

**UNIVERSIDAD DE SANTIAGO DE CHILE**  
**FACULTAD DE INGENIERÍA**  
**Departamento de Ingeniería Informática**



**MODELO DE PREDICCIÓN DEL DESEMPEÑO DE BÚSQUEDA  
DE INFORMACIÓN EN LÍNEA EN ESTUDIANTES DE  
EDUCACIÓN BÁSICA**

**Propuesta de Tesis**

Nombre: Gonzalo Javier Martinez Ramirez

R.U.T.: 18.045.598-1

Año Ingreso: 2010

Teléfono: (+56) 9 96112973

E-mail: gonzalo.martinez@usach.cl

Profesor patrocinador: Roberto Ignacio González Ibañez

Santiago – Chile

2017

## RESUMEN

Durante la última década, debido a los rápidos avances de las tecnologías de la información y comunicación ha aumentado la cantidad de recursos digitales en Internet, la diversidad de fuentes de información, y, además, se ha facilitado el acceso a estos. Asimismo, las búsquedas *web* han pasado a ser parte de las tareas comunes que realizan los estudiantes de los planteles educativos. Considerando la diversidad de fuentes de información y tipos de recursos en línea, resulta necesario desarrollar competencias informacionales durante el proceso de formación en los distintos niveles educativos (primaria, secundaria y universitaria).

En el marco del proyecto iFuCo (*Enhancing learning and teaching future competences of online inquiry in multiple domains*), formado por investigadores de Chile y Finlandia, el cual desea investigar y modelar los comportamientos y competencias de investigación en línea de estudiantes de enseñanza básica, se propone la construcción de un modelo de predicción del comportamiento de búsqueda de información en línea en estudiantes de educación básica el cual se vaya perfeccionando a través del registro de datos históricos y que de un feedback en tiempo real.

La investigación será guiada por la metodología KDD con el fin de descubrir patrones en los datos que permitan la creación de un modelo de predicción del comportamiento de búsqueda. Además, para apoyar el proceso de investigación, se desarrollará una plataforma que funcione como extensión de la plataforma NEURONE (*oNlinE inqUiry expeRimentatiON systEm*). La plataforma propuesta alimentará y perfeccionará el modelo de predicción y entregará predicciones en tiempo real. Esta plataforma se guiará bajo la metodología RAD (*Rapid Application Development*) la cual se orienta a un desarrollo iterativo e incremental para la rápida construcción de prototipos de *software*.

# TABLA DE CONTENIDOS

<b>Capítulo 1. Objetivos y alcances de la solución</b>	<b>1</b>
1.1 Objetivo general . . . . .	1
1.2 Objetivos específicos . . . . .	1
<b>Capítulo 2. Descripción del problema</b>	<b>2</b>
2.1 Motivación . . . . .	2
2.2 Revisión de la literatura . . . . .	2
2.3 Definición del problema . . . . .	2
<b>Capítulo 3. Descripción de la solución propuesta</b>	<b>3</b>
3.1 Características de la solución . . . . .	3
3.2 Propósito de la solución . . . . .	3
3.3 Alcances y limitaciones de la solución . . . . .	3
<b>Capítulo 4. Metodología, herramientas y ambiente de desarrollo</b>	<b>4</b>
4.1 Metodología a usar . . . . .	4
4.2 Herramientas de desarrollo . . . . .	4
4.3 Ambiente . . . . .	4
<b>Capítulo 5. Plan de trabajo</b>	<b>5</b>
<b>Apéndice A. Capítulo Apéndice</b>	<b>6</b>
A.1 Sección del apéndice . . . . .	6
A.1.1 Subseccion del apéndice . . . . .	6
<b>Apéndice B. Another Appendix Chapter</b>	<b>7</b>

## ÍNDICE DE TABLAS

B.1. Ejemplo de una tabla. . . . .	7
------------------------------------	---

## ÍNDICE DE FIGURAS

A.1. A scientific diagram using the <code>pgfplots</code> package by Christian Feuersaenger using the same colors which are also used for the layout . . . . .	6
---	---

# **CAPÍTULO 1. OBJETIVOS Y ALCANCES DE LA SOLUCIÓN**

## **1.1. OBJETIVO GENERAL**

Diseñar y evaluar un modelo de predicción en tiempo real del resultado del comportamiento de búsqueda de información en línea de estudiantes de enseñanza básica.

