. Esta desarrollado en JAVA y se distribuye bajo licencia GNU GPL 3. Resulta ideal para desplegar gráficos representados mediante grafos, complejos gráficos de visualización de datos utilizados en análisis de redes sociales, o jerarquía de datos. Soporta la representación de grafos dirigidos, no dirigidos y mixtos. Uno de los aspectos más importantes cubiertos por Gephi es la interacción en tiempo real, permite modificar propiedades de los nodos y arcos al mismo tiempo que se

modifica la representación o layout del grafo y ofrecérselas al usuario sin largas esperas. Así mismo permite realizar agrupaciones, filtrado, manipulación, navegación y proveer un fácil acceso a los datos.

Gephi dispone del código fuente para su utilización y de una API denominada Gephi Toolkit para desarrollar aplicaciones propias basadas en dicha herramienta.

Gephi admite múltiples formatos de entrada de datos: el suyo propio GEFX (similar XML), GDF, GML, GraphML, Pajek NET, GraphViz DOT, CSV, UCINET DL, Tulip TPL, Netdraw VNA, Hoja de cálculo. Los formatos de salida son SVG, PNG y PDF.

En este trabajo se han tratado los datos utilizando el formato GEFX, el cual es generado a través de un componente en PHP que extrae el léxico desde la BD del modelo del alumno, para determinar los estadígrafos y añadirle los datos necesarios para visualizarlos en XML.

El siguiente código es un ejemplo del formato que asigna el módulo en PHP a los datos:

```
<?xml version="1.0" encoding="UTF-8"?>
<graph defaultedgetype="undirected"
  timeformat="double" mode="dynamic">
  <attributes class="node" mode="static">
    <attribute id="modularity_class" title="Modularity Class"
      type="integer">
      <default>0</default>
    </attribute>
  </attributes>
  <nodes>
    <node id="0.0" label="Rectangulo">
      <attvalues>
        <attvalue for="modularity_class" value="1">
          </attvalue>
        </attvalues>
      </node>
    <edges>
      <edge source="1.0" target="0.0">
        <attvalues>
          <attvalue for="weight" value="1.0"></attvalue>
        </attvalues>
      </edge>
    </edges>
  </graph>
</gexf>
```

En el código es posible identificar los nodos, con los ID que se le asignan y las aristas (edges) que indican el origen y destino de cada nodo según los ID asignados anteriormente, a lo que se le agrega el peso correspondiente a las veces que dos nodos son citados según la secuencia de palabras dadas por los alumnos.

En la figura 4 podemos entonces observar, un ejemplo de un grafo generado por el sistema, siendo en este caso para el centro específico de Geometría, para alumnos de un colegio municipal. En el grafo se ha identificado en distintos colores las mayores asociaciones y a mismo tiempo se ha utilizado un algoritmo proporcionado por Gephi para redistribución, lo que permite visualizar las mayores asociaciones.