## **1.2. OBJETIVOS ESPECÍFICOS**

1. Realizar una revisión bibliográfica sobre trabajos recientes relacionados con minería de datos en el contexto educacional.
2. Realizar una exploración, limpieza, pre-procesamiento y transformación de los datos recopilados por la plataforma NEURONE (oNlinE inqUiry expeRimentatiON systEm, NEURONE desde ahora en adelante).
3. Definir las características de comportamiento de búsqueda de los estudiantes para la construcción de modelos predicción.
4. Comparar, seleccionar e implementar algoritmos de minería de datos, para la construcción de modelos de predicción.

## **CAPÍTULO 2. DESCRIPCIÓN DEL PROBLEMA**

### **2.1. MOTIVACIÓN**

### **2.2. REVISIÓN DE LA LITERATURA**

### **2.3. DEFINICIÓN DEL PROBLEMA**

El proceso de búsqueda de información involucra diferentes procesos cognitivos, habilidades, variables de comportamiento y entorno de una persona. En el contexto de la enseñanza de la alfabetización informacional, las evaluaciones de los cursos se centran principalmente en los resultados de los estudiantes, sin tomar en cuenta el proceso formativo y factores asociados que podrían influir directa o indirectamente sobre los resultados finales y el desempeño de los alumnos.

A partir de lo señalado anteriormente, surgen las siguientes interrogantes (*research questions*, RQ desde ahora en adelante):

**RQ 1 :**

**RQ 2 :**

## **CAPÍTULO 3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA**

### **3.1. CARACTERÍSTICAS DE LA SOLUCIÓN**

### **3.2. PROPÓSITO DE LA SOLUCIÓN**

### **3.3. ALCANCES Y LIMITACIONES DE LA SOLUCIÓN**

Los modelos se construyen a partir de un conjunto de datos específicos, estos datos tienen su propio contexto y origen que limitan la generalización de los modelos a construir. A continuación, se describen las principales limitaciones y alcances de la solución.

1. El curso de alfabetización informacional y sus respectivos registros de datos, pertenecen al proyecto iFuCo [24], el cual es un trabajo colaborativo entre universidades de Finlandia (University of Tampere, University of Jyväskylä y University of Turku) y de Chile (Universidad de Santiago de Chile y Pontificia Universidad Católica de Chile).
2. Los registros de datos provienen de un estudio enmarcado en un curso de alfabetización en información, aplicado al área de Ciencia y Ciencias Sociales, en ambos países.
3. Los datos son recolectados y almacenados por un sistema externo llamado NEURONE (*oNlinE inqUiry expeRimentatiON systEm*), trabajo de memoria de un estudiante de la carrera de Ingeniería de Ejecución en Computación e Informática de la Universidad de Santiago de Chile [25].
4. La solución funciona como un sistema predictor del desempeño de estudiante en la búsqueda de información, sin ofrecer acciones correctivas en caso de bajo desempeño.



## **CAPÍTULO 4. METODOLOGÍA, HERRAMIENTAS Y AMBIENTE DE DESARROLLO**

### **4.1. METODOLOGÍA A USAR**

El presente proyecto presenta una componente de investigación y desarrollo de *software* (I+D). Esto debido a la relación que existe entre ambas componentes, la investigación necesita una herramienta de *software* de apoyo que permita recibir los datos de NEURONE, alimentar el modelo de predicción y que permita al usuario interactuar con resultados de la predicción a realizar.

### **4.2. HERRAMIENTAS DE DESARROLLO**

Las herramientas a utilizar en el trabajo de tesis, se dividen tanto en *hardware* como en *software*.

### **4.3. AMBIENTE**

El ambiente de desarrollo del presente proyecto será tanto en el domicilio particular del candidato a tesista, como también en el Departamento de Ingeniería Informática de la Universidad de Santiago de Chile, específicamente, en el laboratorio de sistemas colaborativos.

Después de finalizar cada iteración, la retroalimentación a este proyecto es ofrecida por miembros del equipo de investigación del proyecto iFuCo y el profesor guía, quien, además brinda apoyo en los aspectos tecnológicos y metodológicos. Finalmente, el equipo de desarrollo de este trabajo de título es unipersonal, con colaboración en fundamentos teóricos de otros tesisas y memoristas involucrados en el proyecto.

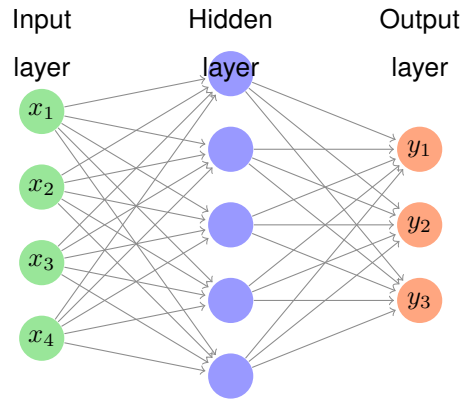
## **CAPÍTULO 5. PLAN DE TRABAJO**

El presente proyecto contempla 612 horas de trabajo efectivas y se realizará en el transcurso del segundo semestre del año 2017, el cual se inicia el 7 de agosto y termina el 7 de diciembre del presente año contemplando 16 semanas de trabajo. Se dispone como día de trabajo todos los días hábiles de la semana, desde las 9:00 hrs. AM hasta las 18:00 hrs. PM considerando una hora de descanso.

El plan de trabajo propuesto se muestra en la Tabla ??, dada las metodologías empleadas, las actividades se realizan de forma secuencial. Cabe destacar que el alumno candidato a tesista a la fecha de entrega del presente informe, ya ha avanzado el estado del arte e investigación de tecnologías.

## APÉNDICE A. CAPÍTULO APÉNDICE

### A.1. SECCIÓN DEL APÉNDICE



**Example Diagram with a Line Break in the Title  
(using the `text width` option in the title style)**

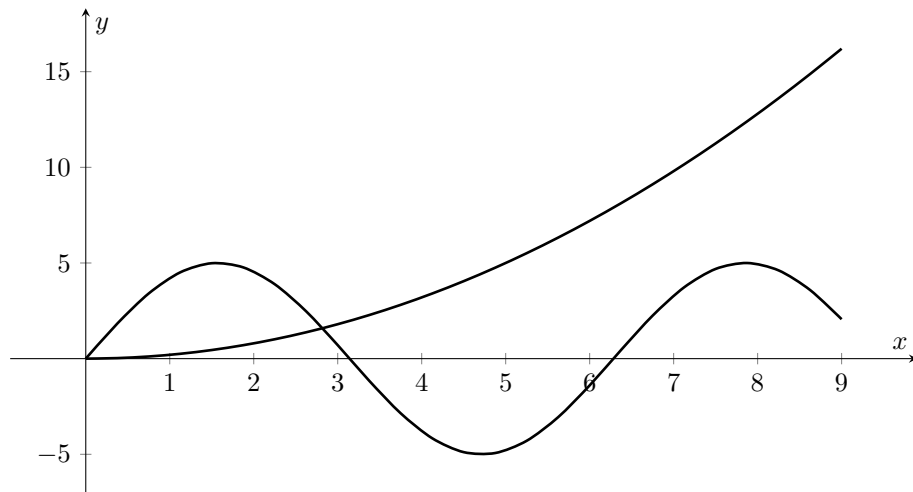


Figura A.1: A scientific diagram using the `pgfplots` package by Christian Feuersaenger using the same colors which are also used for the layout

Fuente: Elaboración propia, (2017).

#### A.1.1. Subseccion del apéndice

## APÉNDICE B. ANOTHER APPENDIX CHAPTER

Como se puede apreciar en la Tabla B.1.

Tabla B.1: Ejemplo de una tabla.  
Fuente: Elaboración propia, (2017).

header1	header2	header3
1	2	3
4	5	6
7	8	9